Київський національний університет
імені Тараса Шевченка
Факультет радіофізики, електроніки
та комп'ютерних систем
Кафедра радіотехніки та радіоелектронних систем

І.О. Анісімов, С.М. Левитський

Сигнали та процеси в радіотехніці

Підручник

Київ 2025

ЗМІСТ

Передмова	12
Вступ	13
В.1. Поняття про радіотехніку (13). В.2. Піддіапазони радіохвиль та їхнє застосування (14). В.3. Тракт передавання-приймання сигналу та його елементи (15). В.4. Короткий нарис з історії розвитку радіотехніки (17). В.5. Структура курсу (18). Контрольні питання до Вступу Література	 19 20
Частина I. Радіосигнали та методи аналізу лінійних кіл	21
Розділ 1. Регулярні сигнали в радіотехніці: часове та спектральне представлення	21
1.1. Модульовані сигнали 1.1.1. Амплітудно-модульовані сигнали (22). 1.1.2. Відеоімпульси та радіоімпульси (22). 1.1.3. Частотна модуляція (24). 1.1.4. Фазова модуляція (25). Контрольні питання до підрозділу 1.1	21 26
1.2. Ряди Фур'є 1.2.1. Повна ортонормована система функцій (27). 1.2.2. Ортогональність гармонік (27). 1.2.3. Ряд Фур'є (28). 1.2.4. Альтернативні форми запису ряду Фур'є (29). 1.2.5. Середня потужність періодичного сигналу (31). Контрольні питання до підрозділу 1.2 Приклади розв'язання задач до підрозділу 1.2 Задачі до підрозділу 1.2	27 31 31 34
1.3. Інтеграли Фур'є 1.3.1. Інтеграл Фур'є (35). 1.3.2. Властивості спектрів аперіодичних функцій (35). 1.3.3. δ-функція Дірака і функція Хевісайда (36). 1.3.4. Енергетичний спектр. Формули Парсеваля і Рейлі (38). 1.3.5. Спектр потужності (39). 1.3.6. Обговорення поняття спектру (40). Контрольні питання до підрозділу 1.3 Приклади розв'язання задач до підрозділу 1.3 Задачі до підрозділу 1.3	35 41 41 44
1.4. Модульовані сигнали: спектри та векторні діаграми 1.4.1. Спектр амплітудно-модульованого сигналу з гармонічною модуляцією (45). 1.4.2. Векторна діаграма сигналу з гармонічною амплітудною модуляцією (46). 1.4.3. Спектр амплітудно-модульованого сигналу з повільною обвідною (47). 1.4.4. Співвідношення між спектрами відеоімпульсу та радіоімпульсу (48). 1.4.5. Спектр фазово-модульованого сигналу з гармонічною модуляцією (48). 1.4.6. Векторна діаграма фазово-модульованого сигналу (50). Контрольні питання до підрозділу 1.4 Приклади розв'язання задач до підрозділу 1.4 Задачі до підрозділу 1.4	45 51 51 52
1.5. Узагальнення поняття спектру 1.5.1. Віконне перетворення Фур'є (53). 1.5.2. Віконне перетворення Фур'є для квазімонохроматичного сигналу (54). 1.5.3. Приклади віконних функцій (56). 1.5.4. Поняття про вейвлет-перетворення (57). 1.5.5. Головні ознаки вейвлету (57). 1.5.6. Приклади вейвлет-спектрів (58). Контрольні питання до підрозділу 1.5 Приклади розв'язання задач до підрозділу 1.5 Задачі до підрозділу 1.5	53 60 60 62

Розділ 2. Методи аналізу лінійних електричних кіл	63
2.1. Лінійні кола та лінійні чотириполюсники 2.1.1. Поняття про лінійні електричні кола (63). 2.1.2. Падіння напруги на резисторі, конденсаторі та котушці індуктивності (64). 2.1.3. Лінійний чотириполюсник (64). Контрольні питання до підрозділу 2.1	63 66
2.2. Лінійні електричні кола, збуджувані гармонічним джерелом 2.2.1. Метод комплексних амплітуд (68). 2.2.2. Закон Ома для пасивних кіл гармонічного струму (68). 2.2.3. Передавальна функція (коефіцієнт передачі) (69). Контрольні питання до підрозділу 2.2 Приклади розв'язання задач до підрозділу 2.2 Задачі до підрозділу 2.2	67 71 72 74
2.3. Лінійні електричні кола, збуджувані довільним джерелом 2.3.1. Метод передавальної функції (75). 2.3.2. Метод перехідної функції (78). 2.3.3. Метод імпульсної функції (81). 2.3.4. Зв'язок між імпульсною, перехідною та передавальною функцією (85). 2.3.5. Врахування ненульових початкових умов (87). 2.3.6. Заміна ненульових початкових умов еквівалентною зовнішньою силою (88). Контрольні питання до підрозділу 2.3 Приклади розв'язання задач до підрозділу 2.3 Задачі до підрозділу 2.3	75 89 90 92
2.4. Операторний метод аналізу лінійних кіл за довільних початкових умов 2.4.1. Інтегральне перетворення Лапласа (93). 2.4.2. Властивості перетворення Лапласа (94). 2.4.3. Операторний метод розв'язування лінійних диференціальних рівнянь (95). 2.4.4. Теорема розкладення Хевісайда (96). 2.4.5. Аналіз перехідних процесів операторним методом (97). Контрольні питання до підрозділу 2.4 Приклади розв'язання задач до підрозділу 2.4 Задачі до підрозділу 2.4	93 99 99 101
Частина II. Елементи радіоелектроніки	103
Розділ 3. Прилади напівпровідникової електроніки	103
3.1. Етапи розвитку елементної бази радіоелектроніки Контрольні питання до підрозділу 3.1	103 105
3.2. Елементи фізики напівпровідників 3.2.1. Механізми протікання струму в напівпровідниках (106). 3.2.2. *p-n* перехід та його вольт-амперна характеристика (106). 3.2.3. Ємність *p-n* переходу (108). 3.2.4. Пробій *p-n* переходу (110). Контрольні питання до підрозділу 3.2 Задачі до підрозділу 3.2	106 110 111
3.3. Напівпровідникові діоди 3.3.1. Випрямні діоди (112). 3.3.2. Діоди Шотткі (112). 3.3.3. Стабілітрони і стабістори (113). 3.3.4. Варикап (114). 3.3.5. Тунельний діод (115). 3.3.6. Умовні позначення діодів (116). Контрольні питання до підрозділу 3.3 Задачі до підрозділу 3.3	112 116 116
3.4. Польові транзистори 3.4.1. Польовий транзистор з керувальним *p-n* переходом (117). 3.4.2. Польові транзистори з ізольованим затвором (120). Контрольні питання до підрозділу 3.4	117 122

3.5. Біполярні транзистори: принцип роботи, характеристики та параметри 3.5.1. Будова та принцип роботи біполярного транзистора (123). 3.5.2. Характеристики біполярного транзистора, увімкненого за схемою зі спільним емітером (125). 3.5.3. Вплив температури на характеристики транзистора (129). 3.5.4. Біполярний транзистор як лінійний чотириполюсник (130). Контрольні питання до підрозділу 3.5 Приклади розв'язання задач до підрозділу 3.5 Задачі до підрозділу 3.5	123 132 132 133
3.6. Частотні властивості біполярного транзистора 3.6.1. Перехідні процеси в транзисторі (134). 3.6.2. Частотна залежність коефіцієнта передачі струму біполярного транзистора (135). 3.6.3. Шляхи поліпшення частотних властивостей транзисторів (137). 3.6.4. Фізична еквівалентна схема транзистора в області високих частот (138). 3.6.5. Зв'язок елементів схеми Джіаколетто з *h*-параметрами транзистора та його граничною частотою (139). Контрольні питання до підрозділу 3.6	134 142
Розділ 4. Лінійні пристрої аналогової радіоелектроніки	143
4.1. Пасивні фільтри 4.1.1. Визначення, класифікація та параметри фільтрів (144). 4.1.2. RC- та RL-ланцюжки як фільтри верхніх та нижніх частот (146). 4.1.3. Коливні контури як фільтри верхніх та нижніх частот (146). 4.1.4. Смугові та загороджувальні фільтри на базі RC- та RL-ланцюжків (147). 4.1.5. Смугові та загороджувальні фільтри на базі коливних контурів (150). Контрольні питання до підрозділу 4.1 Приклади розв'язання задач до підрозділу 4.1 Задачі до підрозділу 4.1	143 156 156 159
4.2. Підсилювачі на транзисторах: загальні питання 4.2.1. Підсилювач на транзисторі як керований подільник напруги (161). 4.2.2. Увімкнення навантаження до джерела гармонічного сигналу (162). 4.2.3. Основні параметри та різновиди підсилювачів (165). Контрольні питання до підрозділу 4.2	161 166
4.3. Зворотний зв'язок у підсилювачах 4.3.1. Послідовний зворотний зв'язок за напругою (167). 4.3.2. Позитивний і негативний зворотний зв'язок (167). 4.3.3. Вплив послідовного зворотного зв'язку за напругою на вхідний опір підсилювача (168). 4.3.4. Стабілізація коефіцієнту підсилення (168). 4.3.5. Вплив зворотного зв'язку на частотну характеристику (169). 4.3.6. Самозбудження підсилювача, охопленого позитивним зворотним зв'язком. Критерій Найквіста (169). 4.3.7. Паралельний зворотний зв'язок за напругою (171). Контрольні питання до підрозділу 4.3 Задачі до підрозділу 4.3	167 172 172
4.4. Підсилювачі на польових транзисторах 4.4.1. Схема підсилювача на МДН-транзисторі з вбудованим каналом (173). 4.4.2. Коефіцієнт підсилення за напругою (174). 4.4.3. Амплітудно-частотна характеристика підсилювача (175). 4.4.4. Схема підсилювача на МДН-транзисторі з індукованим каналом (178). 4.4.5. Схема підсилювача на транзисторі з керувальним *p-n* переходом (179). Контрольні питання до підрозділу 4.4 Приклади розв'язання задач до підрозділу 4.4 Задачі до підрозділу 4.4	173 180 180 181
4.5. Підсилювачі на біполярних транзисторах 4.5.1. Схема найпростішого підсилювача на біполярному транзисторі (182).	182

4.5.2. Параметри підсилювача на біполярному транзисторі (182). 4.5.3. Стабілізація режиму за допомогою негативного зворотного зв'язку (184). 4.5.4. Частотні властивості підсилювача на біполярному транзисторі (185). 4.5.4. Корекція частотної характеристики (189). Контрольні питання до підрозділу 4.5 Приклади розв'язання задач до підрозділу 4.5 Задачі до підрозділу 4.5	 192 192 195
4.6. Спеціальні типи підсилювачів на біполярних транзисторах 4.6.1. Підсилювач мінливої напруги (196). 4.6.2. Селективний підсилювач (198). 4.6.3. Емітерний повторювач (200). Контрольні питання до підрозділу 4.6 Задачі до підрозділу 4.6	196 202 202
4.7. Параметричні підсилювачі 4.7.1. Параметричний підсилювач: принцип роботи (203). 4.7.2. Одноконтурний параметричний підсилювач (205). 4.7.3. Форма амплітудно-частотної характеристики (207). 4.7.4. Пульсації коефіцієнту підсилення (208). 4.7.5. Двоконтурний параметричний підсилювач (209). Контрольні питання до підрозділу 4.7 Задачі до підрозділу 4.7	203 212 212
Розділ 5. Автогенератори та інші нелінійні аналогові схеми	213
5.1. Автогенератори на елементах із негативною диференціальною провідністю 5.1.1. Баланс потужності в колі з негативною диференціальною провідністю (213). 5.1.2. Увімкнення негативної диференціальної провідності паралельно до коливного контуру (214). 4.1.3. Реальні елементи з негативною диференціальною провідністю (216). Контрольні питання до підрозділу 5.1 Задачі до підрозділу 5.1	213 217 217
5.2. LC-автогенератори квазігармонічних коливань 5.2.1. Виведення нелінійного рівняння автогенератора Ван-дер-Поля (218). 5.2.2. Умови самозбудження (219). 5.2.3. Рівняння Рейлі: якісний аналіз розв'язку (219). 5.2.4. Квазілінійна теорія автогенератора (220). 5.2.5. Розв'язання рівняння Рейлі для режиму квазігармонічних коливань (221). 5.2.6. Поняття про фазовий портрет (223). 5.2.7. Фазовий портрет автогенератора Ван-дер-Поля у м'якому режимі (225). 5.2.8. Жорсткий режим самозбудження автогенератора (226). 5.2.9. Вимушена синхронізація автогенератора (227). 5.2.10. Реальні схеми LC-автогенераторів квазігармонічних коливань на біполярних транзисторах (228). Контрольні питання до підрозділу 5.2 Приклади розв'язання задач до підрозділу 5.2 Задачі до підрозділу 5.2	218 230 230 231
5.3. RC-автогенератори 5.3.1. Автогенератор з багатоланковим RC-фільтром (232). 5.3.2. Розрахунок триланкового RC-фільтру для автогенератора (233). 5.3.3. Особливості реальних схем RC-автогенераторів (234). Контрольні питання до підрозділу 5.3 Приклади розв'язання задач до підрозділу 5.3 Задачі до підрозділу 5.3	232 235 235 236
5.4. Автогенератор Ван-дер-Поля в режимі релаксаційних коливань 5.4.1. Умови реалізації релаксаційного режиму (237). 5.4.2. Аналіз релаксаційних коливань (237). 5.4.3. Інтерпретація релаксаційного режиму (238). 5.4.4. Період релаксаційних коливань (240).	237

Контрольні питання до підрозділу 5.4	240
5.5. Мультивібратор 5.5.1. Якісний розгляд роботи мультивібратора (241). 5.5.2. Кількісні оцінки параметрів мультивібратора (243). Контрольні питання до підрозділу 5.5 Приклади розв'язання задач до підрозділу 5.5 Задачі до підрозділу 5.5	241 244 244 245
5.6. Тригери 5.6.1. Схема тригера та його стаціонарні стани (246). 5.6.2. Запуск тригера (248). 5.6.3. Тригер Шмідта (249). Контрольні питання до підрозділу 5.6 Задачі до підрозділу 5.6	246 251 251
5.7. Проходження детермінованих сигналів через нелінійні кола 5.7.1. Спотворення гармонічного сигналу при проходженні через нелінійний елемент (252). 5.7.2. Помножувач частоти (253). 5.7.3. Детектування сигналу (254). 5.7.4. Інтермодуляційні спотворення (255). 5.7.5. Амплітудна модуляція (256). 5.7.6. Отримання фазової модуляції (258). 5.7.7. Перетворювач частоти (258). 5.7.8. Синхронний детектор (258). 5.7.9. Амплітудний обмежувач (259). Контрольні питання до підрозділу 5.7 Приклади розв'язання задач до підрозділу 5.7 Задачі до підрозділу 5.7	252 260 260 262
Частина III. Передавання сигналів на віддаль	263
Розділ 6. Лінії зв'язку	263
6.1. Елементи теорії хвиль 6.1.1. Двопровідна лінія. Телеграфні рівняння (264). 6.1.2. Розв'язок хвильового рівняння. Дисперсія (265). 6.1.3. Поширення хвильового пакету в лінії з дисперсією (266). Контрольні питання до підрозділу 6.1 Приклади розв'язання задач до підрозділу 6.1 Задачі до підрозділу 6.1	264 267 267 269
6.2. Ланцюжкові системи 6.2.1. Ланцюжкова система як частотний фільтр (270). 6.2.2. Хвильовий процес. Непрозорість (272). 6.2.3. Ланцюжкові системи як сповільнювальні структури (273). 6.2.4. Причини виникнення дисперсії (274). 6.2.5. Механізм формування смуг прозорості та непрозорості (275). Контрольні питання до підрозділу 6.2 Приклади розв'язання задач до підрозділу 6.2 Задачі до підрозділу 6.2	270 276 277 278
6.3. Розв'язок телеграфних рівнянь для двопровідної лінії 6.3.1. Розв'язок телеграфних рівнянь для лінії без втрат (279). 6.3.2. Розв'язок телеграфних рівнянь для лінії з втратами (279). 6.3.3. Погонні параметри двопровідної лінії та її фізичні властивості (280). Контрольні питання до підрозділу 6.3 Задачі до підрозділу 6.3	279 282 282
6.4. Режими навантаженої лінії передачі 6.4.1. Відбиття хвиль від навантаження (283). 6.4.2. Режим біжучих хвиль у лінії (284). 6.4.3. Режими стоячих хвиль (285). 6.4.4. Мішаний режим (287). Контрольні питання до підрозділу 6.4 Задачі до підрозділу 6.4	283 288 289

6.5. Вхідний опір ділянки лінії 6.5.1. Вхідний опір ділянки лінії з навантаженням (290). 6.5.2. Фізична інтерпретація вхідного опору лінії передачі (291). 6.5.3. Зв'язок хвилі в лінії з параметрами генератора на її вході (2921). Контрольні питання до підрозділу 6.5 Задачі до підрозділу 6.5	290 293 293
6.6. Узгодження лінії з навантаженням та джерелом сигналу 6.6.1. Узгодження лінії з навантаженням за допомогою чвертьхвильового трансформатора (294). 6.6.2. Узгодження лінії з навантаженням за допомогою шлейфа Татаринова (295). 6.6.3. Узгодження лінії з навантаженням за допомогою двох шлейфів Татаринова (296). 6.6.4. Узгодження джерела сигналу з лінією за допомогою трансформатора (296). Контрольні питання до підрозділу 6.6 Приклади розв'язання задач до підрозділу 6.6 Задачі до підрозділу 6.6	294 298 298 299
6.7. Відрізок довгої лінії як резонатор 6.7.1. Власні частоти і власні моди в загальному випадку (300). 6.7.2. Відрізок лінії з короткозамкненими кінцями (302). 6.7.3. Відрізок лінії з розімкнутими кінцями (303). 6.7.4. Відрізок лінії, у якого один кінець короткозамкнений, а другий розімкнутий (304). 6.7.5. Відрізок лінії з реактивними елементами на обох кінцях (304). 6.7.6. Вплив слабкої дисипації (305). Контрольні питання до підрозділу 6.7 Задачі до підрозділу 6.7	300 305 305
6.8. Хвилеводи 6.8.1. Диференціальні оператори (306). 6.8.2. Характеристики електромагнітного поля та рівняння Максвелла (307). 6.8.3. Хвильове та дисперсійне рівняння для електромагнітних хвиль у вакуумі (308). 6.8.4. Поляризація плоских електромагнітних хвиль у вакуумі (309). 6.8.5. Граничні умови для електромагнітного поля на поверхні ідеального провідника (311). 6.8.6. ТЕ- та ТМ-хвилі (312). 6.8.7. ТМ-моди прямокутного хвилеводу (313). 6.8.8. ТЕ-моди прямокутного хвилеводу (315). 6.8.9. Хвильовий опір хвилеводу. Узгодження хвилеводу з навантаженням (317). 6.8.10. Моди циліндричного хвилеводу (317). Контрольні питання до підрозділу 6.8 Задачі до підрозділу 6.8	306 319 319
Розділ 7. Антени	320
7.1. Випромінювання електромагнітних хвиль системою зарядів, що рухаються 7.1.1. Отримання хвильового рівняння для електромагнітних хвиль у вакуумі за наявності джерел (320). 7.1.2. Поле випромінювання системи зарядів, що рухаються (322). 7.1.3. Дипольне, магнітодипольне та квадрупольне випромінювання (323). Контрольні питання до підрозділу 7.1 Приклади розв'язання задач до підрозділу 7.1 Задачі до підрозділу 7.1	320 325 325 326
7.2. Диполь Герца 7.2.1. Елементарний електричний випромінювач: опис моделі (327). 7.2.2. Поле випромінювання диполя Герца (327). 7.2.3. Близька та далека зони випромінювання (329). 7.2.4. Диполь Герца як антена (331). 7.2.5. Магнітний диполь Герца (332). Контрольні питання до підрозділу 7.2 Задачі до підрозділу 7.2	327 333 333

7.3. Півхвильовий вібратор 7.3.1. Півхвильовий вібратор як резонатор (335). 7.3.2. Параметри півхвильового вібратора (336). 7.3.3. Вплив провідної поверхні на діаграму спрямованості (337). Контрольні питання до підрозділу 7.3 Задачі до підрозділу 7.3	335 340 340
7.4. Основні параметри та характеристики антен 7.4.1. Характеристика спрямованості передавальної антени (341). 7.4.2. Фазова та поляризаційна характеристики (343). 7.4.3. Коефіцієнт спрямованої дії та коефіцієнт підсилення (345). 7.4.4. Інші параметри передавальних антен (346). 7.4.5. Теорема взаємності (347). 7.4.6. Характеристики та параметри приймальних антен (351). Контрольні питання до підрозділу 7.4 Задачі до підрозділу 7.4	341 352 352
7.5. Основні типи антен 7.5.1. Принципи класифікації антен (353). 7.5.2. Прості випромінювачі (353). 7.5.3. Групові випромінювачі (356). 7.5.4. Випромінювальні структури (359). 7.5.5. Апертурні випромінювачі (361). Контрольні питання до підрозділу 7.5 Задачі до підрозділу 7.5	353 364 365
7.6. Антени різних частотних діапазонів 7.6.1. Випромінювальні антени довгих та середніх хвиль (366). 7.6.2. Приймальні антени довгих та середніх хвиль (368). 7.6.3. Особливості антен короткохвильового діапазону (372). 7.6.4. Квазіоптичні антени діапазону ультракоротких хвиль (376). 7.6.5. Хвилеводні та рупорні випромінювачі (379). 7.6.6. Щілинні випромінювачі (381). 7.6.7. Спіральні антени (382). Контрольні питання до підрозділу 7.6 Задачі до підрозділу 7.6	366 383 383
Розділ 8. Поширення радіохвиль у приземному просторі	384
8.1. Деякі питання теорії електромагнітних хвиль у середовищах 8.1.1. Дисперсійне рівняння для електромагнітних хвиль у середовищі (384). 8.1.2. Діелектрична проникність плазми (386). 8.1.3. Діелектрична проникність газу (390). Контрольні питання до підрозділу 8.1 Задачі до підрозділу 8.1	384 392 393
8.2. Електричні властивості земної атмосфери 8.2.1. Структура та властивості атмосфери (394). 8.2.2. Електричні властивості тропосфери та стратосфери (396). 8.2.3. Електричні властивості іоносфери (397). Геомагнітне поле (399) Контрольні питання до підрозділу 8.2 Задачі до підрозділу 8.2	394 400 400
8.3. Поширення радіохвиль у вільному просторі 8.3.1. Електричне поле, створене передавальною антеною в точці приймання сигналу (401). 8.3.2. Поняття про суттєву область (402). 8.3.3. Зони Френеля (402). 8.3.4. Розрахунок напруженості електричного поля в точці приймання сигналу (404). 8.3.5. Форма та розміри суттєвої області (405). Контрольні питання до підрозділу 8.3 Задачі до підрозділу 8.3	401 407 407

8.4. Збудження земної хвилі високо піднятою антеною 8.4.1. Способи розрахунку поля земної хвилі (408). 8.4.2. Розрахунок впливу Землі шляхом урахування відбитих від неї хвиль (409). 8.4.3. Суттєва область відбиття (410). 8.4.4. Поле в освітленій зоні без урахування кривини Землі (413). 8.4.5. Вплив кривини земної поверхні на поле в освітленій зоні (416). 8.4.6. Поле в зонах півтіні й тіні (417). Контрольні питання до підрозділу 8.4 Задачі до підрозділу 8.4	408 419 419
8.5. Поле низько розміщеного випромінювача 8.5.1. Опис моделі без урахування кривини Землі та поняття числової віддалі (420). 8.5.2. Результати розрахунку множника ослаблення сигналу без урахування кривини Землі (422). 8.5.3. Результати розрахунків поля антени з урахуванням кривини Землі (423). Контрольні питання до підрозділу 8.5 Приклади розв'язання задач до підрозділу 8.5 Задачі до підрозділу 8.5	420 424 424 425
8.6. Відбиття електромагнітних хвиль від слабконеоднорідної плазми 8.6.1 Хвильові рівняння для електромагнітних хвиль у плоскошаруватій плазмі (426). 8.6.2. Наближення геометричної оптики (427). 8.6.3. Інтерпретація геометрикооптичного розв'язку (429). 8.6.4. Розподілене відбиття хвилі в слабконеоднорідному середовищі (429). 8.6.5. Відбиття s-поляризованої хвилі від слабконеоднорідної плазми (430). 8.6.6. Відбиття поляризованої хвилі від розмитої межі плазми. Точка локального плазмового резонансу (432). Контрольні питання до підрозділу 8.6 Задачі до підрозділу 8.5	426 434 434
8.7. Поширення радіохвиль в атмосфері Землі 8.7.1. Деякі числові оцінки параметрів іоносфери (436). 8.7.2. Форма променю в приземній атмосфері (437). 8.7.3. Радіус кривини променю в іоносферній плазмі (439). 8.7.4. Траєкторії хвиль у тропосфері (441). 8.7.5. Траєкторії радіохвиль в іоносфері (442). 8.7.6. Поглинання та розсіювання радіохвиль в атмосфері (445). Контрольні питання до підрозділу 8.7 Задачі до підрозділу 8.7	436 447 448
Частина IV. Елементи статистичної радіотехніки	449
Розділ 9. Методи опису та аналізу випадкових процесів	449
9.1. Випадкові процеси та їхні характеристики 9.1.1. Випадкові величини, випадкові процеси та випадкові поля (449). 9.1.2. Густина розподілу ймовірності (451). 9.1.3. Статистичні середні (453). Контрольні питання до підрозділу 9.1 Задачі до підрозділу 9.1	449 454 454
9.2. Багатовимірні статистичні характеристики 9.2.1. Багатовимірні розподіли ймовірності (455). 9.2.2. Умовні ймовірності (456). 9.2.3. Кореляційні та автокореляційні функції (458). 9.2.4. Теорема Вінера – Хінчина (459). Контрольні питання до підрозділу 9.2 Задачі до підрозділу 9.2	455 460 460
9.3. Моделі випадкових процесів 9.3.1. Ергодичні процеси (462). 9.3.2. Нормальний (гауссівський) випадковий процес (465). 9.3.3. Білий шум (466). 9.3.4. Вузькосмуговий випадковий процес (467). 9.3.5. Випадкова імпульсна послідовність (468). 9.3.6. Дифузійний випадковий процес (472). 9.3.7. Коливання, модульовані шумом (475).	462

Контрольні питання до підрозділу 9.3 Задачі до підрозділу 9.2	481 482
9.4. Стохастичні диференціальні рівняння 9.4.1. Розв'язання стохастичних диференціальних рівнянь шляхом усереднення точних розв'язків (483). 9.4.2. Приклад: струм в RL-ланцюжку при надходженні на його вхід випадкового стрибка напруги (484). 9.4.3. Стохастичні методи: отримання рівняння для середніх (485). 9.4.3. Стохастичні методи: рівняння Фоккера – Планка (488). Контрольні питання до підрозділу 9.4 Задачі до підрозділу 9.4	483 491 492
Розділ 10. Шуми та завади в радіотехнічних колах	493
10.1. Шуми пасивних та активних елементів радіотехнічних кіл 10.1.1. Тепловий шум (493). 10.1.2. Дробовий шум (496). 10.1.3. Флікер-шум (498). 10.1.4. Шуми розподілу струмів (500). 10.1.5. Шуми генерації-рекомбінації (500). 10.1.6. Дифузійний шум (501). 10.1.7. Шуми напівпровідникових приладів у різних частотних діапазонах (501). Контрольні питання до підрозділу 10.1 Задачі до підрозділу 10.1	493 502 502
10.2. Флуктуації в автогенераторах 10.2.1. Технічна та природна ширина лінії автогенератора (503). 10.2.2. Рівняння для автогенератора Ван-дер-Поля з урахуванням власних шумів (504). 10.2.3. Вкорочені стохастичні рівняння для автоколивань з урахуванням шумів (505). 10.2.4. Спрощення вкорочених рівнянь (506). 10.2.5. Функція розподілу й моменти амплітуди автоколивань (509). 10.2.6. Спектр флуктуацій амплітуди автоколивань (511). 10.2.7. Флуктуації фази автогенератора та їхній спектр (512). 10.2.8. Форма та природна ширина спектральної лінії автогенератора (514). 10.2.9. Встановлення коливань автогенератора від рівня шумів: розподіл для миттєвої амплітуди (516). 10.2.10. Розподіл для часу встановлення сигналу автогенератора (518). Контрольні питання до підрозділу 10.2 Задачі до підрозділу 10.2	503 518 519
10.3. Генератор шуму Кияшка – Піковського – Рабиновича 10.3.1. Непередбачувана поведінка простих систем (520). 10.3.2. Генератор шуму КПР: схема та рівняння руху (521). 10.3.3. Якісний аналіз рівнянь руху в режимах регулярних коливань (522). 10.3.4. Фазовий портрет стохастичних коливань (523). 10.3.5. Біфуркації генератора КПР при зміні керувального параметра (524). Контрольні питання до підрозділу 10.3 Задачі до підрозділу 10.3	520 525 526
10.4. Поняття про електромагнітну сумісність 10.4.1. Нормативні документи, що визначають вимоги до електромагнітної сумісності (527). 10.4.2. Електромагнітні завади природного походження (528). 10.4.3. Техногенні завади (529). 10.4.4. Електромагнітна обстановка (530). 10.4.5. Радіопередавальні пристрої як джерела завад (531). 10.4.6. Небажані канали приймання супергетеродинних радіоприймачів (532). 10.4.7. Сполучні провідники як передавачі та приймачі завад (533). 10.4.8. Особливості приймальних антен, спрямовані на забезпечення електромагнітної сумісності (535). 10.4.10. Екранування та заземлення (537). 10.4.11. Протизавадні фільтри (540). Контрольні питання до підрозділу 10.4	527 540
Розділ 11. Відокремлення корисних сигналів від завад і шумів	541
11.1. Найпростіші моделі фільтрації сигналу за наявності шумів та завад	541

11.1.1. Лінійні фільтри (541). 11.1.2. Фільтрація монохроматичних сигналів (541). Контрольні питання до підрозділу 11.1 Задачі до підрозділу 11.1	 543 544
11.2. Фільтрація сигналів за наявності шуму: задача виявлення 11.2.1. Постановка задачі про оптимальний фільтр (545). 11.2.2. Виявлення сигналу на фоні білого шуму (545). 11.2.3. Зміна форми сигналу на виході оптимального фільтра (546). 11.2.4. Оптимальний фільтр для шумів з довільним спектром (547). Контрольні питання до підрозділу 11.2 Задачі до підрозділу 11.2	545 548 548
11.3. Фільтрація сигналів за наявності шуму: задача виділення 11.3.1. Постановка задачі (549). 11.3.2. Розрахунок передавальної функції (549). 11.3.3. Характерні режими фільтрації (551). 11.3.4. Рівняння Вінера – Хопфа (551). 11.3.5. Оптимальні фільтри й корелятори (553). Контрольні питання до підрозділу 11.3 Задачі до підрозділу 11.3	549 554 554
11.4. Радіометри 11.4.1. Компенсаційний радіометр: перенесення енергії сигналу в постійну складову (555). 11.4.2. Компенсаційний радіометр: розрахунок відношення сигнал-шум (556). 11.4.3. Модуляційний радіометр (558). 11.4.4. Чутливість радіометра (559). Контрольні питання до підрозділу 11.4 Задачі до підрозділу 11.4	555 559 560
Частина V. Елементи цифрової електроніки	561
Розділ 12. Поняття про цифрове передавання інформації	561
12.1. Дискретизація та квантування сигналів 12.1.1. Поняття про дискретизацію, квантування та оцифрування сигналів (562). 12.1.2. Формування імпульсної послідовності, що відповідає дискретизації за часом (563). 12.1.3. Теорема відліків Віттекера – Найквіста – Котельникова – Шеннона (564). 12.1.4. Відлікові функції (564). 12.1.5. Доведення теореми відліків (567). 12.1.6. Шуми квантування (567). 12.1.7. Основні види імпульсної модуляції (568). Контрольні питання до підрозділу 12.1 Задачі до підрозділу 12.1	562 569 569
12.2. Обсяг цифрових даних 12.2.1. Системи числення (570). 12.2.2. Двійкова система числення (572). 12.2.3. Кількісна міра обсягу цифрових даних (572). Контрольні питання до підрозділу 12.2 Задачі до підрозділу 12.1	570 574 574
12.3. Канали зв'язку та їхня пропускна спроможність 12.3.1. Системи передавання інформації (575). 12.3.2. Ущільнення каналів зв'язку (576). 12.3.3. Частотне ущільнення каналів (577). 12.3.4. Часове ущільнення каналів (578). 12.3.5. Характеристики джерел інформації та каналів зв'язку (579). 12.3.6. Теорема Шеннона для каналу без завад (580). 12.3.7. Вплив завад на канал зв'язку (581). 12.3.8. Пропускна спроможність каналу зв'язку (581). Контрольні питання до підрозділу 12.3 Задачі до підрозділу 12.3	575 583 583

Передмова

Курс «Сигнали та процеси в радіотехніці» викладається студентам спеціальності G5 «Електроніка, електронні комунікації, приладобудування та радіотехніка», чотирирічна бакалаврська програма «Радіоелектроніка та електронні комунікації». Це вибірковий курс блоку дисциплін «Програмно-апаратні засоби інформаційної безпеки». За актуальним навчальним планом він читається в 4-5 семестрах і включає 11 кредитів (98 год. лекцій, 42 год. практичних та 14 год. лабораторних занять).

По суті, дисципліна «Сигнали та процеси в радіотехніці» – це огляд класичної аналогової радіотехніки, включаючи лінії зв'язку, антени та поширення радіохвиль у приземній атмосфері, а також елементи статистичної радіотехніки. Він орієнтований на викладення найбільш загальних принципів радіотехніки (генерація, підсилення, модуляція, детектування, передавання, фільтрація сигналів). Тому в ньому розглядаються лише найпростіші радіотехнічні схеми з використанням елементної бази напівпровідникової електроніки. Складніші й сучасніші схеми студенти вивчають в інших курсах радіоелектронного спрямування, де конкретизуються, розвиваються й поглиблюються провідні ідеї даного курсу.

Ця дисципліна певною мірою може розглядатись і як вступ до проблематики, пов'язаної з технічним захистом інформації – саме тому вона, як уже вказувалося, входить до вибіркового блоку «Програмно-апаратні засоби інформаційної безпеки». Принаймні, без знання основних різновидів сигналів та процесів у радіотехніці важко зрозуміти як природу можливих каналів витоку інформації з радіотехнічних систем, так і методи придушення цих каналів.

При підготовці як загалом програми цього курсу, так і його окремих розділів я широко користувався ідеями та матеріалами (зокрема, підручниками та навчальними посібниками, вказаними в списку використаної літератури) мого вчителя – доктора фізико-математичних наук, професора, багаторічного завідувача кафедри радіоелектроніки радіофізичного факультету Київського університету імені Тараса Шевченка, прекрасного викладача й методиста Сергія Михайловича Левитського (1927-2018). Це дає мені підстави вважати його повноцінним співавтором даного підручника.

* * * * * * * *

В підручнику, як правило, наводяться детальні розрахунки. Вони призначені для тих, хто має потребу (або бажання) в них розібратися. Інші категорії читачів можуть не звертати на них надмірної уваги.

Для зручності сприйняття матеріалу в електронному вигляді деякі рисунки та формули в тексті при повторному посиланні дублюються.

Ігор Анісімов

Підручник рекомендований до друку вченою радою факультету радіофізики, електроніки та комп'ютерних систем Київського національного університету імені Тараса Шевченка, протокол №11 від 11.06.2026 р.

Вступ

В.1. Поняття про радіотехніку

Радіо (від лат. radio — випромінюю) — загальна назва методів безпровідного передавання та отримання інформації за допомогою електромагнітних хвиль радіочастотного діапазону. Передавання інформації здійснюють у системах радіозв'язку, радіомовлення, телебачення, радіокерування, активної радіолокації тощо. У системах пасивної радіолокації, пеленгації і подібних до них інформацію тільки отримують.

Радіотехніка – це галузь науки, що займається (і) застосуванням радіохвиль для передавання сигналів (іі) та розробленням принципів роботи пристроїв, які забезпечують цей процес, а також відповідна галузь техніки і технології.

Радіотехніку можна умовно поділити на такі частини:

- *радіопередавальні пристрої*, або радіопередавачі – це пристрої, які забезпечують формування радіочастотних сигналів, а саме їх генерування, модуляцію та підсилення;
- *антенна техніка*: передавальні антени перетворюють електричні сигнали на радіохвилі й випромінюють їх в ефір, приймальні – приймають радіохвилі та перетворюють їх на електричні коливання;
- *поширення радіохвиль* в атмосфері Землі та космічному просторі;
- *радіоприймальні пристрої*, або радіоприймачі – пристрої, призначені для підсилення радіосигналів, що надходять від приймальної антени та наступного перетворення їх (демодуляція, детектування тощо) до вигляду, потрібного споживачу;
- *телебачення*, або телевізія (від давньогрецького τῆλε – далеко і латинського visio – бачення) – передавання на велику відстань рухомих зображень зі звуковим супроводом у реальному часі;
- *радіолокація* – виявлення об'єктів (цілей) і визначення їхніх просторових координат та параметрів руху за допомогою радіотехнічних методів; призначені для цього пристрої називають радіолокаційними станціями, або радіолокаторами;
- *радіокерування* (дистанційне керування) – керування об'єктами і системами на відстані шляхом передачі до них каналами зв'язку сигналів для увімкнення відповідних виконавчих пристроїв: реле, вимикачів, контакторів, пускачів, вентилів, засувок тощо; частинним випадком радіокерування є радіонавігація, тобто керування рухом транспортних засобів (літальних апаратів, суден тощо) та наведення керованих об'єктів (наприклад, ракет, безпілотних літальних апаратів тощо) за допомогою радіотехнічних засобів;
- *радіоелектронна боротьба* (англ. electronic warfare) – це, по-перше, здобування інформації про розташування радіоелектронних засобів противника, його систем керування військами і зброєю з подальшим їхнім радіоелектронним придушенням або знищенням (виведенням з ладу) всіма видами зброї; по-друге – це захист власних радіоелектронних засобів і систем керування військами та зброєю від застосованих противником радіоелектронної розвідки й радіоелектронної протидії.

Радіотехніка як інструмент, що використовується в експериментальних дослідженнях, знайшла застосування в різних галузях науки, таких як фізика, астрономія, хімія, біологія, медицина тощо.

Радіотехніка тісно пов'язана, з одного боку, з радіофізикою[1], фізикою твердого тіла, електродинамікою, оптикою й статистичною фізикою, а з іншого – з електротехнікою, автоматикою, телемеханікою[2], обчислювальною технікою та інформатикою.

---

[1] Радіофізика – це галузь фізики, що досліджує взаємодію електромагнітних хвиль з речовиною.

[2] Телемеханіка – галузь науки і техніки, присвячена теорії і технічним засобам автоматичного передавання на відстань повідомлень про стан об'єктів та команд керування ними.

Аналогічним до терміну «радіотехніка», але ширшим і, в певному сенсі, більш сучасним є термін «радіоелектроніка».

Радіоелектроніка – це галузь науки і техніки, яка охоплює теорію, методи створення та використання пристроїв для передавання, приймання та перетворення інформації за допомогою електромагнітних сигналів. Як видно з цього означення, радіоелектроніка, на відміну від радіотехніки, розглядає передавання інформації не тільки за допомогою радіохвиль, але й з використанням інших типів електромагнітних сигналів та ліній їхньої передачі.

Сфера використання радіоелектроніки безперервно розширюється, поступово захоплюючи економіку, зокрема, промислове виробництво, сільське господарство, транспорт, а також медицину, побут та інші області людської діяльності.

Інколи радіоелектроніку ототожнюють з електронікою. Це не зовсім правильно.

Як наука, *електроніка* (від грец. ηλεκτρόνιο – електрон) вивчає взаємодію електронів з електромагнітними полями і принципи роботи електронних приладів і пристроїв, у яких ця взаємодія використовується для перетворення електромагнітних коливань та хвиль, в основному – для передавання, оброблення й зберігання інформації.

З науково-технічної точки зору електронікою називають галузь прикладної фізики й техніки, яка розробляє методи утворення пучків (потоків) електронів та/або інших заряджених частинок і керування їхнім рухом у різноманітних середовищах (вакуум, газ, плазма, тверде тіло) та на їхніх межах, а також шляхи створення електронних приладів різного призначення.

Електроніка загалом розбивається на вужчі області. В залежності від середовища, в якому рухаються заряджені частинки, розрізняють вакуумну електроніку, газову електроніку, плазмову електроніку, напівпровідникову електроніку. Ці області, а також емісійну електроніку, об’єднують терміном «фізична електроніка». В залежності від довжин і типів хвиль, що використовуються, виділяють радіоелектроніку, мікрохвильову електроніку, інфрачервону техніку[3], оптоелектроніку[4]. акустоелектроніку[5], спінхвильову електроніку. В залежності від підходів до створення приладів та ефектів, що використовуються, говорять про функціональну електроніку[6], квантову електроніку[7], хемотроніку[8], кріоелектроніку[9]. В залежності від характерних розмірів пристроїв виділяють мікроелектроніку та наноелектроніку[10].

---

[3] Інфрачервона техніка – галузь прикладної фізики і техніки, що включає розробку і вживання в наукових дослідженнях, на виробництві і у військовій справі приладів, дія яких заснована на використанні інфрачервоного випромінювання і його фізичних властивостей (прилади для спостереження й фотографування в темряві, для дистанційного вимірювання температури нагрітих тіл, для виявлення наземних, морських і повітряних цілей, для самонаведення на ціль снарядів і ракет тощо).

[4] Оптоелектроніка – розділ фізики й техніки, пов'язаний з перетворенням світлового випромінювання на електричний струм і навпаки.

[5] Акустоелектроніка – розділ електроніки, що займається дослідженням взаємодії акустичних хвиль з електромагнітними полями і електронами провідності в конденсованих середовищах та створенням пристроїв, що використовують ці ефекти.

[6] Функціональна електроніка базується на фізичній інтеграції, тобто на використанні фізичних явищ (згустків зарядів, хвиль, доменів) для реалізації складних функцій без поділу на окремі елементи (наприклад, резистори або транзистори).

[7] Квантова електроніка вивчає взаємодію електромагнітного випромінювання з речовиною на основі вимушених квантових переходів між енергетичними станами атомів і молекул та використання цих ефектів для генерації та підсилення світлових (лазери) чи радіохвиль (мазери).

[8] Хемотроніка – науково-технічний напрям, що займається розробленням, дослідженням та застосуванням приладів і пристроїв автоматики, вимірювальної та обчислювальної техніки, дія яких заснована на електрохімічних процесах та явищах, що мають місце на межі електрод — електроліт при пропусканні електричного струму.

[9] Кріогенна електроніка – напрямок електроніки, що охоплює дослідження при кріогенних температурах (нижче 120 К) специфічних ефектів взаємодії електромагнітного поля з носіями зарядів в твердому тілі і створення електронних приладів і пристроїв, що працюють на основі цих ефектів.

[10] Оперує з розмірами менше 100 нм, коли визначальну роль відіграють квантові ефекти.

З викладеного вище випливає, що під сигналами в радіотехніці розуміють різні типи електричних та електромагнітних коливань та хвиль радіочастотного діапазону, які використовуються для передавання та оброблення інформації. Відповідно радіотехнічні процеси – це процеси генерування, оброблення, передавання та приймання радіотехнічних сигналів.

В.2. Піддіапазони радіохвиль та їхнє застосування

Класифікація електромагнітних хвиль в залежності від їхньої довжини (частоти) подана в таблиці В.1. Нагадаємо, що межі діапазонів носять умовний характер.

Таблиця В.1. Шкала електромагнітних хвиль

Діапазон	**Довжина хвилі**	**Частота коливань, Гц**
Радіохвилі	1 мм – 10 км	$3\cdot10^{4}$ – $3\cdot10^{11}$
Інфрачервоне випромінювання	780 нм – 1 мм	$3\cdot10^{11}$ – $3.8\cdot10^{14}$
Оптичний діапазон	400 нм – 780 нм	$3.8\cdot10^{14}$ – $7.5\cdot10^{14}$
Ультрафіолетове випромінювання	10 нм – 400 нм	$7.5\cdot10^{14}$ – $3\cdot10^{16}$
Рентгенівське випромінювання (Х-промені)	10 фм – 10 нм	$3\cdot10^{16}$ – $3\cdot10^{20}$
γ- випромінювання	менше 10 фм	більше $3\cdot10^{20}$

За міжнародною класифікацією, до радіохвиль відносять хвилі з частотами від 30 кГц до 3 ТГц (довжини хвиль відповідно від 10 км до 10 мкм) – див. табл. В.2. Зазначимо, що визначення діапазону радіохвиль в табл. В.1 і в табл. В.2 дещо відрізняється.

Таблиця В.2. Піддіапазони радіохвиль та їхні застосування

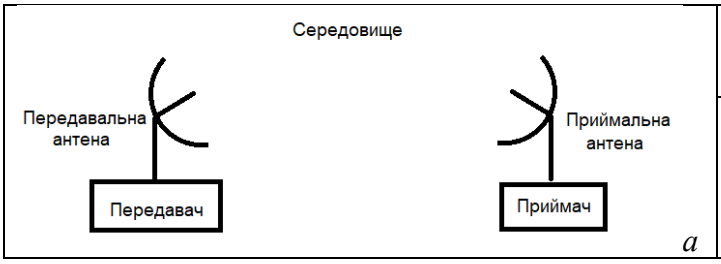
Піддіапазон	**Довжина хвилі, м**	**Частота коливань, Гц**	**Застосування**
Наддовгі хвилі	Більша 10^{4}	Менша $3\cdot10^{4}$	Зв'язок із підводними човнами, служба точного часу, геофізичні дослідження
Довгі хвилі	10^{4} — 10^{3}	$3\cdot10^{4}$ — $3\cdot10^{5}$	Радіомовлення, радіозв'язок земною (поверхневою) хвилею, радіонавігація
Середні хвилі	10^{3} — 10^{2}	$3\cdot10^{5}$ — $3\cdot10^{6}$	Радіомовлення, радіозв'язок земною хвилею та іоносферний
Короткі хвилі	10^{2} — 10	$3\cdot10^{6}$ — $3\cdot10^{7}$	Радіомовлення, іоносферний радіозв'язок, загоризонтна радіолокація, рації
Метрові хвилі	10 — 1	$3\cdot10^{7}$ — $3\cdot10^{8}$	Телебачення, радіомовлення, радіозв'язок тропосферний та прямою хвилею, рації, УВЧ-терапія
Дециметрові хвилі	1 — 10^{-1}	$3\cdot10^{8}$ — $3\cdot10^{9}$	Телебачення, радіомовлення, радіозв'язок тропосферний та прямою хвилею, рації, мобільні телефони, мікрохвильові печі, супутникова навігація
Сантиметрові хвилі	10^{-1} — 10^{-2}	$3\cdot10^{9}$ — $3\cdot10^{10}$	Радіолокація, інтернет, супутникове телебачення, супутниковий та радіозв'язок прямою хвилею, безпровідні комп'ютерні мережі
Міліметрові хвилі	10^{-2} — 10^{-3}	$3\cdot10^{10}$ — $3\cdot10^{11}$	Швидкісний релейний зв'язок, радіолокація (метеорологічна, управління озброєнням), медицина, супутниковий радіозв'язок
Субміліметрові хвилі	10^{-3} — 10^{-5}	$3\cdot10^{11}$ — $3\cdot10^{13}$	Реєстрація зображень у довгохвильових інфрачервоних променях

Мікрохвильове випромінювання, або надвисокочастотне (НВЧ) випромінювання (застарілий термін) – це електромагнітне випромінювання, що включає в себе дециметровий, сантиметровий і міліметровий діапазон радіохвиль.

Надалі в цьому курсі йтиметься переважно про діапазон частот, нижчих від мікрохвильового діапазону (умовно – до 3 ГГц).

В.3. Тракт передавання-приймання сигналу та його елементи

Радіотехніка історично розвивалася навколо задачі про передавання інформації. Блок-схема каналу зв'язку через ефір та через лінію зв'язку подана на рис. В.1.



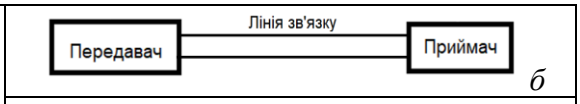


Рис. В.1. Схема тракту передавання – приймання сигналу: а – через ефір; б – через лінію зв'язку

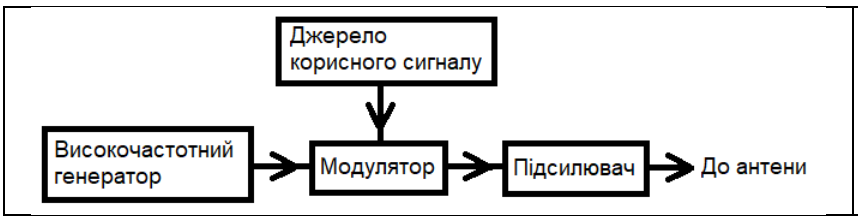


Рис. В.2. Блок-схема передавача

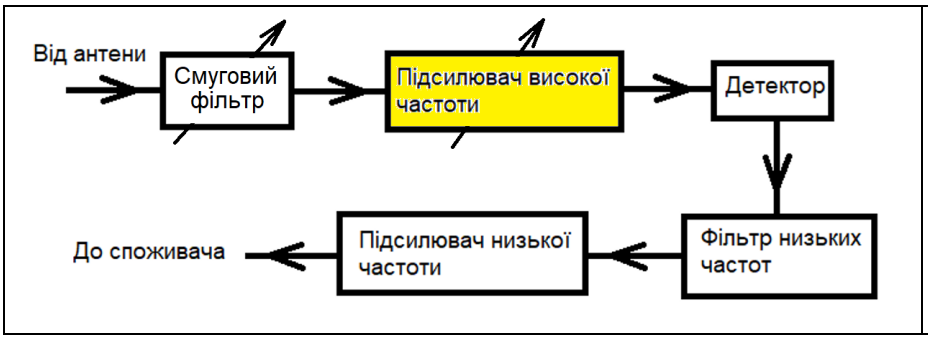


Рис. В.3. Блок-схема приймача прямого підсилення

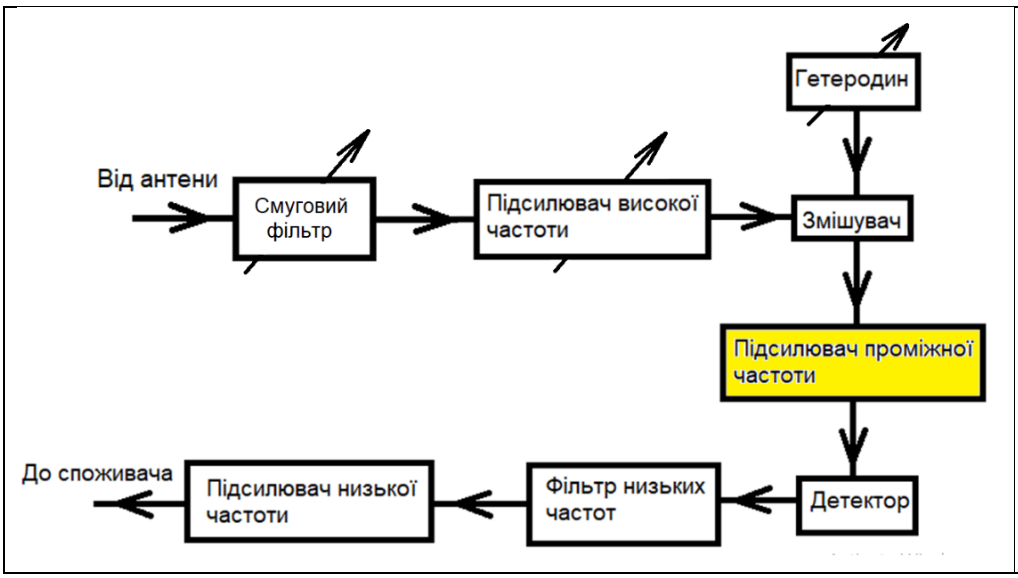


Рис. В.4. Блок-схема супергетеродинного приймача

В аналоговій радіотехніці протягом тривалого часу (принаймні до появи телебачення) одним з найважливіших завдань було передавання та приймання акустичних сигналів. Акустичному діапазону (приблизно від 20 Гц до 20 кГц) відповідають довжини електромагнітних хвиль від 15 тис. км до 15 км. Як ми побачимо далі, для ефективного передавання та приймання сигналів відповідні антени повинні мати розміри порядку довжини хвилі. Через те працювати з електромагнітними хвилями, частоти яких належать до акустичного діапазону, незручно. Для цього зазвичай використовуються радіохвилі з довжинами від одиниць кілометрів до одиниць метрів (відповідну частоту називають частотою-носієм). Для перенесення корисної інформації гармонічний сигнал з частотою-носієм модулюється, тобто один з його параметрів (амплітуда, частота або фаза) змінюються за законом керувального (наприклад, акустичного) сигналу.

На рис. В.2-В.4 показані блок-схеми передавача та різних типів приймачів.

В передавачі (рис. В.2) високочастотний сигнал від генератора модулюється корисним сигналом, підсилюється й передається до передавальної антени, яка перетворює цей сигнал на електромагнітні хвилі й випромінює його в ефір (рис. В1 а). Приймальна антена перетворює прийняті електромагнітні хвилі, що випромінюються передавальною антеною, на електричний сигнал, який надходить на вхід радіоприймача.

Сигнал може надходити з виходу передавача в лінію зв'язку, а потім – на вхід приймача (рис. В1 б). У цьому випадку сигнал повільніше ослаблюється з віддаллю і може бути краще захищений від зовнішніх завад.

Найпростішою схемою радіоприймача є приймач прямого підсилення (рис. В.3). У ньому сигнал від приймальної антени (модульовані корисним сигналом електричні коливання на частоті-носії, а також різноманітні зовнішні завади, наприклад, сигнали від інших передавачів) спершу проходить через смуговий фільтр, який пропускає лише частоти, які займає спектр корисного сигналу (окіл частоти-носія). Цей сигнал підсилюється підсилювачем високої частоти (ПВЧ) і надходить на детектор, який разом з фільтром низьких частот виділяє з модульованого сигналу корисний сигнал. Останній підсилюється підсилювачем низької частоти і надходить до споживача (наприклад, на динамік, який перетворює електричний сигнал на акустичний).

Як правило, радіоприймач повинен мати можливість перестроювання на сигнали з різними частотами-носіями. Для перестоювання приймача прямого підсилення треба перестроювати, зокрема, ПВЧ, який забезпечує основне підсилення сигналу. Через це в такому приймачі не можна використати високоякісні ПВЧ складної конструкції.

Названий недолік долається в супергетеродинному радіоприймачі (рис. В.4), де основне підсилення сигналу забезпечує підсилювач проміжної частоти (ППЧ). При перестроюванні на іншу частоту-носій проміжна частота f_{int} не змінюється, відповідно ППЧ перестроювати не потрібно.

Два перших і три останніх блоки супергетеродинного радіоприймача такі самі, як в приймачі прямого підсилення. Відмінність полягає в додаткових блоках. Один з них – це генератор високої частоти f_g (так званий гетеродин), частоту якого можна перестроювати. Другий – це змішувач, який перетворює модульований сигнал з частотою-носієм f_0 на сигнал з частотою-носієм $f_{\text{int}} = f_g - f_0$ і тією самою модуляцією. При перестроюванні на іншу частоту вхідного сигналу одночасно перестроюють і гетеродин, так що частота f_{int} залишається незмінною.

На практиці саме супергетеродинні приймачі використовуються найчастіше.

Як бачимо, до складу передавачів та приймачів входять фільтри, підсилювачі різних типів, генератори, детектори, модулятори, змішувачі. Принципи роботи та схеми цих пристроїв будуть розглянуті в даному курсі. Крім того, він ознайомить із методами дослідження

лінійних кіл, поняттям про шуми й завади та методи боротьби з ними. Коротко йтиметься про лінії передачі, антени та поширення радіохвиль в атмосфері Землі. Таким чином, у курсі розглядатимуться основні елементи каналу аналогового передавання інформації. Наприкінці курсу ми окремо поговоримо про цифрове передавання та оброблення сигналів.

В.4. Короткий нарис з історії розвитку радіотехніки

Коротко обговоримо етапи розвитку власне радіотехніки.

Радіозв'язок було винайдено в середині 1890-х рр. майже одночасно кількома розробниками з різних країн (Олівер Лодж, 1894, Велика Британія; Джагадіш Чандра Бос, 1894-1895, Індія; Олександр Попов, 1895, Росія; Гульєльмо Марконі, 1895-1896, Італія). На першому етапі розвитку радіотехніки (приблизно до початку 1920-х рр.) радіозв'язок носив характер радіотелеграфу (безпровідного телеграфу), коли інформація кодувалася у формі послідовності імпульсів певної тривалості (наприклад, азбука Морзе).

На початок 1920-х рр. вже існували такі пристрої, як генератори та підсилювачі на основі електронних ламп, а також детектори. Вони поклали початок другому етапу розвитку радіотехніки – радіомовленню, заснованому на використанні модульованих радіосигналів.

Перші експерименти в галузі телебачення спиралися на механічні системи. Вважають, що перша передача рухомого зображення за допомогою електронної системи була продемонстрована в 1928 р. в Ташкенті Борисом Грабовським (сином відомого українського поета й політичного засланця) та Іваном Білянським. Іконоскоп – перша передавальна трубка, яка дозволила створити високоякісне електронне телебачення – був винайдений у 1931 р. Володимиром Зворикіним (США). І вже в 1932 р. в Нью-Йорку почалися перші експериментальні передачі електронного телебачення, що використовували іконоскоп.

На цьому ж етапі розвитку радіотехніки з'явилися радіолокація. Першу радіолокаційну станцію (РЛС) створив ще в 1904 р. Християн Хюльсмейер (Німеччина). Вона використовувалася у судноплавстві на Рейні. Активний розвиток радіолокації почався в 1930-х рр. в зв'язку з розвитком авіації (для потреб протиповітряної оборони). Так, у 1935 р. шотландець Роберт Ватсон-Ватт презентував РЛС, яка використовувала передавання та приймання радіохвиль. Вона дозволяла не тільки виявити об'єкт, але й знайти віддаль до нього, оцінити його розміри та швидкість руху. Радіолокація бурхливо розвивалася в роки Другої світової війни, коли набули поширення генератори мікрохвильового діапазону.

Ідея використання радіомаяків для морських кораблів висловлювалася різними дослідниками ще з кінця 1890-х рр. У 1920-1935 рр. були створені та набули поширення радіомаяки та відповідні системи пеленгації, що поклали початок радіонавігаційним системам. У 1935-1945 рр. були розроблені та впроваджені радіокомпаси для літаків. У 1940-1960 рр. на основі високоточних систем визначення координат були розроблені системи близької та далекої навігації, системи сліпої посадки літаків. До кінця 1970-х рр. була розроблена та запрацювала GPS (Global Positioning System – система глобального позиціонування), що використовує систему супутників (зараз їх 24) для точного (з похибкою в межах 10 м) визначення положення в просторі.

Третій етап розвитку радіотехніки можна умовно датувати 1960-ми – 1970-ми рр. В цей час замість електронних ламп почали використовуватися напівпровідникові елементи – діоди та транзистори, в результаті чого габарити радіотехнічних пристроїв та їхнє енергоспоживання суттєво зменшилися. Це дозволило використовувати складніші схеми і, зокрема, привело до створення перших комп'ютерів.

Четвертий етап розвитку радіотехніки (або, краще сказати, радіоелектроніки) розпочався в 1980-х рр. Він характеризується двома особливостями. По-перше, це перехід від схем, зібраних з дискретних елементів, до так званих інтегральних схем, які виготовляються на одній пластині напівпровідника і в єдиному технологічному циклі. Це дозволило різко

підвищити надійність схем і зробило можливим їхнє подальше значне ускладнення. По-друге, радіотехніка перейшла від роботи з аналоговими (неперервними в часі) сигналами до імпульсних сигналів, що дозволило різко збільшити точність передавання інформації та її захищеність від завад. Цей етап триває досі.

В.5. Структура курсу

Частина I. Радіосигнали та лінійні кола

1. Регулярні сигнали в радіотехніці: часове та спектральне представлення
2. Методи аналізу лінійних електричних кіл

Частина II. Елементи радіоелектроніки

3. Прилади напівпровідникової електроніки
4. Лінійні пристрої аналогової радіоелектроніки
5. Автогенератори та інші нелінійні аналогові схеми

Частина III. Передавання сигналів на віддаль

6. Лінії зв'язку
7. Антени
8. Поширення радіохвиль у приземному просторі

Частина IV. Елементи статистичної радіотехніки

9. Методи опису та аналізу випадкових процесів
10. Шуми та завади в радіотехнічних колах
11. Відокремлення корисних сигналів від завад і шумів

Частина V. Елементи цифрової електроніки

12. Поняття про цифрове передавання інформації

Перші дві частини курсу можна розглядати як вступ до аналогової радіоелектроніки. У третій та четвертій частинах обговорюються суміжні питання, які традиційно розглядає радіотехніка – лінії зв'язку, випромінювання та поширення радіохвиль, а також шуми в радіоелектронних колах та відокремлення від них корисного сигналу. Остання частина служить дуже стислим вступом до теорії цифрового передавання інформації.

Матеріал даної дисципліни використовується (або частково повторюється) в ряді обов'язкових та вибіркових курсів для студентів даної спеціальності.

Обов'язкові курси: «Основи теорії кіл», «Матеріали сучасної електроніки», «Компонентна база радіоелектронних засобів», «Аналогові електронні пристрої», «Генерування, формування та передавання сигналів», «Приймання та оброблення сигналів», «Цифрове оброблення сигналів», «Цифрові пристрої», «Електродинаміка та поширення радіохвиль», «Основи теорії передавання інформації», «Надійність і технічна діагностика радіоелектронних засобів», «Основи теорії автоматичного управління», «Основи радіолокації та радіонавігації».

Вибіркові курси: «Антенні системи та техніка мікрохвильового діапазону», «Електронні та квантові пристрої мікрохвильового діапазону», «Статистична радіотехніка».

Контрольні питання до Вступу

1. Що вивчає радіотехніка? Де вона застосовується?
2. В чому відмінності між термінами «радіотехніка», «радіоелектроніка» та «електроніка»?
3. Яким довжинам хвиль відповідає радіодіапазон? На які піддіапазони його ділять?
4. Опишіть тракт передавання-приймання сигналу.
5. В чому відмінність між радіоприймачем прямого підсилення та супергетеродинним радіоприймачем?

Використана література

1. Н.В. Зернов, В.Г. Карпов. Теорія радіотехнічних кіл. Енергія, 1965.
2. В. Жалуд, В.Н. Кулешов. Шуми в напівпровідникових пристроях. 1977.
3. С.А. Ахманов, Ю.Е. Дьяков, А.С. Чиркін. Вступ до статистичної радіофізики та оптики. Наука, 1981.
4. Л.Д. Ландау, Е.М. Лифшиць. Теорія поля. Наука, 1988.
5. В.В. Нікольський, Т.В. Нікольська. Електродинаміка й поширення радіохвиль. Наука, 1989.
6. А.Л. Драбкін, Е.Б. Коренберг. Антени. Радіо й зв'язок, 1992.
7. Ю.П. Жураковський, В.П. Полторак. Теорія інформації та кодування. К., Вища школа, 2001.
8. Ю.І. Волощук. Сигнали та процеси у радіотехніці: Підручник для студентів вищих навчальних закладів, том 1-2. - Харків: «Компанія СМІТ», 2003.
9. В.В. Лідовский. Теорія інформації. Компанія Супутник+, 2004.
10. ROTHAMMELS ANTENNENBUCH. Neu bearbeitet und erweitert von Alois Krischke. 12 aktualisierte und erweitere Auflage. DARC Verlag Baunatal. 2007.
11. С.І. Баскаков Лекції з теорії кіл. КомКнига, 2005.
12. С.М. Левитський, В.В. Загородній. Вступ до радіозв'язку. Випромінювання та поширення радіохвиль. К., ВПЦ «Київський університет», 2006.
13. С.М. Левитський. Основи радіоелектроніки (підручник). К., ВПЦ «Київський університет», 2007.
14. Н.А. Малков, А.П. Пудовкін. Електромагнітна сумісність радіоелектронних засобів. ТГТУ, 2007.
15. Г.А. Єрохін, О.В. Чернов, Н.Д. Козирев, В.Д. Кочержевський. Антенно-фидерні пристрої та поширення радіохвиль. Гаряча Лінія – Телеком, 2007.
16. І.О. Анісімов. Коливання та хвилі. К., ВПЦ «Київський університет», 2009.
17. В.П. Майданюк. Кодування та захист інформації. Навчальний посібник. Вінниця, ВНТУ, 2009.
18. І.І. Бех, С.М. Левитський. Фізичні основи комп'ютерної електроніки. К., Карбон, 2010.
19. С.М. Левитський. Теорія радіотехнічних кіл. К., 2012.
20. А.В. Якимов. Фізика шумів і флуктуацій параметрів. 2013.
21. О.М. Нікітчин, С.М. Левитський. Сигнали і процеси в радіотехніці. К., Логос, 2014.
22. В.О. Іванов, Є.І. Габрусенко, Л.Я. Ільницький, О.А. Щербина. Електромагнітна сумісність радіоелектронної апаратури. К., НАУ, 2014.
23. Т.А. Цалієв. Електромагнітна сумісність радіоелектронних засобів. Одесса, ОНАС, 2014.
24. О.В. Бирюкова. Антенно-фідерні пристрої. 2015.
25. М.Ю. Артеменко. Спеціальні розділи теорії електричних кіл. К., НТУУ «КПІ», 2016.
26. І.В. Гордієнко. Пристрої НВЧ та антени. Дніпро, 2017.
27. В.Г. Коберниченко. Основи цифрової обробки сигналів. 2018.
28. В.О. Піддубний, І.О. Товкач. Елементна база радіоелектронної апаратури. Багатоперехідні структури. К., НТУУ «КПІ», 2021.
29. В.С. Лазебний, В.В. Пілінський, В.Б. Швайченко. Електромагнітна сумісність радіоелектронних засобів. К., НТУУ «КПІ», 2023.

Частина І. Радіосигнали та методи аналізу лінійних кіл

Ми вже згадували (див. п. В3), що в аналоговій радіотехніці для передавання інформації використовуються різні типи модульованих коливань. У цій частині курсу ми спочатку розглянемо різновиди модульованих коливань. У зв'язку з цим ми детально обговоримо спектральне представлення сигналів, яке в багатьох випадках виявляється зручнішим, ніж часове. Поговоримо про ряди й інтеграли Фур'є, а також про узагальнення поняття спектру – віконне перетворення Фур'є та вейвлет-аналіз (розділ 1).

Значна частина радіотехнічних кіл належить до числа лінійних, тобто їхні параметри не змінюються під дією сигналу, а процеси в них описуються лінійними диференціальними рівняннями зі сталими коефіцієнтами. Тому ми розглянемо методи аналізу лінійних електричних кіл, збуджуваних гармонічними джерелами (з якими часто мають справу в радіотехніці) та джерелами з довільною часовою залежністю, а також операторний метод розв'язання лінійних диференціальних рівнянь з довільними початковими умовами (розділ 2).

Розділ 1. Регулярні сигнали в радіотехніці: часове та спектральне представлення

Спосіб опису процесів у тій чи іншій радіотехнічній системі суттєво залежить від її розміру.

З курсу загальної фізики відомо, що електромагнітні сигнали поширюються в просторі зі скінченною швидкістю $c = 3 \cdot 10^8\ м/с = 3 \cdot 10^{10}\ см/с = 300\ тис.\ км/с$. Іншими словами, поширення сигналу в просторі має характер хвильового процесу (коли йдеться про ефір, це буде електромагнітна хвиля). Хвильові процеси описуються диференціальними рівняннями в частинних похідних, оскільки хвильові поля змінюються як у часі, так і в просторі.

Але нерідко розміри досліджуваної системи (наприклад, електричного кола) можуть бути значно меншими за довжину хвилі. В цьому випадку можна вважати, що в будь-якому перерізі провідника, який входить до складу кола, струм однаковий, а потенціал будь-якої точки такого провідника також один і той самий. Струми й напруги в такому колі змінюються лише в часі, а процеси їхньої зміни описуються диференціальними рівняннями в повних похідних. Системи, описувані такими рівняннями, називають системами із зосередженими параметрами. Переважно саме такі системи й розглядатимуться в даному курсі (за винятком частини ІІІ, присвяченої передаванню сигналів на віддаль, де йдеться про лінії передачі, антени та поширення радіохвиль).

Оцінимо, для яких довжин хвиль електричні кола можна розглядати як системи із зосередженими параметрами. Покладемо характерний розмір схеми 10 см. Тоді довжини хвиль, на яких таку схему можна розглядати як систему із зосередженими параметрами, не повинні бути меншими за 1 м. Цій умові задовольняють радіохвилі від довгих до метрових включно (див. табл. В.2). Відповідно частота таких хвиль має бути меншою від 300 МГц.

1.1. Модульовані сигнали

На першому етапі розвитку радіозв'язку (див. вище п. В4) сигнали, як уже згадувалося, передавались у формі деякого коду типу абетки Морзе (так звані сигнали з амплітудною маніпуляцією, див. нижче). Пізніше були розроблені методи передавання в ефір та приймання акустичних сигналів. Складність при вирішенні цієї задачі, як уже відзначалося (див. п. В3; див. також нижче розділ 7), полягає в тому, що антени для передавання й приймання сигналів найбільш ефективно працюють, коли їхні розміри сумірні з довжиною відповідних хвиль. Але акустичний діапазон простягається приблизно від 20 Гц до 20 кГц, що відповідає довжинам електромагнітних хвиль від 15 км до майже 20 тис. км (див. вище п. В.3). Зрозуміло, що антени таких розмірів або неможливо побудувати взагалі, або вкрай дорого й незруч-

но експлуатувати. Тому виникла ідея використати електромагнітні хвилі значно більшої частоти, але модулювати їх акустичним сигналом. Подивимося, як це в принципі можна зробити.

1.1.1. Амплітудно-модульовані сигнали

Гармонічний (немодульований) сигнал може бути поданий у формі

$$u(t) = u_m \cos(\omega_0 t + \varphi_0), \tag{1.1.1}$$

де u_m, ω_0 і φ_0 – відповідно амплітуда, частота та початкова фаза сигналу, $\varphi(t) = \omega_0 t + \varphi_0$ – поточна (миттєва) фаза.

При модуляції амплітуда, частота або початкова фаза змінюються з часом за законом керувального сигналу. Відповідно розрізняють амплітудну, частотну й фазову модуляцію. Амплітудно-модульований сигнал (рис. 1.1.1) має вигляд

$$u(t) = u_m \left[1 + k u_{\text{mod}}(t)\right] \cos(\omega_0 t), \tag{1.1.2}$$

де $u_{\text{mod}}(t)$ – керувальний сигнал, ω_0 – частота-носій, k – деякий множник з розмірністю В$^{-1}$. Надалі вважатимемо виконаною умову

$$\left| k u_{\text{mod}}(t) \right| < 1 \tag{1.1.3}$$

(вираз у квадратних дужках в (1.1.2) має залишатися додатним). Якщо ця умова не виконується, говорять про перемодуляцію. За перемодуляції однозначно відновити обвідну $u_{\text{mod}}(t)$ за модульованим сигналом (1.1.2) неможливо.

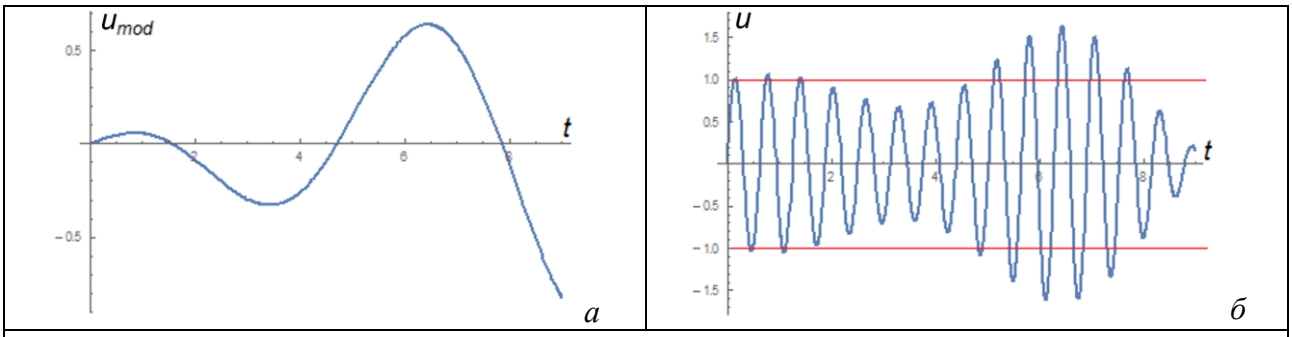


Рис. 1.1.1. Сигнал модуляції (а) та амплітудно-модульований сигнал (б)

Характерний час зміни обвідної можна визначити по-різному, наприклад:

$$\tau = \left| \frac{1}{u_{\text{mod}}} \frac{d u_{\text{mod}}(t)}{dt} \right|^{-1} \tag{1.1.4}$$

(інші можливі визначення дають величину того самого порядку).

Формула (1.1.2) відповідає модульованому сигналу, якщо виконана умова

$$\omega_0 \tau \gg 1, \tag{1.1.5}$$

тобто характерний час зміни обвідної значно більший за період, що відповідає частоті-носію (рис. 1.1.1).

1.1.2. Відеоімпульси та радіоімпульси

Частинним випадком амплітудно-модульованого сигналу є радіоімпульс – амплітудно-модульований сигнал обмеженої (невеликої) тривалості. Його обвідну в цьому випадку називають відеоімпульсом. Зазначимо, що на практиці ми завжди маємо справу із сигналами обмеженої тривалості, тобто їх часто можна розглядати як відеоімпульси або радіоімпульси.

Відеоімпульс $u(t)$ має задовольняти умовам:

$$u(t) \geq 0, \qquad |u(t \to \pm\infty)| = 0. \tag{1.1.6}$$

Радіоімпульс має вигляд:

$$u(t) = u_m(t)\cos\omega_0 t, \qquad u_m(t) \geq 0, \qquad |u_m(t \to \pm\infty)| = 0, \qquad \left|\frac{1}{u_m}\frac{du_m}{dt}\right|^{-1} >> \frac{2\pi}{\omega_0}. \tag{1.1.7}$$

(рис. 1.1.2). Функцію $u_m(t)$ називають обвідною радіоімпульсу.

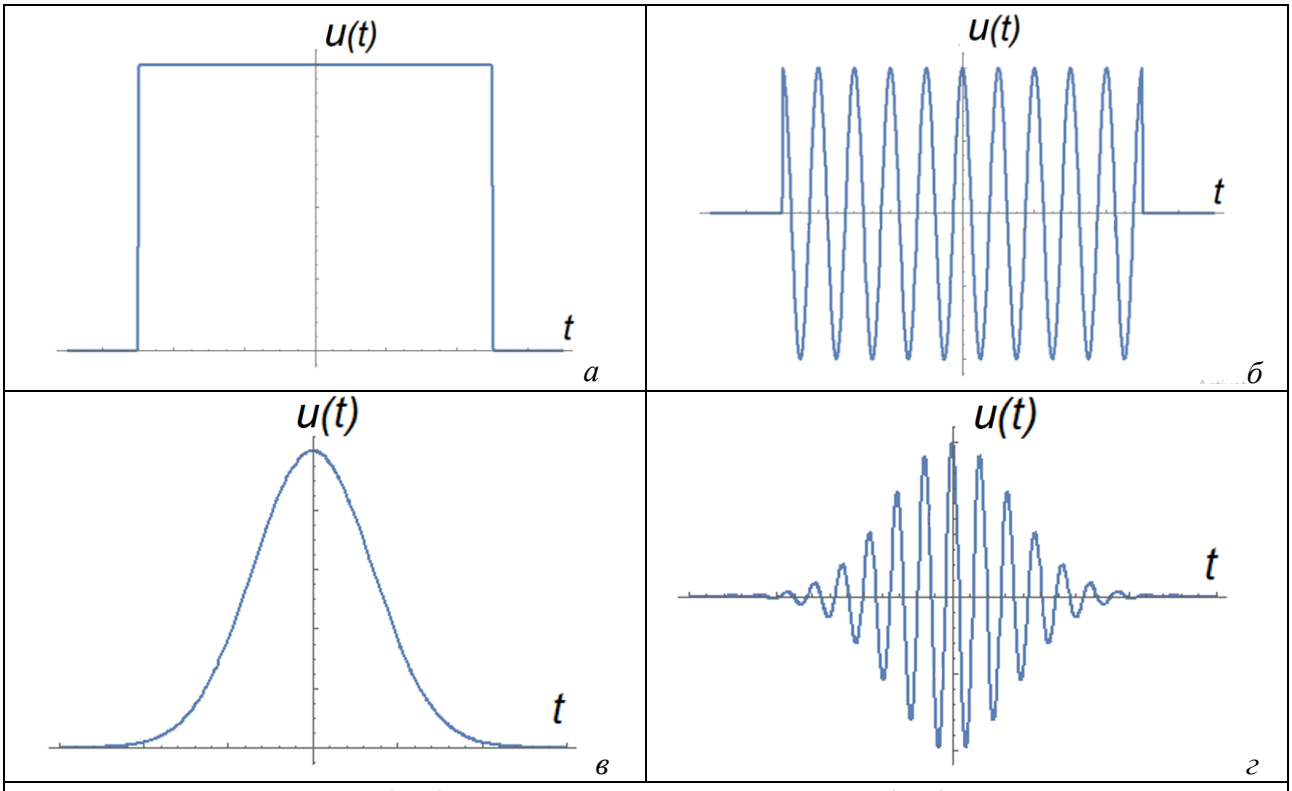


Рис. 1.1.2. Відеоімпульси (а, в) та відповідні їм радіоімпульси (б, г): а-б – імпульси прямокутної форми, в-г – імпульси гауссівської форми

Зокрема, відеоміпульс і радіоімпульс прямокутної форми можуть бути записані у вигляді

$$u(t) = \begin{cases} u_m, & |t| \leq T/2; \\ 0, & |t| > T/2, \end{cases} \tag{1.1.8}$$

та

$$u(t) = \begin{cases} u_m \cos\omega_0 t, & |t| \leq T/2; \\ 0, & |t| > T/2, \end{cases} \tag{1.1.9}$$

(рис. 1.1.2 а-б). Такі відео- та радіоімпульси широко використовуються в цифровій радіоелектроніці, зокрема, прямокутні відеоімпульси – при передаванні сигналів на віддаль (див. нижче частину V).

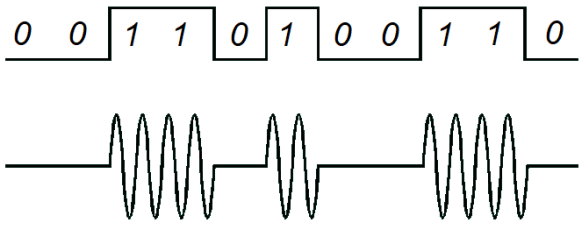


Рис. 1.1.3. Сигнал з амплітудною маніпуляцією для передачі двійкового коду

Послідовність прямокутних відеоімпульсів різної тривалості утворює сигнал з амплітудною маніпуляцією (рис. 1.1.3). Такі сигнали використовувалися свого часу в радіотелеграфії: за допомогою послідовності радіоімпульсів різної тривалості кодувалися літери абетки (так звана абетка Морзе).

1.1.3. Частотна модуляція

Частотно-модульований сигнал може бути поданий у вигляді

$$u(t) = u_m \cos\left\{\left[\omega_0 + \Delta\omega(t)\right]t + \varphi_0\right\}, \qquad \Delta\omega(t) = k u_{\text{mod}}(t) \qquad (1.1.10)$$

(рис. 1.4). Як і в попередньому випадку, керувальний сигнал має бути повільним у масштабі періоду частоти-носія (див. вище формулу (1.1.5)), тобто

$$\left|\Delta\omega(t)\right| \ll \omega_0 . \qquad (1.1.11)$$

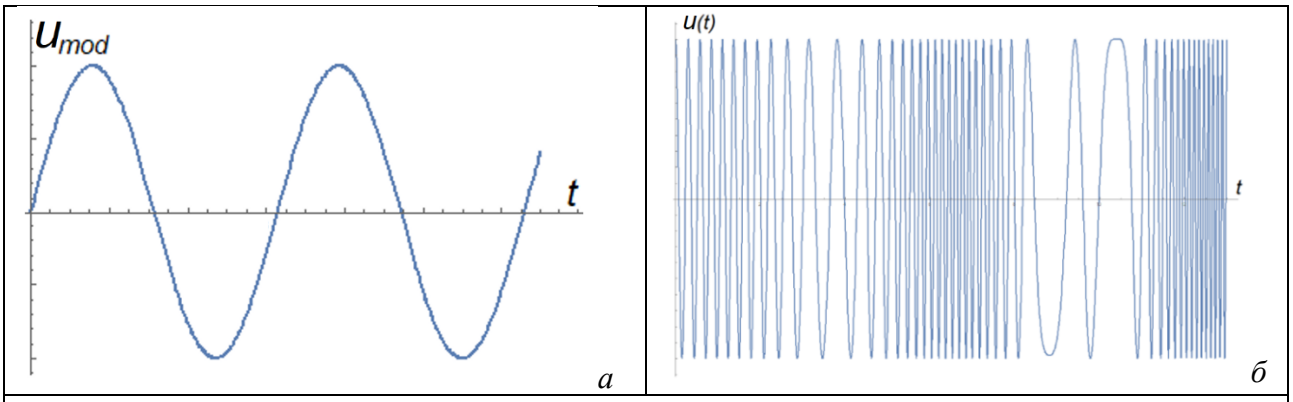


Рис. 1.1.4. Сигнал модуляції (а) та частотно-модульований сигнал (б)

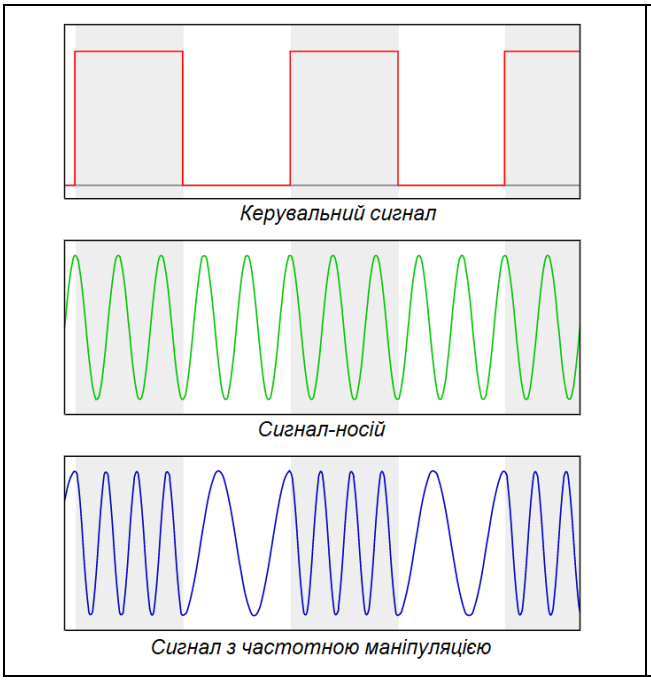


Рис. 1.1.5. Сигнал з частотною маніпуляцією

На рис. 1.1.4 видно, що періодичному закону модуляції відповідає аперіодичний модульований сигнал. Щоб зрозуміти, чому так виходить, знайдемо миттєву частоту сигналу (1.1.10) як похідну від його повної фази:

$$\omega(t) = \frac{d}{dt}\left\{\left[\omega_0 + \Delta\omega(t)\right]t + \varphi\right\} = \omega_0 + \Delta\omega(t) + t\frac{d}{dt}\Delta\omega(t). \qquad (1.1.12)$$

Як бачимо, для періодичної функції $\Delta\omega(t)$ миттєва частота (1.1.12) справді виходить аперіодичною за рахунок останнього доданку, який монотонно зростає з часом. З цієї причини частотна модуляція на практиці не використовується. Винятком є частотна маніпуляція, коли величина $\Delta\omega(t)$ набуває лише двох (сталих) значень, так що останній доданок в (1.1.12) зникає (рис. 1.1.5).

1.1.4. Фазова модуляція

Фазово-модульований сигнал має вигляд

$$u(t) = u_m \cos\left[\omega_0 t + \varphi(t)\right], \qquad \varphi(t) = k u_{\text{mod}}(t) \tag{1.1.13}$$

(рис. 1.1.6). Його миттєва частота завжди змінюється з часом і визначається формулою

$$\omega(t) = \frac{d}{dt}\left[\omega_0 t + \varphi(t)\right] = \omega_0 + \frac{d}{dt}\varphi(t), \tag{1.1.14}$$

тобто буде періодичною функцією часу, якщо керувальний сигнал періодичний. І в цьому випадку керувальний сигнал має бути повільним у масштабі періоду частоти-носія:

$$\left|\frac{d\varphi}{dt}\right| \ll \omega_0 . \tag{1.1.15}$$

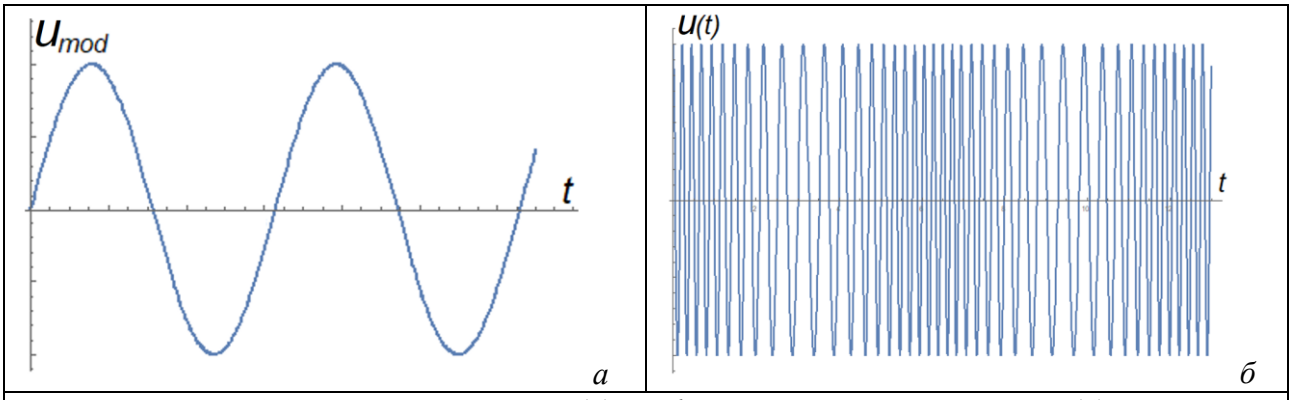


Рис. 1.1.6. Сигнал модуляції (а) та фазово-модульований сигнал (б)

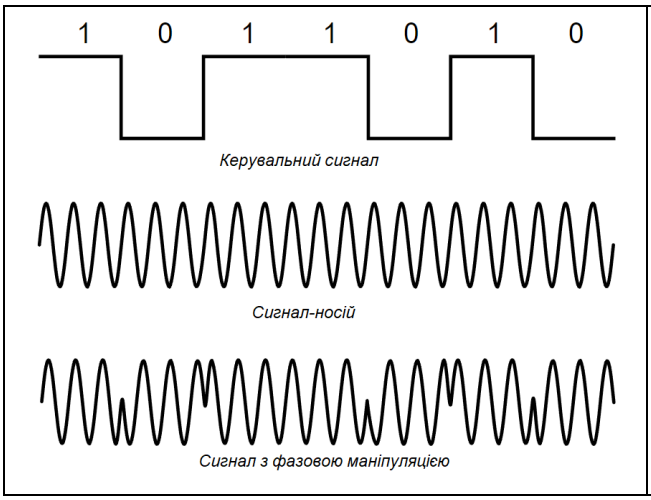


Рис. 1.1.7. Сигнал з фазовою маніпуляцією

Якщо визначити функцію $\varphi(t)$ з умови

$$\frac{d}{dt}\varphi(t) = \Delta\omega(t) + t\frac{d}{dt}\Delta\omega(t) = \frac{d}{dt}\left[t\Delta\omega(t)\right], \tag{1.1.16}$$

або

$$\varphi(t) = t\Delta\omega(t) + \varphi_0, \qquad (1.1.17)$$

то частотна модуляція може бути формально зведена до фазової[1]. Тому фазову модуляцію, яка широко використовувалася в останній період розвитку аналогового радіомовлення, на побутовому рівні інколи неправильно називають частотною.

Сигнал із фазовою маніпуляцією поданий на рис. 1.1.7.

Про те, за допомогою яких схем можна практично реалізувати різні типи модуляції гармонічного сигналу, йтиметься в другій частині даного курсу (див. нижче пп. 5.7.5-5.7.6).

Контрольні питання до підрозділу 1.1

1. Для чого потрібна модуляція низькочастотних сигналів перед передаванням в ефір?
2. В чому відмінність між радіоімпульсом та відеоімпульсом?
3. Чому замість частотно-модульованих сигналів на практиці використовуються фазово-модульовані?
4. Що таке амплітудна, частотна та фазова маніпуляція сигналу?

[1] Зазначимо, що, відповідно до формул (1.1.16)-(1.1.17), періодична зміна частоти не супроводжується періодичною зміною фази, як уже вказувалося вище.

1.2. Ряди Фур'є

Досі (п. 1.1) ми розглядали сигнали як функції часу. Як ми побачимо далі, в багатьох випадках їх зручно представляти у вигляді спектру – дискретного або неперервного набору гармонічних функцій з різними амплітудами та фазами. В багатьох випадках (але не завжди) знаходження спектру відповідає перетворенню Фур'є, тобто розкладанню функцій в ряд або в інтеграл Фур'є.

Фур'є-перетворення є особливо зручним при аналізі лінійних систем – систем, параметри яких не змінюються під дією сигналу. Такі системи описуються лінійними диференціальними рівняннями. Для них справджується принцип суперпозиції: відгук системи на суму впливів дорівнює сумі відгуків на кожен із впливів.

1.2.1. Повна ортонормована система функцій

Для того, щоб довільну (однозначну) функцію можна було однозначно подати як суму деяких інших функцій, останні повинні бути ортогональними й утворювати повну систему. Нескінченна система функцій

$$\varphi_1(x), \varphi_2(x), ..., \varphi_n(x), ... \tag{1.2.1}$$

називається ортогональною на відрізку $[a,b]$, якщо при будь-яких $n \neq k$ виконується рівність

$$\int_a^b \varphi_n(x)\varphi_k(x)\,dx = 0. \tag{1.2.2}$$

При цьому норма окремих функцій не повинна дорівнювати нулю:

$$\int_a^b \left[\varphi_n(x)\right]^2 dx \neq 0. \tag{1.2.3}$$

Ортогональна система функцій називається ортонормованою, якщо норма кожної з функцій дорівнює одиниці:

$$\int_a^b \left[\varphi_n(x)\right]^2 dx = 1. \tag{1.2.4}$$

Ортонормована система функцій (1.2.1) називається повною, якщо для будь-якої (дійсної) функції $f(x)$, що задовольняє умову

$$\int_a^b f^2(x)\,dx < \infty, \tag{1.2.5}$$

можна підібрати такий набір коефіцієнтів c_i, для якого виконується рівність

$$\lim_{n\to\infty}\int_a^b \left[f(x) - \sum_{i=0}^{n} c_i\varphi_i(x)\right]dx = 0. \tag{1.2.6}$$

Іншими словами, середньоквадратичне відхилення суми $\sum_{i=0}^{n} c_i\varphi_i(x)$ від функції $f(x)$ прямує до нуля при $n \to \infty$.

1.2.2. Ортогональність гармонік

Під гармоніками розуміють функції $\cos\omega t$, $\sin\omega t$ та $\exp(i\omega t)$.

Покажемо, що набір тригонометричних функцій $\cos(2\pi nt/T)$ та $\sin(2\pi nt/T)$ ($n = 1,2,3,...$) є ортогональним на проміжку $[-T/2, T/2]$, тобто виконуються такі співвідношення:

$$\int_{-T/2}^{T/2} \cos\left(\frac{2\pi n}{T}t\right)\sin\left(\frac{2\pi n'}{T}t\right)dt = 0 \qquad (1.2.7)$$

(результат очевидний, оскільки маємо інтеграл від непарної функції в симетричних межах);

$$\frac{2}{T}\int_{-T/2}^{T/2} \cos\left(\frac{2\pi n}{T}t\right)\cos\left(\frac{2\pi n'}{T}t\right)dt = \begin{cases} 0, & n \neq n', \\ 1, & n = n', \end{cases} \qquad (1.2.8)$$

$$\frac{2}{T}\int_{-T/2}^{T/2} \sin\left(\frac{2\pi n}{T}t\right)\sin\left(\frac{2\pi n'}{T}t\right)dt = \begin{cases} 0, & n \neq n', \\ 1, & n = n', \end{cases} \qquad (1.2.9)$$

Доведемо, що ці співвідношення справді виконуються. Нехай спершу $n \neq n'$. Маємо:

$$\int_{-T/2}^{T/2} \cos\left(\frac{2\pi n}{T}t\right)\cos\left(\frac{2\pi n'}{T}t\right)dt = \frac{1}{2}\int_{-T/2}^{T/2}\left\{\cos\left[\frac{2\pi t}{T}\left(n+n'\right)\right] + \cos\left[\frac{2\pi t}{T}\left(n-n'\right)\right]\right\}dt = 0$$

(інтеграл від косинуса по цілому числу періодів дорівнює нулеві). Аналогічно

$$\int_{-T/2}^{T/2} \sin\left(\frac{2\pi n}{T}t\right)\sin\left(\frac{2\pi n'}{T}t\right)dt = \frac{1}{2}\int_{-T/2}^{T/2}\left\{\cos\left[\frac{2\pi t}{T}\left(n-n'\right)\right] - \cos\left[\frac{2\pi t}{T}\left(n+n'\right)\right]\right\}dt = 0.$$

Нарешті, при $n = n'$

$$\int_{-T/2}^{T/2} \cos^2\left(\frac{2\pi n}{T}t\right)dt = \frac{1}{2}\int_{-T/2}^{T/2}\left[1+\cos\left(\frac{4\pi n}{T}t\right)\right]dt = \frac{1}{2}\int_{-T/2}^{T/2} dt = \frac{T}{2},$$

$$\int_{-T/2}^{T/2} \sin^2\left(\frac{2\pi n}{T}t\right)dt = \frac{1}{2}\int_{-T/2}^{T/2}\left[1-\cos\left(\frac{4\pi n}{T}t\right)\right]dt = \frac{1}{2}\int_{-T/2}^{T/2} dt = \frac{T}{2}.$$

Отже, співвідношення (1.2.7)-(1.2.9) справді виконуються. Таким чином, функції $\sqrt{2/T}\cos\left(2\pi nt/T\right)$ та $\sqrt{2/T}\sin\left(2\pi nt/T\right)$ (які, як легко зрозуміти, є нормованими) утворюють ортонормований набір на проміжку $\left[-T/2, T/2\right]$. Можна також показати, що цей набір функцій є повним. Тому по них можна розкладати будь-які інші функції, визначені на відрізку $\left[-T/2, T/2\right]$:

$$f\left(t\right) = c_0 + \sum_{n=1}^{\infty}\left[a_n\cos\left(\frac{2\pi n}{T}t\right) + b_n\sin\left(\frac{2\pi n}{T}t\right)\right]. \qquad (1.2.10)$$

Легко бачити, що права частина цієї рівності є періодичною функцією. Отже, таке представлення справедливе не тільки для довільних функцій, визначених на відрізку $\left[-T/2, T/2\right]$, але й для будь-яких періодичних функцій з періодом T, визначених на дійсній осі.

Таке розкладення називають рядом Фур'є.

1.2.3. Ряд Фур'є

Дамо коректніше (з математичної точки зору) визначення ряду Фур'є.

Нехай функція $f\left(t\right)$ є періодичною,

$$f\left(t+nT\right) = f\left(t\right), \qquad T > 0, \qquad n = \pm 1, \pm 2, \pm 3 \ldots, \qquad (1.2.11)$$

і задовольняє умови Діріхлє:

- функція $f\left(t\right)$ – обмежена;
- на періоді вона має скінченну кількість максимумів і мінімумів та скінченну кількість розривів.

Крім того, $f(t)$ – однозначна функція (кожному значенню t відповідає одне єдине значення $f(t)$).

Тоді функція $f(t)$ може бути подана у вигляді ряду Фур'є (1.2.10):

$$f(t)=c_0+\sum_{n=1}^{\infty}\left[a_n\cos\left(\frac{2\pi n}{T}t\right)+b_n\sin\left(\frac{2\pi n}{T}t\right)\right],$$

де

$$a_n=\frac{2}{T}\int_{-T/2}^{T/2}f(t)\cos\left(\frac{2\pi n}{T}t\right)dt, \tag{1.2.12}$$

$$b_n=\frac{2}{T}\int_{-T/2}^{T/2}f(t)\sin\left(\frac{2\pi n}{T}t\right)dt, \tag{1.2.13}$$

$$c_0=\frac{1}{T}\int_{-T/2}^{T/2}f(t)dt. \tag{1.2.14}$$

Щоб довести формулу (1.2.12), помножимо обидві частини рівності (1.2.10) на $\cos(2\pi kt/T)$ і проінтегруємо за часом на проміжку $[-T/2,T/2]$. У лівій частині отримаємо

$$\int_{-T/2}^{T/2}f(t)\cos(2\pi kt/T)dt.$$

У правій частині доданки матимуть вигляд

$$\begin{aligned}&a_n\int_{-T/2}^{T/2}\cos\left(\frac{2\pi n}{T}t\right)\cos\left(\frac{2\pi k}{T}t\right)dt=\frac{a_n}{2}\int_{-T/2}^{T/2}\left\{\cos\left[(n+k)\frac{2\pi}{T}t\right]+\cos\left[(n-k)\frac{2\pi}{T}t\right]\right\}dt=\\&=\frac{a_n}{2}\int_{-T/2}^{T/2}\cos\left[(n-k)\frac{2\pi}{T}t\right]dt=\begin{cases}0, & n\neq k,\\ a_nT/2, & n=k;\end{cases}\end{aligned} \tag{1.2.15}$$

$$b_n\int_{-T/2}^{T/2}\sin\left(\frac{2\pi n}{T}t\right)\cos\left(\frac{2\pi k}{T}t\right)dt=\frac{b_n}{2}\int_{-T/2}^{T/2}\left\{\sin\left[(n+k)\frac{2\pi}{T}t\right]+\sin\left[(n-k)\frac{2\pi}{T}t\right]\right\}dt=0. \tag{1.2.16}$$

(знову враховано, що інтеграл від гармонічної функції по цілій кількості періодів дає нуль).

Прирівнявши ліву й праву частини, отримуємо наведену вище формулу для a_n. Аналогічно отримуємо формулу для b_n.

Таким чином, періодичні функції мають дискретний спектр, що визначається набором коефіцієнтів a_n, b_n, c_0.

Для парних функцій ($f(-t)=f(t)$) усі коефіцієнти $b_n=0$, для непарних ($f(-t)=-f(t)$), навпаки, $a_n=0$, $c_0=0$.

Зазначимо також, що в силу формул (1.2.12)-(1.2.13) $a_{-n}=a_n$, $b_{-n}=-b_n$.

Перетворення Фур'є – лінійна операція. Тому коефіцієнти ряду Фур'є від суми двох функцій з однаковими періодами дорівнюють сумі відповідних коефіцієнтів рядів для кожної окремої функції. При збільшенні функції $f(t)$ в k разів коефіцієнти ряду (2.2) також збільшуються в k разів.

1.2.4. Альтернативні форми запису ряду Фур'є

Ряд Фур'є може бути записаний як сума лише косинусів з різними частотами та фаза-

ми:

$$f(t)=c_0+\sum_{n=1}^{\infty}\left[a_n\cos\left(\frac{2\pi n}{T}t\right)+b_n\sin\left(\frac{2\pi n}{T}t\right)\right]=c_0+\sum_{n=1}^{\infty}d_n\left[\frac{a_n}{d_n}\cos\left(\frac{2\pi n}{T}t\right)+\frac{b_n}{d_n}\sin\left(\frac{2\pi n}{T}t\right)\right]=$$

$$=c_0+\sum_{n=1}^{\infty}d_n\left[\cos\varphi_n\cos\left(\frac{2\pi n}{T}t\right)-\sin\varphi_n\sin\left(\frac{2\pi n}{T}t\right)\right]=c_0+\sum_{n=1}^{\infty}d_n\cos\left(\frac{2\pi n}{T}t+\varphi_n\right), \qquad (1.2.17)$$

де

$$d_n=\sqrt{a_n^2+b_n^2}, \qquad \cos\varphi_n=a_n/d_n, \qquad \sin\varphi_n=-b_n/d_n. \qquad (1.2.18)$$

Тут враховане відоме співвідношення

$\cos\alpha\cos\beta-\sin\alpha\sin\beta=\cos(\alpha+\beta)$.

Ряд Фур'є може бути також записаний через експоненти з уявними показниками[2]. Для отримання такої форми ряду Фур'є скористаємося відомим поданням тригонометричних функцій через експоненти та формулою Ейлера

$$\cos\alpha=\frac{\exp(i\alpha)+\exp(-i\alpha)}{2}, \qquad \sin\alpha=\frac{\exp(i\alpha)-\exp(-i\alpha)}{2i}, \qquad \exp(i\alpha)=\cos\alpha+i\sin\alpha.$$

Маємо:

$$f(t)=c_0+\sum_{n=1}^{\infty}\left[a_n\cos\left(\frac{2\pi n}{T}t\right)+b_n\sin\left(\frac{2\pi n}{T}t\right)\right]=$$

$$=c_0+\sum_{n=1}^{\infty}\left[a_n\frac{\exp\left(i\frac{2\pi n}{T}t\right)+\exp\left(-i\frac{2\pi n}{T}t\right)}{2}+b_n\frac{\exp\left(i\frac{2\pi n}{T}t\right)-\exp\left(-i\frac{2\pi n}{T}t\right)}{2i}\right]=$$

$$=c_0+\sum_{n=1}^{\infty}\left[\frac{a_n-ib_n}{2}\exp\left(i\frac{2\pi n}{T}t\right)+\frac{a_n+ib_n}{2}\exp\left(-i\frac{2\pi n}{T}t\right)\right]= \qquad (1.2.19)$$

$$=c_0+\sum_{n=1}^{\infty}\frac{a_n-ib_n}{2}\exp\left(i\frac{2\pi n}{T}t\right)+\sum_{n=1}^{\infty}\frac{a_n+ib_n}{2}\exp\left(-i\frac{2\pi n}{T}t\right)=$$

$$=c_0+\sum_{n=1}^{\infty}\frac{a_n-ib_n}{2}\exp\left(i\frac{2\pi n}{T}t\right)+\sum_{n=-1}^{-\infty}\frac{a_n-ib_n}{2}\exp\left(i\frac{2\pi n}{T}t\right)=\sum_{n=-\infty}^{\infty}c_n\exp\left(i\frac{2\pi n}{T}t\right),$$

де

$$c_n=\frac{a_n-ib_n}{2}=\frac{1}{T}\int_{-T/2}^{T/2}f(t)\left[\cos\left(\frac{2\pi n}{T}t\right)-i\sin\left(\frac{2\pi n}{T}t\right)\right]dt=\frac{1}{T}\int_{-T/2}^{T/2}f(t)\exp\left(-i\frac{2\pi n}{T}t\right)dt. \qquad (1.2.20)$$

Враховані також співвідношення $a_{-n}=a_n$, $b_{-n}=-b_n$.

Коефіцієнти c_n можна представити також через величини (1.2.18):

$$c_n=\frac{a_n-ib_n}{2}=\frac{d_n}{2}\left(\frac{a_n}{d_n}-i\frac{b_n}{d_n}\right)=\frac{d_n}{2}(\cos\varphi_n+i\sin\varphi_n)=\frac{1}{2}d_n\exp(i\varphi_n). \qquad (1.2.21)$$

Видно, що для дійсних $f(t)$

$$c_{-n}=c_n^* \qquad (1.2.22)$$

(значок * означає комплексне спряження).

Зазначимо, що при поданні ряду Фур'є у формі (1.2.19) ми вводимо до розгляду

[2] В цьому підручнику уявна одиниця позначається символом *i*. Щоб відрізнити від неї змінний струм, всюди, де може виникнути сумнів, позначатимемо струм як *i(t)*.

від'ємні частоти – частоти $2\pi n/T$, яким відповідають від'ємні значення n. Для дійсних $f(t)$ амплітуди додатних і від'ємних гармонік з однаковими модулями частот однакові, а фази – протилежні (див. формулу (1.2.22). При вимірюваннях, звичайно, можна зафіксувати лише додатні частоти, що відповідають ряду Фур'є у формі (1.2.17).

1.2.5. Середня потужність періодичного сигналу

Нехай функція $f(t)$ є періодичною з періодом T. Подамо її у формі ряду Фур'є (1.2.19):

$$f(t)=\sum_{n=-\infty}^{\infty} c_n \exp\left(\frac{i2\pi n}{T}t\right). \tag{1.2.23}$$

Для підрахунку середньої потужності такого сигналу $\overline{P}$ (тут і далі горизонтальна риска над символом означає усереднення відповідної величини за часом) достатньо усереднити її за один період. Для цього запишемо миттєву потужність $f^2(t)$ як добуток двох рядів Фур'є, причому індекси підсумовування позначаємо різними літерами, після чого перетворимо добуток сум на подвійну суму:

$$\overline{P}=\frac{1}{T}\int_{-T/2}^{T/2} f^2(t)dt=\frac{1}{T}\int_{-T/2}^{T/2} f(t)\cdot f(t)dt=\frac{1}{T}\int_{-T/2}^{T/2}\left[\sum_{n=-\infty}^{\infty} c_n \exp\left(\frac{i2\pi n}{T}t\right)\right]\left[\sum_{k=-\infty}^{\infty} c_k \exp\left(\frac{i2\pi k}{T}t\right)\right]dt=$$

$$=\frac{1}{T}\sum_{n=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} c_n c_k \int_{-T/2}^{T/2}\exp\left(\frac{i2\pi n}{T}t\right)\exp\left(\frac{i2\pi k}{T}t\right)dt=\frac{1}{T}\sum_{n=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} c_n c_k \int_{-T/2}^{T/2}\exp\left[\left(\frac{i2\pi(n+k)}{T}t\right)\right]dt=$$

$$=\sum_{n=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} c_n c_k \frac{\exp\left[i\pi(n+k)\right]-\exp\left[-i\pi(n+k)\right]}{i2\pi(n+k)}=\sum_{n=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} c_n c_k \frac{\sin\left[\pi(n+k)\right]}{\pi(n+k)}=\sum_{n=-\infty}^{\infty} c_n c_{-n} \tag{1.2.24}$$

(враховано, що для цілих l вираз $\sin(\pi l)/(\pi l)$ відмінний від нуля лише при $l=0$). Нарешті, врахуємо, що $f(t)$ – дійсна функція; тоді в силу (1.2.22) $c_{-n}=c_n^*$, і остаточно маємо:

$$P=\frac{1}{T}\int_{-T/2}^{T/2} f^2(t)dt=\sum_{n=-\infty}^{\infty}|c_n|^2=\frac{1}{4}\sum_{n=0}^{\infty} d_n^{\,2}\ . \tag{1.2.25}$$

Як бачимо, середню потужність періодичного сигналу можна підрахувати як через інтеграл за часом від його інтенсивності, так і підсумувавши інтенсивності всіх гармонік, що входять до його спектру.

Контрольні питання до підрозділу 1.2

1. Що таке ортонормована система функцій?
2. Який фізичний зміст коефіцієнтів розкладу в ряд Фур'є по косинусах?
3. Як співвідносяться спектри рядів Фур'є по косинах та по експонентах з уявними показниками?
4. Які функції можна розкласти в ряд Фур'є?

Приклади розв'язання задач до підрозділу 1.2

П 1.2.1. Нехай період повторення послідовності – T, тривалість одного імпульсу – τ :

$$u(t)=\begin{cases} U_m, & |t-nT|\le \tau/2; \\ 0, & |t-T/2-nT|<(T-\tau)/2, \end{cases} \qquad n=0,\pm1,\pm2,...$$

(рис. 1.2.1). Знайти розклад цієї функції в ряд Фур'є по тригонометричних функціях.

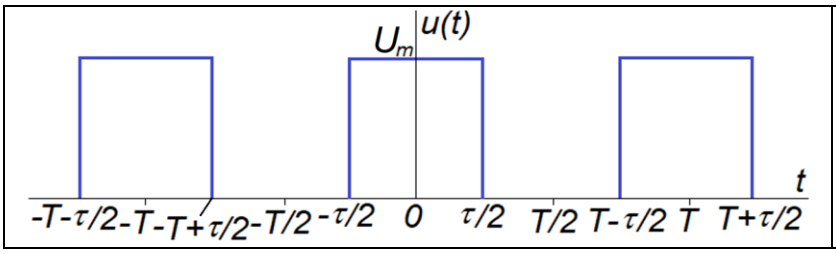


Рис. 1.2.1

Розв'язок

Ряд Фур'є (1.2.10) має вигляд

$$u(t)=c_0+\sum_{n=1}^{\infty}\left[a_n\cos\left(\frac{2\pi n}{T}t\right)+b_n\sin\left(\frac{2\pi n}{T}t\right)\right],$$

його коефіцієнти визначаються формулами (1.2.12)-(1.2.14):

$$a_n=\frac{2}{T}\int_{-T/2}^{T/2}f(t)\cos\left(\frac{2\pi n}{T}t\right)dt,$$

$$b_n=\frac{2}{T}\int_{-T/2}^{T/2}f(t)\sin\left(\frac{2\pi n}{T}t\right)dt,$$

$$c_0=\frac{1}{T}\int_{-T/2}^{T/2}f(t)dt.$$

$u(t)$ – парна функція, тому $b_n=0$.

$$c_0=\frac{1}{T}\int_{-\tau/2}^{\tau/2}U_m dt=U_m\frac{\tau}{T};$$

$$a_n=\frac{2}{T}\int_{-\tau/2}^{\tau/2}U_m\cos\left(\frac{2\pi n}{T}t\right)dt=\frac{2U_m}{T}\frac{T}{2\pi n}\sin\left(\frac{2\pi n}{T}t\right)\Bigg|_{-\tau/2}^{\tau/2}=$$

$$=\frac{2U_m}{\pi n}\sin\frac{\pi n\tau}{T}=\frac{2U_m\tau}{T}\frac{\sin(\pi n\tau/T)}{\pi n\tau/T}=\frac{2U_m\tau}{T}\operatorname{sinc}(\pi n\tau/T),$$

де $\operatorname{sinc}x=\sin x/x$.

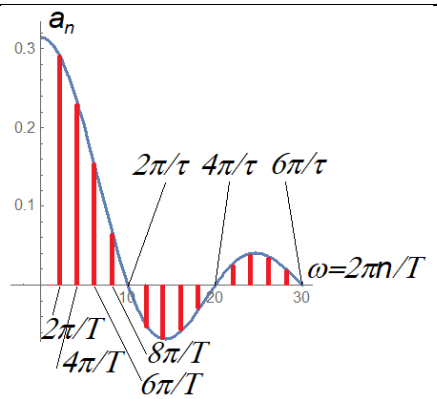


Рис. 1.2.2. Спектр періодичної послідовності прямокутних імпульсів

Спектр періодичної послідовності прямокутних імпульсів поданий на рис. 1.2.2. Видно, що $a_n \to 0$ при $n \to \pm\infty$. Віддаль за частотою між двома сусідніми лініями в спектрі – $2\pi/T$, тобто при $T \to \infty$ спектр стає неперервним.

П 1.2.2. Ідеалізований діод має кусково-лінійну вольт-амперну характеристику

$$i(t) = \begin{cases} u(t)/R, & u(t) \geq 0; \\ 0, & u(t) < 0. \end{cases}$$

На вхід подається косинусоїда й постійний зсув напруги:

$$u(t) = U_0 + U_m \cos \omega t$$

(рис. 1.2.3).

Сигнал на виході може бути поданий у вигляді ($\theta = \omega\tau$ – так званий кут відтину, де 2τ – тривалість імпульсу струму):

$$i(t) = \begin{cases} I_m(\cos \omega t - \cos\theta), & \cos \omega t \geq \cos\theta; \\ 0, & \cos \omega t < \cos\theta. \end{cases}$$

Знайти амплітуди нульової та першої гармонік такого сигналу.

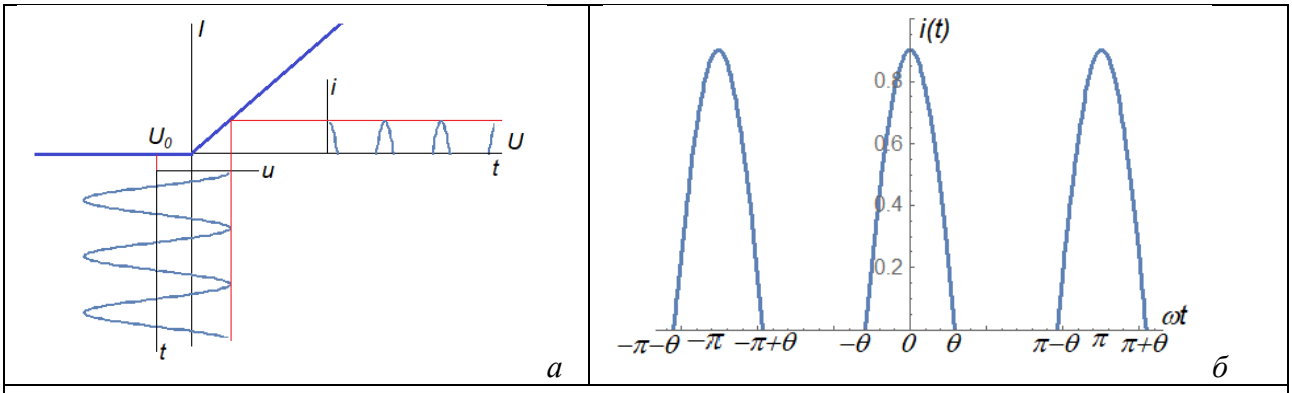


Рис. 1.2.3. Проходження сигналу через ідеалізований діод: а – загальна схема; б – вихідний сигнал

Розв'язок

Функція $u(t)$ – парна, тому $b_n = 0$. Коефіцієнти a_n, c_0 знаходимо за формулами (1.2.12), (1.2.14):

$$c_0 = \frac{1}{T}\int_{-T/2}^{T/2} f(t)dt = \frac{I_m}{\omega T}\int_{-\theta}^{\theta}\left[\cos(\omega t) - \cos\theta\right]d(\omega t) = \frac{I_m}{2\pi}\left[\sin(\omega t) - \omega t\cos\theta\right]\Big|_{-\theta}^{\theta} = \frac{I_m}{\pi}(\sin\theta - \theta\cos\theta).$$

$$a_1 = \frac{2}{T}\int_{-T/2}^{T/2} f(t)\cos\left(\frac{2\pi}{T}t\right)dt = \frac{2I_m}{\omega T}\int_{-\theta}^{\theta}\left[\cos(\omega t) - \cos\theta\right]\cos\left(\frac{2\pi}{T}t\right)d(\omega t) =$$

$$= \frac{I_m}{\pi}\int_{-\theta}^{\theta}(\cos x - \cos\theta)\cos x dx = \frac{I_m}{\pi}\int_{-\theta}^{\theta}\left(\cos^2 x - \cos\theta\cos x\right)dx = \frac{I_m}{\pi}\int_{-\theta}^{\theta}\left(\frac{1}{2} + \frac{1}{2}\cos 2x - \cos\theta\cos x\right)dx =$$

$$= \frac{I_m}{\pi}\left(\frac{1}{2}x + \frac{1}{4}\sin 2x - \cos\theta\sin x\right)_{-\theta}^{\theta} = \frac{I_m}{\pi}\left(\theta + \frac{1}{2}\sin 2\theta - 2\sin\theta\cos\theta\right) = \frac{I_m}{\pi}\left(\theta + \frac{1}{2}\sin 2\theta - \sin 2\theta\right) =$$

$$= \frac{I_m}{\pi}\left(\theta - \frac{1}{2}\sin 2\theta\right) = \frac{I_m}{\pi}(\theta - \sin\theta\cos\theta).$$

Максимальне значення вихідного струму – $I_{\max} = I_m(1 - \cos\theta)$.

Величини

$$\alpha_0 = \frac{c_0}{I_{\max}} = \frac{\sin\theta - \theta\cos\theta}{\pi\left(1-\cos\theta\right)}$$

та

$$\alpha_1 = \frac{a_1}{I_{\max}} = \frac{\theta - \cos\theta\sin\theta}{\pi\left(1-\cos\theta\right)})$$

відомі в літературі як коефіцієнти Берга.

Задачі до підрозділу 1.2

1.2.1. Розкласти послідовність прямокутних імпульсів (рис. 1.2.1) у ряд Фур'є по експонентах з уявними показниками.

1.2.2. Яка частка потужності зосереджена

а) в постійній складовій;

б) у перших трьох гармоніках

спектру послідовності прямокутних імпульсів (рис. 1.2.1)?

1.2.3. Розкласти в ряд Фур'є по тригонометричних функціях таку періодичну послідовність імпульсів:

$$u\left(t\right) = \begin{cases} 2tu_m/\tau, & \left|t-nT\right| \le \tau/2; \\ 0, & \left|t-T/2-nT\right| < \left(T-\tau\right)/2, \end{cases} \qquad n = 0, \pm 1, \pm 2, ... \qquad \tau < T .$$

Підказка: за потреби скористатись формулою інтегрування по частинах.

1.2.4. Побудувати графіки залежності коефіцієнтів Берга від кута відсічки.

1.3. Інтеграли Фур'є

Ми бачили, що для знаходження спектру періодичного сигналу його треба розкласти в ряд Фур'є. Для знаходження спектру сигналу зі скінченною енергією (реально – сигналу скінченної тривалості) користуються інтегральним перетворенням Фур'є.

1.3.1. Інтеграл Фур'є

Періодична функція $f(t)$ з періодом T при $T \to \infty$ стає аперіодичною. При цьому віддаль між сусідніми гармоніками в її спектрі (див. вище п. 1.2) зменшується до нуля, і ряд Фур'є (у формі (1.2.19)-(1.2.20)) переходить в інтеграл Фур'є. Визначимо інтеграл Фур'є таким чином. Нехай однозначна неперіодична функція $f(t)$ є скінченною при довільних t і задовольняє умову

$$\int_{-\infty}^{\infty} \left| f^2(t) \right| dt < \infty \tag{1.3.1}$$

(енергія сигналу скінченна). Тоді вона може бути подана у вигляді

$$f(t) = \int_{-\infty}^{\infty} s(\omega) \exp(i\omega t)\, d\omega, \tag{1.3.2}$$

$$s(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) \exp(-i\omega t)\, dt, \tag{1.3.3}$$

де $s(\omega)$ – (комплексний) спектр функції.

Видно, що неперіодичні функції мають неперервний спектр, який простягається також і в область від'ємних частот.

Для парних дійсних функцій ($f(-t) = f(t)$) спектр є дійсним і парним ($s(-\omega) = s(\omega)$; в цьому випадку внесок до інтегралу в симетричних межах дає тільки парна частина $\exp(-i\omega t)$, тобто $\cos \omega t$), а для непарних дійсних функцій ($f(-t) = -f(t)$) – уявним і непарним ($s(-\omega) = -s(\omega)$). У загальному випадку для дійсних функцій

$$s(-\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) \exp(i\omega t)\, dt = s^*(\omega), \tag{1.3.4}$$

тобто модуль спектру є парною функцією частоти.

Як і при розкладанні в ряд Фур'є по експонентах з уявними показниками, в експерименті можна спостерігати лише додатні частоти.

Зазначимо, що в літературі зустрічаються й інші форми запису інтегрального перетворення Фур'є:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} s(\omega) \exp(i\omega t)\, d\omega, \qquad s(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-i\omega t)\, dt \tag{1.3.5}$$

та

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} s(\omega) \exp(i\omega t)\, d\omega, \qquad s(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) \exp(-i\omega t)\, dt. \tag{1.3.6}$$

Видно, що вони відрізняються числовими множниками при визначенні спектрів. Надалі ми (якщо явно не вказано інше) будемо користуватися перетворенням Фур'є у формі (1.3.2)-(1.3.3).

1.3.2. Властивості спектрів аперіодичних функцій

Нагадаємо ще раз, що для знаходження спектру аперіодичної функції її потрібно подати у формі інтегралу Фур'є.

- Адитивність: функції $af_1(t)+bf_2(t)$ відповідає спектр $as_1(\omega)+bs_2(\omega)$, де $s_{1,2}(\omega)$ спектри відповідних функцій $f_{1,2}(t)$.

- Зміна масштабу часу: функції $f(at)$ відповідає спектр

$$s_1(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty} f(at)\exp(-i\omega t)dt=\frac{1}{2\pi}\frac{1}{a}\int_{-\infty}^{\infty} f(at)\exp\left(-i\frac{\omega}{a}at\right)d(at)=$$

$$=\|at=t'\|=\frac{1}{a}\frac{1}{2\pi}\int_{-\infty}^{\infty} f(t')\exp\left(-i\frac{\omega}{a}t'\right)dt'=\frac{1}{a}s\left(\frac{\omega}{a}\right), \qquad (1.3.7)$$

де $s(\omega)$ – спектр функції $f(t)$.

- Зсув часу: функції $f(t-t_0)$ відповідає спектр $s(\omega)\exp(-i\omega t_0)$, тобто в спектрі з'являється додатковий зсув фази, пропорційний частоті й часу зсуву сигналу (як і раніше, $s(\omega)$ – спектр функції $f(t)$).

- Спектр похідної може бути знайдений безпосередньо з формули (1.3.2) шляхом диференціювання за часом. Маємо:

$$\frac{d}{dt}f(t)=\int_{-\infty}^{\infty}(i\omega)s(\omega)\exp(i\omega t)d\omega,$$

тобто спектр похідної $s_1(\omega)$ пов'язаний зі спектром функції $s(\omega)$ співвідношенням

$$s_1(\omega)=i\omega s(\omega). \qquad (1.3.8)$$

- Спектр інтегралу. Абсолютно аналогічний розрахунок приводить до формули для спектру інтегралу $s_2(\omega)$:

$$s_2(\omega)=\frac{s(\omega)}{i\omega}. \qquad (1.3.9)$$

1.3.3. $\delta-$функція Дірака і функція Хевісайда

Надалі (зокрема, при розрахунку спектрів) ми часто матимемо справу з так званою $\delta-$функцією Дірака.

Початкове (неповне) уявлення про цю функцію, що позначається $\delta(x)$, дає формула

$$\delta(x)=\begin{cases}0, x\neq 0;\\ \infty, x=0.\end{cases} \qquad (1.3.10)$$

Акуратніше можна визначити $\delta-$функцію зі співвідношення

$$\delta(x-x_0)=\lim_{\Delta x\to 0} f(x_0,\Delta x), \qquad (1.3.11)$$

де

$$f(x_0,\Delta x)=\begin{cases}\dfrac{1}{\Delta x},\ x\in\left[x_0-\dfrac{\Delta x}{2},x_0+\dfrac{\Delta x}{2}\right];\\ 0,\ x\overline{\in}\left[x_0-\dfrac{\Delta x}{2},x_0+\dfrac{\Delta x}{2}\right]\end{cases} \qquad (1.3.12)$$

(рис. 1.3.1).

З останнього співвідношення зрозуміло, що площа під графіком $\delta-$функції дорівнює одиниці:

$$\int_{-\infty}^{\infty}\delta(x)dx=1. \qquad (1.3.13)$$

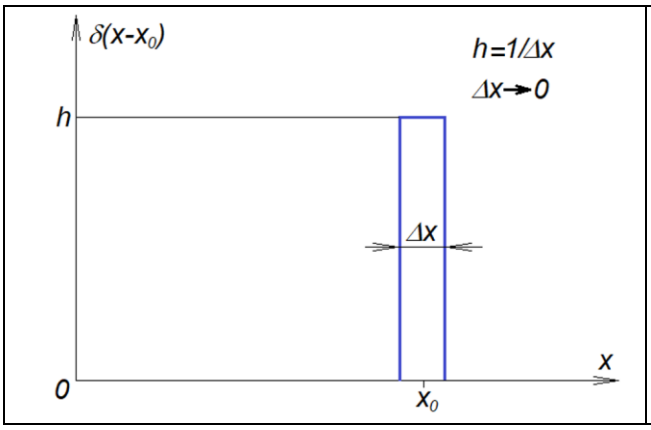


Рис. 1.3.1. До визначення δ -функції Дірака

Справедлива також формула

$$\int_{-\infty}^{\infty} f(x)\delta(x-x_0)dx = f(x_0). \tag{1.3.14}$$

Її складніша версія має вигляд:

$$\int_{a}^{b} f(x)\delta(x-x_0)dx = \begin{cases} f(x_0), a < x_0 < b; \\ \frac{1}{2} f(x_0), x_0 = a \cup x_0 = b; \\ 0, x_0 < a \cup x_0 > b. \end{cases} \tag{1.3.15}$$

Спектр $\delta-$функції:

$$s(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \delta(t)\exp(-i\omega t)dt = \frac{1}{2\pi}, \tag{1.3.16}$$

тобто цей спектр є рівномірним, і, отже, нескінченно широким. Зворотне перетворення Фур'є показує ще одне можливе представлення $\delta-$функції:

$$\delta(t) = \int_{-\infty}^{\infty} s(\omega)\exp(i\omega t)d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(i\omega t)d\omega = \frac{1}{\pi}\int_{0}^{\infty} \cos\omega t\, d\omega \tag{1.3.17}$$

(враховано, що інтеграл від непарної функції в симетричних межах дорівнює нулеві). З останньої формули випливає, що $\delta-$функція парна:

$$\delta(-t) = \delta(t). \tag{1.3.18}$$

Справедливі також співвідношення, що випливають із формули (1.3.15):

$$\delta(at) = \frac{1}{a}\delta(t); \tag{1.3.19}$$

$$\int_{-\infty}^{x} \delta(x-x_0)dx \equiv \Theta(x-x_0) = \begin{cases} 0, & x < x_0; \\ 1/2, & x = x_0; \\ 1, & x > x_0, \end{cases} \tag{1.3.20}$$

де $\theta(x-x_0)$ – так звана функція Хевісайда (її інколи також позначають $1(x-x_0)$ – див. рис. 1.3.2).

Відповідно

$$\frac{d}{dx}\Theta(x-x_0) = \delta(x-x_0). \tag{1.3.21}$$

Справді, можна вважати, що

$$\frac{d}{dx}\Theta(x) = \lim_{\Delta x\to 0}\frac{\Theta(x+\Delta x/2)-\Theta(x-\Delta x/2)}{\Delta x} = \begin{cases} 1/\Delta x, & x \le |\Delta x/2|; \\ 0, & x > |\Delta x/2| \end{cases} \equiv \delta(x). \tag{1.3.22}$$

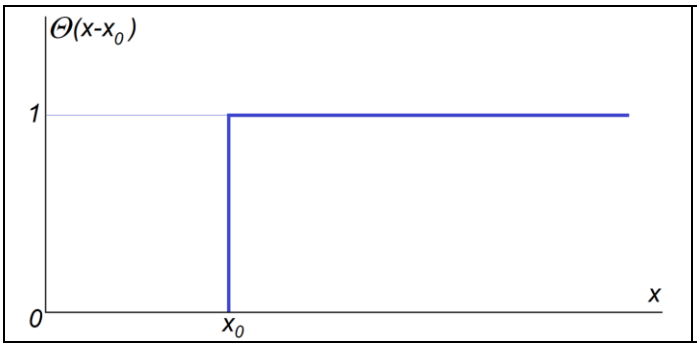


Рис. 1.3.2. Графік функції Хевісайда

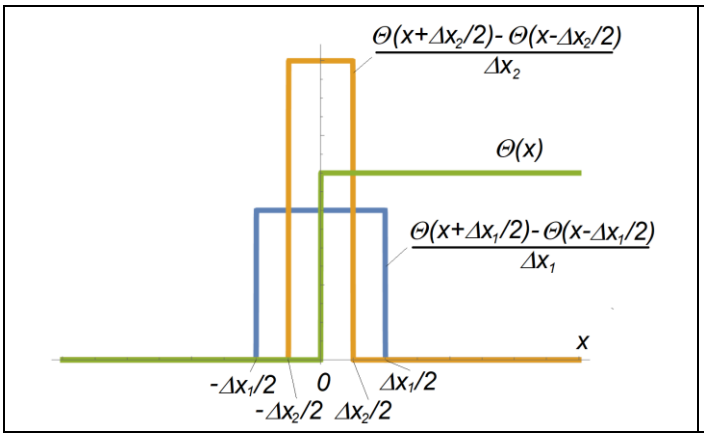


Рис. 1.3.3. До знаходження зв'язку між функцією Хевісайда і δ – функцією Дірака

Під знаком граничного переходу стоїть функція, подана на рис. 1.3.3. Площа, обмежена графіком цієї функції, не залежить від Δx і дорівнює одиниці.

1.3.4. Енергетичний спектр. Формули Парсеваля і Рейлі

Розглядатимемо й далі сигнал обмеженої тривалості. Нехай $f(t)$ – струм через резистор з одиничним опором (дійсна функція часу). Тоді миттєва потужність, яка виділяється на ньому, буде $f^2(t)$, а повна енергія –

$$A = \int_{-\infty}^{\infty} f^2(t)\,dt\,. \tag{1.3.23}$$

Перетворимо цей вираз з урахуванням формули (1.3.3), помінявши порядок інтегрування:

$$A = \int_{-\infty}^{\infty} f(t)\,dt \int_{-\infty}^{\infty} s(\omega)\exp(i\omega t)\,d\omega = \int_{-\infty}^{\infty} s(\omega)\,d\omega \int_{-\infty}^{\infty} f(t)\exp(i\omega t)\,dt =$$
$$= 2\pi \int_{-\infty}^{\infty} s(\omega)s^*(\omega)\,d\omega = 2\pi \int_{-\infty}^{\infty} W(\omega)\,d\omega, \tag{1.3.24}$$

де

$$W(\omega) = s(\omega)s^*(\omega) = |s(\omega)|^2 \tag{1.3.25}$$

– енергетичний спектр сигналу.

Рівність

$$A = \int_{-\infty}^{\infty} f^2(t)\,dt = 2\pi \int_{-\infty}^{\infty} W(\omega)\,d\omega \tag{1.3.26}$$

відома як співвідношення (формула) Парсеваля. Зазначимо, що ми вже отримували аналогічне співвідношення для сигналу з дискретним спектром (див. п. 1.2.5). Як і потужність для

періодичних сигналів, повну енергію сигналів зі скінченною енергією можна підрахувати двома способами – інтегруючи інтенсивність за часом або підсумовуючи (інтегруючи) енергетичний спектр по частотах.

Нехай тепер $i(t)$ – струм через деякий (нелінійний) елемент, $u(t)$ – напруга на ньому (обидві функції дійсні й допускають розкладання в інтеграл Фур'є), $i(\omega)$ та $u(\omega)$ – відповідні спектри. В цьому випадку для енергії, що виділиться на (нелінійному) елементі, можна записати:

$$A=\int_{-\infty}^{\infty} i(t)u(t)dt=\int_{-\infty}^{\infty} dt\int_{-\infty}^{\infty} i(\omega)\exp(i\omega t)d\omega\int_{-\infty}^{\infty} u(\omega')\exp(i\omega' t)d\omega'=$$
$$=\int_{-\infty}^{\infty} i(\omega)d\omega\int_{-\infty}^{\infty} u(\omega')d\omega'\int_{-\infty}^{\infty}\exp\left[i(\omega+\omega')t\right]dt \qquad (1.3.27)$$

Врахуємо, що, відповідно до (1.3.17),

$$\int_{-\infty}^{\infty}\exp(ikx)dx=2\pi\delta(x),$$

тоді

$$A=\int_{-\infty}^{\infty} i(\omega)d\omega\int_{-\infty}^{\infty} u(\omega')d\omega' 2\pi\delta(\omega+\omega')=2\pi\int_{-\infty}^{\infty} i(\omega)u(-\omega)d\omega=2\pi\int_{-\infty}^{\infty} i(\omega)u^*(\omega)d\omega. \qquad (1.3.28)$$

Співвідношення

$$A=\int_{-\infty}^{\infty} i(t)u(t)dt=2\pi\int_{-\infty}^{\infty} i(\omega)u^*(\omega)d\omega \qquad (1.3.29)$$

відоме як формула Рейлі (застаріла транскрипція – Релей). Формула Парсеваля, очевидно, є її частинним випадком.

1.3.5. Спектр потужності

Розглядатимемо тепер неперіодичні сигнали $f(t)$ необмеженої тривалості. Строго кажучи, вони не задовольняють умові розкладання в інтеграл Фур'є (див. формулу (1.3.1)). Вважатимемо, що середня потужність таких сигналів залишається обмеженою і визначається співвідношенням:

$$\overline{P}=\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} f^2(t)dt. \qquad (1.3.30)$$

На проміжку тривалістю T сигнал $f(t)$ несе обмежену енергію, і його можна розкласти в спектр:

$$s_T(\omega)=\frac{1}{2\pi}\int_{-T/2}^{T/2} f(t)\exp(-i\omega t)dt,\quad f_T(t)=\int_{-\infty}^{\infty} s_T(\omega)\exp(i\omega t)d\omega=\begin{cases} f(t), & |t|\le T/2;\\ 0, & |t|\le T/2.\end{cases} \qquad (1.3.31)$$

Для обмеженого в часі сигналу $f_T(t)$ можна записати формулу Парсеваля у формі:

$$A=\int_{-T/2}^{T/2} f^2(t)dt=2\pi\int_{-\infty}^{\infty}\left|s_T(\omega)\right|^2 d\omega. \qquad (1.3.32)$$

Розділимо обидві частини на T і спрямуємо T до нескінченності. Отримаємо вираз для середньої потужності:

$$\bar{P}=\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}f^2(t)\,dt=2\pi\lim_{T\to\infty}\frac{1}{T}\int_{-\infty}^{\infty}\left|s_T(\omega)\right|^2 d\omega=2\pi\int_{-\infty}^{\infty}\lim_{T\to\infty}\frac{\left|s_T(\omega)\right|^2}{T}d\omega=2\pi\int_{-\infty}^{\infty}P(\omega)\,d\omega,\quad (1.3.33)$$

де величина

$$P(\omega)=\lim_{T\to\infty}\frac{\left|s_T(\omega)\right|^2}{T} \qquad (1.3.34)$$

відома в літературі як спектр потужності.

Формулу (1.3.33) можна розглядати як деякий аналог формули Парсеваля для нескінченних неперіодичних сигналів: бачимо, що середню потужність можна розрахувати як шляхом усереднення за часом миттєвої потужності, так і шляхом інтегрування спектру потужності за частотами.

1.3.6. Обговорення поняття спектру

Спробуємо застосувати інтегральне перетворення Фур'є до гармоніки $\exp(i\omega_0 t)$ (що, строго кажучи, некоректно). Отримаємо:

$$s(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(i\omega_0 t)\exp(-i\omega t)\,dt=\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left[i(\omega_0-\omega)t\right]dt=\delta(\omega-\omega_0). \qquad (1.3.35)$$

Отримана нескінченна амплітуда спектральної складової є результатом того, що гармоніка не належить до сигналів з обмеженою енергією. В той же час при розкладенні цієї самої функції в ряд Фур'є по експонентах з уявними показниками останній, очевидно, буде складатися з одного доданка з одиничною амплітудою. Отже, для періодичних сигналів та сигналів з обмеженою енергією поняття спектру відрізняється.

Якщо замість комплексної функції $\exp(i\omega_0 t)$ взяти суто дійсну функцію $\cos\omega_0 t$, інтегральне перетворення Фур'є дасть спектр $\left[\delta(\omega-\omega_0)+\delta(\omega+\omega_0)\right]/2$, а розкладання в ряд Фур'є по експонентах – два доданки з частотами $\pm\omega_0$ і однаковими амплітудами 1/2.

Розмірність спектру періодичного сигналу збігається з розмірністю самого сигналу. Скажемо, амплітуди в ряді Фур'є для струму вимірюються в амперах. Для сигналу з обмеженою енергією він записується як інтеграл від спектру по частотах, тобто, наприклад, для струму спектр має розмірність А·с.

Зазначимо, що поняття спектру застосовується й до таких сигналів, які не можна розкласти ні в ряд, ні в інтеграл Фур'є. Для прикладу розглянемо суму двох гармонічних сигналів з несумірними частотами (тобто з ірраціональним відношенням частот):

$$x(t)=A_1\exp(i\omega_1 t)+A_2\exp(i\omega_2 t), \qquad \omega_1/\omega_2\neq m/n, \qquad (1.3.36)$$

де m,n – довільні цілі числа. Здається очевидним, що до спектру такого сигналу входять частоти ω_1 і ω_2. В той же час через несумірність частот цей сигнал не є періодичним, тому в ряд Фур'є його розкласти не можна. Але він також не належить до сигналів з обмеженою енергією, тому розкласти його в інтеграл Фур'є також не можна.

Надалі під спектром сигналу ми будемо розуміти набір гармонік (дискретний або неперервний), сумою яких можна представити цей сигнал.

В експерименті спектри сигналів радіодіапазону вимірюються за допомогою спеціальних приладів – аналізаторів спектру. По-перше, сигнал аналізується протягом обмеженого проміжку часу, тому виміряний спектр відповідає інтегральному перетворенню Фур'є від сигналу на певному часовому інтервалі (так зване віконне перетворення Фур'є – див. нижче п. 1.5). Крім того, спектр вимірюється лише для додатних частот.

Щоб з'ясувати, який різновид спектру вимірюється в експерименті, розіб'ємо інтеграл

по частотах на дві частини і замінимо від'ємні частоти додатними:

$$f(t)=\int_{-\infty}^{\infty} s(\omega)\exp(i\omega t)\,d\omega=\left[\int_{0}^{\infty} s(\omega)\exp(i\omega t)\,d\omega+\int_{-\infty}^{0} s(\omega)\exp(i\omega t)\,d\omega\right]=$$

$$=\left[\int_{0}^{\infty} s(\omega)\exp(i\omega t)\,d\omega+\int_{0}^{\infty} s(-\omega)\exp(-i\omega t)\,d\omega\right].$$

Тепер врахуємо співвідношення $s(-\omega)=s^{*}(\omega)$ (див. формулу (1.3.4)) та $\operatorname{Re} z=\left(z+z^{*}\right)/2$. Тоді

$$f(t)=\left[\int_{0}^{\infty} s(\omega)\exp(i\omega t)\,d\omega+\int_{0}^{\infty} s^{*}(\omega)\exp(-i\omega t)\,d\omega\right]=$$

$$=\left\{\int_{0}^{\infty} s(\omega)\exp(i\omega t)\,d\omega+\int_{0}^{\infty}\left[s(\omega)\exp(i\omega t)\right]^{*}d\omega\right\}=2\left\{\int_{0}^{\infty}\operatorname{Re}\left[s(\omega)\exp(i\omega t)\right]d\omega\right\}.$$

Нарешті, розпишемо $\exp(i\omega t)$ за формулою Ейлера, введемо модуль $|s(\omega)|$ та фазу $\varphi(\omega)$ спектру $s(\omega)$ і скористаємося формулою $\cos(\alpha+\beta)=\cos\alpha\cos\beta-\sin\alpha\sin\beta$. Остаточно отримуємо:

$$f(t)=2\left[\int_{0}^{\infty} d\omega\cos\omega t\operatorname{Re} s(\omega)-\int_{0}^{\infty} d\omega\sin\omega t\operatorname{Im} s(\omega)\right]=\left\|\begin{matrix}\operatorname{Re} s(\omega)=|s(\omega)|\cos\varphi(\omega)\\ \operatorname{Im} s(\omega)=|s(\omega)|\sin\varphi(\omega)\end{matrix}\right\|=$$

$$=2\int_{0}^{\infty}|s(\omega)|\left\{\cos\omega t\cos\left[\varphi(\omega)\right]-\sin\omega t\sin\left[\varphi(\omega)\right]\right\}d\omega=2\int_{0}^{\infty}|s(\omega)|\cos\left[\omega t+\varphi(\omega)\right]d\omega. \qquad (1.3.37)$$

Отже, в експерименті вимірюється величина $|s(\omega)|$.

Контрольні питання до підрозділу 1.3

1. Коли сигнал розкладають у ряд Фур'є, а коли – в інтеграл Фур'є?
2. Які функції можна розкладати в інтеграл Фур'є?
3. Який фізичний зміст формули Парсеваля?
4. Що таке $\delta-$функція Дірака?
5. Як пов'язані між собою функція Дірака і функція Хевісайда?
6. Чому в спектрі відеоімпульсу є постійна складова, а в спектрі радіоімпульсу вона відсутня (або близька до нуля)?

Приклади розв'язання задач до підрозділу 1.3

П 1.3.1. Подати як інтеграл Фур'є прямокутний відеоімпульс:

$$u(t)=\begin{cases}U_m, & |t|\le\tau/2;\\ 0, & |t|>\tau/2.\end{cases}$$

Розв'язок

Інтеграл Фур'є (1.3.2)-(1.3.3):

$$f(t)=\int_{-\infty}^{\infty} s(\omega)\exp(i\omega t)\,d\omega,$$

$$s(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)\exp(-i\omega t)\,dt\,.$$

В даному разі

$$s(\omega)=\frac{1}{2\pi}\int_{-\tau/2}^{\tau/2} U_m\exp(-i\omega t)\,dt=U_m\left.\frac{\exp(-i\omega t)}{-i2\pi\omega}\right|_{-\tau/2}^{\tau/2}=$$

$$=\frac{U_m}{\pi\omega}\frac{\exp(-i\omega\tau/2)-\exp(i\omega\tau/2)}{-2i}=\frac{U_m}{\pi\omega}\sin\omega\tau/2=\frac{U_m\tau}{2\pi}\frac{\sin\omega\tau/2}{\omega\tau/2}=\frac{U_m\tau}{2\pi}\operatorname{sinc}(\omega\tau/2),$$

де, нагадаємо, $\operatorname{sinc} x\equiv\sin x/x$ (див. рис. 1.3.4 а). Ширина цього спектру (між двома нулями навколо головного максимуму) складає величину $\Omega=\pi/\tau$, амплітуда – $s_m=U_m\tau/2\pi$. Відповідно, добуток цих величин $\Omega s_m=U_m/2$ не залежить від τ. Так само добуток тривалості імпульсу τ на ширину спектру π/τ не залежить від τ. Це – загальна закономірність, яка не залежить від форми відеоімпульсу: ширина його спектру завжди обернено пропорційна його тривалості.

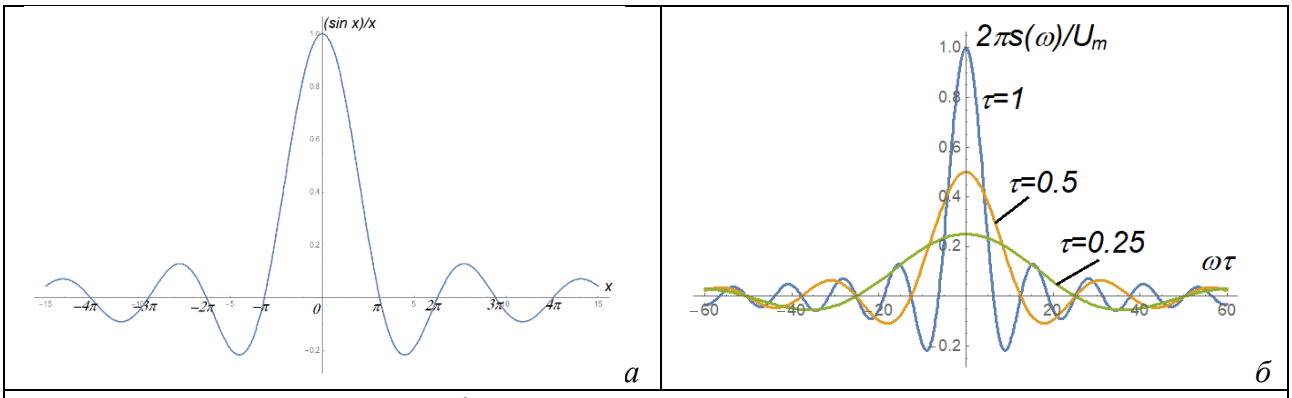


Рис. 1.3.4: а – графік функції $\operatorname{sinc} x$*; б – залежність спектру прямокутного імпульсу від його тривалості*

Порівняємо спектри одиночного прямокутного імпульсу та періодичної послідовності таких імпульсів (див. задачу П 1.2.1 та рис. 1.2.1 – 1.2.2; треба взяти до уваги, що в цій задачі розрахований спектр розкладу по косинусах, а не по експонентах з уявними показниками). В першому випадку спектр є неперервним, а його форма визначається функцією $\operatorname{sinc}(\omega\tau/2)$ з максимальним значенням $U_m\tau/2\pi$. У другому спектр є дискретним, віддаль між сусідніми лініями – $2\pi/T$, а функція $\operatorname{sinc}(\pi n\tau/T)$ з максимальним значенням $2u_m\tau/T$ виступає його обвідною. Величина $2\pi/T$ є аналогом частоти ω неперервного спектру, і в цьому сенсі форми неперервного й дискретного спектрів однакові.

П.1.3.2. Знайти спектр експоненціального імпульсу: $u(t)=\Theta(t)\exp(-\alpha t)$, (рис. 1.3.5), де $\Theta(t)$ – функція Хевісайда (одиничний стрибок) – див. п. 1.3.3 та рис. 1.3.2.

Розв'язок

$$s(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty} u(t)\exp(-i\omega t)\,dt=\frac{1}{2\pi}\int_{0}^{\infty}\exp(-i\omega t-\alpha t)\,dt=$$

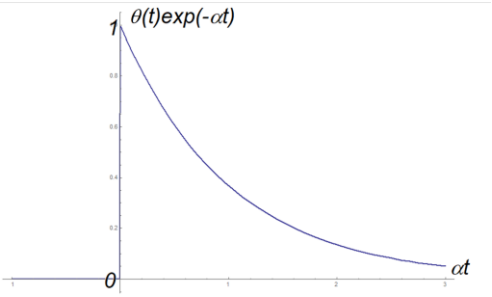


Рис. 1.3.5

$$= \left.\frac{\exp(-i\omega t - \alpha t)}{2\pi(-\alpha - i\omega)}\right|_0^\infty = \frac{1}{2\pi(\alpha + i\omega)}.$$

П 1.3.3. Знайти спектр функції Хевісайда.

Розв'язок.

Функція Хевісайда не є функцією з обмеженою енергією, тому звичайні формули для розкладання в інтеграл Фур'є до неї незастосовні. Через те знайдемо її спектр як границю спектру з попередньої задачі при $\alpha \to 0$.

Будь-яку функцію $f(x)$ можна подати як суму її парної $f_{even}(x)$ та непарної $f_{odd}(x)$ частин:

$$f(x) = f_{even}(x) + f_{odd}(x),$$

де

$$f_{even}(x) = \frac{1}{2}\left[f(x) + f(-x)\right]$$

$$f_{odd}(x) = \frac{1}{2}\left[f(x) - f(-x)\right]$$

В даному разі

$$u_{even}(t) = \frac{1}{2}\left[\Theta(t)\exp(-\alpha t) + \Theta(-t)\exp(\alpha t)\right],$$

$$u_{odd}(t) = \frac{1}{2}\left[\Theta(t)\exp(-\alpha t) - \Theta(-t)\exp(\alpha t)\right]$$

(див. рис. 1.3.6).

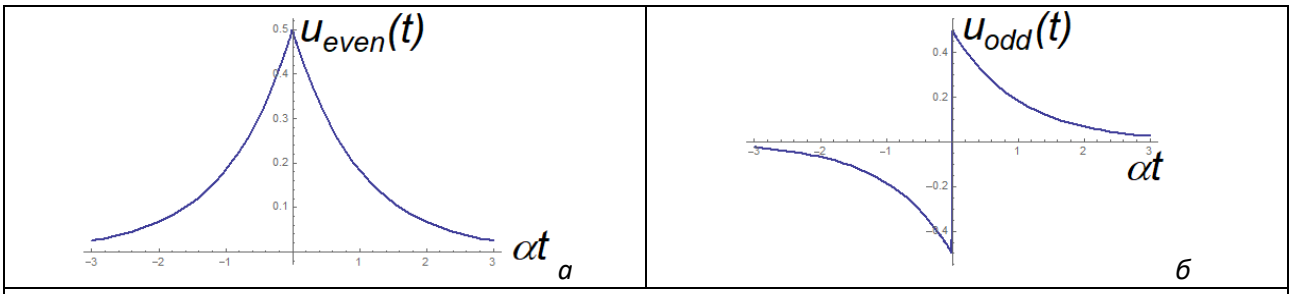


Рис. 1.3.6. Парна (а) та непарна (б) частини функції $u(t) = \Theta(t)\exp(-\alpha t)$

Можна записати функцію Хевісайда у формі:

$$\Theta(t) = \lim_{\alpha \to 0}\left[u_{even}(t) + u_{odd}(t)\right] = \frac{1}{2} + \lim_{\alpha \to 0}\int_{-\infty}^{\infty} s_{odd}(\omega)\exp(i\omega t)\,d\omega.$$

Спектр $s_{odd}(\omega)$ може бути поданий у вигляді

$$s_{odd}(\omega) = \mathrm{Im}[s(\omega)] = -\frac{i\omega}{2\pi(\omega^2 + \alpha^2)},$$

де спектр $s(\omega)$ визначається формулою (1.3.40) і враховане співвідношення (1.3.43).

Поміняємо місцями порядок інтегрування і граничного переходу в другому доданку (1.3.44) і врахуємо формулу (1.3.45). Отримаємо:

$$\Theta(t) = \frac{1}{2} - \frac{i}{2\pi}\int_{-\infty}^{\infty}\exp(i\omega t)\lim_{\alpha \to 0}\frac{\omega}{\alpha^2 + \omega^2}\,d\omega = \frac{1}{2} + \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\exp(i\omega t)}{i\omega}\,d\omega.$$

Останню формулу й можна інтерпретувати як спектральне представлення функції Хевісайда. Постійна складова відбиває той факт, що функція Хевісайда несе нескінченну енер-

гію.

Задачі до підрозділу 1.3

1.3.1. Знайти спектр та його ширину для гауссівського відеоімпульсу $u(t) = u_m \exp\left[-(t/\tau)^2\right]$.

Підказка: скористатися формулою $\int\limits_{-\infty}^{\infty} \exp\left(-x^2\right)\cos(kx)\,dx = \sqrt{\pi}\exp\left(-k^2/4\right)$.

1.3.2. Знайти спектр та його ширину для трикутного відеоімпульсу

$$u(t) = \begin{cases} u_m\left(1 - 2|t|/\tau\right), |t| \le \tau/2; \\ 0, \quad |t| > \tau/2. \end{cases}$$

1.3.3. Знайти спектр та його ширину для параболічного відеоімпульсу:

$$u(t) = \begin{cases} u_m\left[1 - (2t/\tau)^2\right], |t| \le \tau/2; \\ 0, \quad |t| > \tau/2. \end{cases}$$

1.3.4. Яка частка енергії зосереджена в центральній пелюстці спектру прямокутного імпульсу?

1.3.5. Знайти спектр функції $u(t) = \Theta(t)\exp(-\alpha t)\cos\omega_0 t$.

1.3.6. Знайти спектр функції $u(t) = \Theta(t)\cos\omega_0 t$.

1.4. Модульовані сигнали: спектри та векторні діаграми

У цьому розділі ми розглянемо спектри та векторні діаграми найбільш поширених модульованих гармонічних сигналів – амплітудно-модульованих та фазово-модульованих.

1.4.1. Спектр амплітудно-модульованого сигналу з гармонічною модуляцією

Розглянемо амплітудно-модульований сигнал з гармонічною модуляцією:

$$u(t) = u_m (1 + m_a \cos \Omega t) \cos \omega_0 t . \qquad (1.4.1)$$

(рис. 1.4.1 а; безрозмірну величину m_a називають глибиною амплітудної модуляції). Знайдемо його спектр:

$$u(t) = u_m \cos \omega_0 t + m_a u_m \cos \Omega t \cos \omega_0 t = \\ = u_m \cos \omega_0 t + \frac{1}{2} m_a u_m \left[\cos(\omega_0 + \Omega) t + \cos(\omega_0 - \Omega) t \right], \qquad (1.4.2)$$

(рис. 1.4.1 б). Видно, що ширина цього спектру складає величину $\Delta\omega = 2\Omega$. Такий сигнал (1.4.1) можна записати також через експоненти з уявними показниками:

$$u(t) = \frac{1}{2} u_m \{ \exp(i\omega_0 t) + \exp(-i\omega_0 t) + (m_a/2) \times \\ \times \left[\exp\left[i(\omega_0 + \Omega) t \right] + \exp\left[-i(\omega_0 + \Omega) t \right] + \exp\left[i(\omega_0 - \Omega) t \right] + \exp\left[-i(\omega_0 - \Omega) t \right] \right] \}. \qquad (1.4.3)$$

Таке представлення, як і раніше, включає до розгляду від'ємну область частот (рис. 1.4.1 в).

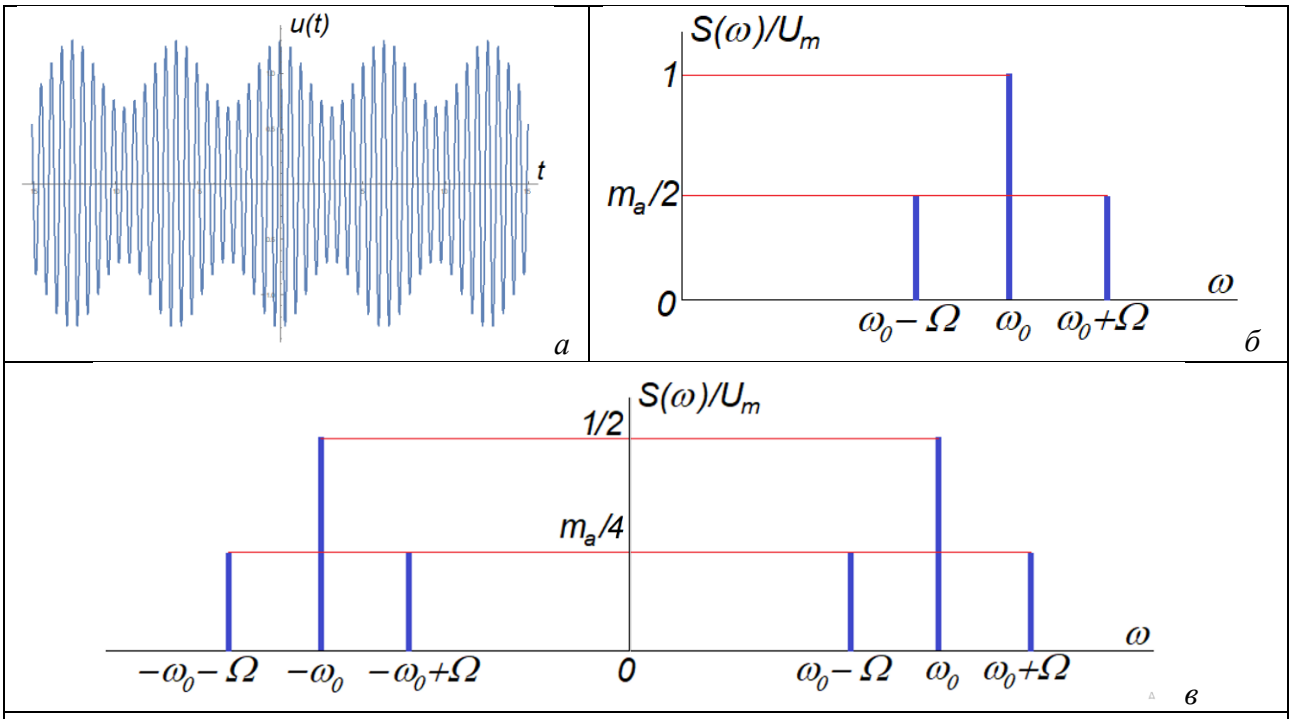


Рис. 1.4.1. Амплітудно-модульований сигнал (а) та його спектр для додатних частот (б) та для довільних частот (в)

Строго кажучи, наведений розрахунок не завжди відповідає саме розкладу в ряд Фур'є. Справді, такий розклад можливий лише для періодичних сигналів, а амплітудно-модульоване коливання (3.1) буде періодичним лише тоді, коли відношення частоти-носія й частоти модуляції буде раціональним: $\omega_0/\Omega = m/n$, де m, n – натуральні числа. В цьому випадку період $2\pi m/\omega_0 = 2\pi n/\Omega$ буде спільним для обох частот. Якщо ж відношення цих частот ірраціональне, $\omega_0/\Omega \neq m/n$, відповідний сигнал уже не буде періодичним (пор. з п.

1.3.6). Тим не менше його подання у формі (1.4.2)-(1.4.3) залишається справедливим. У цьому випадку сигнал (1.4.1) буде квазіперіодичним, а гармоніки в його поданні у формі (1.4.2) – спектром, хоч він і не відповідає розкладу в ряд або в інтеграл Фур'є.

1.4.2. Векторна діаграма сигналу з гармонічною амплітудною модуляцією

Нагадаємо, що таке векторна діаграма.

Векторна діаграма – це графічне зображення напруг і струмів, які гармонічно змінюються з часом, у вигляді векторів, що обертаються проти годинникової стрілки з кутовою швидкістю, що дорівнює кутовій частоті сигналу. Довжина векторів дорівнює амплітудам у відповідному масштабі, а початкове положення задається початковою фазою, відрахованою від осі абсцис. Векторні діаграми наглядно показують фазові співвідношення між струмами і напругами в колі гармонічного струму.

Нехай ми розглядаємо дійсний сигнал вигляду $f(t) = A_m \cos(\omega t + \varphi)$. В багатьох випадках такий сигнал що зручно записувати у формі експонент з уявними показниками, $g(t) = A_m \exp[i(\omega t + \varphi)] = A \exp(i\omega t)$, де величину $A = A_m \exp(i\varphi)$ називають комплексною амплітудою і виконується співвідношення $f(t) = \mathrm{Re}\{g(t)\}$.

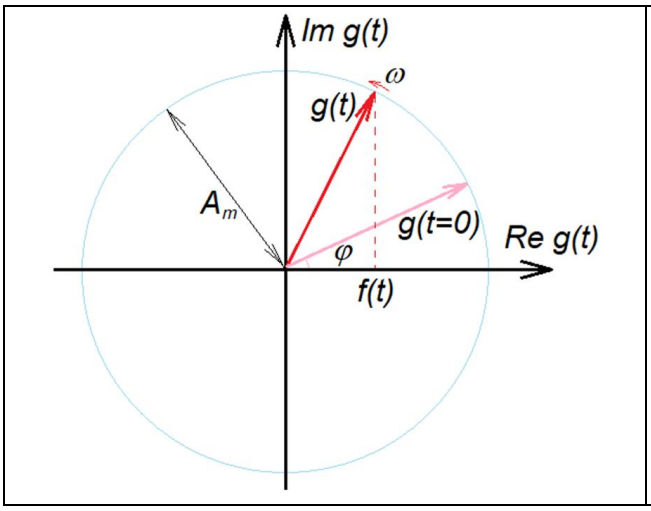


Рис. 1.4.2. До пояснення побудови векторної діаграми

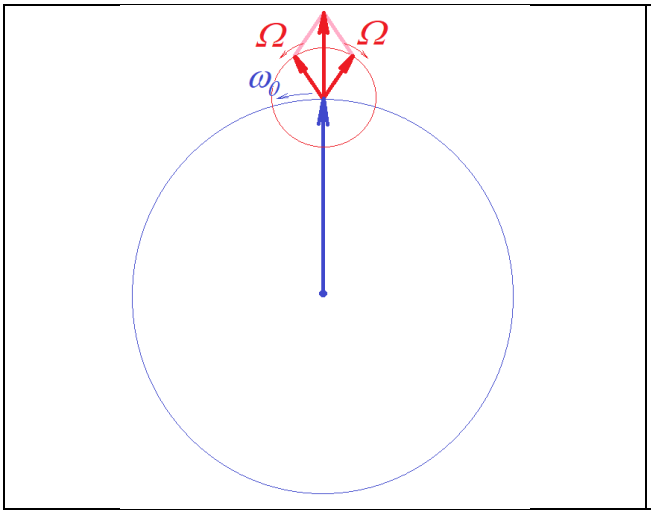


Рис. 1.4.3. Векторна діаграма сигналу з гармонічною амплітудною модуляцією

На комплексній площині $\mathrm{Re}\{g(t)\}, \mathrm{Im}\{g(t)\}$ функцію $g(t) = A_m \exp[i(\omega t + \varphi)]$ зображає вектор завдовжки A_m, кінець якого обертається навколо початку координат проти

годинникової стрілки з кутовою швидкістю ω. В момент часу $t=0$ він утворює кут φ з віссю абсцис. Очевидно, проєкція цього вектора на вісь абсцис – це функція $f(t)$ (рис. 1.4.2). Нарешті, перейдемо в систему координат, що обертається проти годинникової стрілки. В такій системі вектор, що зображає функцію $g(t)$, буде нерухомим це й буде векторна діаграма.

Звичайно на векторній діаграмі зображають кілька напруг та/або струмів, щоб наочно продемонструвати спввідношення між їхніми фазами та/або амплітудами.

Але зараз наша мета – побудувати векторну діаграму, що показує гармонічну зміну амплітуди сигналу в часі. Для цього в кінці вектора завдовжки u_m, що відповідає частоті-носію, треба намалювати два вектори, що обертаються в протилежних напрямках із частотою модуляції і утворюють однакові за величиною, але протилежні за знаком кути з вектором, що відповідає частоті-носію (рис. 1.4.3). Амплітуди цих векторів однакові й дорівнюють $m_a u_m/2$, тому їхня сума завжди паралельна вектору, що відповідає частоті-носію, а проєкція на напрямок цього вектора змінюється від $-m_a U_m$ до $m_a u_m$.

1.4.3. Спектр амплітудно-модульованого сигналу з повільною періодичною обвідною

Нехай тепер частота-носій модулюється за амплітудою сигналом з повільною (в масштабі періоду частоти-носія) обвідною, яка періодично залежить від часу:

$$u(t)=u_m\left[1+m_a(t)\right]\cos\omega_0 t, \qquad m_a(t+T)=m_a(t) \qquad \omega_0 T \gg 1 \tag{1.4.4}$$

(пор. з формулою (1.1.2); очевидно, $m_a(t)=ku_{\text{mod}}(t)$).

Розкладемо функцію $m_a(t)$ в ряд Фур'є:

$$m_a(t)=\sum_{n=-\infty}^{\infty} c_n \exp\left(i\frac{2\pi n}{T}t\right), \tag{1.4.5}$$

причому $c_n \to 0$ при $|n|>n_{\max}$ такому, що $n_{\max} \ll \omega_0 T/2\pi$, тобто ширина спектру $m_a(t)$ значно менша від частоти-носія ω_0. Тоді маємо:

$$\begin{aligned} u(t)&=\frac{1}{2}u_m\left[1+\sum_{n=-\infty}^{\infty} c_n \exp\left(i\frac{2\pi n}{T}t\right)\right]\left[\exp(i\omega_0 t)+\exp(-i\omega_0 t)\right]=\frac{1}{2}u_m\times \\ &\times\left\{\exp(i\omega_0 t)+\sum_{n=-\infty}^{\infty} c_n \exp\left[i\left(\omega_0+\frac{2\pi n}{T}\right)t\right]+\exp(-i\omega_0 t)+\sum_{n=-\infty}^{\infty} c_n \exp\left[i\left(-\omega_0+\frac{2\pi n}{T}\right)t\right]\right\}. \end{aligned} \tag{1.4.6}$$

Схематичний вигляд спектрів сигналів $m_a(t)$ і $u(t)$ подані на рис. 1.4.4 а-б.

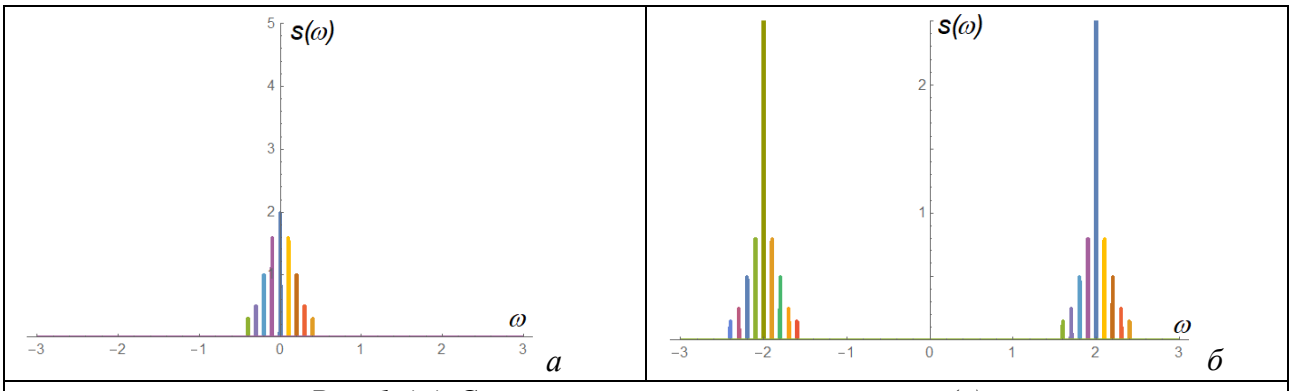


Рис. 1.4.4. Спектр низькочастотного сигналу (а) і відповідного сигналу з амплітудною модуляцією (б)

1.4.4. Співвідношення між спектрами відеоімпульсу та радіоімпульсу

Розглянемо тепер спектр амплітудно-модульованого імпульсу скінченної тривалості. Нехай $f(t)$ – відеоімпульс (немодульований імпульс), $f(t)\cos\omega_0 t$ – відповідний йому радіоімпульс (його частота-носій – ω_0, а обвідною служить відеоімпульс, див. вище п. 1.1.2). Вони подані на рис. 1.4.5 а-б. Спектр радіоімпульсу:

$$s_1(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)\cos\omega_0 t\exp(-i\omega t)\,dt=\frac{1}{2}\frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)dt\times$$
$$\times\left\{\exp\left[-i(\omega+\omega_0)t\right]+\exp\left[-i(\omega-\omega_0)t\right]\right\}=\frac{1}{2}\left[s(\omega+\omega_0)+s(\omega-\omega_0)\right], \qquad (1.4.7)$$

де $s(\omega)$ – спектр відповідного відеоімпульсу (див. рис. 1.4.5 в, г).

Зазначимо, що відеоімпульс має ненульову постійну складову, а радіомпульс її не має. Відповідно для відеоімпульсу $s(\omega=0)\neq 0$, а для радіоімпульсу $s_1(\omega=0)=0$ (за умови, що частота-носій ω_0 не потрапляє до спектру $s(\omega)$).

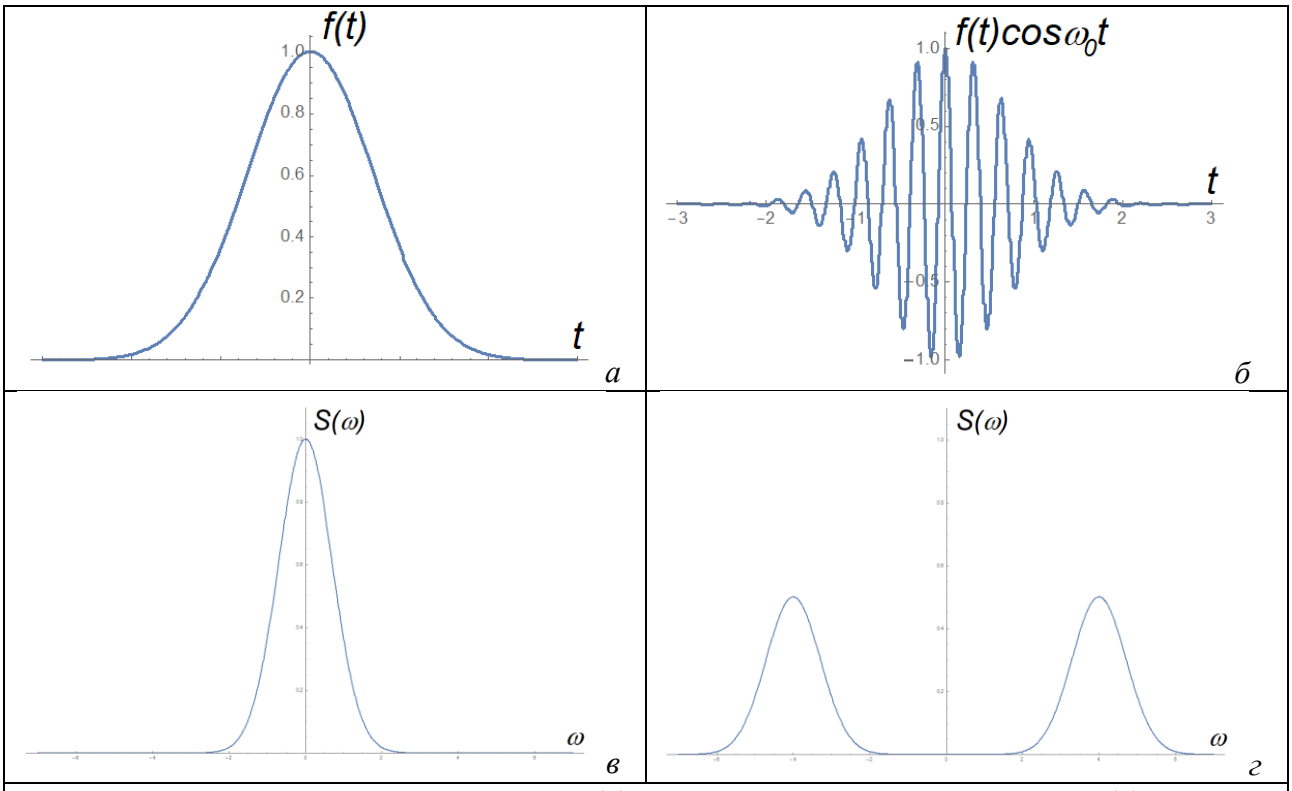


Рис. 1.4.5. Епюри відеоімпульсу (а) і відповідного йому радіоімпульсу (б) та їхні спектри (в, г)

1.4.5. Спектр фазово-модульованого сигналу з гармонічною модуляцією

Розглянемо фазово-модульований сигнал з гармонічною модуляцією на частоті Ω:

$$u(t)=u_m\sin\left(\omega_0 t+m_\varphi\sin\Omega t\right) \qquad (1.4.8)$$

(пор. з формулою (1.1.13); в даному разі $\varphi(t)=ku_{\text{mod}}(t)=m_\varphi\sin\Omega t$). Безрозмірну величину m_φ, що входить до (1.4.8), називають індексом фазової модуляції.

Представимо цей сигнал у вигляді набору гармонік:

$$u(t)=u_m\sin\left(\omega_0 t+m_\varphi\sin\Omega t\right)=u_m\left[\sin\omega_0 t\cos\left(m_\varphi\sin\Omega t\right)+\cos\omega_0 t\sin\left(m_\varphi\sin\Omega t\right)\right]. \qquad (1.4.9)$$

Розкладемо періодичні функції $\cos\left(m_\varphi\sin\Omega t\right)$ та $\sin\left(m_\varphi\sin\Omega t\right)$ з періодом $T=2\pi/\Omega$ (рис. 1.4.6) у ряд Фур'є, скориставшись довідником з математики:

$$\cos\left(m_\varphi\sin\Omega t\right)=J_0\left(m_\varphi\right)+2\sum_{n=1}^{\infty}J_{2n}\left(m_\varphi\right)\cos 2n\Omega t\,; \qquad (1.4.10)$$

$$\sin\left(m_\varphi\sin\Omega t\right)=2\sum_{n=1}^{\infty}J_{2n-1}\left(m_\varphi\right)\sin\left(2n-1\right)\Omega t\,. \qquad (1.4.11)$$

Тут $J_n\left(x\right)$ – спеціальні функції, які називаються функціями Бесселя порядку n аргументу x (рис. 1.4.7).

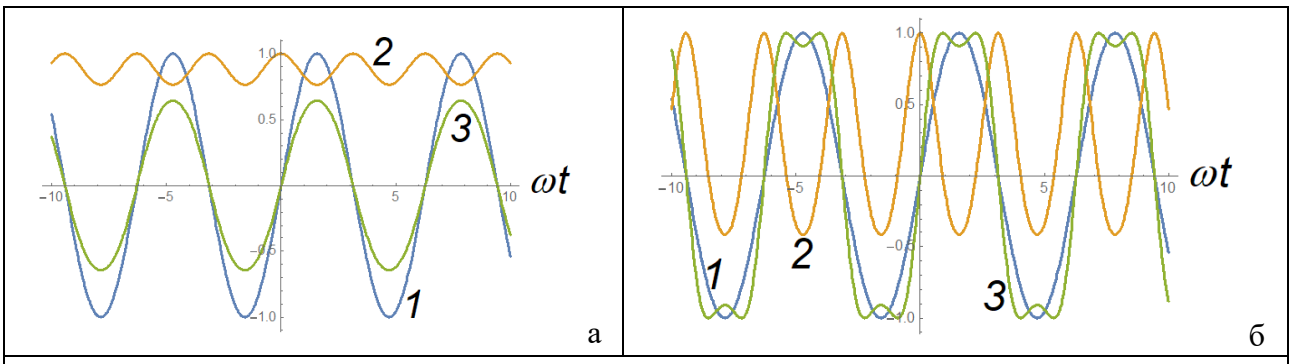


Рис. 1.4.6. Графіки функцій $\sin\Omega t$ (1), $\cos\left(m_\varphi\sin\Omega t\right)$ (2) та $\sin\left(m_\varphi\sin\Omega t\right)$ (3) для випадків $m_\varphi=0.7$ *(а) та* $m_\varphi=2$ *(б)*

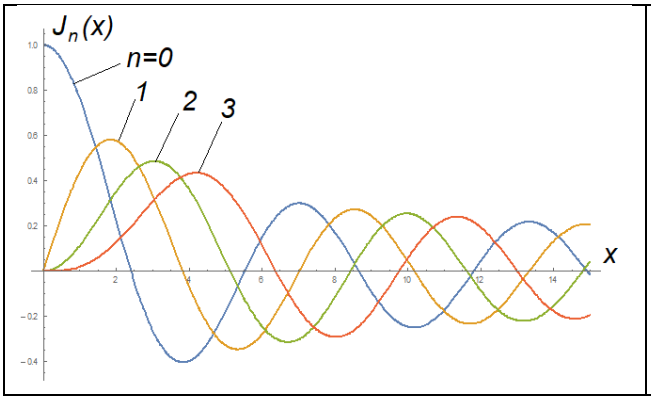


Рис. 1.4.7. Графіки функцій Бесселя дійсного аргументу

Після підстановки (1.4.10)-(1.4.11) до формули (1.4.9) отримаємо:

$$u(t)=u_m\left\{\sin\omega_0 t\left[J_0\left(m_\varphi\right)+2\sum_{n=1}^{\infty}J_{2n}\left(m_\varphi\right)\cos 2n\Omega t\right]+\cos\omega_0 t\left[2\sum_{n=1}^{\infty}J_{2n-1}\left(m_\varphi\right)\sin\left(2n-1\right)\Omega t\right]\right\}=$$

$$=u_m\left\{J_0\left(m_\varphi\right)\sin\omega_0 t+2\sum_{n=1}^{\infty}\left[J_{2n}\left(m_\varphi\right)\sin\omega_0 t\cos 2n\Omega t+J_{2n-1}\left(m_\varphi\right)\cos\omega_0 t\sin\left(2n-1\right)\Omega t\right]\right\}=$$

$$=u_m J_0\left(m_\varphi\right)\sin\omega_0 t+u_m\sum_{n=1}^{\infty}\left\{J_{2n}\left(m_\varphi\right)\left[\sin\left(\omega_0+2n\Omega\right)t+\sin\left(\omega_0-2n\Omega\right)t\right]+\right.$$

$$\left.+J_{2n-1}\left(m_\varphi\right)\left[\sin\left(\omega_0+\left(2n-1\right)\Omega\right)t-\sin\left(\omega_0-\left(2n-1\right)\Omega\right)t\right]\right\}=u_m\sum_{k=-\infty}^{\infty}c_k\sin\left(\omega_0+k\Omega\right)t, \qquad (1.4.12)$$

де

$$c_0=J_0\left(m_\varphi\right),\ c_k=J_k\left(m_\varphi\right),\quad k=\pm 2n\,,\qquad c_k=\pm J_k\left(m_\varphi\right),\quad k=\pm\left(2n-1\right),\qquad n=1,2,3,...$$

(1.4.13)

Таким чином, спектр фазово-модульованого сигналу містить дискретний набір частот вигляду $\omega_0 + k\Omega$, амплітуди яких, взагалі кажучи, спадають зі зростанням абсолютної величини k (рис. 1.4.8).

Функції Бесселя $J_k\left(m_\varphi\right)$ різко спадають за умови $k > m_\varphi + 1$. Тому на практиці ширина спектру фазово-модульованого сигналу може бути оцінена як $\Delta\omega = 2\left(m_\varphi + 1\right)\Omega$.

Якщо $m_\varphi << 1$, вираз для спектру різко спрощується. В цьому випадку, як видно зі щойно наведеної формули, $\Delta\omega \simeq 2\Omega$, тобто в спектрі на практиці відмінні від нуля лише коефіцієнти c_0 і $c_{\pm 1}$. Візьмемо до уваги, що при $m_\varphi << 1$ справедливі наближені формули $J_0\left(m_\varphi\right) \simeq 1$, $J_1\left(m_\varphi\right) \simeq m_\varphi / 2$. Тоді ряд (3.4) спрощується і набуває вигляду:

$$u(t) = u_m \left\{ \sin\omega_0 t + \frac{m_\varphi}{2}\left[\sin\left(\omega_0 + \Omega\right)t - \sin\left(\omega_0 - \Omega\right)t\right] \right\}. \qquad (1.4.14)$$

Формула (1.4.14) нагадує амплітудну модуляцію, з тією тільки відмінністю, що коливання на бічних частотах мають комплексні амплітуди протилежних знаків.

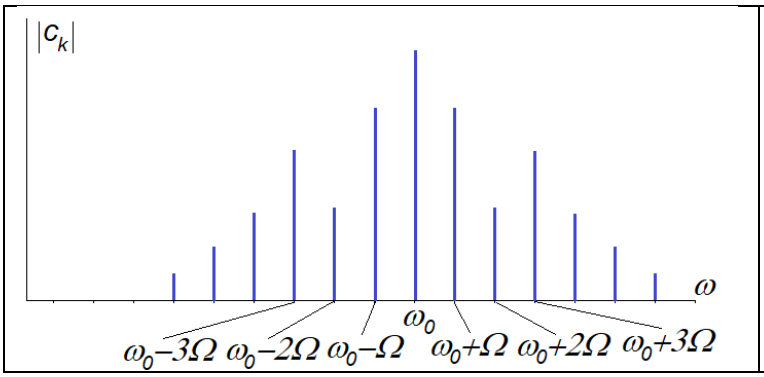


Рис. 1.4.8. Спектр сигналу з гармонічною фазовою модуляцією

1.4.6. Векторна діаграма фазово-модульованого сигналу

Векторна діаграма сигналу з довільним законом фазової модуляції та з гармонічною фазовою модуляцією для випадку $m_\varphi << 1$ подана на рис. 1.4.9 а-б.

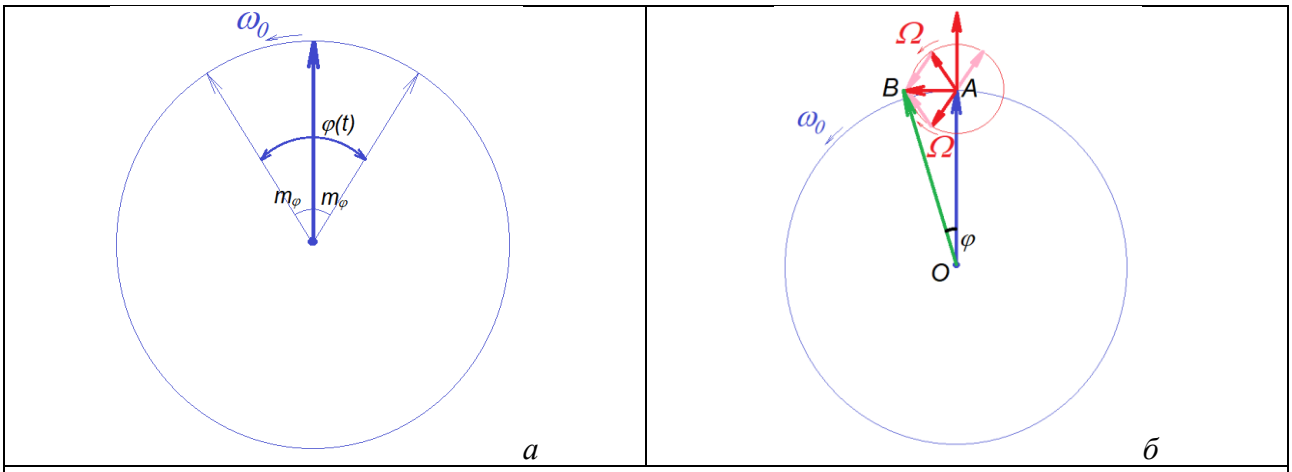


Рис. 1.4.9. Векторна діаграма сигналу з фазовою модуляцією: а – для довільного закону модуляції; б – для гармонічного закону модуляції ($m_\varphi << 1$)

Видно, що сумарний сигнал АВ на частоті Ω зсунутий за фазою щодо сигналу несу-

чої частоти ОА на $\pi/2$ (рис. 1.4.9 б). В силу мализни вектора АВ довжини векторів ОА і ОВ практично однакові, а малий кут між ним складає $\varphi(t) \approx AB(t)/OA$.

Якщо індекс фазової модуляції не буде малим, кількість нетривіальних доданків у ряду (1.4.12) збільшується. Це приведе й до відповідного ускладнення векторної діаграми.

Контрольні питання до підрозділу 1.4

1. Коли амплітудно-модульований сигнал із гармонічною модуляцією можна розкласти в ряд Фур'є?
2. Чому в спектрі фазово-модульованого сигналу присутні комбінаційні частоти вигляду $\omega_0 + k\Omega,\ k = 0, \pm 1, \pm 2...$?
3. Поясніть вигляд векторної діаграми сигналу з гармонічною модуляцією фази ($m_\varphi << 1$).

Приклади розв'язання задач до підрозділу 1.4

П 1.4.1. Розрахувати спектр періодичного амплітудно-маніпульованого сигналу вигляду:

$$u(t) = \begin{cases} u_m \cos(\omega_0 t - nT), & |t - nT| \le \tau/2; \\ 0, & |t - T/2 - nT| < (T - \tau)/2, \end{cases} \qquad n = 0, \pm 1, \pm 2, ... \qquad \tau < T.$$

Розв'язок

Відповідно до формули (1.2.20)

$$c_n = (1/T) \int_{-T/2}^{T/2} f(t) \exp(-i 2\pi n t/T) dt = (1/T) \int_{-\tau/2}^{\tau/2} u_m \cos\omega_0 t \exp(-i 2\pi n t/T) dt =$$

$$= (u_m/2T) \left\{ \int_{-\tau/2}^{\tau/2} \exp\left[i(\omega_0 - 2\pi n/T) t \right] dt + \int_{-\tau/2}^{\tau/2} \exp\left[-i(\omega_0 + 2\pi n/T) t \right] dt \right\} =$$

$$= \frac{u_m}{2T} \left\{ \frac{\exp\left[i(\omega_0 - 2\pi n/T) t \right]_{-\tau/2}^{\tau/2}}{i(\omega_0 - 2\pi n/T)} + \frac{\exp\left[-i(\omega_0 + 2\pi n/T) t \right]_{-\tau/2}^{\tau/2}}{-i(\omega_0 + 2\pi n/T)} \right\} =$$

$$= \frac{u_m}{2T} \left\{ \frac{\exp\left[i\left(\omega_0 - \frac{2\pi n}{T}\right)\frac{\tau}{2} \right] - \exp\left[-i\left(\omega_0 - \frac{2\pi n}{T}\right)\frac{\tau}{2} \right]}{i\left(\omega_0 - \frac{2\pi n}{T}\right)} + \frac{\exp\left[-i\left(\omega_0 + \frac{2\pi n}{T}\right)\frac{\tau}{2} \right] - \exp\left[i\left(\omega_0 + \frac{2\pi n}{T}\right)\frac{\tau}{2} \right]}{-i\left(\omega_0 + \frac{2\pi n}{T}\right)} \right\} =$$

$$= \frac{u_m}{T} \left\{ \frac{\sin\left[\left(\omega_0 - \frac{2\pi n}{T}\right)\frac{\tau}{2} \right]}{\left(\omega_0 - \frac{2\pi n}{T}\right)} + \frac{\sin\left[\left(\omega_0 + \frac{2\pi n}{T}\right)\frac{\tau}{2} \right]}{\left(\omega_0 + \frac{2\pi n}{T}\right)} \right\} = \frac{u_m \tau}{2T} \left\{ \operatorname{sinc}\left[\left(\frac{\omega_0}{2} - \frac{\pi n}{T}\right)\tau \right] + \operatorname{sinc}\left[\left(\frac{\omega_0}{2} + \frac{\pi n}{T}\right)\tau \right] \right\}.$$

Максимуми обвідної відповідають умові $n \approx \pm \omega_0 T/2\pi$ (рис. 1.4.10).

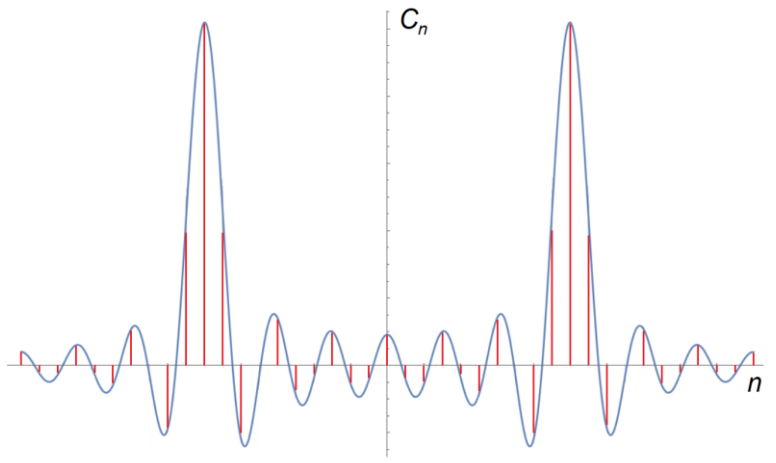


Рис. 1.4.10

Задачі до підрозділу 1.4

1.4.1. Знайти спектр прямокутного радіоімпульсу

$$u(t)=\begin{cases} u_m \cos\omega_0 t, & |t|\le \tau/2; \\ 0, & |t|>\tau/2. \end{cases}$$

і порівняти його зі спектром відповідного відеоімпульсу (див. задачу П 1.3.1).

1.4.2. Знайти спектр частотно-маніпульованого сигналу вигляду

$$u(t)=\begin{cases} u_m \cos(\omega_1 t-nT), & |t-nT|\le \tau/2; \\ u_m \cos(\omega_2 t-nT), & |t-T/2-nT|<(T-\tau)/2, \end{cases} \qquad n=0,\pm1,\pm2,\ldots \qquad \tau<T\,.$$

1.4.3. Знайти спектр фазово-маніпульованого сигналу вигляду

$$u(t)=\begin{cases} u_m \cos(\omega_0 t-nT), & |t-nT|\le \tau/2; \\ -u_m \cos(\omega_0 t-nT), & |t-T/2-nT|<(T-\tau)/2, \end{cases} \qquad n=0,\pm1,\pm2,\ldots \qquad \tau<T\,.$$

1.5. Узагальнення поняття спектру

Як було з'ясовано вище, спектр звичайно знаходять шляхом розкладання сигналу в ряд або інтеграл Фур'є. В певному сенсі лабораторним аналогом цих перетворень є аналізатор спектру. Однак у багатьох випадках час аналізу спектру менший від тривалості сигналу. Математично ці випадки описуються так званим віконним перетворенням Фур'є.

Узагальненням віконного перетворення Фур'є виступає вейвлет-перетворення, яке дозволяє бачити зміну віконного спектру за одночасної зміни ширини вікна та його положення в часі.

1.5.1. Віконне перетворення Фур'є

Формально в прямому інтегральному перетворенні Фур'є (1.3.3)

$$s(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)\exp(-i\omega t)\,dt$$

інтегрування за часом ведеться по нескінченному інтервалу $]-\infty,\infty[$. Тому говорити про спектр сигналу в деякий момент часу, на перший погляд, некоректно.

Однак в експерименті завжди мають справу зі скінченними проміжками часу. Зокрема, спектроаналізатори, як уже вказувалося (див. вище підрозділ 1.3.6), вимірюють спектр якраз на деякому скінченному проміжку часу і дозволяють простежити за змінами спектру сигналу з часом, якщо такі відбуваються.

Математичним вираженням такого підходу є так зване віконне перетворення Фур'є:

$$s(\omega,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} f(t')W(t-t')\exp(-i\omega t')\,dt', \qquad W(\tau)\underset{\tau\to\pm\infty}{\to}0, \qquad (1.5.1)$$

$W(\tau)$ – віконна функція, відмінна від нуля на скінченному проміжку часу. Зрозуміло, що внесок до інтегралу (1.5.1) в момент часу t дають лише ті значення t', для яких функції $f(t')$ та $W(t-t')$ одночасно відмінні від нуля. Іншими словами, віконна функція «вирізає» деяку частину сигналу, яка й дає внесок до віконного перетворення Фур'є (рис. 1.5.1).

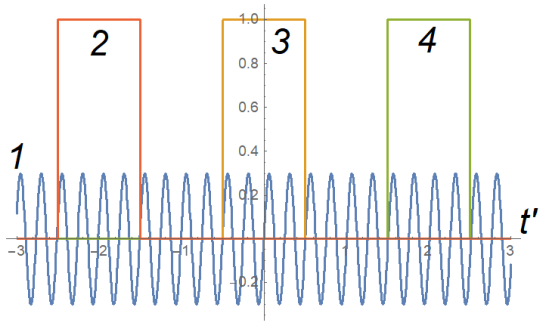


Рис. 1.5.1. До інтерпретації віконного перетворення Фур'є: 1 – сигнал $f(t')$, 2-4 – віконна функція $W(t-t')$ для моментів часу $t=-2$ (крива 2), $t=0$ (крива 3), $t=2$ (крива 4)

Якщо аналізований сигнал $f(t)$ є квазігармонічним, тобто йому можна приписати деяку частоту-носій ω_0, а його амплітуда, частота або фаза змінюються за характерний час T_0 ($\omega_0 T_0 \gg 1$), то тривалість вікна τ_0 зазвичай обирають з умови $T_0 \gg \tau_0 \gg 2\pi/\omega_0$. Тоді віконне перетворення дає можливість прослідкувати зміну властивостей сигналу з часом.

Віконне перетворення Фур'є можна записати також через спектри сигналу $s(\omega)$ та віконної функції $w(\omega)$:

$$s(\omega,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} f(t')W(t-t')\exp(-i\omega t')\,dt'=$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} dt' \exp(-i\omega t')\int_{-\infty}^{\infty} s(\omega')\exp(i\omega' t')d\omega' \int_{-\infty}^{\infty} w(\omega'')\exp\left[i\omega''(t-t')\right]d\omega'' =$$

$$= \int_{-\infty}^{\infty} s(\omega')d\omega' \int_{-\infty}^{\infty} w(\omega'')\exp(i\omega'' t)\,d\omega'' \frac{1}{2\pi}\int_{-\infty}^{\infty} dt' \exp\left[i(\omega'-\omega''-\omega)t'\right] =$$

$$= \int_{-\infty}^{\infty} s(\omega')d\omega' \int_{-\infty}^{\infty} \delta(\omega'-\omega''-\omega)\,w(\omega'')\exp(i\omega'' t)\,d\omega'' =$$

$$= \int_{-\infty}^{\infty} s(\omega')\,w(\omega'-\omega)\exp\left[i(\omega'-\omega)t\right]d\omega'. \qquad (1.5.2)$$

Отриманий результат можна інтерпретувати як аналог згортки функцій $s(\omega)$ та $w(\omega)\exp(i\omega t)$. Нагадаємо, що згорткою функцій $f(t)$ та $g(t)$ в математиці називають інтеграл вигляду

$$f * g \equiv \int_{-\infty}^{\infty} f(x-x')\,g(x')\,dx'. \qquad (1.5.3)$$

З формули (1.5.3) видно, що до віконного перетворення Фур'є роблять внесок тільки ті значення частоти ω', для яких спектри $s(\omega')$ та $w(\omega'-\omega)$ одночасно відмінні від нуля.

В найпростішому випадку віконна функція $W(\tau)$ має прямокутну форму:

$$W(\tau) = \begin{cases} 1, & |\tau| \le \tau_0/2; \\ 0, & |\tau| > \tau_0/2. \end{cases} \qquad (1.5.4)$$

При цьому інтеграл (1.5.1) набуває вигляду

$$s(\omega,t) = \frac{1}{2\pi}\int_{t-\tau_0/2}^{t+\tau_0/2} f(t')\exp(-i\omega t')\,dt'. \qquad (1.5.5)$$

Зазначимо, що підбором віконної функції можна задовольнити умову існування віконного перетворення Фур'є

$$\int_{-\infty}^{\infty} \left|f(t')W(t-t')\right|^2 dt' < \infty \qquad (1.5.6)$$

навіть тоді, коли для самої функції $f(t)$ аналогічна умова (1.3.1)

$$\int_{-\infty}^{\infty} \left|f(t)\right|^2 dt < \infty$$

не виконана.

Зазначимо також, що для віконного перетворення Фур'є спектр визначений як у додатній, так і у від'ємній області частот. Як уже зазначалося (див. вище підрозділ 1.3.6), спектроаналізатори вимірюють спектр лише в додатній області частот.

1.5.2. Віконне перетворення Фур'є для квазімонохроматичного сигналу

Щоб зрозуміти ідею використання віконного перетворення Фур'є, застосуємо його до монохроматичного сигналу, а вікно оберемо прямокутним. Маємо:

$$s(\omega,t) = \frac{u_m}{2\pi}\int_{t-T/2}^{t+T/2} \cos\omega_0 t' \exp(-i\omega t')\,dt' = \frac{u_m}{4\pi}\left\{\int_{t-T/2}^{t+T/2} \exp\left[i(\omega_0-\omega)t'\right]dt' + \int_{t-T/2}^{t+T/2} \exp\left[i(-\omega_0-\omega)t'\right]dt'\right\} =$$

$$= \frac{u_m}{4\pi}\left\{ \frac{\exp\left[i\left(\omega_0-\omega\right)t'\right]}{i\left(\omega_0-\omega\right)}\Bigg|_{t-T/2}^{t+T/2} + \frac{\exp\left[i\left(-\omega_0-\omega\right)t'\right]}{i\left(-\omega_0-\omega\right)}\Bigg|_{t-T/2}^{t+T/2} \right\} =$$

$$= \frac{u_m}{4\pi}\left\{ \frac{\exp\left[i\left(\omega_0-\omega\right)\left(t+T/2\right)\right]-\exp\left[i\left(\omega_0-\omega\right)\left(t-T/2\right)\right]}{i\left(\omega_0-\omega\right)} + \right.$$

$$\left. + \frac{\exp\left[i\left(-\omega_0-\omega\right)\left(t+T/2\right)\right]-\exp\left[i\left(-\omega_0-\omega\right)\left(t-T/2\right)\right]}{i\left(-\omega_0-\omega\right)} \right\} =$$

$$= \frac{u_m}{2\pi}\left\{ \exp\left[i\left(\omega_0-\omega\right)t\right]\frac{\sin\left[\left(\omega_0-\omega\right)T/2\right]}{\left(\omega_0-\omega\right)} + \exp\left[-i\left(\omega_0+\omega\right)t\right]\frac{\sin\left[\left(\omega_0+\omega\right)T/2\right]}{\left(\omega_0+\omega\right)} \right\} =$$

$$= \frac{u_m T}{4\pi}\left\{ \operatorname{sinc}\frac{\left(\omega-\omega_0\right)T}{2}\exp\left[-i\left(\omega-\omega_0\right)t\right] + \operatorname{sinc}\frac{\left(\omega+\omega_0\right)T}{2}\exp\left[-i\left(\omega+\omega_0\right)t\right] \right\}, \qquad (1.5.7)$$

де, як і раніше, $\operatorname{sinc} x = \sin x / x$. Графік функції $\operatorname{sinc}\left(x-x_0\right)$ поданий на рис. 1.5.2 а, спектр функції $\cos\omega_0 t$ – на рис. 1.5.2.б (в даному разі амплітуда гармоніки не змінюється з часом, тому її віконний спектр не залежить від часу).

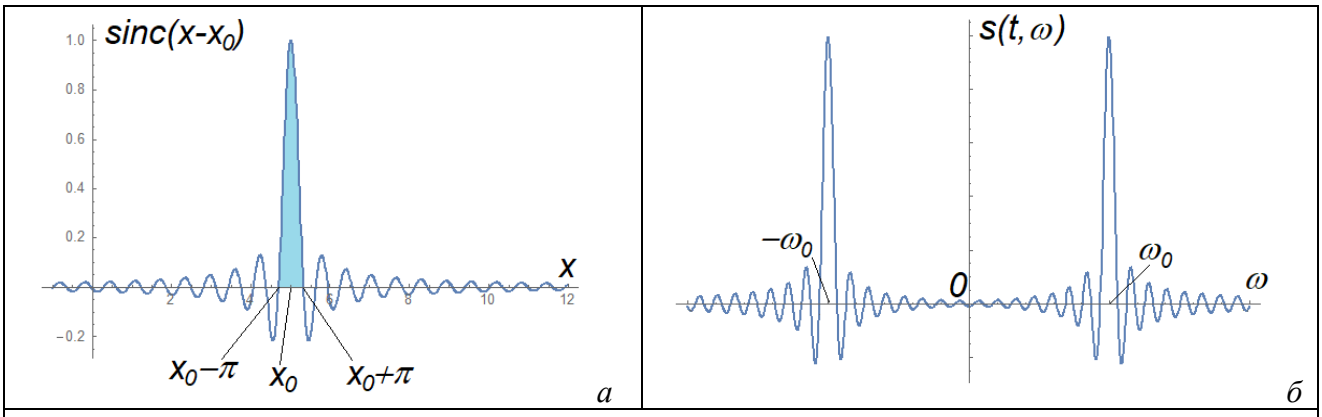


Рис. 1.5.2. Віконне перетворення Фур'є з прямокутним вікном:
а – графік функції $\operatorname{sinc}\left(x-x_0\right)$ *(основна частина виділена кольором);*
б – віконний спектр гармонічної функції $\cos\omega_0 t$

Видно, що спектр (1.5.7) локалізований в околі частот $\omega = \pm\omega_0$. Основні частини спектру мають ширину за частотою $T\Delta\omega/2 = 2\pi$, або $\Delta\omega = 4\pi/T$ і висоту $T/2$, тобто площа заштрихованої області (величина порядку 2π) зберігається при $T \to \infty$. Це означає, що за цієї умови функція $\operatorname{sinc}\left(x-x_0\right)$ стає схожою на δ-функцію Дірака.

Припустимо тепер, що частота сигналу $\omega_0 = \omega_0(t)$ повільно змінюється з часом – повільно в масштабі періоду самої частоти $\omega_0 = \omega_0(t)$, тобто характерний час її зміни $\tau = \left|\left(1/\omega_0\right)\left(d\omega_0/dt\right)\right|^{-1}$ задовольняє умову $\tau \gg 2\pi/\omega_0$, або, що те саме, $\omega_0\tau \gg 1$. Нехай тривалість вікна задовольняє умові

$$\tau \gg T \gg 2\pi/\omega_0 . \qquad (1.5.8)$$

В цьому випадку, з одного боку, неточність визначення частоти буде $\Delta\omega = 4\pi/T \ll \omega_0$, тобто відносна похибка виявляється малою, $\Delta\omega/\omega_0 \ll 1$. З іншого боку, за час, що дорівнює тривалості вікна, частота $\omega_0(t)$ не встигає помітно змінитися.

Очевидно, за неточність визначення моменту часу, в який вимірюється частота, слід узяти половину тривалості вікна, $\Delta t = T/2$. Таким чином, у даному разі виконується умова $\Delta t \times \Delta\omega = 2\pi$. Вибір невизначеності за частотою і за часом носить дещо умовний характер, тому значення константи в правій частині також є умовним (з точністю до порядку величини). Але сам результат – що ширина спектру імпульсу обернено пропорційна до його тривалості – носить універсальний характер. З ним пов'язане й знамените співвідношення невизначеності Ґайзенберга в квантовій механіці (в експерименті неможливо одночасно визначити координату точки та швидкість її руху).

Повернемося до віконного перетворення Фур'є. Як видно з рис. 1.5.2, при виборі прямокутного вікна значна частина енергії спектру лежить за межами основного максимуму. Максимальний рівень бічних пелюсток його спектру щодо центральної складає лише 13 дБ (рис. 1.5.3 а). Тому на практиці використовують інші віконні функції, деякі з яких наведені в таблиці 1.5.1. В цифровій техніці віконне перетворення Фур'є використовується для синтезу частотних фільтрів.

1.5.3. Приклади віконних функцій

Деякі приклади поширених віконних функцій наведені в табл. 1.5.1 (див. також рис. 1.5.3).

Таблиця 1.5.1. Деякі поширені віконні функції

Назва	Аналітична формула	Максимальний рівень бічних пелюсток, дБ
вікно Ханна	$W(t) = \begin{cases} 0.5\left[1-\cos(2\pi t/T)\right], & \lvert t\rvert \le T/2; \\ 0, & \lvert t\rvert > T/2. \end{cases}$	-31.5
вікно Хеммінга	$W(t) = \begin{cases} 0{,}53836 - 0{,}46164\cos(2\pi t/T), & \lvert t\rvert \le T/2; \\ 0, & \lvert t\rvert > T/2. \end{cases}$	-42
вікно Блекмена	$W(t) = \begin{cases} \frac{1-\alpha}{2} - \frac{1}{2}\cos(2\pi t/T) + \frac{\alpha}{2}\cos(4\pi t/T), & \lvert t\rvert \le T/2; \\ 0, & \lvert t\rvert > T/2. \end{cases}$	-58 ($\alpha = 0.16$)

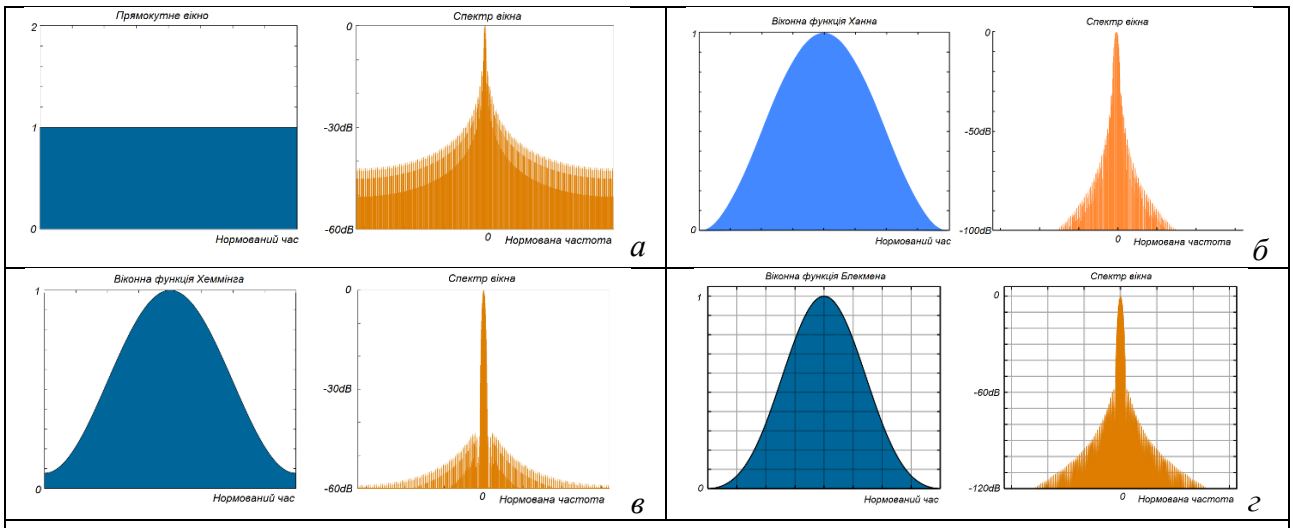


Рис. 1.5.3. Вигляд графіків деяких поширених віконних функцій (ліворуч – часовий хід $W(\tau)$, праворуч – спектр $w(\omega)$ у напівлогарифмічному масштабі): а – прямокутне вікно; б – вікно Ханна; в – вікно Хеммінга; г – вікно Блекмена

1.5.4. Поняття про вейвлет-перетворення

Одним з варіантів подальшого узагальнення віконного перетворення Фур'є виступає вейвлет-перетворення (wavelet – маленька або коротка хвиля). Пряме вейвлет-перетворення функції $f(t)$ має вигляд

$$W(a,b)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right)dt=\int_{-\infty}^{\infty} f(t)\psi_{ab}(t)dt, \qquad (1.5.9)$$

зворотне –

$$f(t)=\frac{1}{C_{\psi}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(a,b)\psi_{ab}(t)\frac{da\,db}{a^2}, \qquad (1.5.10)$$

де використані позначення

$$\psi_{ab}(t)=\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \qquad (1.5.11)$$

$$C_{\psi}=\int_{-\infty}^{\infty}\left|\frac{\psi^2(\omega)}{\omega}\right|d\omega<\infty, \qquad (1.5.12)$$

функція $\psi(t)$ – вихідний вейвлет (про його властивості мова буде далі), аналогом якого у віконному перетворенні Фур'є виступає добуток $W(t-t')\exp(-i\omega t')$. Параметр b визначає момент дії вейвлету (аналог t у віконному перетворенні Фур'є), параметр a визначає часовий масштаб перетворення (аналог тривалості вікна).

Отже, при $a=const$ інтеграл (1.5.9) виступає деяким аналогом віконного перетворення Фур'є, а зміна цього параметра означає зміну тривалості вікна (часового масштабу перетворення).

З іншого боку, інтеграл (1.5.10) – це розкладання сигналу по системі базисних функцій (1.5.11), сконструйованих із вихідного вейвлету $\psi(t)$.

1.5.5. Головні ознаки вейвлету

1. Обмеженість норми:

$$\|\psi\|\equiv\int_{-\infty}^{\infty}|\psi(t)|^2\,dt<\infty. \qquad (1.5.13)$$

2. Локалізація в часі й за частотою:

$$|\psi(t)|\le\frac{C}{(1+|t|)^{1+\varepsilon}};\quad S_{\psi}(\omega)=\left|\frac{1}{2\pi}\int_{-\infty}^{\infty}\psi(t)\exp(-i\omega t)dt\right|<\frac{C}{(1+|\omega|)^{1+\varepsilon}};\quad \varepsilon>0 \qquad (1.5.14)$$

(див. рис. 1.5.3-1.5.4).

3. Нульове середнє:

$$\int_{-\infty}^{\infty}\psi(t)dt=0. \qquad (1.5.15)$$

4. Автомодельність (самоподібність). Усі вейвлети сім'ї $\psi_{ab}(t)$, отримані з того самого вихідного вейвлету $\psi(t)$, мають однакову кількість осциляцій (рис. 1.5.4).

Аналітично перетворення (1.5.9)-(1.5.10) важко виконати навіть для найпростіших вейвлетів і гармонічних сигналів. Однак деякі якісні висновки можна зробити й без таких розрахунків.

Нехай характерний період осциляцій вейвлета дорівнює $\tau=\tau(a)$, а характерна час-

тота сигналу $f(t)$ – ω. Очевидно, інтеграл вигляду (1.5.9) за умови $\omega\tau(a) \ll 1$ обертається на нуль.

Справді, в цьому випадку $f(t)$ можна винести з-під інтеграла, оскільки протягом тривалості вейвлету вона майже не зміниться, а інтеграл від вейвлета за часом дає нуль. В протилежному випадку $\omega\tau(a) \gg 1$ отримаємо швидкі осциляції з повільною амплітудою, а інтеграл від такої функції теж близький до нуля. Таким чином, інтеграл (1.5.9) має помітну величину лише для тих значень a, для яких $\omega\tau(a) \simeq 1$.

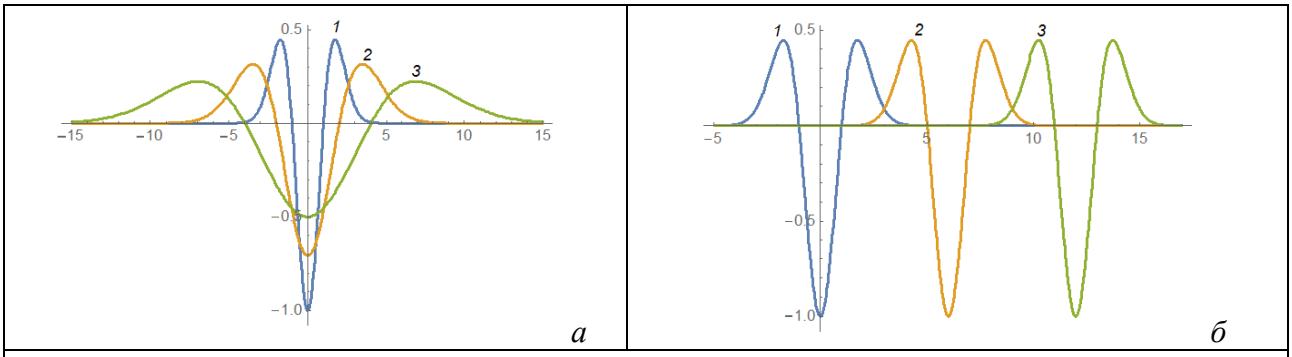


Рис. 1.5.4. Графіки сім'ї вейвлетів $\psi_{ab}(t)$, породженої вихідним вейвлетом $\psi(x) = (x^2 - 1)\exp(-x^2/2)$ ("мексиканський капелюх"): а – $a = 1,\ b = 0$ (крива 1), $a = 2,\ b = 0$ (крива 2), $a = 4,\ b = 0$ (крива 3); б – $a = 1,\ b = 0$ (крива 1), $a = 1,\ b = 6$ (крива 2), $a = 1,\ b = 12$ (крива 3)

Обговоримо тепер вейвлет-перетворення для сигналів обмеженої тривалості. Нехай сигнал $f(t)$ має тривалість T_1, вейвлет $\psi[(t-b)/a]$ (при $a = 1$) – тривалість T_2. Тоді перша з цих функцій відмінна від нуля при $|t| \le T_1$, друга – при $|t - b| \le aT_2$, тобто в області $b - aT_2 \le t \le b + aT_2$. Обидва множники під інтегралом будуть одночасно відмінні від нуля за умови $|b| \le T_1 + aT_2$. Ця умова визначає область, де функція $W(a,b)$ відмінна від нуля.

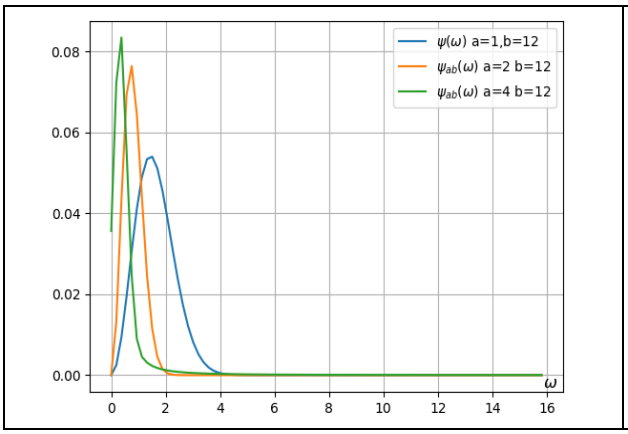


Рис. 1.5.5. Графіки сім'ї спектрів вейвлетів $\psi_{ab}(\omega)$, породжених вихідним вейвлетом $\psi(x) = (x^2 - 1)\exp(-x^2/2)$ ("мексиканський капелюх")

1.5.6. Приклади вейвлет-спектрів

На рис. 1.5.6-1.5.7 наведені вейвлет-спектри для вихідного вейвлету типу «мексиканський капелюх» і двох варіантів вхідних сигналів – монохроматичного та двочастотного – які підтверджують наведені міркування. Видно, що зі зсувом вейвлету (зміною параметра b) картинка періодично повторюється. Це наслідок періодичності функції $f(t)$. Видно, що на рис. 1.5.6 а період вхідного сигналу $T \approx 45$. Відповідно на рис. 1.5.7 б-в максимуми сигналів

задовільно відповідають значенню $a = T/\pi$.

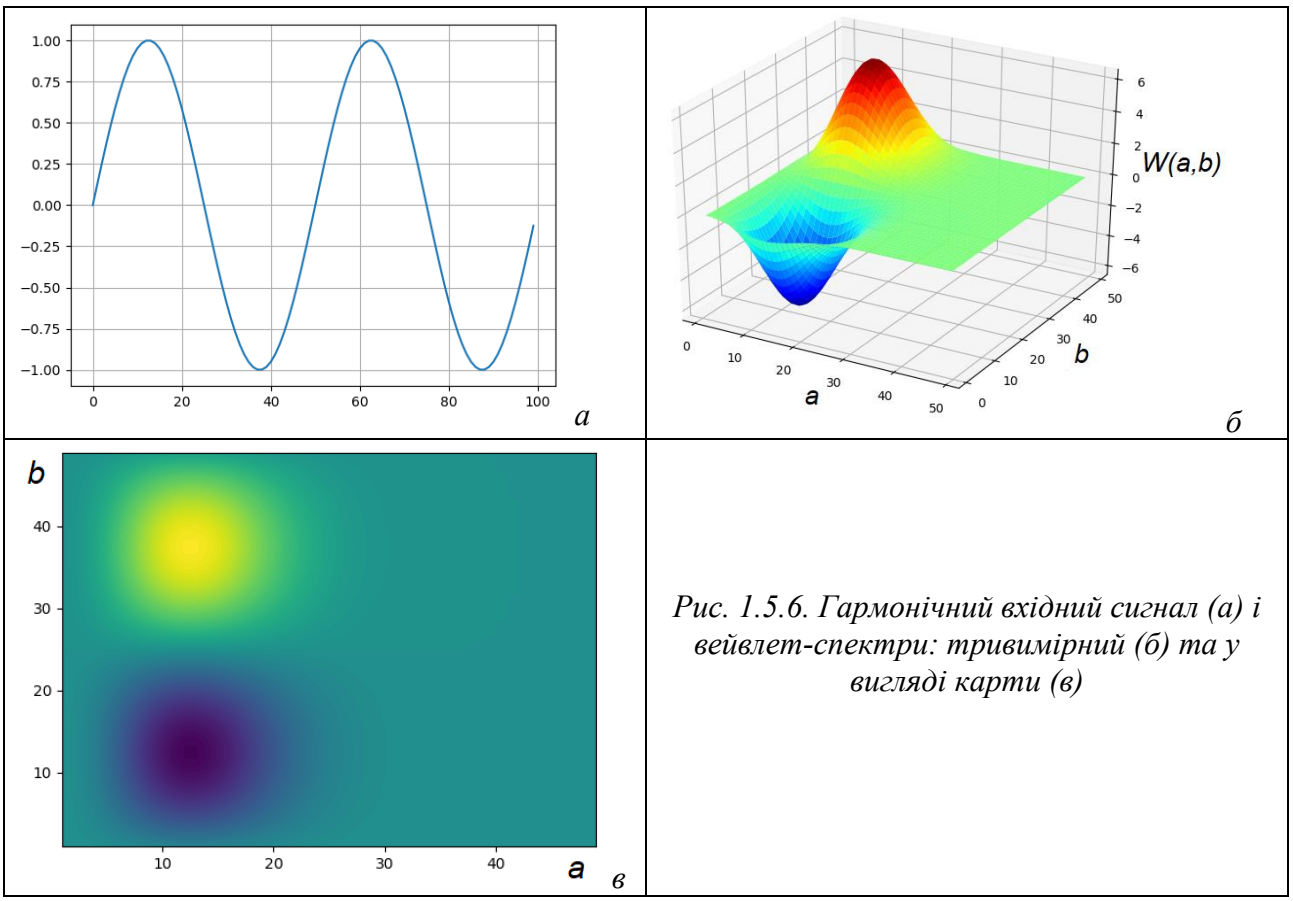


Рис. 1.5.6. Гармонічний вхідний сигнал (а) і вейвлет-спектри: тривимірний (б) та у вигляді карти (в)

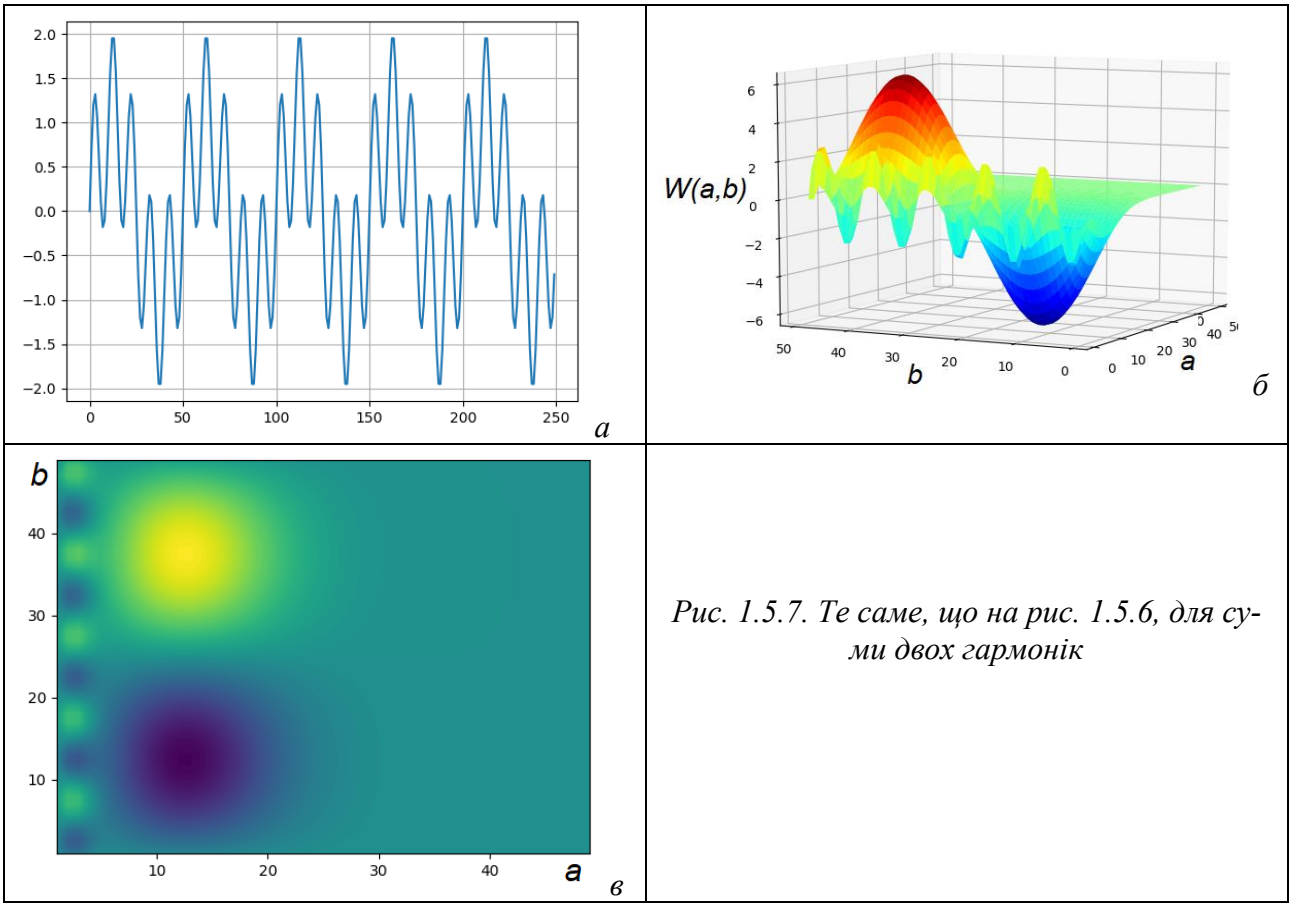


Рис. 1.5.7. Те саме, що на рис. 1.5.6, для суми двох гармонік

Контрольні питання до підрозділу 1.5

1. Чим віконне перетворення Фур'є відрізняється від звичайного?
2. Які можливості надає віконне перетворення Фур'є при аналізі експериментальних даних?
3. Що таке вейвлет-перетворення?

Приклади розв'язання задач до підрозділу 1.5

П 1.5.1. Оберемо вихідний вейвлет у модельній формі:

$$\psi(x)=\begin{cases}\cos x, & |x|\le 2\pi\\ 0, & |x|>2\pi.\end{cases}$$

Легко переконатися, що ця модельна функція (рис. 1.5.7) не задовольняє другій з умов (1.5.14), але задовольняє іншим умовам, указаним у підрозділі 1.5.5.

а) Записати для модельного вихідного вейвлета сім'ю вейвлетів $\psi_{ab}(t)$ та побудувати їхні графіки для різних значень параметрів $a,\ b$.

Розв'язок

Скористаємося формулою (1.5.11):

$$\psi_{ab}(t)=\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)=\begin{cases}\dfrac{1}{\sqrt{a}}\cos\left(\dfrac{t-b}{a}\right), & \left|\dfrac{t-b}{a}\right|\le 2\pi;\\ 0, & \left|\dfrac{t-b}{a}\right|>2\pi\end{cases}$$

(рис. 1.5.8).

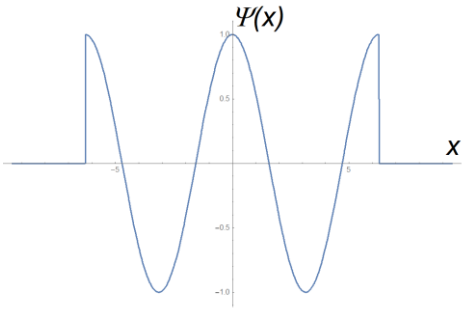
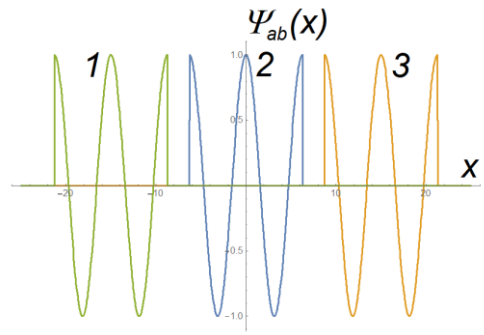
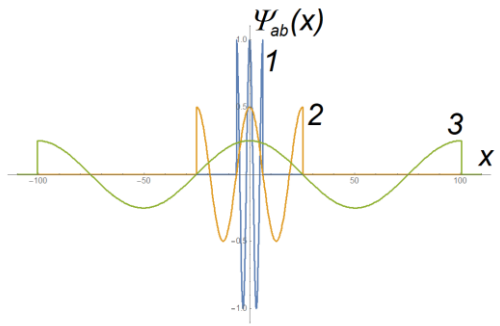


Рис. 1.5.7. Вихідний вейвлет



а



б

Рис. 1.5.8. Сім'я вейвлетів $\psi_{ab}(t)$ (1.5.17): а – $a=1$, $b=-15$ (крива 1), $b=0$ (крива 2), $b=15$ (крива 3); б – $b=0$, $a=1$ (крива 1), $a=4$ (крива 2), $a=16$ (крива 3)

б) Розрахувати пряме вейвлет-перетворення функції $f(t)=\sin t$ і побудувати його двовимірний графік.

Розв'язок

$$W(a,b)=\int_{-\infty}^{\infty} f(t)\psi_{ab}(t)dt=\frac{1}{\sqrt{a}}\int_{-2\pi}^{2\pi}\sin t\cos\left(\frac{t-b}{a}\right)dt=$$

$$=\frac{1}{2\sqrt{a}}\int_{-2\pi}^{2\pi}\left[\sin\left(t+\frac{t-b}{a}\right)+\sin\left(t-\frac{t-b}{a}\right)\right]dt=$$

$$=\frac{1}{2\sqrt{a}}\left\{\int_{-2\pi}^{2\pi}\sin\left[\left(1+\frac{1}{a}\right)t-\frac{b}{a}\right]dt+\int_{-2\pi}^{2\pi}\sin\left[\left(1-\frac{1}{a}\right)t+\frac{b}{a}\right]dt\right\}=$$

$$=\frac{\sqrt{a}}{2(a+1)}\int_{-2\pi}^{2\pi}\sin\left(\frac{a+1}{a}t-\frac{b}{a}\right)d\left(\frac{a+1}{a}t-\frac{b}{a}\right)+$$

$$+\frac{\sqrt{a}}{2(a-1)}\int_{-2\pi}^{2\pi}\sin\left(\frac{a-1}{a}t+\frac{b}{a}\right)d\left(\frac{a-1}{a}t+\frac{b}{a}\right)=$$

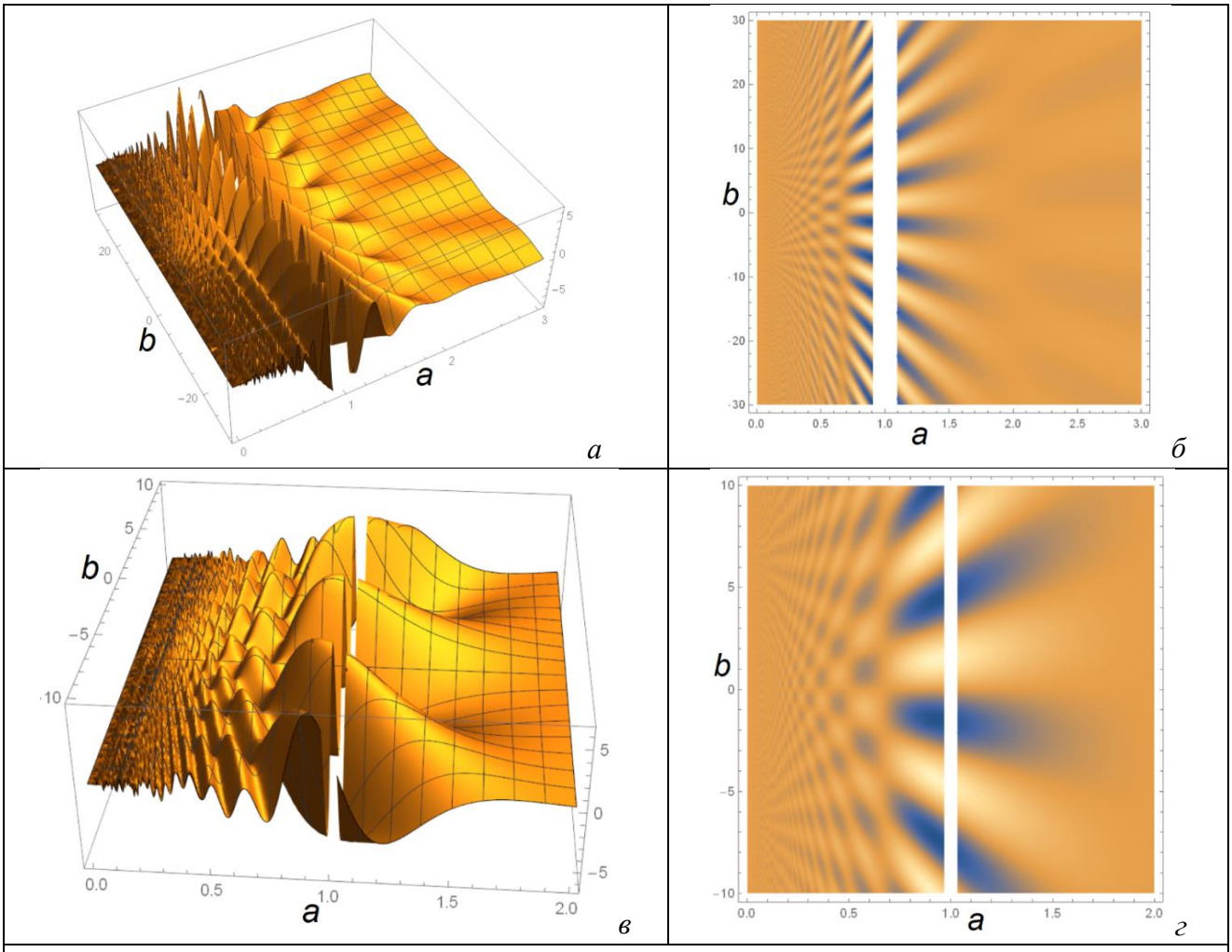


Рис. 1.5.9. Вейвлет-перетворення функції $f(t)=\sin t$: а – тривимірне зображення, б – кольорова карта, в, г – збільшені фрагменти рисунків а, б. Розриви на рис. а, в та білі смуги на рис. б, г – результат недосконалості числового розрахунку

$$= -\frac{\sqrt{a}}{2(a+1)}\cos\left(\frac{a+1}{a}t-\frac{b}{a}\right)\Bigg|_{-2\pi}^{2\pi} - \frac{\sqrt{a}}{2(a-1)}\cos\left(\frac{a-1}{a}t+\frac{b}{a}\right)\Bigg|_{-2\pi}^{2\pi} =$$

$$= \frac{\sqrt{a}}{2(a+1)}\left[\cos\left(2\pi\frac{a+1}{a}+\frac{b}{a}\right)-\cos\left(2\pi\frac{a+1}{a}-\frac{b}{a}\right)\right]+$$

$$+\frac{\sqrt{a}}{2(a-1)}\left\{\cos\left(2\pi\frac{a-1}{a}-\frac{b}{a}\right)-\cos\left(2\pi\frac{a-1}{a}+\frac{b}{a}\right)\right\}=$$

$$= \frac{\sqrt{a}}{a+1}\sin\left(2\pi\frac{a+1}{a}\right)\sin\frac{b}{a}+\frac{\sqrt{a}}{a-1}\sin\left(2\pi\frac{a-1}{a}\right)\sin\frac{b}{a}=$$

$$= \sqrt{a}\sin\left(\frac{b}{a}\right)\left[\frac{1}{a+1}\sin\left(2\pi\frac{a+1}{a}\right)+\frac{1}{a-1}\sin\left(2\pi\frac{a-1}{a}\right)\right]$$

(див. рис. 1.5.9).

Порівняння рис. 1.5.9 та рис. 1.5.6 наочно демонструє переваги віконної функції типу «мексиканський капелюх»: рельєф на рис. 1.5.6 б спостерігається лише там, де значення параметра a збігається з частотою сигналу, тоді як на рис. 1.5.9 а порізаною виявляється також область малих значень a. Але й на цьому рисунку рельєф найбільш виявлений там, де значення параметра a збігається з частотою сигналу.

Задачі до підрозділу 1.5

1.5.1. Для прямокутної віконної функції тривалістю T виконати віконне перетворення Фур'є функції

$$f(t)=\begin{cases}\cos t, |t|\le \theta/2;\\ 0,\ |t|>\theta/2.\end{cases}$$

1.5.2. Для гауссівської віконної функції виконати віконне перетворення Фур'є функції $\cos t$.

1.5.3. Для модельного вейвлета (1.5.16) розрахувати пряме вейвлет-перетворення функції $f(t)=\cos t$ і побудувати його двовимірний графік.

Розділ 2. Методи аналізу лінійних електричних кіл

В багатьох випадках радіоелектронні схеми (або, принаймні, їхні частини) можна розглядати як лінійні (див. нижче п. 2.1.1). Наприклад, лінійними, як правило, є фільтри та підсилювачі.

У цій частині курсу розглядаються методи аналізу лінійних електричних кіл (лінійних чотириполюсників) під дією гармонічних та негармонічних сигналів.

2.1. Лінійні кола та лінійні чотириполюсники

У цьому підрозділі дається означення лінійного електричного кола та вводиться поняття лінійного чотириполюсника. Як ми переконаємося далі, значна частина радіотехнічних кіл (не лише пасивних, але й активних) можуть розглядатися як лінійні, а моделлю лінійного чотириполюсника зручно користуватися як моделлю деяких активних елементів (зокрема, біполярних транзисторів).

2.1.1. Поняття про лінійні електричні кола

Ми вже згадували, що лінійні електричні кола – це кола, складені з елементів, параметри яких (опір, провідність, ємність, індуктивність тощо) не залежать від форми та величини сигналу (наприклад, прикладеної напруги або струму, що протікає через них). Така модель зазвичай справедлива для сигналів невеликої амплітуди. При великих амплітудах сигналів можуть проявлятись нелінійні ефекти: наприклад, може стати помітною зміна опору резистора внаслідок його нагрівання, зумовленого протіканням струму або зміна індуктивності котушки з феромагнітним осердям через залежність магнітної проникності останнього від сили струму в котушці.

Лінійні кола описуються лінійними диференціальними рівняннями зі сталими коефіцієнтами. Як уже вказувалося (див. вступ до частини I), ми розглядаємо лише кола з геометричними розмірами, значно меншими за довжину хвилі, тому це будуть рівняння в повних похідних. Отже, далі в цій частині курсу йтиметься про лінійні диференціальні рівняння в повних похідних зі сталими коефіцієнтами.

Зазначимо, що в радіотехніці часто розглядають наближення малого сигналу, коли струми або напруги x складаються з довільної постійної складової X_0 та малої змінної складової $x(t)$: $x = X_0 + x(t)$, $|x(t)| << X_0$. У цьому випадку функцію такого сигналу (наприклад, залежність вихідного струму польового транзистора від поданої на його вхід напруги) можна розкласти в ряд Тейлора, обмежившись лінійним доданком:

$$f(x) = f\left[X_0 + x(t)\right] \approx f(X_0) + \left.\frac{df}{dx}\right|_{x=x_0} x(t). \qquad (2.1.1)$$

Тоді для малого сигналу навіть нелінійні елементи (польові та біполярні транзистори, варикапи, діоди тощо) можна розглядати як лінійні. В таких випадках говорять про роботу на лінійній ділянці характеристики нелінійного елемента.

Як випливає з викладеного вище, рівняння, що описують відгук лінійних кіл на зовнішній вплив – це лінійні неоднорідні диференціальні рівняння в повних похідних зі сталими коефіцієнтами. В найзагальнішому випадку рівняння такого типу можна подати у вигляді

$$a_0 \frac{d^n x}{dt^n} + a_1 \frac{d^{n-1} x}{dt^{n-1}} + \ldots + a_{n-1} \frac{dx}{dt} + a_n x = f(t), \qquad (2.1.2)$$

де a_i – дійсні сталі коефіцієнти. В ролі невідомої функції $x(t)$ звичайно виступає струм або напруга в колі, які нас цікавлять, а функція $f(t)$ в правій частині рівняння відіграє роль вхідного сигналу (зовнішньої сили).

Теорія диференціальних рівнянь стверджує, що загальний розв’язок рівняння вигляду

(2.1.2) записується як сума частинного розв'язку неоднорідного рівняння і загального розв'язку відповідного однорідного рівняння (коли в правій частині стоїть нуль). Загальний розв'язок однорідного рівняння залежить від довільних сталих, кількість яких дорівнює порядку диференціального рівняння (в нашому випадку – n). Ці сталі визначаються з початкових умов.

Ми, як правило, обмежуватимемося випадком, коли за відсутності зовнішньої сили $x(t=0)=0$ (нульові початкові умови) і надалі шукатимемо лише розв'язок неоднорідного рівняння.

Для лінійних кіл справджується принцип суперпозиції: якщо для зовнішніх сил $f_1(t)$ та $f_2(t)$ розв'язки рівняння (2.1.2) є відповідно $x_1(t)$ та $x_2(t)$, то сумарній сили $f_1(t)+f_2(t)$ відповідає сума розв'язків $x_1(t)+x_2(t)$.

Таким чином, методи аналізу лінійних електричних кіл відіграють важливу роль у радіотехніці, й володіння ними необхідне для відповідних фахівців.

2.1.2. Падіння напруги на резисторі, конденсаторі та котушці індуктивності

Поки що ми обмежимося електричними колами, складеними лише з пасивних елементів – резисторів, конденсаторів та котушок індуктивності. Нагадаємо, що для цих елементів струм та напруга пов'язані між собою лінійними співвідношеннями. Зокрема, для активного опору R справджується закон Ома для ділянки кола:

$$u(t)=i(t)R. \tag{2.1.3}$$

Напруга на ємності C прямо пропорційна її заряду:

$$q(t)=u(t)C, \tag{2.1.4}$$

або

$$u(t)=\frac{1}{C}\int i(t)dt. \tag{2.1.5}$$

Тут враховано, що струм – це похідна за часом від заряду q, або, навпаки,

$$q(t)=\int i(t)dt. \tag{2.1.6}$$

Нарешті, падіння напруги на котушці індуктивності визначається зміною магнітного потоку Φ через неї:

$$u(t)=\frac{d\Phi}{dt}. \tag{2.1.7}$$

З урахуванням лінійного зв'язку магнітного потоку зі струмом,

$$\Phi(t)=Li(t), \tag{2.1.8}$$

де L – індуктивність котушки, остаточно отримуємо:

$$u(t)=L\frac{di}{dt}. \tag{2.1.9}$$

Співвідношення (2.1.3), (2.1.5) і (2.1.9) і використовуються, зокрема, при аналізі пасивних електричних кіл.

2.1.3. Лінійний чотириполюсник

У багатьох випадках для опису лінійних систем зручно користуватися моделлю лінійного чотириполюсника (рис. 2.1.1 а). Ця модель характеризується вхідним та вихідним струмами i_1, i_2 і вхідною та вихідною напругою u_1, u_2 (рис. 2.1.1 б). Вважаємо, що між цими величинами існує лінійний зв'язок.

В курсі теорії електричних кіл звичайно вважають напруги і струми i_1, i_2, u_1, u_2 на вході й виході лінійного чотириполюсника (рис. 2.1.1 б) сталими величинами (або майже сталими, тобто такими, що повільно змінюються з часом). Зв'язок між цими величинами записують у вигляді лінійних алгебраїчних рівнянь, коефіцієнти яких називаються параметрами лінійного чотириполюсника.

В цьому курсі ми користуватимемося системою h-параметрів. При такому описі заданими величинами (впливами) вважаються вхідний струм i_1 і вихідна напруга u_2, шуканими (відгуками) – відповідно вхідна напруга u_1 й вихідний струм i_2. Таким чином, рівняння лінійного чотириполюсника в h-параметрах мають вигляд

$$\begin{cases} u_1 = h_{11}i_1 + h_{12}u_2; \\ i_2 = h_{21}i_1 + h_{22}u_2. \end{cases} \qquad (2.1.10)$$

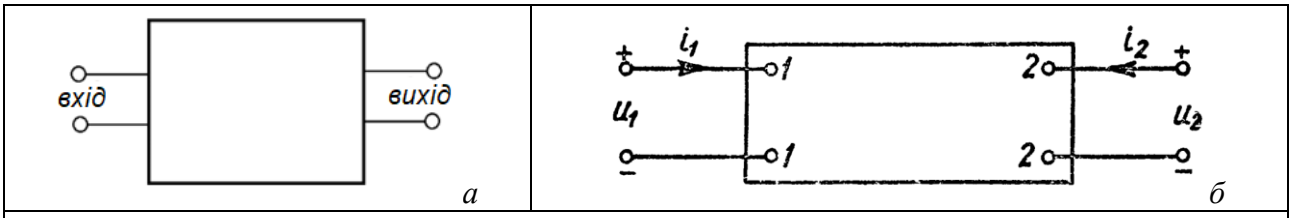


Рис. 2.1.1. Умовне зображення лінійного чотириполюсника (а) та його вхідні й вихідні напруги й струми (б)

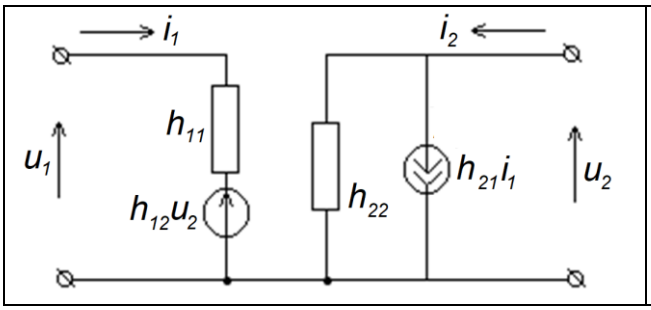


Рис. 2.1.2. Еквівалентна схема лінійного чотириполюсника в h-парметрах

Еквівалентна схема лінійного чотириполюсника в h-параметрах подана на рис. 2.1.2. Вона повністю відповідає рівнянням (2.1.10) і складається з резисторів та керованих джерел струму й напруги.

Ми розумітимемо лінійний чотириполюсник ширше – як деяку лінійну схему, описувану диференціальним рівнянням вигляду (2.1.10), причому під вхідним сигналом (впливом) розумітимемо $f(t)$, а під вихідним (відгуком) – $x(t)$.

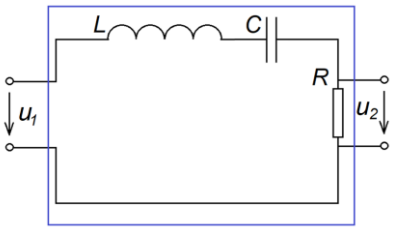


Рис. 2.1.3. Варіант подання схеми послідовного коливного контуру з джерелом змінної напруги u_1 у вигляді лінійного чотириполюсника

На рис. 2.1.3 у формі лінійного чотириполюсника подана схема послідовного коливного контуру. Вихідна напруга u_2 пропорційна вхідному струму. Зв'язок між вхідною напругою u_1 та вхідним струмом i_1 (а фактично – й вихідною напругою u_2) в ненавантаженому

режимі ($i_2 = 0$) задається інтегро-диференціальним рівнянням вигляду

$$L\frac{di_1}{dt} + Ri_1 + \frac{1}{C}\int i_1 dt = u_1(t) \qquad (2.1.11)$$

або еквівалентним йому диференціальним рівнянням

$$L\frac{d^2 i_1}{dt^2} + R\frac{di_1}{dt} + \frac{1}{C}i_1 = \frac{du_1}{dt}. \qquad (2.1.12)$$

Нерідко модель лінійного чотириполюсника може бути використана й для опису нелінійних елементів. Це можливо, якщо у напруг і струмів можна виділити малі змінні (сигнальні) складові – див. вище п. 2.1.1. Тоді під i_1, i_2, u_1, u_2 у рівняннях типу (2.1.10) слід розуміти саме сигнальні складові відповідних струмів і напруг.

Контрольні питання до підрозділу 2.1

1. Які електричні кола називаються лінійними?
2. Чи можна біполярний транзистор розглядати як лінійний елемент?
3. Що таке лінійний чотириполюсник?

2.2. Лінійні електричні кола, збуджувані гармонічним джерелом

В радіоелектроніці часто мають справу із гармонічними (або квазігармонічними) сигналами. Крім того, за гармонічного зовнішнього впливу можна легко знаходити розв'язки лінійних диференціальних рівнянь, якими описуються різноманітні радіотехнічні кола. Нарешті, як відзначалося в першій частині цього курсу, сигнали довільної (або майже довільної) форми можна подати як суму гармонік (розкладення в спектр), що дозволяє будувати розв'язки неоднорідних диференціальних рівнянь навіть тоді, коли вплив (зовнішня сила) є негармонічним.

Тому розгляд методів аналізу лінійних електричних кіл ми розпочнемо з кіл, збуджуваних гармонічним джерелом. Такі кола можна описати лінійними неоднорідними диференціальними рівняннями вигляду (2.1.2), в правій частині яких стоїть гармонічна функція часу. Розв'язок такого рівняння, який відповідає усталеному режиму[1], можна шукати у вигляді гармонічних коливань на частоті зовнішньої сили.

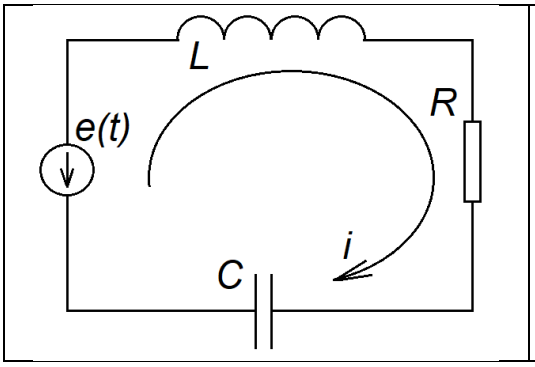


Рис. 2.2.1. Приклад схеми електричного кола, що містить різні пасивні елементи і описується лінійним диференціальним рівнянням зі сталими коефіцієнтами

Наприклад, для кола, схема якого подана на рис. 2.2.1, виходячи з другого правила Кірхгофа, можна записати:

$$L\frac{di}{dt}+Ri+\frac{1}{C}\int idt=e(t), \qquad (2.2.1)$$

де $e(t)$ – е.р.с. джерела напруги, яку ми надалі вважатимемо гармонічною:

$$e(t)=e_m\cos\omega t. \qquad (2.2.2)$$

Зазначимо, що рівняння, подібне до (2.2.1), можна записати для будь-якого електричного кола, складеного з лінійних пасивних елементів.

Рівняння (2.2.1) є інтегро-диференціальним. Працювати з такими рівняннями незручно. Тому його потрібно звести до диференціального рівняння. Для цього можна або продиференціювати його за часом, або перейти від струму до заряду q на конденсаторі, скориставшись співвідношенням (2.1.6). В останньому випадку отримаємо:

$$L\frac{d^2q}{dt^2}+R\frac{dq}{dt}+\frac{1}{C}q=e_m\cos\omega t. \qquad (2.2.3)$$

Зазначимо, що рівняння (2.2.3) є частинним випадком більш загального рівняння (2.1.2). Розв'язок $q(t)$ рівняння (2.2.3), що відповідає вимушеним коливанням, буде дійсною функцією часу.

Розв'язувати рівняння (2.2.12) не дуже зручно, оскільки його розв'язок запишеться не тільки через $\cos\omega t$, але й через $\sin\omega t$. Щоб обійти цю незручність, користуються методом комплексних амплітуд.

---

[1] Усталений режим завжди встановлюється в колах з активними опором через деякий час після увімкнення зовнішнього джерела – після того, як закінчяться перехідні процеси. Таким чином, усталений режим не залежить від початкових умов і описується асимптотикою будь-якого розв'язку рівняння (2.1.2) при $t\to\infty$.

2.2.1. Метод комплексних амплітуд

Неоднорідне диференціальне рівняння (2.2.3) є лінійним, тобто для нього справджується принцип суперпозиції: відгук на суму впливів (зовнішніх сил – того, що стоїть у правій частині) дорівнює сумі відгуків на кожен із впливів.

Додамо до правої частини доданок $iE_m \sin\omega t$. Тоді вона (в силу формули Ейлера $\exp(i\alpha) = \cos\alpha + i\sin\alpha$) набуде вигляду $E_m \exp(i\omega t)$.

В результаті ми отримаємо рівняння вигляду

$$L\frac{d^2Q}{dt^2} + R\frac{dQ}{dt} + \frac{1}{C}Q = e_m \exp(i\omega t), \tag{2.2.4}$$

розв'язок якого пов'язаний із розв'язком рівняння (2.2.3) співвідношенням

$$q(t) = \mathrm{Re}\,Q(t). \tag{2.2.5}$$

Справді, уявній частині зовнішньої сили $iE_m \sin\omega t$ в силу лінійності рівняння (2.2.13) буде, очевидно, відповідати суто уявний розв'язок $i\,\mathrm{Im}\,Q(t)$.

Рівняння (2.2.4) значно зручніше розв'язувати порівняно з рівнянням (2.2.3), оскільки вимушені коливання описуються тією самою функцією $\exp(i\omega t)$, що й зовнішня сила. Але амплітуда цих коливань Q_m буде комплексною:

$$Q_m = |Q_m|\exp(i\varphi), \tag{2.2.6}$$

де величина φ – це зсув фази функції $Q(t)$ (і $q(t)$) щодо зовнішньої сили (в нашому прикладі – е.р.с.) $e(t)$. Справді, дійсна величина $q(t)$ запишеться у вигляді:

$$q(t) = \mathrm{Re}\left[|Q_m|\exp(i\varphi)\exp(i\omega t)\right] = |Q_m|\cos(\omega t + \varphi). \tag{2.2.7}$$

Очевидно, комплексна амплітуда струму I_m, як випливає зі співвідношення $i = dq/dt$, може бути записана у вигляді

$$I_m = i\omega Q_m = \omega|Q_m|\exp\left[i\left(\varphi + \frac{\pi}{2}\right)\right], \tag{2.2.8}$$

звідки

$$i(t) = \mathrm{Re}\left\{I_m \exp\left[i\left(\omega t + \varphi + \frac{\pi}{2}\right)\right]\right\} = \omega|Q_m|\cos\left(\omega t + \varphi + \frac{\pi}{2}\right) = -|I_m|\sin(\omega t + \varphi). \tag{2.2.9}$$

Миттєва потужність, що виділяється при протіканні цього струму через активний опір R, відповідно до закону Джоуля – Ленца, складає

$$P(t) = i^2(t)R = |i_m^2|R\sin^2(\omega t + \varphi) = \frac{1}{2}|i_m^2|R\left\{1 - \cos\left[2(\omega t + \varphi)\right]\right\}, \tag{2.2.10}$$

а середня за часом (середня за період) потужність –

$$\overline{P} \equiv \frac{2\pi}{\omega}\int_0^{2\pi/\omega} P(t)dt = \frac{1}{2}|i_m^2|R = \frac{1}{2}i_m i_m^* R = \frac{1}{2}i_m u_m^*, \qquad U_m = I_m R. \tag{2.2.11}$$

2.2.2 Закон Ома для пасивних кіл гармонічного струму

Повернемося до диференціального рівняння (2.2.4). Шукатимемо його розв'язок, який відповідає суто вимушеним коливанням (коливанням на частоті зовнішньої сили) у формі

$$Q(t) = Q_m \exp(i\omega t). \tag{2.2.12}$$

Підстановка цього розв'язку і скорочення на множник $\exp(i\omega t)$ перетворює диференціальне рівняння (2.2.4) на лінійне неоднорідне алгебраїчне рівняння щодо комплексної амп-

літуди Q_m:

$$-\omega^2 L Q_m + i\omega R Q_m + \frac{1}{C} Q_m = E_m, \tag{2.2.13}$$

або, з урахуванням (2.2.8),

$$i\omega L I_m + R I_m + \frac{1}{i\omega C} I_m = E_m. \tag{2.2.14}$$

Рівняння (2.2.24) можна розглядати як аналог закону Ома для кола з реактивними елементами, якщо ввести позначення для уявних (реактивних) опорів індуктивності та ємності:

$$Z_L = i\omega L, \qquad Z_C = \frac{1}{i\omega C} \tag{2.2.15}$$

(пор. із формулами (2.1.4) та (2.1.9)).

Формально уявний опір означає, що струм і напруга на реактивному елементі зсунуті за фазою на $\pi/2$ (нагадаємо, що $\exp(\pm i\pi/2) = \pm i$). Ми приймали, що

$$u(t) = u_m \cos\omega t = u_m \operatorname{Re}\{\exp(i\omega t)\}$$

(амплітуду U_m вважаємо дійсною й додатною). Тоді

$$i(t) = \frac{u(t)}{Z} = \frac{u_m}{|Z|}\operatorname{Re}\{\pm i\exp(i\omega t)\} = \frac{u_m}{|Z|}\operatorname{Re}\left\{\exp(i\omega t) \pm i\frac{\pi}{2}\right\} = \frac{u_m}{|Z|}\cos\left(\omega t \pm \frac{\pi}{2}\right) = \mp\frac{u_m}{|Z|}\sin\omega t, \tag{2.2.16}$$

або, відповідно до методу комплексних амплітуд,

$$i(t) = \operatorname{Re}\left\{\pm i\frac{u_m}{|Z|}\exp(i\omega t)\right\} = \operatorname{Re}\{i_m \exp(i\omega t)\}, \qquad i_m = \pm i\frac{u_m}{|Z|}. \tag{2.2.17}$$

В результаті для середньої потужності, що виділяється на реактивному елементі, маємо:

$$\overline{P} \equiv \frac{2\pi}{\omega}\int_0^{2\pi/\omega} P(t)dt = \frac{2\pi}{\omega}\int_0^{2\pi/\omega} u_m \cos\omega t \frac{u_m}{|Z|}\sin\omega t\, dt = \frac{\pi}{\omega}\frac{u_m^2}{|Z|}\int_0^{2\pi/\omega}\sin 2\omega t\, dt = 0. \tag{2.2.18}$$

Повна ж середня потужність, записана через комплексні амплітуди, відповідно до формули (2.2.11), дорівнює

$$\overline{P} = \frac{1}{2} i_m u_m^* = \pm i\frac{u_m^2}{|Z|}. \tag{2.2.19}$$

Бачимо, що реактивним опорам відповідають уявні потужності – на реактивних елементах енергія не виділяється у вигляді тепла, а періодично накопичується у вигляді електричного (ємність) та магнітного (індуктивність) полів.

Таким чином, для електричних кіл гармонічного струму вводиться поняття комплексного опору, або імпедансу (від лат. impedio – перешкоджаю).

При аналізі лінійних кіл гармонічного струму методом комплексних амплітуд немає необхідності починати із запису диференціальних рівнянь. Можна відразу записувати алгебраїчні рівняння для комплексних амплітуд, подібні до закону Ома для повного кола. При цьому індуктивностям та ємностям приписуються уявні опори (2.2.15), а замість е.р.с. або джерел струму записуються їхні амплітуди – в загальному випадку комплексні.

2.2.3. Передавальна функція (коефіцієнт передачі)

В силу лінійності вихідних диференціальних рівнянь комплексна амплітуда вимушених коливань (наприклад, амплітуда струму в рівнянні (2.2.14) виявляється прямо пропорційною комплексній амплітуді зовнішньої сили (в даному разі – е.р.с.). Зокрема, з рівняння

(2.2.14) випливає такий вираз для комплексної амплітуди струму:

$$i_m = \frac{e_m}{i\omega L + R + \dfrac{1}{i\omega C}} = K(\omega) e_m . \qquad (2.2.20)$$

Коефіцієнт пропорційності між комплексною амплітудою вимушених коливань та комплексною амплітудою зовнішньої сили називають (комплексним) коефіцієнтом передачі, або передавальною функцією. Очевидно, передавальна функція залежить від вигляду електричного кола, від частоти і від того, яку величину ми розглядаємо як вихідний сигналі (для рівняння (2.2.13) це заряд конденсатора, для рівняння (2.2.14) – струм у колі). В той же час коефіцієнт передачі не залежить від амплітуди зовнішньої сили, оскільки ми розглядаємо лінійні кола.

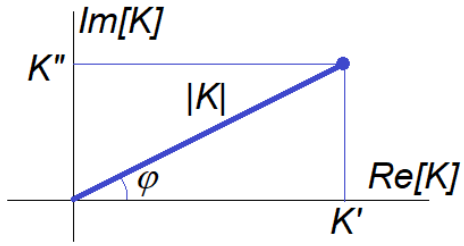


Рис. 2.2.2. Коефіцієнт передачі на комплексній площині

Комплексний коефіцієнт передачі також можна записати через модуль і фазу:

$$K(\omega) = |K(\omega)| \exp\left[i\varphi_K(\omega)\right],$$

$$\varphi_K(\omega) = \arcsin \frac{\mathrm{Im}\left[K(\omega)\right]}{|K(\omega)|} = \arccos \frac{\mathrm{Re}\left[K(\omega)\right]}{|K(\omega)|} = arctg \frac{\mathrm{Im}\left[K(\omega)\right]}{\mathrm{Re}\left[K(\omega)\right]} \qquad (2.2.21)$$

(рис. 2.2.2). Залежність $|K(\omega)|$ називають амплітудно-частотною характеристикою (АЧХ) кола, залежність $\varphi_K(\omega)$ – його фазово-частотною характеристикою (ФЧХ).

Поняття передавальної функції легко узагальнюється на випадок рівняння (2.1.2)

$$a_0 \frac{d^n x}{dt^n} + a_1 \frac{d^{n-1} x}{dt^{n-1}} + \ldots + a_{n-1} \frac{dx}{dt} + a_n x = f(t)$$

з правою частиною вигляду $f(t) = F_m \cos \omega t$. Скориставшись методом комплексних амплітуд, замінимо її на $f(t) = F_m \exp(i\omega t)$. Тоді частинний розв'язок рівняння (2.1.2). що описує відгук на цей вплив, можна шукати у формі $X_m \exp(i\omega t)$. Підстановка цього розв'язку до рівняння (2.1.2) дозволяє відразу знайти комплексну амплітуду відгуку

$$X_m = \frac{F_m}{a_0 (i\omega)^n + a_1 (i\omega)^{n-1} + \ldots + a_{n-1}(i\omega) + a_n} \qquad (2.2.22)$$

і передавальну функцію

$$K(\omega) = \frac{X_m}{F_m} = \frac{1}{a_0 (i\omega)^n + a_1 (i\omega)^{n-1} + \ldots + a_{n-1}(i\omega) + a_n} . \qquad (2.2.23)$$

Оскільки коефіцієнти a_i дійсні, то безпосередньо з формули (2.2.28) випливає, що

$$K(-\omega) = K^*(\omega). \qquad (2.2.24)$$

Для прикладу побудуємо АЧХ і ФЧХ струму в колі, схема якого подана на рис. 2.2.1.

Для цього формулу (2.2.20) запишемо у вигляді[2]

$$K(\omega)=\frac{1}{i\omega L+R+\dfrac{1}{i\omega C}}=\frac{1}{\dfrac{iL}{\omega}\left(\omega^2-\dfrac{1}{LC}\right)+R}=\frac{\omega/L}{i\left(\omega^2-\omega_0^2\right)+2\delta\omega},\quad \omega_0^2=\frac{1}{LC},\qquad \delta=\frac{R}{2L}. \tag{2.2.25}$$

Тоді

$$\left|K(\omega)\right|=\frac{\omega/L}{\sqrt{\left(\omega^2-\omega_0^2\right)^2+\left(2\delta\omega\right)^2}}=\frac{1}{\omega_0 L}\frac{\omega/\omega_0}{\sqrt{\left[\left(\omega/\omega_0\right)^2-1\right]^2+\left[\left(2\delta/\omega_0\right)\left(\omega/\omega_0\right)\right]^2}}, \tag{2.2.26}$$

$$\varphi_K(\omega)=\operatorname{arctg}\frac{\omega_0^2-\omega^2}{2\delta\omega}=\operatorname{arctg}\frac{1-\left(\omega/\omega_0\right)^2}{\left(2\delta/\omega_0\right)\left(\omega/\omega_0\right)}. \tag{2.2.27}$$

Графіки АЧХ і ФЧХ для різних значень безрозмірного параметра $2\delta/\omega_0$ подані на рис. 2.2.3.

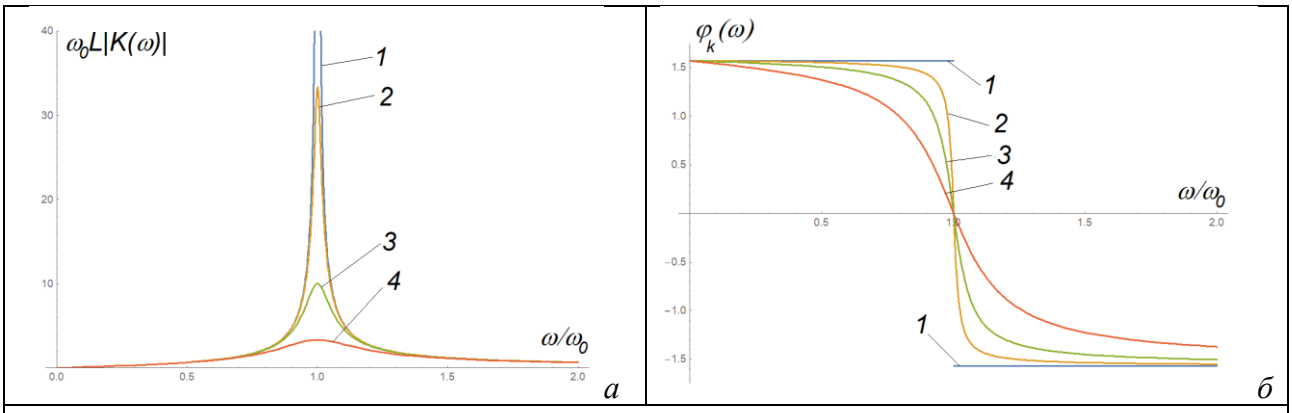


Рис. 2.2.3. Графіки амплітудно-частотної (а) та фазово-частотної характеристики, що відповідають формулам (2.2.26)-(2.2.27): 1 – $2\delta/\omega_0=0$, 2 – $2\delta/\omega_0=0.03$, 3 – $2\delta/\omega_0=0.1$, 4 – $2\delta/\omega_0=0.3$

Схема, подана на рис. 2.2.1, якій відповідають графіки на рис. 2.2.3 – це послідовний коливний контур. Параметром сімей кривих служить величина, обернена до добротності контуру. Нагадаємо, що максимуми кривих на рис. 2.2.3 а відповідають (амплітудному) резонансу, тобто збігу частоти зовнішньої сили ω й частоти власних коливань ω_0. Резонанс тим дужче виявлений, чим більша добротність контуру $Q=\omega_0/2\delta$, якій прямо пропорційна максимальна амплітуда коливань. ФЧХ (рис. 2.2.3 б) стає все більш плавною при зменшенні добротності й проходить через нуль на частоті ω_0.

Контрольні питання до підрозділу 2.2

1. В чому перевага методу комплексних амплітуд порівняно з використанням дійсних гармонічних функцій?
2. Для аналізу яких кіл можна застосувати метод комплексних амплітуд?
3. За яких умов ємність можна розглядати як деякий (реактивний) опір?

---

[2] Тут і далі під δ (без аргументу) розуміємо параметр дисипації (декремент), під $\delta(x)$ (з аргументом) δ - функцію Дірака (див. вище п. 1.3.3).

Приклади розв'язання задач до підрозділу 2.2

П 2.2.1. Скласти диференціальне рівняння, що описує поведінку кола, схема якого подана на рис. 2.2.4.

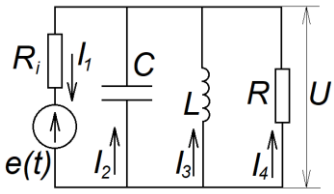


Рис. 2.2.4

Розв'язок

Запишемо струми в гілках через напругу U.

Гілка 1:

$$U = e(t) - i_1 R_i,$$

звідки

$$i_1 = \frac{e(t) - U}{R_i}.$$

Гілка 2:

$$U = \frac{1}{C}\int i_2 dt,$$

звідки

$$i_2 = C\frac{dU}{dt}.$$

Гілка 3:

$$U = L\frac{di_3}{dt},$$

звідки

$$i_3 = \frac{1}{L}\int U dt.$$

Гілка 4:

$$i_4 = \frac{U}{R}.$$

Тепер запишемо рівняння за першим правилом Кірхгофа:

$$i_1 = i_2 + i_3 + i_4,$$

звідки

$$\frac{e(t) - U}{R_i} = C\frac{dU}{dt} + \frac{1}{L}\int U dt + \frac{U}{R},$$

або

$$R_i C\frac{dU}{dt} + \frac{R_i}{L}\int U dt + U\left(\frac{R_i}{R} + 1\right) = e(t).$$

Остаточно маємо:

$$R_i C\frac{d^2U}{dt^2} + \left(\frac{R_i}{R} + 1\right)\frac{dU}{dt} + \frac{R_i}{L}U = \frac{d}{dt}e(t).$$

П 2.2.2. Джерело гармонічної е.р.с. $e(t) = e_m \cos\omega t$ увімкнене в коло, схема якого показана на рис. 2.2.5. Користуючись методом комплексних амплітуд, записати для цього кола закон Ома. Вважаючи, що вихідна напруга знімається з опору навантаження R_l, побудувати графіки АЧХ і ФЧХ для випадку, коли $R_l = \rho \equiv \sqrt{L/C}$.

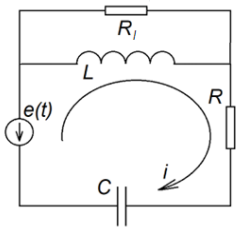


Рис. 2.2.5

Розв'язок

Еквівалентний опір індуктивності та опору навантаження, сполучених паралельно, складає

$$Z_1 = \frac{i\omega L R_l}{i\omega L + R_l}.$$

Тоді, відповідно до закону Ома для повного кола, можна записати для комплексних амплітуд струму та е.р.с. таке співвідношення:

$$i_m\left(R + Z_1 + 1/i\omega C\right) = e_m,$$

звідки

$$i_m = \frac{e_m}{R + Z_1 + 1/i\omega C}.$$

Вихідна напруга:

$$u_m = i_m Z_1 = \frac{e_m Z_1}{R + Z_1 + 1/i\omega C} = \frac{i\omega L R_l/\left(i\omega L + R_l\right)}{R + i\omega L R_l/\left(i\omega L + R_l\right) + 1/i\omega C} e_m =$$

$$= \frac{i\omega L R_l i\omega C}{R i\omega C\left(i\omega L + R_l\right) + i\omega L R_l i\omega C + i\omega L + R_l} e_m = \frac{-\omega^2 L C R_l}{-\omega^2 L C R + i\omega C R R_l - \omega^2 L C R_l + i\omega L + R_l} e_m =$$

$$= \frac{\omega^2 L C R_l}{\omega^2 L C\left(R + R_l\right) - R_l - i\omega\left(C R R_l + L\right)} e_m = \frac{\omega^2/\omega_0^2}{\left(\omega/\omega_0\right)^2\left(R/R_l + 1\right) - 1 - i\left(\omega/\omega_0\right)\left(\omega_0 C R + \omega_0 L/R_l\right)} e_m.$$

За умовою задачі $R_l = \rho$, тому

$$u_m = \frac{\omega^2/\omega_0^2}{\left[\left(\omega/\omega_0\right)^2\left(R/\rho + 1\right) - 1\right] - i\left(\omega/\omega_0\right)\left(R/\rho + 1\right)} e_m.$$

Коефіцієнт передачі:

$$K(\omega) = \frac{u_m}{e_m} = \frac{\omega^2/\omega_0^2}{\left[\left(\omega/\omega_0\right)^2\left(R/\rho + 1\right) - 1\right] - i\left(\omega/\omega_0\right)\left(R/\rho + 1\right)}.$$

АЧХ та ФЧХ:

$$\left|K(\omega)\right| = \frac{\omega^2/\omega_0^2}{\sqrt{\left[\left(\omega/\omega_0\right)^2\left(R/\rho + 1\right)\right]^2 - \left(\omega/\omega_0\right)^2\left(R/\rho + 1\right) + 1}};$$

$$\varphi_K(\omega) = \operatorname{arctg}\frac{\left(\omega/\omega_0\right)\left(R/\rho + 1\right)}{\left(\omega/\omega_0\right)^2\left(R/\rho + 1\right) - 1}$$

(рис. 2.2.6).

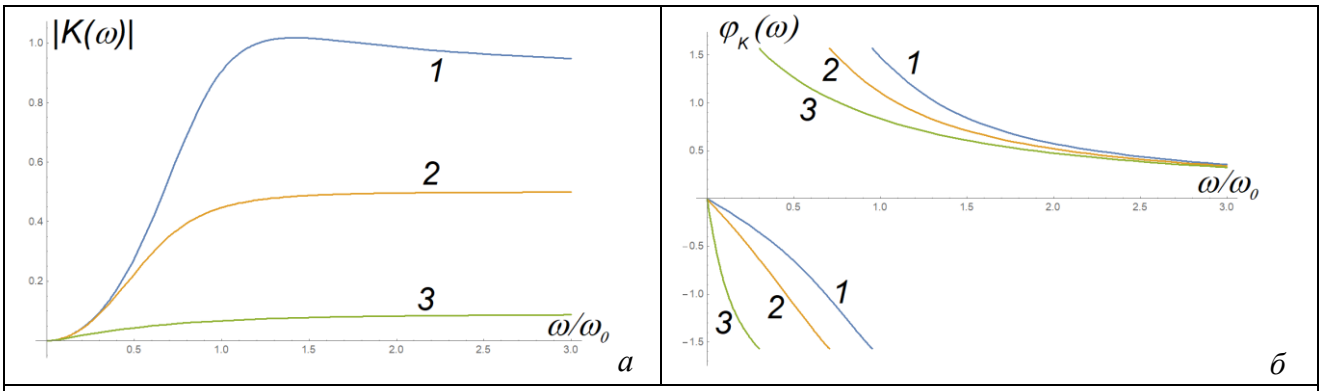


Рис. 2.2.6

П 2.2.3. На вхід кола, схема якого подана на рис. 2.2.3, прикладена напруга $u_1(t) = u_{m1}\cos\omega t$. Знайти h-параметри цього чотириполюсника в ненавантаженому режимі[3], розуміючи під $u_{1,2}$ та $i_{1,2}$ відповідні комплексні амплітуди $u_{m1,2}$ та $i_{m1,2}$.

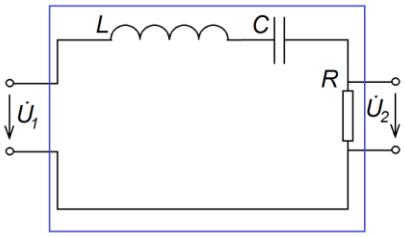


Рис. 2.2.3

Розв'язок

Рівняння (2.1.11)

$$L\frac{di_1}{dt} + Ri_1 + \frac{1}{C}\int i_1 dt = u_1(t)$$

з урахуванням явного вигляду $u_1(t)$ та застосування методу комплексних амплітуд має вигляд

$$i\omega L i_{m1} + Ri_{m1} + i_{m1}/i\omega C = u_{m1}.$$

Візьмемо до уваги, що

$$u_{m2} = Ri_{m1}.$$

Перепишемо попередні рівняння у формі

$$\begin{cases} u_{m1} = i(\omega L - 1/\omega C)i_{m1} + u_{m2}; \\ i_{m2} = 0\cdot i_{m1} + 0\cdot u_{m2}. \end{cases}$$

Порівнюючи отриманий результат із системою (2.1.10)

$$\begin{cases} U_1 = h_{11}I_1 + h_{21}U_2; \\ I_2 = h_{21}I_1 + h_{22}U_2, \end{cases}$$

бачимо, що

$$h_{11} = i(\omega L - 1/\omega C),\ h_{21} = 1,\ h_{21} = 0,\ h_{22} = 0. \qquad (2.2.44)$$

Задачі до підрозділу 2.2

2.2.1. Записати диференціальні рівняння, що описують коло, схему якого подано на рис. 2.2.7.

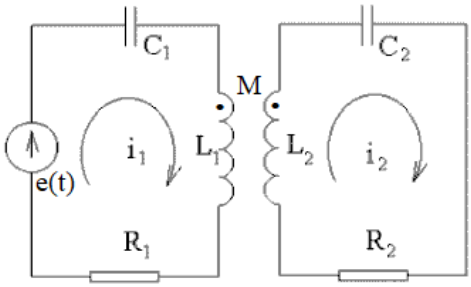


Рис. 2.2.7

2.2.2. Побудувати АЧХ та ФЧХ кіл, схеми яких подані на рис. 2.2.8 а-б.

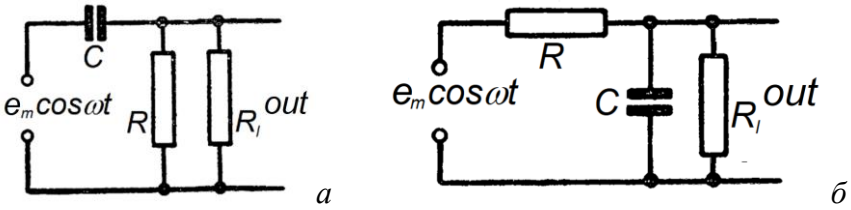


Рис. 2.2.8

[3] Тобто з розімкненим виходом, коли $i_2 = 0$. Його називають також режимом холостого ходу.

2.3. Лінійні електричні кола, збуджувані довільним джерелом

У багатьох випадках сигнали, що надходять на вхід електричного кола, не є гармонічними. Наприклад, такими є різноманітні послідовності прямокутних імпульсів, які використовуються в цифровій електроніці. З негармонічними сигналами зазвичай мають справу при обробленні різного роду експериментальних даних.

Для аналізу електричних кіл, на вхід яких надходять такі сигнали, треба вміти розв'язувати диференціальні рівняння вигляду (2.1.2)

$$a_0 \frac{d^n x}{dt^n} + a_1 \frac{d^{n-1} x}{dt^{n-1}} + \ldots + a_{n-1} \frac{dx}{dt} + a_n x = f(t)$$

з довільною функцією часу в правій частині (довільним впливом, довільною зовнішньою силою).

Для нелінійних кіл і, відповідно, нелінійних диференціальних рівнянь така задача є вельми складною. Однак для лінійних кіл вона різко спрощується, оскільки для лінійних диференціальних рівнянь справджується принцип суперпозиції, про який уже йшлося в п. 2.1.1). Нагадаємо, що, відповідно до цього принципу, відгук на суму впливів дорівнює сумі відгуків на кожен із впливів.

Відповідно до принципу суперпозиції, довільну зовнішню силу, що діє на систему, представляють у вигляді суми деяких «зручних» функцій, для яких можна розв'язати відповідне диференціальне рівняння. Потім розв'язки, що відповідають доданкам у такій сумі, додаються, і ми отримуємо розв'язок рівняння для зовнішньої сили, яка нас цікавить. Прикладом таких «зручних» функцій можуть бути гармоніки, які розглядалися в попередньому підрозділі.

Спочатку ми проілюструємо методи аналізу відгуку лінійних кіл на негармонічні впливи, вважаючи, що від моменту, коли почався вплив, до моменту спостереження пройшло багато часу, й перехідні процеси, зумовлені початковими умовами, встигли згаснути завдяки наявній у системі дисипації (пор. зі вступом до п. 2.2). Формально в цьому випадку можна вважати, що зовнішній вплив почався при $t = -\infty$.

А на закінчення ми розглянемо вплив початкових умов, вважаючи, що вони задані в момент часу $t = 0$, і з цього самого моменту на систему починає діяти зовнішній вплив.

2.3.1.Метод передавальної функції

Як зазначалось у вступі до цього розділу, для знаходження відгуку лінійного кола на довільний зовнішній вплив функцію, яка описує цей вплив, подають у вигляді суми деяких «зручних» функцій, для яких можна розв'язати відповідне лінійне неоднорідне диференціальне рівняння.

В попередньому розділі йшлося про те, що лінійні неоднорідні диференціальні рівняння легко розв'язуються, якщо зовнішня сила є гармонічною.

Тому можливий шлях пошуку відгуку на довільний вплив – це розкладання функції, що описує цей вплив, по гармоніках (див. вище розділи 1.2 – 1.3).

В залежності від властивостей функції $f(t)$, яка описує вхідний сигнал, тут можливі два основні підходи.

<u>Метод ряду Фур'є</u>

Нехай дійсна функція $f(t)$ – періодична з періодом T і задовольняє умови Діріхлє (див. підрозділ 1.2.3). Тоді її можна подати у вигляді ряду Фур'є (1.2.19)

$$f(t) = \sum_{n=-\infty}^{\infty} c_n \exp\left(i \frac{2\pi n}{T} t \right)$$

з коефіцієнтами (1.2.20)

$$c_n = \frac{1}{T}\int_{-T/2}^{T/2} f(t)\exp\left(-i\frac{2\pi n}{T}t\right)dt.$$

Розв'язок рівняння (2.1.2), що відповідає кожному з доданків у правій частині (1.2.19), може бути записаний через передавальну функцію (2.2.28) для відповідної частоти:

$$x_n = K\left(\frac{2\pi n}{T}\right)c_n \exp\left(i\frac{2\pi n}{T}t\right). \qquad (2.3.1)$$

Тоді шуканий розв'язок рівняння (2.1.2) з правою частиною $f(t)$ записується у формі ряду Фур'є:

$$x(t) = \sum_{n=-\infty}^{\infty} K\left(\frac{2\pi n}{T}\right)c_n \exp\left(i\frac{2\pi n}{T}t\right). \qquad (2.3.2)$$

Оскільки $c_{-n} = c_n^*$ (див. розділ 1.2.4) і в силу (2.2.24) $K(-n) = K^*(n)$, то сума кожної пари доданків, у яких n однакові за величиною і протилежні за знаком, дійсна. В результаті відгук $x(t)$ на дійсний вплив $f(t)$ також дійсний, як і має бути.

Щоб проілюструвати цей метод, знайдемо струм у коливному контурі (рис. 2.2.1) під дією нескінченної періодичної послідовності коротких імпульсів вхідної напруги:

$$f(t) = u_m T \sum_{k=-\infty}^{\infty} \delta(t - kT). \qquad (2.3.3)$$

(рис. 2.3.1).

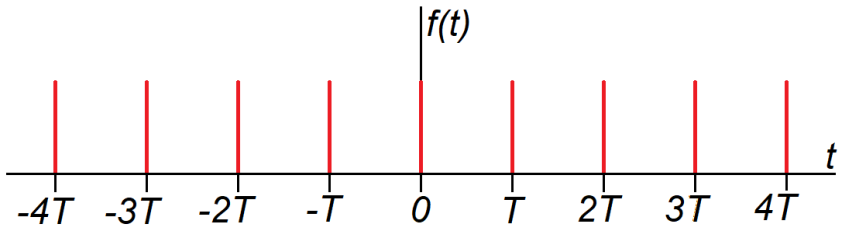


Рис. 2.3.1.Графік зовнішньої сили, що відповідає формулі (2.3.3)

Коефіцієнти ряду Фур'є для цієї функції:

$$c_k = \frac{1}{T}\int_{-T/2}^{T/2} u_m T\delta(t)\exp\left(-i\frac{2\pi k}{T}t\right)dt = u_m \qquad (2.3.4)$$

(враховано формулу 1.3.15), звідки

$$f(t) = u_m \sum_{n=-\infty}^{\infty} \exp\left(-i\frac{2\pi n}{T}t\right). \qquad (2.3.5)$$

Передавальна функція має вигляд (2.2.25):

$$K(\omega) = \frac{\omega/L}{i\left(\omega^2 - \omega_0^2\right) + 2\delta\omega}.$$

Підставивши її та коефіцієнти (2.3.4) до формули (2.3.2), маємо:

$$x(t) = \frac{2\pi u_m}{LT}\sum_{n=-\infty}^{\infty} \frac{n\exp(-i\,2\pi nt/T)}{i\left[(2\pi n/T)^2 - \omega_0^2\right] + 4\pi n\delta/T}. \qquad (2.3.6)$$

Якщо параметр дисипації δ малий, то амплітуда коливань різко зростає за виконання умов

$$\omega_0 = 2\pi n/T,\ n = 1, 2, 3, \ldots \qquad (2.3.7)$$

Це пов'язано з тим, що зовнішня сила (2.3.3) є суттєво негармонічною, і частота влас-

них коливань ω_0 присутня в її спектрі за умови (2.3.7) не тільки при $n=1$, коли вона є основною гармонікою зовнішньої сили, але й при більших значеннях n (тобто при більших значеннях періоду зовнішньої сили *T*).

Описаний вище підхід легко узагальнюється на випадок, коли дискретний спектр зовнішньої сили містить несумірні частоти, тобто ця сила не є періодичною функцією часу (див. вище п. 1.3.6).

Метод інтегралу Фур'є

Нехай тепер дійсна функція $f(t)$ описує сигнал з обмеженою енергією (див. умову (1.3.1)), або, що приблизно те саме, сигнал з обмеженою тривалістю. Тоді вона може бути подана у вигляді інтегралу Фур'є (1.3.2)

$$f(t)=\int_{-\infty}^{\infty} s(\omega)\exp(i\omega t)\,d\omega$$

зі спектром (1.3.3)

$$s(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)\exp(-i\omega t)\,dt\,.$$

За аналогією з (2.3.2) можна, очевидно, в цьому випадку записати розв'язок рівняння (2.1.2) у формі

$$x(t)=\int_{-\infty}^{\infty} K(\omega)s(\omega)\exp(i\omega t)\,d\omega\,. \qquad (2.3.8)$$

Як і в попередньому випадку, дійсному впливу $f(t)$ відповідає дійсний відгук $x(t)$. Інтеграли вигляду (2.3.8) звичайно розраховуються з використанням методу лишків, оскільки передавальна функція $K(\omega)$ містить особливі точки в знаменнику (в загальному випадку – комплексні), в яких її модуль прямує до нескінченності (резонанс на власних частотах).

RC-коло як диференціювальна ланка

Покажемо, що *RC*-коло (див. рис. 2.3.2) може служити диференціювальною ланкою і з'ясуємо, за якої умови це справедливо (розглядаємо випадок, коли вихідна напруга збігається з напругою на резисторі).

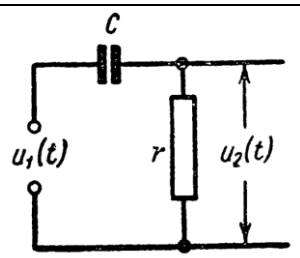


Рис. 2.3.2. Диференціювальна RC*-ланка*

З'ясуємо спершу, яку передавальну функцію $K(\omega)$ повинен мати лінійний чотириполюсник, щоб сигнал на його виході був пропорційним до похідної від вхідного сигналу $u_1(t)$ зі спектром $s(\omega)$. Для цього прирівняємо вихідний сигнал, записаний через передавальну функцію та через спектр похідної (див. вище формулу (1.3.8)):

$$u_2(t)=\int_{-\infty}^{\infty} s(\omega)K(\omega)\exp(i\omega t)\,d\omega=A\frac{d}{dt}\int_{-\infty}^{\infty} s(\omega)\exp(i\omega t)\,d\omega=\int_{-\infty}^{\infty} Ai\omega s(\omega)\exp(i\omega t)\,d\omega\,. \quad (2.3.9)$$

Тут A – деяка стала. Оскільки Фур'є-перетворення однозначне, можемо прирівняти спектри в першому і третьому інтегралах, звідки

$$s(\omega)K(\omega)=i\omega As(\omega)\,,$$

або

$$K(\omega) = i\omega A. \qquad (2.3.10)$$

Тепер звернемося до ланцюжка, заданого в умові задачі. Для нього

$$K(\omega) = \frac{R}{R + 1/i\omega C} = \frac{i\omega RC}{1 + i\omega RC} \approx i\omega RC, \qquad (2.3.11)$$

якщо $\omega RC << 1$.

Отже, цей ланцюжок буде диференціювальним для сигналів, частоти зі спектру яких задовольняють умові $\omega << 1/RC$.

<u>RC-коло як інтегрувальна ланка</u>

Показжемо, що RC-коло (див. рис. 2.3.3) може служити інтегрувальною ланкою (тепер вихідна напруга знімається з конденсатора) і з'ясуємо, за якої умови це справедливо.

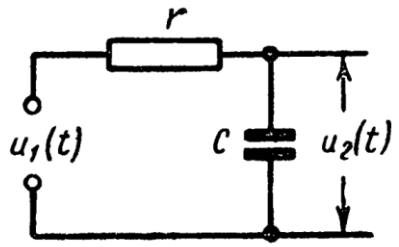


Рис. 2.3.3. Інтегрувальна RC-ланка

Аналогічно до попередньої задачі (тут ми користуємося формулою (1.3.9))

$$u_2(t) = \int_{-\infty}^{\infty} s(\omega) K(\omega) \exp(i\omega t)\, d\omega = A \int dt \int_{-\infty}^{\infty} s(\omega) \exp(i\omega t)\, d\omega = \int_{-\infty}^{\infty} A \frac{s(\omega)}{i\omega} \exp(i\omega t)\, d\omega, \qquad (2.3.12)$$

звідки

$$K(\omega) = \frac{A}{i\omega}. \qquad (2.3.13)$$

Для кола, поданого на рис. 2.3.3,

$$K(\omega) = \frac{1/i\omega C}{R + 1/i\omega C} = \frac{1}{1 + i\omega RC} \approx \frac{1}{i\omega RC}, \qquad (2.3.14)$$

якщо

$\omega RC >> 1$.

Отже, цей ланцюжок буде інтегрувальним для сигналів, частоти зі спектру яких задовольняють умові $\omega \gg 1/RC$.

2.3.2. Метод перехідної функції

Досі ми бачили, що вхідні сигнали лінійного чотириполюсника майже будь-якої форми можна розкласти в ряд або в інтеграл Фур'є і знайти вихідний сигнал за допомогою передавальної функції цього чотириполюсника.

Але в деяких випадках зручніше розкладати вхідний сигнал $f(t)$ по функціях Хевісайда (рис. 2.3.4)

$$\theta(t') = \begin{cases} 1, & t' \geq 0; \\ 0, & t' < 0 \end{cases} \qquad (2.3.15)$$

(пор. з формулою (1.3.20)).

Розіб'ємо інтервал $]-\infty, t[$ на осі абсцис на k ($k = 0, \pm 1, \pm 2, \ldots$) інтервалів тривалістю $\Delta t'$. Тоді значення функції $f(t)$ можна наближено записати у формі

$$f(t) \approx \sum_{k'=-\infty}^{k} \left. \frac{df(t')}{dt'} \right|_{t'=k\Delta t'} \theta(t - k'\Delta t')\Delta t', \qquad (2.3.16)$$

де враховано, що

$$\Delta f\left(t'=k\Delta t'\right)\approx\left.\frac{df\left(t'\right)}{dt'}\right|_{t'=k\Delta\theta}\Delta t'. \qquad (2.3.17)$$

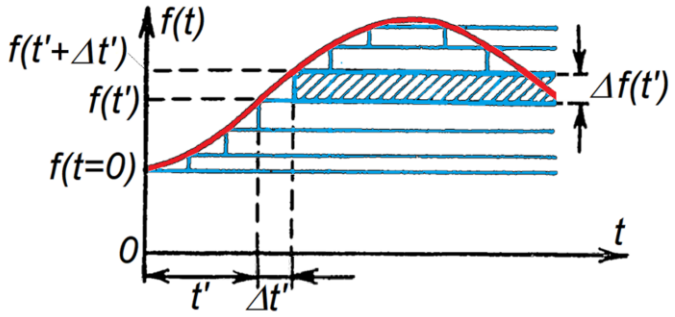


Рис. 2.3.4. Розкладення сигналу по функціях Хевісайда: червоним показана функція, блакитним – її апроксимація набором функцій Хевісайда

Спрямовуючи $\Delta t'\to 0$ і, відповідно, $k\to\infty$ та перейшовши в (2.3.16) від суми до інтеграла, отримаємо розкладення довільної функції по функціях Хевісайда:

$$f\left(t\right)=\int_{-\infty}^{t}\frac{df\left(t'\right)}{dt'}\theta\left(t-t'\right)dt'. \qquad (2.3.18)$$

Відгук системи на зовнішній вплив у формі функції Хевісайда $\theta\left(t-t'\right)$ називається її перехідною функцією $a\left(t,t'\right)$, де t' – момент стрибка сигналу (напруги або струму) на вході лінійного чотириполюсника, t – момент спостереження. Очевидно, $a\left(t<t'\right)=0$, оскільки, відповідно до принципу причинності, відгук не може передувати впливу.

Тоді сигнал $x\left(t\right)$ на виході чотириполюсника, спричинений вхідним сигналом (2.3.18), може також бути поданий у формі інтегралу, аналогічного (2.3.18), але замість функції Хевісайда в ньому фігурує перехідна функція:

$$x\left(t\right)=\int_{-\infty}^{t}\frac{df\left(t'\right)}{dt'}a\left(t,t'\right)dt'. \qquad (2.3.19)$$

Формула (2.3.19) відома в літературі як інтеграл Дюамеля II роду.

Якщо система описується лінійним диференціальним рівнянням зі сталими коефіцієнтами, то її перехідна функція (відгук на стрибок заданої величини) залежить тільки від проміжку часу, який пройшов після стрибка, тобто $a\left(t,t'\right)=a\left(t-t'\right)$.

Для радіотехнічних кіл $f\left(t\right)$ – вхідна напруга або струм, $x\left(t\right)$ – вихідна напруга або струм. Тоді $a\left(t,t'\right)$ – це перехідний процес, спричинений стрибком вхідної напруги одиничної величини, який надійшов на вхід системи в момент часу t' (тому її й називають перехідною функцією).

Знайдемо для прикладу перехідну функцію RC-ланцюжка, поданого на рис. 2.3.3. Вона описується диференціальним рівнянням, яке для заряду на конденсаторі $q\left(t\right)$ має вигляд

$$\frac{1}{C}q+R\frac{dq}{dt}=u_1\left(t\right)=u_m\Theta\left(t\right). \qquad (2.3.20)$$

Вважатимемо, що $q\left(t<0\right)=0$. Відповідно при $t\geq 0$

$$\frac{1}{C}q+R\frac{dq}{dt}=u_m. \qquad (2.3.21)$$

Введемо нову змінну $x=q-Cu_m$. Для неї рівняння (2.3.22) можна подати у формі

$$x+\tau\frac{dx}{dt}=0, \qquad \tau=RC, \tag{2.3.22}$$

звідки

$$x(t)=A\exp(-t/\tau), \tag{2.3.23}$$

або

$$q(t)=Cu_m+A\exp(-t/\tau). \tag{2.3.24}$$

Сталу інтегрування A знаходимо з початкової умови $q(t<0)=0$ і остаточно маємо:

$$q(t)=Cu_m\left[1-\exp(-t/\tau)\right]. \tag{2.3.25}$$

$$u_2(t)=\frac{q}{C}=u_m\left[1-\exp(-t/\tau)\right]. \tag{2.3.26}$$

Перехідна функція:

$$a(t\geq 0)=\frac{u_2}{u_m}=1-\exp(-t/\tau) \tag{2.3.27}$$

(рис. 2.3.5).

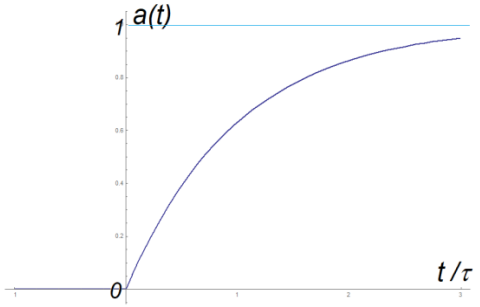


Рис. 2.3.5. Перехідна характеристика RC-ланки

Розглянемо тепер приклад застосування методу перехідної функції.

Припустимо, що на вхід того самого ланцюжка подається напруга, яка лінійно змінюється з часом,

$$u_1(t)=\begin{cases} u_m t/T, & t\geq 0; \\ 0, & t<0. \end{cases} \tag{2.3.28}$$

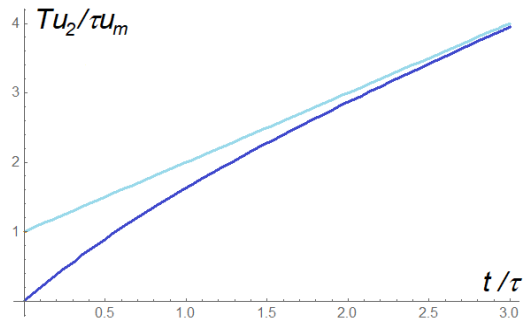


Рис. 2.3.6. Вихідна напруга як функція часу (світлим тоном показана асимптота)

Знайдемо напругу на виході, користуючись формулою (2.3.19).

$$u_2(t)=\int_0^t dt'\frac{u_m}{T}\left\{1-\exp\left[-(t-t')/\tau\right]\right\}=\frac{u_m}{T}\int_0^t dt'\left\{1-\exp\left[-(t-t')/\tau\right]\right\}=$$
$$=\frac{u_m}{T}\left\{t'+\tau\exp\left[-(t-t')/\tau\right]\right\}\Big|_0^t=\frac{u_m}{T}\left\{t+\tau\left[1-\exp(-t/\tau)\right]\right\}=u_m\frac{\tau}{T}\left[\frac{t}{\tau}+1-\exp\left(-\frac{t}{\tau}\right)\right] \tag{2.3.29}$$

(рис. 2.3.6). Через час порядку τ зарядка конденсатора вже відбувається сталим струмом, а на резисторі падає стала напруга.

2.3.3. Метод імпульсної функції

В деяких випадках вхідний сигнал зручно розкладати по коротких імпульсах, які діють на систему не одночасно (як гармоніки або функції Хевісайда), а послідовно в часі (рис. 2.3.7 а).

Розіб'ємо часовий інтервал $]-\infty,\infty[$ на проміжки тривалістю $\Delta t'$, при цьому $t'=k\Delta t'$, $k=0,\pm1,\pm2,\ldots$. Введемо до розгляду функцію

$$\Pi(t,k)\equiv\Theta(t-k\Delta t')-\Theta\left[t-(k+1)\Delta t'\right]=\begin{cases}1, & k\Delta t'<t<(k+1)\Delta t';\\ 0, & t\le k\Delta t'\cup t\ge(k+1)\Delta t'.\end{cases} \tag{2.3.30}$$

Можна записати, що для довільної функції $f(t')$ справджується наближена рівність

$$f(t)\approx\sum_{k=-\infty}^{\infty}f(k\Delta t')\Pi(t,k)=\sum_{k=-\infty}^{\infty}f(k\Delta t')\frac{\Pi(t,k)}{\Delta t'}\Delta t' \tag{2.3.31}$$

(ми, як і в попередньому випадку, замінили неперервну функцію $f(t')$ ступінчастою, див. рис. 2.3.7 а). В сумі, яка стоїть у правій частині, в кожний момент часу відмінний від нуля лише один доданок.

Якщо $\Delta t'\to0$, то

$$\lim_{\Delta t'\to0}\frac{\Pi(t,k)}{\Delta t'}=\lim_{\Delta t'\to0}\frac{\Theta(t-k\Delta t')-\Theta\left[t-(k+1)\Delta t'\right]}{\Delta t'}=\frac{d\Theta(t-k\Delta t')}{dt'}=\delta(t-k\Delta t') \tag{2.3.32}$$

(див. п. 1.3.3). Тоді замість суми (2.3.31) отримаємо інтеграл вигляду

$$f(t)=\int_{-\infty}^{\infty}f(t')\delta(t-t')dt'. \tag{2.3.33}$$

Формально (2.3.33) можна інтерпретувати як розкладання функції $f(t)$ по δ-функціях Дірака.

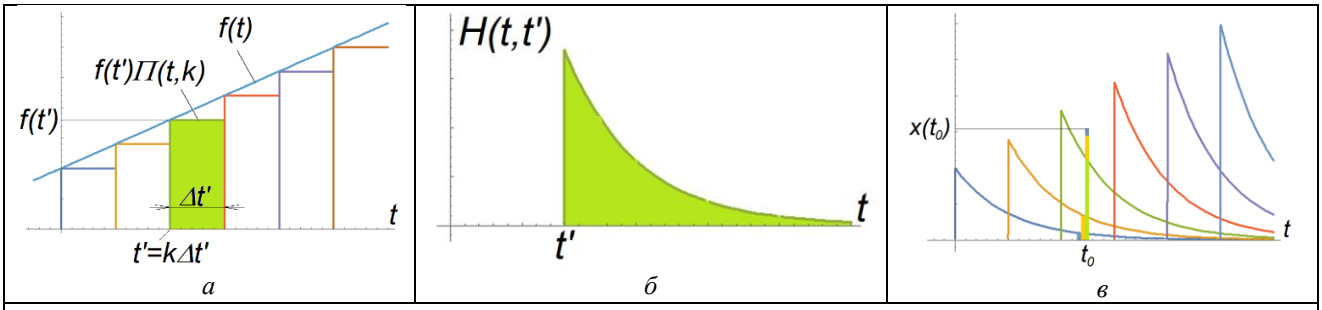


Рис. 2.3.7. До пояснення змісту імпульсної функції (функції Гріна): а – розкладення вхідного сигналу по коротких імпульсах; б – функція Гріна (показана умовно); в – вихідний сигнал як сума відгуків на окремі вхідні імпульси

Відгук лінійної системи на одиночний короткий імпульс називають імпульсною функцією (функцією Гріна) системи і зазвичай позначають як $H(t,t')$, де t' – момент дії короткого імпульсу, а t – момент спостереження. Оскільки, як уже вказувалося, відгук системи спостерігається лише після дії на неї, то $H(t<t',t')=0$ (рис. 2.3.7 б; пор. з перехідною функцією, див. п. 2.3.2). Якщо властивості системи не змінюються з часом, тобто вона описується диференціальним рівнянням зі сталими коефіцієнтами, то відгук на дію короткого імпуль-

су залежить лише від проміжку часу після його дії, тобто $H(t,t') = H(t-t')$ (як уже вказувалося, аналогічну властивість має й перехідна функція).

Очевидно, при поданні на вхід довільного сигналу, який розкладений по δ-функціях, вихідний сигнал можна подати як суму відгуків на такі функції, які встигли подіяти на сигнал до моменту спостереження:

$$x(t) = \int_{-\infty}^{t} f(t')H(t,t')dt'. \qquad (2.3.34)$$

Запис вихідного сигналу у формі (2.3.34) називають інтегралом Дюамеля I роду. Для систем, властивості яких не змінюються з часом, він набуває вигляду

$$x(t) = \int_{-\infty}^{t} f(t')H(t-t')dt'. \qquad (2.3.34\ а)$$

Нехай досліджувана система описується диференціальним рівнянням вигляду (2.1.2):

$$a_0\frac{d^n x}{dt^n} + a_1\frac{d^{n-1}x}{dt^{n-1}} + \ldots + a_{n-1}\frac{dx}{dt} + a_n x = f(t), \qquad a_0 \neq 0.$$

Нехай відповідне однорідне рівняння має набір лінійно незалежних розв'язків $x_1(t), x_2(t), x_3(t)\ldots$ Тоді імпульсна функція (функція Гріна) даної системи може бути записана у вигляді:

$$H(t>t',t') = \frac{1}{a_0}\frac{\begin{vmatrix} x_1(t') & x_2(t') & \ldots & x_n(t') \\ \dot{x}_1(t') & \dot{x}_2(t') & \ldots & \dot{x}_n(t') \\ \ldots & \ldots & \ldots & \ldots \\ x_1^{(n-2)}(t') & x_2^{(n-2)}(t') & \ldots & x_n^{(n-2)}(t') \\ x_1(t) & x_2(t) & \ldots & x_n(t) \end{vmatrix}}{\begin{vmatrix} x_1(t') & x_2(t') & \ldots & x_n(t') \\ \dot{x}_1(t') & \dot{x}_2(t') & \ldots & \dot{x}_n(t') \\ \ldots & \ldots & \ldots & \ldots \\ x_1^{(n-1)}(t') & x_2^{(n-1)}(t') & \ldots & x_n^{(n-1)}(t') \end{vmatrix}}, \qquad \dot{x} = \frac{dx}{dt}, \quad x^{(k)} = \frac{d^k x}{dt^k}. \qquad (2.3.35)$$

Розглянемо для прикладу коло, про яке йшлося у розділі 2.1 (рис. 2.2.1). Для нього ми записали диференціальне рівняння у формі (2.2.10)

$$L\frac{di}{dt} + Ri + \frac{1}{C}\int idt = e(t).$$

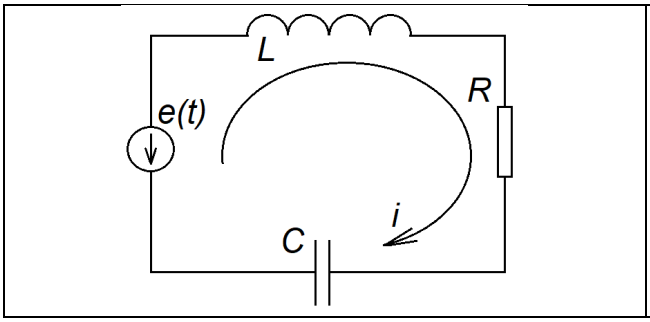


Рис. 2.2.1. Схема послідовного коливного контуру з джерелом е.р.с.

Продиференціювавши рівняння (2.2.1) перепишемо його у формі

$$\frac{d^2 i}{dt^2} + 2\delta\frac{di}{dt} + \omega_0^2 i = \frac{1}{L}\frac{de(t)}{dt}, \qquad \delta = r/2L,\ \omega_0^2 = 1/LC. \qquad (2.3.36)$$

Рівняння (2.3.36) – лінійне зі сталими коефіцієнтами, але неоднорідне. Йому відповідає однорідне рівняння

$$\frac{d^2 i}{dt^2} + 2\delta \frac{di}{dt} + \omega_0^2 i = 0 . \qquad (2.3.37)$$

Знайдемо спочатку розв'язок рівняння (2.3.7). Як відомо, розв'язки лінійного однорідного диференціального рівняння зі сталими коефіцієнтами завжди можна шукати в експоненціальній формі:

$$i(t) = A\exp(\lambda t) . \qquad (2.3.38)$$

Підставимо розв'язок (2.3.38) до рівняння (2.3.37). Отримаємо:

$$\lambda^2 A\exp(\lambda t) + 2\delta\lambda A\exp(\lambda t) + \omega_0^2 A\exp(\lambda t) = 0 ,$$

або, після скорочення на $A\exp(\lambda t)$,

$$\lambda^2 + 2\delta\lambda + \omega_0^2 = 0 . \qquad (2.3.39)$$

Це так зване характеристичне рівняння, з якого можна знайти невідому величину λ:

$$\lambda_{1,2} = -\delta \pm \sqrt{\delta^2 - \omega_0^2} = -\delta \pm i\omega, \quad \omega = \sqrt{\omega_0^2 - \delta^2} \qquad (2.3.40)$$

(вважатимемо для визначеності, що $\delta \le \omega_0$). Тоді лінійно незалежні розв'язки рівняння (2.3.37) можна записати у формі

$$i_{1,2}(t) = \exp(-\delta t)\exp(\pm i\omega t), \qquad (2.3.41)$$

або

$$i_1(t) = \cos(\omega t)\exp(-\delta t), \qquad i_2(t) = \sin(\omega t)\exp(-\delta t) . \qquad (2.3.42)$$

Тепер знайдемо імпульсну функцію (функцію Гріна), яка відповідає неоднорідному рівнянню (2.3.36). Для цього його лінійно незалежні розв'язки (2.3.41) підставимо до формули (2.3.35). Порівнявши рівняння (2.3.37) і (2.1.2), бачимо, що $a_0 = 1$. Рівняння (2.3.37) – другого порядку, тому визначники у формулі (2.3.35) мають розмір 2×2. Маємо:

$$H(t > t', t') = \frac{\begin{vmatrix} \exp(-\delta t')\exp(i\omega t') & \exp(-\delta t')\exp(-i\omega t') \\ \exp(-\delta t)\exp(i\omega t) & \exp(-\delta t)\exp(-i\omega t) \end{vmatrix}}{\begin{vmatrix} \exp(-\delta t')\exp(i\omega t') & \exp(-\delta t')\exp(-i\omega t') \\ (-\delta + i\omega)\exp(-\delta t')\exp(i\omega t') & (-\delta - i\omega)\exp(-\delta t')\exp(-i\omega t') \end{vmatrix}} . \qquad (2.3.43)$$

Винесемо із кожного стовпчика перший елемент:

$$H(t > t', t') = \frac{\begin{vmatrix} 1 & 1 \\ \exp[-\delta(t-t')]\exp[i\omega(t-t')] & \exp[-\delta(t-t')]\exp[-i\omega(t-t')] \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ (-\delta + i\omega) & (-\delta - i\omega) \end{vmatrix}} \times \qquad (2.3.44)$$

$$\times \frac{\exp(-\delta t')\exp(i\omega t')\exp(-\delta t')\exp(-i\omega t')}{\exp(-\delta t')\exp(i\omega t')\exp(-\delta t')\exp(-i\omega t')} .$$

Тепер винесемо спільний множник із другого рядка верхнього визначника:

$$H(t > t', t') = \frac{\begin{vmatrix} 1 & 1 \\ \exp[i\omega(t-t')] & \exp[-i\omega(t-t')] \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ (-\delta + i\omega) & (-\delta - i\omega) \end{vmatrix}} \exp[-\delta(t-t')] =$$

$$= \frac{\exp[-i\omega(t-t')] - \exp[i\omega(t-t')]}{-\delta - i\omega - (-\delta + i\omega)} \exp[-\delta(t-t')] =$$

$$= \frac{\exp\left[i\omega\left(t-t'\right)\right]-\exp\left[-i\omega\left(t-t'\right)\right]}{2i\omega}\exp\left[-\delta\left(t-t'\right)\right]=$$
$$= \frac{1}{\omega}\sin\left[\omega\left(t-t'\right)\right]\exp\left[-\delta\left(t-t'\right)\right]. \qquad (2.3.45)$$

Формула (2.3.45) отримана для рівняння (2.3.36). Для еквівалентного рівняння (2.2.1)

$$H\left(t>t',t'\right)=\frac{1}{\omega L}\sin\left[\omega\left(t-t'\right)\right]\exp\left[-\delta\left(t-t'\right)\right]. \qquad (2.3.45\text{ а})$$

Знайдемо, як змінюватиметься струм у контурі, зображеному на рис. 2.2.1, при поданні на його вхід напруги вигляду (2.3.28), яка лінійно зростає з часом:

$$e(t)=\begin{cases} u_m t/T, & t\ge 0; \\ 0, & t<0. \end{cases}$$

Права частина рівняння (2.3.36):

$$\frac{1}{L}\frac{de(t)}{dt}=\begin{cases} u_m/TL, & t\ge 0; \\ 0, & t<0. \end{cases} \qquad (2.3.46)$$

Підставивши (2.3.45) та (2.3.46) до інтегралу (2.3.34), отримуємо, вважаючи початкові умови нульовими й скориставшись представленням синуса через експоненти з уявними показниками:

$$i(t)=i(t=0)+\int_0^t \frac{u_m}{TL}\frac{1}{\omega}\sin\left[\omega\left(t-t'\right)\right]\exp\left[-\delta\left(t-t'\right)\right]dt'=$$
$$=\frac{u_m}{\omega TL}\int_0^t \sin\left[\omega\left(t-t'\right)\right]\exp\left[-\delta\left(t-t'\right)\right]dt'=$$
$$=\frac{u_m}{2i\omega TL}\left\{\int_0^t \exp\left[\left(-\delta+i\omega\right)\left(t-t'\right)\right]dt'-\int_0^t \exp\left[\left(-\delta-i\omega\right)\left(t-t'\right)\right]dt'\right\}=\left\|\begin{matrix} t-t'=\tau \\ dt'=-d\tau \end{matrix}\right\|=$$
$$=\frac{u_m}{2i\omega TL}\left\{\int_0^t \exp\left[\left(-\delta+i\omega\right)\tau\right]d\tau-\int_0^t \exp\left[\left(-\delta-i\omega\right)\tau\right]d\tau\right\}=$$
$$=\frac{u_m}{2i\omega TL}\left\{\frac{\exp\left[\left(\delta+i\omega\right)\tau\right]}{-\delta+i\omega}-\frac{\exp\left[\left(-\delta-i\omega\right)\tau\right]}{-\delta-i\omega}\right\}\Bigg|_0^t=\frac{u_m}{2i\omega TL}\left\{\frac{\exp\left[\left(-\delta+i\omega\right)t\right]-1}{-\delta+i\omega}-\frac{\exp\left[\left(-\delta-i\omega\right)t\right]-1}{-\delta-i\omega}\right\}=$$
$$=\frac{u_m}{2i\omega TL}2i\,\mathrm{Im}\left\{\frac{\exp\left[\left(-\delta+i\omega\right)t\right]-1}{-\delta+i\omega}\right\}=\frac{u_m}{\omega TL}\mathrm{Im}\left\{\frac{\left(-\delta-i\omega\right)\left[\exp\left[\left(-\delta+i\omega\right)t\right]-1\right]}{\delta^2+\omega^2}\right\}=$$
$$=\frac{u_m}{\omega TL\left(\delta^2+\omega^2\right)}\mathrm{Im}\left\{\left(-\delta-i\omega\right)\left[\exp\left(-\delta t\right)\left(\cos\omega t+i\sin\omega t\right)-1\right]\right\}= \qquad (2.3.47)$$
$$=\frac{u_m}{\omega\omega_0^2 TL}\left\{-\omega\left[\cos\omega t\exp\left(-\delta t\right)-1\right]-\delta\sin\omega t\exp\left(-\delta t\right)\right\}=-\frac{Cu_m}{T}\left\{\exp\left(-\delta t\right)\left[\cos\omega t+\frac{\delta}{\omega}\sin\omega t\right]+1\right\}$$

(рис. 2.3.8; враховано, що $\tau=0$ при $t'=t$ і $\tau=t$ при $t'=0$).

У пізні моменти часу ($\delta t>>1$) коливання згасають, і встановлюється постійний струм $i=-Cu_m/T$.

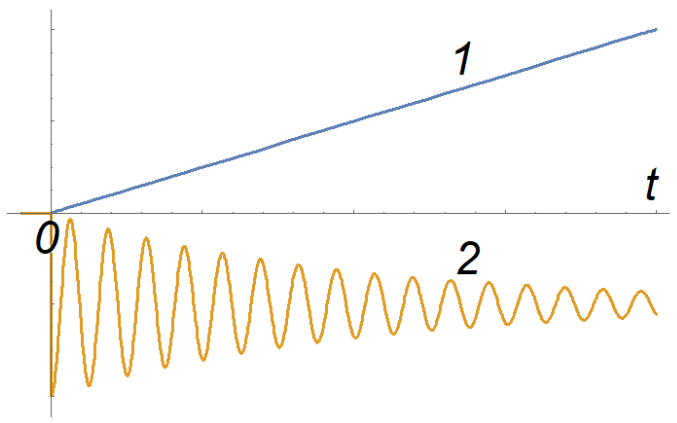


Рис. 2.3.8. Часовий хід вхідної напруги (1) та струму в контурі (2)

2.3.4. Зв'язок між імпульсною, перехідною та передавальною функцією

Всі введені раніше функції чотириполюсника – передавальна, перехідна та імпульсна –описують зміну сигналу при його проходженні через цей чотириполюсник і, отже, повинні бути пов'язані між собою.

Зв'язок між імпульсною та перехідною функціями

Як уже вказувалося, імпульсна функція (функція Гріна) лінійного чотириполюсника – це реакція на вхідний сигнал у вигляді δ-функції. Скористаємось поданням останньої у формі (1.3.11)-(1.3.12):

$$\delta(t)=\lim_{\Delta t\to 0}\frac{\Theta(t)-\Theta(t-\Delta t)}{\Delta t}.$$

Тоді функцію Гріна відповідно до принципу суперпозиції можна записати у вигляді:

$$H(t,t')=\lim_{\Delta t'\to 0}\frac{a(t,t')-a(t,t'-\Delta t')}{\Delta t'}=\frac{da(t,t')}{dt'}. \qquad (2.3.48)$$

Для лінійних стаціонарних систем (описуваних лінійними диференціальними рівняннями зі сталими коефіцієнтами), як ми вказували раніше, $H(t,t')=H(t-t')$ і $a(t,t')=a(t-t')$. Тому в цьому випадку

$$H(t-t')=a'(t-t'), \qquad t\ge t'. \qquad (2.3.49)$$

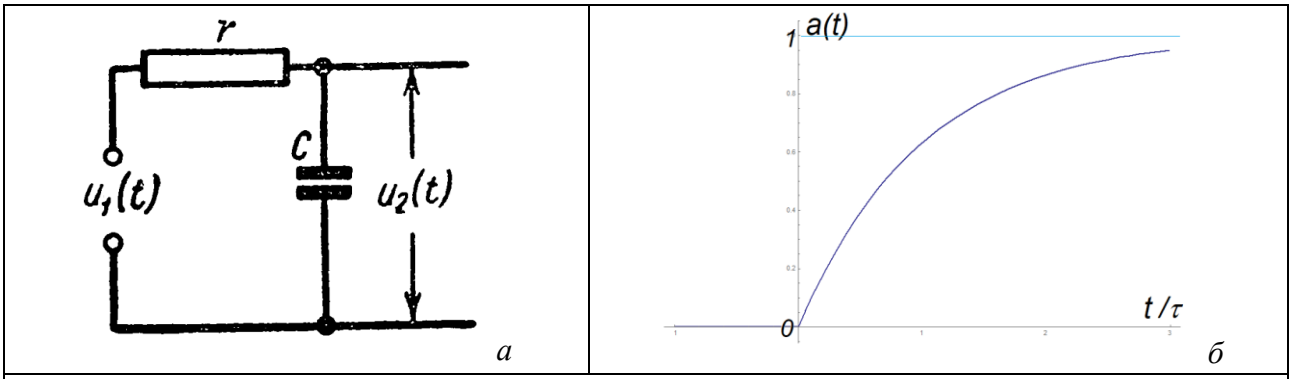


Рис. 2.3.9. Схема інтегрувального RC-ланцюжка(а) та графік його перехідної функції (б)

Проілюструємо ці результати на найпростішому прикладі *RC*-ланцюжка, який розглядався у п. 2.3.3 (рис. 2.3.9 а). Перехідна функція для нього має вигляд (2.3.27) – див. рис. 2.3.9 б:

$$a(t-t')=1-\exp\left[-(t-t')/\tau\right], \qquad \tau=RC.$$

Відповідно

$$a'(t-t')=\frac{1}{\tau}\exp\left[-(t-t')/\tau\right]. \qquad (2.3.50)$$

З іншого боку, однорідне диференціальне рівняння, яке описує цей ланцюжок при закороченому вході й розімкнутому виході, має вигляд

$$q+\tau\frac{dq}{dt}=0,$$

його розв'язок –

$$q(t)=A\exp(-t/\tau),$$

а функція Гріна, відповідно до формули (2.3.35) –

$$H(t,t')=\frac{1}{\tau}\exp\left[-(t-t')/\tau\right] \qquad (2.3.51)$$

(враховано, що $a_0=1/\tau$). Ця формула описує часовий хід заряду конденсатора й збігається з отриманою раніше формулою (2.3.50).

Отже, ми переконалися на цьому прикладі, що загальне співвідношення (2.3.49) справджується.

<u>Зв'язок між передавальною та імпульсною функцією</u>

Для простоти обмежимося розглядом лінійних стаціонарних систем. Запишемо вихідний сигнал через імпульсну функцію (див. формулу (2.3.34 а)), а потім замінимо вхідний сигнал його спектром:

$$x(t)=\int_{-\infty}^{t}H(t-t')f(t')dt'=\int_{-\infty}^{t}H(t-t')dt'\int_{-\infty}^{\infty}s(\omega)\exp(i\omega t')d\omega. \qquad (2.3.52)$$

Оскільки $H(t-t')=0$ при $t'>t$, ми можемо верхню межу інтегрування замінити на нескінченність. Крім того, поміняємо місцями інтегрування за часом та за частотою. Маємо:

$$x(t)=\int_{-\infty}^{\infty}H(t-t')dt'\int_{-\infty}^{\infty}s(\omega)\exp(i\omega t')d\omega=\int_{-\infty}^{\infty}s(\omega)d\omega\int_{-\infty}^{\infty}H(t-t')\exp(i\omega t')dt'. \qquad (2.3.53)$$

Замість t' введемо нову змінну інтегрування $\tau=t-t'$, звідки $dt'=-d\tau$, $t'=t-\tau$, $\tau\to\mp\infty$ при $t'\to\pm\infty$. Отримуємо:

$$x(t)=\int_{-\infty}^{\infty}s(\omega)d\omega\int_{-\infty}^{\infty}H(\tau)\exp\left[i\omega(t-\tau)\right]d\tau=\int_{-\infty}^{\infty}s(\omega)\exp(i\omega t)d\omega\int_{-\infty}^{\infty}H(\tau)\exp(-i\omega\tau)d\tau. \qquad (2.3.54)$$

З іншого боку, відповідно до формули (2.3.8),

$$x(t)=\int_{-\infty}^{\infty}s(\omega)K(\omega)\exp(i\omega t)d\omega. \qquad (2.3.55)$$

В силу однозначності перетворення Фур'є однакові сигнали мають однакові спектри, тобто з формул (2.3.54)-(2.3.55) випливає рівність

$$K(\omega)=\int_{-\infty}^{\infty}H(\tau)\exp(-i\omega\tau)d\tau=\int_{0}^{\infty}H(\tau)\exp(-i\omega\tau)d\tau \qquad (2.3.56)$$

(враховано, що $H(\tau<0)=0$). Формально з рівності (2.3.56) випливає, що функцію Гріна можна розглядати як спектр передавальної функції. Тоді для них можна написати й зворотне перетворення Фур'є:

$$H(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}K(\omega)\exp(i\omega\tau)d\omega. \qquad (2.3.57)$$

Для прикладу знову розглянемо RC-ланцюжок, описуваний рівнянням (2.3.20) (див.

рис. 2.3.3), імпульсна функція (функція Гріна) якого має вигляд (2.3.51). Його передавальну функцію легко знайти, замінивши праву частину на $\exp(i\omega t)$. Шукатимемо розв'язок такого рівняння у формі $K(\omega)\exp(i\omega t)$. Маємо:

$$K(\omega)\exp(i\omega t)+i\omega\tau K(\omega)\exp(i\omega t)=\exp(i\omega t),$$

або

$$K(\omega)=\frac{1}{1+i\omega\tau}, \qquad \tau=RC. \tag{2.3.58}$$

З іншого боку, права частина (2.3.56) набуває вигляду

$$K(\omega)=\int_0^\infty \frac{1}{\tau}\exp(-\theta/\tau)\exp(-i\omega\theta)d\theta=\left.\frac{\exp\left[-(i\omega+1/\tau)\theta\right]}{-(i\omega+1/\tau)\tau}\right|_0^\infty=\frac{1}{1+i\omega\tau}. \tag{2.3.59}$$

Отже, в розглянутому випадку співвідношення (2.3.52) справджується.

<u>Зв'язок між передавальною та перехідною функціями</u>

Для знаходження такого зв'язку скористаємося співвідношеннями (2.3.48) та (2.3.56):

$$K(\omega)=\int_0^\infty H(\tau)\exp(-i\omega\tau)d\tau=\int_0^\infty \frac{da(\tau)}{d\tau}\exp(-i\omega\tau)d\tau=$$

$$=\int_0^\infty \exp(-i\omega\tau)da(\tau)=a(\tau)\exp(-i\omega\tau)\Big|_0^\infty-\int_0^\infty a(\tau)d\left[\exp(-i\omega\tau)\right]= \tag{2.3.60}$$

$$=a(\tau)\exp(-i\omega\tau)\Big|_0^\infty-(-i\omega)\int_0^\infty a(\tau)\exp(-i\omega)d\tau.$$

Реальні системи завжди є дисипативними, тому для них відгук на короткий імпульс чи стрибок напруги триває протягом обмеженого часу. Це означає, що $a(\infty)=0$. В момент часу $t=0$ коло ще не встигло відреагувати на вплив, тому можна вважати, що $a(0)=0$. В результаті з (2.3.60) маємо:

$$K(\omega)=i\omega\int_0^\infty a(\tau)\exp(-i\omega\tau)d\tau. \tag{2.3.61}$$

Відповідно

$$a(\tau)=\int_0^\infty \frac{K(\omega)}{i\omega}\exp(i\omega\tau)d\omega. \tag{2.3.62}$$

2.3.5. Врахування ненульових початкових умов

Розглянемо тепер, відповідно до вказаного у вступній частині до цього підрозділу, дещо іншу постановку задачі. Вважатимемо, що зовнішня сила починає діяти на систему в момент часу $t=0$. Крім того, врахуємо початкові умови (взагалі кажучи, ненульові) для диференціального рівняння *n*-го порядку (2.1.2)

$$a_0\frac{d^nx}{dt^n}+a_1\frac{d^{n-1}x}{dt^{n-1}}+\ldots+a_{n-1}\frac{dx}{dt}+a_nx=f(t)$$

в цей момент часу:

$$x(t=0)=X_0,\ x'(t=0)=X_1,\ x''(t=0)=X_2,\ldots\ x^{(n-1)}(t=0)=X_n. \tag{2.3.63}$$

В цьому випадку розв'язки, записані через передавальну функцію (див. формулу (2.3.8)), через перехідну функцію (див. формулу (2.3.19)) та через імпульсну функцію (див. формулу (2.3.34)), зазнають змін. По-перше, в двох останніх випадках нижня межа інтегрування заміняється з $t'=-\infty$ на $t'=0$. По-друге, до цих розв'язків, що описують суто вимушені коливання під дією зовнішньої сили, треба додати доданки вигляду

$$x_{tr}(t)=\sum_{k=1}^{n}C_k x_k(t), \tag{2.3.64}$$

що враховують перехідні процеси, зумовлені ненульовими початковими умовами (індекс – від англ. transitional, перехідний). У формулі (2.3.64) функції $x_k(t)$ ($k=1,2,3,...n$), як і в формулі (2.3.35) – розв'язки відповідного (2.1.2) однорідного рівняння

$$a_0\frac{d^n x}{dt^n}+a_1\frac{d^{n-1}x}{dt^{n-1}}+...+a_{n-1}\frac{dx}{dt}+a_n x=0,$$

а коефіцієнти C_k знаходяться шляхом підстановки розв'язку (2.3.64) до граничних умов.

Таким чином, замість формул (2.3.8), (2.3.19) та (2.3.34) отримуємо відповідно

$$x(t)=\sum_{k=1}^{n}C_k x_k(t)+\int_{-\infty}^{\infty}K(\omega)s(\omega)\exp(i\omega t)d\omega; \tag{2.3.65}$$

$$x(t)=\sum_{k=1}^{n}C_k x_k(t)+\int_{0}^{t}\frac{df(t')}{dt'}a(t,t')dt'; \tag{2.3.66}$$

$$x(t)=\sum_{k=1}^{n}C_k x_k(t)+\int_{0}^{t}f(t')H(t,t')dt'. \tag{2.3.67}$$

У двох останніх формулах враховано, що зовнішня сила починає діяти на систему з моменту часу $t=0$.

2.3.6. Заміна ненульових початкових умов еквівалентною зовнішньою силою

Розглянемо ще один підхід до розв'язання рівняння (2.1.2) з ненульовими початковими умовами – заміну таких умов зовнішньою силою у формі короткого імпульсу, що діє на систему в початковий момент часу.

Обмежимося максимально простим випадком – однорідним диференціальним рівнянням першого порядку вигляду

$$a_0\frac{dx}{dt}+a_1 x=0 \tag{2.3.68}$$

з початковою умовою

$$x(t=0)=X_0. \tag{2.3.69}$$

Знайдемо еквівалентну зовнішню силу $f(t)$, яка дозволяє звести співвідношення (2.3.68)-(2.3.69) до рівняння вигляду

$$a_0\frac{dx}{dt}+a_1 x=f_{eq}(t) \tag{2.3.70}$$

з нульовими початковими умовами. З урахуванням (2.3.69) можна вважати, що еквівалентна зовнішня сила $f_{eq}(t)$ починає діяти на систему з моменту $t=0$, тобто

$$f_{eq}(t<0)=0. \tag{2.3.71}$$

Відповідно й $x(t<0)=0$.

Введемо до розгляду функцію

$$x^{+}(t)=x(t)\Theta(t), \tag{2.3.72}$$

де, як і раніше, $\Theta(t)$ – функція Хевісайда (одиничний стрибок). Легко бачити, що функція $x^{+}(t)$ відповідає нульовим початковим умовам. Знайдемо похідну цієї функції:

$$\frac{dx^{+}(t)}{dt}=\frac{dx(t)}{dt}+x(t)\frac{d\Theta(t)}{dt}=\frac{dx(t)}{dt}+\delta(t)x(t)=\frac{dx(t)}{dt}+\delta(t)x(t=0)=\frac{dx(t)}{dt}+\delta(t)X_0.$$

(2.3.73)

(тут враховано початкову умову (2.3.69)), звідки

$$\frac{dx(t)}{dt} = \frac{dx^{+}(t)}{dt} - \delta(t) X_0 . \qquad (2.3.74)$$

Тепер отримаємо диференціальне рівняння для функції $x^{+}(t)$, врахувавши її явний вигляд (2.3.73) та співвідношення (2.3.75):

$$a_0 \frac{dx^{+}}{dt} + a_1 x^{+} = a_0 X_0 \delta(t), \qquad (2.3.75)$$

причому функція $x^{+}(t)$, як уже зазначалося, задовольняє нульові початкові умови. Таким чином,

$$f_{eq}(t) = a_0 X_0 \delta(t). \qquad (2.3.76)$$

Таким чином, формулу (2.3.75) можна розглядати як приклад зведення ненульових початкових умов до еквівалентної зовнішньої сили.

Якщо тепер розглядати неоднорідне диференціальне рівняння вигляду

$$a_0 \frac{dx}{dt} + a_1 x = f(t) \qquad (2.3.77)$$

з тими самими ненульовими початковими умовами (2.3.69), то в силу принципу суперпозиції воно буде еквівалентне до рівняння

$$a_0 \frac{dx}{dt} + a_1 x = f(t) + f_{eq}(t), \qquad (2.3.78)$$

але вже з нульовими початковими умовами.

Запропонований підхід можна узагальнити на випадок диференціального рівняння довільного порядку. Такий прийом буває зручним при розв'язанні рівнянь вигляду (2.1.2) методом інтеграла Дюамеля (особливо – інтеграла Дюамеля другого роду).

По суті, ця сама ідея – можливість заміни ненульових початкових умов еквівалентною зовнішньою силою – використовується і в так званому операторному методі, при який буде мова в наступному підрозділі.

Контрольні питання до підрозділу 2.3

1. Опишіть основні етапи аналізу лінійних електричних кіл методом комплексних амплітуд. Коли цей метод застосовний?
2. Що таке коефіцієнт передачі (передавальна функція) лінійного чотириполюсника?
3. Що таке диферецiювальна та інтегрувальна ланка?
4. За якої умови RC-коло може виступати як інтегрувальна ланка? З якого елемента RC-кола зніматиметься проінтегрована напруга?
5. За якої умови RC-коло може виступати як диференціювальна ланка? З якого елемента RC-кола зніматиметься продиференційована напруга?
6. Що таке перехідна функція лінійного чотириполюсника?
7. На вхід лінійного чотириполюсника з відомою перехідною функцією подається сигнал у вигляді імпульсу. Запишіть сигнал на його виході.
8. Що таке імпульсна функція (функція Гріна) лінійного чотириполюсника?
9. На вхід лінійного чотириполюсника з відомою імпульсною функцією подається сигнал у вигляді імпульсу. Запишіть сигнал на його виході.
10. Як пов'язані між собою імпульсна та перехідна функції?
11. Як пов'язані між собою імпульсна та передавальна функції?

Приклади розв'язання задач до підрозділу 2.3

П 2.3.1. На вхід ланцюжка, схема якого подана на рис. 2.3.10, надходить послідовність коротких імпульсів вигляду (2.3.3):

$$f(t) = u_m T \sum_{k=-\infty}^{\infty} \delta(t - kT).$$

(рис. 2.3.1). Записати напругу на виході ланцюжка в ненавантаженому режимі.

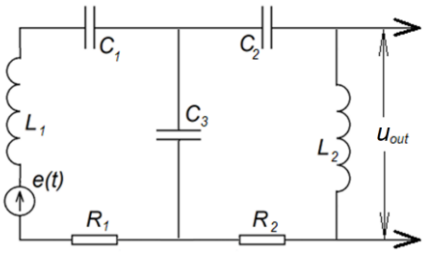


Рис. 2.3.10

Розв'язок

Оскільки вхідний сигнал є періодичним, скористаємося методом інтегралу Фур'є.

Знайдемо спочатку передавальну функцію $K(\omega) = u_{m\,out} / e_m$, прийнявши для цього, що $e(t) = e_m \exp(i\omega t)$.

Нехай у лівому контурі протікає за годинниковою стрілкою струм i_1, у правому в тому самому напрямку – i_2. Тоді рівняння для контурів за другим правилом Кірхгофа мають вигляд:

$$\begin{cases} i_{m1}\left(i\omega L_1 + R_1 + \dfrac{1}{i\omega C_1}\right) + \dfrac{1}{i\omega C_3}(i_{m1} - i_{m2}) = e_m; \\ i_{m2}\left(i\omega L_2 + R_2 + \dfrac{1}{i\omega C_2}\right) + \dfrac{1}{i\omega C_3}(i_{m2} - i_{m1}) = 0. \end{cases}$$

Після перетворень

$$\begin{cases} \left(i\omega L_1 + R_1 + \dfrac{1}{i\omega C_1'}\right) i_{m1} - \dfrac{1}{i\omega C_3} i_{m2} = e_m; \\ \left(i\omega L_2 + R_2 + \dfrac{1}{i\omega C_2'}\right) i_{m2} - \dfrac{1}{i\omega C_3} i_{m1} = 0, \end{cases} \qquad \frac{1}{C_{1,2}'} = \frac{1}{C_{1,2}} + \frac{1}{C_3}.$$

або

$$\begin{cases} \left[i\omega C_3 \left(i\omega L_2 + R_2 + \dfrac{1}{i\omega C_2'}\right)\left(i\omega L_1 + R_1 + \dfrac{1}{i\omega C_1'}\right) - \dfrac{1}{i\omega C_3} \right] i_{m2} = e_m; \\ i_{m1} = i\omega C_3 \left(i\omega L_2 + R_2 + \dfrac{1}{i\omega C_2'}\right) i_{m2}. \end{cases}$$

Таким чином,

$$i_{m2} = \frac{e_m}{i\omega C_3 \left(i\omega L_2 + R_2 + \dfrac{1}{i\omega C_2'}\right)\left(i\omega L_1 + R_1 + \dfrac{1}{i\omega C_1'}\right) - \dfrac{1}{i\omega C_3}} =$$

$$= \frac{-i\omega C_3 e_m}{(\omega C_3)^2 \left(i\omega L_2 + R_2 + \dfrac{1}{i\omega C_2'}\right)\left(i\omega L_1 + R_1 + \dfrac{1}{i\omega C_1'}\right) + 1} =$$

$$=\frac{-i\omega C_3 e_m}{L_1 L_2 \left(C_3\right)^2 \left(i\omega^2+\omega\frac{R_2}{L_2}+\frac{1}{iC_2' L_2}\right)\left(i\omega^2+\omega\frac{R_1}{L_1}+\frac{1}{iC_1' L_1}\right)+1}=$$

$$=\frac{-i\omega C_3 e_m}{L_1 L_2 C_1' C_2' \frac{C_3^2}{C_1' C_2'}\left(i\omega^2+2\omega\delta_2-i\omega_{02}^2\right)\left(i\omega^2+2\omega\delta_1-i\omega_{01}^2\right)+1}=$$

$$=\frac{-i\omega_{01}^2\omega_{02}^2\omega C_3 e_m}{\omega_{01}^2\omega_{02}^2-\kappa^2\left[\left(\omega^2-\omega_{01}^2\right)-i2\omega\delta_1\right]\left[\left(\omega^2-\omega_{02}^2\right)-i2\omega\delta_2\right]},$$

де використано позначення

$$\omega_{01,02}^2=\frac{1}{L_{1,2}C_{1,2}'},\qquad \delta_{1,2}=\frac{R_{1,2}}{2L_{1,2}},\qquad \kappa^2=\frac{C_3^2}{C_1' C_2'}.$$

Передавальна функція:

$$K\left(\omega\right)=\frac{u_{m\,out}}{e_m}=\frac{i\omega L_2 i_{m2}}{e_m}=\frac{\omega_{01}^2\omega_{02}^2\omega^2 C_3 L_2}{\omega_{01}^2\omega_{02}^2-\kappa^2\left[\left(\omega^2-\omega_{01}^2\right)-i2\omega\delta_1\right]\left[\left(\omega^2-\omega_{02}^2\right)-i2\omega\delta_2\right]}.$$

Скориставшись представленням прикладеної напруги (2.3.3) у формі (2.3.5),

$$f\left(t\right)=u_m\sum_{n=-\infty}^{\infty}\exp\left(-i\frac{2\pi n}{T}t\right),$$

можна відразу записати напругу на виході:

$$u_{out}\left(t\right)=\left(\frac{2\pi}{T}\right)^2\omega_{01}^2\omega_{02}^2 C_3 L_2 u_m\times$$

$$\times\sum_{n=-\infty}^{\infty}\frac{n^2\exp\left(-i\frac{2\pi n}{T}t\right)}{\omega_{01}^2\omega_{02}^2-\kappa^2\left\{\left[\left(\frac{2\pi n}{T}\right)^2-\omega_{01}^2\right]-i\frac{4\pi n}{T}\delta_1\right\}\left\{\left[\left(\frac{2\pi n}{T}\right)^2-\omega_{02}^2\right]-i\frac{4\pi n}{T}\delta_2\right\}}.$$

При малих $\delta_{1,2}$ резонанс спостерігається за виконання умов

$$\frac{2\pi n}{T}=\omega_{01}$$

або

$$\frac{2\pi n}{T}=\omega_{02}.$$

П 2.3.2. Знайти форму сигналу на виході RC-ланцюжка (рис. 2.3.9 а) при поданні на його вхід прямокутного імпульсу. Розглянути різні співвідношення між тривалістю імпульсу T й сталою часу ланцюжка $\tau=RC$. Задачу розв'язати методом перехідної функції.

Розв'язок

Перехідна функція ланцюжка має вигляд

$$a\left(t-\theta\right)=\begin{cases}1-\exp\left[-\left(t-\theta\right)/\tau\right], & t\ge\theta\\ 0, & t<\theta,\end{cases}\qquad \tau=RC$$

(див. формулу (2.3.27)).

Вхідний сигнал представимо у формі:

$$u_1\left(t\right)=u_m\left[\Theta\left(t\right)-\Theta\left(t-T\right)\right]=\begin{cases}u_m, & 0\le t\le T;\\ 0, & t<0\cup t>T.\end{cases}$$

Тоді, відповідно до формули (2.3.19),

$$u_2(t) = u_m\left[a(t) - a(t-T)\right] = \begin{cases} 0, \ t < 0; \\ u_m\left[1 - \exp(-t/\tau)\right], 0 \le t \le T; \\ u_m \exp(-t/\tau)\left[\exp(T/\tau) - 1\right] t > T \end{cases}$$

(враховано, що $d\left[\theta(t)\right]/dt = \delta(t)$). Графіки функцій $u_1(t)$ та $u_2(t)$ для різних значень T/τ подані на рис. 2.3.11.

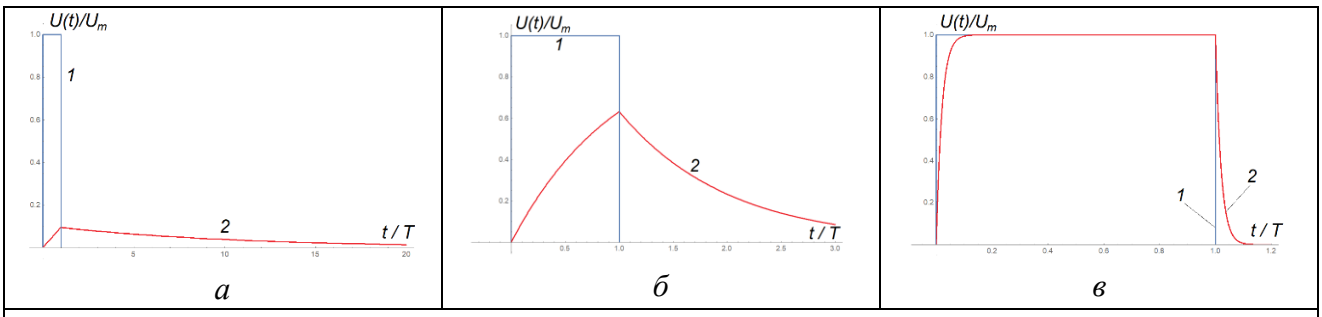


Рис. 2.3.11. Вхідна (1) та вихідна (2) напруги при $T/\tau = 0.1$ (а), $T/\tau = 1$ (б), $T/\tau = 50$ (в)

У випадку $T \ll \tau$ (рис. 2.3.11 а) експоненту малого аргументу можна розкласти в ряд Тейлора, обмежившись малим доданком:

$$u_2(t) = u_m\left[a(t) - a(t-T)\right] = \begin{cases} 0, \ t < 0; \\ u_m t/\tau, \ 0 \le t \le T; \\ u_m T/\tau \exp(-t/\tau), \ t > T \end{cases} \qquad (2.3.78)$$

У протилежному випадку (рис. 2.3.10 в) сигнал на виході відрізняється від сигналу на вході лише дещо розмитими переднім і заднім фронтами.

Задачі до підрозділу 2.3

2.3.1. За якої умови *RL*-коло може служити диференціювальною ланкою? З якого елемента зніматиметься продиференційована напруга?

2.3.2. За якої умови *RL*-коло може служити інтегрувальною ланкою? З якого елемента зніматиметься проінтегрована напруга?

2.3.3. Знайти перехідну функцію послідовного коливного контуру, вважаючи вихідним сигналом струм контуру.

2.3.4. Знайти перехідну функцію *RC*-ланцюжка, якщо вихідна напруга знімається з резистора.

2.3.5. Знайти форму сигналу на виході *RC*-ланцюжка при поданні на його вхід прямокутного імпульсу, якщо вихідна напруга знімається з резистора. Розглянути різні співвідношення між тривалістю імпульсу T й сталою часу ланцюжка $\tau = RC$. Задачу розв'язати методом перехідної функції.

2.4. Операторний метод аналізу лінійних кіл за довільних початкових умов

Спектральний метод аналізу поведінки лінійних електричних кіл під дією зовнішнього впливу (див. п. 2.3.1) не є універсальним. Це пов'язано, зокрема, з тим, що інтегральне перетворення Фур'є застосовне лише для сигналів $f(t)$ з обмеженою енергією, тобто таких, що задовольняють умові (1.3.1):

$$\int_{-\infty}^{\infty} \left| f^2(t) \right| dt < \infty .$$

У багатьох випадках функції, які являють інтерес, не задовольняють цій умові. Прикладом такого роду, який ми вже розглядали, є функція Хевісайда.

Крім того, аналіз за допомогою перетворення Фур'є орієнтований на нульові початкові умови, що також не завжди відповідає дійсності. Звичайно, до розв'язків такого вигляду завжди можна дописати частину, що відповідає вільним коливанням, і накласти відповідні початкові умови (див. вище п. 2.3.5), але це робить розв'язок більш громіздким.

Вказаних вад позбавлений так званий операторний метод розв'язання лінійних диференціальних рівнянь, заснований на інтегральному перетворенні Лапласа. Він є альтернативним щодо розглянутих вище методів перехідної та імпульсної функцій (пп. 2.3.2-2.3.3) з урахуванням ненульових початкових умов.

2.4.1. Інтегральне перетворення Лапласа

Припустимо, що функція $f(t)$ не задовольняє умові (1.3.1). Розглядатимемо замість неї функцію $f_1(t) = f(t)\theta(t)\exp(-\sigma t)$, $\sigma > 0$. На практиці майже для кожної функції $f(t)$ можна підібрати таке значення σ, що для функції $f_1(t)$ умова вигляду (1.3.1) задовольняється, і її можна представити як інтеграл Фур'є зі спектром

$$\int_{-\infty}^{\infty} f_1(t)\exp(-i\omega t)dt = \int_{0}^{\infty} f(t)\exp\left[-(\sigma + i\omega)t\right]dt \equiv F(\sigma + i\omega). \qquad (2.4.1)$$

Зворотне Фур'є-перетворення для функції $f_1(t)$ має вигляд:

$$f_1(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\sigma + i\omega)\exp(i\omega t)d\omega . \qquad (2.4.2)$$

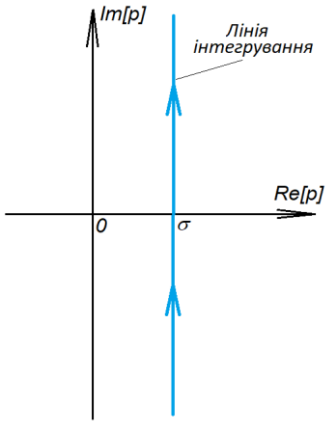


Рис. 2.4.1. Лінія інтегрування на площині комплексної змінної p в інтегралі Лапласа (2.4.4)

Введемо нову змінну $p = \sigma + i\omega$ $(\sigma = const)$, $d\omega = dp/i$. Тоді співвідношення (2.4.1)-(2.4.2) можна переписати у формі:

$$F(p) = \int_{0}^{\infty} f(t)\exp(-pt)dt = \int_{-\infty}^{\infty} \theta(t) f(t)\exp(-pt)dt ; \qquad (2.4.3)$$

$$f(t)=\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty}F(p)\exp(pt)dp. \qquad (2.4.4)$$

Останній інтеграл можна інтерпретувати як інтеграл уздовж прямої на комплексній площині (рис. 2.4.1).

Співвідношення (2.4.3)-(2.4.4) відомі інтегральні перетворення Лапласа. Називатимемо інтеграл (2.4.3) прямим перетворенням Лапласа, а (2.4.4) – зворотним перетворенням. Функцію $f(t)$ називають у цьому випадку оригіналом, функцію $F(p)$ – зображенням. Символічне позначення перетворення Лапласа:

$f(t)\to F(p)$,

або

$F(p)=L\{f(t)\}, \qquad f(t)=L^{-1}\{F(p)\}$.

З точки зору математики перетворення Лапласа можна розглядати як деяку модифікацію інтегрального перетворення Фур'є. Відмінність полягає в тому, що:

- у прямому перетворенні під інтегралом з'являється функція Хевісайда (або, що те саме, інтегрування за часом відбувається лише вздовж додатної півосі);
- замість дійсної частоти ω фігурує комплексна величина $p=\sigma+i\omega$.

В результаті пряме перетворення Лапласа можна уподібнити до віконного перетворення Фур'є зі стаціонарним асиметричним вікном вигляду $\exp(-\sigma t)\cdot\theta(t)$. Як уже вказувалося (див. вище п. пп. 1.5.1-1.5.3), віконне перетворення Фур'є застосовне й до функцій, що не задовольняють умові (1.3.1). Тому це справедливо й для перетворення Лапласа. Але через комплексність змінної $p=\sigma+i\omega$ і наявність функції Хевісайда під інтегралом зображення функції суттєво відрізняється від її спектру, який ми обговорювали раніше (див. пп. 1.2-1.3, 1.5).

2.4.2. Властивості перетворення Лапласа

Перетворення Лапласа – це лінійна операція. Звідси випливає таке:

1. Якщо $f(t)\to F(p)$, то

$$Cf(t)\to CF(p), \qquad (2.4.5)$$

де C – довільна стала.

2. Якщо $f_{1,2}(t)\to F_{1,2}(p)$, то

$$f_2(t)+f_1(t)\to F_1(p)+F_2(p). \qquad (2.4.6)$$

3. Знайдемо зображення похідної $\varphi(t)=df(t)/dt$, виконавши інтегрування частинами:

$$\varphi(t)\to\int_0^\infty\frac{df(t)}{dt}\exp(-pt)dt=\int_0^\infty df(t)\exp(-pt)=$$
$$=f(t)\exp(-pt)\Big|_0^\infty-\int_0^\infty f(t)d\left[\exp(-pt)\right]=-f(0)+p\int_0^\infty f(t)\exp(-pt)dt. \qquad (2.4.7)$$

Отже,

$$\frac{df(t)}{dt}\to pF(p)-f(0), \qquad (2.4.8)$$

де $f(t)\to F(p)$.

Для другої похідної

$$\frac{d^2 f(t)}{dt^2} \to p^2 F(p) - pf(0) - \frac{df(0)}{dt}, \tag{2.4.9}$$

для похідної n-го порядку

$$\frac{d^n f(t)}{dt^n} \to p^n F(p) - p^{n-1} f(0) - \ldots - \frac{d^{n-1} f(0)}{dt^{n-1}}. \tag{2.4.10}$$

Як бачимо, формула (2.4.10) містить n початкових умов, які повністю визначають розв'язок диференціального рівняння n-го порядку. У правій частині формули (2.4.10) перший доданок відповідає зовнішній силі, а інші – ненульовим початковим умовам. Таким чином, цю формулу в певному сенсі можна розглядати як заміну ненульових початкових умов деякою еквівалентною зовнішньою силою (див. вище п. 1.3.6).

Для нульових початкових умов формули (2.4.8)-(2.4.10) максимально спрощуються, зокрема,

$$\frac{d^n f(t)}{dt^n} \to p^n F(p), \quad n = 1, 2, 3, \ldots \tag{2.4.11}$$

4. Знайдемо тепер зображення інтегралу, $\varphi(t) = \int f(t)\,dt$. Очевидно, при цьому справедлива рівність $f(t) = d\varphi(t)/dt$. Знову інтегруючи частинами, отримаємо:

$$\begin{aligned} F(p) &= \int_0^\infty f(t)\exp(-pt)\,dt = \int_0^\infty \frac{d\varphi}{dt}\exp(-pt)\,dt = \int_0^\infty d\varphi \exp(-pt) = \\ &= \varphi\exp(-pt)\Big|_0^\infty - \int_0^\infty \varphi\, d\left[\exp(-pt)\right] = -\varphi(0) + p\int_0^\infty \varphi\exp(-pt)\,dp. \end{aligned} \tag{2.4.12}$$

Оскільки $\int_0^\infty \varphi\exp(-pt)\,dp$ – це і є шукане зображення інтегралу, то з (2.4.12) маємо:

$$\varphi(t) = \int f(t)\,dt \to \frac{F(p) + \varphi(0)}{p}. \tag{2.4.13}$$

За потреби формула (2.4.13) може бути узагальнена на випадок кратних інтегралів.

5. Зсув початку відліку за часом. Нехай $f(t) \to F(p)$ Знайдемо зображення функції $\varphi(t) \equiv f(t - t_0)$.

$$\begin{aligned} &\int_0^\infty \varphi(t)\exp(-pt)\,dt = \int_0^\infty f(t - t_0)\exp(-pt)\,dt = \left\| \begin{matrix} t - t_0 = \tau \\ t = \tau + t_0 \end{matrix} \right\| = \\ &= \int_0^\infty f(\tau)\exp\left[-p(\tau + t_0)\right]d\tau = \exp(-pt_0)\int_0^\infty f(\tau)\exp(-p\tau)\,d\tau = \\ &= F(p)\exp(-pt_0). \end{aligned} \tag{2.4.14}$$

2.4.3. Операторний метод розв'язування лінійних диференціальних рівнянь

Інтегральні перетворення Лапласа лежать в основі так званого операторного методу розв'язання лінійних диференціальних рівнянь вигляду (2.1.2)

$$a_0 \frac{d^n x}{dt^n} + a_1 \frac{d^{n-1} x}{dt^{n-1}} + \ldots + a_{n-1}\frac{dx}{dt} + a_n x = f(t).$$

Цей метод застосовний і до деяких інтегро-диференціальних рівнянь.

Перш ніж описувати операторний метод, зазначимо, що для поширеніших функцій перетворення Лапласа давно виконані, і їхні зображення добре відомі. Вони представлені в

таблиці 1.3.1.

Таблиця 1.3.1. Оригінали та зображення поширених функцій

Оригінал $f(t)$	Зображення $F(p)$
$U_m \cdot \theta(t)$	$U_m \frac{1}{p}$
$U_m \cdot \theta(t-t_0)$	$U_m \frac{1}{p}\exp(-pt_0)$
$\exp(-\alpha t)\cdot\theta(t)$	$\frac{1}{p+\alpha}$
$\sin\omega_0 t\cdot\theta(t)$	$\frac{\omega_0}{p^2+\omega_0^2}$
$\cos\omega_0 t\cdot\theta(t)$	$\frac{p}{p^2+\omega_0^2}$
$\sin(\omega_0 t)\exp(-\alpha t)\cdot\theta(t)$	$\frac{\omega_0}{(p+\alpha)^2+\omega_0^2}$
$\cos(\omega_0 t)\exp(-\alpha t)\cdot\theta(t)$	$\frac{p}{(p+\alpha)^2+\omega_0^2}$
$\sin(\omega_0 t+\psi)\cdot\theta(t)$	$\frac{p\sin\psi+\omega_0\cos\psi}{p^2+\omega_0^2}$
$\cos(\omega_0 t+\psi)\cdot\theta(t)$	$\frac{p\cos\psi-\omega_0\sin\psi}{p^2+\omega_0^2}$
$t\cdot\theta(t)$	$\frac{1}{p^2}$
$t^n\cdot\theta(t),\quad n=1,2,3,...$	$\frac{n!}{p^{n+1}}$

Ідея операторного методу полягає в переході від неоднорідних лінійних диференціальних (або інтегро-диференціальних) рівнянь та лінійних диференціальних рівнянь із ненульовими початковими умовами для оригіналів до алгебраїчних рівнянь для зображень. Це робиться шляхом застосування до вихідних рівнянь типу (2.1.2) перетворення Лапласа з використанням формули (2.4.10). Для інтегро-диференціальних рівнянь використовується також формула (2.4.13) та її узагальнення.

Отримані алгебраїчні рівняння розв'язуються щодо зображення невідомої функції $x(t)$, після чого виконується зворотний перехід від зображення до оригіналу. При цьому автоматично враховуються початкові умови.

2.4.4. Теорема розкладення Хевісайда

При розв'язанні лінійних диференціальних (інтегро-диференціальних) рівнянь вигляду (2.1.2)

$$a_0\frac{d^n x}{dt^n}+a_1\frac{d^{n-1}x}{dt^{n-1}}+...+a_{n-1}\frac{dx}{dt}+a_n x=f(t),$$

з ненульовими початковими умовами або ненульовою правою частиною можна отримати вираз для зображення $X(t)$ шуканої функції $x(t)$ у формі

$$X(p)=\frac{X_m(p)}{X_n(p)}, \tag{2.4.15}$$

де $X_n(p) = a_0 p^n + a_1 p^{n-1} + ... + a_{n-1} p + a_n$, $m < n$. Вигляд поліному $X_m(p)$ визначається зовнішньою силою та початковими умовами (див. нижче п. 2.4.5).

Нехай поліном $X_n(p)$ має n відмінних між собою коренів (втім, деякі корені можуть повторюватися): $p_1, p_2, ... p_l, ..., p_n$. Тоді (2.4.15) можна подати у формі:

$$X(p) = \sum_{k=1}^{n} \frac{B_k}{p - p_k}. \tag{2.4.16}$$

Щоб знайти коефіцієнт B_l[4], домножимо обидві сторони рівності (2.4.16) на $p - p_l$:

$$(p - p_l) X(p) = (p - p_l) \sum_{k=1}^{n} \frac{B_k}{p - p_k}. \tag{2.4.17}$$

При $x \to x_l$ співвідношення (2.4.17) з урахуванням (2.4.15) набуває вигляду:

$$B_l = \lim_{p \to p_l} (p - p_l) \frac{X_m(p)}{X_n(p)} = X_m(p_l) \lim_{p \to p_l} \frac{p - p_l}{X_n(p) - X_n(p_l)} = \frac{X_m(p_l)}{X'_n(p_l)}. \tag{2.4.18}$$

(враховано, що $X_n(p_l) = 0$). Таким чином,

$$X(p) = \sum_{k=1}^{n} \frac{X_m(p_k)}{(p - p_k) X'_n(p_k)}. \tag{2.4.19}$$

Оскільки $\exp(p_i t) \to 1/(p - p_i)$, то в силу властивостей перетворення Лапласа, оригіналом зображення (2.4.19), тобто розв'язком рівняння (2.1.2), буде функція

$$x(t) = \sum_{k=1}^{n} \frac{X_m(p_k)}{X'_n(p_k)} \exp(p_k t). \tag{2.4.20}$$

Зазначимо тут, що експоненти $\exp(p_k t)$ – це набір власних розв'язків рівняння (2.1.2). Якщо воно описує дисипативну систему, то всі ці розв'язки з часом згасають.

Якщо один з n коренів поліному $X_n(p)$ дорівнює нулеві, формула (2.4.20) з урахуванням того, що $\theta(t) \to 1/p$, набуває вигляду

$$x(t) = \frac{X_m(0)}{\Phi(0)} \theta(t) + \sum_{k=1}^{n-1} \frac{X_m(p_k)}{X'_n(p_k)} \exp(p_k t), \tag{2.4.21}$$

де $\Phi(0) \equiv \Phi(p = 0)$, $\Phi(p) = X_n(p)/p$ – поліном, що не має нульових коренів. Перший доданок відповідає стаціонарному стану, який з часом встановлюється в системі, другий – аналогу перехідного процесу, зумовленого ненульовими початковими умовами. Формула (2.4.21) може бути узагальнена на випадок, коли кратність нульового кореню перевищує одиницю.

2.4.5. Аналіз перехідних процесів операторним методом

Операторний метод дозволяє легко розв'язувати лінійні неоднорідні диференціальні рівняння з довільними початковими умовами. Проілюструємо його застосування на прикладі дослідження перехідних процесів у коливному контурі (рис. 2.2.1), який описується інтегро-диференціальним рівнянням (2.2.1)

$$L \frac{di}{dt} + Ri + \frac{1}{C} \int i dt = e(t).$$

Вважатимемо, що на вхід кола в момент часу $t = 0$ надходить стрибок напруги величиною E_0, тобто $e(t) = E_0 \theta(t)$. Значення струму $i(t)$ та заряду $q(t) = \int i dt$ при $t = 0$ – від-

[4] У формулі (2.4.16) k – індекс підсумовування, l – деяке конкретне значення цього індексу.

повідно $i(0)$ та $q(0)$.

Нехай $i(t) \to I(p)$, $e(t) \to E(p)$. Тоді з урахуванням (2.4.8) та (2.4.13)

$$L\frac{di}{dt} \to L\left[pI(p) - i(0)\right], \tag{2.4.22}$$

$$\frac{1}{C}\int idt \to \frac{1}{C}\left[\frac{I(p)}{p} + \frac{q(0)}{p}\right], \qquad q(t) = \int idt\,; \tag{2.4.23}$$

$$e(t) = E_0\theta(t) \to E_0/p\,. \tag{2.4.24}$$

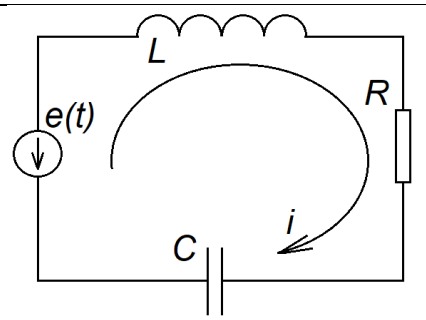


Рис. 2.2.1. Послідовний коливний контур, до якого прикладена зовнішня напруга

Таким чином, рівняння для зображення $I(p)$ шуканої функції $i(t)$ має вигляд

$$L\left[pI(p) - i(0)\right] + RI(p) + \frac{1}{C}\left[\frac{I(p)}{p} + \frac{q(0)}{p}\right] = \frac{E_0}{p}, \tag{2.4.25}$$

або

$$I(p)\left(pL + R + \frac{1}{pC}\right) = \frac{E_0}{p} + Li(0) - \frac{1}{pC}q(0), \tag{2.4.26}$$

звідки

$$I(p) = \frac{\dfrac{E_0}{p} + Li(0) - \dfrac{1}{pC}q(0)}{pL + R + \dfrac{1}{pC}} = \frac{C\left[E_0 + pLi(0)\right] - q(0)}{pC(pL+R)+1} =$$

$$= \frac{CL\left\{\left[E_0/L + pi(0)\right] - q(0)/LC\right\}}{CL(p-p_1)(p-p_2)} = \frac{\left[E_0/L + pi(0)\right] - q(0)/CL}{(p-p_1)(p-p_2)}, \tag{2.4.27}$$

$$p_{1,2} = -\delta \pm i\omega, \qquad \omega = \sqrt{\omega_0^2 - \delta^2}, \qquad \delta = \frac{R}{2L}, \qquad \omega_0^2 = \frac{1}{LC} \tag{2.4.28}$$

(вважаємо, що $\omega_0 > \delta$). Тут в ролі поліному $X_n(p)$ виступає вираз $p(p + R/L) + 1/CL$, аналогом поліному $X_m(p)$ є вираз $\left[E_0/L + pi(0)\right] - q(0)/CL$, вигляд якого залежить від зовнішньої сили та початкових умов (див. вище формулу (2.4.15)).

Накладемо на систему нульові початкові умови:

$$i(0) = 0, \quad q(0) = 0\,. \tag{2.4.29}$$

Тоді формула (2.4.27) помітно спрощується:

$$I(p) = \frac{E_0}{L}\frac{1}{(p-p_1)(p-p_2)}\,. \tag{2.4.30}$$

В такому разі, відповідно до формули (2.4.20),

$$i(t)=\frac{E_0}{L}\left[\frac{1}{p_1-p_2}\exp(i\omega t)+\frac{1}{p_2-p_1}\exp(-i\omega t)\right]\exp(-\delta t)=$$
$$=\frac{E_0\exp(-\delta t)}{2i\omega L}\left[\exp(i\omega t)-\exp(-i\omega t)\right]=\frac{E_0}{\omega L}\sin(\omega t)\exp(-\delta t) \qquad (2.4.31)$$

(рис. 2.4.2). Враховано, що

$$\frac{d}{dp}(p-p_1)(p-p_2)\Big|_{p=p_{1,2}}=\left[(p-p_2)+(p-p_1)\right]\Big|_{p=p_{1,2}}=\pm(p_1-p_2)). \qquad (2.4.32)$$

Інтегро-диферрнціальне рівняння (2.2.1), як уже відзначалося, еквівалентне диференціальному рівнянню (2.3.36)

$$\frac{d^2i}{dt^2}+2\delta\frac{di}{dt}+\omega_0^2 i=\frac{1}{L}\frac{de(t)}{dt}, \qquad \delta=r/2L,\ \omega_0^2=1/LC.$$

Тому при $E_0=1$ знайдена нами перехідна функція (2.4.31) для рівняння (2.2.1) збігається з імпульсною функцією (2.3.45 а) для рівняння (2.3.36).

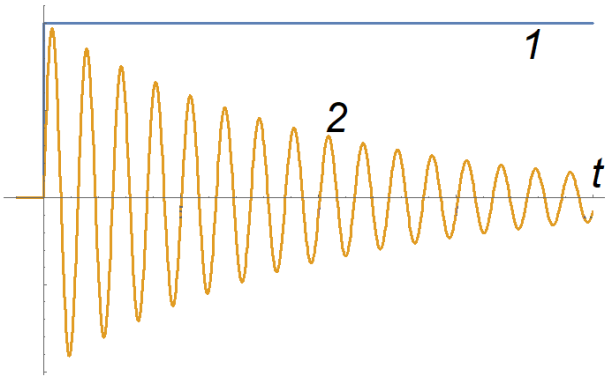


Рис. 2.4.2. Перехідний процес у послідовному коливному контурі:
1 – вхідна напруга, 2 – струм

Контрольні питання до підрозділу 2.4

1. Чим перетворення Лапласа відрізняється від перетворення Фур'є?
2. У чому зручність розв'язання диференціальних рівнянь операторним методом?
3. Чому зворотне перетворення Лапласа звичайно не виконують при розв'язанні кожного нового диференціального рівняння?

Приклади розв'язання задач до підрозділу 2.4

П 2.4.1. Знайти зображення функції Хевісайда $\Theta(t-t_0)$.

Розв'язок

$$F(p)=\int_0^\infty f(t)\exp(-pt)dt=\int_0^\infty \theta(t-t_0)\exp(-pt)dt=\left\|\begin{matrix}t-t_0=\tau\\ t=\tau+t_0\end{matrix}\right\|=\int_0^\infty \theta(\tau)\exp\left[-p(\tau+t_0)\right]d\tau=$$
$$=\exp(-pt_0)\int_0^\infty \exp(-p\tau)d\tau=\exp(-pt_0)\frac{\exp(-p\tau)}{-p}\Bigg|_0^\infty=\frac{\exp(-pt_0)}{p}.$$

П 2.4.2. Знайти зображення функції $\Theta(t)\cos\omega t$.

Розв'язок

$$F(p)=\int_0^\infty f(t)\exp(-pt)dt=\int_0^\infty \cos\omega t\exp(-pt)dt=$$

$$=\frac{1}{2}\left\{\int_0^\infty \exp\left[(i\omega-p)t\right]dt+\int_0^\infty \exp\left[(-i\omega-p)t\right]dt\right\}=\frac{1}{2}\left\{\frac{\exp\left[(i\omega-p)t\right]}{i\omega-p}+\frac{\exp\left[(-i\omega-p)t\right]}{-i\omega-p}\right\}\Bigg|_0^\infty=$$

$$=\frac{1}{2}\left\{\frac{1}{p-i\omega}+\frac{1}{p+i\omega}\right\}=\frac{1}{2}\frac{p+i\omega+p-i\omega}{p^2+\omega^2}=\frac{p}{p^2+\omega^2}.$$

П 2.4.3. Знайти зображення функції $\Theta(t)\cos\omega t\exp(-\alpha t)$.

Розв'язок

$$F(p)=\int_0^\infty f(t)\exp(-pt)dt=\int_0^\infty \cos\omega t\exp\left[-(p+\alpha)t\right]dt=\frac{p+\alpha}{(p+\alpha)^2+\omega^2}.$$

П 2.4.4. Знайти зображення функції $\Theta(t)\cos(\omega t+\psi)$.

Розв'язок

$$F(p)=\int_0^\infty f(t)\exp(-pt)dt=\int_0^\infty \cos(\omega t+\psi)\exp(-pt)dt=$$

$$=\frac{1}{2}\left\{\int_0^\infty \exp\left[(i\omega-p)t+i\psi\right]dt+\int_0^\infty \exp\left[(-i\omega-p)t-i\psi\right]dt\right\}=$$

$$=\frac{1}{2}\left\{\exp(i\psi)\frac{\exp\left[(i\omega-p)t\right]}{i\omega-p}\Bigg|_0^\infty+\exp(-i\psi)\frac{\exp\left[(-i\omega-p)t\right]}{-i\omega-p}\Bigg|_0^\infty\right\}=$$

$$=\frac{1}{2}\left\{\frac{\exp(i\psi)}{p-i\omega}+\frac{\exp(-i\psi)}{p+i\omega}\right\}=\frac{1}{2}\frac{(p+i\omega)\exp(i\psi)+(p-i\omega)\exp(-i\psi)}{p^2+\omega^2}=$$

$$=\frac{1}{2}\frac{p\left[\exp(i\psi)+\exp(-i\psi)\right]+i\omega\left[\exp(i\psi)-\exp(-i\psi)\right]}{p^2+\omega^2}=\frac{p\cos\psi-\omega\sin\psi}{p^2+\omega^2}.$$

П 2.4.5. Знайти зображення функції $t\Theta(t)$.

Розв'язок

$$F(p)=\int_{-0}^\infty f(t)\exp(-pt)dt=\int_{-0}^\infty t\exp(-pt)dt=$$

$$=\left\|\int_a^b udv=uv\Big|_a^b-\int_a^b vdu\right\|=\left\|\begin{matrix} t=u & du=dt \\ \exp(-pt)dt=dv & v=-\dfrac{\exp(-pt)}{p}\end{matrix}\right\|=$$

$$=-\frac{t\exp(-pt)}{p}\Bigg|_0^\infty+\frac{1}{p}\int_{-0}^\infty \exp(-pt)dt=-\frac{\exp(-pt)}{p^2}\Bigg|_0^\infty=\frac{1}{p^2}$$

П 2.4.6. Проаналізувати перехідні процеси в RC-ланцюжку, користуючись операторним методом (рис. 2.4.3).

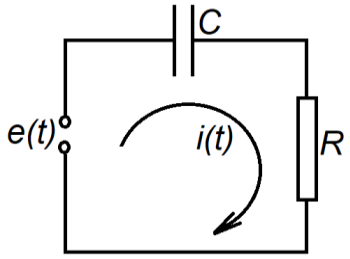


Рис. 2.4.3

Розв'язок

Рівняння за II правилом Кірхгофа записується у формі:

$$Ri(t)+\frac{1}{C}\int idt=e(t).$$

Як і раніше, вважатимемо, що $i(t)\to I(p)$, $e(t)\to E(p)$, $e(t)=E_0\theta(t)$.

Початкові умови вважаємо нульовими. Тоді рівняння для зображень набуває вигляду:

$$RI(p)+\frac{I(p)}{pC}=\frac{E}{p},)$$

звідки

$$I(p)=\frac{E}{p(R+1/pC)}=\frac{E}{R}\frac{1}{p+1/\tau},\quad \tau=RC.$$

Отже,

$$i(t)=\frac{E_0}{R}\Theta(t)\exp\left(-\frac{t}{\tau}\right).$$

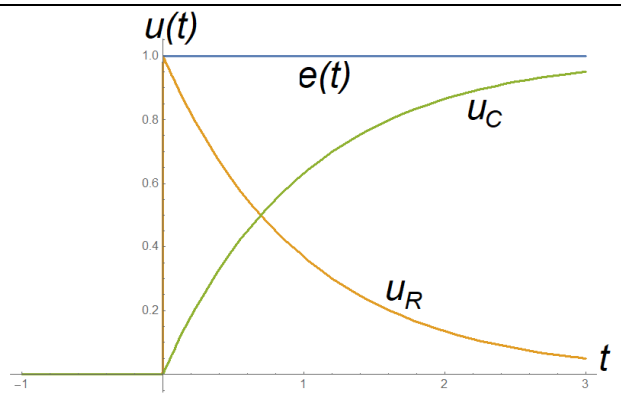


Рис. 2.4.3. Перехідні процеси в RC-ланцюжку

Далі знаходимо інші величини, що характеризують RC-ланцюжок:

$$u_R(t)=i(t)R=E_0\Theta(t)\exp\left(-\frac{t}{\tau}\right);$$

$$q(t)=\int_0^t i(t)dt=\frac{E_0}{R}\int_0^t\exp\left(-\frac{t}{\tau}\right)dt=-\tau\frac{E_0}{R}\exp\left(-\frac{t}{\tau}\right)\Bigg|_0^t=CE_0\left[1-\exp\left(-\frac{t}{\tau}\right)\right];$$

$$u_C(t)=\frac{q(t)}{C}=E_0\left[1-\exp\left(-\frac{t}{\tau}\right)\right]$$

(рис. 2.4.3).

Задачі до підрозділу 2.4

2.4.1. Знайти зображення функції $\exp(-\alpha t)\Theta(t)$.

2.4.2. Знайти зображення функції $\Theta(t)\sin\omega t$.

2.4.3. Знайти зображення функції $\Theta(t)\sin\omega t\exp(-\alpha t)$.

2.4.4. Знайти зображення функції $\Theta(t)\sin(\omega t+\psi)$.

2.4.5. Знайти зображення функції $t^2\Theta(t)$.

2.4.6. Користуючись отриманим результатом дослідження перехідного процесів у послідовному коливному контурі, записати формули та побудувати графіки часових залежностей напруг на активному опорі, на індуктивності та на ємності.

2.4.7. Проаналізувати перехідні процеси в RL-ланцюжку, користуючись операторним методом.

2.4.8. За яких початкових умов в RL-ланцюжку при поданні на його вхід стрибка напруги не виникне перехідний процес?

2.4.9. За яких початкових умов в RC-ланцюжку при поданні на його вхід стрибка напруги не виникне перехідний процес?

2.4.10. За яких початкових умов у послідовному коливному контурі при поданні на його вхід стрибка напруги не виникне перехідний процес?

2.4.11. Знайти сигнал на виході RC-ланцюжка при поданні на його вхід прямокутного імпульсу. Вихідна напруга знімається з конденсатора. Задачу розв’язати операторним методом. Порівняти результат із розв’язком задачі П 2.2.2.

2.4.12. Знайти сигнал на виході RC-ланцюжка при поданні на його вхід прямокутного імпульсу. Вихідна напруга знімається з резистора. Задачу розв’язати операторним методом.

Частина II. Елементи радіоелектроніки

У цій частині розглядаються основні схеми аналогової радіоелектроніки.

Розділ III присвячений напівпровідниковим приладам, які використовує радіоелектроніка – різним типам діодів, польовим та біполярним транзисторам. У розділі IV оглянуто схеми, які можна вважати лінійними – пасивні фільтри та малосигнальні підсилювачі різних типів. Нарешті, в розділі V описано схеми, де нелінійність є принциповою – автогенератори квазігармонічних та імпульсних коливань, а також помножувачі та перетворювачі частоти, модулятори, детектори.

Розділ 3. Прилади напівпровідникової електроніки

У цьому розділі ми опишемо принцип роботи найбільш поширених напівпровідникових приладів. Але спершу оглянемо історичні етапи розвитку елементної бази радіоелектроніки за сто років її існування.

3.1. Етапи розвитку елементної бази радіоелектроніки

У розвитку елементної бази радіоелектроніки можна виділити три основних етапи. Перший етап (приблизно 1920-ті – 1950-ті рр.) пов'язаний з використанням електронних ламп, другий (умовно 1960-ті – 1970-ті рр.) – з використанням дискретних напівпровідникових приладів, третій (умовно з 1980-х рр. дотепер) – з використанням інтегральних мікросхем.

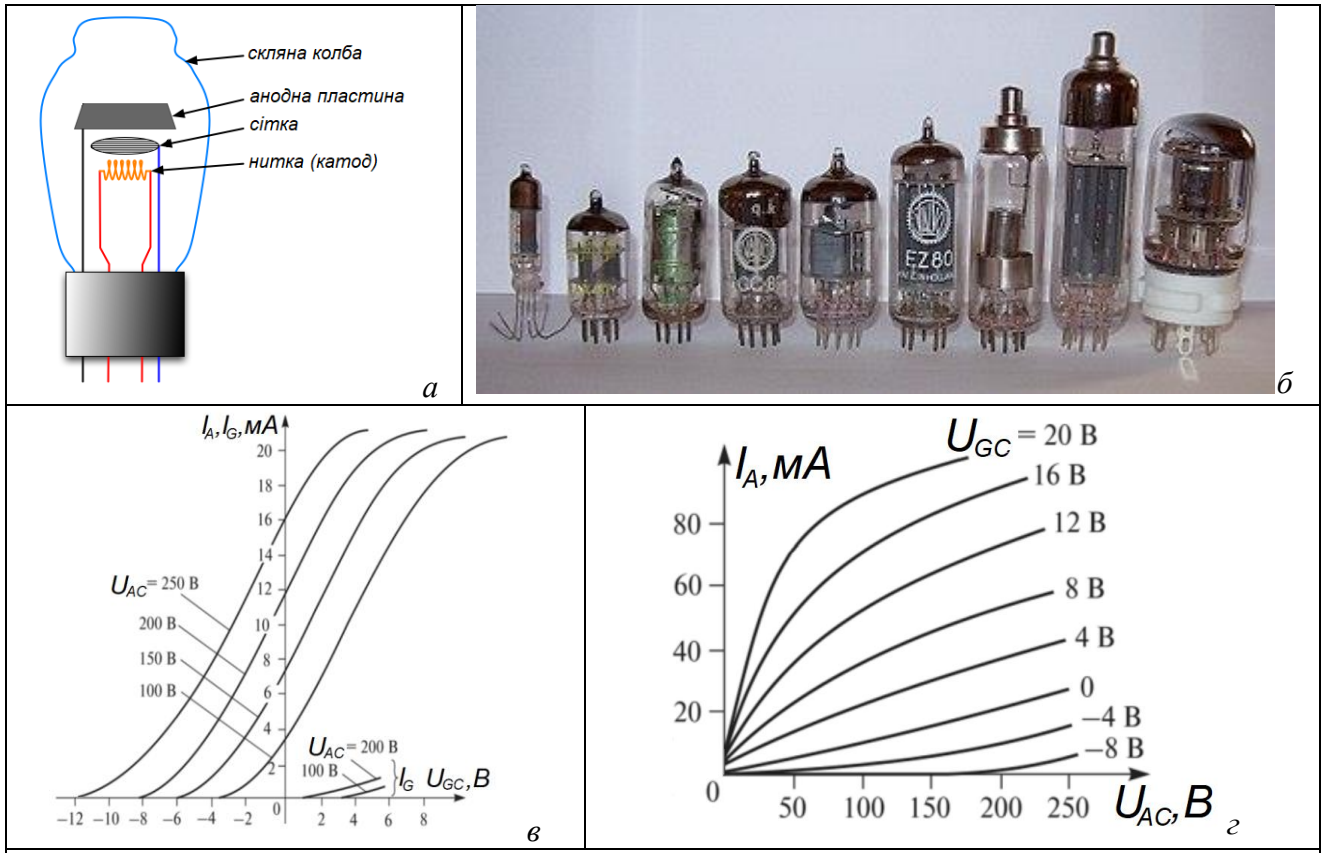


Рис. 3.1.1. Електронні лампи: а – структурна схема тріода; б – фото сучасних ламп, переважно мініатюрних; в-г – вхідні, прохідні (в) та вихідні (г) характеристики тріода, увімкненого зі спільним катодом (А – анод, С – катод, G – сітка, від англ. greed)

Перша електронна лампа – вакуумний діод – була винайдена в 1883 р. Томасом Едісоном (США). Електровакуумний тріод – лампа, на основі якої були створені підсилювачі та генератори – була винайдена в 1906 р. Лі де Форестом (США). До початку 1920-х рр. були

розроблені лампові генератори та підсилювачі. Зараз електронні лампи використовуються в пристроях, призначених для роботи при високих температурах та високому рівні радіації (атомні електростанції, космічні апарати тощо) – рис. 3.1.1.

Принцип роботи напівпровідникового діода запропонував у 1874 р. (тобто ще до винайдення електровакуумних діодів) Карл Фердинанд Браун (Німеччина). Піонерські експериментальні дослідження *p-n* переходу виконав у Києві в 1941 р. Вадим Лашкарьов. Перший напівпровідниковий тріод, або транзистор, запропонували в 1947 р. Джон Бардін і Волтер Браттейн, що працювали під керівництвом Вільяма Шоклі (США, Бел Лабораторіз). Вже в 1956 р. ця робота була відзначена Нобелівською премією. Перші польові транзистори з'явилися в 1960 р., хоча принцип їхньої роботи запропонував ще в 1926-1928 рр. Юліус Лілієнфельд (Німеччина, США) – до речі, уродженець Львова. Напівпровідникові прилади були значно меншими за електровакуумні та споживали помітно менше енергії. Саме на основі напівпровідникових елементів були розроблені, зокрема, кольорові телевізори.

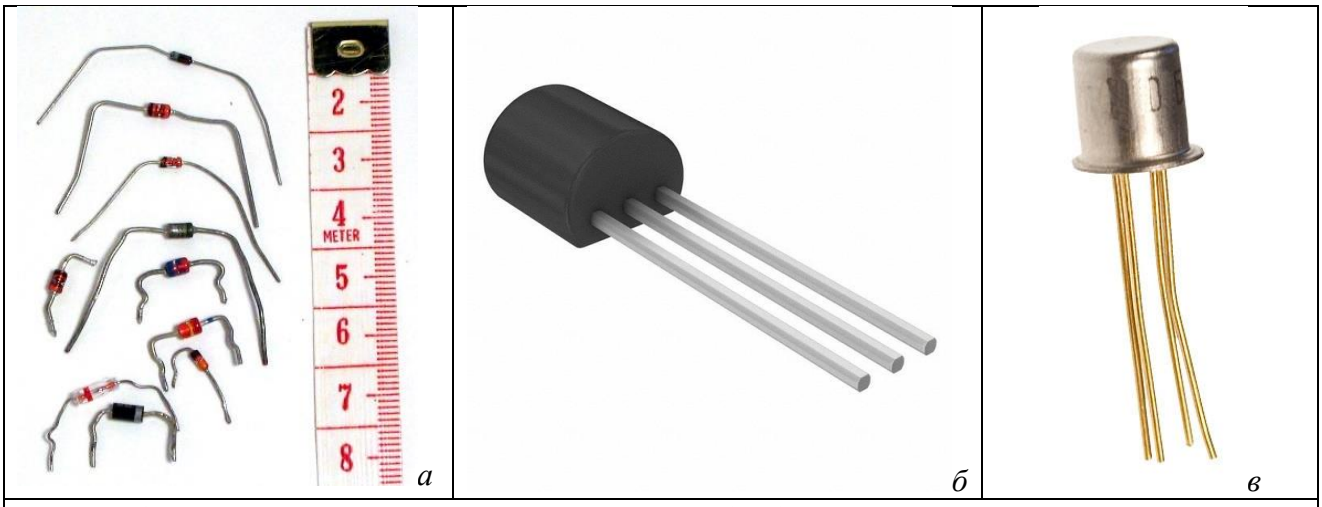


Рис. 3.1.2. Напівпровідникові прилади: а – діоди; б – біполярний транзистор; в – польовий транзистор

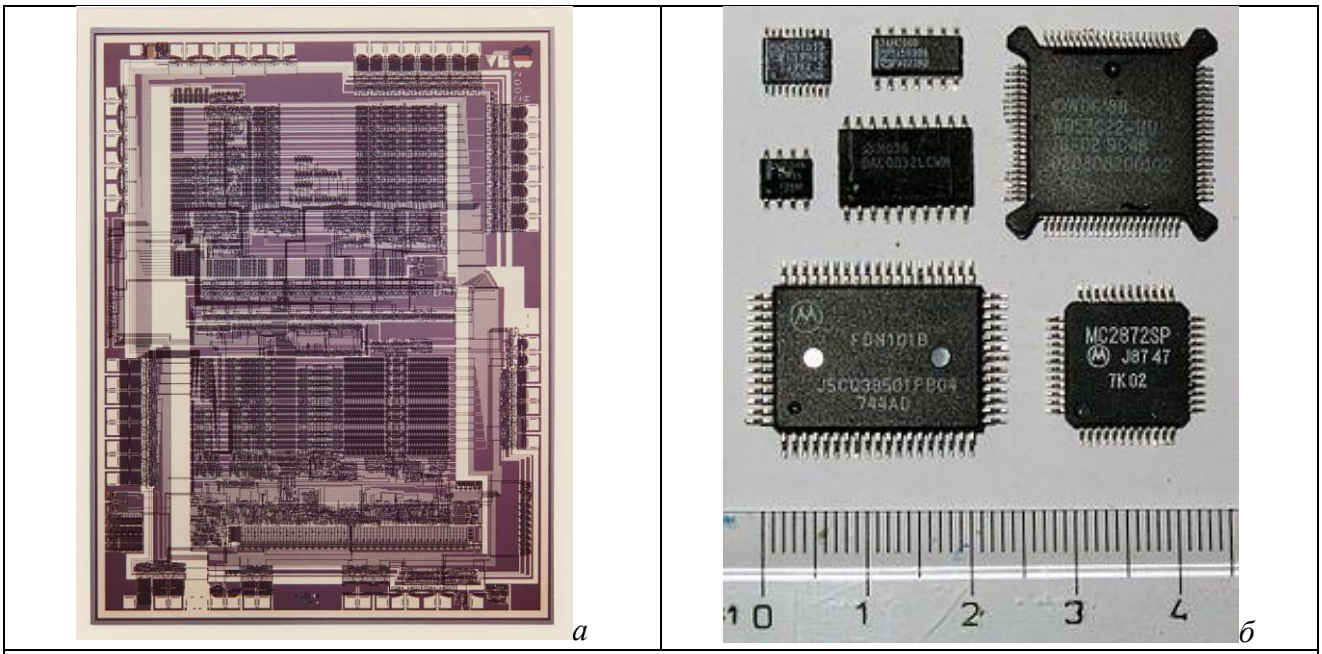


Рис. 3.1.3. Інтегральні мікросхеми: а – вигляд окремої мікросхеми; б – мікросхеми в корпусах

Ускладнення принципових схем радіоелектронних пристроїв привело до відповідного ускладнення технології їхнього виробництва та відповідного зменшення надійності (через велику кількість контактів між елементами). Але вже в 1958 р., всього через 11 років після

винайдення біполярного транзистора, було запропоновано принцип створення інтегральних мікросхем (Джек Кілбі та Роберт Нойс, США). Ідея інтегральних мікросхем полягає в тому, що цілий пристрій з багатьох окремих елементів виготовляється на одній пластині напівпровідника в єдиному технологічному циклі. Тим самим досягається подальша мініатюризація, зростання швидкодії, зменшення енергоспоживання та радикальне збільшення надійності. Так, якщо в 1970-х рр. мінімальний розмір елемента мікросхеми складав 2-8 мкм, то в 2022 р. – 3 нм, тобто на три порядки менше. Сучасні інтегральні схеми включають до 2 млрд. елементів на одному кристалі. Саме інтегральні мікросхеми лежать, зокрема, в основі конструкції сучасних комп'ютерів.

Розвиток елементної бази радіоелектроніки відбувався на базі створення відповідних технологій: отримання чистого кремнію (який сьогодні є основним матеріалом напівпровідникової електроніки), його легування, методи вирощування плівок, методи обробки поверхні тощо.

Зараз на практиці, як правило, використовуються інтегральні мікросхеми (принаймні, в масовому виробництві). Помітне місце серед них займають аналогові схеми (крім аналогових, існують також цифрові й аналогово-цифрові мікросхеми). В основі їхньої роботи лежать ті самі принципи, що й в основі роботи лампових та напівпровідникових схем. У цьому курсі ми розглянемо основні схеми аналогової радіоелектроніки, реалізовані на основі напівпровідникових приладів. Тому розгляд елементів радіоелектроніки ми почнемо саме з напівпровідникових приладів.

Найпоширеніші напівпровідникові активні елементи (елементи, що забезпечують, зокрема, підсилення сигналів) – це польові й біполярні транзистори. Часто використовуються також різні типи діодів. Щоб рухатися від простішого до складнішого, ми спершу розглянемо напівпровідникові діоди, які містять один *p-n* перехід, а вже потім перейдемо до різних типів транзисторів, будова яких є більш складною.

Контрольні питання до підрозділу 3.1

1. Де зараз використовуються електронні лампи?
2. Що таке інтегральна мікросхема?
3. В чому перевага інтегральних схем над схемами з дискретних елементів?

3.2. Елементи фізики напівпровідників

Перш ніж розглядати роботу конкретних напівпровідникових приладів, нагадаємо (в найпростішому вигляді) деякі елементи фізики напівпровідників – зокрема, те, що стосується *p-n* переходу.

3.2.1. Механізми протікання струму в напівпровідниках

Напівпровідники (англ. semiconductors) займають проміжне становище між діелектриками, у яких електрони знаходяться в зв'язаному стані, і металами, в яких електрони можуть вільно рухатися в об'ємі кристала. В напівпровідниках при кімнатній температурі лише незначна частина електронів відірвана від атомів (вузлів кристалічної ґратки) і може переміщуватися по об'єму кристала. Однак кількість таких електронів – так званих електронів провідності – різко зростає зі зростанням температури.

На місці електрона, відірваного від атома, утворюється вакансія, тобто нескомпенсований позитивний заряд – дірка (англ. hole). В електричному полі електрон від сусіднього атому може перестрибнути на вільне місце (вакансію). Зовні це виглядає як рух дірки. Тому говорять, що заряд у напівпровідниках переноситься електронами та дірками. Очевидно, в чистому напівпровіднику концентрація електронів і дірок однакова.

Однак у напівпровідниковій електроніці звичайно використовуються так звані домішкові напівпровідники, леговані атомами, валентність яких на одиницю більша (донори) або на одиницю менша (акцептори) від валентності атомів напівпровідника. Донорні атоми в кристалічній ґратці значно легше віддають «зайвий» електрон, ніж атоми власне напівпровідника. Навпаки, акцепторні атоми легко «прив'язують» до себе додатковий електрон. Як результат, при легуванні донорами в напівпровіднику носіями заряду виступають переважно електрони (напівпровідник *n*-типу), при легуванні акцепторами – дірки (напівпровідник *p*-типу). Концентрація носіїв, створених за рахунок домішок (основних носіїв, англ. main carriers), на порядки переважає концентрацію власних носіїв. Тим не менше, в напівпровіднику завжди наявні так звані неосновні носії (minor carriers), тобто дірки в напівпровіднику *n*-типу і електрони в напівпровіднику *p*-типу.

Рухаючись а електричному полі, носії заряду зазнають зіткнень із вузлами кристалічної ґратки, В результаті вони втрачають спрямовану швидкість, набуту в зовнішньому електричному полі, і їхня кінетична енергія переходить у теплову. Так проявляється закон Джоуля – Ленца на мікроскопічному рівні.

Зустріч вільного електрона з діркою може привести до їхньої рекомбінації (коли вільний електрон займає вакансію, що існувала в кристалі). З іншого боку, за рахунок енергії теплових коливань ґратки можлива генерація електронно-діркових пар (відрив електрона від вузла кристалічної гратки). Процеси генерації та рекомбінації носіїв перебувають у динамічній рівновазі.

3.2.2. p-n перехід та його вольт-амперна характеристика

Розглянемо контакт двох напівпровідників з різними типами провідності. Такий контакт в літературі дістав назву *p-n* переходу (англ. *p-n* junction). В перші моменти часу після утворення такого контакту через нього виникає дифузійний (зумовлений градієнтом концентрації) потік електронів з *n*-області до *p*-області й аналогічний потік дірок у протилежному напрямку. Обидва потоки приведуть до протікання струму з *p*-області до *n*-області. Як наслідок, перша з них зарядиться негативно, а друга – позитивно. Таким чином, виникає електричне поле, яке відтісняє основні носії від поверхні контакту, по обидва боки від якої в результаті утворяться збіднені шари (англ. depletion layers) З іншого боку, це поле породжує дрейфовий струм, зумовлений неосновними носіями і спрямований назустріч дифузійному струму. В стані рівноваги дрейфовий і дифузійний струми компенсуються. Електричне поле в області *p-n* переходу породжує контактну різницю потенціалів.

Тепер прикладемо до *p-n* переходу зовнішнє електричне поле. Якщо воно спрямоване так само, як власне поле *p-n* переходу, воно зменшує дифузійний струм і може збільшувати дрейфовий. Однак останній залишається малим, оскільки він зумовлений неосновними носі-

ями. Таке увімкнення *p-n* переходу називають зворотним (англ. reverse turn-on of the *p-n* junction), див. рис. 3.2.1 а.

При зворотному увімкненні кількість неосновних носіїв в областях, що прилягають до *p-n* переходу, зменшується. В цьому випадку говорять про екстракцію (англ. extraction – видобування, витягання) неосновних носіїв (дірок із *n*-області та електронів із *p*-області).

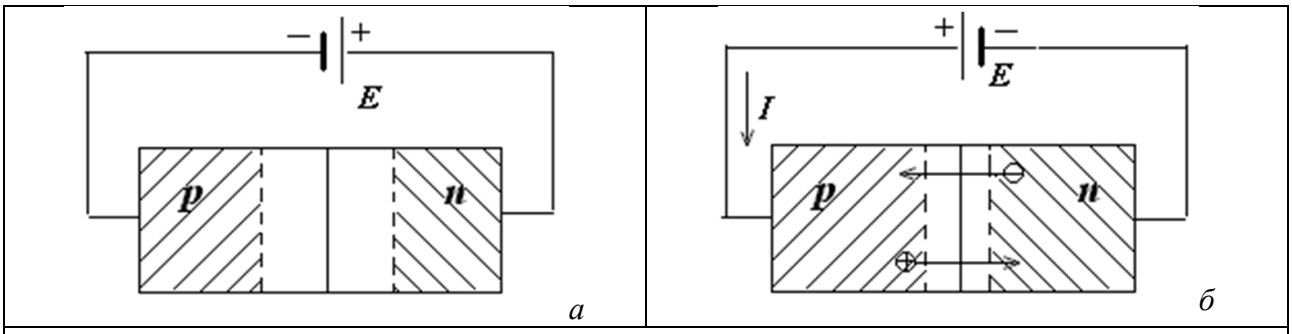


Рис. 3.2.1. Зворотне (а) та пряме (б) увімкнення p-n переходу

Навпаки, якщо зовнішнє поле спрямоване проти внутрішнього (пряме увімкнення *p-n* переходу, англ. direct activation of the *p-n* junction, див. рис. 3.2.1 б), то за достатньої різниці потенціалів воно може породити струм основних носіїв через контакт. Очевидно, цей струм буде значно більшим від струму при зворотному увімкненні, оскільки концентрація основних носіїв значно більша, ніж неосновних.

При прямому увімкненні кількість неосновних носіїв в областях, що прилягають до *p-n* переходу, збільшується. В цьому випадку говорять про інжекцію (англ. injection– ін'єкція, впорскування) неосновних носіїв (дірок в *n*-область та електронів у *p*-область).

Аналітично вольт-амперна характеристика (ВАХ) *p-n* переходу може бути записана у вигляді

$$I = I_0\left[\exp\left(\frac{eU}{k_B T}\right) - 1\right], \qquad (3.2.1)$$

де I_0 – струм насичення зворотно увімкненого *p-n* переходу, зумовлений неосновними носіями, U – пряма напруга, прикладена до *p-n* переходу, T – абсолютна температура (в Кельвінах), k_B – стала Больцмана, e – елементарний заряд. Ця формула справедлива при помірних значеннях $|U|$. При великих прямих напругах *p-n* перехід може згоріти, при великих зворотних напругах – пробитися (див. нижче п. 3.2.4).

Зазвичай концентрація основних носіїв у напівпровідниках, які утворюють *p-n* перехід, неоднакова (через різний ступінь легування), і струм через *p-n* перехід, увімкнений у прямому напрямку, створюється переважно одним типом носіїв. Тоді область, з якої витікають ці носії, називають емітером (від англ. emit – випромінювати, видавати, вивергати), а область, де вони опиняються, пройшовши через контакт – базою.

Графік ВАХ *p-n* переходу поданий на рис. 3.2.2 а. Видно, що *p-n* перехід є нелінійним елементом.

В багатьох випадках до нелінійних елементів прикладається сума постійної U та малої змінної $u(t)$ напруги, $u = U + u(t)$. Відповідно струм через нелінійний елемент у цьому випадку також складається з постійної I та малої змінної $i(t)$ складових, $i = I + i(t)$ (пор. з п. 2.1.1). Говорять, що значення U, I визначають робочу точку на вольт-амперній характеристиці нелінійного елемента. В обраній робочій точці відношення $R_O(U) = U/I$ називають омічним опором, а величину $R_d = du/di$ – диференціальним опором. Величина, обернена до диференціального опору, відома як диференціальна провідність.

Диференціальний опір *p-n* переходу

$$R_d(U)=\left.\frac{du}{di}\right|_{U,I}=\left.\left(\frac{di}{du}\right)^{-1}\right|_{U,I}=\left[\frac{eI_0}{k_BT}\exp\left(\frac{eU}{k_BT}\right)\right]^{-1}=\frac{k_BT}{eI_0}\exp\left(-\frac{eU}{k_BT}\right) \qquad (3.2.2)$$

є функцією прикладеної напруги (як і в інших нелінійних елементів). Те саме стосується й омічного опору

$$R_O(U)=\frac{U}{I}=\frac{U}{I_0\left[\exp\left(eU/k_BT\right)-1\right]}. \qquad (3.2.3)$$

Якщо до *p-n* переходу прикладена невелика напруга, так що $e|U|\ll k_BT$, то

$$R_d(U)\approx R_O(U)\approx\frac{k_BT}{eI_0}=const. \qquad (3.2.4)$$

Якщо *p-n* перехід відкритий ($eU>0$) і $eU\gg k_BT$, то

$$\frac{R_d(U)}{R_O(U)}=\frac{k_BT}{eI_0}\exp\left(-\frac{eU}{k_BT}\right)\frac{I_0}{U}\left[\exp\left(\frac{eU}{k_BT}\right)-1\right]\approx\frac{k_BT}{eU}\ll 1, \qquad (3.2.5)$$

тобто диференціальний опір значно менший від омічного, і обидва вони спадають зі зростанням прикладеної напруги, як випливає з формул (3.2.2)-(3.2.3).

Нарешті, коли *p-n* перехід закритий ($eU<0$) і $e|U|\gg k_BT$, то, як випливає з формул (3.2.2)-(3.2.3), зі зростанням $|U|$ диференціальний опір зростає експоненціально, а омічний – лінійно, тому в цій області $R_O\ll R_d$. Залежності диференціального та омічного опорів від напруги на *p-n* переході показані на рис. 3.2.2 б-в. Геометрично диференціальний опір – це котангенс кута нахилу дотичної, проведеної до робочої точки, омічний – кута нахилу січної графіка, що проходить через робочу точку (рис. 3.2.2 а).

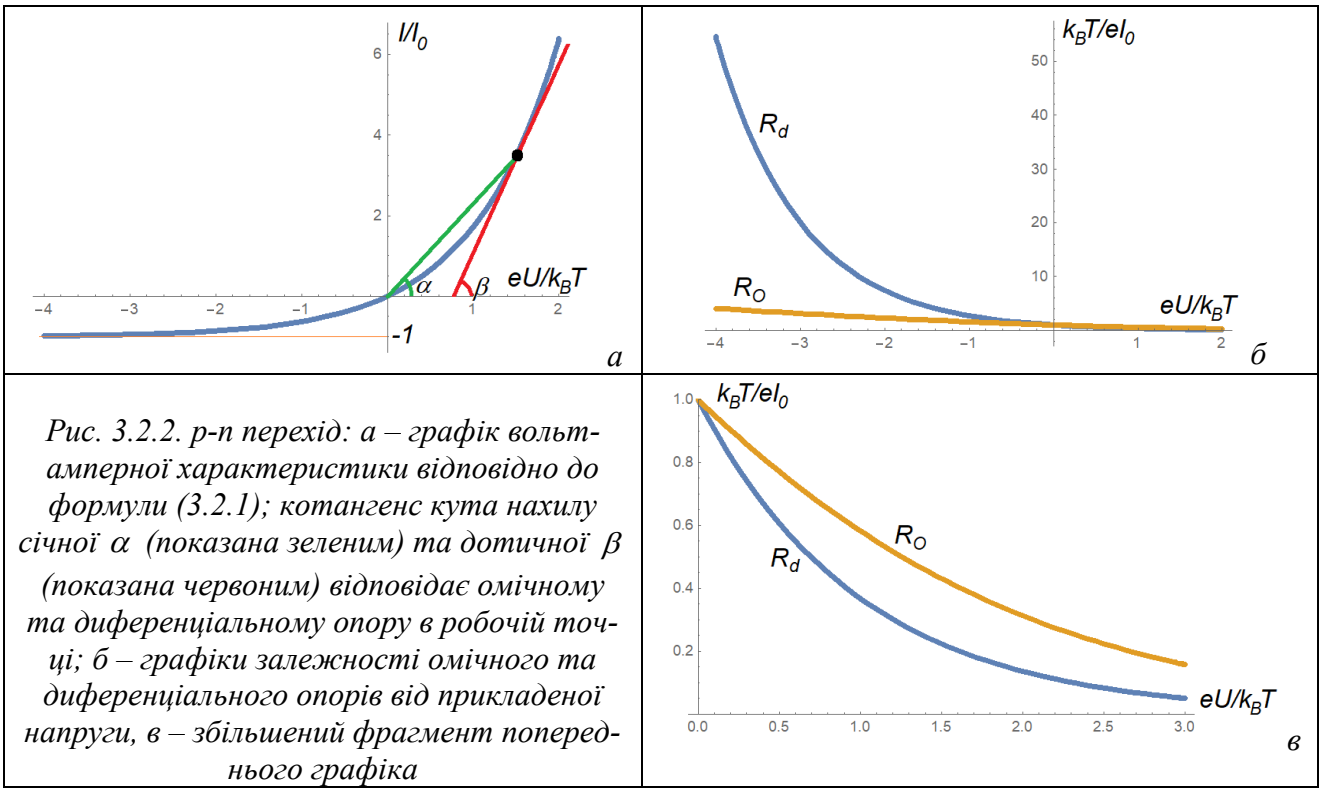


Рис. 3.2.2. p-n перехід: а – графік вольт-амперної характеристики відповідно до формули (3.2.1); котангенс кута нахилу січної α (показана зеленим) та дотичної β (показана червоним) відповідає омічному та диференціальному опору в робочій точці; б – графіки залежності омічного та диференціального опорів від прикладеної напруги, в – збільшений фрагмент попереднього графіка

3.2.3. Ємність p-n переходу

В об'ємі напівпровідника заряд вузлів кристалічної ґратки скомпенсований зарядом носіїв. Але коли біля контакту утворюється збіднений шар, там заряд вузлів ґратки виявляється нескомпенсованим, і з обох боків від контакту виникає об'ємний заряд протилежного

знаку (рис. 3.2.3 а). Загалом ця структура нагадує плоский конденсатор, і їй можна приписати деяку ємність, що називається бар'єрною ємністю *p-n* переходу. Зі зростанням прикладеної зворотної напруги ширина збідненого шару, яка виступає аналогом віддалі між пластинами конденсатора, зростає, а бар'єрна ємність відповідно зменшується (рис. 3.2.5 а).

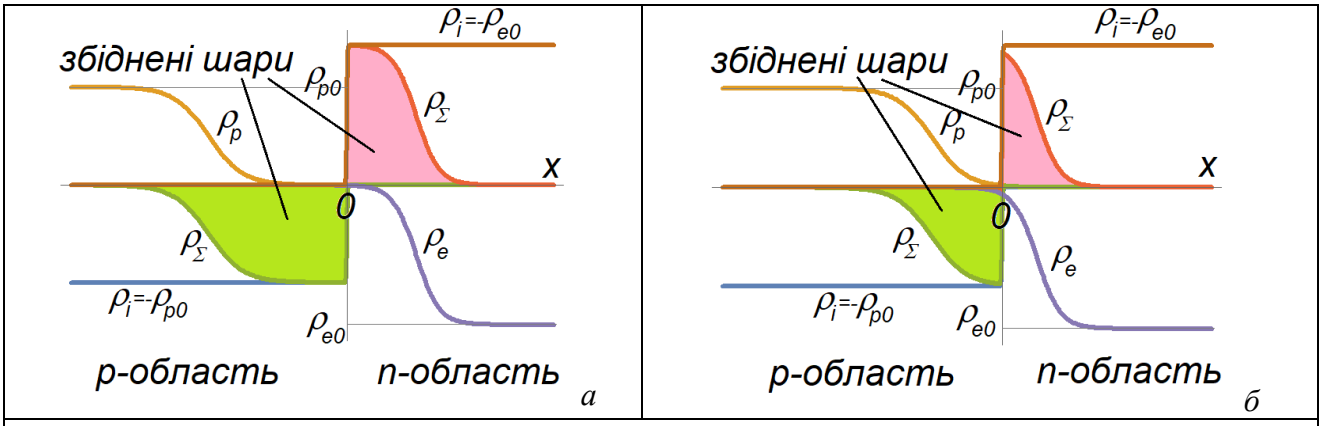


Рис. 3.2.3. Схема формування бар'єрної ємності p-n переходу при більшій (а) та меншій (б) прикладеній зворотній напрузі: ρ_i – густина заряду іонів, ρ_e та ρ_p – густини заряду електронів та дірок, ρ_Σ – сумарна густина заряду

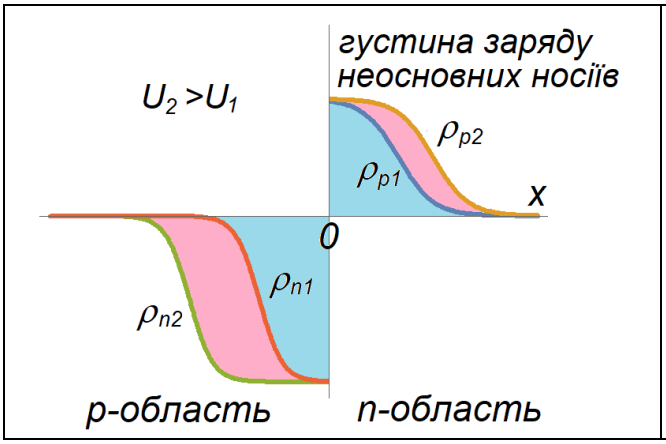


Рис. 3.2.4. Розподіл неосновних носіїв при прямій полярності увімкнення p-n переходу. Блакитним кольором показана площа, що відповідає повному заряду неосновних носіїв по обидва боки p-n переходу, рожевим – зміна цього заряду при збільшенні прикладеної напруги

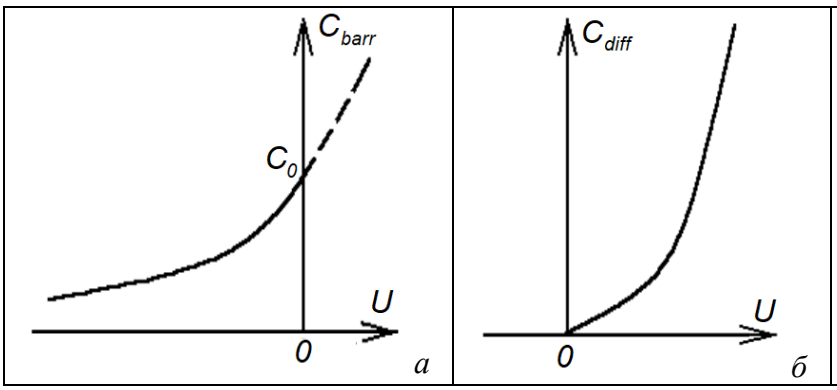


Рис. 3.2.5. Графіки залежностей бар'єрної (а) та дифузійної (б) ємностей p-n переходу від прикладеної напруги

При прямому увімкненні *p-n* переходу основні носії рухаються через контакт у сусідню область. Оскільки прикладене електричне поле зосереджене переважно в *p-n* переході, на деякій віддалі від контакту вони вже не дрейфують, а дифундують і поступово рекомбінують, але безпосередньо в приконтактній області утворюють нескомпенсований об'ємний заряд (рис. 3.2.4). Відповідна цьому ефекту ємність носить назву дифузійної. Точніше, дифузійна ємність визначається, як відношення зміни цього заряду до відповідної зміни напруги на *p-n* переході. В цьому сенсі дифузійну ємність можна розглядати як аналог диференціального опору. Ця ємність зростає зі зростанням прямої напруги (рис. 3.2.5.б).

3.2.4. Пробій p-n переходу

При великих зворотних напругах струм через *p-n* перехід різко зростає (рис. 3.2.6). Цей процес називають пробоєм (англ. breakdown) *p-n* переходу.

Відомо кілька механізмів пробою.

Лавинний пробій реалізується, якщо неосновний носій, який прискорюється електричним полем у збідненому шарі, до зіткнення з вузлом кристалічної ґратки встигає набрати достатню для іонізації енергію. Тоді в результаті такого зіткнення з'являється ще один електрон, і далі прискорюються вже два електрони. Вони, в свою чергу, знову зазнають зіткнень, і з'являється вже чотири електрони. Якщо довжина, на якій електрон-ініціатор набирає енергію, достатню для іонізації нейтрального атома, менша за довжину його вільного пробігу, а довжина вільного пробігу набагато менша за товщину збідненого шару, то в результаті виникає ціла лавина вторинних електронів. Зі сказаного зрозуміло, що лавинний пробій відбувається в широких *p-n* переходах (тобто при помірних концентраціях домішок). Лавинний пробій є оборотним процесом і не призводить до руйнування *p-n* переходу.

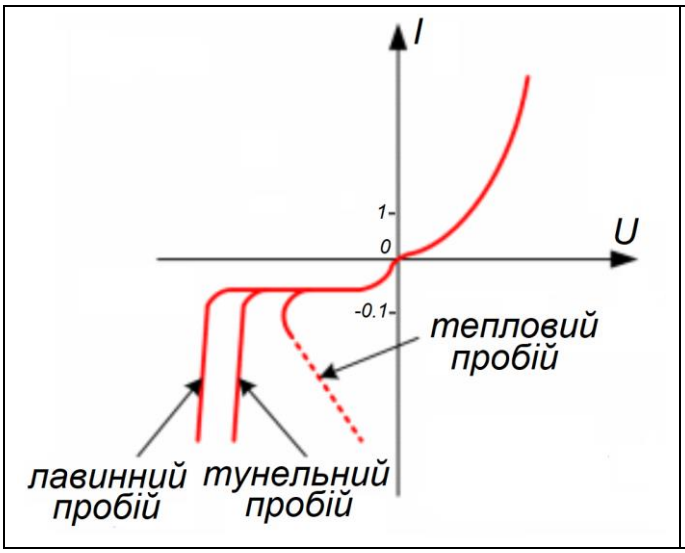


Рис. 3.2.6. Графік вольт-амперних характеристик, що відповідають різним типам пробою p-n переходу (зворотна гілка розтягнута по вертикалі порівняно з прямою)

У вузьких *p-n* переходах (при великих концентраціях домішок) може спостерігатися тунельний пробій. Під дією сильного електричного поля носії заряду можуть переходити через *p-n* перехід без зміни енергії (так званий тунельний ефект). Коректне пояснення цього явища дає так звана зонна теорія напівпровідників, розгляд якої виходить за межі цього курсу. Напруга пробою (напруга, з якої починається стрімке зростання зворотного струму) для тунельного пробою менша, ніж для лавинного, оскільки для вузького *p-n* переходу та сама напруженість електричного поля досягається при менших значеннях прикладеної напруги (рис. 3.2.6). Тунельний пробій, подібно до лавинного, є оборотним процесом і не призводить до руйнування *p-n* переходу.

Більш небезпечним є тепловий пробій. Якщо струм у перерізі *p-n* переходу розподілений нерівномірно, то в області з більшою густиною струму відбувається локальне розігрівання напівпровідника. В результаті концентрація носіїв у цій області зростає, і густина струму збільшується ще дужче. В кінцевому підсумку *p-n* перехід руйнується (згорає).

Контрольні питання до підрозділу 3.2

1. Що таке домішковий напівпровідник, напівпровідник n-типу, p-типу?
2. Яка електропровідність залежить від температури сильніше – власна чи домішкова?
3. Що називають основними та неосновними носіями заряду в напівпровіднику?

4. Чи існує збіднений шар на межі *p-n* переходу, якщо до нього не прикладена зовнішня напруга?
5. Що таке інжекція та екстракція неосновних носіїв заряду?
6. Що таке зворотний струм *p-n* переходу, чим він обумовлений, як він залежить від температури?
7. В чому різниця між омічним та диференціальним опорами відкритого $p-n$ переходу? Який з них більше?
8. Що таке бар'єрна ємність *p-n* переходу? Як вона залежить від знаку та величини прикладеної напруги?

Задачі до підрозділу 3.2

3.2.1. На скільки градусів повинна підвищитись температура, щоб концентрація електронів у чистому кремнії зросла вдвічі? Вихідною температурою вважати 300 К. *Вказівк*а: шуканий приріст температури можна вважати малим порівняно з вихідною температурою.

3.2.2. У кремній введено діркову (акцепторну) домішку в кількості 10^{-5} %. Якою буде концентрація дірок? Питома вага кремнію 2.5 г/см^3, атомна вага 2.81. *Вказівка*: можна вважати, що концентрація дірок, які утворилися, дорівнює концентрації атомів-акцепторів.

3.2.3. Якою буде рівноважна концентрація дірок у n-легованому германії при температурі 300 К, якщо концентрація донорних атомів становить 10^{16} см$^{-3}$?

3.2.4. Концентрація неосновних носіїв у германію становить $2\cdot10^8$ см$^{-3}$, середня швидкість їхнього теплового хаотичного руху 10^7 см/с. Якою буде обумовлена ними густина зворотного струму через *p-n* перехід?

3.2.5. Зворотний струм *p-n* переходу при температурі 300 К дорівнює 5 мкА. Чому дорівнюють омічний та диференціальний опори цього ж переходу у відкритому стані, якщо прикладено напругу +0.3 В?

3.2.6. Чому дорівнюватиме омічний та диференціальний опір ідеального *p-n* переходу при U = 0.1 В? Зворотний струм *p-n* переходу I_0 =10 мкА, температура 350 К.

3.3. Напівпровідникові діоди

Напівпровідниковим діодом називається електронний прилад з двома виводами, який містить один *p-n* перехід (рис. 3.1.2 а).

3.3.1. Випрямні діоди

Для випрямлення струму (тобто перетворення змінного струму на постійний) використовуються германієві або кремнієві площинні діоди, у яких поперечні розміри контакту набагато більші від товщини збідненого шару й глибини, на яку проникають за рахунок дифузії інжектовані в базу неосновні носії. Значна площа контакту забезпечує можливість проходження через такі діоди великих струмів.

Типові графіки вольт-амперних характеристик для германієвого (1) та кремнієвого (2) випрямних діодів зображені на рис. 3.3.1. Області прямої та зворотної напруги й відповідних їм струмів зображені в різних масштабах.

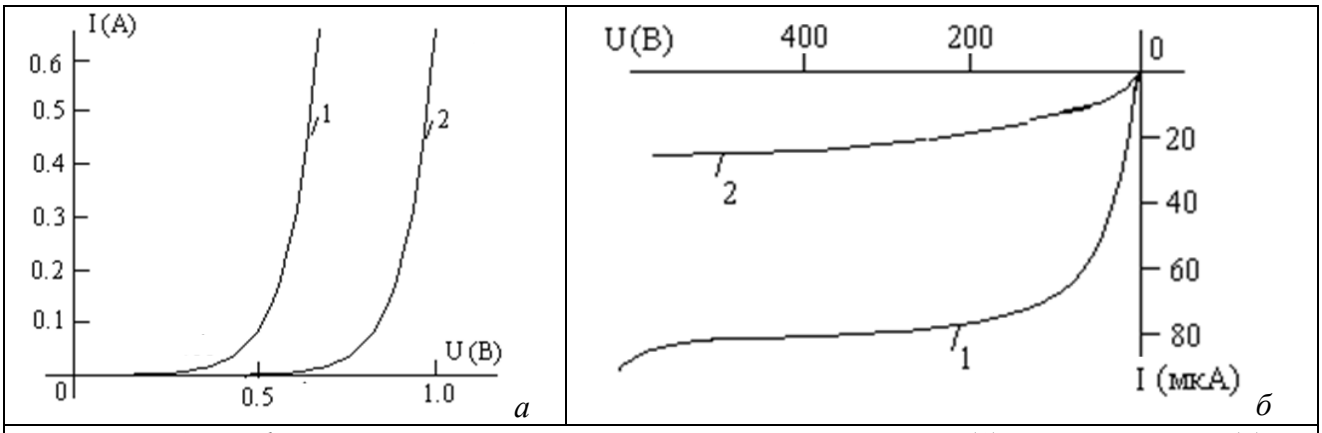


Рис. 3.3.1. Графіки вольт-амперних характеристик германієвого (1) та кремнієвого (2) випрямних діодів: а – пряме увімкнення, б – зворотне увімкнення

3.3.2. Діоди Шотткі

Велика площа контакту випрямних діодів зумовлює велике значення їхньої бар'єрної й дифузійної ємності. В результаті на високих частотах через такий діод протікає вже не струм провідності, а переважно струм зміщення, як через конденсатори, і вони втрачають свої випрямні властивості. Тому для детектування високочастотних сигналів (понад 1 ГГц) використовуються спеціальні високочастотні діоди, у яких площа контакту є малою. Часто в ролі високочастотних діодів використовують так звані діоди Шотткі, побудовані на основі контакту метал-напівпровідник.

Як відомо, концентрація вільних електронів у металах лежить у межах 10^{22} – 10^{23}см$^{-3}$. Напівпровідники мають значно меншу концентрацію вільних електронів. Ця величина змінюється в широкому діапазоні (при нормальній температурі для власних носіїв – 10^{16} – 10^{17}см$^{-3}$, з урахуванням домішок – 10^{18} – 10^{19}см$^{-3}$). Таким чином, струм через контакт метал-напівпровідник буде значно більшим тоді, коли він створюється електронами металу, тобто коли напівпровідник має позитивний потенціал щодо металу.

При прямому увімкненні діода Шотткі в напівпровідник не інжектуються неосновні носії (дірки). Тому в ньому не утворюється об'ємний заряд неосновних носіїв, відсутня дифузійна ємність і всі пов'язані з нею інерційні ефекти. Таким чином, інерційність діодів Шотткі визначається лише бар'єрною ємністю і набагато менша від інерційності діодів на основі *p-n* переходу.

Вольт-амперна характеристика діода Шотткі для прямого струму починається практично з нуля і стрімко йде вгору (рис. 3.3.2, пор. з рис. 3.3.1 а). Тому в такому режимі падіння напруги на діоді Шотткі невелике. Отже, й омічне нагрівання таких переходів виявляється меншим, аніж у *p-n* переходів на основі кремнію. Один з електродів переходу Шотткі є мета-

лом, завдяки чому забезпечується краще відведення тепла, і це дає змогу працювати з великими густинами струму.

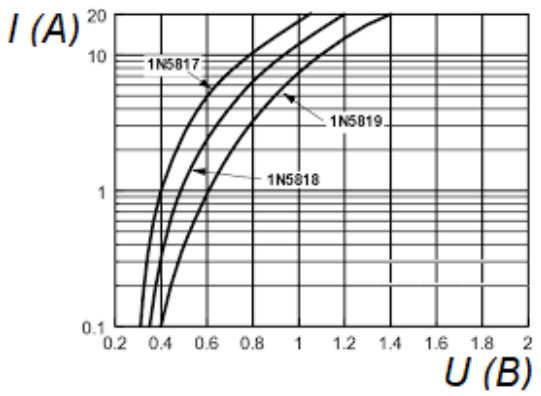
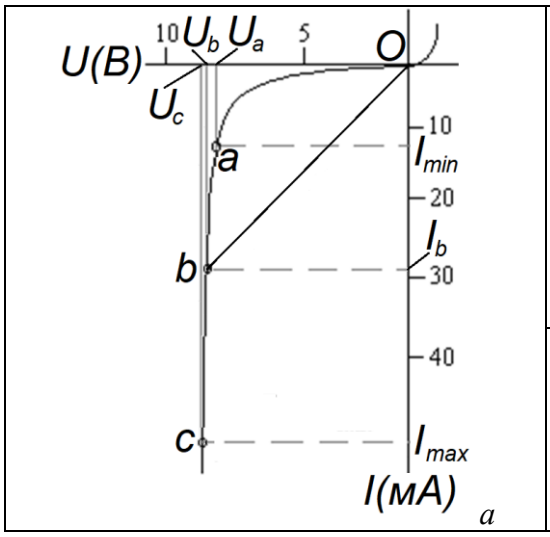


Рис. 3.3.2. Пряма гілка графіка вольт-амперної характеристики діода Шотткі (в логарифмічному масштабі, для різних марок діодів)

3.3.3. Стабілітрони і стабістори

Стабілітрон – це спеціальний кремнієві площинний діод, робоча точка якого розташована на ділянці вольт-амперної характеристики, яка відповідає лавинному пробою (див. вище п. 3.2.4) – рис. 3.3.3 а. Назва цих приладів пов'язана з їхнім використанням для стабілізації напруги від одиниць до десятків вольт.

На ділянці пробою *ac* (точка *a* відповідає початку пробою, точка *c* – максимально припустимому струму), починаючи з деякої напруги, позначеної на рис. 3.3.3 а як U_a, спостерігається швидке зростання зворотного струму. На цій ділянці (точніше, в точці *b*, що відповідає її середині) диференціальний опір $R_d = dU/dI \approx (U_a - U_c)/(I_{\max} - I_{\min})$, який визначає нахил характеристики до осі ординат, набагато менший від омічного опору $R_O = U_b/I_b$, якому відповідає нахил січної *Ob* до тієї самої осі. У стабілітронах ця відмінність складає один-два порядки.



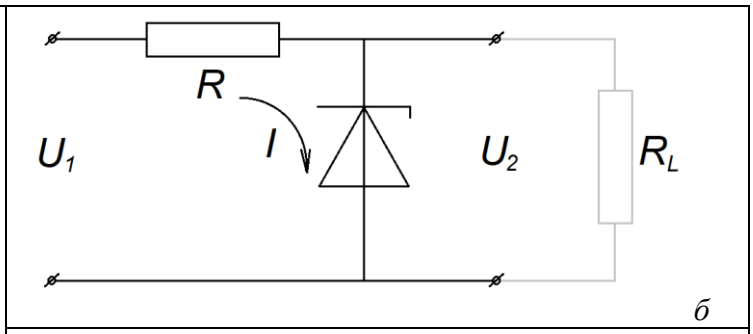


Рис. 3.3.3. Зворотна галка графіка вольт-амперної характеристика стабілітрона (а) і найпростіша схема стабілізатора напруги (б)

Стабілітрони найчастіше використовуються в стабілізаторах – пристроях, призначених для зменшення пульсацій напруги живлення радіоелектронних схем. Найпростіша схема стабілізатора на стабілітроні подана на рис. 3.3.3 б. По суті, це подільник вхідної напруги, а вихідна напруга знімається зі стабілітрона. Через нелінійність вольт-амперної характеристика стабілітрона цей подільник має різний коефіцієнт передачі для постійної напруги, який пропорційний R_O, і для малої змінної складової напруги – останній, очевидно, пропорційний R_d. Оскільки $R_d \ll R_O$, то за умови $R_d \ll R \ll R_O \ll R_L$ коефіцієнт передачі постійної напруги

$$K_0 \equiv \frac{U_{0out}}{U_{0in}} \approx \frac{R_O}{R_O + R} \approx 1, \qquad (3.3.1)$$

а коефіцієнт передачі змінної наруги

$$K_1 \equiv \frac{\Delta U_{out}}{\Delta U_{in}} \approx \frac{R_d}{R_d + R} \approx \frac{R_d}{R} \ll 1. \qquad (3.3.2)$$

Коефіцієнт стабілізації стабілізатора показує, наскільки зменшується відносний рівень пульсацій на виході порівняно зі входом:

$$K_{st} \equiv \frac{\Delta U_{in}/U_{0in}}{\Delta U_{out}/U_{0out}} = \frac{K_0}{K_1} \approx \frac{R_O}{R_O + R}\frac{R_d + R}{R_d} \approx \frac{R}{R_d} \gg 1. \qquad (3.3.3)$$

Ще раз підкреслимо, що формула (3.3.3) справедлива лише за умови, коли опір навантаження (рис. 3.3.3 б) значно перевищує омічний опір стабілітрона, $R_O \ll R_L$.

Стабістор, ВАХ якого подана на рис. 3.3.4, являє собою звичайний діод, увімкнений у прямому напрямку. В основі його роботи також лежить відмінність між диференціальним та омічним опорами діода, але вже для прямого струму (див. п. 3.2.2 та рис. 3.2.2 в). Ця відмінність особливо помітна для малих струмів. За допомогою стабістора стабілізують малі напруги порядку десятих часток вольта.

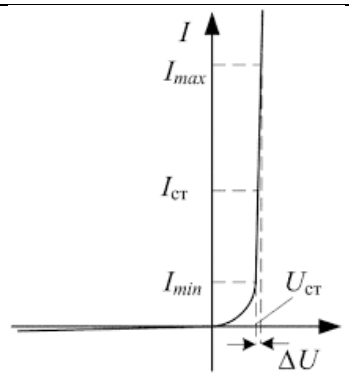


Рис. 3.3.4. Графік вольт-амперної характеристики стабістора

3.3.4. Варикап

Як уже йшлося, при запірній полярності прикладеної напруги по обидва боки від *p-n* переходу утворюються збіднені шари, які зумовлюють його бар'єрну ємність (див. вище п. 3.2.3). Оскільки товщина збідненого шару зростає при збільшенні прикладеної напруги, то відповідна йому бар'єрна ємність зменшується (рис. 3.3.5). Цей ефект використовують у спеціальних напівпровідникових приладах, так званих варикапах (від англ. varied capacity – змінна ємність) – приладах для створення електрично керованої ємності. При зміні прикладеної напруги ємність варикапа може змінюватись у півтора-два рази, а її середнє значення складає одиниці-сотні пФ. Варикапи використовуються, наприклад, для електричного налаштування коливних контурів у радіоприймальних пристроях, для забезпечення частотної модуляції сигналів, у параметричних підсилювачах тощо.

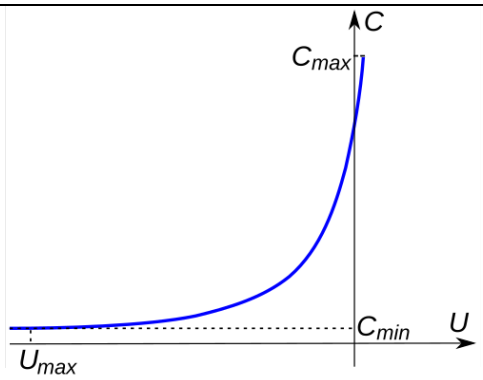


Рис. 3.3.5. Графік типової вольт-фарадної характеристика варикапа

3.3.5. Тунельний діод

В основі роботи тунельного діода лежить ефект тунелювання електронів через потенціальний бар'єр (про цей ефект уже згадувалося в п. 3.2.4), який пов'язаний із хвильовими властивостями електрона (та інших мікрочастинок)[1]. В тунельному діоді використовується сильно легований *p-n* перехід (пор. з п. 3.2.4), крізь який носії заряду можуть тунелювати за певних умов, які залежать від прикладеної напруги. Струм через *p-n* перехід зростає для значень напруги, при яких такі умови виконані, але зменшується при більшій напрузі, створюючи ділянку *bc* з від'ємною диференційною провідністю (рис. 3.3.6).

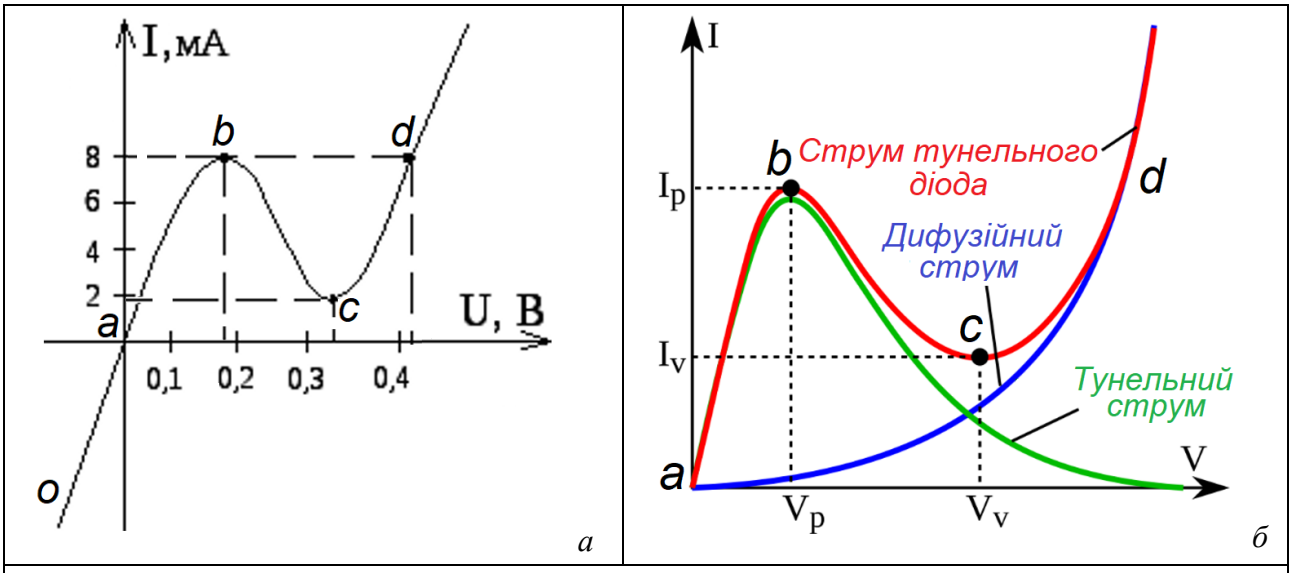


Рис. 3.3.6. Тунельний діод: графік вольт-амперної характеристики (а) та схема формування його прямого струму (б)

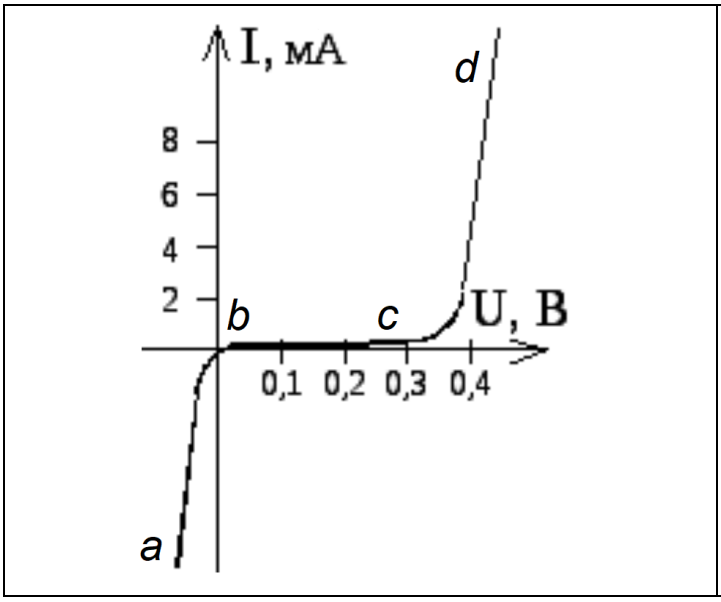


Рис. 3.3.7. Графік вольт-амперної характеристики оберненого діода

При зворотній полярності тунельний діод відкритий (ділянка *ao*), що зумовлено сильним легуванням *p-n* переходу (фактично спостерігається тунельний пробій, див. п. 3.2.4). При прямій полярності ділянка зростання струму *ab* якраз і створюється тунельним струмом (рис. 3.3.6 б). Вона переходить у ділянку *bc*, де умови тунелювання поступово порушуються, і збільшення напруги супроводжується зменшенням струму. На цій ділянці диференціальна провідність тунельного діода (та його диференціальний опір) від'ємні, хоча омічний опір

[1] Це та інші подібні явища вивчає розділ фізики, який називають квантовою механікою.

залишається додатним. Далі, починаючи з точки *c*, спостерігається нове зростання струму і на ділянці *cd* диференціальна провідність знову стає позитивною. Зростання струму на ділянці *cd* відповідає відкритому *p-n* переходу.

Наявність ділянки з від'ємним диференціальним опором обумовлює можливість використання тунельного діода для генерації високочастотних коливань.

Різновидом тунельного діода є обернений діод, у якому ділянка *bc* вольт-амперної характеристики йде практичного горизонтально (рис. 3.3.7). Обернені діоди використовують для детектування сигналів мікрохвильового діапазону.

3.3.6. Умовні позначення діодів

На рис. 3.3.8 наведено умовні позначення різних типів діодів, які застосовуються на радіоелектронних схемах.

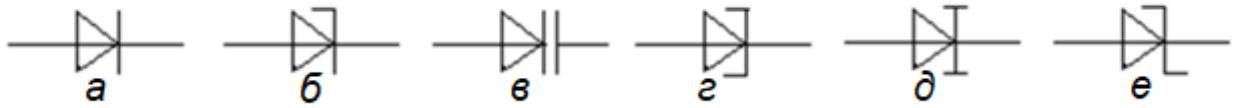


Рис. 3.3.8. Умовні позначення різних типів діодів: а – діоди випрямний, детекторний, змішувальний; б – стабілітрон; в – варикап; г – тунельний діод; д – обернений діод; е – діод Шотткі.

Контрольні питання до підрозділу 3.3

1. Яка властивість *p-n* переходу використовується для побудови випрямних діодів?
2. Яке фізичне явище лежить в основі роботи стабілітрона?
3. Що розуміють під терміном "стабілізація напруги"?
4. Чім визначаються межі області робочих струмів стабілітрона?
5. Яке фізичне явище лежить в основі роботи варикапа?
6. Як залежить ємність варикапа від величини прикладеної напруги?
7. Які основні особливості вольт-амперної характеристики тунельного діода?

Задачі до підрозділу 3.3

3.3.1. Яким буде граничний прямий допустимий струм для германієвого діода, вольт-амперна характеристика якого зображена на рис. 3.3.1 а, якщо він здатний розсіяти потужність в 0,1 Вт?

3.3.2. Якими будуть омічний та диференціальний опори стабілітрона, вольт-амперна характеристика якого зображена на рис. 3.3.3 а в робочій точці *b*?

3.3.3. Визначте крутість вольт-фарадної характеристики варикапа dC/dU, зображеної на рис. 3.3.5, в околі точки U = 5 В.

3.3.4. Для стабілізатора, зображеного на рис. 3.3.3 б, знайдіть значення опору R та коефіцієнта стабілізації при U_1=18 В та U_2=10 В. Параметри стабілітрона: U_b=10 В, $I_{\min}$=3 мА, $I_{\max}$=23 мА, R_d= I5 Ом.

3.3.5. Знайдіть межу стабілізації за вхідною напругою для схеми, розрахованої в завданні 3.3.4. Як вона зміниться, якщо стабілізатор навантажити опором R_L= 1 кОм?

3.4. Польові транзистори

Як уже згадувалося, в сучасній електроніці в ролі активних елементів, які забезпечують підсилення напруг та струмів, виступають польові та біполярні транзистори. Ми розпочнемо розгляд з польових транзисторів, оскільки принцип їхньої дії простіший порівняно з біполярними транзисторами. Польові транзистори широко застосовуються як в аналоговій, так і особливо в цифровій електроніці, зазвичай як елементи інтегральних мікросхем.

Транзи́стор (англ. transistor, скорочення від transfer resistance – перехідний, або змінюваний опір) – це напівпровідниковий елемент, який дозволяє керувати струмом, що протікає крізь нього, за допомогою зміни вхідної напруги (тобто електричного поля) або вхідного струму. Невелика зміна вхідного сигналу може приводити до суттєво більшої зміни вихідного сигналу (напруги та/або струму).

Назва польових транзисторів (англ. – field-effect transistor, FET) пов'язана з тим, що вони керуються електричним полем (вхідною анпругою). Це означає, що в робочому режимі їхній вхідний струм можна вважати практично нульовим. Завдяки цьому вони фактично не споживають потужність вхідного сигналу. Це дозволило, зокрема, використати їх у таких пристроях, як електронні годинники чи пульти керування телевізором, які живляться від малопотужної батарейки, й зумовило надзвичайно широке застосування польових транзисторів у цифровій електроніці.

Польові транзистори також називають уніполярними (англ. – unipolar transistor). Це пов'язано з тим, що в цих приладах протікання струму забезпечується лише одним типом носіїв заряду в напівпровіднику з одним типом провідності. Ще одна їхня назва – канальні – відображає той факт, що рух носіїв відбувається через провідну область, яка має назву каналу, переріз і, відповідно, загальний опір котрого регулюються вхідною напругою.

Існує кілька конструкцій польових транзисторів. Найвідоміші з них – це польові транзистори з керувальним *p-n* переходом та польові транзистори з ізольованим затвором (або МДН-транзистори), які, в свою чергу, діляться на транзистори з вбудованим каналом та з індукованим каналом. Розглянемо їхню роботу послідовно.

3.4.1. Польовий транзистор з керувальним p-n переходом

Принцип роботи польового транзистора з керувальним *p-n* переходом легко зрозуміти з його схематичної моделі, зображеної на рис. 3.4.1. Його основною частиною є об'єм слабколегованого напівпровідника (у даному разі n-типу; для спрощення вважаємо, що він має форму витягнутого паралелепіпеда), до торців якого прикладена напруга U_{DS}.

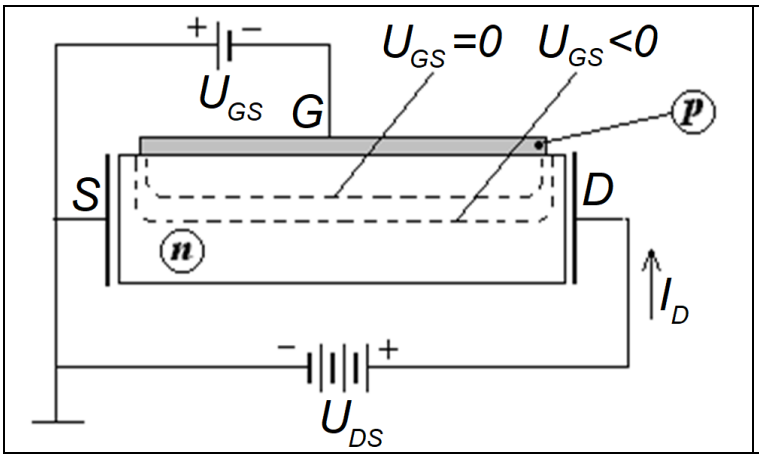


Рис. 3.4.1. Схематична модель польового транзистора з керувальним p-n переходом та n-каналом

В результаті руху електронів від електрода S (витоку, англ. source) до електрода D (стоку, англ. drain) виникає наскрізний струм $I_D = U_{DS}/R$, величина котрого визначається омічним опором R об'єму напівпровідника. Верхня грань об'єму межує з шаром *p*-напівпровідника, на який ззовні нанесений шар металу – затвор (англ. gate). Між напівпровідниками з різним типом провідності виникає *p-n* перехід. Збіднений шар, що існує навколо

p-n переходу, дещо звужує переріз *n*-області, по якій тече струм I_D, і збільшує її омічний опір. Якщо до затвору прикласти напругу $U_{GS} < 0$ запірної полярності, то можна збільшити товщину збідненого шару, зменшивши тим самим переріз області, через яку тече струм I_D, і збільшивши її опір. При великому значенні затворної напруги межа збідненого шару може досягти нижньої грані паралелепіпеда і повністю перекрити шлях струму I_D. Та напруга, затвора, при якій струм I_D зникає, зветься запірною напругою і позначається як U_{G0}.

Таким чином, з'являється можливість керування наскрізним струмом I_D між витоком та стоком шляхом зміни затворної напруги U_{GS}. Це керування здійснюється, як уже згадувалося, електричним полем, яке існує у збідненому шарі.

В певному сенсі польовий транзистор можна розглядати як змінний опір (між витоком і стоком), величиною якого можна керувати за допомогою напруги затвор-виток.

Із викладеного вище зрозуміло, що вхідна напруга, яка керує роботою польового транзистора, прикладається між затвором та витоком. Оскільки між цими контактами ввімкнений запертий *p-n* перехід, то вхідний струм є дуже малим. Для кремнієвих польових транзисторів він може складати 10^{-8}-10^{-9}А, так що на керування вхідним струмом потужність практично не витрачається. Іншими словами, вхідний опір польового транзистора R_{GS} звичайно складає 0.1-1 ГОм. Підключений до джерела вхідного сигналу польовий транзистор практично не навантажує це джерело і не впливає на його роботу. В цьому й полягає основна перевага польових транзисторів над біполярними.

В реальності у польових транзисторів об'єм напівпровідника, в якому протікає струм, має форму не паралелепіпеда, а циліндра. Відповідно контактна поверхня *p-n* переходу збігається з бічною поверхнею цього циліндра (рис. 3.4.2 а).

На рис. 3.4.2 б показана сім'я вихідних характеристик польового транзистора, тобто залежностей вихідного (стокового) струму від вихідної напруги (напруги стік-виток). Параметром сім'ї характеристик служить вхідна напруга (напруга затвор-виток). Видно, що закон Ома, який ми використовували для пояснення принципу роботи транзистора, описує лише початкові ділянки цих характеристик. При подальшому зростанні стокової напруги нахил характеристик зменшується, тобто опір області між стоком та витоком зростає. Іншими словами, польовий транзистор є нелінійним елементом.

Зростання його вихідного омічного опору при великих стокових напругах U_{DS} пов'язано з тим, що при $|U_{DS}| > |U_{GS}|$ до різних ділянок *p-n* переходу прикладається різна напруга. Оскільки весь затвор укритий металевим контактом, прилеглу до нього область можна вважати еквіпотенціальною з потенціалом $-U_{GS}$ щодо затвору (вважаємо його заземленим). А потенціал на осі провідної області зростає від нуля на витоку до U_{DS} на стоку. Отже, падіння напруги на *p-n* переході зростає від U_{GS} біля витоку до $U_{GS} + U_{DS}$ біля стоку. В результаті біля стоку провідний канал звужується зі зростанням U_{DS} (рис. 3.4.2 а), а його опір, відповідно, зростає, чим і пояснюється хід вихідних характеристик, показаних на рис. 3.4.2 б.

Диференціальний вихідний опір польового транзистора R_{DS} на ділянках вихідних характеристик, що відповідають зростанню опору провідного каналу, звичайно складає 10-100 кОм.

За вихідними характеристиками можна побудувати так звані статичні[2] прохідні характеристики польового транзистора – залежності вихідного (стокового) струму від вхідної (затворної) напруги при фіксованій напрузі витоку (рис. 3.4.2 в). Для цього треба на графіку сім'ї вихідних характеристик (рис. 3.4.2 б) провести вертикальну лінію, що відповідає обра-

[2] Під статичним режимом польового транзистора розуміють увімкнення джерела живлення (джерела вихідної напруги) безпосередньо між витоком і стоком.

ному значенню стокової напруги, і записати значення стокового струму, що відповідають різним значенням напруги на затворі, тобто точкам перетину вертикалі з різними вихідними характеристиками (рис. 3.4.2 г).

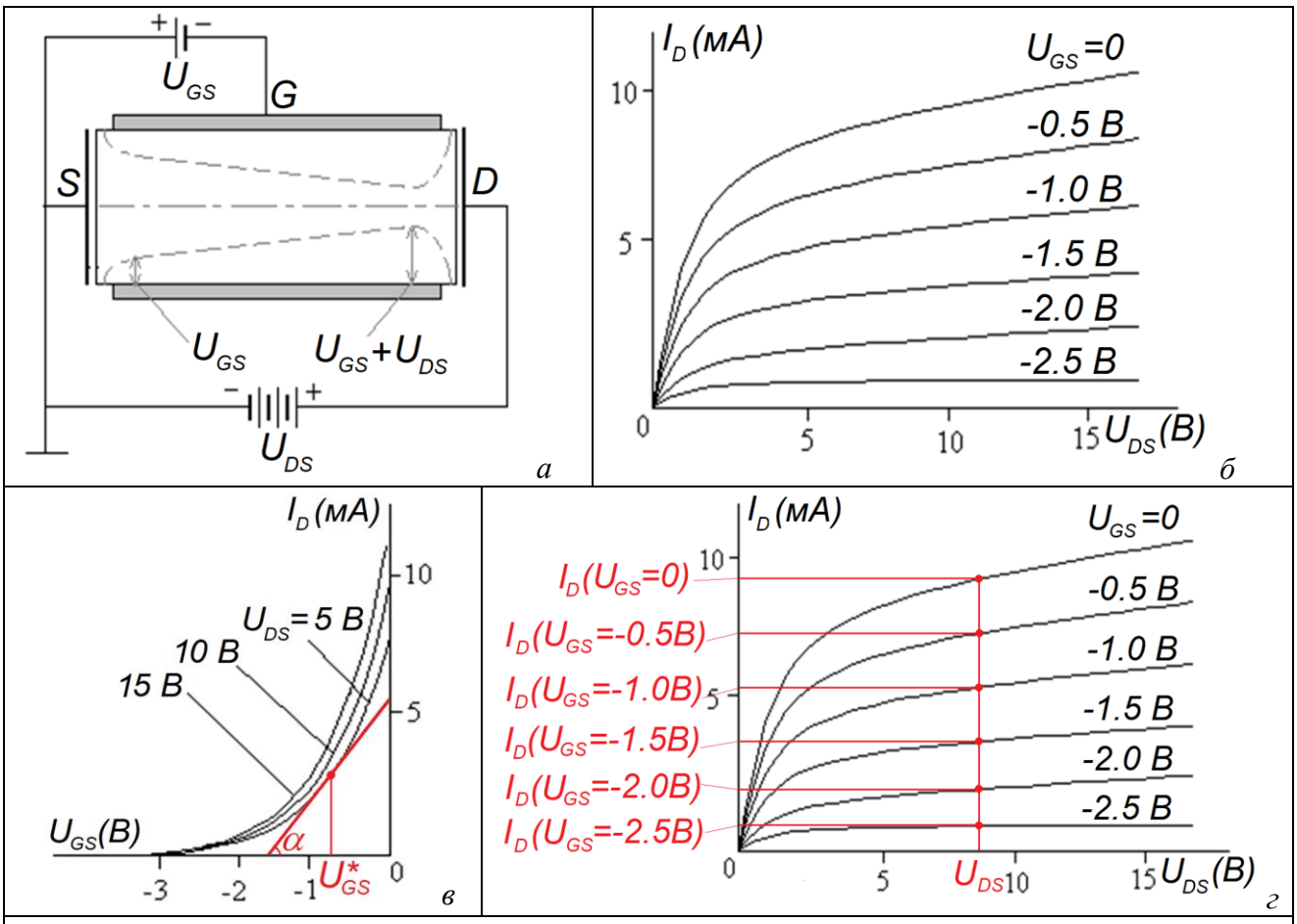


Рис. 3.4.2. Польовий транзистор з керувальним p-n переходом та n-каналом: а – схема аксіально-симетричного транзистора; б, в – графіки типової сім'ї вихідних (б) та прохідних (в) характеристик; г – схема побудови графіка статичних прохідних характеристик на основі вихідних характеристик

З рис. 3.4.2 в видно, що прохідні характеристики слабко залежать від затворної напруги. Їхній нахил $S = dI_D / dU_{GS}\big|_{U^*_{GS}} = tg\,\alpha$ називають крутістю прохідної характеристики (тут α – кут нахилу дотичної до прохідної характеристики в обраній робочій точці U^*_{GS}, див. рис. 3.4.2 в). Для польових транзисторів з керувальним *p-n* переходом вона звичайно складає 1 - 10 мА/В. Зазначимо, що робоча точка (середина ділянки, близької до лінійної) на прохідній характеристиці транзисторів такого типу лежить в області негативних затворних напруг (для транзисторів з *n*-каналом).

Зазначимо, що прохідні та вихідні характеристики польового транзистора з керувальним *p-n* переходом та *n*-каналом мають практично такий самий вигляд, як аналогічні характеристики тріода (пор. рис. 3.4.2 б-в та рис. 3.1.1 в-г).

Таким чином, еквівалентну схему польового транзистора за сигнальною складовою можна подати у вигляді, показаному на рис. 3.4.3 а. На практиці звичайно можна вважати, що $R_{GS} \to \infty$ і, відповідно, $i_G = 0$.

Польовому транзистору можна приписати еквівалентну вхідну ємність C_{in} та вихідну ємність C_{out} (рис. 3.4.3 б). Їхня величина складає одиниці пФ. Фактично частотні властивості польових транзисторів визначаються часом перезарядки бар'єрної ємності затворного *p-n*

переходу. Дійсно, при зміні керувальної напруги U_{GS} струм, який заряджає цю ємність, повинен пройти від витоку до межі збідненого шару через матеріал каналу, омічний опір якого порядку 1-10 кОм. Для бар'єрної ємності 1 пФ маємо сталу часу $10^{-8} - 10^{-9}$ с, що відповідає граничній частоті роботи транзистора 0.1-1 ГГц. Це дещо менше за граничні частоти кращих зразків біполярних транзисторів.

Максимальна потужність розсіювання для польових транзисторів малої потужності становить 10-100 мВт, а для потужних – до 10 Вт і більше.

Умовне зображення польових транзисторів з керувальним $p-n$ переходом подане на рис. 3.4.4.

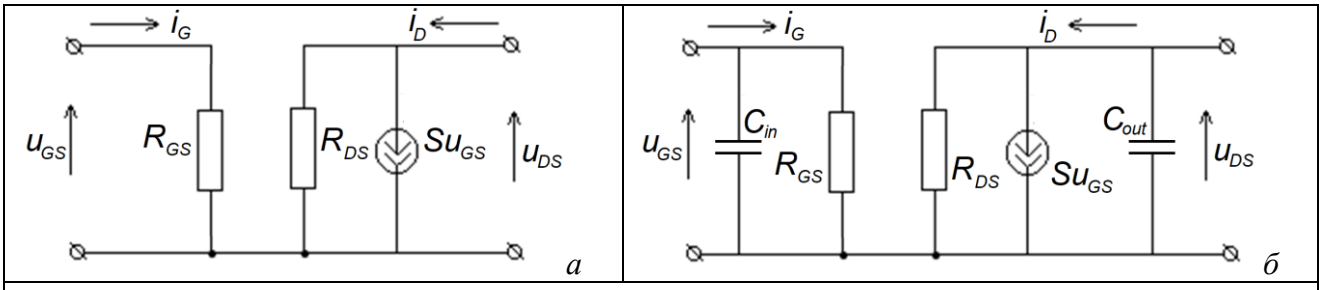


Рис. 3.4.3. Еквівалентна схема польового транзистора за сигнальними складовими напруги та струму на низьких (а) та на високих частотах

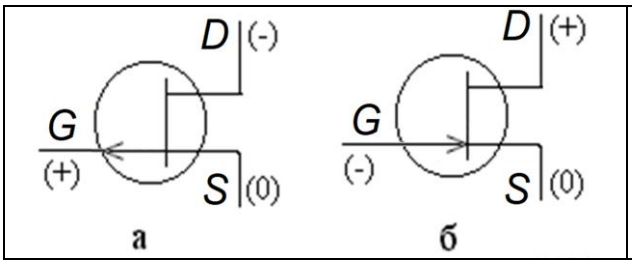


Рис. 3.4.4. Умовне зображення польових транзисторів з керувальним p-n переходом: а – з p-каналом; б – з n-каналом (S – від англ. source, виток; D – від англ. drain, стік; G – від англ. gate, затвор)

3.4.2. Польові транзистори з ізольованим затвором

Схематична модель польового транзистора з ізольованим затвором та вбудованим каналом подана на рис. 3.4.5. Основна структура, через яку протікає струм, виготовлена в даному разі з напівпровідника *p*-типу, оточеного напівпровідником *n*-типу (його називають підкладкою, англ. substrate). Бічні частини каналу виготовлені із сильнолегованого *p*-напівпровідника і через отвори в тонкому (частки мікрона) шарі діелектрика сполучені з металевими контактами – стоком та витоком. Затвор відділений від каналу шаром діелектрика.

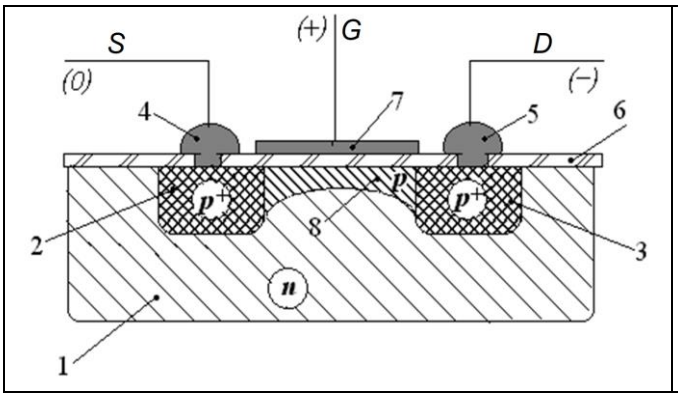


Рис. 3.4.5. Схематична модель польового транзистора з ізольованим затвором та вбудованим p-каналом:1 – напівпровідник n-типу; 2, 3 – сильнолеговані області p-типу, сполучені відповідно з витоком 4 і стоком 5; 6 – шар діелектрика; 7 – ізольований затвор; 8 – канал з напівпровідника p-типу, в якому протікає стоковий струм

Коли затвор заземлений, при прикладанні напруги між витоком та стоком через канал протікає струм. В такому режимі *p-n* перехід між каналом і підкладкою закритий. Подання негативної напруги між затвором і витоком приводить до того, що в канал всмоктуються додаткові дірки з сильно легованих *p*-областей, і тим самим опір каналу зменшується. При по-

данні на затвор позитивної напруги електричне поле, навпаки, відганяє дірки з каналу і збільшує його опір. В результаті за допомогою напруги на затворі вдається керувати стоковим струмом.

Пристрої такого типу відомі як польові транзистори з вбудованим каналом. Оскільки за своєю структурою вони виглядають як чергування шарів металу, діелектрика й напівпровідника, їх називають також МДН-транзисторами (англ. MIS transistor, скорочено від metal-insulator-semiconductor). Нерідко (зокрема, для кремнієвих транзисторів) діелектриком служить плівка оксиду самого кремнію. В цьому випадку використовують назву «МОН-транзистор» (англ. MOSFET – metal-oxide-semiconductor field effect transistor).

Прохідна та вихідні характеристики МДН-транзистора з вбудованим *p*-каналом схематично показані на рис. 3.4.6.

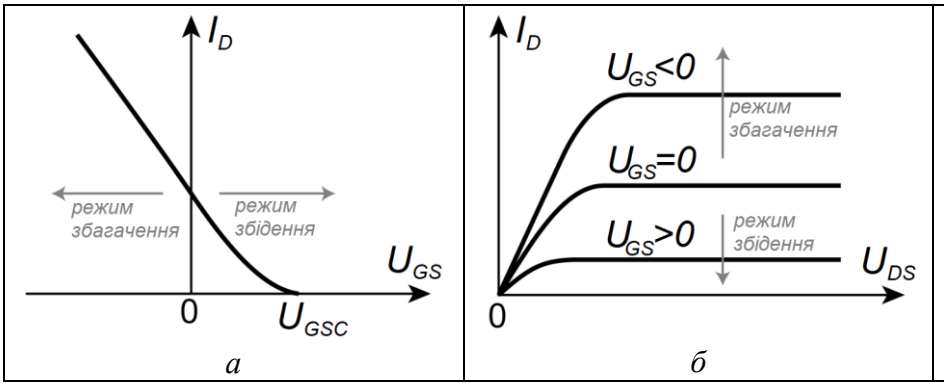


Рис. 3.4.6. Графіки прохідної (а) та вихідних (б) характеристик МДН-транзистора з вбудованим p-каналом (показані схематично); U_{GSC} – напруга відтину (cut-off), або запирання на затворі

Поширеним різновидом МДН-транзистора є транзистор з індукованим каналом. Від зображеного на рис. 3.4.4 він відрізняється тим, що спеціально виготовлений канал 8, який сполучає виток зі стоком, у ньому відсутній. Тому при нульовій (і тим більше при позитивній) напрузі на затворі струм через канал відсутній. Однак при негативній напрузі на затворі дірки з підкладки та (головним чином) із сильно легованих *p*-областей притягаються до затвору, утворюючи індукований канал. Через цей канал протікає струм, величина якого залежить від затворної напруги. Прохідна та вихідні характеристики МДН-транзистора з індукованим *p*-каналом подані на рис. 3.4.7.

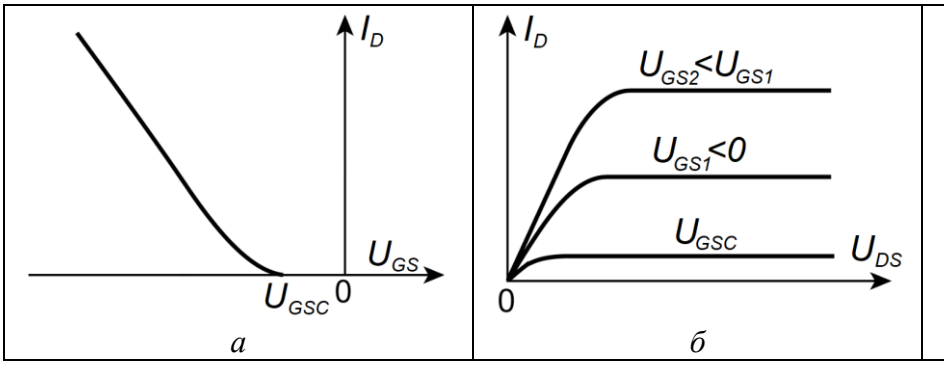


Рис. 3.4.7. Графіки прохідної (а) та вихідних (б) характеристик МДН-транзистора з індукованим p-каналом (показані схематично)

Наведені вище міркування сточуються транзистора з каналом *p*-типу. Зрозуміло, що в транзисторі з індукованим *n*-каналом робочий режим забезпечується при позитивних напругах на затворі.

МДН-транзистори характеризуються тими самими параметрами, що й польові транзистори з керувальним *p-n* переходом. Вихідний опір, крутість, вхідна та вихідна ємність МДН-транзисторів мають той самий порядок величини, що й у польових транзисторів з *p-n* переходом. Істотно відрізняється лише вхідний опір, який для МДН-транзисторів може досягати 10^{14}-10^{15}Ом. Великий вхідний опір є перевагою МДН-транзисторів, але одночасно й їхнім недоліком, оскільки ці транзистори дуже чутливі до статичної електрики. Необережний дотик до затвору інструментами чи пальцями, на яких є заряд статичної електрики, може приз-

вести до пробою тонкого шару діелектрика і виходу транзистора з ладу. Тому при роботі з МДН-транзисторами потрібно завжди заземлювати як тіло працівника, так і інструменти, якими він користується.

На рис. 3.4.8 для порівняння подані статичні прохідні характеристики різних типів польових транзисторів з *p*-каналом. Для МДН-транзисторів із індукованим каналом (1) робочу точку звичайно обирають при деякій негативній напрузі на затворі, для транзисторів із вбудованим каналом (2) – при нульовій напрузі. Для транзистора з керувальним *p-n* переходом (3) робочу точку обирають при позитивній затворній напрузі.

Еквівалентна схема МДН-транзисторів за сигналом така сама, як у транзисторів з керувальним *p-n* переходом (рис. 3.4.3).

Умовні позначення транзисторів із вбудованим та індукованим каналами подані на рис. 3.4.9.

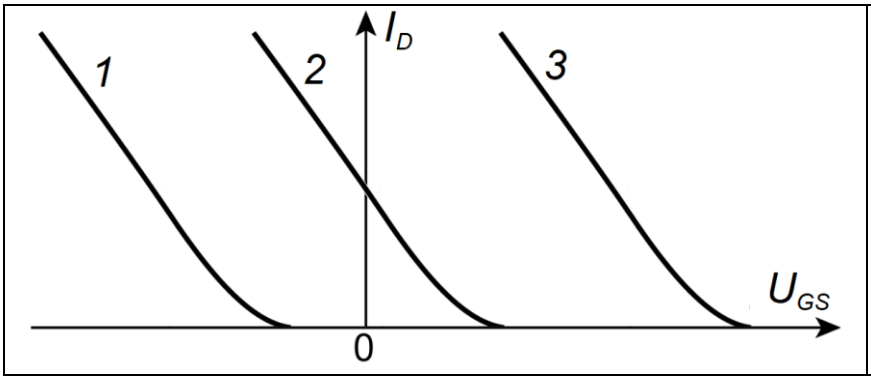


Рис. 3.4.8. Схематичне порівняння графіків статичних прохідних характеристик польових транзисторів з p-каналом: 1 – МДН-транзистор з індукованим каналом; 2 – МДН-транзистор з вбудованим каналом; 3 – транзистор з керувальним p-n переходом

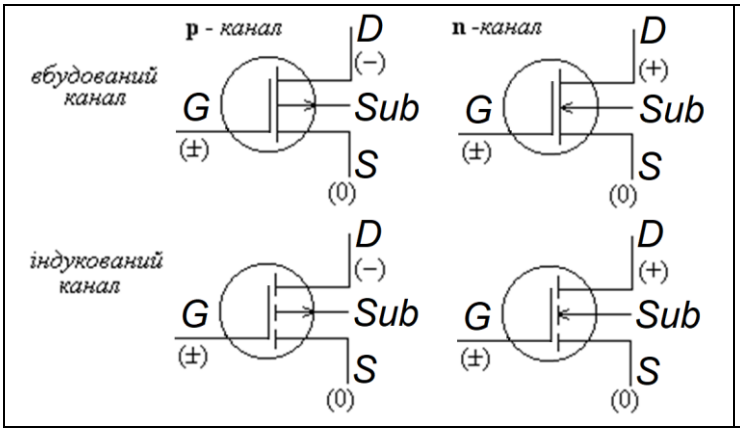


Рис. 3.4.9. Умовні позначення транзисторів із вбудованим та індукованим каналами різних типів (Sub – від англ. sudstrate, підкладка, решта позначень – ті самі, що на рис. 3.4.11) та полярність робочих напруг на затворі й на витоку

Контрольні питання до підрозділу 3.4

1. Опишіть принцип роботи транзистора з керувальним *p-n* переходом.
2. Як працює польовий транзистор з ізольованим затвором та вбудованим каналом?
3. Як працює польовий транзистор з ізольованим затвором та індукованим каналом?
4. Чим відрізняються характеристики різних типів польових транзисторів?
5. Чим відрізняються прохідні характеристики транзисторів з *n*-каналом та з *p*-каналом?

3.5. Біполярні транзистори: принцип роботи, характеристики та параметри

До числа найбільш поширених напівпровідникових приладів належать біполярні транзистори, що являють собою комбінацію двох *p-n* переходів. Вони використовуються в аналогових схемах підсилювачів, автогенераторів, модуляторів, детекторів тощо, а також в імпульсній техніці в ключовому режимі (як елементи з двома стійкими станами). Біполярні транзистори є активними елементами багатьох інтегральних мікросхем.

До переваг біполярних транзисторів у порівнянні з польовими належать більша потужність вихідного сигналу й можливість отримати більший коефіцієнт підсилення.

Назва біполярного транзистора пов'язана з тим, що протікання струмів у ньому забезпечується (на відміну від уніполярних транзисторів – див. вище п. 3.4) двома типами носіїв, тобто водночас електронами й дірками.

3.5.1. Будова та принцип роботи біполярного транзистора

При зворотній полярності прикладеної напруги струм *p-n* переходу визначається неосновними носіями по обидва боки від переходу. Можливість змінювати їхню концентрацію дає змогу керувати величиною струму через закритий *p-n* перехід. Так, вводячи ззовні (інжектуючи) дірки в *n*-область (рис. 3.5.1 а), можна збільшувати цей струм. При збільшенні кількості інжектованих дірок струм насичення зростає (рис. 3.5.1 б).

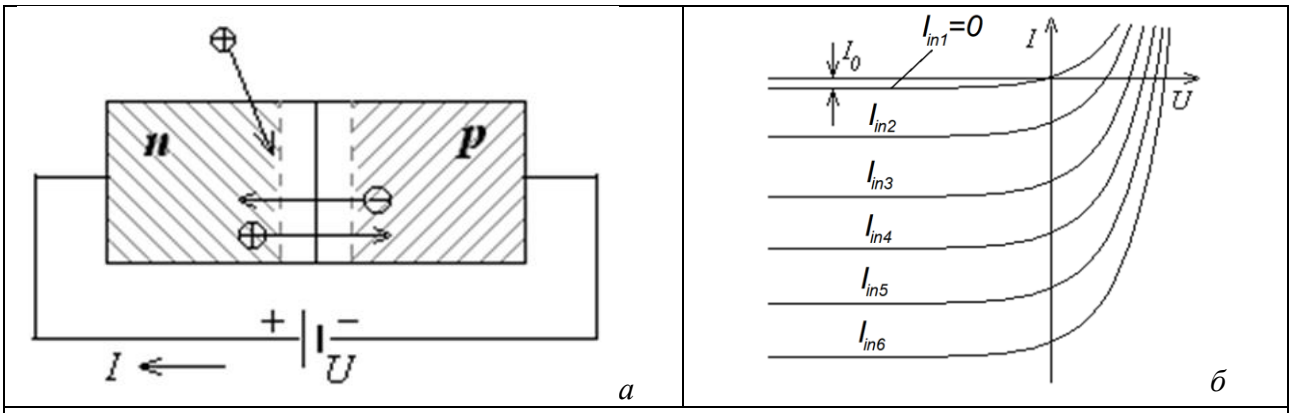


Рис. 3.5.1. До пояснення принципу роботи біполярного транзистора: а – протікання струму неосновних носіїв через закритий p-n перехід; б – графіки вольт-амперних характеристик закритого p-n переходу при додатковій інжекції неосновних носіїв в емітер (струми інжекції задовольняють умові $I_{in2}<I_{in3}<I_{in4}<I_{in5}<I_{in6}$)

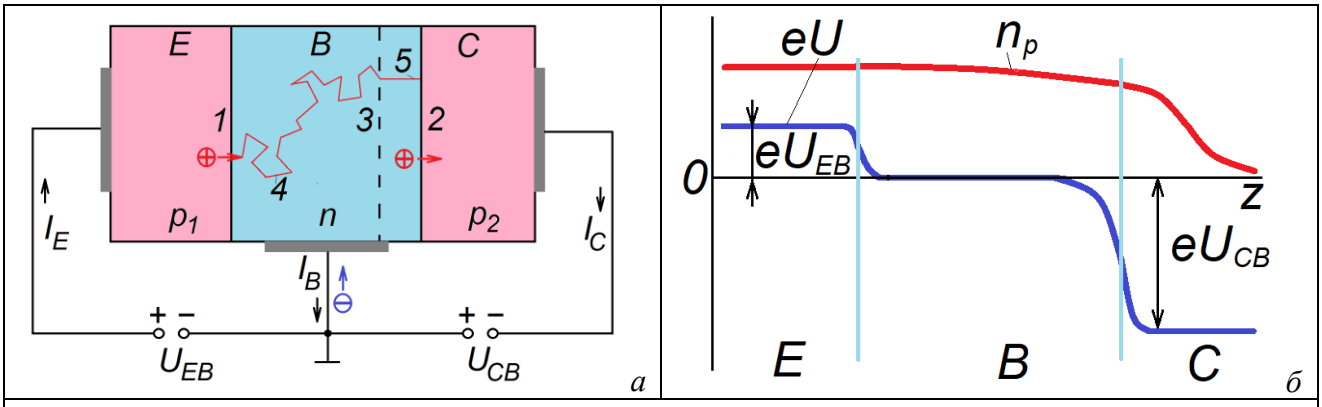


Рис. 3.5.2. Біполярний транзистор: а – схематична структура транзистора p-n-p типу (E – емітер, B – база, C – колектор), 1-2 – емітерний та колекторний p-n переходи, 3 – межа збідненого шару біля колекторного p-n переходу, 4-5 – траєкторії дифузійного та дрейфового руху дірки через базу (показані умовно); б – розподіл потенціалу та концентрації дірок уздовж транзистора. Транзистор увімкнений за схемою зі спільною базою

Таку інжекцію можна реалізувати в приладі, відомому як біполярний транзистор. Його структура схематично зображена на рис.3.5.2 а.

Біполярний транзистор складається з двох *p-n* переходів. На рис. 3.5.2 а подано схематичну структуру *p-n-p* транзистора, в якому *n*-область оточена двома *p*-областями. Полярність джерел живлення підбирається такою, щоб перехід 1 був відкритим, а перехід 2 – закритим. Таким чином, перехід 2 виступає аналогом закритого *p-n* переходу, показаного на рис. 3.5.1 а. Нехай ліва *p*-область легована значно сильніше за *n*-область. Тоді остання є базою (від англ. base – основа) щодо області p_1, яку називають емітером (від англ. emit – випромінювати, випускати). Область p_2 називають колектором (від англ. collect – збирати). За виконання всіх названих умов колекторний струм у біполярному транзисторі створюється переважно дірками.

Дірки, що інжектуються через відкритий емітерний *p-n* перехід 1, у базі є неосновними носіями, і їхня концентрація залежить від емітерного струму I_E. Отже, змінюючи цей струм, можна керувати концентрацією неосновних носіїв в базі. Дірки, що потрапили в базу, дифундують у напрямі до колекторного переходу і, досягнувши його, екстрагуються в закритий колекторний *p-n* перехід 2, формуючи колекторний струм I_C (рис. 3.5.2 а-б).

Зазначимо, що сім'я характеристик, показана на рис. 3.5.1 б, може бути інтерпретована як залежність колекторного струму I_C від напруги U_{CE} на колекторному *p-n* переході. В ролі параметра сім'ї – струму інжекції – виступає емітерний струм I_E.

Таким чином, колекторний струм I_C складається з двох компонент. Некерована компонента I_{CB0} створюється неосновними носіями колектора і завжди тече через закритий *p-n* перехід, в тому числі й за відсутності емітерного струму. В позначенні I_{CB0} індекс С вказує на колекторний струм, індекс В – на увімкнення транзистора зі спільною базою, індекс 0 – на те, що йдеться про зворотний струм у режимі насичення. Керована компонента пропорційна струму емітера. Отже,

$$I_C = I_{CB0} + \alpha I_E . \qquad (3.5.1)$$

Оскільки потік дірок, що екстрагуються до колектора з бази, не може перевищувати потік дірок, що інжектуються до неї з емітера, то завжди виконується умова $\alpha < 1$. Для ефективного керування колекторним струмом бажано, щоб коефіцієнт α був якомога ближчим до одиниці.

Під час руху через базу дірка може зустрітися з вільним електроном і рекомбінувати з ним. Дефіцит електронів, що при цьому виникає, поповнюється за рахунок базового струму. Оскільки траєкторія електронів на ділянці дрейфового руху має складну форму і, відповідно, велику довжину, то рекомбінації відбуваються саме під час дрейфового руху (рис. 3.5.2 а). Для зменшення втрати дірок на шляху від емітера до колектора через базу (тобто для збільшення коефіцієнта α) потрібно, щоб середній час їхньої дифузії через базу був значно менший за середній час їхньої рекомбінації. Досягти цього можна зменшенням товщини бази (рис. 3.5.2 а) та зниженням у ній концентрації основних носіїв (електронів), тобто зменшенням її легування. У сучасних транзисторах товщина бази становить частки мікрона, а легування забезпечує електронну провідність, яка лише трохи перевищує власну. В результаті лише мала частка дірок, які дифундують крізь базу, зазнає рекомбінації. Відповідно малим виявляється й базовий струм I_B, який визначається з першого правила Кірхгофа:

$$I_B = I_E - I_C . \qquad (3.5.2)$$

Завдяки слабкому легуванню бази і сильному легуванню емітера струм через емітерний перехід переважно формується дірками, які інжектуються з емітера, а не електронами, які екстрагуються з бази. Справді, електронна компонента емітерного струму не бере участі в створенні колекторного струму, а лише зменшує величину α.

Завдяки виконанню викладених умов значення α у більшості сучасних транзисторів лежать у межах від 0.9 до 0.995.

Як уже вказувалося, в біполярному транзисторі струми створюються обома типами носіїв – і електронами, й дірками. Так, у розглянутому вище *p-n-p* транзисторі в активному

режимі емітерний та колекторний струми утворюються переважно дірками, а базовий – електронами. Формула (3.5.1) наочно показує, що, на відміну від польового транзистора, вихідний (колекторний) струм біполярного транзистора зі спільною базою безпосередньо керується вхідним (емітерним) струмом.

Аналогічно до *p-n-p* транзистора працює й *n-p-n* транзистор, у якому колекторний струм створюється переважно електронами (рис. 3.5.3).

Умовні позначення біполярних транзисторів показані на рис.3.5.4, а схеми їхнього увімкнення – на рис. 3.5.5. Надалі ми, як правило, розглядатимемо транзистори, увімкнені за схемою зі спільним емітером. Одна з причин цього – це нездатність транзистора, увімкненого зі спільною базою, підсилювати струм. Справді, з формули (3.5.1) випливає, що відношення колекторного струму до емітерного (якщо знехтувати некерованою компонентою) дорівнює α, а ця величина завжди менша від одиниці.

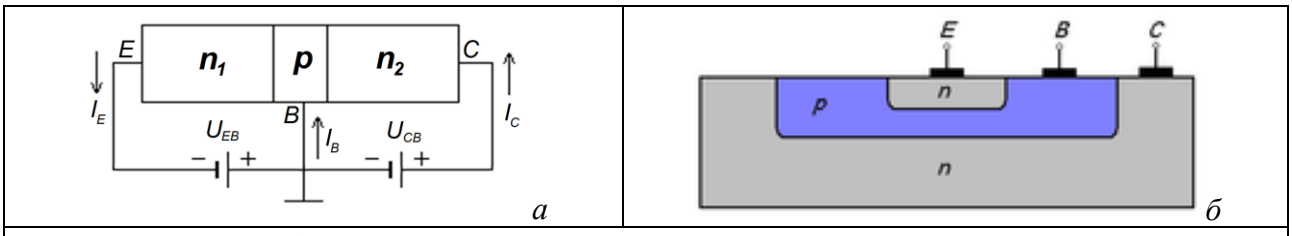


Рис. 3.5.3. Будова біполярного транзистора типу n-p-n: а – схема транзистора, увімкненого зі спільною базою (показані напруги й струми); б – переріз реального транзистора

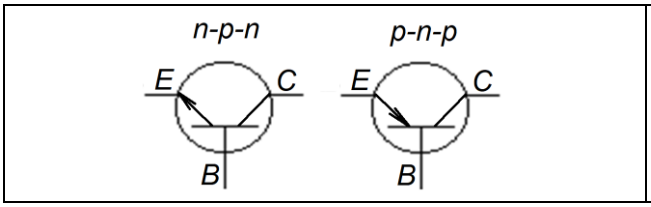


Рис. 3.5.4. Умовні позначення біполярних транзисторів

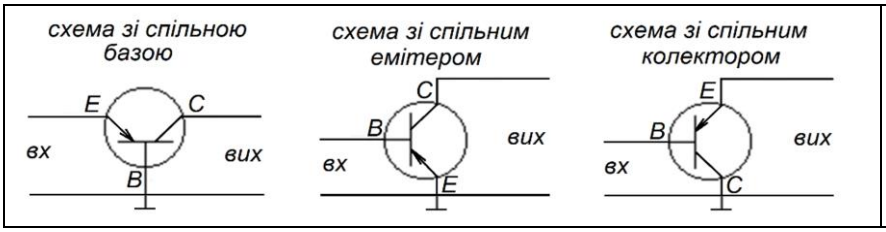


Рис. 3.5.5. Схеми увімкнення біполярних p-n-p транзисторів

В той же час на вході транзистора зі спільною базою увімкнений *p-n* перехід у прямій полярності, а на виході – у зворотній. Тому вихідний опір транзистора при такому увімкненні виявляється значно (наприклад, на два порядки) більшим від вхідного. Ця обставина дає змогу підсилювати напругу за допомогою транзистора, увімкненого зі спільною базою.

3.5.2. Характеристики біполярного транзистора, увімкненого за схемою зі спільним емітером

Як видно з рис. 3.5.5-3.5.6, при увімкненні транзистора за схемою зі спільним емітером в ролі вхідних напруги та струму виступають U_{BE} та I_B, в ролі вихідних – відповідно U_{CE} та I_C.

Знайдемо спершу співвідношення між вихідним та вхідним струмом. Тоді з рівнянь (3.5.1)-(3.5.2)

$$I_C = I_{CB0} + \alpha I_E, \qquad I_B = I_E - I_C$$

маємо:

$$I_E = \frac{I_C - I_{CB0}}{\alpha}, \tag{3.5.3}$$

та, виключаючи з (3.5.2) емітерний струм за допомогою рівняння (3.5.3),

$$I_B = I_E - I_C = \frac{I_C - I_{CB0}}{\alpha} - I_C = \frac{1-\alpha}{\alpha} I_C - \frac{I_{CB0}}{\alpha}. \qquad (3.5.4)$$

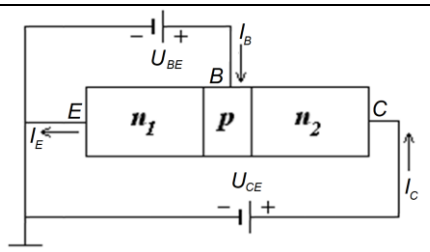


Рис. 3.5.6. Живлення n-p-n транзистора зі спільним емітером в активному режимі

Звідси випливає, що

$$I_C = \left(I_B + \frac{I_{CB0}}{\alpha} \right) \frac{\alpha}{1-\alpha} = \frac{\alpha}{1-\alpha} I_B + \frac{I_{CB0}}{1-\alpha} = \beta I_B + I_{CE0}, \qquad (3.5.5)$$

де

$$\beta = \frac{\alpha}{1-\alpha} \gg 1, \qquad (3.5.6)$$

оскільки $0 < 1-\alpha \ll 1$ (див. вище п.3.5.1),

$$I_{CE0} = \frac{I_{CB0}}{1-\alpha} \approx \beta I_{CB0}. \qquad (3.5.7)$$

Перший доданок у правій частині формули (3.5.5) – це керована складова колекторного струму транзистора, увімкненого зі спільним емітером, другий – некерована складова цього струму. Оскільки $\alpha < 1$, а $\beta \gg 1$ (пор. формули (3.5.1) та (3.5.5)), то суттєва перевага транзистора, увімкненого за схемою зі спільним емітером, порівняно з увімкненням зі спільною базою, полягає в можливості підсилення струму.

Як випливає з формули (3.5.7), некерована складова значно більша, ніж при увімкненні транзистора зі спільною базою. Дамо пояснення цьому результату на прикладі *n-p-n* транзистора, увімкнення якого в активному режимі показане на рис. 3.5.3 а. В схемі зі спільною базою струм I_{CB0} відповідно до формули (3.5.1) визначається за умови $I_E = 0$ (рис. 3.5.7 а). При цьому напруга U_{CB} прикладена лише до колекторного *p-n* переходу, і струм I_{CB0} відповідає струму цього переходу в режимі запирання. У схемі зі спільним емітером струм I_{CE0} відповідно до формули (3.5.1) визначається за умови $I_B = 0$ (рис. 3.5.6 б). Тепер напруга U_{CE} ділиться між емітерним та колекторним *p-n* переходами пропорційно їхньому омічному опору, тобто до емітерного *p-n* переходу прикладена невелика пряма напруга. В результаті емітерний струм відмінний від нуля, що й приводить до зростання колекторного струму порівняно з попереднім випадком.

Графіки вхідних та вихідних характеристик *n-p-n* транзистора зі спільним емітером показані на рис. 3.5.8.

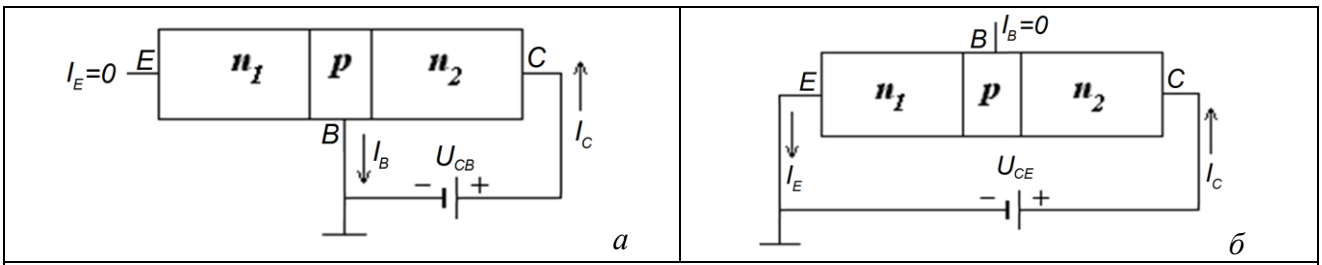


Рис. 3.5.7. Увімкнення транзистора з нульовим вхідним струмом за схемою зі спільною базою (а) та зі спільним емітером (б)

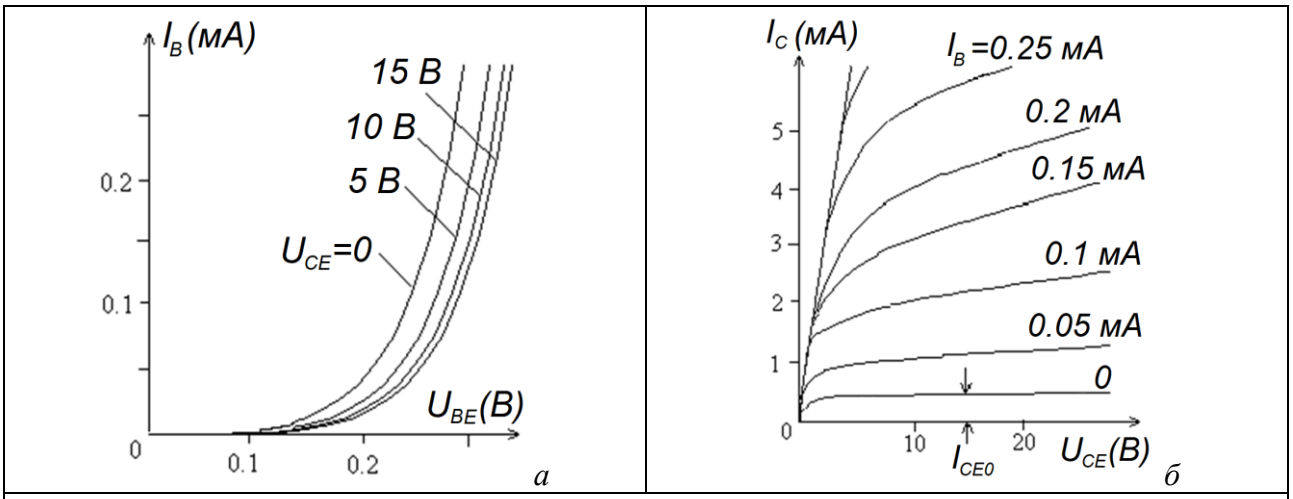


Рис. 3.5.8. Типові графіки вхідних (а) та вихідних (б) характеристик n-p-n транзистора, увімкненого за схемою зі спільним емітером

Параметром сім'ї вхідних характеристик транзистора зі спільним емітером (рис. 3.5.8 б) служить колекторна напруга. При нульовій колекторній напрузі відповідна залежність – це, по суті, пряма гілка вольт-амперної характеристики відкритого *p-n* переходу. Як і пряма гілка вольт-амперної характеристики діода (див. рис. 3.3.1 а), вона починається з деякої ненульової напруги U_{BE}^{*}, яка становить 0.1-0.2 В для германієвих та 0.5-0.7 В для кремнієвих транзисторів. Базова напруга, при якій з'являється помітний базовий струм, практично не залежить від напруги на колекторі.

Зростання колекторної напруги супроводжується зменшенням базового струму, тобто вхідні характеристики зсуваються праворуч. Це пояснюється так званим ефектом модуляції товщини бази. Він полягає в тому, що при зростанні колекторної напруги зростає і її падіння на запертому колекторному *p-n* переході. При цьому збіднений шар з боку слабко легованої бази помітно розширюється, відповідно довжина ділянки, на якій електронний струм спричинений дифузією, зменшується. Оскільки рекомбінація неосновних носіїв у базі відбувається переважно під час їхньої дифузії (швидкість їхнього спрямованого руху в полі запертого колекторного *p-n* переходу різко зростає, див. рис. 3.5.2 а), то в результаті зменшення довжини ділянки дифузії зменшується кількість рекомбінацій, що й приводить до зменшення базового струму. З графіків вхідних характеристик (рис. 3.5.8 а) видно, що вхідний диференціальний опір зменшується зі зростанням базового струму, а його величина має порядок 10^2 - 10^3 Ом.

З графіків на рис. 3.5.8 а видно, що зсув вхідних характеристик праворуч відбувається переважно при невеликих значеннях U_{CE}. Тому в довідниках звичайно наводяться лише дві вхідні характеристики: одна для $U_{CE} = 0$ і друга – для великого значення U_{CE}; ця характеристика є граничною для всієї сім'ї.

Перейдемо до обговорення вихідних характеристик, графіки яких подані на рис. 3.5.8 б. Видно, що при малих значеннях U_{CE} всі характеристики зливаються, і струм колектора стає незалежним від струму бази (втрачається керувальна дія бази). В цьому режимі напруга на колекторі менша від напруги на базі. В результаті колекторний перехід відкривається, і робочий режим протікання струму в транзисторі, описаний раніше, порушується. Ця ділянка вихідних характеристик, по суті, відповідає початковій ділянці вольт-амперної характеристики відкритого *p-n* переходу. Відповідний режим роботи називають *режимом насичення*. Він характеризується малим вихідним опором транзистора (одиниці Ом і менше).

Вихідна характеристика, для якої $I_B = 0$, відповідає так званому *режиму запирання* транзистора, який обговорювався раніше.

В робочому діапазоні колекторної напруги й базового струму (тобто між областями насичення та запирання) колекторний струм переважно визначається базовим струмом, а

опосередковано – емітерним струмом (див. вище п. 3.5.1), зі збільшенням якого збільшується й струм бази. Однак вихідні характеристики мають помітний нахил, який зростає із збільшенням I_B. З графіків вхідних характеристик видно, що для підтримання фіксованого значення базового струму при зростанні колекторної напруги необхідно збільшувати базову напругу, прикладену до емітерного *p-n* переходу. В результаті емітерний струм зростає, й відповідно зростає колекторний струм. Таким чином, вихідний диференціальний опір $R_{d\,out} = dU_{CE}/dI_C$ виявляється порівняно невеликим і зменшується зі збільшенням I_B. Звичайно $R_{d\,out}$ для транзисторів, увімкнених зі спільним емітером, становить 1-10 кОм.

У підсумку вихідний і, особливо, вхідний диференціальний опір біполярного транзистора помітно нижчі, ніж у польових транзисторів (див. вище п. 3.4). Як буде видно далі, це недолік біполярних транзисторів порівняно з польовими.

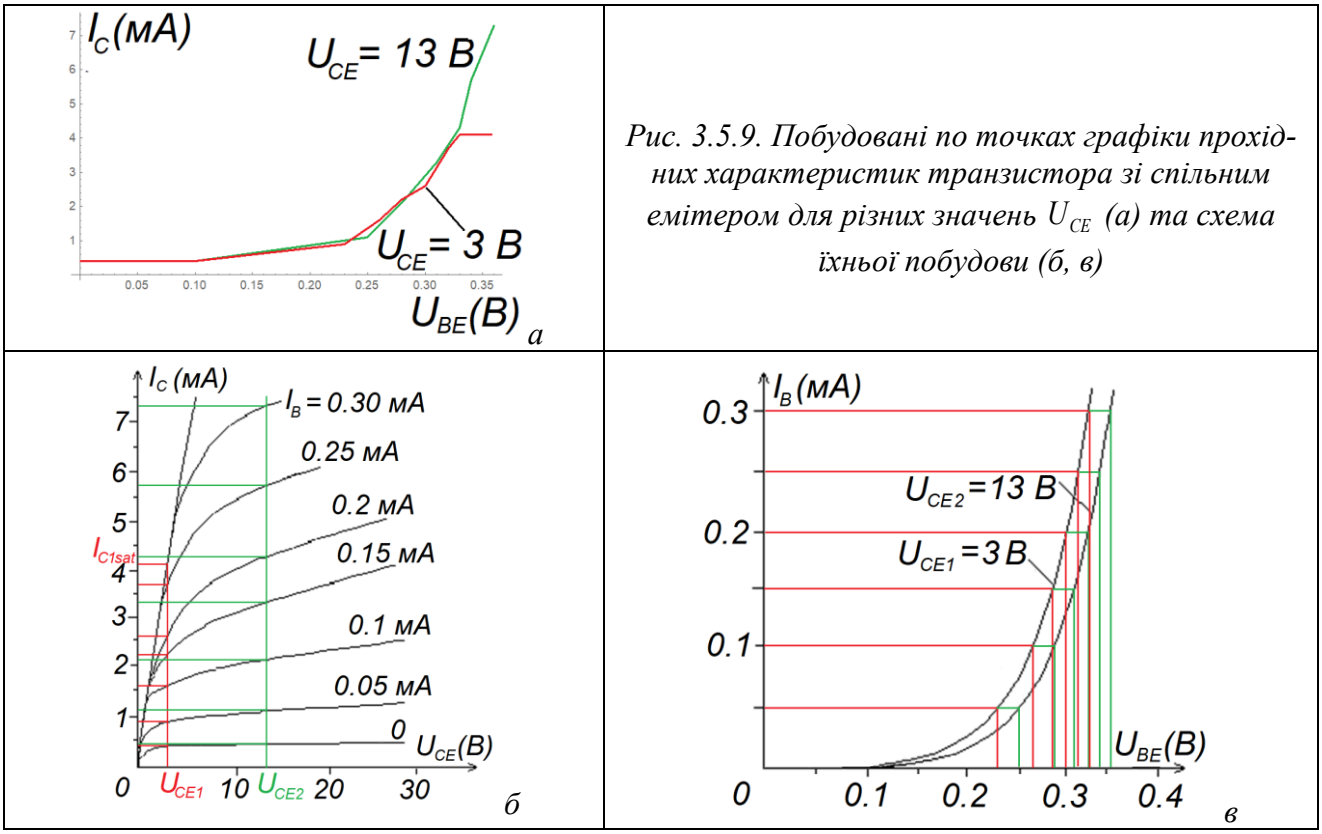


Рис. 3.5.9. Побудовані по точках графіки прохідних характеристик транзистора зі спільним емітером для різних значень U_{CE} (а) та схема їхньої побудови (б, в)

Таблиця 3.5.1

I_B, мА	0	0.05	0.10	0.15	0.20	0.25	0.3
***I_C, мА при U_{CE} = 3* В**	0.4	0.9	1.6	2.2	2.6	3.7	4.1
***U_{BE}, В при U_{CE} = 3* В**	0.10	0.23	0.26	0.28	0.30	0.32	0.33
***I_C, мА при U_{CE} = 13* В**	0.4	1.1	2.1	3.3	4.3	5.7	7.3
***U_{BE}, В при U_{CE} = 13* В**	0.10	0.25	0.28	0.31	0.33	0.34	0.36

При $I_B = 0$ транзистор не закривається повністю (тобто колекторний струм залишається відмінним від нуля). Некерована компонента колекторного струму I_{CE0} залишається порівняно великою (див. вище формулу (3.5.7)).

Отже, вихід біполярного транзистора, увімкненого зі спільним емітером, також (як і польовий транзистор) може розглядатись як керований диференціальний опір. Але тепер керування забезпечується не вхідною напругою, а вхідним струмом (струмом бази).

Для аналізу роботи і розрахунку радіоелектронних схем часто буває корисною статична прохідна характеристика, яку можна визначити як залежність вихідного (колекторного)

струму від вхідної (базової) напруги при заданій колекторній напрузі (рис. 3.5.9 а). Вона будується на основі вхідних та вихідних характеристик шляхом виключення базового струму. Спочатку за допомогою вихідних характеристик знаходиться залежність I_C від I_B при фіксованому значенні U_{CE} (див. рис. 3.5.9 б, пор. з рис. 3.4.2 г), а потім за допомогою вхідної характеристики для даного U_{CE} знаходяться значення U_{BE}, відповідні потрібним значенням I_B (рис. 3.5.9 в). Відповідні числові значення наведені в табл. 3.5.1. З їхнім використанням і побудовані графіки, подані на рис. 3.5.9 а.

Якщо при $I_B = 0$ прикладена напруга U_{CE} достатня для досягнення колекторного струму насичення I_{CE0}, то колекторний струм зберігає цю величину в діапазоні напруг U_{BE} від нуля до значення, при якому починається помітне зростання базового струму. Починаючи з цієї точки, струм I_C зростає майже експоненціально. При невеликих напругах U_{CE} зростаюча ділянка прохідної характеристики змінюється на горизонтальну, що відповідає режиму насичення транзистора.

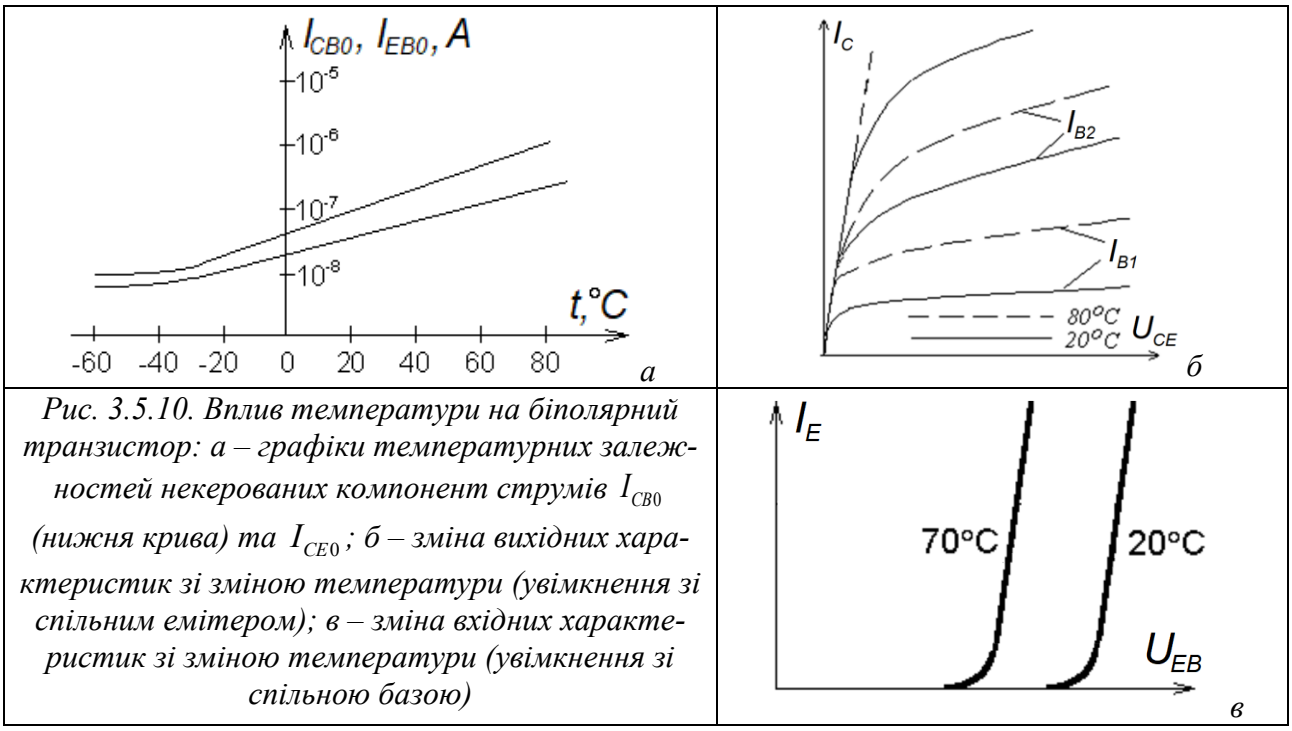


Рис. 3.5.10. Вплив температури на біполярний транзистор: а – графіки температурних залежностей некерованих компонент струмів I_{CB0} (нижня крива) та I_{CE0}; б – зміна вихідних характеристик зі зміною температури (увімкнення зі спільним емітером); в – зміна вхідних характеристик зі зміною температури (увімкнення зі спільною базою)

3.5.3. Вплив температури на характеристики транзистора

Зворотний струм через закритий *p-n* перехід дуже сильно залежить від температури і подвоюється через кожні 5 – 10°С. В результаті доданок I_{CB0} у формулі (3.5.1), що при кімнатній температурі є малим, при високих температурах може стати сумірним з керованою компонентою αI_E. При цьому вся сім'я вихідних характеристик суттєво зсувається вгору. До такого самого зсуву може призвести й зростання α з підвищенням температури, зумовлене відповідним зростанням швидкості дифузії.

Для транзистора, увімкненого за схемою зі спільним емітером, ці ефекти виражені сильніше, оскільки тут некерований струм колектора I_{CE0} у β разів більший за I_{CB0} – некерований струм колектора при увімкненні транзистора зі спільною базою (рис. 3.5.10 а).

Підвищення температури на кілька десятків градусів може істотно змінити вигляд сім'ї вихідних характеристик транзистора зі спільним емітером (рис. 3.5.10 б). Зі зростанням температури зменшується напруга, з якої починаються вхідні характеристики, так що вони зсуваються ліворуч (рис. 3.5.10 в).

Зсуви вхідних та вихідних характеристик викликають істотні зміни режимів роботи

напівпровідникових приладів в радіоелектронних схемах. Це обмежує можливість застосування біполярних транзисторів в умовах, коли температура навколишнього середовища може значно змінюватися.

3.5.4. Біполярний транзистор як лінійний чотириполюсник

При ввімкненні транзистора в схему підсилювача його вхідна напруга u_{in} складається з постійної складової U_{in}, що визначає режим роботи транзистора, та змінної напруги $u_{in}(t)$, що виступає в ролі сигналу, який потрібно підсилити (пор. з пп. 2.1.1, 3.2.2):

$$u_{in} = U_{in} + u_{in}(t). \qquad (3.5.6)$$

Аналогічний вигляд мають і всі інші струми та напруги:

$$u_{out} = U_{out} + u_{out}(t); \quad i_{in} = I_{in} + i_{in}(t); \quad i_{out} = I_{out} + i_{out}(t). \qquad (3.5.7)$$

Вважатимемо для спрощення (хоч це й не завжди відповідає дійсності), що змінні складові струмів і напруг значно менші відповідних постійних складових (наближення малого сигналу). Тоді будь-яку нелінійну функцію струмів або напруг (зокрема, функції, які відповідають тим чи іншим характеристикам транзистора) можна розкласти в ряд Тейлора за їхньою змінною складовою, обмежившись лінійним доданком (пор. з п. 2.1.1, див. формулу (2.1.1)). У такому наближенні режимні (постійні) й сигнальні (змінні) складові безпосередньо не пов'язані між собою і розглядати їх можна як незалежні набори струмів і напруг. На цьому ґрунтується метод аналізу роботи електронних схем у режимі підсилення малих сигналів.

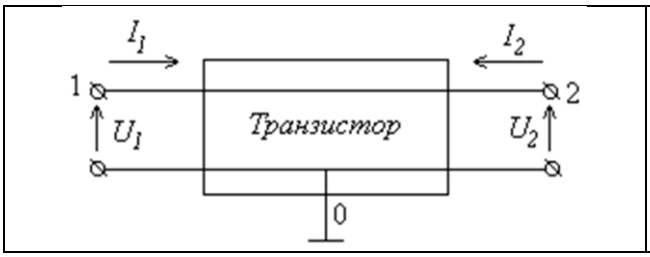


Рис. 3.5.11. Представлення транзистора як лінійного чотириполюсника

Конкретизуємо висловлені міркування. Зобразимо транзистор у вигляді чотириполюсника (рис. 3.5.11). Клема 1 тут є входом транзистора (це може бути база або емітер, залежно від способу його увімкнення), клема 2 – вихід транзистора (колектор), а клема 0 – його заземлений електрод. З набору величин I_1, I_2, U_1, U_2 дві ми вважаємо впливами, а дві інші – відгуками, так що кожен з відгуків однозначно визначається впливами. Так, наприклад, при увімкненні біполярного транзистора за схемою зі спільним емітером напруга на базі U_{BE} та колекторний струм I_C однозначно визначається вхідним струмом I_B та вихідною напругою U_{CE}:

$$U_{BE} = U_{BE}(I_B, U_{CE}), \qquad I_C = I_C(I_B, U_{CE}). \qquad (3.5.8)$$

Перша з цих залежностей визначається виглядом вхідних характеристик транзистора, друга – виглядом вихідних характеристик (див. рис. 3.5.8).

Подамо вхідний струм I_B та вихідну напругу U_{CE} як суму постійної (режимної) та малої змінної (сигнальної) складових:

$$I_B = I_{B0} + i_B(t); \quad U_{CE} = U_{CE0} + u_{CE}(t), \qquad |i_B(t)| \ll I_{B0}; \quad |u_{CE}(t)| \ll U_{CE0}. \qquad (3.5.9)$$

Розклавши функції $U_{BE} = U_{BE}(I_B, U_{CE})$, $I_C = I_C(I_B, U_{CE})$ в подвійний ряд Тейлора за малими змінними складовим i_B та u_{CE}, дістанемо:

$$U_{BE} = U_{BE}(I_{B0}, U_{CE0}) + \left.\frac{\partial U_{BE}}{\partial I_B}\right|_{U_{CE}=U_{CE0}} i_B(t) + \left.\frac{\partial U_{BE}}{\partial U_{CE}}\right|_{I_B=I_{B0}} u_{CE}(t), \qquad (3.5.10)$$

$$I_C = I_C\left(I_{B0}, U_{CE0}\right) + \left.\frac{\partial I_C}{\partial I_B}\right|_{U_{CE}=U_{CE0}} i_B\left(t\right) + \left.\frac{\partial I_C}{\partial U_{CE}}\right|_{I_B=I_{B0}} u_{CE}\left(t\right). \qquad (3.5.11)$$

Ми обірвали ряд на доданках першого порядку, оскільки через їхню мализну наступні доданки в ряду значно менші від них.

У свою чергу подамо ліві частини рівнянь як суму режимної та малої сигнальної складових. Тоді в рівняннях (3.5.10)-(3.5.11) можна окремо прирівняти між собою доданки нульового та першого порядків мализни. Для доданків першого порядку маємо:

$$\begin{cases} u_{BE}\left(t\right) = h_{11E} i_B\left(t\right) + h_{12E} u_{CE}\left(t\right); \\ i_C\left(t\right) = h_{21E} i_B\left(t\right) + h_{22E} u_{CE}\left(t\right), \end{cases} \qquad (3.5.12)$$

де використано позначення

$$h_{11E} = \left.\frac{\partial U_{BE}}{\partial I_B}\right|_{U_{CE}=U_{CE0}}, \; h_{12E} = \left.\frac{\partial U_{BE}}{\partial U_{CE}}\right|_{I_B=I_{B0}}, \; h_{21E} = \left.\frac{\partial I_C}{\partial I_B}\right|_{U_{CE}=U_{CE0}}, \; h_{22E} = \left.\frac{\partial I_C}{\partial U_{CE}}\right|_{I_B=I_{B0}}. \qquad (3.5.13)$$

Система (3.5.12) називається рівняннями транзистора, увімкненого за схемою зі спільним емітером, в h-параметрах, а співвідношення (3.5.13) визначають h-параметри транзистора при такому увімкненні (пор. з п. 2.1.3, див. формулу (2.1.10)).

Еквівалентна схема біполярного транзистора, що відповідає системі рівнянь (3.5.12), показана на рис. 3.5.12 (пор. з рис. 2.2.2). Вона складніша від аналогічної схеми польового транзистора (див. вище рис. 3.4.3 а).

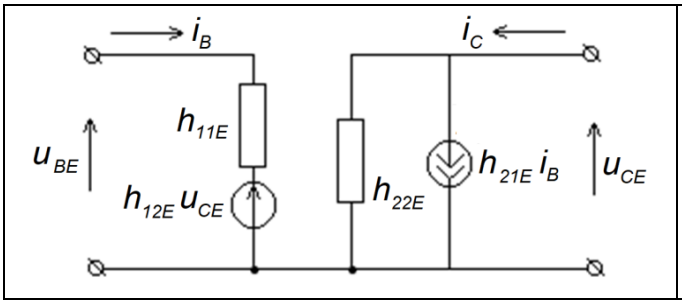


Рис. 3.5.12. Еквівалентна схема транзистора, увімкненого зі спільним емітером, в h-парметрах

h-параметрам можна приписати певний фізичний зміст. Так, h_{11E} і h_{21E} – це відповідно вхідний диференціальний опір транзистора і його коефіцієнт передачі струму при короткому замкненні за сигналом на виході, h_{22E} і h_{12E} – відповідно вихідна диференціальна провідність за відсутності вхідного змінного струму (тобто при розімкненому вході) і коефіцієнт внутрішнього зворотного зв'язку. Внутрішній зворотний зв'язок зумовлений зсувом вхідних характеристик праворуч (тобто зменшенням базового струму) при зростанні колекторної напруги. Як уже відзначалося, це пов'язано з ефектом модуляції товщини бази. Зазначимо, що в польових транзисторах внутрішній зворотний зв'язок відсутній.

Очевидно, ща параметри транзистора залежать від його типу, положення робочої точки на вхідних та вихідних характеристиках, температури тощо.

Ми ввели h-параметри для транзистора увімкненого зі спільним емітером. Очевидно, це можна зробити й для інших можливих схем увімкнення транзистора. h-параметри для однієї схеми увімкнення можна виразити через h-параметри для іншої схеми увімкнення.

Таким чином, зробивши припущення про відносну мализну змінних складових струмів і напруг, ми представили транзистор, що є нелінійним елементом, як лінійний чотириполюсник, що описується системою рівнянь (3.5.12).

Обравши за задані величини інші пари змінних струмів чи напруг, можна описати транзистор іншими наборами параметрів. Але ми надалі для опису режиму малого сигналу користуватимемося саме h-параметрами.

Контрольні питання до підрозділу 3.5

1. Опишіть принцип роботи біполярного транзистора.
2. Чим відрізняються характеристики *p-n-p* і *n-p-n* транзисторів?
3. Чим визначається струм бази біполярного транзистора?
4. Чому біполярні транзистори найчастіше вмикаються за схемою зі спільним емітером?
5. Чому властивості біполярного транзистора суттєво залежать від температури?
6. Біполярний транзистор – принципово нелінійний елемент. Як його можна описувати моделлю лінійного чотириполюсника?

Приклади розв'язання задач до підрозділу 3.5

П 3.5.1. Вхідні та вихідні характеристики транзистора зі спільним емітером подані на рис. 3.5.13. Напруга живлення – 20 В, опір навантаження – 4 кОм. Побудувати навантажувальну пряму. Обрати робочу точку на середині її робочої ділянки і записати її координати (вхідні й вихідні напруги та струми).

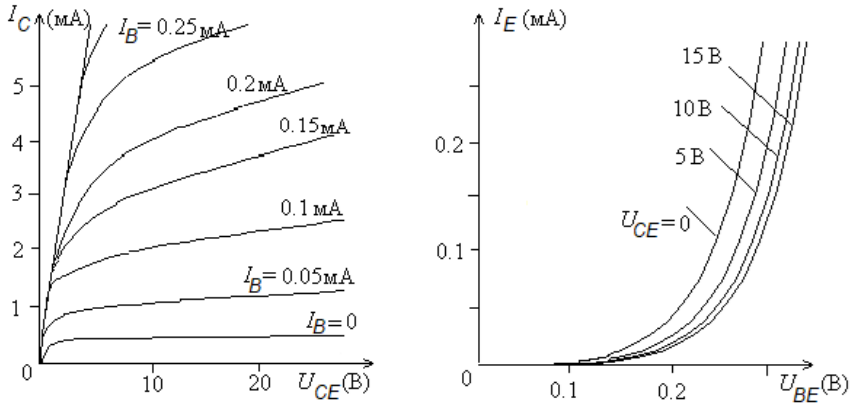


Рис. 3.5.13.

Розв'язок

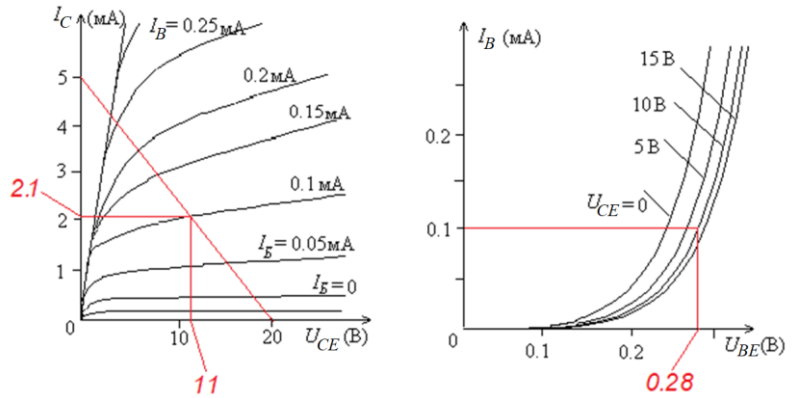


Рис. 3.5.14. Побудова навантажувальної прямої, вибір робочої точки та знаходження її координат

Побудуємо навантажувальну пряму $E = I_C R_l + U_{CE}$. Вона перетинає осі координат у точках $U_{CE} = E$ та $I_C = E/R_l$ (рис. 3.5.14 а). Координати робочої точки (посередині між точками запирання та насичення): $I = 0.1$ мА, $I_C = 2.1$ мА, $U_{BE} = 0.28$ В, $U_{CE} = 11$ В.

П 3.5.2. Скориставшись даними попередньої задачі, знайти h-параметри транзистора в околі обраної робочої точки.

Розв'язок

З графіків вихідних характеристик (рис. 3.5.15) знаходимо параметри h_{21E} та h_{22E}:

$$h_{21E} = \left.\frac{\Delta I_C}{\Delta I_B}\right|_{u_{CE}=11B} = \frac{3.2-1.1}{0.15-0.05} = \frac{2.1}{0.1} = 21;$$

$$h_{22E} = \left.\frac{\Delta i_C}{\Delta u_{CE}}\right|_{i_B=0.1мА} = \frac{2.2-1.7}{15-5} = \frac{0.5}{10} = 0.05\,\text{мСм}, \qquad \frac{1}{h_{22E}} = 20\,\text{кОм}.$$

З графіків вхідних характеристик знаходимо параметри h_{11E} та h_{12E}:

$$h_{11E} = \left.\frac{\Delta u_{BE}}{\Delta i_B}\right|_{u_{CE}=11В} = \frac{0.285-0.26}{0.125-0.075} = \frac{0.025}{0.05} = 0.5\,\text{кОм};$$

$$h_{12E} = \left.\frac{\Delta u_{BE}}{\Delta u_{CE}}\right|_{i_B=0.1мА} = \frac{0.29-0.26}{15-5} = \frac{0.03}{10} = 3\cdot 10^{-3}.$$

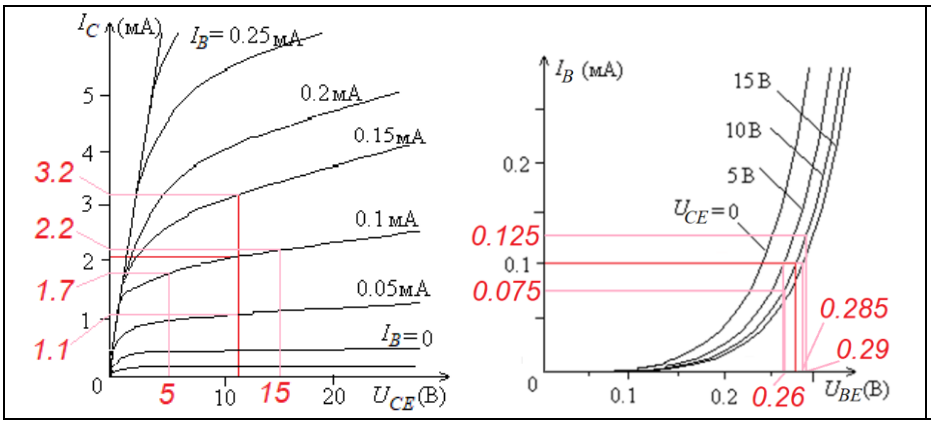


Рис. 3.5.15. До визначення h-параметрів

Задачі до підрозділу 3.5

3.5.1. Для деякого біполярного транзистора $\alpha = 0.98$. Знайти для нього параметр β.

3.5.2. Побудувати навантажувальну пряму, яка забезпечить максимальні струм та напругу на виході транзистора, характеристики якого подані на рис. 3.5.17. Обрати робочу точку посередині та визначити в її околі h-параметри.

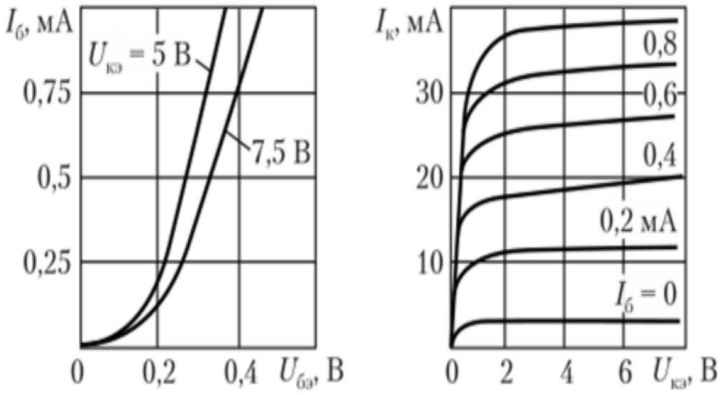


Рис. 3.5.16.

3.6. Частотні властивості біполярного транзистора

Важливою характеристикою будь-якого підсилювача є його смуга пропускання – область частот, в якій він забезпечує підсилення сигналу.

Верхня межа смуги пропускання підсилювача на біполярному транзисторі значною мірою визначається частотними властивостями такого транзистора. Тому ці властивості вимагають детальнішого обговорення.

3.6.1. Перехідні процеси в транзисторі

Як указувалося вище, колекторний струм біполярного транзистора створюється неосновними носіями, інжектованими в базу через емітерний *p-n* перехід. В результаті дифузії через базу ці носії доходять до колекторного *p-n* переходу. На дифузію через базу потрібен деякий час. Тому при стрибкоподібному збільшенні напруги на емітерному *n-p* переході і відповідному збільшенні емітерного струму *p-n-p* транзистора (рис. 3.6.1 а-б) колекторний струм зростає не одразу, а з деяким запізненням на час τ_1, протягом якого перші дірки доходять до колектора (рис. 3.6.1 в). Потім струм I_C поступово зростає, прямуючи до усталеного значення αI_E. Середній час дифузії через базу визначається звичайною для процесів дифузії формулою

$$\tau_D = \frac{d_B^2}{D}, \qquad (3.6.1)$$

де d_B – товщина бази, D – коефіцієнт дифузії неосновних носіїв в матеріалі бази. Так, наприклад, для дірок у кремнії $D = 13$ см^2/с, отже при $d_B = 1$ мкм середній час дифузії дорівнює $8 \cdot 10^{-10}$ с. Саме за такий час струм колектора наближається до насичення. Час затримки τ_1 (рис. 3.6.1 в) звичайно набагато менший від τ_D, тому ним можна нехтувати. Хід зростання колекторного струму добре апроксимується експоненціальним законом

$$I_C(t) \approx \alpha I_E \left[1 - \exp\left(-\frac{t}{\tau_D} \right) \right] \qquad (3.6.2)$$

(див. рис. 3.6.1 в).

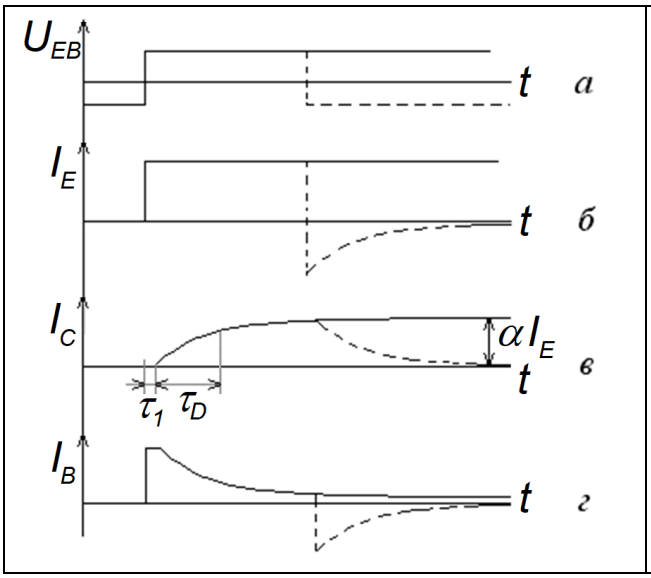


Рис. 3.6.1. Часові епюри струмів біполярного транзистора при стрибкоподібній зміні напруги на емітерному p-n переході

При стрибкоподібному запиранні емітерного *p-n* переходу (штрихові лінії на рис. 3.6.1) колекторний струм спадає не стрибком, а плавно, в міру розсмоктування (за рахунок дифузії та подальшої екстракції) надлишкових неосновних носіїв, які накопичилися в базі. При цьому дірки дифундують не тільки в бік колектора, але й у бік емітера, до якого тепер також прикладена запірна напруга. Тому в момент запирання емітерний струм змінює

свій напрям і далі поступово зменшується. Спадання емітерного та колекторного струмів також можна описати експоненціальним законом зі сталою часу τ_D.

В момент часу, коли починається інжекція неосновних носіїв у базу, базовий струм різко зростає (рис. 3.6.1 г). Адже об'ємний заряд інжектованих у базу дірок повинен компенсуватися електронами, які надходять від джерела живлення. Згідно першого правила Кірхгофа базовий струм має змінюватися за законом

$$I_B = I_E - I_C = I_E\left\{1-\alpha\left[1-\exp\left(-\frac{t}{\tau_D}\right)\right]\right\} \approx I_E \exp\left(-\frac{t}{\tau_D}\right). \qquad (3.6.3)$$

З цієї формули випливає, що в перші моменти після відкриття емітерного *p-n* переходу базовий струм сумірний з емітерним, тоді як в усталеному режимі базовий струм завжди набагато менший від емітерного. Після припинення емітерного струму об'ємний заряд неосновних носіїв у базі розсмоктується, і основні носії, котрі цей струм компенсували, тепер ідуть з бази, утворюючи експоненціальний імпульс базового струму зворотної полярності.

3.6.2. Частотна залежність коефіцієнта передачі струму біполярного транзистора

Якщо до емітерного *p-n* переходу прикладати періодичну (з періодом T, сумірним з τ_D) послідовність прямокутних імпульсів прямої та зворотної напруги, то колекторний струм кожного разу не встигатиме ані дорости до усталеного значення αI_E, ані зменшитися до нуля. В результаті часова залежність колекторного струму $i_C(t)$ набуває вигляду пилкоподібних коливань (рис. 3.6.2), розмах яких зменшується зі зменшенням відношення T/τ_D.

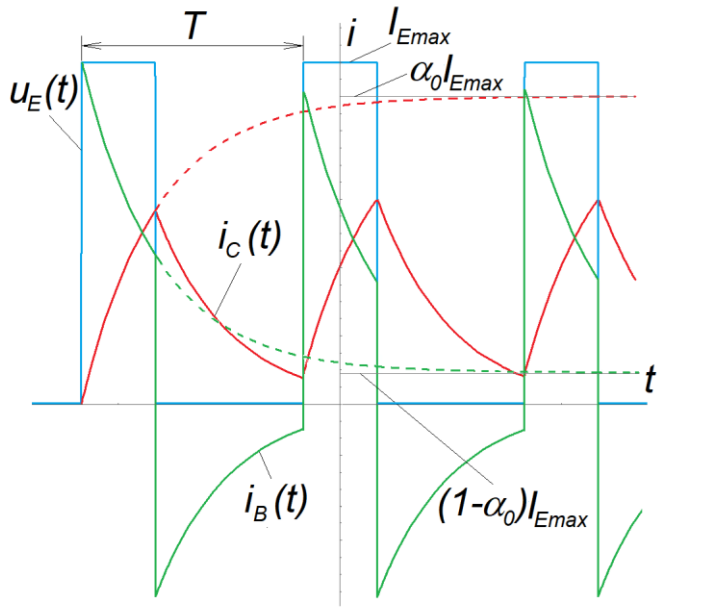


Рис. 3.6.2. Часові епюри струмів, збуджуваних у транзисторі при періодичному відкритті та закритті емітерного p-n переходу

Очевидно, з підвищенням частоти (емітерної) напруги коефіцієнт передачі струму α зменшується. Справді, залежність (3.6.2) можна розглядати як перехідну функцію $a(t)$ транзистора (з точністю до константи), який описується моделлю лінійного чотириполюсника, вхідним сигналом якого служить колекторний струм, а вихідним – емітерний:

$$a(\tau) = \alpha_0\left[1-\exp\left(-\frac{\tau}{\tau_D}\right)\right], \qquad (3.6.4)$$

де α_0 – відношення амплітуд колекторного та емітерного струму на низьких частотах.

Тоді, скориставшись формулою (2.2.57), можемо легко знайти відповідну передавальну функцію:

$$\alpha(\omega)=\int_0^\infty \frac{da(\tau)}{d\tau}\exp(-i\omega\tau)\,d\tau=\alpha_0\int_0^\infty \frac{1}{\tau_D}\exp\left(-\frac{\tau}{\tau_D}\right)\exp(-i\omega\tau)d\tau=$$

$$=\frac{\alpha_0}{\tau_D}\int_0^\infty \exp\left[\left(-i\omega-\frac{1}{\tau_D}\right)\tau\right]d\tau=\frac{\alpha_0}{\tau_D}\left.\frac{\exp\left[\left(-i\omega-1/\tau_D\right)\tau\right]}{-i\omega-1/\tau_D}\right|_0^\infty=\frac{\alpha_0}{1+i\omega\tau_D}=\frac{\alpha_0}{1+i\omega/\omega_\alpha}, \qquad (3.6.5)$$

де $\omega_\alpha=1/\tau_D$ – верхня гранична частота транзистора, увімкненого за схемою зі спільною базою. Відповідно

$$\left|\alpha(\omega)\right|=\frac{\alpha_0}{\sqrt{1+\left(\omega/\omega_\alpha\right)^2}}. \qquad (3.6.6)$$

(рис. 3.6.3). З графіка видно, що модуль коефіцієнта передачі в широкому діапазоні низьких частот майже незмінний і близький до α_0, а потім починає плавно зменшуватись, асимптотично прямуючи до нуля на високих частотах.

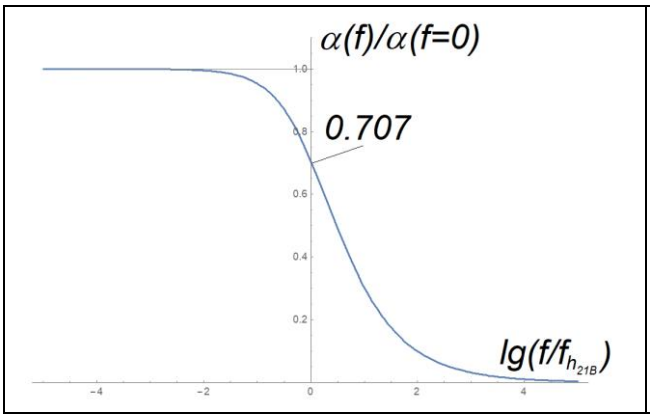


Рис. 3.6.3. Графік амплітудно-частотної характеристики біполярного транзистора зі спільною базою: вхідним сигналом вважається струм емітера, вихідним – струм колектора

Частота ω_α (відповідну їй лінійну частоту в довідниках позначають $f_{h_{21B}}=\omega_\alpha/2\pi$) характеризує зміну з частотою коефіцієнта α – коефіцієнта передачі струму транзистора, увімкненого зі спільною базою. Але, як уже вказувалося, в більшості випадків транзистори вмикаються за схемою зі спільним емітером. При такому увімкненні в ролі коефіцієнта передачі струму виступає параметр β. Визначимо його частотну залежність.

$$\beta(\omega)=\frac{\alpha(\omega)}{1-\alpha(\omega)}=\frac{\dfrac{\alpha_0}{1+i\,\omega/\omega_\alpha}}{1-\dfrac{\alpha_0}{1+i\,\omega/\omega_\alpha}}=\frac{\alpha_0}{1+i\,\omega/\omega_\alpha-\alpha_0}=\frac{\dfrac{\alpha_0}{1-\alpha_0}}{1+\dfrac{i\omega}{\omega_\alpha\left(1-\alpha_0\right)}}=\frac{\beta_0}{1+i\,\omega/\omega_\beta}, \qquad (3.6.7)$$

$$\left|\beta(\omega)\right|=\frac{\beta_0}{\sqrt{1+\left(\omega/\omega_\beta\right)^2}}$$

Тут використані позначення

$$\beta_0=\frac{\alpha_0}{1-\alpha_0}, \qquad \omega_\beta=\omega_\alpha\left(1-\alpha_0\right). \qquad (3.6.8)$$

Видно, що верхня гранична частота для коефіцієнта β значно менша, ніж для коефіцієнта α, оскільки $0<1-\alpha\ll 1$ і, отже, $\omega_\beta\ll\omega_\alpha$. У довідниках лінійна частота, що відповідає ω_β, позначається як $f_{h_{21E}}$.

Фізичну причину погіршення частотних властивостей транзистора при його увімкненні за схемою зі спільним емітером можна зрозуміти з рис. 3.6.1 г: при швидких перемиканнях (коли $T\ll\tau_D$) базовий струм виявляється сумірним з емітерним, а колекторний, як

уже відзначалося, суттєво зменшується, так що коефіцієнт β стає набагато меншим, аніж у стаціонарному режимі чи на низьких частотах.

Частота, на якій параметр β дорівнює одиниці (тобто транзистор втрачає підсилювальні властивості за струмом) позначається як f_T і має назву частоти одиничного підсилення. Покладемо у формулі (3.6.7) $\left|\beta\left(\omega_T\right)\right| = 1$. Тоді

$$\beta_0^2 = 1 + \left(\omega_T / \omega_\beta\right)^2,$$

звідки

$$\omega_T = \omega_\beta \sqrt{\beta_0^2 - 1} \approx \beta_0 \omega_\beta \approx \frac{\omega_\beta}{1 - \alpha_0} \approx \omega_\alpha . \qquad (3.6.9)$$

Окрім уже згаданих граничних частот, транзистор ще характеризують максимальною частотою генерації $f_{\max}$. На цій частоті коефіцієнт підсилення транзистора за потужністю зменшується до одиниці, і транзистор втрачає здатність до підсилення сигналів.

3.6.3. Шляхи поліпшення частотних властивостей транзисторів

За частотними властивостями всі транзистори поділяються на низькочастотні (з робочою частотою до 3 МГц), середньочастотні (до 30 МГц), високочастотні (до 300 МГц) та надвисокочастотні (з робочою частотою понад 300 МГц).

Очевидно, для поліпшення частотних властивостей біполярного транзистора треба зменшити час дифузії неосновних носіїв у базі.

Найбільш прямолінійний шлях вирішення цього завдання – це мікромініатюризація, тобто зменшення розмірів бази. Ще один спосіб – це використання *n-p-n* транзисторів, оскільки електрони мають більший коефіцієнт дифузії, ніж дірки.

Верхня межа частот надвисокочастних транзисторів з вихідною потужністю понад 1 Вт складає близько 10 ГГц. Конструктивно практично всі вони належать до так званих багатоемітерних транзисторів, тобто мають кілька емітерних областей, об’єднаних одним зовнішнім електродом[3]. Такі транзистори характеризуються великим значенням відношення периметра емітера до його площі, що забезпечує малий опір бази транзистора та високі значення густини його емітерного струму. Емітер надвисокочастотного транзистора зазвичай має структуру гребінки або сітки.

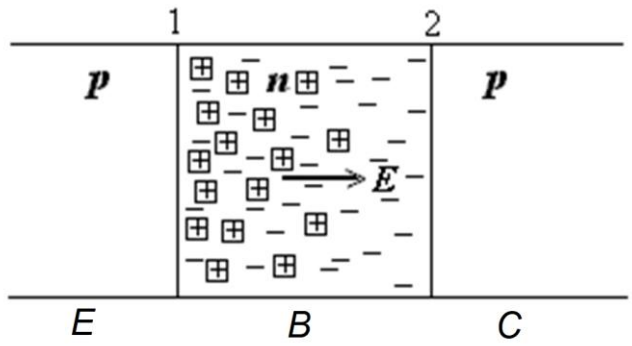


Рис. 3.6.4. До пояснення роботи дрейфового транзистора

Щоб прискорити рух неосновних носіїв (і тим самим зменшити інерційність транзистора), в об’ємі бази іноді створюють внутрішнє електричне поле шляхом її нерівномірного легування. Так, наприклад, забезпечивши в базі підвищену концентрацію донорних домішок поблизу емітерного *p-n* переходу (рис. 3.6.4; атоми донорних домішок тут зображені квадратиками з хрестиками), можна добитися того, що вільні електрони (якщо їхня температура не

[3] Існують також багатоемітерні транзистори, у яких кожна емітерна область має окремий зовнішній вивід. Вони використовуються у транзисторно-транзисторній логіці як частина логічного елементу «І».

дуже мала) за рахунок дифузії більш-менш рівномірно розподіляються по об'єму бази, тоді як позитивно заряджені домішкові атоми, закріплені в кристалічній ґратці, створюють біля емітера об'ємний заряд. Цей заряд породжує електричне поле, що прискорює рух інжектованих у базу дірок у напрямку колектора.

Транзистори такого типу називають дрейфовими, оскільки проходження неосновних носіїв через базу відбувається не за рахунок дифузії (як у звичайних транзисторах, див. вище рис. 3.5.2), а за рахунок дрейфу у внутрішньому електричному полі. За рахунок дрейфу можна скоротити час руху неосновних носіїв у базі в кілька разів порівняно з дифузією і тим самим істотно підвищити швидкодію транзистора.

3.6.4. Фізична еквівалентна схема транзистора в області високих частот

Як було показано, для малих сигналів транзистор можна розглядати як лінійний елемент. Отже, його можна представити у вигляді деякої лінійної схеми. Приклад такої схеми ми вже розглядали (рис. 3.5.12). Однак схема, показана на цьому рисунку, по-перше, не відповідає будові транзистора, і, по-друге, не описує його поведінку на високих частотах. Цих недоліків позбавлена фізична еквівалентна схема транзистора в області високих частот (схема Джіаколетто), представлена на рис. 3.6.5[4]. Тут точки B, E, C відповідають базі, емітеру та колектору. Видно, що ця еквівалентна схема відповідає увімкненню транзистора зі спільним емітером.

Оскільки схема Джіаколетто описує транзистор у наближенні малого сигналу, то всі резистори, що входять до її складу, відповідають диференціальним (а не омічним) опорам реального транзистора.

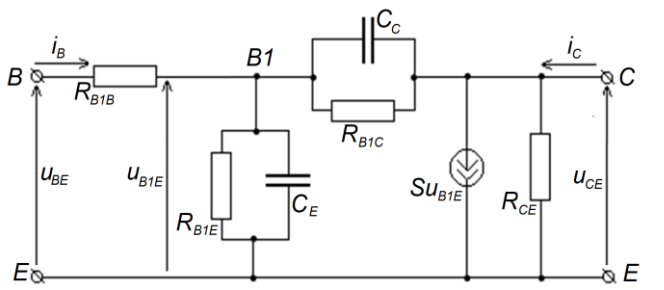


Рис. 3.6.5. Фізична еквівалентна схема біполярного транзистора в області високих частот (схема Джіаколетто)

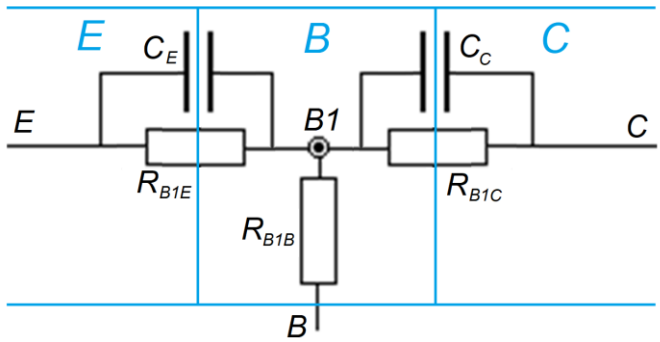


Рис. 3.6.6. До пояснення змісту елементів схеми Джіаколетто

Точка B1 знаходиться всередині об'єму бази. Саме її потенціал u_{B1E} (щодо емітера) визначає колекторний струм. Резистор R_{B1B} відповідає об'ємному опору бази (нагадаємо, що вона виготовляється зі слабколегованого напівпровідника з порівняно великим об'ємним опором). Таким чином, напруга, прикладена між колектором та емітером, розподіляється між об'ємним опором бази та емітерним *p-n* переходом з активним опором R_{B1E} (перерахованим до базового струму, оскільки він виступає як вхідний) і дифузійною ємністю C_E. Конденса-

[4] Аналогом схеми Джіаколетто для польових транзисторів можна вважати схему, показану на рис. 3.4.3 б.

тор C_C відповідає бар'єрній ємності колекторного *p-n* переходу. Співвідношення описаних вище елементів еквівалентної схеми з реальною структурою транзистора демонструє рис. 3.6.6.

Резистор R_{B1C} моделює внутрішній зворотний зв'язок у транзисторі, тобто вплив вихідного сигналу на вхідний. Оскільки цей вплив малий, відповідний опір виходить великим.

Джерело струму Su_{B1E} моделює здатність транзистора, увімкненого за схемою зі спільним емітером, підсилювати струм. Величина S знаходиться зі співвідношення, що справджується, коли вихід схеми замкнений накоротко:

$$S = \frac{i_C}{u_{B1E}} = \frac{\alpha i_E}{u_{B1E}} = \frac{\alpha}{R_d} \qquad (3.6.10)$$

де R_d – диференціальний опір відкритого емітерного *p-n* переходу, тобто R_{B1E} (див. формулу (3.2.2)). На низьких частотах, нехтуючи впливом ємностей і вважаючи опір R_{B1C} дуже великим, можна записати, що

$$u_{B1E} = u_{BE}\frac{R_{B1E}}{R_{B1B} + R_{B1E}},$$

звідки

$$i_C = Su_{B1E} = u_{BE}\frac{SR_{B1E}}{R_{B1B} + R_{B1E}} = S_1 u_{BE}, \qquad S_1 = \frac{SR_{B1E}}{R_{B1B} + R_{B1E}}. \qquad (3.6.11)$$

Величина S_1 в першому наближенні відповідає крутості прохідної характеристики транзистора.

3.6.5. Зв'язок елементів схеми Джіаколетто з h-параметрами транзистора та його граничною частотою

Зазначимо, що на низьких частотах схема Джіаколетто не збігається зі еквівалентною схемою транзистора в h-параметрах (рис. 3.6.7, що повторює рис. 3.5.12). В останній внутрішній зворотний зв'язок моделюється керованим джерелом напруги $h_{12E}U_{CE}$, а в схемі Джіаколетто на низьких частотах його забезпечує опір R_{B1C}, увімкнений між колектором та середньою точкою бази $B1$.

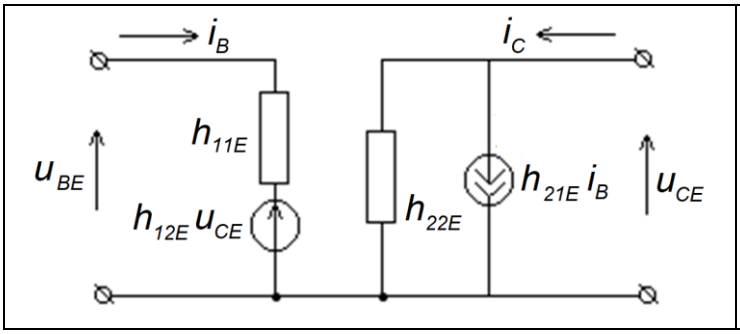


Рис. 3.6.7. Еквівалентна схема транзистора, увімкненого зі спільним емітером, в h-парметрах

Тим не менше, номінали елементів схеми Джіаколетто можна пов'язати з h-параметрами транзистора та його граничними частотами. Підкреслимо, що значення h-параметрів залежать від вибору робочої точки транзистора. Отже, це стосується й елементів схеми Джіаколетто.

Означення h_E-параметрів (3.4.13) в режимі малого сигналу можна подати у формі

$$h_{11E} = \left.\frac{u_{BE}}{i_B}\right|_{u_{CE}=0}, \quad h_{12E} = \left.\frac{u_{BE}}{u_{CE}}\right|_{i_B=0}, \quad h_{21E} = \left.\frac{i_C}{i_B}\right|_{u_{CE}=0}, \quad h_{22E} = \left.\frac{i_C}{u_{CE}}\right|_{i_B=0}. \qquad (3.6.12)$$

Для знаходження їхнього зв'язку з елементами схеми Джіаколетто розглянемо остан-

ню на низьких частотах в режимі розімкнутого входу, коли $i_B = 0$ (рис. 3.6.8 а), та короткозамкненого виходу, коли $u_{CE} = 0$ (рис. 3.6.8 б).

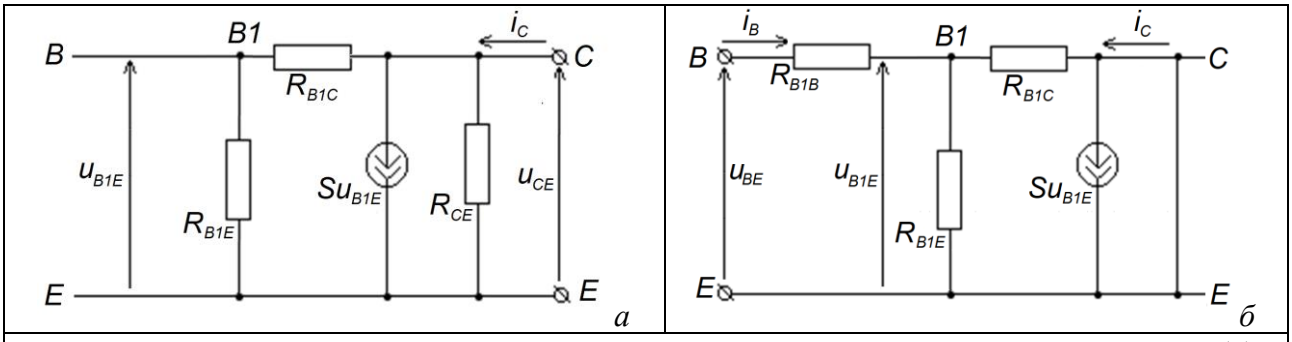


Рис. 3.6.8. Схема Джіаколетто на низьких частотах у режимах розімкнутого входу (а) та короткозамкненого виходу (б)

Для схеми, поданої на рис. 3.6.7 а,

$$h_{12E} = \frac{u_{BE}}{u_{CE}} = \frac{u_{B1E}}{u_{CE}} = \frac{R_{B1E}}{R_{B1E} + R_{B1C}} \approx \frac{R_{B1E}}{R_{B1C}}, \tag{3.6.13}$$

оскільки $h_{12E} \ll 1$ (див. вище задачу П 3.5.1), а, отже, $R_{B1E} \ll R_{B1C}$.

Для знаходження параметра h_{22E} будемо напругу u_{CE} вважати заданою. Тоді струм керованого джерела можна, виходячи з рис. 3.6.8 а, записати як

$$Su_{B1E} = Su_{CE} \frac{R_{B1E}}{R_{B1E} + R_{B1C}} = \frac{\alpha u_{CE}}{R_{B1E} + R_{B1C}}. \tag{3.6.14}$$

(враховано формулу (3.6.10)). Запишемо еквівалентний опір між точками С і Е, що складається з паралельних опорів R_{CE} та $R_{B1E} + R_{B1C}$:

$$R_{eq} = \frac{R_{CE}\left(R_{B1E} + R_{B1C}\right)}{R_{CE} + R_{B1E} + R_{B1C}} \approx \frac{R_{CE} R_{B1C}}{R_{CE} + R_{B1C}} \approx R_{CE} \tag{3.6.15}$$

(враховано, що $R_{B1E} \ll R_{CE} \ll R_{B1C}$). В результаті для вихідної напруги отримуємо:

$$u_{CE} = R_{eq}\left(i_C - \frac{\alpha u_{CE}}{R_{B1E} + R_{B1C}}\right), \tag{3.6.16}$$

звідки

$$i_C = u_{CE}\left(\frac{1}{R_{eq}} + \frac{\alpha}{R_{B1E} + R_{B1C}}\right). \tag{3.6.17}$$

Таким чином, відповідно до формули (3.6.12)

$$h_{22E} = \frac{i_C}{u_{CE}} = \frac{1}{R_{eq}} + \frac{\alpha}{R_{B1E} + R_{B1C}} \approx \frac{1}{R_{CE}} + \frac{\alpha}{R_{B1E} + R_{B1C}} \approx \frac{1}{R_{CE}}. \tag{3.6.18}$$

Для схеми, поданої на рис. 3.6.8 б, можна, очевидно, відразу записати, що її вхідний опір має вигляд

$$h_{11E} = R_{B1B} + \frac{R_{B1E} R_{B1C}}{R_{B1E} + R_{B1C}} \approx R_{B1B} + R_{B1E}. \tag{3.6.19}$$

Нарешті, відповідно до формули (3.6.11) і схеми (рис. 3.6.8 б),

$$i_C = Su_{B1E} = Si_B \frac{\dfrac{1}{R_{B1E}}}{\dfrac{1}{R_{B1E}} + \dfrac{1}{R_{B1C}}} R_{B1E} = i_B \frac{SR_{B1E} R_{B1C}}{R_{B1E} + R_{B1C}}, \tag{3.6.20}$$

звідки

$$h_{21E} = \frac{i_C}{i_B} = S\frac{R_{B1E}R_{B1C}}{R_{B1E}+R_{B1C}} \approx SR_{B1E}\,. \qquad (3.6.21)$$

Тепер подивимось, як поводить себе коефіцієнт $h_{21E} = \beta$ на високих частотах. Еквівалентна схема Джіаколетто з короткозамкненим виходом подана на рис. 3.6.9.

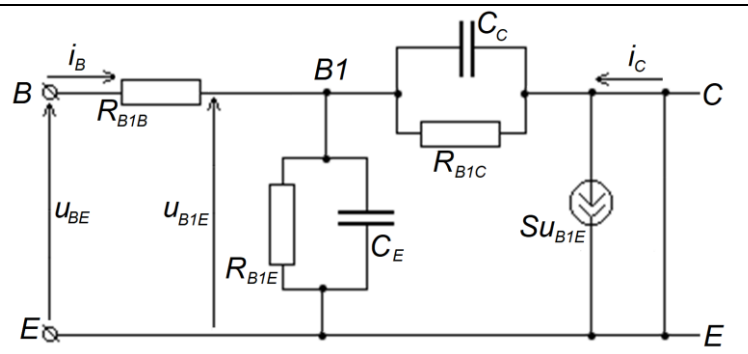


Рис. 3.6.9. Схема Джіаколетто з короткозамкненим виходом на високих частотах

Видно, що тепер резистори R_{B1E} і R_{B1C} та конденсатори C_C і C_E увімкнені паралельно і утворюють імпеданс

$$Z_1 = \left[\frac{1}{R_{B1E}} + \frac{1}{R_{B1C}} + i\omega\left(C_E + C_C\right)\right]^{-1} \approx \left[\frac{1}{R_{B1E}} + i\omega\left(C_E + C_C\right)\right]^{-1} = \frac{R_{B1E}}{1 + i\omega R_{B1E}\left(C_E + C_C\right)}\,. \qquad (3.6.22)$$

На цьому імпедансі падає напруга, яка керує колекторним струмом, тому

$$u_{B1E} = i_B Z_1 \approx \frac{i_B R_{B1E}}{1 + i\omega R_{B1E}\left(C_E + C_C\right)}\,, \qquad (3.6.23)$$

звідки

$$h_{21E}(\omega) = \frac{Su_{B1E}}{i_B} = SZ_1 \approx \frac{SR_{B1E}}{1 + i\omega R_{B1E}\left(C_E + C_C\right)}\,. \qquad (3.6.24)$$

При $\omega \to 0$ формула (3.6.24) збігається з (3.6.21), як і має бути.

З іншого боку, оскільки за означенням $h_{21E} = \beta$, то формула (3.6.24) збігається й із формулою (3.6.7),

$$\beta(\omega) = \frac{\beta_0}{1 + i\,\omega/\omega_\beta}\,,$$

якщо покласти

$$\omega_\beta = \frac{1}{R_{B1E}\left(C_E + C_C\right)}\,. \qquad (3.6.25)$$

Візьмемо типовий набір h_E-параметрів і спробуємо розрахувати за ним елементи схеми Джіаколетто. Отже, нехай $h_{11E} = 1500$ Ом; $h_{12E} = 10^{-4}$; h_{21E} =50; h_{22E} = $5\cdot10^{-5}$ Сим. Нехай $f_{h_{21E}}$ =2 МГц, I_E = 1 мА.

Для визначення крутості скористаємося формулою (3.6.9)

$$S = \alpha/R_d$$

Перепишемо формулу (3.1.2) для диференціального опору R_d *p-n* переходу у вигляді

$$R_d = \frac{k_B T}{eI_0}\exp\left(-\frac{eU}{k_B T}\right) \approx \frac{k_B T}{eI_E} = \frac{1.38\cdot10^{-23}\cdot300}{1.60\cdot10^{-19}\cdot10\cdot10^{-3}} = \frac{1.38\cdot3}{1.60} \approx 2.59\ \text{Ом},$$

де враховано, що емітерний перехід відкритий, тому $I_E \gg I_0$. Згідно (3.5.6),

$$\alpha \approx 1-\frac{1}{\beta}=1-\frac{1}{h_{21E}}=1-\frac{1}{50}=0.98.$$

Таким чином,

$$S=\frac{\alpha}{R_d}=\frac{0.98}{2.59}=0.38\,\text{А/В}=380\,\text{мА/В}.$$

Відповідно до формули (3.6.21)

$$R_{B1E}\approx\frac{h_{21E}}{S}=\frac{h_{21E}R_d}{\alpha}=\frac{50\cdot 2.59}{0.98}=132\,\text{Ом}.$$

$$R_{B1B}=h_{11E}-R_{B1E}=1500-132\approx 1.37\,\text{кОм}.$$

Згідно формули (3.6.13),

$$R_{B1C}\approx\frac{R_{B1E}}{h_{12E}}=\frac{1370}{10^{-4}}=13.7\,\text{МОм}.$$

З формули (3.6.36) випливає, що

$$R_{CE}\approx\frac{1}{h_{22E}}=\frac{1}{5\cdot 10^{-5}}=20\,\text{кОм}.$$

Бачимо, що нерівність $R_{B1E}\ll R_{CE}\ll R_{B1C}$, використана для спрощення кількох розрахункових формул, справді виконується.

Нарешті, з формули (3.6.25) випливає, що

$$C_E+C_C=\frac{1}{\omega_\beta R_{B1E}}=\frac{1}{2\pi f_{h21E}R_{B1E}}=\frac{1}{6.3\cdot 2\cdot 10^6\cdot 132}=\frac{10^9}{0.63\cdot 2\cdot 1.32}=760\,\text{пФ}.$$

Контрольні питання до підрозділу 3.6

1. Чим визначаються частотні властивості біполярного транзистора?
2. Що таке схема Джіаколетто? Для чого вона використовується?
3. Як можна розрахувати елементи схеми Джіаколетто в заданій робочій точці за вхідними й вихідними характеристиками біполярного транзистора?

Розділ 4. Лінійні пристрої аналогової радіоелектроніки

Дві перші розділи курсу були присвячені сигналам у радіотехніці та їхньому проходженню через лінійні кола. Тепер ми перейдемо до процесів у радіотехнічних колах. Як ми вже говорили (див. Вступ), до таких процесів належать генерація, підсилення, фільтрація, модуляція, детектування, перетворення частоти та інші подібні процеси. У цьому та наступному розділах ми, базуючись на наведеній у розділі 3 інформації про напівпровідникові прилади, розглянемо основні схеми аналогової радіоелектроніки, призначені для реалізації різних типів згаданих процесів. Ми будемо розглядати, як правило, найпростіші схеми, що дають можливість зрозуміти принцип їхньої роботи.

У розділі 2 цього курсу розглядалися методи аналізу лінійних електричних кіл. В цьому розділі ми також розглядатимемо лінійні кола, за допомогою яких здійснюються процеси фільтрації та підсилення сигналів. До цієї категорії належать, перш за все, пасивні фільтри, що складаються з лінійних елементів (резисторів, конденсаторів, котушок індуктивності).

До лінійних схем можна віднести також підсилювачі електричних сигналів. Звичайно від таких пристроїв вимагають, щоб форма корисних сигналів на виході була такою самою, як на вході (в робочому діапазоні частот). До складу підсилювачів, як правило, входять активні елементи (не плутати з активним опором) – електронні лампи, польові та біполярні транзистори (див. п. 3.1). Активні елементи завжди є нелінійними елементами, їхні параметри в загальному випадку суттєво залежать від сигналів, які на них подаються. Однак, як уже відзначалося (див. вище пп. 2.2.1, 3.4, 3.5.4, 3.6), для сигналів невеликої амплітуди (для невеликих відхилень від робочої точки) в багатьох випадках ці пристрої можуть розглядатися як лінійні чотириполюсники. Тоді й схеми, до складу яких входять такі активні елементи, можна вважати лінійними. Саме такі схеми розглядатимуться в цій частині курсу.

Ми розглянемо тут також параметричні підсилювачі, які не містять активних елементів і засновані на іншому принципі роботи, але також описуються лінійними диференціальними рівняннями (зі змінними коефіцієнтами).

Всі розрахунки в цій частині курсу використовують метод комплексних амплітуд (див. вище п. 2.2), тобто сигнали вважаються гармонічними функціями часу.

Аналогові радіотехнічні схеми, для роботи яких врахування нелінійних ефектів є принциповим, розглядатимуться в наступній частині курсу. Це автогенератори гармонічних коливань, різні типи імпульсних генераторів, а також нелінійні перетворювачі сигналів – модулятори, детектори, перетворювачі частоти тощо.

4.1. Пасивні фільтри

Під фільтрами розуміють пристрої, які пропускають сигнали (або складові їхнього спектру) в певному діапазоні частот і придушують їх в інших частотних діапазонах.

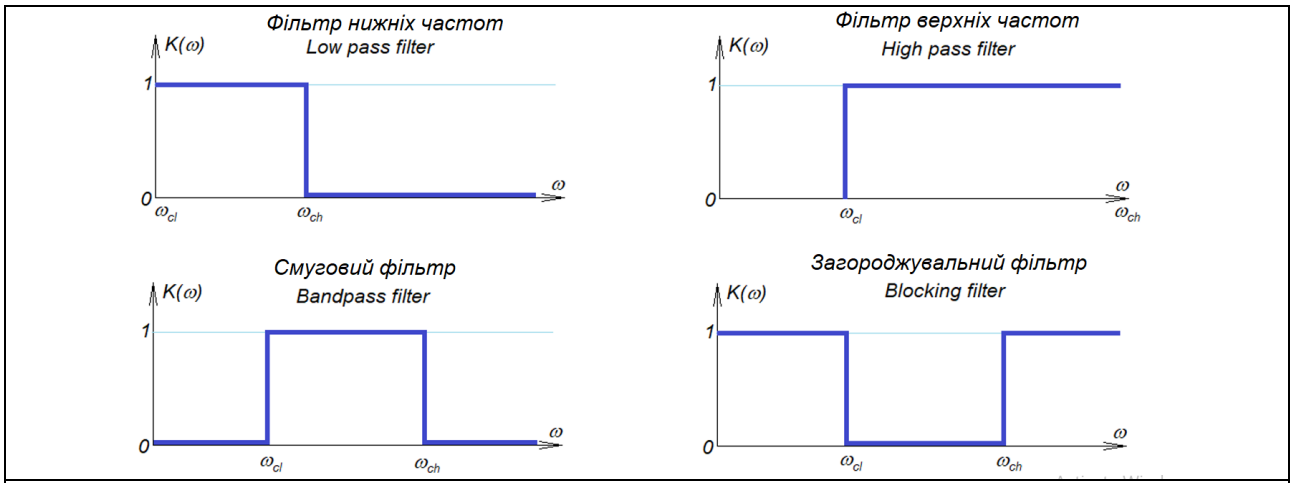


Рис. 4.1.1. Графіки ідеалізованих амплітудно-частотних характеристик різних типів фільтрів

4.1.1. Визначення, класифікація та параметри фільтрів

Назви фільтрів відповідають тим частотам, які вони пропускають (їхній смузі пропускання). Наприклад, фільтр нижніх частот (ФНЧ) пропускає сигнали з частотами від нуля до деякої ω_C (індекс – від англ. cutoff frequency, частота відтину, або гранична частота). Відповідно частоти від ω_C до нескінченності такий фільтр придушує.

Найбільш поширеними типами фільтрів є вже згаданий ФНЧ, фільтр верхніх частот (ФВЧ), смуговий фільтр та загороджувальний фільтр. Графіки ідеалізованих АЧХ цих фільтрів подані на рис. 4.1.1.

АЧХ реальних фільтрів мають розмиті межі смуги пропускання (рис. 4.1.2). Для таких фільтрів межі смуги пропускання визначаються з умови, що на відповідних частотах потужність вихідного сигналу зменшується вдвічі порівняно з максимальним значенням, а амплітуда – в $\sqrt{2}$ разів, тобто

$$\left|\frac{K(\omega_C)}{K_{\max}}\right| = \frac{1}{\sqrt{2}} \simeq 0.707 \qquad (4.1.1)$$

(враховано, що потужність сигналу пропорційна квадрату його амплітуди).

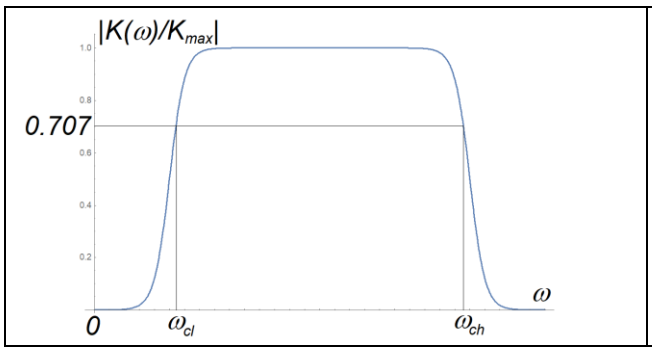


Рис. 4.1.2. До визначення граничних частот фільтра

Розмитість межі смуги пропускання характеризується крутістю зрізу фільтру. Подивимося, як вводиться це поняття для верхньої межі смуги пропускання.

Ослаблення сигналу часто вимірюють у децибелах:

$$|K(\omega)|(\partial Б) = 10\lg\frac{P}{P_{\max}} = 20\lg\left|\frac{K(\omega)}{K_{\max}}\right|. \qquad (4.1.2)$$

Тоді межам смуги пропускання відповідає ослаблення сигналу $10\lg(1/2) \approx -3.01\ \partial Б$.

Розглянемо для прикладу RC-ланцюжок для випадку, коли вихідний сигнал знімається з конденсатора (рис. 4.1.3 а). Коефіцієнт передачі для цього випадку визначається формулою (2.2.15):

$$K(\omega) = \frac{u_{out1}}{u_{in}} = \frac{I/i\omega C}{IR + I/i\omega C} = \frac{1}{1+i\omega\tau}, \quad \tau = RC. \qquad (4.1.3)$$

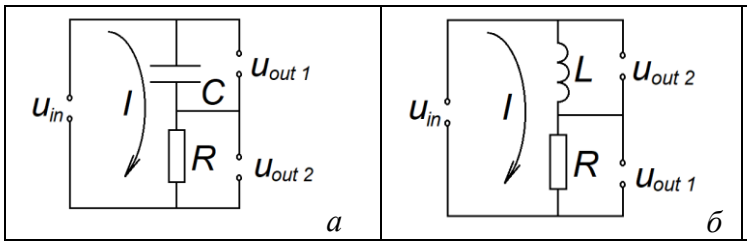


Рис. 4.1.3. Схеми найпростіших фільтрів на основі RC-ланцюжка (а) та RL-ланцюжка (б): ФНЧ (вихід 1) та ФВЧ (вихід 2)

Відповідно

$$\left|K(\omega)\right| = \frac{1}{\sqrt{1+(\omega\tau)^2}}, \qquad (4.1.4)$$

$$\varphi_K(\omega) = -\mathrm{arctg}\,\omega\tau \qquad (4.1.5)$$

(рис. 4.1.4).

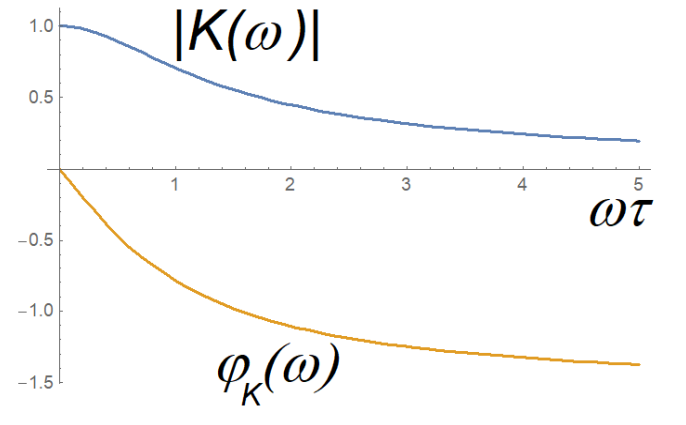


Рис. 4.1.4. Графіки АЧХ та ФЧХ RC-ланцюжка, поданого на рис. 4.1.3 а (вихідна напруга знімається з конденсатора)

Очевидно, такий ланцюжок є фільтром низьких частот із граничною частотою $\omega_C = 1/\tau$. Якщо $\omega \gg \omega_C$, то $\left|K(\omega)\right| \approx (\omega\tau)^{-1}$. Тоді на графіку АЧХ, побудованому в логарифмічному масштабі, на частотах $\omega \gg \omega_C$ спостерігається лінійне спадання (рис. 4.1.5).Справді, якщо записати частоту у вигляді $\omega = \omega_C \cdot 10^x$, отримаємо в цьому випадку:

$$\left|K(\omega)\right|(\partial Б) = 20\lg\frac{1}{10^x} = -20x. \qquad (4.1.6)$$

Зміна величини x на одиницю відповідає зміні частоти в 10 разів.

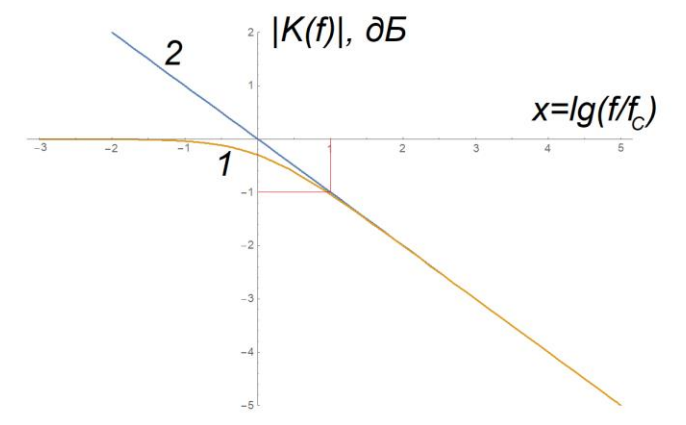


Рис. 4.1.5. До визначення крутості зрізу фільтру: 1 – графік АЧХ в логарифмічному масштабі, 2 – його асимптота, нахил якої відповідає крутості зрізу, а точка $x = 1$ – частоті зрізу

Кількісно швидкість спадання графіка на ділянці $\omega \gg \omega_C$, або, як кажуть, крутість зрізу АЧХ, визначається нахилом графіка на цій ділянці $\Delta\left|K(\omega)\right|(\partial Б)/\Delta x$. У даному разі ми отримали крутість зрізу 20 децибел на декаду. Якщо записати частоту сигналу у формі $\omega = \omega_C \cdot 2^y$, то величина $\Delta\left|K(\omega)\right|(\partial Б)/\Delta y$ буде крутістю зрізу в децибелах на октаву[1]. Легко зрозуміти, що має виконуватися співвідношення $10^x = 2^y$, звідки $x = y\lg 2 \approx 0.301y$. Отже, для розглянутого прикладу крутість зрізу складає приблизно 6 децибел на октаву.

Фільтр з асимптотикою $\left|K(\omega)\right| \approx (\omega\tau)^{-1}$ називають фільтром першого порядку (саме

[1] В музиці та акустиці октава – це інтервал звукових тонів, частоти яких на кінцях інтервалу відрізняються вдвічі.

таким є розглянутий вище приклад фільтра на основі RC-ланцюжка), з асимптотикою $\left|K(\omega)\right| \approx (\omega\tau)^{-2}$ – другого порядку, з асимптотикою $\left|K(\omega)\right| \approx (\omega\tau)^{-3}$ – третього порядку тощо.

Якщо спектр вхідного сигналу лише частково потрапляє в смугу пропускання, то форма сигналу на виході спотворюється. Такі спотворення називаються лінійними. При лінійних спотвореннях у спектрі не виникає нових частот, але змінюється співвідношення між наявними спектральними складовими.

Прикладом лінійних спотворень може бути зміна форми імпульсу (зокрема, прямокутного) при проходженні через лінійний фільтр[2] (приклади такої зміни показані на рис. 4.1.6, пор. з рис. 2.3.10). Ця зміна дозволяє судити про амплітудно-частотну характеристику останнього.

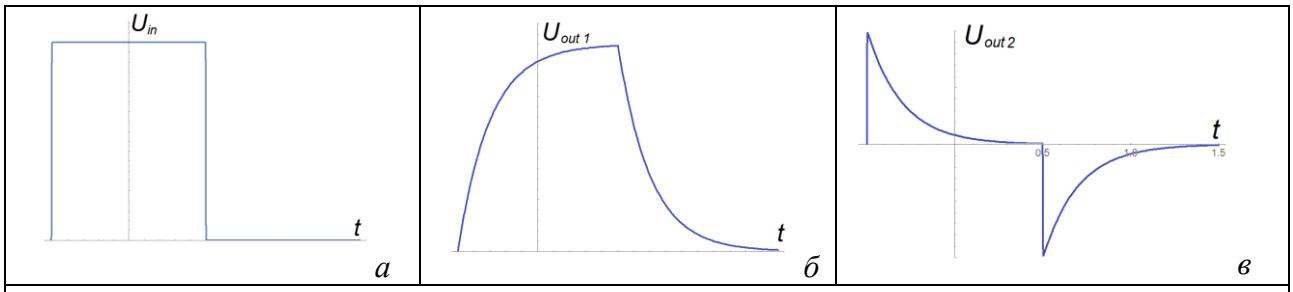


Рис. 4.1.6. Зміна форми прямокутного імпульсу (а) після проходження через ФНЧ (б) та через ФВЧ (в), показані на рис. 4.1.3

4.1.2. RC- та RL-ланцюжки як фільтри верхніх та нижніх частот

Вище ми розглянули RC-ланцюжок і переконалися, що він служить фільтром низьких частот за умови, що вихідний сигнал знімається з конденсатора. Легко показати, що такий самий ланцюжок може бути фільтром верхніх частот за умови, що вихідна напруга знімається з резистора (рис. 4.1.3 а)

Аналогічно RL-ланцюжок може виступати в ролі ФНЧ, якщо вихідна напруга знімається з резистора, і в ролі ФВЧ, якщо вона знімається з котушки індуктивності (рис. 4.1.3 б). Легко показати, що ФНЧ на основі RL-ланцюжка також належить до фільтрів першого порядку.

4.1.3. Коливні контури як фільтри верхніх та нижніх частот

В ролі ФНЧ або ФВЧ можуть виступати й послідовні коливні контури. Неважко зрозуміти, що при знятті вихідного сигналу з ємності можна отримати ФНЧ, а з індуктивності – ФВЧ (рис. 4.1.7).

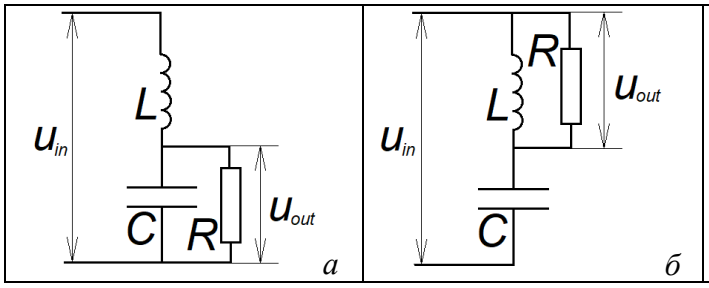


Рис. 4.1.7. Схеми ФНЧ (а) та ФВЧ (б) на основі навантаженого послідовного контуру

Розглянемо, наприклад, схему ФНЧ. Її верхня ланка – це котушка індуктивності; нижня, з якої й знімається вихідна напруга – паралельне сполучення конденсатора й резистора навантаження. Очевидно, її передавальна функція може бути подана у вигляді

$$K(\omega) = \frac{R\cdot(1/i\omega C)/(R+1/i\omega C)}{i\omega L + R\cdot(1/i\omega C)/(R+1/i\omega C)} = \frac{R\cdot(1/i\omega C)}{i\omega L(R+1/i\omega C)+R\cdot(1/i\omega C)} =$$

[2] Нагадаємо, що параметри лінійних кіл не залежать від поданого сигналу (див. вище п. 2.1.1).

$$= \frac{R}{i\omega L\left(i\omega CR+1\right)+R} = \frac{1}{1-\omega^2/\omega_0^2+i\omega\tau}, \qquad (4.1.7)$$

де $\omega_0^2 = 1/LC$, $\tau = L/R$.

Відповідно

$$\left|K\left(\omega\right)\right| = \frac{1}{\sqrt{\left(1-\omega^2/\omega_0^2\right)^2+\left(\omega\tau\right)^2}}, \qquad (4.1.8)$$

$$\varphi_K\left(\omega\right) = -\mathrm{arctg}\frac{\omega\tau}{1-\omega^2/\omega_0^2} \qquad (4.1.9)$$

(рис. 4.1.8). Параметром сім'ї характеристик, поданої на рисунку, є величина $\omega_0\tau = \sqrt{L}/R\sqrt{C} = \rho/R = Q$ – добротність коливного контуру (в даному разі втрати енергії, що визначають добротність, зумовлені скінченним опором навантаження).

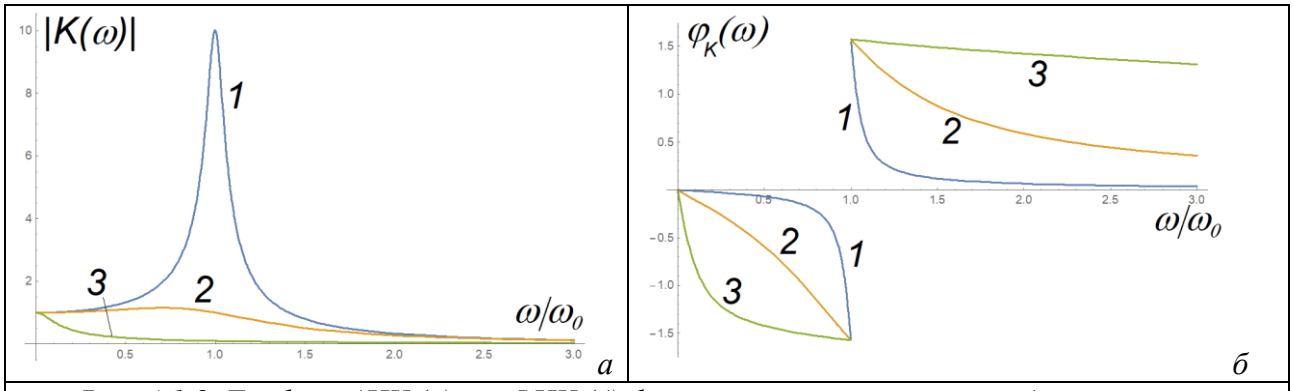


Рис. 4.1.8. Графіки АЧХ (а) та ФЧХ (б) фільтра низьких частот на базі коливного контуру, поданого на рис. 4.1.7 а: 1 – $Q = 10$, 2 – $Q = 1$, 3 – $Q = 0.1$.

Випадок $Q = \infty$ відповідає ненавантаженому режиму на виході, тобто контуру без втрат. Зазначимо, що на власній частоті контуру $\left|K\left(\omega_0\right)\right| = Q$. Відповідно при великих Q резонанс сильно виявлений (криві 1 на рис. 4.1.8). При малих Q, коли конденсатор зашунтований низькоомним резистором, резонанс на АЧХ взагалі не проявляється, а модуль коефіцієнта передачі спадає до значень, менших від одиниці, вже на частотах, помітно менших від ω_0 (криві 3). Нарешті, при $Q = 1$ АЧХ в діапазоні від нуля до ω_0 близька до одиниці, а далі спадає (криві 2), тобто її форма найближча до ідеального ФНЧ.

З формули для коефіцієнту передачі видно, що за умов $\omega \gg \omega_0$ та $\omega \gg \omega_0^2\tau$ він спадає як $\left(\omega_0/\omega\right)^2$, тобто ця схема є фільтром другого порядку з крутістю зрізу 12 дБ на октаву. З цієї точки зору ФНЧ на основі послідовного контуру кращий за аналогічний фільтр на основі RC-ланцюжка.

4.1.4. Смугові та загороджувальні фільтри на базі RC- та RL-ланцюжків

Смугові та загороджувальні (режекторні, від англ. reject – вирізати) фільтри звичайно складніші за ФНЧ та ФВЧ. В принципі, їх можна сконструювати з ідеалізованих ФНЧ та ФВЧ (рис. 4.1.9). Але в реальності каскади впливають один на одного, що призводить до погіршення сумарних характеристик. Крім того, такі схеми виходять досить складними.

Для ілюстрації цих тверджень розглянемо найпростіший смуговий RC-фільтр, схема якого подана на рис. 4.1.10 а. Подамо його коефіцієнт передачі у вигляді

$$K\left(\omega\right) \equiv \frac{u_{out}}{u_{in}} = \frac{u_1}{u_{in}}\frac{u_{out}}{u_1} = K_{in}\left(\omega\right)K_{out}\left(\omega\right). \qquad (4.1.10)$$

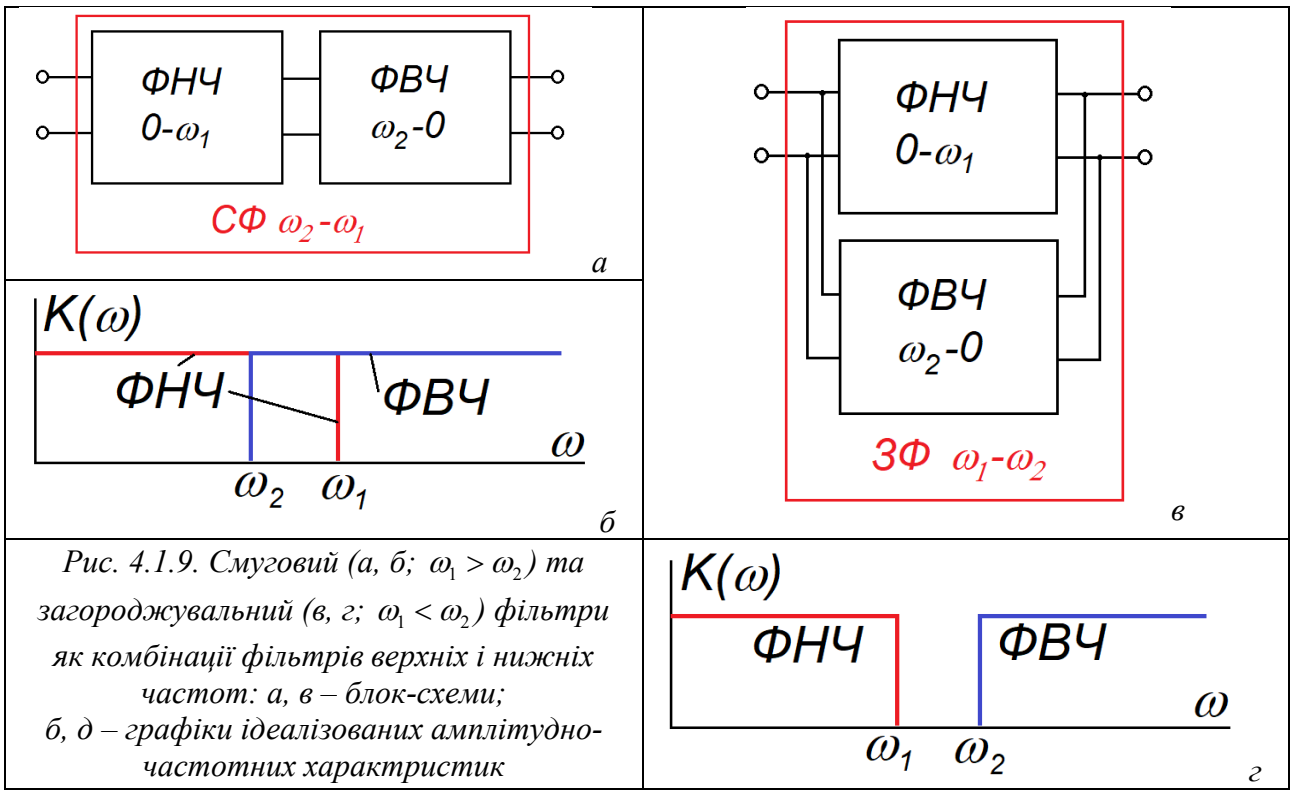


Рис. 4.1.9. Смуговий (а, б; $\omega_1 > \omega_2$) та загороджувальний (в, г; $\omega_1 < \omega_2$) фільтри як комбінації фільтрів верхніх і нижніх частот: а, в – блок-схеми; б, д – графіки ідеалізованих амплітудно-частотних характристик

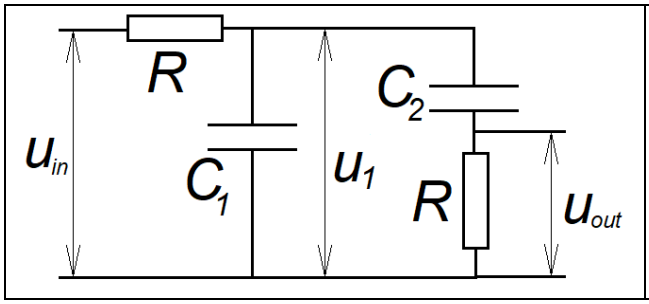


Рис. 4.1.10. Схема смугового RC-фільтру в ненавантаженому режимі

Перший каскад схеми – це ФНЧ. На його вхід подається напруга U_{in}, вихідна напруга – U_1. Якби він працював у ненавантаженому режимі, його коефіцієнт передачі був би

$$K'_{in} = \frac{1/i\omega C_1}{R + 1/i\omega C_1} = \frac{1}{1 + i\omega/\omega_1} = \frac{1}{1 + i\Omega/k} \qquad (4.1.11)$$

з граничною частотою $\omega_1 = 1/RC_1$, або $\Omega'_{c1} = k$ (тут введено нормовану частоту $\Omega = \omega/\omega_2$, де $\omega_2 = 1/RC_2$, і коефіцієнт $k = \omega_1/\omega_2 = C_2/C_1$).

Насправді нижня ланка подільника першого каскаду являє собою паралельне сполучення ємності C_1 та ланцюжка RC_2. Її еквівалентний опір складає

$$Z_{eq} = \frac{(1/i\omega C_1)(R + 1/i\omega C_2)}{1/i\omega C_1 + R + 1/i\omega C_2} = \frac{1}{i\omega C_1}\frac{1 + i\omega/\omega_2}{(1 + C_2/C_1 + i\omega/\omega_2)} = \frac{R}{ik\Omega}\frac{1 + i\Omega}{(1 + k + i\Omega)}. \qquad (4.1.12)$$

З урахуванням цього коефіцієнт передачі першого каскаду набуває вигляду

$$K_{in} = \frac{U_1}{U_{in}} = \frac{Z_{eq}}{R + Z_{eq}} = \frac{(R/ik\Omega)\left[(1 + i\Omega)/(1 + k + i\Omega)\right]}{R + (R/ik\Omega)\left[(1 + i\Omega)/(1 + k + i\Omega)\right]} =$$
$$= \frac{1 + i\Omega}{ik\Omega(1 + k + i\Omega) + 1 + i\Omega} = \frac{1 + i\Omega}{1 - k\Omega^2 + i\Omega(1 + k + k^2)}. \qquad (4.1.13)$$

Його граничну частоту знаходимо з умови $|K_{in}|^2 = 1/2$. В загальному випадку відповідна

формула виходить громіздкою:

$$\Omega_{c1}^2 = \frac{1}{2k^2}\left[-\left(3k^2+2k^3+k^4-1\right)+\sqrt{\left(3k^2+2k^3+k^4-1\right)^2+4k^2}\right]. \qquad (4.1.14)$$

Вона спрощується для двох крайніх випадків:

$$\Omega_{c1} \approx \begin{cases} 1/k >> 1, & k << 1; \\ 1/k^2 << 1, & k >> 1 \end{cases} \qquad (4.1.15)$$

(рис. 4.1.11).

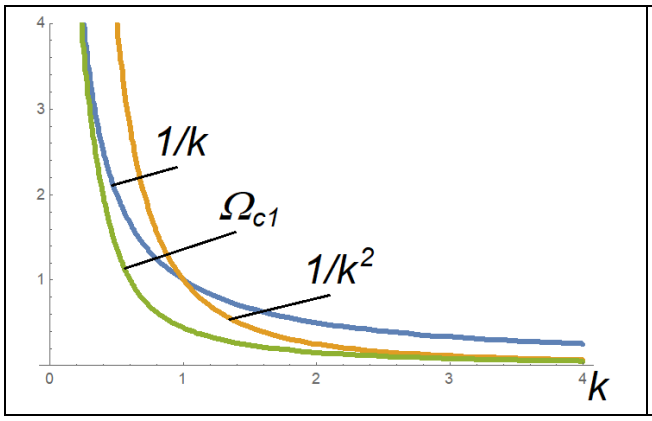


Рис. 4.1.11. Залежність $\Omega_{c1}(k)$ та її апроксимації при малих та великих значеннях k

Другий каскад схеми – це ФВЧ з коефіцієнтом передачі

$$K_{out} = \frac{U_{out}}{U_1} = \frac{R}{R+1/i\omega C_2} = \frac{i\omega/\omega_2}{1+i\omega/\omega_2} = \frac{i\Omega}{1+i\Omega} \qquad (4.1.16)$$

і граничною частотою $\omega = \omega_2$, або $\Omega_{c2} = 1$.

Для того, щоб схема, подана на рис. 4.1.10, була смуговим фільтром, повинна виконуватися умова $\Omega_{c2} = 1 \le \Omega_{c1}$ (пор. з рис. 4.1.9 б). Ця умова справджується при $k << 1$, тобто при $\omega_1 \ll \omega_2$ (рис. 4.1.11).

Коефіцієнт передачі двох послідовно увімкнених каскадів має вигляд

$$K = K_{in}K_{out} = \frac{1+i\Omega}{\left[1-k\Omega^2+i\Omega\left(1+k+k^2\right)\right]}\frac{i\Omega}{(1+i\Omega)} = \frac{i\Omega}{1-k\Omega^2+i\Omega\left(1+k+k^2\right)}. \qquad (4.1.17)$$

Звідси можна знайти його АЧХ та ФЧХ:

$$|K| = \left|\frac{i\Omega}{1-k\Omega^2+i\Omega\left(1+k+k^2\right)}\right| = \frac{\Omega}{\sqrt{\left(1-k\Omega^2\right)^2+\Omega^2\left(1+k+k^2\right)^2}}, \qquad (4.1.18)$$

$$\varphi_K = \frac{\pi}{2} - \operatorname{arctg}\frac{\Omega\left(1+k+k^2\right)}{1-k\Omega^2}. \qquad (4.1.19)$$

На рис. 4.12 подані графіки частотних залежностей $|K|$, а також допоміжних величин – $|K'_{in}|$, $|K_{in}|$ та $|K_{out}|$. Вигляд графіків для різних значень параметра k відповідає висловленим вище міркуванням. На рис. 4.13 подано сім'ю ФЧХ.

Частоту, що відповідає максимуму АЧХ, знаходимо з умови $d|K(\Omega)|^2 / d\left(\Omega^2\right) = 0$:

$$\Omega_{opt} = 1/\sqrt{k}. \qquad (4.1.20)$$

На цій частоті коефіцієнт передачі складає величину

$$|K|_{\max} = \frac{1}{1+k+k^2}. \qquad (4.1.21)$$

Видно, що $|K|_{\max} \to 1$ при $k \to 0$ (пор. з рис. 4.12)

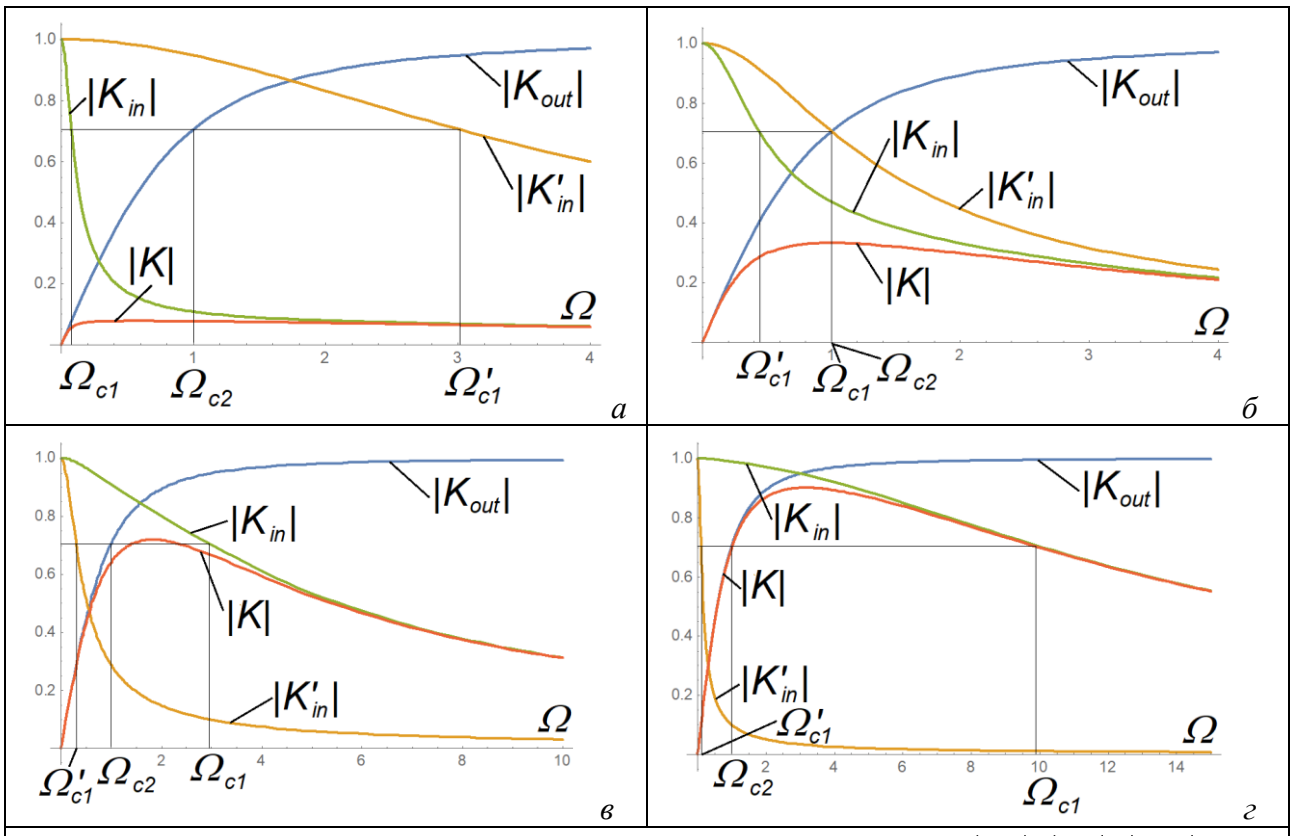


Рис. 4.12. Графіки частотних залежностей коефіцієнтів передачі $|K'_{in}|$, $|K_{in}|$, $|K_{out}|$ *та* $|K| = |K_{in}K_{out}|$ *для* $k = 3$ *(а),* $k = 1$ *(б),* $k = 0.3$ *(в),* $k = 0.1$ *(г)*

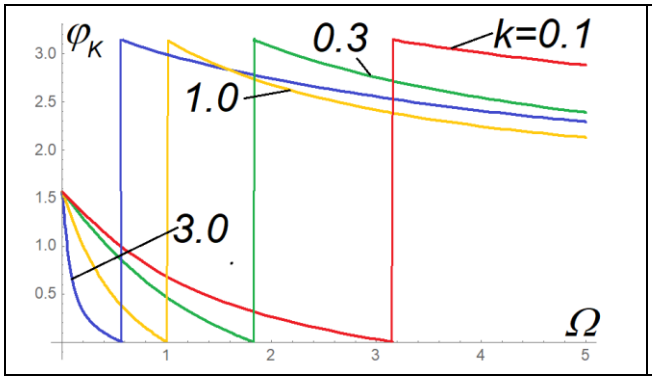


Рис. 4.13. Графіки фазово-частотних характеристик смугового RC-фільтра

Отримані результати добре узгоджуються з результатом числового розрахунку для цього випадку (крива 3 на рис. 4.1.10 б).

Асимптотику АЧХ при $\Omega \to \infty$ видно з формули (4.1.18): $|K(\Omega)| \approx (k\Omega)^{-1}$, тобто це фільтр першого порядку.

Як бачимо, навіть для порівняно простої схеми, поданої на рис. 4.1.10 а, аналітичний розрахунок виявляється досить громіздким, і для отримання результатів доводиться використовувати наближені методи.

4.1.5. Смугові та загороджувальні фільтри на базі коливних контурів

Одиночний контур. Очевидно, високодобротні контури (в розглянутому вище прикладі – контури з $\omega_0\tau \ll 1$) можуть використовуватися як смугові фільтри. Як випливає з формули (4.1.8)

$$\left|K(\omega)\right| = \frac{1}{\sqrt{\left(1-\omega^2/\omega_0^2\right)^2 + (\omega\tau)^2}}$$

(де, як і раніше, $\omega_0 = 1/\sqrt{LC}$, $\tau = L/R$), середня частота такого фільтру – ω_0 (рис. 4.1.14), а смуга пропускання знаходиться з умови

$$\frac{K^2(\omega)}{K^2(\omega_0)} = \frac{1}{2}, \tag{4.1.22}$$

тобто

$$\left(1-\frac{\omega^2}{\omega_0^2}\right)^2 + (\omega\tau)^2 = 2(\omega_0\tau)^2, \tag{4.1.23}$$

або

$$\left(1-\frac{\omega^2}{\omega_0^2}\right)^2 + \left(\frac{\omega^2}{\omega_0^2}-2\right)(\omega_0\tau)^2 = 0. \tag{4.1.24}$$

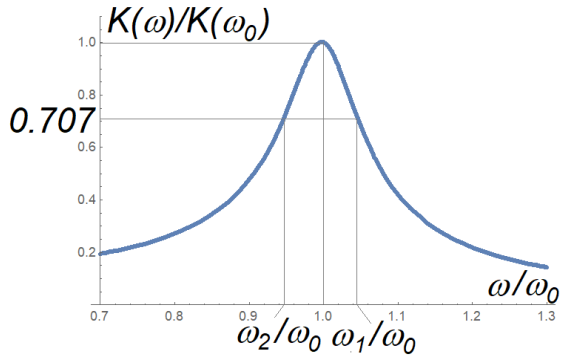


Рис. 4.1.14. Графік амплітудно-частотної характеристики смугового фільтру на основі послідовного коливного контуру

Перепишемо це рівняння у формі

$$\left(\frac{\omega^2}{\omega_0^2}\right)^2 - \frac{\omega^2}{\omega_0^2}\left[2-(\omega_0\tau)^2\right] + \left[1-2(\omega_0\tau)^2\right] = 0. \tag{4.1.25}$$

Корені цього рівняння відповідають мажам смуги пропускання:

$$\begin{aligned}\omega_{1,2}^2/\omega_0^2 &= \frac{1}{2}\left\{\left[2-(\omega_0\tau)^2\right] \pm \sqrt{\left[2-(\omega_0\tau)^2\right]^2 - 4\left[1-2(\omega_0\tau)^2\right]}\right\} = \\ &= \frac{1}{2}\left\{\left[2-(\omega_0\tau)^2\right] \pm \sqrt{\left[4+(\omega_0\tau)^2\right](\omega_0\tau)^2}\right\} \approx \frac{1}{2}\left\{\left[2-(\omega_0\tau)^2\right] \pm 2\omega_0\tau\right\} \approx 1 \pm \omega_0\tau\end{aligned} \tag{4.1.26}$$

(враховано, що $\omega_0\tau \ll 1$).

Нехай $\omega_{1,2} = \omega_0 \pm \Delta\omega$, де $\Delta\omega$ – півширина смуги пропускання. Тоді $\omega_{1,2}^2 \approx \omega_0^2\left(1 \pm 2\Delta\omega/\omega_0\right)$, і з формули (4.1.26) випливає, що

$$\Delta\omega = \frac{1}{2}\omega_0^2\tau. \tag{4.1.27}$$

Резонансні контури є найпоширенішими в аналоговій радіоелектроніці смуговими фільтрами. Але їхній недолік – це форма АЧХ, яка більше схожа на трикутник, ніж на прямокутник (рис. 4.1.8, крива 1).

Зв'язані контури. Поліпшити форму АЧХ смугового фільтра можна, використавши зв'язані контури.

Для контурів із ємнісним зв'язком, схема яких подана на рис. 4.1.15, можна записати систему рівнянь за II правилом Кірхгофа. Вважаючи зовнішню напругу гармонічною, можна отримати такі рівняння для комплексних амплітуд:

$$\begin{cases} i_{m1}\left(R_1 + i\omega L_1 + \dfrac{1}{i\omega C_1}\right) - \dfrac{i_{m2}}{i\omega C_c} = e_m; \\ -\dfrac{i_{m1}}{i\omega C_c} + i_{m2}\left(R_2 + i\omega L_2 + \dfrac{1}{i\omega C_2}\right) = 0, \end{cases} \qquad \frac{1}{C_{1,2}} \equiv \frac{1}{C'_{1,2}} + \frac{1}{C_c}. \tag{4.1.28}$$

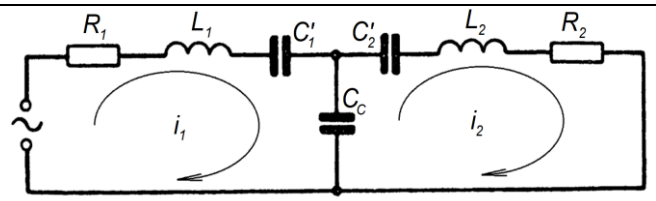


Рис. 4.1.15. Схема коливних контурів з ємнісним зв'язком

Для найцікавішого випадку ідентичних контурів ($R_1 = R_2 = R$, $L_1 = L_2 = L$, $C_1 = C_2 = C$) можна отримати такий вираз для передавальної функції (вважаємо, що вихідний сигнал знімається з ємності або з індуктивності другого контуру):

$$\frac{\rho i_{m2}}{e_m} = \frac{ikx}{\left\{2\Delta x + i\left[x^2 - (1+k)\right]\right\}\left\{2\Delta x + i\left[x^2 - (1-k)\right]\right\}}, \tag{4.1.29}$$

$$x = \frac{\omega}{\omega_0}, \quad k = \frac{C}{C_c}, \quad \Delta = \frac{\delta}{\omega_0}, \quad \delta = \frac{R}{2L}, \quad \omega_0 = \frac{1}{\sqrt{LC}}, \quad \rho = \sqrt{\frac{L}{C}}. \tag{4.1.30}$$

З формули випливає асмптотика для нормованого струму в другому контурі вигляду x^{-3}. Іншими словами, якщо сигнал знімається з резистора другого контуру, отримуємо фільтр третього порядку.

У цьому випадку можна отримати передавальну характеристику, форма якої при оптимальному підборі параметра зв'язку k ближча до прямокутника (рис. 4.1.16). Щоб переконатися в цьому, знайдемо максимальну амплітуду струму в другому контурі на частоті ω_0 (тобто при x=1). Маємо:

$$\left.\frac{\rho i_{m2}}{e_m}\right|_{x=1} = \frac{ik}{(2\Delta - ik)(2\Delta + ik)} = \frac{ik}{4\Delta^2 + k^2}.$$

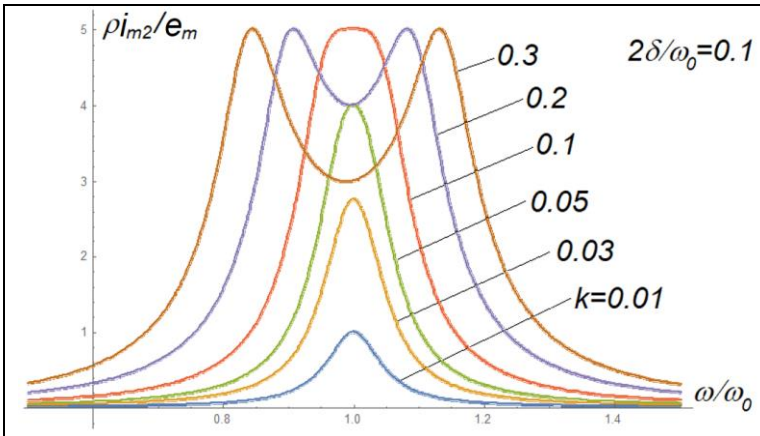


Рис. 4.1.16. Графік передавальної функції зв'язаних контурів

Прирівняємо до нуля похідну від цієї величини:

$$\left|\frac{d}{dk}\frac{\rho i_{m2}}{e_m}\right|_{x=1} = \frac{4\Delta^2 + k^2 - k\cdot 2k}{\left(4\Delta^2 + k^2\right)^2} = 0,$$

звідки

$$k_{opt} = 2\Delta\,, \qquad \left(\left|\frac{\rho i_{m2}}{e_m}\right|_{x=1}\right)_{\max} = \frac{1}{4\Delta}\,.$$

Коротко пояснимо вигляд кривих, показаних на цьому рисунку. Величина k, що служить параметром сім'ї кривих, називається коефіцієнтом зв'язку між контурами. Зі схеми (рис. 4.1.12) видно, що при $k \to 0$, тобто при $C_c \to \infty$, зв'язок між контурами зникає, і струм у другому контурі відсутній. Відповідно при малих k зростання k приводить до збільшення струму в другому контурі.

При великих k (точніше, при $k > k_{opt}$) спостерігається явище розштовхування власних частот[3]. За відсутності дисипації та за ідентичності контурів практично за будь-яких початкових умов між контурами відбувається періодичний обмін енергією, який можна інтерпретувати як биття між коливаннями з близькими частотами. Як видно з рис. 4.1.16, розштовхування частот зростає зі збільшенням k. Цей ефект придушується дисипацією, тому він не проявляється при $k < k_{opt}$.

<u>Т-подібний смуговий фільтр на основі коливних контурів</u>. Ще кращий вигляд передавальних характеристик мають фільтри, сконструйовані з послідовних та паралельних коливних контурів (рис. 4.1.17).

Щоб зрозуміти, як працюють ці схеми, спершу розглянемо еквівалентний опір послідовного та паралельного контурів з урахуванням їхнього активного опору (рис. 4.1.18). Вважатимемо, що виконано умову $R \ll \rho \equiv \sqrt{L/C}$.

Для послідовного контуру (рис. 4.1.18 а) повний імпеданс складає величину

$$Z_1(\omega) = \frac{1}{i\omega C} + i\omega L + R = i\left(\omega L - \frac{1}{\omega C}\right) + R\,. \qquad (4.1.31)$$

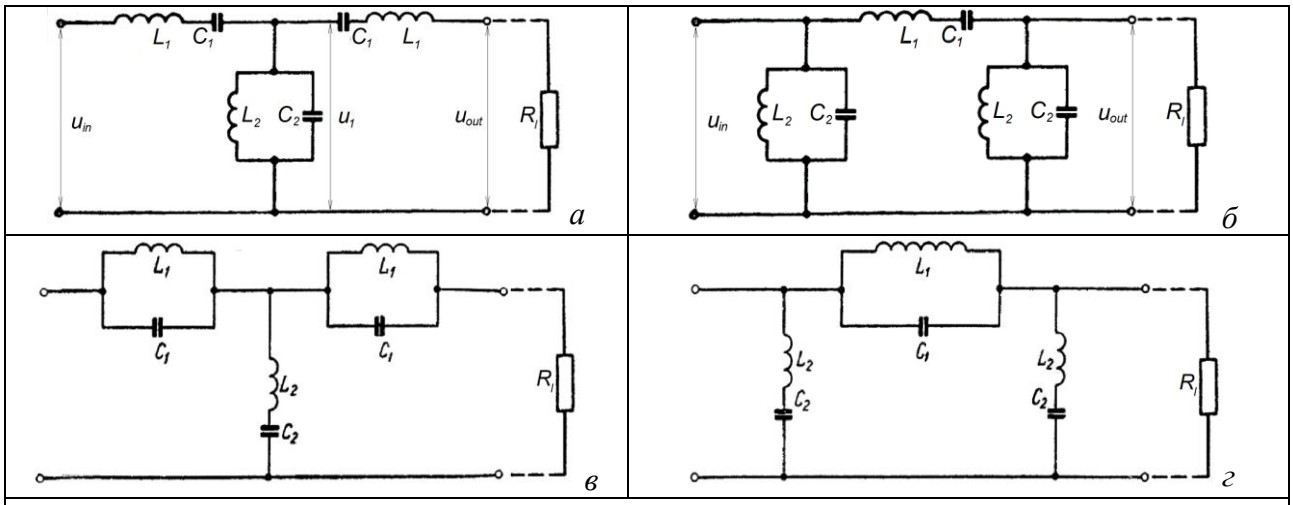


Рис. 4.1.17. Т-подібна (а, в)) та П-подібна (б, г) ідеалізовані схеми смугового (а, б) та загороджувального (в, г) фільтрів

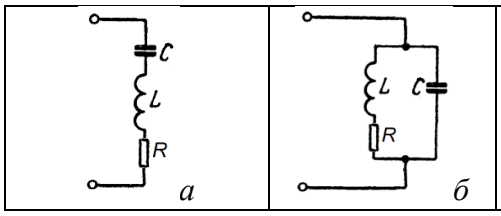


Рис. 4.1.18. Послідовний (а) та паралельний (б) коливні контури з урахуванням їхнього активного опору

[3] Детальніше див., наприклад: Анісімов І.О. Коливання та хвилі. К., КНУ, 2009.

На резонансній частоті $\omega_0 = 1/\sqrt{LC}$

$$Z_1(\omega_0) = R, \tag{4.1.32}$$

тобто еквівалентний опір ідеального ($R = 0$) послідовного контуру прямує до нуля.

Імпеданс паралельного контуру (рис. 4.1.18 б):

$$Z_2(\omega) = \left(i\omega C + \frac{1}{i\omega L + R} \right)^{-1} = \frac{i\omega L + R}{1 - \omega^2 LC + i\omega RC}. \tag{4.1.33}$$

На резонансній частоті

$$Z_2(\omega_0) = \frac{i\omega_0 L + R}{i\omega_0 RC} = \frac{L}{RC} - \frac{i}{\omega_0 C} = \rho\left(\frac{\rho}{R} - i \right) \approx \frac{\rho^2}{R}, \tag{4.1.34}$$

тобто еквівалентний опір ідеального паралельного контуру прямує до нескінченності.

Нехай послідовні та паралельні контури (рис. 4.1.17) налаштовані на однакову частоту. Тоді на цій частоті в смугових фільтрах (рис. 4.1.17 а, б) сигнал проходить на вихід, а в загороджувальних (рис. 4.1.17 в, г) – не проходить.

Для навантаженого Т-подібного смугового фільтра (рис. 4.1.17 а) коефіцієнт передачі можна записати у вигляді (4.1.10):

$$K(\omega) \equiv \frac{u_{out}}{u_{in}} = \frac{u_1}{u_{in}} \frac{u_{out}}{u_1} = K_{in}(\omega) K_{out}(\omega),$$

$$K_{out}(\omega) = \frac{R_l}{R_l + Z_1(\omega)}, \qquad K_{in}(\omega) = \frac{\left\{ 1/Z_2(\omega) + 1/\left[R_l + Z_1(\omega) \right] \right\}^{-1}}{Z_1(\omega) + \left\{ 1/Z_2(\omega) + 1/\left[R_l + Z_1(\omega) \right] \right\}^{-1}}, \tag{4.1.35}$$

Для імпедансів $Z_{1,2}(\omega)$ скористаємося формулами (4.1.31) та (4.1.33), тобто індуктивності та ємності послідовного й паралельного контурів вважатимемо однаковими; братимемо до уваги власний опір контурів R. Для спрощення вважатимемо, що цей опір однаковий у паралельного та послідовного контурів (рис. 4.1.18 а, б). Тоді маємо:

$$K_{out}(\omega) = \frac{R_l}{R_l + i\left(\omega L - \dfrac{1}{\omega C} \right) + R} = \frac{\omega C R_l}{\omega C(R_l + R) + i\left(\omega^2 LC - 1 \right)} =$$

$$= \frac{(\omega/\omega_0)\omega_0 C R_l}{(\omega/\omega_0)(\omega_0 C R_l + \omega_0 C R) + i\left(\omega^2/\omega_0^2 - 1 \right)} = \frac{\alpha x}{x(\alpha + \beta) + i\left(x^2 - 1 \right)}, \tag{4.1.36}$$

$$\omega_0^2 = \frac{1}{LC}, \qquad x = \frac{\omega}{\omega_0}, \qquad \alpha = \omega_0 R_l C = \frac{R_l}{\rho}, \qquad \beta = \omega_0 RC = \frac{R}{\rho} = \frac{1}{Q}. \tag{4.1.37}$$

$$K_{in}(\omega) = \frac{\left[1/Z_2(\omega) + 1/\left[R_l + Z_1(\omega) \right] \right]^{-1}}{Z_1(\omega) + \left[1/Z_2(\omega) + 1/\left[R_l + Z_1(\omega) \right] \right]^{-1}} = \frac{Z_2(\omega)\left[R_l + Z_1(\omega) \right] / \left[R_l + Z_1(\omega) + Z_2(\omega) \right]}{Z_1(\omega) + Z_2(\omega)\left[R_l + Z_1(\omega) \right] / \left[R_l + Z_1(\omega) + Z_2(\omega) \right]} =$$

$$= \frac{Z_2(\omega)\left[R_l + Z_1(\omega) \right]}{Z_1(\omega)\left[R_l + Z_1(\omega) + Z_2(\omega) \right] + Z_2(\omega)\left[R_l + Z_1(\omega) \right]} == \frac{Z_2(\omega)\left[R_l + Z_1(\omega) \right]}{\left[R_l + Z_1(\omega) \right]\left[Z_1(\omega) + Z_2(\omega) \right] + Z_1(\omega) Z_2(\omega)}.$$

Врахувавши співвідношення (4.1.31), (4.1.34) та (4.1.37), отримаємо:

$$K_{in}(\omega) = \frac{i\omega^2 LC + R\omega C}{1 - \omega^2 LC + i\omega RC}\left[(R_l + R)\omega C + i\left(\omega^2 LC - 1 \right) \right] \times$$

$$\times \left\{ \left[(R_l + R)\omega C + i\left(\omega^2 LC - 1 \right) \right] \left[i\left(\omega^2 LC - 1 \right) + R\omega C + \frac{i\omega^2 LC + R\omega C}{1 - \omega^2 LC + i\omega RC} \right] + \right.$$

$$+\left[i\left(\omega^2 LC-1\right)+R\omega C\right]\frac{i\omega^2 LC+R\omega C}{1-\omega^2 LC+i\omega RC}\Bigg\}^{-1}=$$

$$=\frac{\dfrac{x^2-ix\beta}{i\left(x^2-1\right)+x\beta}\left[x\left(\alpha+\beta\right)+i\left(x^2-1\right)\right]}{\left[x\left(\alpha+\beta\right)+i\left(x^2-1\right)\right]\left[i\left(x^2-1\right)+x\beta+\dfrac{x^2-ix\beta}{i\left(x^2-1\right)+x\beta}\right]+\left[i\left(x^2-1\right)+x\beta\right]\dfrac{x^2-ix\beta}{i\left(x^2-1\right)+x\beta}}=$$

$$=\frac{\left(x^2-ix\beta\right)\left[x\left(\alpha+\beta\right)+i\left(x^2-1\right)\right]}{\left[x\left(\alpha+\beta\right)+i\left(x^2-1\right)\right]\left\{\left[i\left(x^2-1\right)+x\beta\right]^2+\left(x^2-ix\beta\right)\right\}+\left[i\left(x^2-1\right)+x\beta\right]\left(x^2-ix\beta\right)}=$$

$$=\frac{\left(x^2-ix\beta\right)}{\left\{\left[i\left(x^2-1\right)+x\beta\right]^2+\left(x^2-ix\beta\right)\right\}+\left(x^2-ix\beta\right)} \qquad (4.1.38)$$

Таким чином,

$$K\left(x\right)=K_{in}\left(\omega\right)K_{out}\left(\omega\right)=$$

$$=\frac{\left(x^2-ix\beta\right)}{\left\{\left[i\left(x^2-1\right)+x\beta\right]^2+\left(x^2-ix\beta\right)\right\}+\left[i\left(x^2-1\right)+x\beta\right]\left(x^2-ix\beta\right)}\frac{x\alpha}{\left[x\left(\alpha+\beta\right)+i\left(x^2-1\right)\right]}. \qquad (4.1.39)$$

Асимптотика цієї залежності при $x\to\infty$ буде $\left|K\left(x\right)\right|=\alpha x^{-3}$, тобто це фільтр третього порядку. Сім'я його АЧХ (для різних значень β; навантаження обирається узгодженим, тобто $\alpha=1$) подана на рис. 4.1.19 а. Видно, що при малих t АЧХ в смузі прозорості містить три локальних максимуми, тобто форма АЧХ ближче до прямокутної, ніж у зв'язаних контурів.

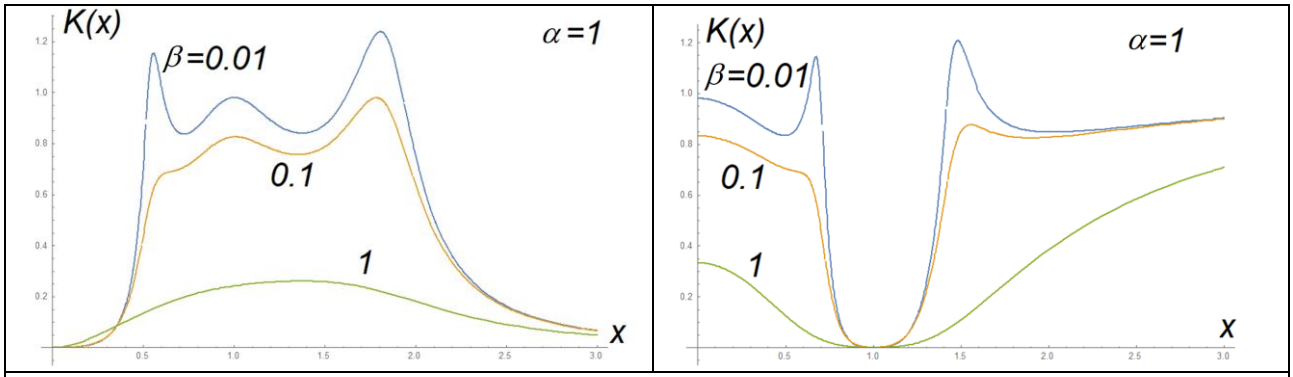


Рис. 4.1.19. Графіки АЧХ смугового (а) та загороджувального (б) навантажених Т-подібних фільтрів

<u>Т-подібний загороджувальний фільтр на основі коливних контурів.</u> Аналогічний розрахунок для навантаженого Т-подібного загороджувального фільтру (рис. 4.1.17 в) дає:

$$K\left(\omega\right)=\frac{R_l}{R_l+Z_2}\frac{Z_1\left(R_l+Z_2\right)/\left(Z_1+R_l+Z_2\right)}{Z_2+Z_1\left(R_l+Z_2\right)/\left(Z_1+R_l+Z_2\right)}=\frac{R_l}{R_l+Z_2}\frac{Z_1\left(R_l+Z_2\right)}{Z_2\left(Z_1+R_l+Z_2\right)+Z_1\left(R_l+Z_2\right)}=$$

$$=\frac{Z_1R_l}{Z_2\left(Z_1+R_l+Z_2\right)+Z_1\left(R_l+Z_2\right)}=\frac{Z_1R_l}{Z_1Z_2+\left(Z_1+Z_2\right)\left(R_l+Z_2\right)}. \qquad (4.1.40)$$

Підставивши Z_1 та Z_2 з формул (4.1.31) та (4.1.33) і вважаючи справедливими ті самі припущення, що й у попередньому випадку, отримуємо:

$$K(\omega)=R_l\left[R+\frac{i}{\omega C}\left(\omega^2LC-1\right)\right]\times\left\{\left[R+\frac{i}{\omega C}\left(\omega^2LC-1\right)\right]\frac{i\omega L+R}{1-\omega^2LC+i\omega RC}+\right.$$

$$\left.+\left[R+\frac{i}{\omega C}\left(\omega^2LC-1\right)+\frac{i\omega L+R}{1-\omega^2LC+i\omega RC}\right]\left(R_l+\frac{i\omega L+R}{1-\omega^2LC+i\omega RC}\right)\right\}^{-1}=$$

$$=R_l\left(1-\omega^2LC+i\omega RC\right)\left\{\left(1-\omega^2LC+i\omega RC\right)\frac{i\omega L+R}{1-\omega^2LC+i\omega RC}+\right.$$

$$\left.+\left[1-\omega^2LC+i\omega RC+\frac{-\omega^2LC+i\omega RC}{1-\omega^2LC+i\omega RC}\right]\left(R_l+\frac{i\omega L+R}{1-\omega^2LC+i\omega RC}\right)\right\}^{-1}=$$

$$=R_l\left(1-\omega^2LC+i\omega RC\right)^3\left\{\left(1-\omega^2LC+i\omega RC\right)^2\left(i\omega L+R\right)+\right. \tag{4.1.41}$$

$$\left.+\left[\left(1-\omega^2LC+i\omega RC\right)^2-\omega^2LC+i\omega RC\right]\left[R_l\left(1-\omega^2LC+i\omega RC\right)+i\omega L+R\right]\right\}^{-1}.$$

Скористаємося позначеннями (4.1.37):

$$K(x)=\frac{\left(1-x^2+itx\right)^3R_l}{\left(1-x^2+itx\right)^2\left(i\omega L+r\right)+\left[\left(1-x^2+itx\right)^2-x^2+itx\right]\left[R_l\left(1-x^2+itx\right)+i\omega L+r\right]}=$$

$$=\frac{\left(1-x^2+itx\right)^3ixT}{\left(1-x^2+itx\right)^2\left(-x^2+ixt\right)+\left[\left(1-x^2+itx\right)^2-x^2+itx\right]\left[ixT\left(1-x^2+itx\right)-x^2+ixt\right]}. \tag{4.1.50}$$

Сім’я АЧХ, що відповідають формулі (4.1.50), подана на рис. 4.1.19 б.

Поняття про активні фільтри. Частотні фільтри можуть бути складені з резисторів, конденсаторів та котушок індуктивності (пасивні фільтри). Але останнім часом для покращання форми АЧХ та компенсації втрат сигналу *RC*-фільтри стали комбінувати з підсилювальними каскадами, застосовуючи при цьому позитивний або негативний зворотний зв’язок. Такі фільтри одержали назву активних *RC*-фільтрів, або, як їх частіше називають, просто активних фільтрів. Як правило, активні фільтри будуються на основі операційних підсилювачів.

Контрольні питання до підрозділу 4.1

1. Які радіотехнічні кола можна вважати лінійними?
2. Опишіть основні типи частотних фільтрів (за виглядом смуги пропускання) та намалюйте їхні амплітудно-частотні характеристики (для ідеального випадку).
3. Що таке крутість зрізу фільтра? В яких одиницях її вимірюють?
4. Опишіть якісно роботу RC-фільтру низьких частот.
5. Опишіть якісно роботу LC-фільтру високих частот.
6. Опишіть якісно роботу LRC-фільтру низьких частот. Коли його амплітудно-частотна характеристика наближається до характеристики ідеального фільтру?
7. Порівняйте між собою фільтри нижніх частот на основі RC-ланцюжків, RL-ланцюжків та коливних контурів.
8. Запропонуйте схему загороджувального фільтра на основі RC-ланцюжків.
9. Яка схема смугового фільтра є, на вашу думку, найкращою?

Приклади розв'язання задач до підрозділу 4.1

П 4.1.1. Отримати АЧХ та ФЧХ зображеного на рис. 4.1.20 ланцюжка і побудувати їхні графіки.

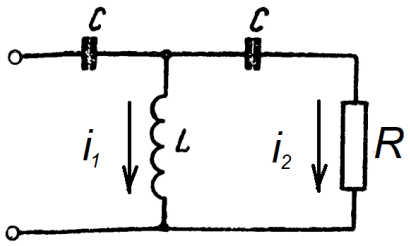


Рис. 4.1.20.

Розв'язок

Запишемо систему рівнянь за другим правилом Кірхгофа:

$$e_m = \left(i_{m1} + i_{m2}\right)\frac{1}{i\omega c} + i\omega L i_{m1}$$

$$i\omega L i_{m1} = i_{m2}\left(R + 1/i\omega C\right)$$

З другого рівняння

$$i_{m1} = i_{m2}\frac{1}{i\omega L}\left(R + 1/i\omega C\right)$$

Підставивши до першого, отримаємо:

$$e_m = i_{m2}\left(\frac{1}{i\omega L}\left(R + \frac{1}{i\omega C}\right) + 1\right)\frac{1}{i\omega C} + i_{m2}\left(R + \frac{1}{i\omega C}\right) = i_{m2}\left\{\frac{1}{i\omega C}\left[\frac{1}{i\omega L}\left(R + \frac{1}{i\omega C}\right) + 1\right] + \left(R + \frac{1}{i\omega C}\right)\right\}$$

$$i_{m2} = \frac{e_m}{\left\{\frac{1}{i\omega C}\left[\frac{1}{i\omega L}\left(R + \frac{1}{i\omega C}\right) + 1\right] + \left(R + \frac{1}{i\omega C}\right)\right\}}$$

$$K(\omega) = \frac{i_{m2}R}{e_m} = \frac{R}{\left\{\frac{1}{i\omega C}\left[\frac{1}{i\omega L}\left(R + \frac{1}{i\omega C}\right) + 1\right] + \left(R + \frac{1}{i\omega C}\right)\right\}} =$$

$$= \frac{i\omega RC}{\left[\frac{1}{i\omega L}\left(R + \frac{1}{i\omega C}\right) + 1\right] + \left(1 + i\omega RC\right)} = \frac{i\omega RC}{\left[-\frac{1}{\omega^2 LC}\left(1 + i\omega RC\right) + 1\right] + \left(1 + i\omega RC\right)} =$$

$$= \frac{i\omega RC}{-\frac{1}{\omega^2 LC}\left(1 + i\omega RC\right) + 1 + \left(1 + i\omega RC\right)} = \frac{i\omega RC}{2 - \frac{1}{\omega^2 LC} + i\omega RC\left(1 - \frac{1}{\omega^2 LC}\right)} =$$

$$= \frac{i\omega\tau}{\left(2 - \omega_0^2/\omega^2\right) + i\omega\tau\left(1 - \omega_0^2/\omega^2\right)}$$

$$\omega_0^2 = \frac{1}{LC}, \quad \tau = RC, \quad x = \frac{\omega}{\omega_0}, \quad \omega = x\omega_0.$$

$$K(x) = \frac{i\omega_0\tau x}{\left(2 - 1/x^2\right) + i\omega_0\tau x\left(1 - 1/x^2\right)} = \frac{i\omega_0\tau x^3}{\left(2x^2 - 1\right) + i\omega_0\tau x\left(x^2 - 1\right)};$$

$$\left|K(\omega)\right| = \frac{\omega_0\tau x^3}{\sqrt{\left(2x^2 - 1\right)^2 + \left(\omega_0\tau\right)^2 x^2\left(x^2 - 1\right)^2}}.$$

$$\varphi(\omega) = arctg\frac{\left(2x^2 - 1\right)}{\omega_0\tau x\left(x^2 - 1\right)}.$$

Відповідні залежності подані на рис. 4.1.21.

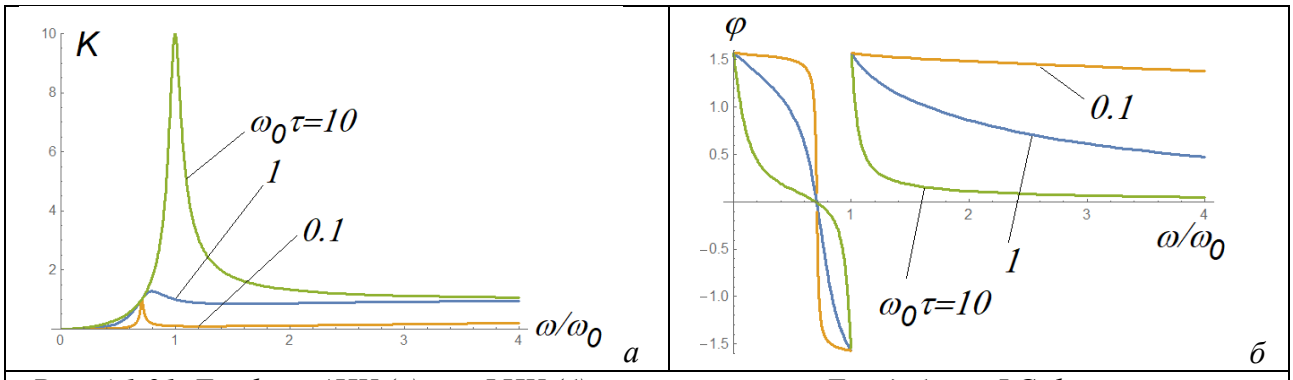


Рис. 4.1.21. Графіки АЧХ (а) та ФЧХ (б) навантаженого Т-подібного LC-фільтру верхніх частот

Зазначимо, що

$$\omega_0\tau = \frac{RC}{\sqrt{LC}} = \frac{R}{\sqrt{L/C}} = \frac{R}{\rho}, \quad \rho = \sqrt{L/C}\,.$$

Тоді умова $\omega_0\tau = 10$ означає, що $R \gg \rho$, тобто це практично ненавантажений режим. Іншими словами, $I_{m2} \to 0$, і вихідна напруга – це напруга на індуктивності послідовного коливного контуру. Справді, на кривій маємо різкий максимум на частоті $\omega = \omega_0$.

Навпаки, при $\omega_0\tau = 0.1$ маємо $R \ll \rho$, тобто вихід майже закорочується (рис. 4.1.22). При цьому вираз для АЧХ набуває вигляду

$$|K(\omega)| = \frac{\omega_0\tau x^3}{|2x^2 - 1|}\,.$$

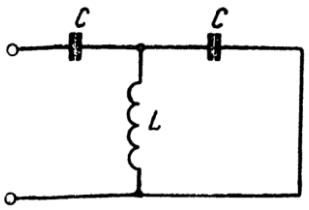


Рис. 4.1.22. Схема Т-подібного LC-фільтру верхніх частот у режимі короткого замкнення на виході

Отже, максимум коефіцієнта передачі відповідає резонансу в такому «послідовно-паралельному» контурі на частоті $\omega = \omega_0/\sqrt{2}$.

Нарешті, при $\omega_0\tau = 1$ досягається режим узгодження, $R = \rho$. При цьому вигляд АЧХ найближчий до ідеалізованого фільтра низьких частот.

П 4.1.2. Отримати АЧХ та ФЧХ зображеного на рис. 4.1.23 ланцюжка і побудувати їхні графіки.

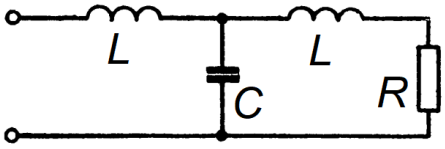


Рис. 4.1.23

Розв'язок

Дана схема відрізняється від розглянутої в попередній задачі заміною ємностей на індуктивності та індуктивностей на ємності. Тому можна взяти готову формулу для коефіцієнта передачі і зробити в ній заміни $i\omega L \to 1/i\omega C$, $1/i\omega C \to i\omega L$. Отримаємо:

$$K(\omega)=\frac{i_{m2}R}{e_m}=\frac{R}{\left\{i\omega L\left[i\omega C\left(R+i\omega L\right)+1\right]+\left(R+i\omega L\right)\right\}}=$$

$$=\frac{R/i\omega L}{i\omega L\left[i\omega C\left(R/i\omega L+1\right)+1/i\omega L\right]+\left(R/i\omega L+1\right)}=$$

$$=\frac{1/i\omega\tau}{-\omega^2/\omega_0^2\left(1+1/i\omega\tau\right)+1+\left(1+1/i\omega\tau\right)}=\frac{1}{-\omega^2/\omega_0^2\left(1+i\omega\tau\right)+i\omega\tau+\left(1+i\omega\tau\right)}=$$

$$=\frac{1}{\left(1-\omega^2/\omega_0^2\right)+i\omega\tau\left(2-\omega^2/\omega_0^2\right)}=\frac{1}{\left(1-x^2\right)+i\omega_0\tau x\left(2-x^2\right)}$$

$$\omega_0^2=\frac{1}{LC},\qquad \tau=\frac{L}{R},\qquad\qquad x=\frac{\omega}{\omega_0}.$$

$$\left|K(x)\right|=\frac{1}{\sqrt{\left(1-x^2\right)^2+\left(\omega_0\tau x\right)^2\left(2-x^2\right)^2}}.$$

$$\varphi(x)=-arctg\frac{\omega_0\tau x\left(2-x^2\right)}{\left(1-x^2\right)}.$$

Графіки цих залежностей подані на рис. 4.1.24.

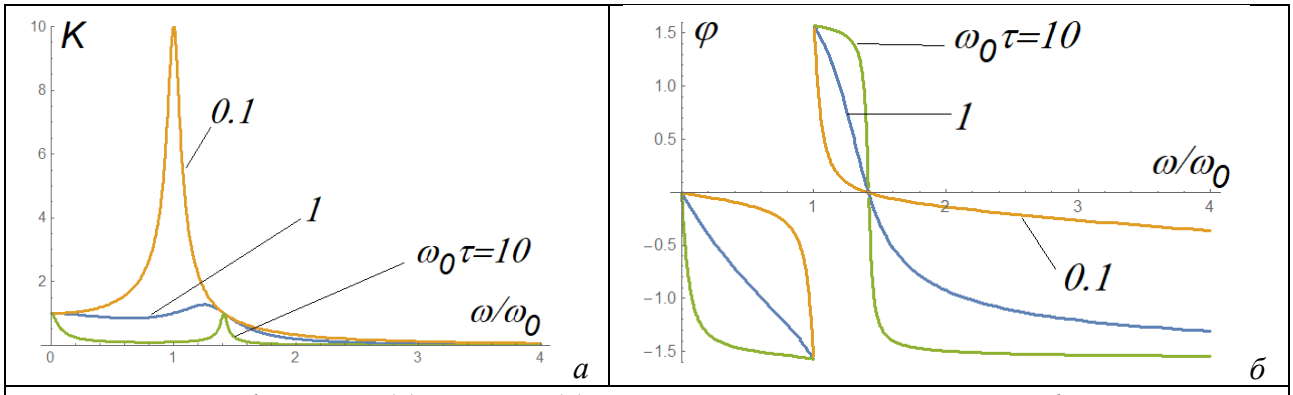


Рис. 4.1.24. Графіки АЧХ (а) та ФЧХ (б) навантаженого Т-подібного LC-фільтру нижніх частот

Для даної схеми

$$\omega_0\tau=\frac{L}{R\sqrt{LC}}=\frac{\rho}{R}.$$

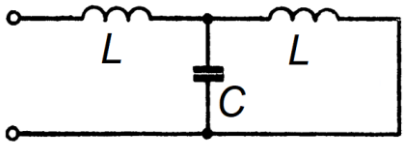
Отже, ненавантаженому режиму відповідає випадок $\omega_0\tau=0.1$. Вихідний сигнал при цьому знімається з ємності послідовного коливного контуру, і на частоті $\omega=\omega_0$ спостерігається резонанс.

При $\omega_0\tau=10$ маємо коротке замкнення на виході (див. рис. 4.1.25), а коефіцієнт передачі набуває вигляду

$$\left|K(x)\right|=\frac{1}{\omega_0\tau x\left|2-x^2\right|}$$

і досягає максимуму при $\omega=\sqrt{2}\omega_0$.

Нарешті, при $\omega_0\tau=1$ досягається режим узгодження, і фільтр найбільше уподібнюється до ідеального.

	*Рис. 4.1.25. Схема Т-подібного LC-фільтру нижніх частот у режимі короткого замкнення на виході*

Задачі до підрозділу 4.1

4.1.1. Отримати АЧХ та ФЧХ зображеного на рис. 4.1.3 б ланцюжка і побудувати їхні графіки.

4.1.2. Отримати АЧХ та ФЧХ зображеного на рис. 4.1.26 ланцюжка і побудувати їхні графіки.

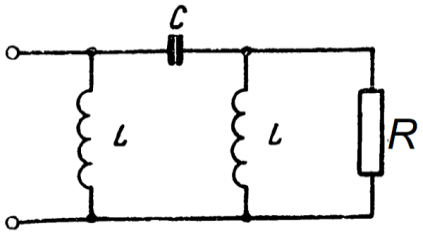


Рис. 4.1.26

4.1.3. Отримати АЧХ та ФЧХ зображеного на рис. 4.1.27 ланцюжка і побудувати їхні графіки.

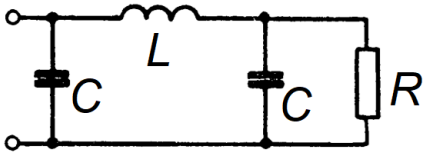


Рис. 4.1.27.

4.2. Підсилювачі на транзисторах: загальні питання

У цьому підрозділі ми розберемо загальні питання, що стосуються підсилювачів на польових та біполярних транзисторах: розглянемо загальний принцип дії підсилювача як керованого подільника напруги, проаналізуємо питання про роль навантаження, яке в тому чи іншому вигляді завжди присутнє на виході схеми підсилювача, визначимо основні параметри підсилювачів та наведемо їхню класифікацію[4].

4.2.1. Підсилювач на транзисторі як керований подільник напруги

Підсилювачі на транзисторах можна розглядати як керовані подільники напруги (рис. 4.2.1), складені зі сталого резистора R_l[5] (верхня ланка) та виходу транзистора (нижня ланка). З вихідних характеристик польових та біполярних транзисторів видно, що за допомогою вхідного сигналу (затворної напруги для польового транзистора та базового струму для біполярного) можна змінювати їхній вихідний омічний опір $R_o = U_{out}/I_{out}$ (нагадаємо, що сам термін «транзистор» – це абревіатура від англ. «змінюваний опір»). Оскільки напруга в середній точці подільника може бути записана як

$$U = E\frac{R_o}{R_l + R_o}, \qquad (4.2.1)$$

то зміна вихідного омічного опору транзистора супроводжуватиметься зміною цієї напруги. Якщо напруга живлення E значно перевищує за величиною сигнальну складову напруги на вході транзистора, то в результаті можна одержати підсилення сигналу за напругою[6].

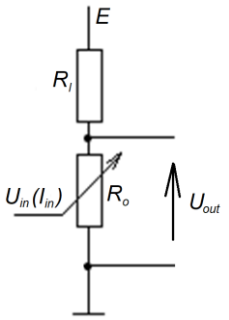


Рис. 4.2.1. До пояснення принципу роботи підсилювача

Розглянемо тепер цю саму задачу акуратніше – з урахуванням того, що напруга U є одночасно вихідною напругою транзистора U_{out}, яка також (як і вхідний сигнал) впливає на його вихідний струм I_{out}. Тому увімкнення транзистора до джерела живлення послідовно з резистором навантаження називають динамічним режимом транзистора.

Отже, з одного боку вихідні струм I_{out} та напруга U_{out} транзистора пов'язані між собою вихідними характеристиками. З іншого для кола напруга живлення – резистор – вихід транзистора – земля можна записати рівняння за другим правилом Кірхгофа

$$U_{out} = E - I_{out}R_l, \qquad (4.2.2)$$

яке також пов'язує між собою величини I_{out} та U_{out}. Графік залежності $I_{out}(U_{out})$, побудований відповідно до рівняння (4.2.2), називають навантажувадьною прямою. Навантажувальні прямі показані на рис. 4.2.2 червоними лініями. Задавши вхідний сигнал (напругу для польового транзистора чи струм для біполярного), ми тим самим однозначно задаємо стан транзистора, який відповідає так званій *робочій точці* – точці перетину навантажувальної прямої і відповідної вихідної характеристики. При зміні вхідного сигналу

---

[4] Ці міркування стосуються й підсилювачів на електронних лампах.

[5] Цей резистор називають резистором навантаження (індекс від англ. load, навантаження).

[6] Викладені тут міркування цілком застосовні й до підсилювачів на електронних лампах, пор. рис. 4.2.2 та рис.3.1.1 в-г.

робоча точка зсувається по навантажувальній прямій у той чи інший бік. При цьому вихідна напруга змінюється в часі за законом зміни вхідного сигналу. Якщо вхідний сигнал гармонічний, то фаза вихідного сигналу зсунута щодо вхідного на π.

Нагадаємо, що графічно омічний опір визначається кутом нахилу січної, проведеної з початку координат у робочу точку (див. вище п. 3.2.2): $R_o = U_{out} / I_{out} = ctg\,\alpha$. Таким чином, зміна кута α при зміні вхідного сигналу, показана на рис. 4.2.2, наочно ілюструє ефкт керування вихідним опором транзистора, про який говорилося раніше.

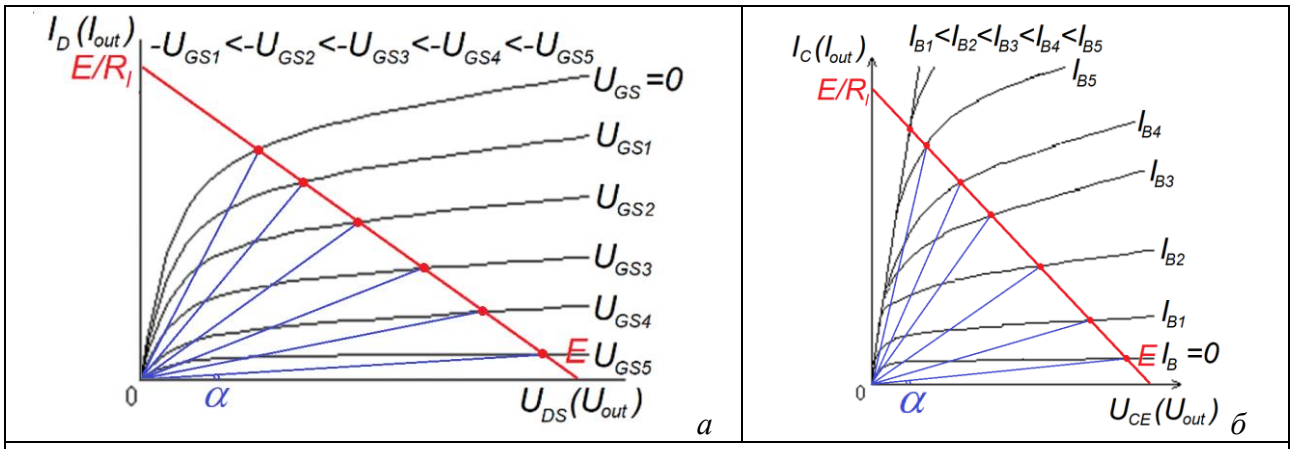


Рис. 4.2.2. Графіки сімей вихідних характеристик польового (а) та біполярного (б) транзисторів. Червоним показані навантажувальні прямі, синім – січні, проведені в робочі точки, котангенс кута нахилу яких до горизонталі відповідає вихідному омічному опору

Користуючись навантажувальною прямою, можна побудувати динамічну прохідну характкристику польового (біполярного) транзистора – залежність стокового (колекторного) струму від затвірної (базової) напруги в динамічному режимі. Її крутість буде меншою, ніж у статичних прохідних характристик, розглянутих у пп. 3.4.1-3.4.2 та 3.5.2.

Той факт, що напруга на шині живлення підтримується сталою, означає, що її можна вважати заземленою за сигналом (таке заземлення можна забезпечити, наприклад, увімкнувши між шиною живлення та заземленням конденсатор великої ємності). Тому за сигналом вихід транзистора і резистор R_l увімкнені паралельно.

4.2.2. Увімкнення навантаження до джерела гармонічного сигналу

В режимі малого сигналу ми представляли біполярний транзистор у вигляді еквівалентної схеми в *h*-параметрах (рис. 4.2.3, який повторює рис. 3.5.12; див. вище п. 3.5.4). Зараз нас цікавить вихідна частина цієї схеми, до якої приєднаний резистор навантаження R_l (рис. 4.2.4 а). Порівнюючи два рисунки, бачимо, що $i_{out} = i_C = h_{21E} i_B$, $R = 1/h_{22E}$.

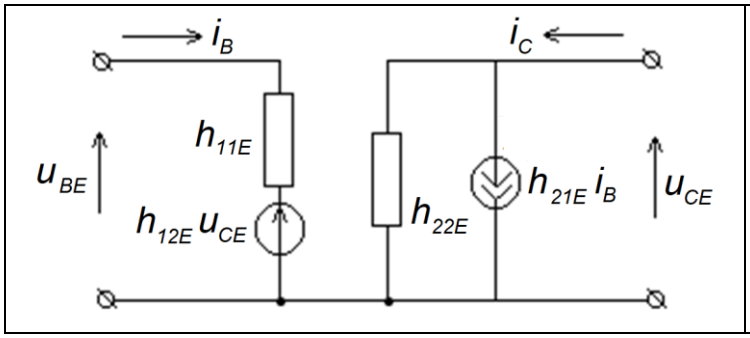
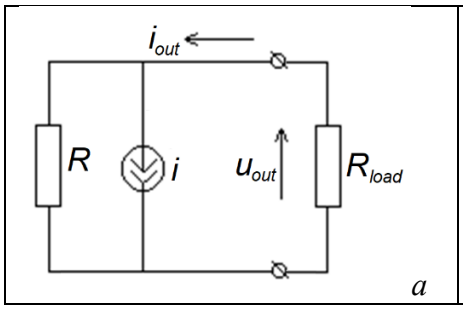


Рис. 4.2.3. Еквівалентна схема транзистора, увімкненого зі спільним емітером, в h-парметрах

Оскільки робочі ділянки вихідних характеристик польового та біполярного транзисторів подібні між собою і відрізняються лише характром керувального сигналу, то вихід польового транзистора, очевидно, також можна подати у вигляді еквівалентної схеми,

показаної на рис. 4.2.4 а. Тепер, очевидно, $i_{out} = i_D = Su_{GS}$, $R = R_{d\,out}$, де $R_{d\,out}$ – вихідний диференціальний опір польового транзистора, S – крутість його прохідної характеристики.

Як уже вказувалося, при ввімкненні транзистора до схеми підсилювача опір навантаження R_l (рис. 4.2.1) вмикається паралельно до виходу транзистора. В цьому випадку в схемі, поданій на рис. 4.2.4 а, під R_{load} слід розуміти паралельне сполучення R_l та зовнішнього навантаження.



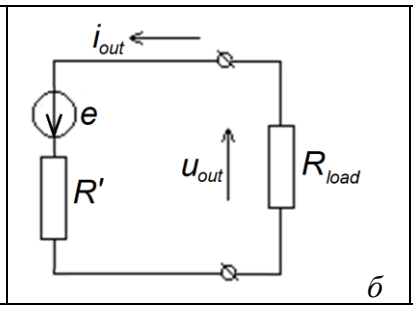


Рис. 4.2.4. Вихід транзистора як джерело сигналу (а – струму, б – напруги) з урахуванням його внутрішнього опору та з увімкненим навантаженням

Надалі нам буде зручніше користуватись схемою транзистора як джерела сигналу, поданою на рис. 4.2.4 б. Прирівнявши струми i_{out} в обох схемах, отримаємо співвідношення

$$i_{out} = i\frac{1/R_{load}}{1/R_{load} + 1/R} = \frac{iR}{R + R_{load}} = \frac{e}{R' + R_{load}} \qquad (4.2.3)$$

(враховано, що в схемі, поданій на рис. 4.2.1 а, струми через опори R_{load} та R пропорційні їхній провідності). Рівність (4.2.3) повинна задовольнятися при довільних значеннях R_{load}, тому з неї випливає, що

$$e = iR, \quad R' = R. \qquad (4.2.4)$$

Ми бачили, що на високих частотах треба брати до уваги вихідну ємність транзистора (див. вище пп. 3.4.1, 3.6), тобто розглядати не його активний вихідний опір, а його вихідний імпеданс. Для повноти картини вважатимемо, що навантаження також має реактивну складову. Приходимо до схеми, показаної на рис. 4.2.5.

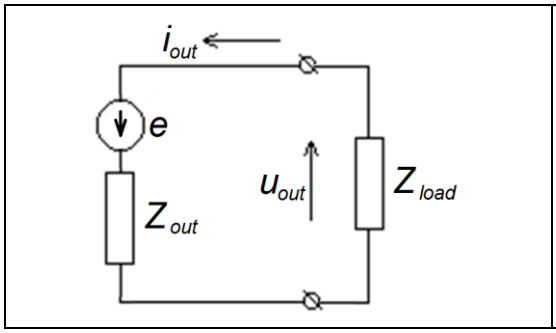


Рис. 4.2.5. Джерело сигналу з вихідним імпедансом та комплексним навантаженням

З'ясуємо, за яких умов у такій схемі досягається максимальний струм, максимальна вихідна напруга і максимальна потужність на навантаженні. Скористаємося позначеннями $Z_{out,load} = X_{out,load} + iY_{out,load}$.

Струм у колі складає

$$i_{out} = \frac{e}{Z_{out} + Z_{load}} = \frac{e}{X_{out} + X_{load} + i\left(Y_{out} + Y_{load}\right)}, \qquad (4.2.5)$$

$$\left|i_{out}\right| = \frac{e}{\sqrt{\left(X_{out} + X_{load}\right)^2 + \left(Y_{out} + Y_{load}\right)^2}}. \qquad (4.2.6)$$

Максимальна амплітуда струму

$$\left|i_{out\max}\right| = \frac{e}{X_{out}}$$

досягається за умов

$$Y_{load} = -Y_{out}, \qquad X_{load} = 0 \tag{4.2.7}$$

Першу з умов (4.2.7) можна назвати умовою компенсації ємнісного та індуктивного опорів. Якщо реактивні складові відсутні, то друга умова (4.2.7) означає звичайний режим короткого замикання на виході.

Вихідна напруга:

$$u_{out} = i_{out} Z_{load} = \frac{eZ_{load}}{Z_{out} + Z_{load}}, \tag{4.2.8}$$

$$\left|u_{out}\right| = \frac{e\sqrt{X_{load}^2 + Y_{load}^2}}{\sqrt{\left(X_{out} + X_{load}\right)^2 + \left(Y_{out} + Y_{load}\right)^2}} = \frac{e\sqrt{\left(X_{load}/X_{out}\right)^2 + \left(Y_{load}/X_{out}\right)^2}}{\sqrt{\left(1 + X_{load}/X_{out}\right)^2 + \left(Y_{out}/X_{out} + Y_{load}/X_{out}\right)^2}}. \tag{4.2.9}$$

За відсутності реактивних складових максимум $\left|u_{out}\right| = e$ досягається при $X_{load} \to \infty$, тобто в ненавантаженому режимі. За наявності реактивних складових максимум

$$\left|u_{out}\right| = \frac{e\sqrt{X_{load}^2 + Y_{out}^2}}{\sqrt{\left(X_{out} + X_{load}\right)^2}} = \frac{e\sqrt{\left(X_{load}/X_{out}\right)^2 + \left(Y_{out}/X_{out}\right)^2}}{\sqrt{\left(1 + X_{load}/X_{out}\right)^2}} \tag{4.2.10}$$

досягається за умови $Y_{load} = -Y_{out}$ (див. рис. 4.2.6) при $X_{load} \to \infty$ (якщо $Y_{out}/X_{out} < 1$) або при $X_{load} = 0$ (якщо $Y_{out}/X_{out} > 1$) – див. рис. 4.2.6.

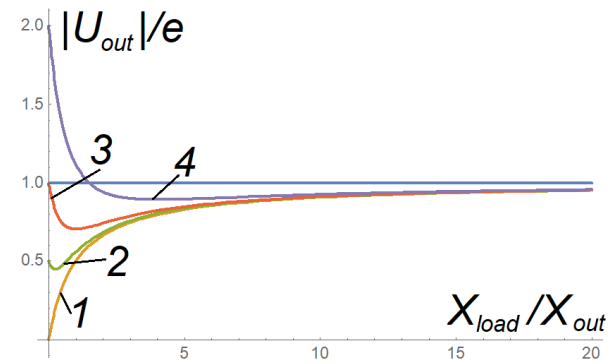


Рис. 4.2.6. Графіки залежності амплітуди вихідної напруги від активної складової опору навантаження в узгодженому режимі ($Y_{load} = -Y_{out}$): 1 – $Y_{out}/X_{out} = 0$; 2 – $Y_{out}/X_{out} = 0.5$; 3– $Y_{out}/X_{out} = 1$; 4 – $Y_{out}/X_{out} = 2$

Активна потужність (середня за період), що виділяється на навантаженні, відповідно до формули (2.1.20) складає

$$\bar{P} = \frac{1}{2}\mathrm{Re}\left\{i_{out} u_{out}^*\right\} = \frac{1}{2}\mathrm{Re}\left\{\frac{e}{\left[X_{out} + X_{load} + i\left(Y_{out} + Y_{load}\right)\right]} \frac{e\left(X_{load} - iY_{load}\right)}{\left[X_{out} + X_{load} - i\left(Y_{out} + Y_{load}\right)\right]}\right\} = \frac{1}{2}\frac{e^2 X_{load}}{\left(X_{out} + X_{load}\right)^2 + \left(Y_{out} + Y_{load}\right)^2} \tag{4.2.12}$$

і досягає максимуму

$$\bar{P}_{\max} = \frac{1}{8}\frac{e^2}{X_{load}} \tag{4.2.13}$$

за виконання так званих умов узгодження джерела з навантаженням;

$$Y_{load} = -Y_{out}, \qquad X_{load} = X_{out}. \tag{4.2.14}$$

Отримані результати справедливі для довільних джерел гармонічних сигналів, які можна подати у вигляді джерела е.р.с. із внутрішнім опором.

4.2.3. Основні параметри та різновиди підсилювачів

Як уже вказувалося, в режимі малого сигналу підсилювач можна розглядати як лінійний чотириполюсник, де в ролі вхідних та вихідних напруг і струмів виступають відповідні сигнальні складові (рис. 4.2.7).

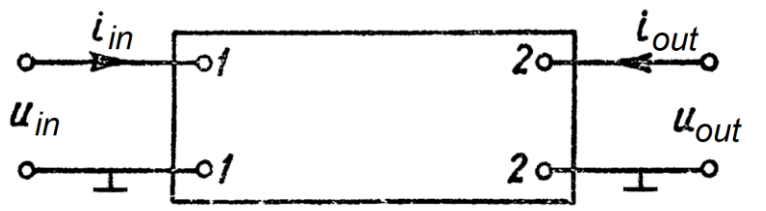


Рис. 4.2.7. Умовне зображення підсилювача як лінійного чотириполюсника та його вхідні й вихідні напруги й струми

Основними параметрами, які характеризують будь-який підсилювач, є його коефіцієнти підсилення:

- коефіцієнт підсилення за напругою;
- коефіцієнт підсилення за струмом;
- коефіцієнт підсилення за потужністю.

Коефіцієнтом підсилення за напругою називають відношення вихідної напруги до вхідної напруги:

$$K_u = \frac{u_{out}}{u_{in}}. \qquad (4.2.15)$$

Аналогічно коефіцієнт підсилення за струмом визначається як

$$K_i = \frac{i_{out}}{i_{in}}. \qquad (4.2.16)$$

Коефіцієнт підсилення за потужністю приймається відношення потужності, яка виділяється на резисторі навантаження до потужності, вживаної вхідним опором підсилювача:

$$K_P = \frac{\overline{P}_{out}}{\overline{P}_{in}} = \frac{\frac{1}{2} i_{out} u_{out}^*}{\frac{1}{2} i_{in} u_{in}^*} = K_i K_u^*. \qquad (4.2.17)$$

Іноді зручно виражати підсилення в децибелах:

$$K_u(\partial Б) = 20 \lg K_u, \qquad K_i(\partial Б) = 20 \lg K_i.$$

Оскільки потужність пропорційна квадрати напруги або струму, то

$$K_P(\partial Б) = 10 \lg K_P.$$

Як правило, підсилювач можна розглядати як деякий смуговий фільтр (див. вище п 4.1), причому в смузі прозорості коефіцієнт передачі набагато більший за одиницю. Тому важливими параметрами підсилювача є нижня і верхня межі смуги пропускання. У будь-якого підсилювача на високих частотах коефіцієнт підсилення зменшується через інерційність транзисторів. У більшості типів підсилювачів смуга пропускання обмежена й з боку нижніх частот.

На вхід підсилювача завжди вмикається джерело сигналу з певним вихідним опором, а на вихід – деяке зовнішнє навантаження (або наступний каскад). Тому важливими параметрами підсилювача є також його вхідний та вихідний опори. Для підсилювачів напруги вхідний опір бажано мати якомога більшим, а вихідний – якомога меншим, для підсилювачів потужності – забезпечити узгодження виходу із зовнішнім навантаженням (див. вище п. 4.2.2).

За тими задачами, що ставляться перед підсилювачами, вони поділяються на три основні групи:

- підсилювачі напруги;
- підсилювачі струму;

- підсилювачі потужності.

В залежності від діапазону підсилюваних ними частот підсилювачі поділяються на:

- *підсилювачі низької частоти* (ПНЧ). Такі підсилювачі забезпечують підсилення частот від кількох десятків герц до 10 - 20 кілогерц. Оскільки цей діапазон відповідає сприйманому людиною діапазону звукових коливань, то такі підсилювачі називають також підсилювачами звукових частот. Вони використовуються в електроакустичній апаратурі (радіомовних приймачах, програвачах, магнітофонах, тощо). Зв'язок між каскадами у них здебільшого ємнісний;

- *підсилювачі мінливої напруги та мінливого струму* (застарілий термін – підсилювачі постійної напруги та постійного струму; насправді йдеться про напруги й струми, які повільно змінюються у часі) забезпечують підсилення в діапазоні від нульової частоти до деякої верхньої граничної частоти f_h. Реально смуга пропускання таких підсилювачів простягається від кількох герц до кількох десятків і навіть сотень кілогерц. Такі підсилювачі знаходять своє застосування у вимірювальній радіоапаратурі, пристроях автоматики та обчислювальної техніки. Ними можна підсилювати як змінну, так і постійну складову сигналу. Міжкаскадний зв'язок в таких підсилювачах гальванічний;

- *широкосмугові підсилювачі*, які охоплюють дуже широкий діапазон частот – від кількох герц до кількох мегагерц. Такий широкий спектр частот притаманний коротким імпульсним сигналам та телевізійним відеосигналам. Тому такі підсилювачі ще називають імпульсними підсилювачами або відеопідсилювачами;

- *вибіркові (або селективні) підсилювачі*. Такі підсилювачі забезпечують підсилення у вузькій смузі частот і застосовуються там, де потрібно з широкого спектра частот виділити та підсилити сигнали заданої частоти. Отже, такі підсилювачі поєднують властивості власне підсилювачів і смугових фільтрів. Вибіркові підсилювачі застосовують у радіоприймачах, де вони виділяють і підсилюють лише смугу частот тієї радіостанції, на яку налаштований приймач, а також у спеціальній радіовимірювальній апаратурі.

Контрольні питання до підрозділу 4.2

1. Яка властивість транзистора лежить в основі роботи підсилювачів?
2. Що таке навантажувальна пряма?
3. Джерело сигналу приєднане до навантаження. За яких умов на навантаженні досягаються максимальні значення струму, напруги та потужності?
4. Що розуміють під узгодженням джерела сигналу з навантаженням?
5. Назвіть основні параметри підсилювачів.
6. Які ознаки кладуть в основу класифікації підсилювачів?

4.3. Зворотний зв'язок у підсилювачах

Про зворотний зв'язок[7] (англ. feedback) говорять, коли частина вихідного сигналу схеми подається на її вхід. Розрізняють зворотний зв'язок за напругою та за струмом – у залежності від того, яка з цих величин визначає сигнал зворотного зв'язку. Говорять також про послідовний та паралельний зворотний зв'язок. У першому випадку сигнал зворотного зв'язку вмикається до входу підсилювача послідовно з джерелом зовнішнього сигналу, в другому – паралельно до входу, через деякий елемент.

4.3.1. Послідовний зворотний зв'язок за напругою

Блок схема підсилювача з послідовним зворотним зв'язком за напругою подана на рис. 4.3.1.

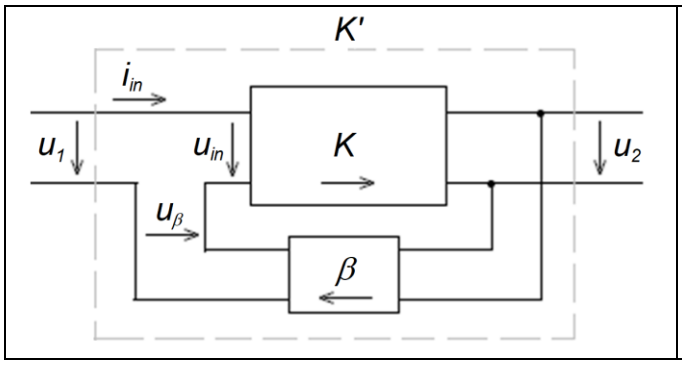


Рис. 4.3.1. Структурна схема підсилювача, охопленого послідовним зворотним зв'язком за напругою

Нехай підсилювач і чотириполюсник зворотного зв'язку характеризуються комплексними коефіцієнтами передачі $K = |K|\exp(i\varphi_K)$, $\beta = |\beta|\exp(i\varphi_\beta)$. Знайдемо коефіцієнт передачі $K' = u_2/u_1$ підсилювача, охопленого зворотним зв'язком.

Сигнал безпосередньо на вході підсилювача (рис. 4.3.1):

$$u_{in} = u_1 + u_\beta = u_1 + \beta u_2 . \qquad (4.3.1)$$

Але $u_2 = Ku_{in}$, тому

$$u_2 = Ku_1 + K\beta u_2 . \qquad (4.3.2)$$

Отже, шуканий коефіцієнт передачі має вигляд

$$K' = \frac{u_2}{u_1} = \frac{K}{1 - K\beta} = \frac{|K|\exp(i\varphi_K)}{1 - |K\beta|\exp(i\varphi)}, \qquad \varphi = \varphi_K + \varphi_\beta , \qquad (4.3.3)$$

а його абсолютне значення –

$$K' = \frac{K}{\sqrt{\left(1 - |K\beta|\cos\varphi\right)^2 + \left(|K\beta|\sin\varphi\right)^2}} = \frac{K}{\sqrt{1 - 2|K\beta|\cos\varphi + |K\beta|^2}} . \qquad (4.3.4)$$

4.3.2. Позитивний і негативний зворотний зв'язок

Позитивним зворотним зв'язком (ПЗЗ) називається випадок, коли $\varphi = 2n\pi$, де $n = 0, \pm 1, \pm 2...$ При цьому сигнал зворотного зв'язку u_β та вхідний сигнал u_1 синфазні, $K\beta > 0$ і ефективний коефіцієнт передачі дорівнює

$$K' = \frac{K}{1 - K\beta}, \qquad |K'| > |K| . \qquad (4.3.5)$$

Збільшення коефіцієнту підсилення за рахунок ПЗЗ називають регенеративним

[7] Зараз поняття зворотного зв'язку вийшло далеко за межі радіотехніки і використовується, наприклад, у суспільствознавстві (політологія, державне управління тощо).

підсиленням.

При $\varphi = (2n-1)\pi$ справджується умова $K\beta < 0$, сигнал зворотного зв'язку u_β протифазний до вхідного сигналу u_1, в результаті чого сигнал u_{in} ослаблюється порівняно з u_1. Такий зворотний зв'язок має назву негативного (НЗЗ). Ефективний коефіцієнт передачі підсилювача, охопленого НЗЗ дорівнює

$$K' = \frac{K}{1+|K\beta|}, \qquad |K'| < |K| \tag{4.3.6}$$

тобто він завжди менший від коефіцієнта підсилення підсилювача без зворотного зв'язку.

4.3.3. Вплив послідовного зворотного зв'язку за напругою на вхідний опір підсилювача

Вхідний опір підсилювача без зворотного зв'язку визначається виразом $Z_{in} = u_{in} / i_{in}$, а підсилювача, охопленого зворотним зв'язком – як $Z'_{in} = u_1 / i_{in}$. Підставивши до (4.3.1) вираз $u_2 = Ku_{in}$, дістанемо

$$u_1 = u_{in}\left(1-|K\beta|\right) \tag{4.3.7}$$

Отже, вхідний опір Z'_{in} дорівнює

$$Z'_{in} = \frac{u_{in}}{i_{in}}\left(1-|K\beta|\right) = Z_{in}\left(1-|K\beta|\right) \tag{4.3.8}$$

При позитивному зворотному зв'язку ($K\beta > 0$) ефективний вхідний опір Z'_{in} зменшується порівняно зі вхідним опором «чистого» підсилювача, а при негативному ($K\beta < 0$) – навпаки, збільшується:

$$Z'_{in} = Z_{in}\left(1+|K\beta|\right) \tag{4.38 а}$$

Останнє здебільшого буває вельми бажаним, оскільки підсилювач з великим вхідним опором зменшує частку вхідної сигнальної напруги, що падає на внутрішньому опорі її джерела (див. вище п. 4.2.2).

4.3.4. Стабілізація коефіцієнту підсилення

Внаслідок різноманітних причин (коливання температури, нестабільність напруги джерел живлення тощо) параметри транзисторів зазнають змін, які позначаються на величині коефіцієнта передачі K підсилювача. Зворотний зв'язок може посилити або, навпаки, ослабити цю нестабільність. Вважатимемо величини $K\beta$ дійсним. Розглянемо для прикладу випадок позитивного зворотного зв'язку. Вважаючи, що власний коефіцієнт підсилення K підсилювача змінюється, візьмемо диференціал від K' (4.3.5):

$$dK' = d\left(\frac{K}{1-K\beta}\right) = \frac{(1-K\beta)+K\beta}{(1-K\beta)^2}dK = \frac{dK}{(1-K\beta)^2}. \tag{4.3.9}$$

Замінимо диференціали скінченими приростами, які відповідають рівню пульсацій за наявності й за відсутності зворотного зв'язку, і поділимо обидві сторони (4.3.9) на K'. Отримаємо:

$$\frac{\Delta K'}{K'} = \frac{1}{1-K\beta}\frac{\Delta K}{K}, \tag{4.3.10}$$

тобто відносний рівень пульсацій із позитивним зворотним зв'язком буде в $1/(1-K\beta)$ разів більшим, ніж без нього.

Навпаки, при негативному зворотному зв'язку відносний рівень пульсацій зменшується в $1+|K\beta|$ разів:

$$\frac{\Delta K'}{K'} = \frac{1}{1+|K\beta|}\frac{\Delta K}{K}. \qquad (4.3.11)$$

Для глибокого НЗЗ, коли $K\beta >> 1$, вираз для K' набуває вигляду $K' = 1/\beta$, так що K' взагалі перестає залежати від коефіцієнта підсилення K. Оскільки коло зворотного зв'язку звичайно є пасивним чотириполюсником, коефіцієнт передачі якого не залежить від режиму і умов роботи підсилювача, то ефективний коефіцієнт підсилення K' вдається в цьому випадку підтримувати з високим ступенем стабільності, однак ціною значного зниження коефіцієнта підсилення (в $K\beta >> 1$ разів).

4.3.5. Вплив зворотного зв'язку на амплітудно-частотну характеристику

Негативний зворотний зв'язок стабілізує також коефіцієнт підсилення в робочому діапазоні частот. Як видно з формули (4.3.6), коефіцієнт підсилення зменшується більшою мірою на тих частотах, де його значення великі, і майже не змінюється на частотах, де його величина мала. В результаті амплітудно-частотна характеристика стає більш рівномірною, хоча в цілому розташована нижче такої ж характеристики за відсутності зворотного зв'язку (рис. 4.3.2). Верхня гранична частота при цьому підвищується, а нижня, навпаки, знижується і загалом смуга пропускання підсилювача розширюється.

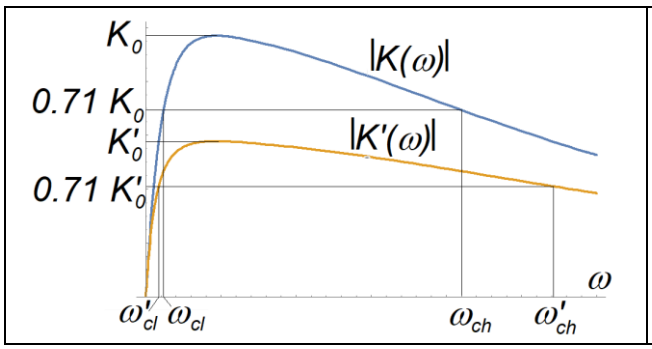


Рис. 4.3.2. Вплив негативного зворотного зв'язку на АЧХ підсилювача

Ці міркування можна кількісно підтвердити на прикладі однокаскадного підсилювача, частотна характеристика якого в області високих частот може бути подана у формі (детальніше про це див. нижче пп. 4.4.3, 4.5.4)

$$K(\omega) = -\frac{K_0}{1+i\,\omega/\omega_{ch}} \qquad (4.3.12)$$

де K_0 – коефіцієнт підсилення на низьких частотах, ω_{ch} – верхня гранична частота.

Якщо зворотний зв'язок не залежить від частоти, вираз для ефективного коефіцієнта підсилення (з урахуванням НЗЗ) набирає вигляду

$$K'(\omega) = \frac{K(\omega)}{1+\beta K(\omega)} = \frac{K_0/(1+i\,\omega/\omega_{ch})}{1+|\beta K_0|/(1+i\,\omega/\omega_{ch})} = \frac{K_0}{1+|\beta K_0|+i\,\omega/\omega_{ch}} =$$
$$= \frac{K_0/(1+|\beta_0 K_0|)}{1+i\omega/\left[\omega_{ch}(1+|\beta_0 K_0|)\right]} = \frac{K'_0}{1+i\,\omega/\omega'_{ch}}, \qquad K'_0 = \frac{K_0}{1+|\beta_0 K_0|}, \qquad \omega'_{ch} = \omega_{ch}(1+|\beta_0 K_0|). \qquad (4.3.13)$$

Видно, що верхня гранична частота за рахунок негативного зворотного зв'язку зросла в $1+\beta_0 K_0$ разів – ціною зменшення в таку саму кількість разів коефіцієнта підсилення на низьких частотах.

4.3.6. Самозбудження підсилювача, охопленого позитивним зворотним зв'язком. Критерій Найквіста

З формули (4.3.5)

$$K' = \frac{K}{1 - K\beta}$$

видно, що при наближенні знизу до одиниці добутку $K\beta$ коефіцієнт підсилення з урахуванням ПЗЗ необмежено зростає, $K' \to \infty$. Це означає, що при $K\beta = 1$ (реально при $K\beta > 1$) сигнал на виході схеми існує навіть за відсутності вхідного сигналу, тобто підсилювач перетворюється на генератор. Говорять, що відбувається самозбудження підсилювача.

Щоб відбулося самозбудження охопленого зворотним зв'язком підсилювача, необхідно і достатньо виконання двох умов:

а) фазової умови, що має вигляд

$$\varphi = \varphi_K + \varphi_\beta = 2\pi n \quad (n = 0, \pm 1, \pm 2, ...); \tag{4.3.14}$$

ця умова означає, що сигнал, який пройшов через підсилювач, коло зворотного зв'язку і повернувся назад на його вхід (див. рис. 4.3.1), повинен мати ту ж саму фазу, що і вхідний сигнал;

б) амплітудної умови, що має вигляд

$$K\beta > 1; \tag{4.3.15}$$

ця умова означає, що сигнал, який пройшов через підсилювач, коло зворотного зв'язку і повернувся на вхід, повинен бути хоч трохи більшим за первісний. Тоді після кожного обходу кола зворотного зв'язку початковий сигнал (у цій ролі звичайно виступають флуктуації) зростатиме в геометричній прогресії, тобто в $K\beta > 1$ разів. Якщо ж $K\beta < 1$, то прогресія буде спадною, і початковий сигнал буде згасати.

Самозбудження відбувається на тій частоті, для якої обидві умови (4.3.14) і (4.3.15) виконуються одночасно.

Альтернативний (щодо перевірки виконання фазової та амплітудної умов самозбудження) підхід до дослідження стійкості підсилювача, охопленого зворотним зв'язком, полягає в перевірці дотримання так званого критерію Найквіста. З цією метою для частот, на яких коефіцієнт підсилення відмінний від нуля, обчислюють значення комплексної величини

$$K(\omega)\beta(\omega) = |K(\omega)\beta(\omega)| \exp\left[i\left(\varphi_K(\omega) + \varphi_\beta(\omega)\right)\right] = |K(\omega)\beta(\omega)| \exp\left(i\varphi(\omega)\right). \tag{4.3.16}$$

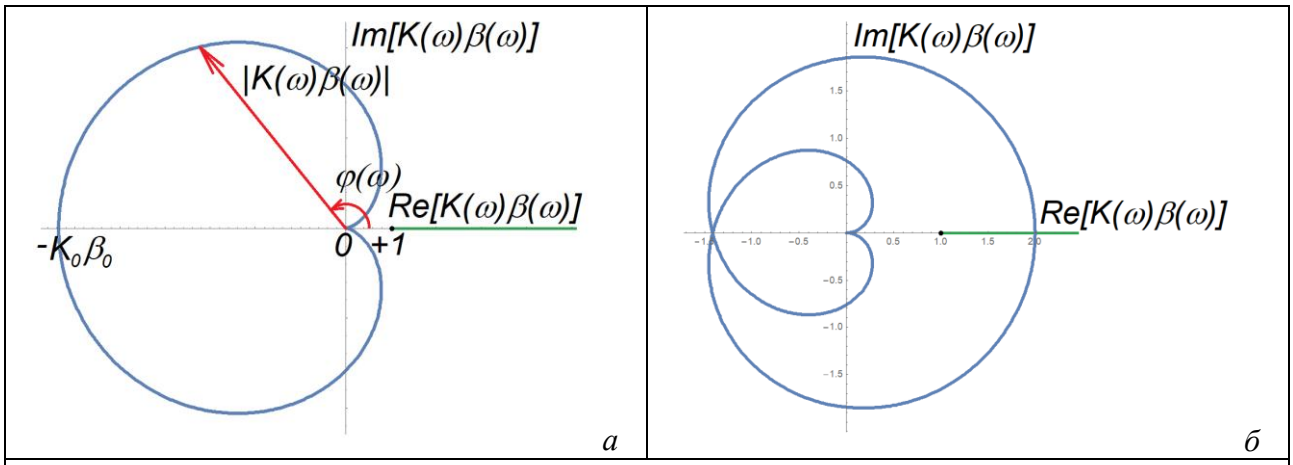


Рис. 4.3.3. До пояснення змісту критерію Найквіста (зеленим кольором показана область виконання умов самозбудження)

На комплексній площині будується лінія, що показує, як змінюється комплексна величина $K\beta$ в діапазоні частот від нуля до нескінченності (рис. 4.3.3). Отримана крива має назву діаграми Найквіста. Оскільки звичайно $K = 0$ при $\omega = 0$ та $\omega \to \infty$, то ця крива починається і закінчується в початку координат. Умовою відсутності самозбудження є критерій Найквіста: підсилювач буде стійким, якщо точка +1 на дійсній осі знаходиться за межами діаграми Найквіста (рис. 4.3.3 а), і нестійким у протилежному випадку (рис. 4.3.3 б).

Очевидно, виконання критерію Найквіста означає неможливість одночасного виконання фазової та амплітудної умов самозбудження.

4.3.7. Паралельний зворотний зв'язок за напругою

Щойно розглянутий (пп. 4.3.1-4.3.6) послідовний зворотний зв'язок за напругою в багатьох випадках (але не завжди) створюється штучно з метою керування властивостями охопленого ним підсилювача. На відміну від нього, паралельний зворотний зв'язок зазвичай виникає спонтанно (наприклад, у біполярних транзисторах та створених на їхній основі підсилювачах[8]).

Про паралельний зворотний зв'язок за напругою (рис. 4.3.4) говорять, коли вихід підсилювача приєднується паралельно до його входу через опір зворотного зв'язку R_f (індекс – від англ. feedback, зворотний зв'язок; в загальному випадку цей опір може мати реактивну складову). Він, зокрема, зменшує вхідний опір підсилювача, тобто погіршує його властивості (див. вище п. 4.2.2).

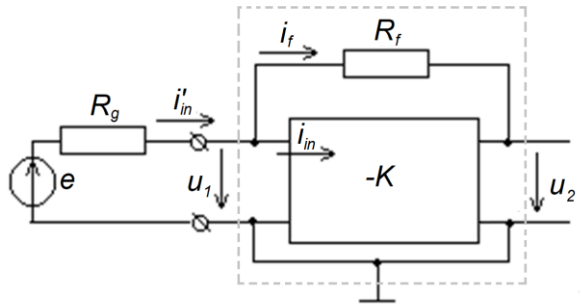


Рис. 4.3.4. Блок-схема підсилювача, охопленого паралельним зворотним зв'язком за напругою

Якщо зворотний зв'язок негативний, тобто $u_2 = -Ku_1$ (це характерно, наприклад, для однокаскадних підсилювачів), до опору R_f прикладається напруга $u_1 - u_2 = (1+K)u_1$. В результаті через нього протікає струм

$$i_f = u_1(1+K)/R_f\,, \tag{4.3.17}$$

і вхідний струм підсилювача складає величину

$$i'_{in} = \frac{u_1}{R_{in}} + u_1\frac{1+K}{R_f} = u_1\left(\frac{1}{R_{in}} + \frac{1+K}{R_f}\right) = \frac{u_1}{R_{in} \parallel \left[R_f/(1+K)\right]}. \tag{4.3.18}$$

(R_{in} – вхідний опір підсилювача без зворотного зв'язку). Таким чином, з урахуванням паралельного зворотного зв'язку

$$R'_{in} = R_{in} \parallel \left[R_f/(1+K)\right]. \tag{4.3.19}$$

Звідси видно, що при паралельному НЗЗ вхідний опір підсилювача при великих K може істотно знизитись.

Сам по собі паралельний НЗЗ не знижує коефіцієнт підсилення: відношення вихідної напруги до вхідної, як і раніше, дорівнює K. Однак якщо розглядати ефективний коефіцієнт підсилення K' як відношення вихідної напруги u_2 до напруги джерела вхідного сигналу e, то слід враховувати ділення напруги e на подільнику, утвореному внутрішнім опором джерела – генератора R_g – та вхідним опором підсилювача (див. вище п. 4.2.2). Тоді

$$u_1 = e\frac{R'_{in}}{R'_{in} + R_g}. \tag{4.3.20}$$

Відповідно ефективний коефіцієнт підсилення дорівнює

$$K' = \frac{u_2}{e} = -\frac{Ku_1}{e} = -\frac{K}{1 + R_g / R'_{in}} \tag{4.3.21}$$

---

[8] Нагадаємо, що в лінійній еквівалентній схемі транзистора в h-параметрах він визначається параметром h_{12}.

Якщо $R_{in} >> R_f / (1+K)$ та $K >> 1$, то

$$R'_{in} \approx R_f / (1+K) \approx R_f / K . \qquad (4.3.22)$$

Тепер

$$| K' | = \frac{K}{1+K\left(R_g / R_f\right)} = \frac{K}{1+K\beta} . \qquad (4.3.23)$$

Таким чином, вираз для ефективного коефіцієнта підсилення зводиться до вигляду (4.3.8), тобто відповідає вхідному опору підсилювача з послідовним негативним зворотним зв'язком за напругою. Роль коефіцієнта зворотного зв'язку відіграє величина $\beta = R_g / R_f$. Оскільки реальні джерела сигналів завжди мають певний внутрішній опір, то цей ефект треба враховувати.

Особливим є випадок, коли паралельний зворотний зв'язок здійснюється не через активний опір R_f, а через конденсатор C_f. Тоді внесена на вхід підсилювача через коло негативного паралельного зворотного зв'язку провідність $i\omega C_f(1+K)$ приєднується до вхідної провідності підсилювача, а його повна вхідна ємність складає

$$C'_{in} = C_{in} + C_f\left(1+K\right) \qquad (4.3.24)$$

і може набагато перевищувати власну вхідну ємність C_{in}.

У багатокаскадних підсилювачах збільшена вхідна ємність наступного каскаду приєднується до вихідної (паразитної) ємності попереднього каскаду і може істотно знизити його верхню граничну частоту.

Контрольні питання до підрозділу 4.3

1. Які типи зворотного зв'язку вам відомі?
2. Що таке позитивний і негативний зворотний зв'язок?
3. Як впливає негативний зворотний зв'язок на властивості підсилювача?
4. За яких умов зворотний зв'язок приводить до самозбудження підсилювача?
5. Як паралельний негативний зворотний зв'язок впливає на вхідний імпеданс підсилювача?

Задачі до підрозділу 4.3

4.3.1. Підсилювач без зворотного зв'язку має коефіцієнт підсилення 50. Коефіцієнт зворотного зв'язку – 0.01. Чому дорівнює коефіцієнт підсилення підсилювача, охопленого

а) позитивним

б) негативним

зворотним зв'язком?

4.3.2. Вхідний опір підсилювача без зворотного зв'язку складає 2 кОм, коефіцієнт підсилення – 20, коефіцієнт зворотного зв'язку – 0.3. Чому дорівнює вхідний опір підсилювача, охопленого

а) позитивним

б) негативним

зворотним зв'язком?

4.3.3. Підсилювач без зворотного зв'язку має коефіцієнт підсилення 100. При якому мінімальному значенні коефіцієнту зворотного зв'язку відбудеться його самозбудження? Фазову умову самозбудження вважати виконаною.

4.4. Підсилювачі на польових транзисторах

Розгляд схем підсилювачів ми розпочнемо з підсилювачів на польових транзисторах, які простіші порівняно з аналогічними схемами на біполярних транзисторах. Це пов'язано з тим, що, як уже вказувалося, польові транзистори порівняно з біполярними мають значно більший вхідний опір (особливо МДН-транзистори) і, відповідно, близький до нуля вхідний струм, і через це практично не поглинають потужність вхідного сигналу. Крім того, характеристики польових транзисторів не так сильно залежать від температури, як у біполярних. Нарешті, в польових транзисторах відсутній внутрішній зворотний зв'язок, завдяки чому їхня еквівалентна схема за сигналом (рис. 4.4.1 а-б) простіша, ніж у біполярних (рис. 4.4.1 в-г). Слід також вказати, що еквівалентні схеми, подані на рис. 4.4.1 а-б, придатні для опису всіх типів польових транзисторів (з керувальним *p-n* переходом, МДН-транзисторів з індукованим та вбудованим каналами).

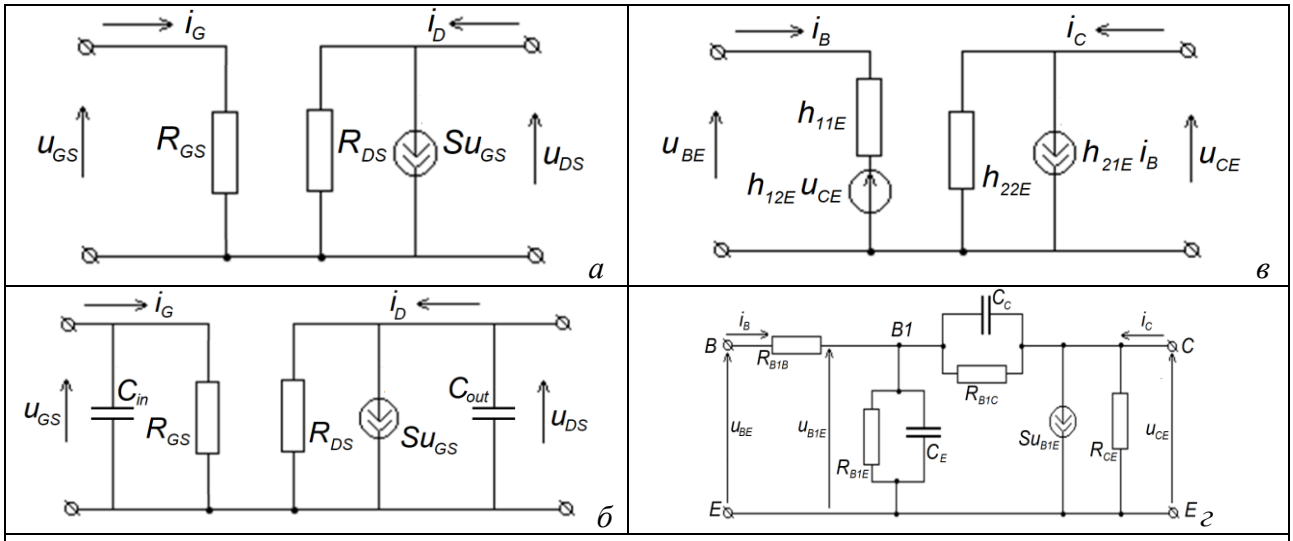


Рис. 4.4.1. Порівняння еквівалентних схем польового (а, б) та біполярного (в, г) транзисторів на низьких (а, в) та на високих (б, г) частотах

Ми вже відзначали, що прохідні характеристики польових транзисторів подібні одна до одної за формою, але зсунуті по горизонталі, тому робочій точці (середині прохідної характеристики) відповідають різні зсуви напруги (voltage offset) на затворі. Для МДН-транзисторів із вбудованим каналом цей зсув дорівнює нулю (рис. 3.4.8, крива 2), для МДН-транзисторів з індукованим *p*-каналом він негативний (рис. 3.4.8, крива 1), для транзисторів із керувальним *p-n* переходом та *p*-каналом – позитивний (рис. 3.4.8 в, крива 3). При заміні *p*-каналу на *n*-канал знак зсуву напруги змінюється на протилежний. Схема підсилювача на польовому транзисторі залежить від того, який постійний зсув напруги треба подати на затвор, щоб забезпечити потрібний вибір робочої точки.

4.4.1. Схема підсилювача на МДН-транзисторі з вбудованим каналом

Найпростіший вигляд має схема підсилювача на МДН-транзисторі з вбудованим каналом, показана на рис. 4.4.2 б. Прохідна характеристика цього транзистора подана на рис. 4.4.2 а. Вихід транзистора увімкнений між землею й шиною живлення послідовно з резистором R_D в колі стоку (пор. з рис. 4.2.1). Затвор польового транзистора заземлений через опір R_G, який визначає вхідний опір підсилювача (він обирається принаймні на порядок меншим від вхідного опору транзистора). Зазначимо, що джерело вхідного сигналу завжди може бути представлене як джерело напруги з деяким внутрішнім опором R_i (рис. 4.2.4 б, де в ролі внутрішнього опору виступає резистор R'), тому для отримання максимальної вхідної напруги повинна бути виконана умова $R_i << R_G$. Нарешті, розділові конденсатори (англ. separation

capacitors) $C_{s1,s2}$ відділяють підсилювач від джерела вхідного сигналу та опору навантаження за постійною складовою напруги (за режимом). В робочому діапазоні частот їхній імпеданс має бути малим порівняно зі вхідним та вихідним опором підсилювача.

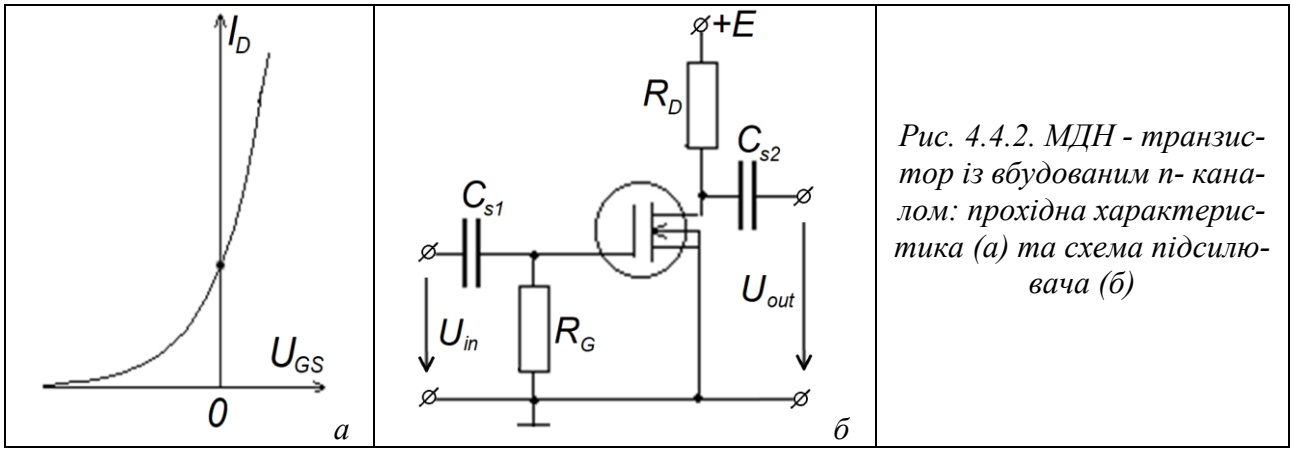


Рис. 4.4.2. МДН - транзистор із вбудованим n- каналом: прохідна характеристика (а) та схема підсилювача (б)

Оскільки, як уже зазначалося, шина живлення заземлена за змінною напругою, то резистор R_D увімкнений паралельно до виходу транзистора (R_{DS} – його вихідний диференціальний опір, S – крутість прохідної характеристики), то еквівалентна схема такого підсилювача для малого сигналу в ненавантаженому режимі має вигляд, показаний на рис. 4.4.3 а (пор. з рис. 3.4.3 а та з рис. 4.2.3 а). Нагадаємо, що на практиці вхідний опір R_{GS} транзистора можна вважати нескінченним, а вхідний струм i_G – нульовим, тому на рис. 4.4.3 а вхідне коло не показане.

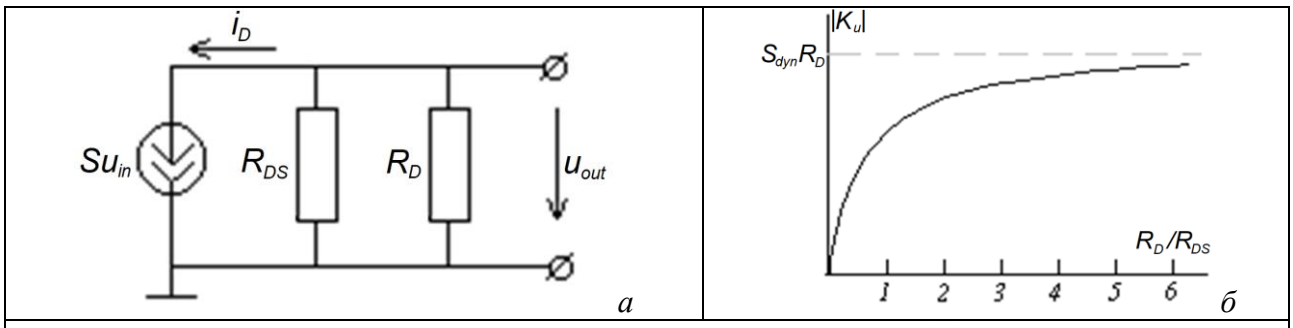


Рис. 4.4.3. Еквівалентна схема підсилювача, поданого на рис. 4.4.1 б, на середніх частотах (а) та залежність його коефіцієнту підсилення за напругою від опору навантаження (б)

4.4.2. Коефіцієнт підсилення за напругою

Розрахуємо коефіцієнт підсилення такого підсилювача за напругою в ненавантаженому режимі. Для цього скористаємося еквівалентною схемою, поданою на рис. 4.4.3 а. Її вихідна напруга може бути записана у формі

$$u_{out} = -Su_{in}\left(R_{DS} \parallel R_D\right) = -Su_{in}\frac{R_{DS}R_D}{R_{DS}+R_D} = -Su_{in}\frac{R_D}{1+R_D/R_{DS}}\ . \qquad (4.4.1)$$

Знак «мінус» вказує на те, що вихідна напруга змінюється в протифазі щодо вхідної (див. вище п.4.2.1).

Відповідно

$$K_u = \frac{u_{out}(t)}{u_{in}(t)} = -\frac{SR_D}{1+R_D/R_{DS}}\ . \qquad (4.4.2)$$

Величину

$$S_{dyn} = \frac{S}{1 + R_D / R_{DS}} \le S \qquad (4.4.3)$$

називають динамічною крутістю польового транзистора. Скориставшись нею, коефіцієнт підсилення (4.4.2) можна записати у формі

$$K_U = -S_{dyn} R_D . \qquad (4.4.4)$$

Залежність коефіцієнта підсилення від опору навантаження подана на рис. 4.4.2 б.

Для оцінки візьмемо S =5 мА/В, R_{DS} =20 кОм, R_D =2 кОм. Тоді з формули (4.4.2) випливає, що K_U =9.1. Як бачимо, коефіцієнт підсилення порівняно невеликий.

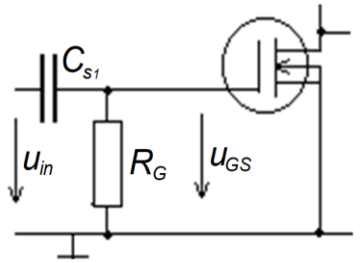


Рис. 4.4.4. До пояснення форми АЧХ підсилювача на низьких частотах

4.4.3. Амплітудно-частотна характеристика підсилювача

Зручно окремо розглянути нижню та верхню межі смуги пропускання підсилювача. На низьких частотах вона визначається розділовою ємністю на вході (рис. 4.4.1 б, рис. 4.4.4). Справді, конденсатор C_{s1} і резистор R_G утворюють фільтр високих частот, навантажений входом транзистора (див. вище п. 4.1.2). З урахуванням вхідного опору польового транзистора R_{GS}, увімкненого паралельно R_G (нагадаємо, що $R_{GS} \gg R_G$), коефіцієнт передачі цього фільтра можна подати у формі

$$K_1 = \frac{R_G R_{GS} / (R_G + R_{GS})}{1/i\omega C_{s1} + R_G R_{GS} / (R_G + R_{GS})} \approx \frac{R_G}{1/i\omega C_{s1} + R_G} = \frac{1}{1 - i/\omega C_{s1} R_G} = \frac{1}{1 - i\omega_{cl}/\omega}, \qquad (4.4.5)$$

$$|K_1| = \frac{1}{\sqrt{1 + 1/(\omega C_{s1} R_G)^2}} = \frac{1}{\sqrt{1 + (\omega_{cl}/\omega)^2}}, \qquad (4.4.6)$$

де

$$\omega_{cl} = 1/R_G C_{s1} \qquad (4.4.7)$$

– нижня гранична частота.

Нехай R_G =100 кОм, C_{s1} =10 нФ. Тоді ω_{cl} =10^3 с$^{-1}$, відповідно f_{cl} = 160 Гц.

Еквівалентна схема підсилювача на польовому транзисторі для області високих частот подана на рис. 4.4.5 (пор. з рис. 3.4.3 б). Розглянемо окремо її вхідну та вихідну частини, почавши з останньої.

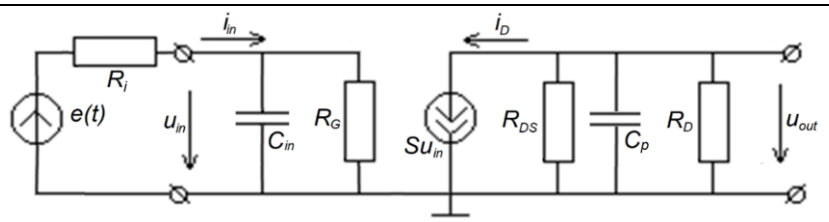


Рис. 4.4.5. Еквівалентна схема підсилювача на польовому транзисторі на високих частотах

Проаналізуємо спочатку спрощений випадок ідеалізованого джерела вхідного сигналу

без внутрішнього опору ($R_i = 0$). Тоді на високих частотах основну роль відіграє паразитна ємність C_p, яка шунтує вихід транзистора (див. рис. 3.4.3 б). В результаті у формулі (4.4.2) для коефіцієнту передачі треба зробити заміну

$$R_{DS} \to \frac{R_{DS}/i\omega C_p}{R_{DS} + 1/i\omega C_p} = \frac{R_{DS}}{1 + i\omega C_p R_{DS}}. \qquad (4.4.8)$$

Отримаємо з формули (4.4.2):

$$K_2(\omega) = -\frac{SR_D}{1 + (R_D/R_{DS})(1 + i\omega C_p R_{DS})} =$$
$$= -\frac{SR_D}{1 + R_D/R_{DS} + i\omega C_p R_D} = -\frac{SR'_D}{1 + i\omega C_p R'_D} = -\frac{K_{\max}}{1 + i\omega/\omega_{ch}}, \qquad (4.4.9)$$

$$K_{\max} = SR'_D, \qquad R'_D = \frac{R_D R_{DS}}{R_D + R_{DS}}, \qquad (4.4.10)$$

$$\omega_{ch1} = 1/C_p R'_D. \qquad (4.4.11)$$

Тут $K_{\max}$ – максимальний коефіцієнт підсилення за напругою (досягається на середніх частотах), ω_{ch1} – верхня гранична частота, зумовлена вихідною паразитною ємністю.

Для типових значень $R'_D = 1$ кОм та $C_p = 10$ пФ ця частота дорівнює $\omega_{ch1} = 10^8$ с$^{-1}$, що відповідає $f_{ch1} = \omega_{ch1}/2\pi = 16$ МГц.

Формулу для коефіцієнта підсилення в усьому діапазоні частот (без урахування внутрішнього опору джерела сигналу) можна подати у формі

$$K(\omega) = K_1(\omega)K_2(\omega) = -\frac{K_{\max}}{(1 - i\omega_{cl}/\omega)(1 + i\omega/\omega_{ch1})} \qquad (4.4.12)$$

або

$$|K(\omega)| = \frac{K_{\max}}{\sqrt{\left[1 + (\omega_{cl}/\omega)^2\right]\left[1 + (\omega/\omega_{ch1})^2\right]}}, \qquad (4.4.13)$$

$$\varphi(\omega) = arctg(\omega_{cl}/\omega) - arctg(\omega/\omega_{ch1}). \qquad (4.4.14)$$

Повернемося тепер до загального випадку, коли джерело вхідного сигналу має ненульовий внутрішній опір ($R_i \neq 0$, див. рис. 4.4.5). В цьому випадку вхідна паразитна ємність може бути суттєвою, оскільки вхідну частину схеми можна розглядати як RC-фільтр низьких частот, навантажений на резистор R_G (пор. з пп. 4.1.1-4.1.2). Тоді напруга вхідного джерела ділиться між резистором R_i і вхідним опором схеми. Останній шунтується паразитною ємністю C_{in}, тому його імпеданс на високих частотах зменшується. Це приводить до зменшення вхідної напруги транзистора u_{in} (за фіксованої амплітуди сигналу $e(t)$), до відповідного зменшення вихідної напруги u_{out} і ефективного коефіцієнту підсилення u_{out}/e.

Для кількісної оцінки даного ефекту розрахуємо коефіцієнт передачі вхідного подільника напруги, складеного з резистора R_i та вхідного імпедансу підсилювача з урахуванням ємності C_{in}, вважаючи, як і раніше, виконаною умову $R_{GS} \gg R_G$:

$$\frac{u_{in}}{e} = \frac{R_G \parallel \frac{1}{i\omega C_{in}}}{R_G \parallel \frac{1}{i\omega C_{in}} + R_i} = \frac{\frac{R_G}{i\omega C_{in}(R_G + 1/i\omega C_{in})}}{\frac{R_G}{i\omega C_{in}(R_G + 1/i\omega C_{in})} + R_i} = \frac{\frac{R_G}{1 + i\omega C_{in} R_G}}{\frac{R_G}{1 + i\omega C_{in} R_G} + R_i} =$$

$$=\frac{R_G}{R_G+R_i\left(1+i\omega C_{in}R_G\right)}=\frac{R_G}{R_G+R_i+i\omega C_{in}R_GR_i}=\frac{R_G}{R_G+R_i}\frac{1}{1+i\omega C_{in}\dfrac{R_GR_i}{R_G+R_i}}=K_0K_3\left(\omega\right), \quad (4.4.15)$$

$$K_0=\frac{R_G}{R_G+R_i}, \quad K_3\left(\omega\right)=\frac{1}{1+i\,\omega/\omega_{ch2}}, \quad \omega_{ch2}=\frac{R_G+R_i}{C_{in}R_GR_i}=\frac{1}{C_{in}\left(R_G\,||\,R_i\right)}. \quad (4.4.16)$$

Візьмемо для оцінки R_G=100 кОм, R_i=0.5 кОм, C_{in}=10 пФ. Тоді $R_G\,||\,R_i$=0.5 кОм, $\omega_{ch2}=2\cdot10^8$ с$^{-1}$, що відповідає $f_{ch2}=\omega_{ch2}/2\pi=32$ МГц. Як бачимо, f_{ch1} та f_{ch2} вийшли однаковими за порядком величини.

З урахуванням вхідного кола можна записати ефективний коефіцієнт підсилення на верхніх частотах у формі

$$K_{2eff}\left(\omega\right)=K_2\left(\omega\right)K_3\left(\omega\right)=\frac{K_0K_2\left(\omega=0\right)}{\left(1+i\,\omega/\omega_{ch1}\right)\left(1+i\,\omega/\omega_{ch2}\right)}; \quad (4.4.17)$$

$$\left|K_{2eff}\left(\omega\right)\right|=\frac{K_0K_2\left(\omega=0\right)}{\sqrt{\left(1-\omega^2/\omega_{ch1}\omega_{ch2}\right)^2+\left(\omega/\omega_{ch1}+\omega/\omega_{ch2}\right)^2}}. \quad (4.4.18)$$

Верхня гранична частота, очевидно, визначається з рівняння

$$\left(1-\omega_{ch}^2/\omega_{ch1}\omega_{ch2}\right)^2+\left(\omega_{ch}/\omega_{ch1}+\omega_{ch}/\omega_{ch2}\right)^2=2,$$

або

$$1-2\frac{\omega_{ch}^2}{\omega_{ch1}\omega_{ch2}}+\frac{\omega_{ch}^4}{\left(\omega_{ch1}\omega_{ch2}\right)^2}+\left(\frac{\omega_{ch}}{\omega_{ch1}}\right)^2+2\frac{\omega_{ch}^2}{\omega_{ch1}\omega_{ch2}}+\left(\frac{\omega_{ch}}{\omega_{ch2}}\right)^2=2,$$

або

$$\frac{\omega_{ch}^4}{\left(\omega_{ch1}\omega_{ch2}\right)^2}+\left(\frac{\omega_{ch}}{\omega_{ch1}}\right)^2+\left(\frac{\omega_{ch}}{\omega_{ch2}}\right)^2=1. \quad (4.4.19)$$

Перейдемо до безрозмірних величин

$$\frac{\omega_{ch}^2}{\omega_{ch1}\omega_{ch2}}=x, \quad \frac{\omega_{ch1}}{\omega_{ch2}}=t. \quad (4.4.20)$$

Тоді біквадратне рівняння (4.4.19) набуває вигляду

$$x^2+x\left(t+\frac{1}{t}\right)-1=0,$$

звідки

$$x=\frac{1}{2}\left[\sqrt{\left(t+\frac{1}{t}\right)^2+4}-\left(t+\frac{1}{t}\right)\right] \quad (4.4.21)$$

(за змістом задачі залишаємо лише додатний корінь). Для розмірної частоти отримуємо:

$$\omega_{ch}^2=\frac{\omega_{ch1}\omega_{ch2}}{2}\left[\sqrt{\left(\frac{\omega_{ch1}}{\omega_{ch2}}+\frac{\omega_{ch2}}{\omega_{ch1}}\right)^2+4}-\left(\frac{\omega_{ch1}}{\omega_{ch2}}+\frac{\omega_{ch2}}{\omega_{ch1}}\right)\right]=$$

$$=\frac{\omega_{ch1}\omega_{ch2}}{2}\left[\sqrt{\left(\frac{\omega_{ch1}^2+\omega_{ch2}^2}{\omega_{ch1}\omega_{ch2}}\right)^2+4}-\left(\frac{\omega_{ch1}^2+\omega_{ch2}^2}{\omega_{ch1}\omega_{ch2}}\right)\right]= \quad (4.4.22)$$

$$=\frac{1}{2}\left[\sqrt{\left(\omega_{ch1}^2+\omega_{ch2}^2\right)^2+4\left(\omega_{ch1}\omega_{ch2}\right)^2}-\left(\omega_{ch1}^2+\omega_{ch2}^2\right)\right]=$$

$$=\frac{1}{2}\left[\sqrt{\omega_{ch1}^4+\omega_{ch2}^4+6\left(\omega_{ch1}\omega_{ch2}\right)^2}-\left(\omega_{ch1}^2+\omega_{ch2}^2\right)\right].$$

Виконані раніше числові оцінки дали $\omega_{ch1} = 10^8\,\text{с}^{-1}$, $\omega_{ch2} = 2\cdot 10^8\,\text{с}^{-1}$. Тоді, відповідно до формули (4.4.42), маємо: $\omega_{ch} = 0.7\cdot 10^8\,\text{с}^{-1}$, $f_{ch} = 11\,\text{МГц}$.

Нехай $\omega_{ch1} \ll \omega_{ch2}$. Тоді, домноживши рівняння (4.4.22) на спряжену величину, маємо:

$$\omega_{ch}^2 = \frac{1}{2}\frac{4\left(\omega_{ch1}\omega_{ch2}\right)^2}{\sqrt{\omega_{ch1}^4 + \omega_{ch2}^4 + 6\left(\omega_{ch1}\omega_{ch2}\right)^2} + \left(\omega_{ch1}^2 + \omega_{ch2}^2\right)} \approx \frac{2\left(\omega_{ch1}\omega_{ch2}\right)^2}{2\omega_{ch2}^2} = \omega_{ch1}^2 . \qquad (4.4.23)$$

Як і слід було чекати, за умови, коли одна з граничних частот значно менша за іншу, впливом більшої граничної частоти можна знехтувати.

Навпаки, якщо $\omega_{ch1} = \omega_{ch2}$, то

$$\omega_{ch}^2 = \omega_{ch1}^2\left(\sqrt{2} - 1\right). \qquad (4.4.24)$$

Графік залежності ω_{ch} від ω_{ch1}, ω_{ch2}, що задається формулою (4.4.22), поданий на рис. 4.4.7.

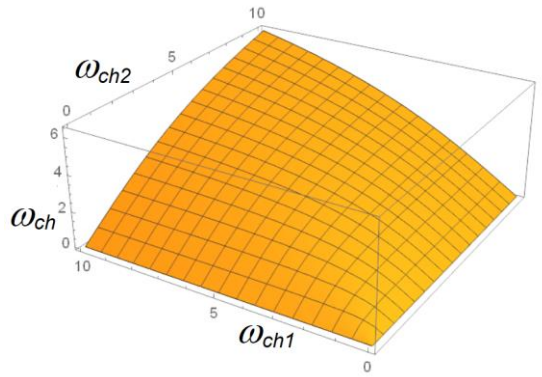


Рис. 4.4.6. Залежність граничної частоти ω_{ch} підсилювача від граничних частот вихідної ω_{ch1} та вхідної ω_{ch2} частин схеми. Всі частоти – у відносних одиницях

Формула (4.4.17) враховує спадання коефіцієнту передачі лише на високих частотах. Щоб врахувати спадання цього коефіцієнту також і на низьких частотах, її треба домножити на K_1 (див. формулу (4.4.5)). Маємо:

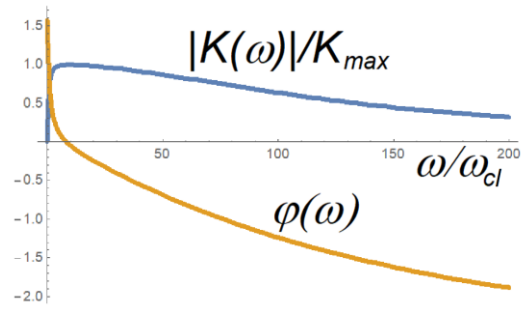
$$K_{eff}(\omega) = K_1(\omega)K_{2eff}(\omega) = -\frac{K_{\max}}{\left(1 - i\omega_{cl}/\omega\right)\left(1 + i\omega/\omega_{ch1}\right)\left(1 + i\omega/\omega_{ch2}\right)} \qquad (4.4.25)$$

або

$$\left|K_{eff}(\omega)\right| = \frac{K_{\max}}{\sqrt{\left[1 + \left(\omega_{cl}/\omega\right)^2\right]\left[1 + \left(\omega/\omega_{ch1}\right)^2\right]\left[1 + \left(\omega/\omega_{ch2}\right)^2\right]}}, \qquad (4.4.26)$$

$$\varphi(\omega) = arctg\left(\omega_{cl}/\omega\right) - arctg\left(\omega/\omega_{ch1}\right) - arctg\left(\omega/\omega_{ch2}\right). \qquad (4.4.27)$$

Графіки залежностей (4.4.13)-(4.4.14) подані на рис. 4.4.6.



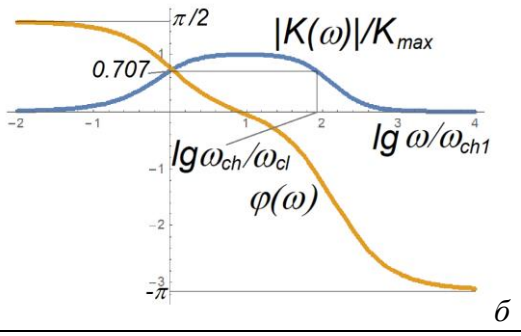


Рис. 4.4.7. Графіки АЧХ та ФЧХ підсилювача на польовому транзисторі в лінійному (а) та напівлогарифмічному (б) масштабах

4.4.4. Схема підсилювача на МДН-транзисторі з індукованим каналом

Нагадаємо, що для забезпечення робочого режиму МДН-транзистора з індукованим *n*-каналом на його затвор треба подати позитивний зсув напруги (рис. 4.4.8 а). Звичайно це роблять за допомогою подільника R_{G1}, R_{G2} (рис. 4.4.8 б). В усьому іншому (призначення елементів, еквівалентна схема в режимі малого сигналу, коефіцієнт підсилення, частотні властивості) схема, показана на рис. 4.4.8 б, цілком аналогічна розглянутій раніше (див. рис. 4.4.2 б).

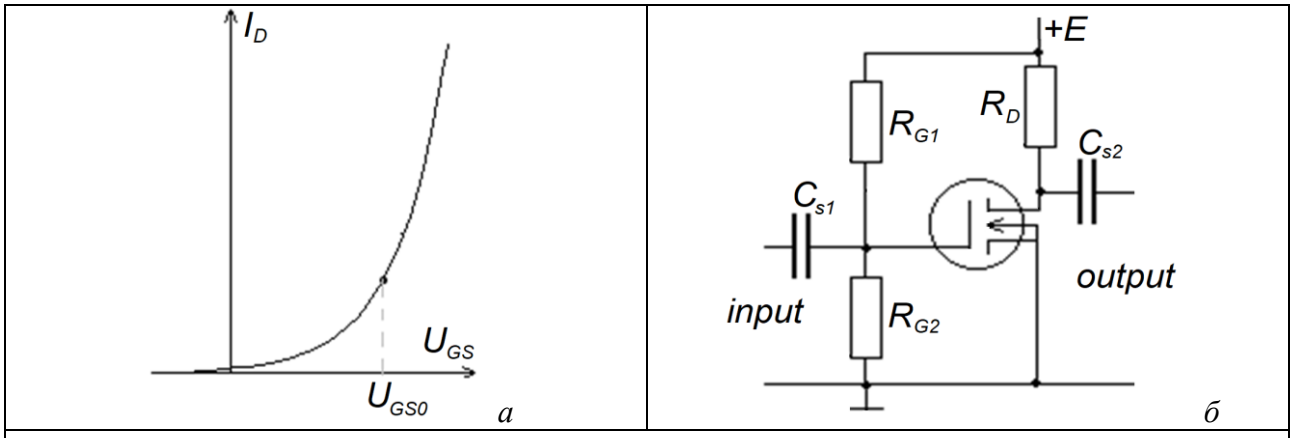


Рис. 4.4.8. МДН - транзистор з індукованим n- каналом: прохідна характеристика (а) та схема підсилювача (б)

4.4.5. Схема підсилювача на транзисторі з керувальним p-n переходом

Вище вже йшлося, що для забезпечення робочого режиму транзистора з керувальним *p-n* переходом та *n*-каналом на його затвор треба подати негативний зсув напруги (рис. 4.4.9 а), при тому, що на виток подається позитивна напруга. Використовувати двополярне живлення вкрай незручно. Тому в схемі (рис. 4.4.9 б), замість того, щоб подавати негативний потенціал на затвор, подають позитивний потенціал на виток. Для цього між витоком і землею включають додатковий опір R_S, через який протікає струм витоку. А затвор, як і в схемі на МДН-транзисторі з вбудованим каналом, заземлений через резистор R_G. Наявність резистора R_S призводить до появи в схемі негативного зворотного зв'язку, через який зменшується коефіцієнт підсилення. Щоб уникнути цього, резистор R_S шунтується конденсатором великої ємності. Завдяки цьому виток транзистора за сигналом виявляється заземленим, і негативний зворотний зв'язок за сигналом зникає.

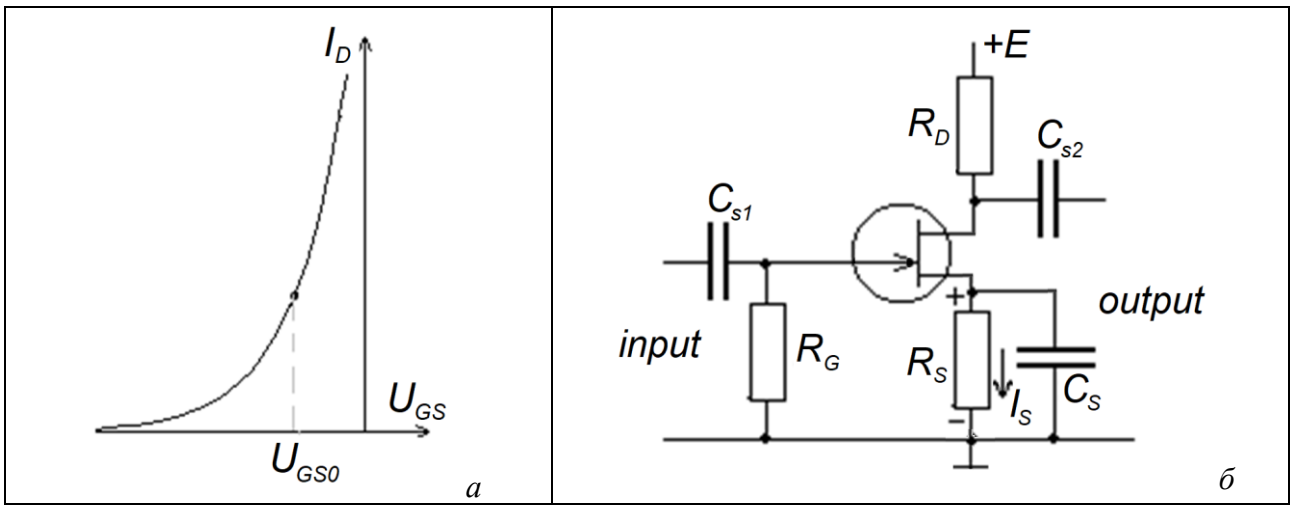


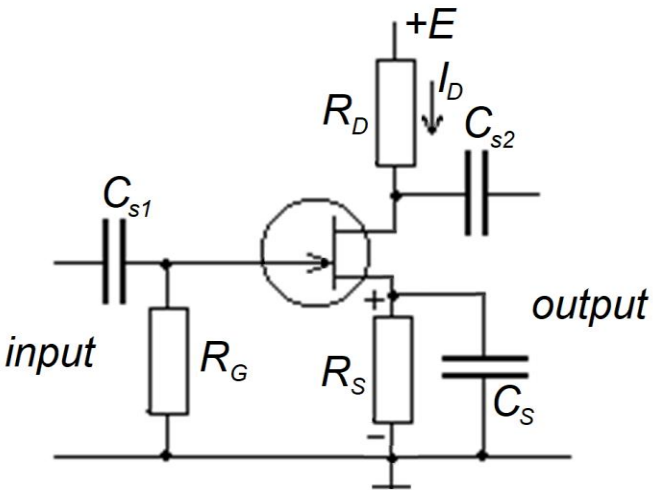
Рис. 4.4.9. Польовий транзистор із керувальним p-n переходом та n-каналом: прохідна характеристика (а) та схема підсилювача (б)

Контрольні питання до підрозділу 4.4

1. Чим відрізняються схеми підсилювачів на різних типах польових транзисторів?
2. Чим динамічна крутість польового транзистора відрізняється від статичної?
3. Чим визначається амплітудно-частотна характеристика підсилювача на польовому транзисторі на низьких частотах?
4. Чим визначається амплітудно-частотна характеристика підсилювача на польовому транзисторі на високих частотах?

Приклади розв'язання задач до підрозділу 4.4

П 4.4.1. Робоча точка польового транзистора: $U_{DS} = 10$ В, $U_{GS} = -1.5$ В, $I_D = 10$ мА. Напруга живлення $E = 25$ В. Розрахувати резистори R_S та R_D.



Розв'язок

Затворний струм транзистора близький до нуля, тому можна вважати, що затвор заземлений через опір R_G. Тому $R_S I_D = |U_{GS}|$, звідки $R_S = \dfrac{U_{GS}}{I_D} = \dfrac{1.5}{0.01} = 150$ Ом.

Тепер запишемо рівняння за другим правилом Кірхгофа для кола земля – R_S – транзистор – R_D – шина живлення:

$$R_S I_D + U_{DS} + R_D I_D = E,$$

Звідки

$$R_D = \frac{E - U_{DS}}{I_D} - R_S = \frac{25-10}{0.01} - 150 = 1350\ Ом \approx 1.4\ кОм.$$

П 4.4.2. В тій самій схемі крутість прохідної характеристики польового транзистора $S = 5$ мА/В, $R_{GS} = 12$ кОм. При якому значенні R_D досягається коефіцієнт підсилення $K = 20$? Як зміниться коефіцієнт підсилення, якщо R_D збільшити вдвічі? Вказівка: скористатися формулою $K = SR_D / \left(1 + R_D / R_{DS}\right)$.

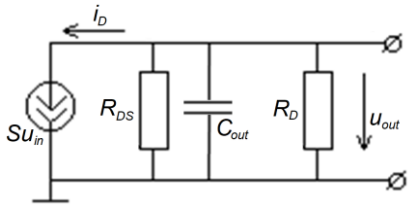
Розв'язок

$K\left(1 + R_D / R_{DS}\right) = SR_D$, звідки $SR_D - K\,R_D / R_{DS} = K$. Тоді

$$R_D = \frac{K}{S - K/R_{DS}} = \frac{KR_{DS}}{SR_{DS} - K} = \frac{20 \times 12}{5 \times 12 - 20} = \frac{240}{40} = 6\ кОм$$

$$K\left(2R_c\right) = \frac{2SR_D}{1 + 2R_D / R_{DS}} = \frac{2 \times 5 \times 6}{1 + 2 \times 6/12} = 30.$$

П 4.4.3. У підсилювачі на польовому транзисторі Знайти верхню граничну частоту підсилювача, скориставшись його еквівалентною схемою на високих частотах. $R_{DS} = 10$ кОм, $R_D = 5$ кОм, $C_{out} = 30$ пФ.



Розв'язок

$$\omega_{ch} = \frac{1}{\left(R_{DS} \parallel R_D\right) C_{out}} = \frac{R_{DS} + R_D}{R_{DS} R_D C_{out}} = \frac{\left(5+10\right)\times 10^3}{\left(5\times 10\right)\times 10^6 \times 30 \times 10^{-12}} = \frac{15\times 10^3}{15\times 10^{-4}} = 1\times 10^7\, c^{-1}$$

$$f_{ch} = \frac{\omega_{ch}}{2\pi} = \frac{10^7}{6.3} = 1.6\,\text{МГц}.$$

Задачі до підрозділу 4.4

4.4.1. Робоча точка польового транзистора: $U_{DS} = 12$ В, $U_{GS} = -2$ В, $I_D = 15$ мА. Напруга живлення $E = 30$ В. Розрахувати резистори R_S та R_D.

4.4.2. В тій самій схемі крутість прохідної характеристики польового транзистора $S = 8$ мА/В, $R_{DS} = 15$ кОм. При якому значенні R_D досягається коефіцієнт підсилення $K = 30$? Як зміниться коефіцієнт підсилення, якщо R_D збільшити вдвічі?

4.4.3. У підсилювачі на польовому транзисторі $R_{DS} = 15$ кОм, $R_D = 7$ кОм, $C_{out} = 20$ пФ. Знайти верхню граничну частоту підсилювача, скориставшись його еквівалентною схемою на високих частотах.

4.5. Підсилювачі на біполярних транзисторах

У сучасній радіоелектроніці поширеним активним елементом, поруч із польовими, є біполярний транзистор.

4.5.1. Схема найпростішого підсилювача на біполярному транзисторі

На рис. 4.5.1 а подано графіки сім'ї вихідних характеристик біполярного транзистора, увімкненого зі спільним еміткром, та навантажувальну пряму $U_{CE} = E - I_C R_C$ (пор. з рис. 4.2.2 б). Робоча точка *O* обирається на середині доступної ділянки АВ навантажувальної прямої.

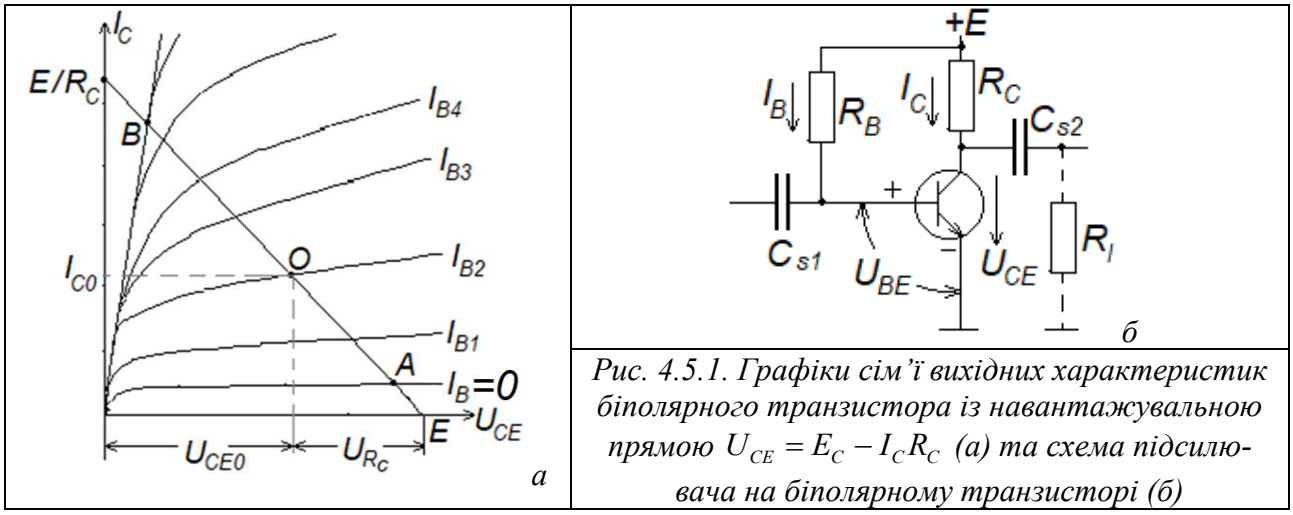


Рис. 4.5.1. Графіки сім'ї вихідних характеристик біполярного транзистора із навантажувальною прямою $U_{CE} = E_C - I_C R_C$ (а) та схема підсилювача на біполярному транзисторі (б)

На рис. 4.5.1 б подано найпростішу схему підсилювача на біполярному транзисторі. Тут R_C – резистор навантаження, резистор $R_B = (E - U_{BE0})/I_{B0} \approx E/I_{B0}$ забезпечує базовий струм, що відповідає робочій точці (струм I_{B2} на рис. 4.5.1 а), резистор R_l – зовнішнє навантаження, $C_{s1,\,s2}$ – розділові конденсатори, що запобігають впливу попереднього (джерела вхідного сигналу) та наступного (зовнішнього навантаження) каскадів на режим (тобто положення робочої точки) підсилювача.

4.5.2. Параметри підсилювача на біполярному транзисторі

Для знаходження параметрів підсилювача скористаємося еквівалентною схемою транзистора в h-параметрах (рис. 3.5.12), врахувавши, що на вході увімкнене джерело сигналу з його внутрішнім опором, а на виході – зовнішнє навантаження (нагадаємо, що шина живлення заземлена за сигналом, тому резистор R_C увімкнений між колектором і базою). Для спрощення приймемо, що $R_l >> R_C$ та $R_B >> h_{11E}$, тобто формально вважатимемо опори резисторів R_l та R_B нескінченними. На середніх частотах ($h_{11E} >> 1/\omega C_{s1}$) знехтуємо впливом розділового конденсатора C_{s1}. Отримаємо еквівалентну схему підсилювача на середніх частотах, показану на рис. 4.5.2, де $u_1 = u_{BE}$ – вхідна напруга, $u_2 = u_{CE}$ – вихідна напруга.

<u>Коефіцієнт підсилення за струмом</u>. Вихідним опором на еквівалентній схемі (рис. 4.5.2) служить паралельне сполучення колекторного резистора R_C і вихідної провідності транзистора h_{22E}, між якими ділиться пропорційно до їхніх провідностей струм керованого джерела $h_{21E} i_1$. Тоді вихідний струм можна записати у формі

$$i_2 = -h_{21E} i_1 \frac{1/R_C}{1/R_C + h_{22E}} = -\frac{h_{21E} i_1}{1 + h_{22E} R_C}. \qquad (4.5.1)$$

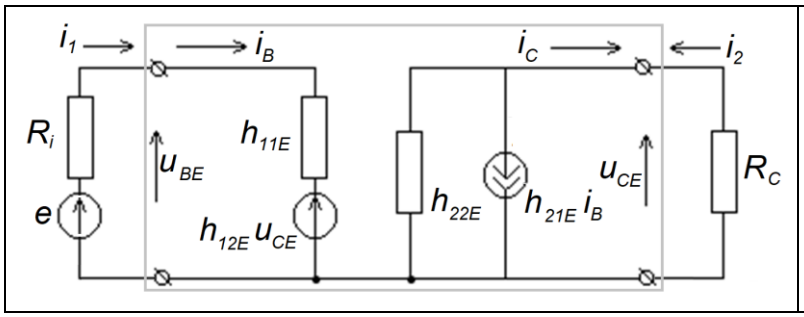


Рис. 4.5.2. Еквівалентна схема підсилювача на біполярному транзисторі на середніх частотах в режимі малого сигналу

Маємо такий коефіцієнт підсилення за струмом:

$$K_i \equiv \frac{i_2}{i_1} = -\frac{h_{21E}}{1+h_{22E}R_C}. \qquad (4.5.2)$$

Підкреслимо, що, як видно з рис. 4.5.2, вихідним струмом вважається струм через колекторний резистор R_C, а не через резистор R_l, який служить навантаженням підсилювача (див. рис. 4.5.1 б).

Вхідний опір підсилювача. Виходячи з еквівалентної схеми,

$$R_{in} \equiv \frac{u_1}{i_1} = h_{11E} + \frac{h_{12E}u_2}{i_1}. \qquad (4.5.3)$$

Приймемо до уваги, що $u_2 = i_2R_C = i_1K_iR_C$. Тоді з урахуванням (4.5.2) отримаємо

$$R_{in} \equiv \frac{u_1}{i_1} = h_{11E} - h_{12E}R_C\frac{h_{21E}}{1+h_{22E}R_C} = h_{11E} - \frac{h_{12E}h_{21E}}{h_{22E}+1/R_C} \approx h_{11E} \qquad (4.5.4)$$

(тут і в наступних формулах враховано мализну параметра h_{12E}).

Коефіцієнт підсилення за напругою. Подамо вхідну та вихідну напруги через відповідні струми: $u_1 = i_1R_{in}$, $u_2 = i_2R_C$. Тоді

$$K_u \equiv \frac{u_2}{u_1} = \frac{i_2R_C}{i_1R_{in}} = K_i\frac{R_C}{R_{in}} = -\frac{h_{21E}}{\left(1+h_{22E}R_C\right)}\frac{R_C}{\left[h_{11E} - h_{12E}h_{21E}/\left(h_{22E}+1/R_C\right)\right]} =$$

$$= -\frac{h_{21E}R_C}{h_{11E}\left(1+h_{22E}R_C\right) - h_{12E}h_{21E}R_C} \approx -\frac{h_{21E}R_C}{h_{11E}\left(1+h_{22E}R_C\right)}. \qquad (4.5.5)$$

Слід пам’ятати, що визначений таким чином коефіцієнт підсилення за напругою не враховує того, що е.р.с. e джерела вхідної напруги ділиться між вхідним опором транзистора R_{in} та внутрішнім опором джерела сигналу R_g (див. далі п. 4.5.4). Цим ефектом можна знехтувати, коли $R_{in} \ll R_g$.

Вихідний опір. Розглядатимемо весь підсилювач як деякий лінійний чотириполюсник. Його параметр h_{22} (величина, обернена до вихідного опору) визначається як

$$h_{22} \equiv \frac{1}{R_{out}} = \left.\frac{\partial I_2}{\partial U_2}\right|_{I_1=const} = \left.\frac{i_2}{u_2}\right|_{i_1=0}$$

(пор. з четвертою з формул (3.5.13)). Справді, умова $i_1 = 0$ еквівалентна умові $I_1 = const$.

Якщо на схемі (рис. 4.5.2) перерахувати джерело е.р.с. e з послідовно увімкненим внутрішнім опором R_i у паралельне сполучення джерела струму $i = e/R_i$ з тим самим опором (зворотну операцію ми виконали в п. 4.2.2), то умова $e = 0$ збігається з умовою $i = 0$.

Отже, умова $i_1 = 0$ означає, що вхідне джерело напруги закорочене, але залишений його внутрішній опір. Відповідно напруга u_2 розглядається як вплив. В результаті отримуємо еквівалентну схему, подану на рис. 4.5.3.

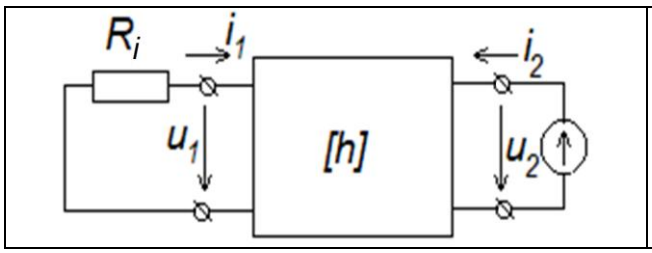


Рис. 4.5.3. До знаходження вихідного опору підсилювача

Для розрахунку R_{out} скористаємося рівняннями чотириполюсника в h -параметрах, узявши до уваги, що чотириполюсник на рис. 4.5.3 – це фактично еквівалентна схема транзистора (пор. з рис. 4.5.2):

$$\begin{cases} u_1 = h_{11E} i_1 + h_{12E} u_2; \\ i_2 = h_{21E} i_1 + h_{22E} u_2. \end{cases} \tag{4.5.6}$$

Врахуємо, що $u_1 = -i_1 R_i$ (див. рис. 4.5.2). Тоді з першого рівняння маємо:

$$i_1 = -\frac{h_{12E} u_2}{R_i + h_{11E}}, \tag{4.5.7}$$

і після підстановки до другого рівняння отримуємо

$$i_2 = -\frac{h_{21E} h_{12E}}{R_i + h_{11E}} u_2 + h_{22E} u_2, \tag{4.5.8}$$

звідки

$$\frac{1}{R_{out}} = \frac{i_2}{u_2} = h_{22E} - \frac{h_{12E} h_{21E}}{h_{11E} + R_i}. \tag{4.5.9}$$

Для оцінки можливих значень всіх цих величин візьмемо як приклад типові значення h -параметрів біполярного транзистора середньої чи малої потужності, увімкненого зі спільним емітером: $h_{11_E} = 1$кОм, $h_{12_E} = 3.10^{-4}$, $h_{21_E} = 40$ та $h_{22_E} = 5 \cdot 10^{-5}$Сим. Нехай опір колекторного резистора становить $R_C = 5$ кОм, а внутрішній опір джерела сигналу – $R_i = 200$ Ом. Підставивши ці значення до отриманих формул, одержимо: $K_i = -32$, $K_u = -167$, $R_{in} = 952$ Ом, $R_{out} = 25$ кОм.

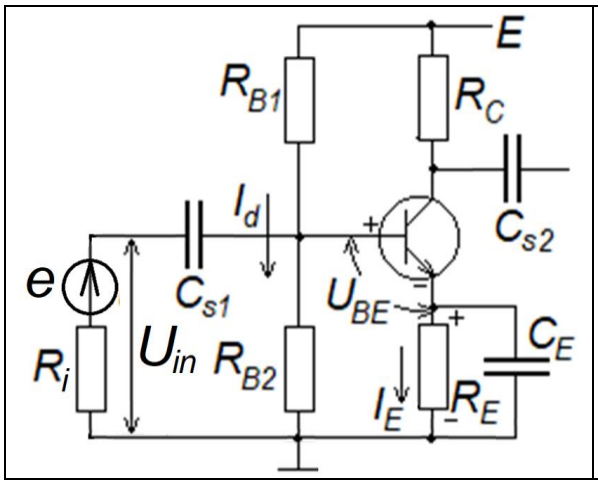


Рис. 4.5.4. Схема підсилювача з ланцюжком термостабілізації в емітерному колі

4.5.3. Стабілізація режиму за допомогою негативного зворотного зв'язку

Негативний зворотній зв'язок за постійною складовою напруги (рис. 4.5.4) дозволяє зменшити небажаний вплив температури на режим транзистора (див. вище п. 3.5.3). Оскільки напруга на вході транзистора складає $u_{BE} = u_{R_{B2}} - u_{R_E}$, а на емітерному резисторі – $u_{R_E} = u_{R_C}\left(R_E / R_C\right)$, де u_{R_C} – вихідна напруга, то u_{R_E} якраз і виступає в ролі сигналу негативного зворотного зв'язку (пор. з рис. 4.3.1). Цей НЗЗ частково компенсує зміну режиму (постійні складові струмів та напруг транзистора), зумовлену зміною температури. В цій схемі постійний струм подільника R_{B1}, R_{B2} вибирається значно більшим від (постійного) вхідного струму транзистора I_{B0}, аби зміна останнього не впливала на потенціал бази.

Щоб запобігти виникненню негативного зворотного зв'язку за сигналом і відповідному зменшенню коефіцієнта підсилення, емітерний опір шунтується конденсатором великої ємності.

4.5.4. Частотні властивості підсилювача на біполярному транзисторі

Для аналізу частотних властивостей підсилювача на біполярному транзисторі (рис. 4.5.1 б) останній можна замінити еквівалентною схемою Джіаколетто. Як і раніше (див. рис. 4.5.2), вважаємо виконаними умови $R_l >> R_C$ та $R_B >> h_{11E}$ і враховуємо, що за сигналом опір R_C увімкнений між колектором і емітером. Але, на відміну від рис. 4.5.2, братимемо до уваги вплив розділового конденсатора C_{s1}. На вході підсилювача увімкнене джерело е.р.с. e із деяким внутрішнім опором R_g. Отриману таким чином еквівалентну схему подано на рис. 4.5.5 (тут $u_{CE} = u_{out}$).

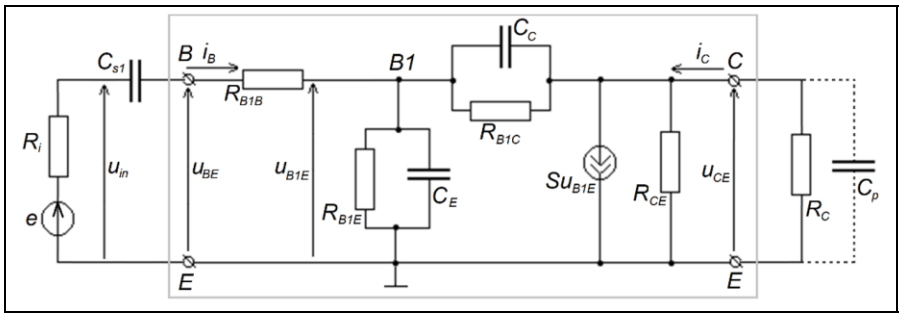


Рис. 4.5.5. Еквівалентна схема підсилювача на біполярному транзисторі в режимі малого сигналу

Коефіцієнт підсилення за напругою можна визначити різними способами, в залежності від того, який сигнал вважати вхідним.

Величину $K_{eff} = u_{out}/e$ будемо називати ефективним коефіцієнтом підсилення.

Напруга u_{in} на вході підсилювача ($u_{in} < e$, оскільки частина е.р.с. падає на внутрішньому опорі джерела R_g, бо вхідний струм відмінний від нуля) служить його вхідним сигналом. Величину $K = u_{out}/u_{in}$ називатимемо повним коефіцієнтом підсилення.

На нижніх частотах помітна частина напруги u_{in} падає на розділовому конденсаторі C_{s1}, тобто між базою і емітером прикладена напруга u_{BE} ($u_{BE} \le u_{in}$). Величину $K_{ext} = u_{out}/u_{BE}$ будемо називати зовнішнім (тобто зовнішнім щодо ефективної бази B1) коефіцієнтом підсилення.

Нарешті, відповідно до схеми Джіаколетто, джерело струму у вихідному колі керується напругою u_{B1E}, тобто потенціалом ефективної бази ($u_{B1E} < u_{BE}$, оскільки частина напруги u_{BE} падає на об'ємному опорі бази). Величину $K_{in} = u_{out}/u_{B1E}$ назвемо внутрішнім коефіцієнтом підсилення.

Очевидно, виконується нерівність

$K_{in} \geq K_{ext} \geq K \geq K_{eff}$.

У смузі пропускання підсилювача (на середніх частотах) можна, з одного боку, знехтувати падінням напруги на розділовому конденсаторі (тобто вважати, що $u_{BE} \approx u_{in}$), а з іншого – не зважати на ємності *p-n* переходів транзистора і вихідну ємність схеми. Тоді коефіцієнт підсилення (повний і зовнішній коефіцієнти підсилення в цьому разі збігаються, $K \approx K_{ext}$) у цьому діапазоні частот визначається раніше отриманою формулою (4.5.5)

$$K = -\frac{h_{21E}}{h_{11E}} \frac{R_C}{1 + h_{22E} R_C} \equiv K_0 .$$

На низьких частотах повний коефіцієнт підсилення $K(\omega) = u_{in} / u_{out}$ зменшується через падіння напруги на розділовому конденсаторі.

На високих частотах резистор R_C шунтується вихідною паразитною ємністю, як це має місце і в підсилювачах на польових транзисторах (див. підрозділ 4.4.1). Це призводить до зменшення внутрішнього коефіцієнта підсилення $K_{in} = u_{out} / u_{B1E}$.

Вхідний імпеданс транзистора, розрахований між ефективною базою та емітером, на високих частотах також зменшується через ємності колекторного та емітерного *p-n* переходів. Тому збільшується частка е.р.с. e, яка падає не лише на об'ємному опорі бази R_{B1B}, але й на внутрішньому опорі джерела сигналу R_i. Останній ефект враховує ефективний коефіцієнт підсилення $K_{eff} = u_{out} / e$.

<u>Коефіцієнт підсилення за напругою в області низьких частот</u>. Поведінка коефіцієнта підсилення в області низьких частот така сама, як у підсилювачі на польовому транзисторі. Коротко розглянемо її для підсилювального каскаду на біполярному транзисторі, схема якого подана на рис. 4.5.1 б.

В області низьких частот коефіцієнт підсилення подібного підсилювача знижується через наявність на його вході розділового конденсатора C_{s1}, який разом із вхідним опором каскаду $R_{in} \approx h_{11e}$ утворює подільник (рис. 4.5.5 – 4.5.6). Він зменшує напругу u_{BE} в $\gamma_1(\omega)$ разів порівняно з напругою u_{in}, де

$$\gamma_1(\omega) = \frac{R_{in}}{R_{in} + 1/i\omega C_{s1}} . \qquad (4.5.9)$$

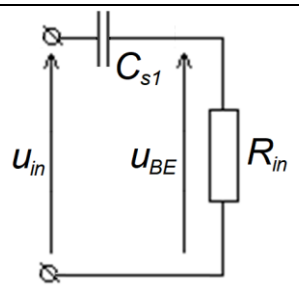


Рис. 4.5.6. Еквівалентна схема вхідного кола підсилювача на біполярному транзисторі на низьких частотах

В результаті повний коефіцієнт підсилення $K(\omega) = u_{out} / u_{in}$ виявляється в γ_1 разів меншим від зовнішнього коефіцієнта підсилення $K_{ext} = u_{out} / u_{BE}$, значення якого на середніх частотах ми позначили вище через K_0. Таким чином, модуль повного коефіцієнта підсилення $|K(\omega)|$ дорівнює

$$|K(\omega)| = K_0 \left|\gamma_1(\omega)\right| = \frac{K_0}{\sqrt{1 + 1/(\omega\theta)^2}} , \qquad (4.5.10)$$

де $\theta = C_{s1}R_{in}$ – стала часу вхідного кола (рис. 4.5.5). Нижня гранична частота складає $\omega_{cl} = 1/\theta$. На цій частоті коефіцієнт підсилення $|K(\omega)|$ знижується в $1/\sqrt{2} \approx 0.71$ разів порівняно з K_0.

Оскільки вхідний опір біполярного транзистора звичайно невеликий, то для нормальної роботи каскаду аж до низьких частот ємність розділового конденсатора доводиться обирати великою (до десятків - сотень мікрофарад).

Справді, візьмемо $R_{in} = 0.5$ кОм, $f_{cl} = 100$ Гц, відповідно $\omega_{cl} \approx 630$ с$^{-1}$. Тоді $C_{s1} = 1/\omega_{cl}R_{in} = 1/630 \cdot 500 = 3.2$ мкФ.

<u>Еквівалентна схема підсилювача в області високих частот</u>. Еквівалентну схему підсилювача на біполярному транзисторі в області високих частот (рис. 4.5.7 а) отримуємо зі схеми, поданої на рис. 4.5.5, замкнувши накоротко розділовий конденсатор, впливом якого на високих частотах можна знехтувати.

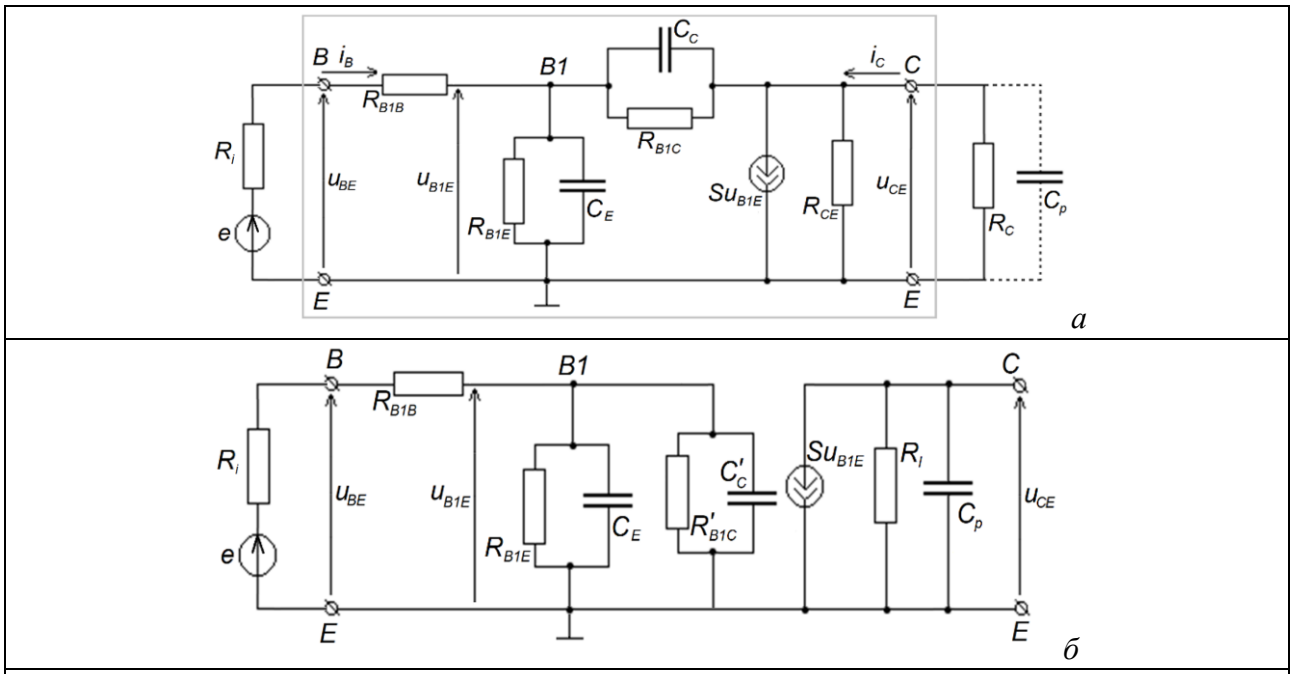


Рис. 4.5.7. Еквівалентна схема підсилювача на біполярному транзисторі в області високих частот (а) та її перетворений варіант (б)

Для аналізу властивостей підсилювача на високих частотах спочатку перетворимо його еквівалентну схему (рис.4.5.7 а) таким чином, щоб позбутися елементів R_{B1C} і C_C, які зв'язують вхідне та вихідне кола транзистора, і звести її до вигляду, подібного до еквівалентної схеми транзистора в h-параметрах.

Елементи R_{B1C} і C_C увімкнені між виходом і ефективною базою нашого підсилювача таким чином, що вони створюють внутрішній паралельний негативний зворотний зв'язок, розглянутий раніше (див. підрозділ 4.3.8, пор. рис. 4.5.7 а та рис. 4.3.4). Було показано, що при такому зв'язку провідність $1/Z_f$ елементів, що його створюють, перераховується у вхідне коло множенням на величину $(1+K)$ (див. формули (4.3.22), (4.3.24)). З порівняння рис. 4.5.7 а та рис. 4.3.4 зрозуміло, що в даному разі під K слід розуміти величину $u_{CE}/u_{B1E} = K_{in}$.

На рис. 4.5.7 б подано еквівалентну схему підсилювача, де елементи R_{B1C} і C_C перераховано до вхідного кола як

$$R'_{B1C} = R_{B1C}/(1+K_{in}), \qquad C'_C = C_C(1+K_{in}). \qquad (4.5.11)$$

Крім того, паралельно сполучені резистори R_{CE} та R_C об'єднані в один елемент $R_l = R_{CE} \parallel R_C$.

Тим самим ми розділили вхідне та вихідне кола. На рис. 4.5.7 б реактивні елементи (ємності C_E та C'_C), що характеризують транзистор, опинилися у вхідному колі, а реактивні елементи зовнішньої схеми (ємність C_p) – у вихідному. В принципі ця еквівалентна схема така сама, як у підсилювача на польовому транзисторі (рис. 4.4.5).

<u>Частотні властивості підсилювача на біполярному транзисторі на високих частотах</u>. Запишемо ефективний коефіцієнт підсилення у формі

$$K_{eff}(\omega) \equiv \frac{u_{out}}{e} = \frac{u_{B1B}}{e}\frac{u_{out}}{u_{B1B}} = \gamma_2 K_{in}, \tag{4.5.12}$$

де K_{in} – внутрішній коефіцієнт підсилення, а γ_2 – коефіцієнт передачі подільника, в якому вхідною напругою служить e, а вихідною – u_{B1B}.

Розрахуємо спершу внутрішній коефіцієнт підсилення, взявши до уваги, що навантаженням керованого джерела струму у вихідному колі служить паралельне сполучення резистора R_l і ємності C_p:

$$K_{in} = \frac{u_{out}}{u_{B1B}} = S\frac{R_l/i\omega C_p}{R_l + 1/i\omega C_p} = \frac{SR_l}{1+i\omega\tau_{out}}, \qquad \tau_{out} = R_l C_p. \tag{4.5.13}$$

Візьмемо $R_l = 2$ кОм, $C_p = 50$ пФ. Тоді $\tau_{out} = 100$ нс, $f_{out} = 1/2\pi\tau_{out} = 1.6$ МГц.

Тепер знайдемо коефіцієнт передачі γ_2 подільника напруги, верхня ланка якого включає послідовно сполучені резистори R_{B1B} та R_i, а нижня – паралельно сполучені опори R_{B1E} та $R'_{B1C} = R_{B1C}(1+K')$ і конденсатори $C'_C = C_C(1+K')$ та C_E. Позначивши $C'_E = C_E + C_C(1+K_{in}) \simeq C_E + K_{in}C_C$ (враховано, що $K_{in} = SR_l \gg 1$) і $R'_{B1E} = R_{B1E} \parallel R'_{B1C} \approx R_{B1E}$ (звичайно $R'_{B1C} \gg R_{B1E}$), запишемо:

$$\begin{aligned}\gamma_2 &= \frac{u_{B1E}}{e} = \frac{R'_{B1E} \parallel (1/i\omega C'_E)}{R_{B1B} + R_i + R'_{B1E} \parallel (1/i\omega C'_E)} = \frac{R'_{B1E}/(1+i\omega C'_E R'_{B1E})}{R_{B1B} + R_i + R'_{B1E}/(1+i\omega C'_E R'_{B1E})} = \\ &= \frac{R'_{B1E}}{(R_{B1B}+R_i)(1+i\omega C'_E R'_{B1E}) + R'_{B1E}} = \frac{R'_{B1E}}{i\omega C'_E R'_{B1E}(R_{B1B}+R_i) + R_{B1B} + R_i + R'_{B1E}} = \\ &= \frac{R'_{B1E}}{(R_{B1B}+R_i+R'_{B1E})}\frac{1}{\left[1+i\omega C'_E R'_{B1E}(R_{B1B}+R_i)/(R_{B1B}+R_i+R'_{B1E})\right]},\end{aligned} \tag{4.5.14}$$

або

$$\gamma_2 = \gamma_3/(1+i\omega\tau_{in}), \tag{4.5.15}$$

де

$$\gamma_3 = \gamma_2(\omega=0) = \frac{R'_{B1E}}{R'_{B1E}+R_{B1B}+R_i}, \quad R''_{B1E} = \frac{R_{B1E}(R_{B1B}+R_i)}{R_{B1E}+R_{B1B}+R_i}, \quad \tau_{in} = C'_E R''_{B1E}. \tag{4.5.16}$$

Нехай $C_E = C_C = 400$ пФ, $K_{in} = 100$, $R_{B1E} = 150$ Ом, $R_{B1B} = 1.6$ кОм, $R_i = 0.35$ кОм. Тоді $C' = 40$ нФ, $R''_{B1E} = 0.4$ кОм і $\tau_{in} = 1.6$ мкс. Відповідно $f_{in} = 1/2\pi\tau_{in} = 0.1$ МГц. Для обраних параметрів, як бачимо, $f_{out} \gg f_{in}$, тобто інерційність підсилювача визначається вхідним колом на еквівалентній схемі, поданій на рис. 4.5.7 б.

Остаточно маємо:

$$K_{eff}(\omega) = K'\gamma_2 = \frac{K_{eff\,0}}{(1+i\omega\tau_{out})(1+i\omega\tau_{in})}, \quad K_{eff\,0} = \gamma_3 SR_l, \tag{4.5.17}$$

$$\left|K_{eff}(\omega)\right| = \frac{K_{eff\,0}}{\sqrt{\left[1+(\omega\tau_{in})^2\right]\left[1+(\omega\tau_{out})^2\right]}}. \qquad (4.5.18)$$

Графік амплітудно-частотної характеристики підсилювача на біполярному транзисторі в області високих частот показана на рис. 4.5.8.

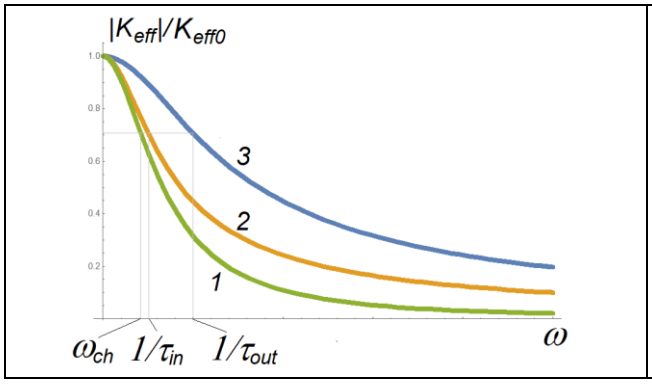


Рис. 4.5.8. Амплітудно-частотна характеристика підсилювача на біполярному транзисторі в області високих частот для випадку $\tau_{out} = 2\tau_{in}$: 1 – загальний випадок; 2 – $\tau_{out} = 0$; 3 – $\tau_{in} = 0$

Формула (4.5.17) з точністю до позначень збігається з раніше отриманою формулою (4.4.17). Тому верхню граничну частоту підсилювача знаходимо з формули (4.4.22)

$$\omega_{ch}^2 = \frac{1}{2}\left[\sqrt{\omega_{ch1}^4 + \omega_{ch2}^4 + 6(\omega_{ch1}\omega_{ch2})^2} - \left(\omega_{ch1}^2 + \omega_{ch2}^2\right)\right],$$

де тепер $\omega_{ch1} = \omega_{in}$, $\omega_{ch2} = \omega_{out}$, або

$$f_{ch} = \sqrt{\frac{1}{2}\left[\sqrt{f_{ch1}^4 + f_{ch2}^4 + 6(f_{ch1}f_{ch2})^2} - \left(f_{ch1}^2 + f_{ch2}^2\right)\right]}. \qquad (4.5.19)$$

Виконані раніше оцінки дали $f_{out} = 1.6$ МГц, $f_{in} = 0.1$ МГц. Тоді $f_{ch} = 0.1$ МГц, як це й випливає з наближеної формули (4.4.23).

4.5.4. Корекція частотної характеристики

Спад АЧХ на високих частотах, спричинений паразитною ємністю на виході транзистора, можна частково компенсувати, увімкнувши індуктивністю послідовно з резистором навантаження транзистора (рис. 4.5.9 а). Тоді провідність паралельного сполучення $1/h_{22E}$ та C_p з частотою зростає, а провідність паралельної гілки $R_C L$ зменшується, частково компенсуючи це зростання.

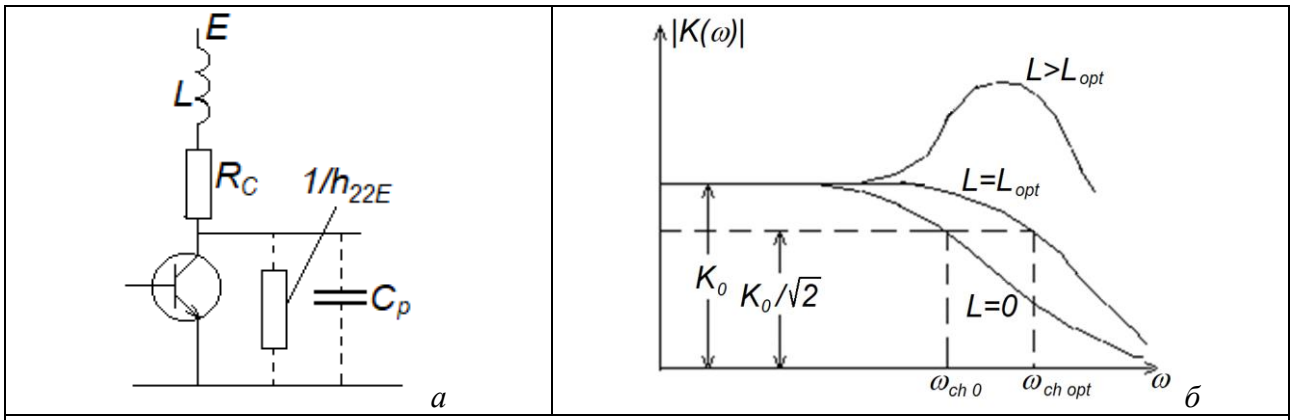


Рис. 4.5.9. Схема підсилювача з корекцією верхніх частот (а) та графіки сім'ї його АЧХ для різних значень коригувальної індуктивності

За відсутності коригувальної індуктивності $\omega_{ch0} = 1/R_l C_p$, де, як і раніше, $R_l = R_C/(1 + h_{22E} R_C)$.

З урахуванням коригувальної індуктивності

$$\left|K(\omega)\right| \sim \left|h_{22E} + i\omega C_p + \frac{1}{R_C + i\omega L}\right|^{-1} = \left|h_{22E} + i\omega C_p + \frac{R_C - i\omega L}{R_C^2 + (\omega L)^2}\right|^{-1}. \tag{4.5.20}$$

Таким чином,

$$\left|K^2(\omega)\right| \sim \left|\left(h_{22E} + \frac{R_C}{R_C^2 + \omega^2 L^2}\right) + i\left(\omega C_p - \frac{\omega L}{R_C^2 + \omega^2 L^2}\right)\right|^{-2} =$$

$$= \left[\left(h_{22E} + \frac{R_C}{R_C^2 + \omega^2 L^2}\right)^2 + \left(\omega C_p - \frac{\omega L}{R_C^2 + \omega^2 L^2}\right)^2\right]^{-1} =$$

$$= \left[(h_{22E})^2 + 2\frac{h_{22E} R_C}{R_C^2 + \omega^2 L^2} + \left(\frac{R_C}{R_C^2 + \omega^2 L^2}\right)^2 + (\omega C_p)^2 - 2\frac{\omega^2 L C_p}{R_C^2 + \omega^2 L^2} + \left(\frac{\omega L}{R_C^2 + \omega^2 L^2}\right)^2\right]^{-1} =$$

$$= \left[(h_{22E})^2 + (\omega C_p)^2 + 2\frac{h_{22E} R_C - \omega^2 L C_p}{R_C^2 + \omega^2 L^2} + \frac{1}{R_C^2 + \omega^2 L^2}\right]^{-1} =$$

$$= \frac{R_C^2 + \omega^2 L^2}{\left[(h_{22E})^2 + (\omega C_p)^2\right]\left(R_C^2 + \omega^2 L^2\right) + 2\left(h_{22E} R_C - \omega^2 L C_p\right) + 1} =$$

$$= \frac{R_C^2\left(1 + \omega^2 \frac{L^2}{R_C^2}\right)}{\left[(h_{22E} R_C)^2 + (\omega R_C C_p)^2\right]\left(1 + \omega^2 \frac{L^2}{R_C^2}\right) + 2\left(h_{22E} R_C - \omega^2 \frac{L}{R_C} R_C C_p\right) + 1} \sim$$

$$\sim \frac{1 + (\omega\tau_L)^2}{\left[a^2 + (\omega\tau_C)^2\right]\left[1 + (\omega\tau_L)^2\right] + 2\left(a - \omega^2 \tau_L \tau_C\right) + 1} = \tag{4.5.21}$$

$$= \frac{1 + x}{\left(a^2 + b^2 x\right)(1 + x) + 2(a - bx) + 1} = \frac{1 + x}{b^2 x^2 + x\left(a^2 + b^2 - 2b\right) + (a + 1)^2},$$

де

$$\tau_L = \frac{L}{R_C}, \quad \tau_C = R_C C_p, \quad a = h_{22E} R_C, \quad b = \frac{\tau_C}{\tau_L}, \quad x = (\omega\tau_L)^2. \tag{4.5.22}$$

Знайдемо оптимальну коригувальну індуктивність з умови, щоб нахил АЧХ на її початку (тобто при $\omega = 0$) був нульовим (рис. 4.5.9 б). Маємо:

$$\frac{d}{d\omega^2} K^2(\omega)\Big|_{\omega=0} \sim \frac{d}{dx} \frac{1 + x}{b^2 x^2 + x\left(a^2 + b^2 - 2b\right) + (a + 1)^2}\Bigg|_{x=0} =$$

$$= \frac{b^2 x^2 + x\left(a^2 + b^2 - 2b\right) + (a + 1)^2 - (1 + x)\left[2b^2 x + \left(a^2 + b^2 - 2b\right)\right]}{\left[b^2 x^2 + x\left(a^2 + b^2 - 2b\right) + (a + 1)^2\right]^2}\Bigg|_{x=0} = \tag{4.5.23}$$

$$=\frac{(a+1)^2-\left(a^2+b^2-2b\right)}{(a+1)^4}\sim 2a+1+2b-b^2=0,$$

або

$$b=1+\sqrt{2(a+1)} \tag{4.5.24}$$

(обираємо додатний корінь), звідки

$$L_{opt}=\frac{R_C^2C_p}{1+\sqrt{2\left(1+h_{22E}R_C\right)}}=\frac{R_C^2C_p\left[\sqrt{2\left(1+h_{22E}R_C\right)}-1\right]}{1+2h_{22E}R_C}. \tag{4.5.25}$$

Нехай $R_C=4$ кОм, $C_p=50$ нФ, $h_{22E}=0.5$ мСим. Тоді $L_{opt}=0.23$ мГн.

Очевидно, при оптимальній корекції

$$\frac{K^2(\omega)}{K^2(0)}=\left.\frac{(1+x)(a+1)^2}{bx(bx-2)+\left(a^2+b^2\right)x+(a+1)^2}\right|_{b=1+\sqrt{2(a+1)}}.$$

Позначимо

$$y=b^2x=(\omega\tau_C)^2,\qquad a+1=c.$$

Тоді

$$\frac{K^2(\omega)}{K^2(0)}=\left.\frac{\left(1+y/b^2\right)c^2}{b\,y/b^2\left(b\,y/b^2-2\right)+\left[(c-1)^2+b^2\right]y/b^2+c^2}\right|_{b=1+\sqrt{2(a+1)}}=$$

$$=\left.\frac{\left(y+b^2\right)c^2}{y\left[y-2b+b^2+(c-1)^2\right]+b^2c^2}\right|_{b=1+\sqrt{2c}}=\left.\frac{\left(y+b^2\right)c^2}{y\left[y-2c+(b-1)^2\right]+\left(y+b^2\right)c^2}\right|_{b=1+\sqrt{2c}}= \tag{4.5.26}$$

$$=\frac{\left[y+\left(1+\sqrt{2c}\right)^2\right]c^2}{y[y-2c+2c]+\left[y+\left(1+\sqrt{2c}\right)^2\right]c^2}=\frac{\left[y+\left(1+\sqrt{2c}\right)^2\right]c^2}{y^2+\left[y+\left(1+\sqrt{2c}\right)^2\right]c^2}=\left\{1+\frac{y^2}{c^2\left[y+\left(1+\sqrt{2c}\right)^2\right]}\right\}^{-1}.$$

(рис. 4.5.10).

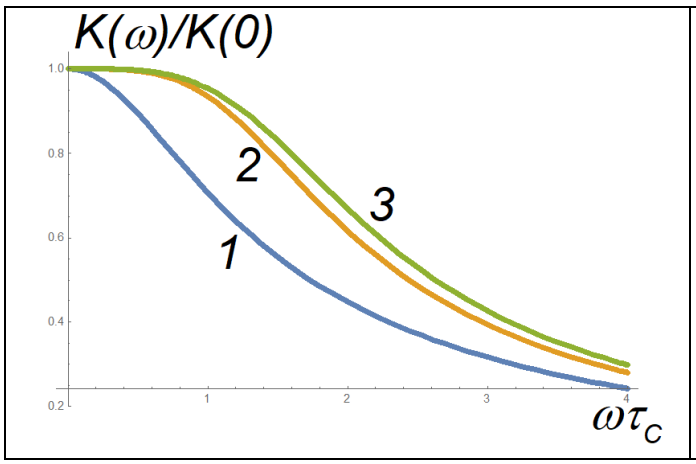


Рис. 4.5.10. Нормовані АЧХ підсилювача: 1 – без корекції, 2-3 – з оптимальною корекцією (2 – $h_{22E}R_C=0,\ c=1$, 3– $h_{22E}R_C=1,\ \ c=2$)

Знайдемо верхню граничну частоту каскаду з оптимальною корекцією з умови $K^2(\omega)/K^2(0)=1/2$, звідки

$$y=\frac{c}{2}\left[c+\sqrt{c^2+4\left(1+2\sqrt{2c}+2c\right)}\right], \qquad (4.5.27)$$

або

$$\left(\omega_{ch}\tau_C\right)^2=\frac{1+h_{22E}R_C}{2}\left[1+h_{22E}R_C+\sqrt{\left(1+h_{22E}R_C\right)^2+4\left(1+\sqrt{2\left(1+h_{22E}R_C\right)}\right)^2}\right]. \qquad (4.5.28)$$

З формули (4.5.28) видно, що зі зростанням параметра $h_{22E}R_c$ частота ω_{ch} зростає.

У найпростішому випадку, коли $1/h_{22E} \gg R_c$, або $h_{22E}R_c \ll 1$, маємо:

$$L_{opt}=\frac{R_C^2C_p}{1+\sqrt{2}}\approx 0,41R_C^2C_p, \qquad (4.5.25\ а)$$

$$\left(\omega_{ch}\tau_C\right)^2=0.5\left[1+\sqrt{1+4\left(1+\sqrt{2}\right)^2}\right]\approx 2.97,$$

$$\omega_{ch}=1.72/\tau_C, \qquad \tau_C=R_CC_p. \qquad (4.5.28\ а)$$

Відносне збільшення верхньої граничної частоти підсилювача за рахунок коригувальної індуктивності складає $\omega_{ch}/\omega_{ch0}=1,72\,R_l/R_C=1,72/\left(1+h_{22E}R_C\right)\approx 1,72$ в силу використаної умови $h_{22E}R_c \ll 1$.

Насправді в цьому розрахунку ми враховували лише вихідну паразитну ємність і не враховували ємності у вхідному колі еквівалентної схеми підсилювача (рис. 4.5.7 б). Тому виконана оцінка характеризує лише зменшення сталої $\tau_{out}\approx\tau_C/1.9\approx 0.53\,\tau_C$ у формулі (4.5.20) і відповідне збільшення f_{out} до значення 3 МГц. Оскільки для обраних нами параметрів $f_{out}\gg f_{in}$, то коригувальна індуктивність практично не впливає верхню граничну частоту. Отже, коригувальну індуктивність варто застосовувати тоді, коли $f_{out}\le f_{in}$.

Контрольні питання до підрозділу 4.5

1. Опишіть призначення елементів схеми підсилювача на біполярному транзисторі.
2. Як послідовний зворотний зв'язок за напругою впливає на властивості підсилювача?
3. Що таке позитивний і негативний зворотний зв'язок?
4. Для чого в схемі підсилювача на біполярному транзисторі потрібна термостабілізація? Як вона забезпечується?
5. Як можна підняти верхню граничну частоту підсилювача на біполярному транзисторі?

Приклади розв'язання задач до підрозділу 4.5

П 4.5.1. В схемі підсилювача (рис. 4.5.11) робоча точка біполярного транзистора характеризується параметрами $U_{CE}=3$ В, $U_{BE}=0.4$ В, $I_C=10$ мА, $I_B=0.1$ мА. Напруга живлення $E=8$ В. Розрахувати резистори R_B та R_C.

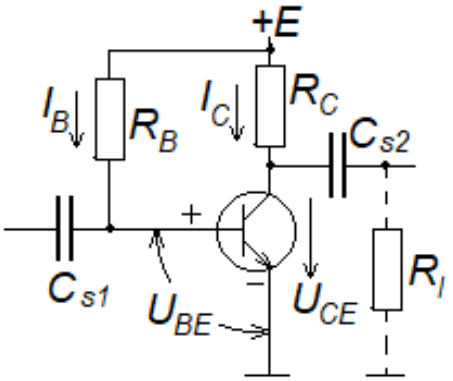


Рис. 4.5.11

Розв'язок

Запишемо рівняння за другим законом Кірхгофа для кола земля – емітер-колектор - R_C - шина живлення:

$U_{CE}+I_CR_C=E$,

звідки

$$R_C = \frac{E - U_{CE}}{I_C} = \frac{8-3}{0.01} = 500\,\text{Ом}.$$

Аналогічно для кола земля – емітер-база - R_B - шина живлення:

$$U_{BE} + I_B R_B = E;$$

$$R_B = \frac{E - U_{BE}}{I_B} = \frac{8-0.4}{10^{-4}} = 76\,\text{кОм}.$$

П 4.5.2. Параметри біполярного транзистора в околі робочої точки $h_{11E} = 1.5$ кОм, $h_{12E} = 6 \cdot 10^{-5}$, $h_{21E} = 20$, $h_{22E} = 50$ мкСим, опір навантаження $R_C = 5$ кОм, внутрішній опір джерела вхідного сигналу $R_i = 0.5$ кОм. Знайти коефіцієнти підсилення за струмом та за напругою, вхідний і вихідний опори підсилювача.

Розв'язок

$$\frac{1}{R_{out}} = h_{22E} - \frac{h_{12E}h_{21E}}{h_{11E} + R_i} = 50\times10^{-6} - \frac{6\times10^{-5}\times20}{1.5\times10^3 + 0.5\times10^3} = 50\times10^{-6} - 0.6\times10^{-6} \approx 50\,\text{мкСим}.$$

$$R_{out} = 20\,\text{кОм}.$$

$$R_{in} = h_{11E} - \frac{h_{12E}h_{21E}}{h_{22E} + 1/R_C} = 1500 - \frac{20\times6\times10^{-5}}{5\times10^{-5} + 1/5000} =$$

$$= 1500 - \frac{1.2\times10^{-3}}{5\times10^{-5} + 2\times10^{-4}} = 1500 - \frac{1.2\times10^{-3}}{2.5\times10^{-4}} = 1500 - 4.8 \approx 1.5\,\text{кОм}.$$

$$k_i = -\frac{h_{21E}}{1 + h_{22E}R_C} = -\frac{20}{1 + 5\times10^{-5}\times5\times10^3} = -\frac{20}{1.25} = -16.$$

$$k_U = k_i\frac{R_C}{R_i} \approx -\frac{16\times5}{1.5} \approx 53.$$

П 4.5.3. Два однакових підсилювача, розглянутих у задачі П 4.5.2, увімкнені послідовно. Розрахувати загальний коефіцієнт підсилення двох каскадів за напругою.

Розв'язок

Навантаженням першого підсилювача служить опір R_C, паралельно якому увімкнений вихід транзистора. Тому

$$R_{l1} = \frac{R_C\left(1/h_{22E}\right)}{R_C + 1/h_{22E}} = \frac{5\times20}{25} = 4\,\text{кОм}.$$

Відповідно коефіцієнт підсилення цього каскаду за струмом та за напругою буде

$$k_i = -\frac{h_{21E}}{1 + h_{22E}R_{l1}} = -\frac{20}{1 + 5\times10^{-5}\times4\times10^3} = -\frac{20}{1.2} \approx -17$$

$$k_{U1} = k_{i1}\frac{R_{l1}}{R_{im}} \approx -\frac{17\times4}{1.5} \approx -45.$$

$$k_{U1}k_U = 53\times45 \approx 2{,}4\times10^3.$$

П 4.5.4. У схемі підсилювача (рис. 4.5.12) E_C =20 В, R_C =4 кОм, I_C =2,1 мА, I_B =0,1 мА, U_{CE} =11 В, U_{BE} =0,28 В, h_{11E} =0,4 кОм, h_{12E} =1,5·10^{-3}, h_{21E} =21, h_{22E} =0,05 мСим. Знайти елементи схеми підсилювача, поданої на рисунку. Нижня гранична частота підсилювача – 50 Гц.

Розв'язок

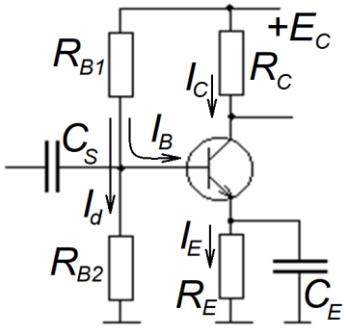


Рис. 4.5.12

1. Оберемо струм подільника I_d з умови $I_B \ll I_d \ll I_C$:

$I_d = \sqrt{I_B I_C} = \sqrt{0{,}1 \cdot 2{,}1} = 0.45$ мА.

2. Запишемо рівняння за другим правилом Кірхгофа для кола $+E_C$ – R_C – U_{CE} – R_E – земля:

$E_C = I_C R_C + U_{CE} + (I_C + I_B) R_E$.

Тоді

$$R_E = \frac{E_C - I_C R_C - U_{CE}}{I_C + I_B} = \frac{20 - 2.1 \cdot 4 - 11}{2.1 + 0.1} = 0.28 \text{ кОм.}$$

3. Запишемо рівняння за другим правилом Кірхгофа для кола земля – R_{B2} – U_{BE} – R_E – земля:

$I_d R_{B2} = U_{BE} + (I_C + I_B) R_E$.

Тоді

$$R_{B2} = \frac{U_{BE} + (I_C + I_B) R_E}{I_d} = \frac{0.28 + 2.2 \cdot 0.28}{0.45} = 1.96 \text{ кОм} \approx 2 \text{ кОм.}$$

4. Запишемо рівняння за другим правилом Кірхгофа для кола земля $+E_C$ – R_{B1} – R_{B2} – земля:

$E_C = (I_d + I_B) R_{B1} + I_d R_{B2}$.

Тоді

$$R_{B1} = \frac{E_C - I_d R_{B2}}{I_d + I_B} = \frac{20 - 0.45 \cdot 2}{0.45 + 0.1} = 34.7 \text{ кОм} \approx 35 \text{ кОм.}$$

5. Знайдемо ємність конденсатора C_E з умови $1/\omega_{cl} C_E \ll R_E$:

$$C_E = \frac{1}{f_{cl} R_E} = \frac{1}{50 \cdot 280} = 71 \text{ мкФ} \approx 70 \text{ мкФ.}$$

6. Знайдемо ємність конденсатора C_S з умови $1/\omega_{cl} C_S \ll R_{in}$, де $R_{in} = h_{11E} \parallel R_{B1} \parallel R_{B2} \approx$ $\approx h_{11E} R_{B2} / (h_{11E} + R_{B2}) = 0.4 \cdot 2 / (0.4 + 2) = 0.33$ кОм (враховано, що $R_{B1} \gg R_{B2}, h_{11E}$):

$$C_S = \frac{1}{f_{cl} R_{in}} = \frac{1}{50 \cdot 330} = 60 \text{ мкФ.}$$

П 4.5.5. Користуючись результатами попередньої задачі, розрахувати параметри підсилювача на біполярному транзисторі.

Розв'язок

1. Коефіцієнт підсилення за струмом:

$$k_i = -\frac{h_{21E}}{1 + h_{22E} R_C} = -\frac{21}{1 + 4/20} = \frac{21}{1.2} = 17.5.$$

2. Коефіцієнт підсилення за напругою:

$$k_u = k_i \frac{R_C}{h_{11E}} = 17.5 \frac{4}{0.4} = 175$$

3. Коефіцієнт підсилення потужності:

$k_P = k_u k_i \approx 3 \cdot 10^3$.

4. Вхідний опір розраховано вище: $R_{in} \approx 0.33$ кОм.

5. Вихідний опір:

$$R_o = \left(h_{22E} + \frac{1}{R_C} \right)^{-1} = \left(0.05 + 0.25\right)^{-1} = \frac{1}{0.3} \approx 3.3\,\text{кОм}.$$

П 4.5.6. Вважаючи, що $C_P = 200\,\text{пФ}$, оцінити коригувальну індуктивність і верхню граничну частоту підсилювача з урахуванням цієї корекції. Параметри підсилювача взяти з попередніх задач.

Розв'язок

1. Перевіримо виконання умови $h_{22E}R_c \ll 1$: $h_{22E}R_c = 0.05 \cdot 4 = 0.2 \ll 1$. Тому для подальших оцінок можна скористатися спрощеними формулами (4.5.25 а) та (4.5.28 а) і вважати, що $R_l \approx R_C$.

2. Верхня гранична частота за відсутності коригувальної індуктивності

$$f_{ch0} = \frac{1}{2\pi R_l C_P} \approx \frac{1}{2\pi R_C C_P} = \frac{1}{6.3 \cdot 2 \cdot 10^{-10} \cdot 4 \cdot 10^3} = \frac{10^7}{42} \approx 0.20\,\text{МГц}.$$

3. Індуктивність коригувальної котушки:

$$L_{opt} = 0.41 R_C^2 C_p = 0.41\left(4 \cdot 10^3\right)^2 2 \cdot 10^{-10} = 0.41 \cdot 32 \cdot 10^{-4} \approx 1.3\,\text{мГн}.$$

3. Верхня гранична частота за наявності коригувальної індуктивності:

$$\tau_C = R_C C_p = 4 \cdot 10^3 \cdot 200 \cdot 10^{-12} = 8 \cdot 10^{-7} = 0.8\ \text{мкс}.$$

$$f_{ch} = \frac{1.72}{2\pi\tau_C} = \frac{0.27}{\tau_C} = \frac{0.27}{0,8} 10^6 = 0.34\ \text{МГц}.$$

Таким чином, за умови $h_{22E}R_c \ll 1$ верхня гранична частота справді зросла приблизно на 70%.

Задачі до підрозділу 4.5

4.5.1. Користуючись результатами задач П 4.5.1 – П 4.5.5 і вважаючи, що $C_P = 50\,\text{пФ}$, розрахувати верхню граничну частоту підсилювача, коригувальну індуктивність і верхню граничну частоту за її наявності. Скористатися спрощеними формулами (4.5.25 а) та (4.5.28 а) і вважати, що $R_l \approx R_C$.

4.5.2. Розв'язати задачу П 4.5.4, якщо $U_{CE} = 5\,\text{В}$, $U_{BE} = 0.5\,\text{В}$, $I_C = 15\,\text{мА}$, $I_B = 0.2\,\text{мА}$, $E = 10\,\text{В}$.

4.5.3. Розв'язати задачу П 4.5.5, якщо $h_{11E} = 1\,\text{кОм}$, $h_{12E} = 2 \cdot 10^{-5}$, $h_{21E} = 30$, $h_{22E} = 20\,\text{мкСим}$, $R_C = 8\,\text{кОм}$, $R_i = 0.2\,\text{кОм}$.

4.5.4. Розв'язати задачу П 4.5.6, скориставшись даними задачі 4.5.3.

4.6. Спеціальні типи підсилювачів на біполярних транзисторах

В підрозділі 4.5 були детально розглянуті підсилювачі низької частоти на біполярних транзисторах. Ми бачили, що їхні амплітудно-частотні характеристики мають спад як на низьких, так і на високих частотах, а середня частота смуги пропускання сумірна з її шириною.

В цьому підрозділі ми розглянемо спеціальні типи підсилювачів, які мають смугу пропускання особливого вигляду. Це так звані підсилювачі мінливої напруги, у яких відсутній спад АЧХ на низьких частотах, та селективні підсилювачі, ширина смуги пропускання яких значно менша за середню частоту цієї смуги.

До спеціальних типів підсилювачів можна віднести й так званий емітерний повторювач – схему, яка не підсилює напругу, але натомість має великий вхідний і малий вихідний опір. Вона зазвичай використовується для узгодження джерела сигналу з великим внутрішнім опором та низькоомного навантаження.

Ми розглянемо спеціальні типи підсилювачів на біполярних транзисторах. Однак аналогічні схеми можна побудувати також на лампах чи на польових транзисторах.

4.6.1. Підсилювач мінливої напруги

Підсилювач мінливої напруги (ПМН), як уже згадувалося, призначений для підсилення повільних сигналів (з характерними частотами одиниці Герц і нижче).

Як уже вказувалося (див. вище пп. 4.4.3, 4.5.4), зменшення коефіцієнту підсилення на низьких частотах зумовлене розділовими конденсаторами. Тому в схемі ПМН (рис. 4.6.1 а) такі ємності відсутні. Графік відповідної амплітудно-частотної характеристики поданий на рис. 4.6.1 б.

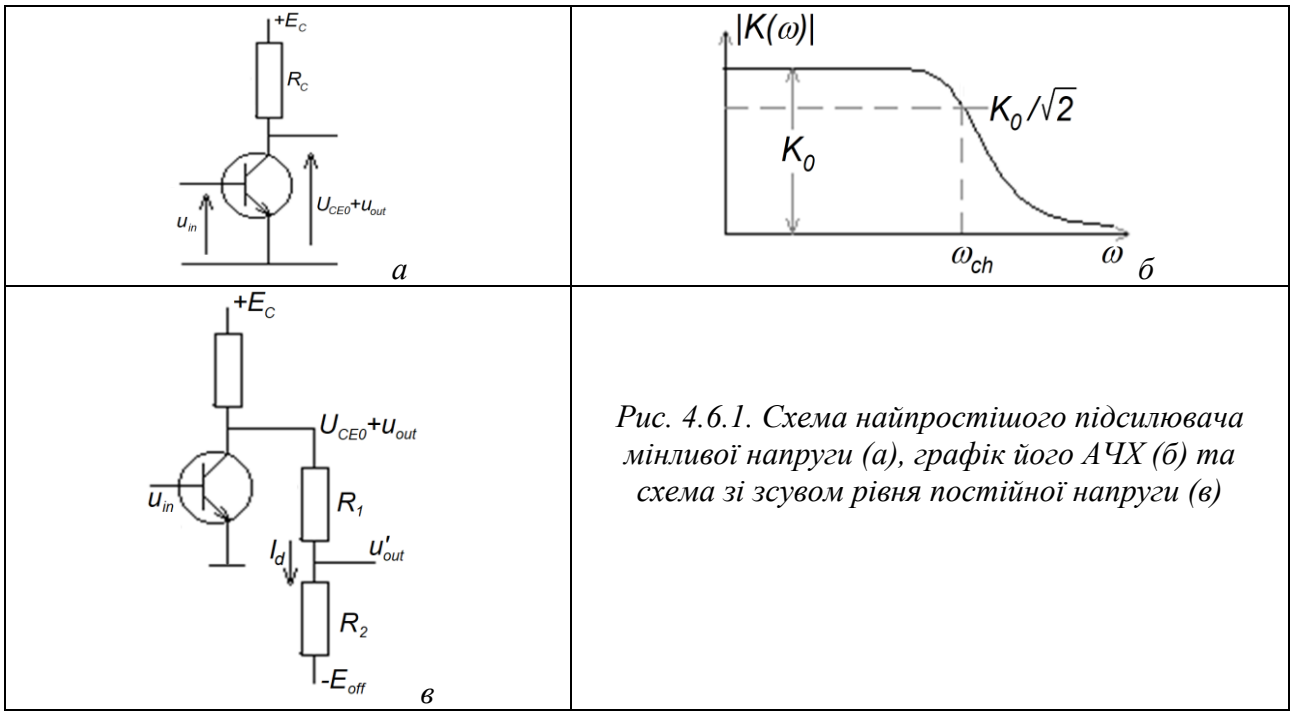


Рис. 4.6.1. Схема найпростішого підсилювача мінливої напруги (а), графік його АЧХ (б) та схема зі зсувом рівня постійної напруги (в)

Але в такій схемі виникає проблема, як на виході відділити сигнальну складову напруги u_{out} від режимної U_{CE0} (див. рис. 4.6.1 а). Необхідно добитися того, щоб при поданні на вхід нульової напруги вихідна напруга також дорівнювала нулеві. Найпростішим шляхом є використання додаткового джерела негативної напруги (див. рис. 4.6.1 в). Плечі подільника (voltage transmitter, voltage divider) R_1 та R_2 підбираються так, щоб за відсутності вхідного сигналу

падіння напруги на резисторі R_2 дорівнювало U_{CE0}. При цьому напруга на виході дорівнює нулеві. Недолік даної схеми – потреба в двополярному джерелі живлення та зменшення коефіцієнта підсилення в $R_2/(R_1+R_2)$ разів.

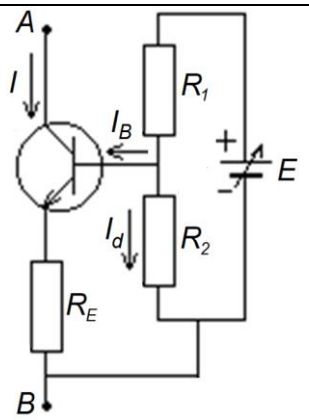


Рис. 4.6.2. Схема струмового дзеркала: омічний опір між точками А та В значно менший від диференціального

Другого з названих недоліків можна позбутися, використавши в подільнику нелінійний елемент, у якого диференціальний опір R_d значно відрізняється від омічного R_o. Одна з можливостей – замість резистора R_1 увімкнути стабілітрон (або набір стабілітронів), у якого $R_d \ll R_o$. Альтернативний шлях – замість резистора R_2 взяти елемент із великим диференціальним опором (при помірному омічному опорі). Таким елементом може бути транзистор з емітерним опором – так зване струмове дзеркало (рис. 4.6.2). Назва схеми пов'язана з тим, що в ній колекторний струм транзистора дорівнює струму подільника.

Справді, напруга на резисторі R_2 (рис. 4.6.2) може бути записана через струм подільника i_d та через емітерний струм транзистора $i+i_B$:

$$i_d R_2 = u_{BE} + (i+i_B)R_E \approx iR_E \qquad (4.6.1)$$

(враховано, що зазвичай $u_{BE} \ll i_d R_2$ та $i_B \ll i$). Тому $i \approx i_d R_2/R_E$, а при $R_2 = R_E$ маємо $i \approx i_d$. Таким чином, вихідний сигнальний струм схеми i (струм колектора) визначається переважно струмом подільника і практично не залежить від вихідної напруги u_{AB}, тобто

$$\frac{1}{R_{AB\,d}} = \frac{\partial i}{\partial u_{AB}} \approx 0, \qquad (4.6.2)$$

і вихідний диференціальний опір схеми виявляється дуже великим. Вихідний омічний опір $R_{AB\,o} = R_E + U_{CE0}/I_{C0}$ виявляється значно меншим.

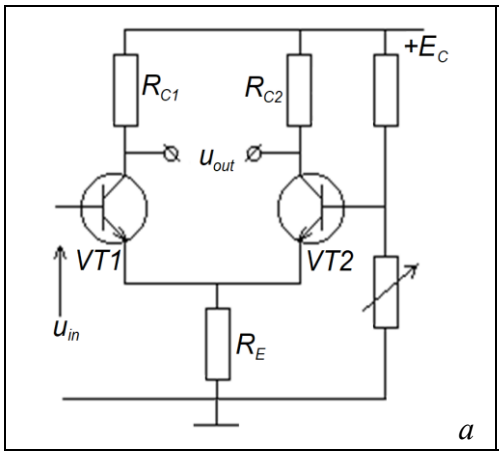
Ще одна проблема ПМН – це так званий дрейф нуля: величини I_{C0} та U_{CE0} з часом повільно змінюються через неконтрольовані випадкові зміни напруги джерел живлення, температури та інших чинників. При підсиленні високочастотних сигналів з цим можна не рахуватись, бо розділовий конденсатор на виході каскаду завжди відділяє сигнал від режимної постійної (або повільної) компоненти. Але в ПМН це неможливо, тому доводиться час від часу закорочувати вхід і виставляти на нуль вихідну напругу (в схемі, зображеній на рис. 4.6.1 в, це можна робити, змінюючи величину E_{off} або співвідношення між опорами R_1 і R_2).

Можливі шляхи боротьби з дрейфом нуля – стабілізація напруги джерел живлення, застосування термокомпенсації, використання негативного зворотного зв'язку за режимом (стабілізація робочої точки).

Однак застосування НЗЗ призводить до зменшення коефіцієнта підсилення ПМН, і вирішити завдання названими методами не вдається. Радикальним шляхом зменшення дрейфу нуля є застосування балансної схеми ПМН.

В такій схемі, зображеній на рис. 4.6.3 а, крім робочого транзистора $VT1$, на якій подається вхідний сигнал, є допоміжний транзистор $VT2$, ідентичний до $VT1$. Вихідний сигнал знімається між колекторами цих транзисторів.

Для кращого розуміння суті справ схема балансного підсилювача може бути зображена як звичайний міст (рис. 4.6.3 б). У ній транзистори $VT1$ і $VT2$ зображені як еквівалентні резистори R_{T1} і R_{T2}. При балансі моста зміни напруги джерела живлення E_C, увімкненого у вертикальну діагональ, не впливають на величину u_{out}, що знімається з горизонтальної діагоналі. Баланс моста зберігається й тоді, коли через зміну температури або з інших причин змінюються властивості і параметри транзисторів $VT1$ і $VT2$ (при цьому рівність R_{T1} і R_{T2} зберігається). Вихідний сигнал з'являється тільки внаслідок розбалансування моста, викликаного зміною режиму $VT1$ за рахунок дії вхідного сигналу. Балансні схеми зменшують дрейф нуля на 2 - 3 порядки і тому широко застосовуються в підсилювачах мінливої напруги.



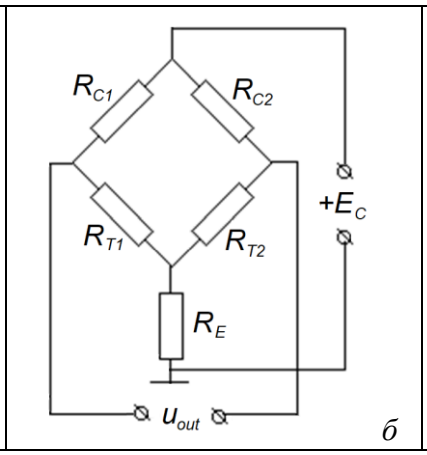


Рис. 4.6.3. Балансна схема підсилювача мінливих напруг (а) та її еквівалентна схема (б)

Зазначимо, що в балансних схемах ПМН одночасно розв'язується й проблема відсутності постійної напруги на виході за відсутності вхідного сигналу.

Недоліком балансної схеми є те, що вихідний сигнал виділяється між симетричними точками, жодна з яких не приєднана до землі. Тому безпосередньо підключати до них іншу схему, де одна з вхідних клем заземлена (а таких схем більшість), неможливо. Втім, ця проблема може бути розв'язана за рахунок використання спеціального пристрою – так званої оптронної пари, що складається зі світлодіода й фотодіода.

4.6.2. Селективний підсилювач

Складовою частиною приймачів прямого підсилення та супергетеродинних приймачів є підсилювач високої частоти, який повинен підсилювати сигнал у вузькій смузі навколо частоти-носія, а сигнали інших частот повинні придушуватися. Такі підсилювачі, які водночас виступають у ролі смугових фільтрів, називаються селективними.

Одна з можливостей реалізації селективного підсилювача полягає в тому, що в ролі опору навантаження для транзистора використовується паралельний коливний контур (рис. 4.6.4). Як було показано вище (п. 4.1.5), опір такого контуру досягає максимуму на його резонансній частоті, а при віддаленні від неї швидко спадає (див. формулу 4.1.33):

$$Z_{//}(\omega)=\frac{(i\omega L+R)(1/i\omega C)}{i\omega L+R+1/i\omega C}=\frac{i\omega L+R}{1-\omega^2LC+i\omega RC}=\frac{i\omega L+R}{1-\omega^2/\omega_0^2+i\omega RC}=\frac{L}{RC}\frac{1-iR/\omega L}{1-i\left(1-\omega^2/\omega_0^2\right)/\omega RC}=$$

$$=\frac{\rho^2}{R}\frac{1-\omega_0R/\omega\rho}{1-i\left(1-\omega^2/\omega_0^2\right)\omega_0\rho/\omega R}=\frac{\rho^2}{R}\frac{1-i\omega_0/\omega Q}{1-i\left(1-\omega^2/\omega_0^2\right)\omega_0Q/\omega}\approx\frac{R_{eq}}{1-2iQ\Delta\omega/\omega_0}, \qquad (4.6.3)$$

де $\Delta\omega=\omega_0-\omega,\ \ |\Delta\omega|\ll\omega_0$,

$$\rho=\sqrt{\frac{L}{C}};\qquad Q=\frac{\rho}{R}>>1;\qquad 1-\frac{\omega^2}{\omega_0^2}=\left(1-\frac{\omega}{\omega_0}\right)\left(1+\frac{\omega}{\omega_0}\right)\approx 2\frac{\Delta\omega}{\omega_0}<<1;\qquad R_{eq}=\frac{\rho^2}{R} \tag{4.6.4}$$

(під R тут слід розуміти активний опір паралельного контуру, перерахований до індуктивної гілки). Такі підсилювачі називають резонансними.

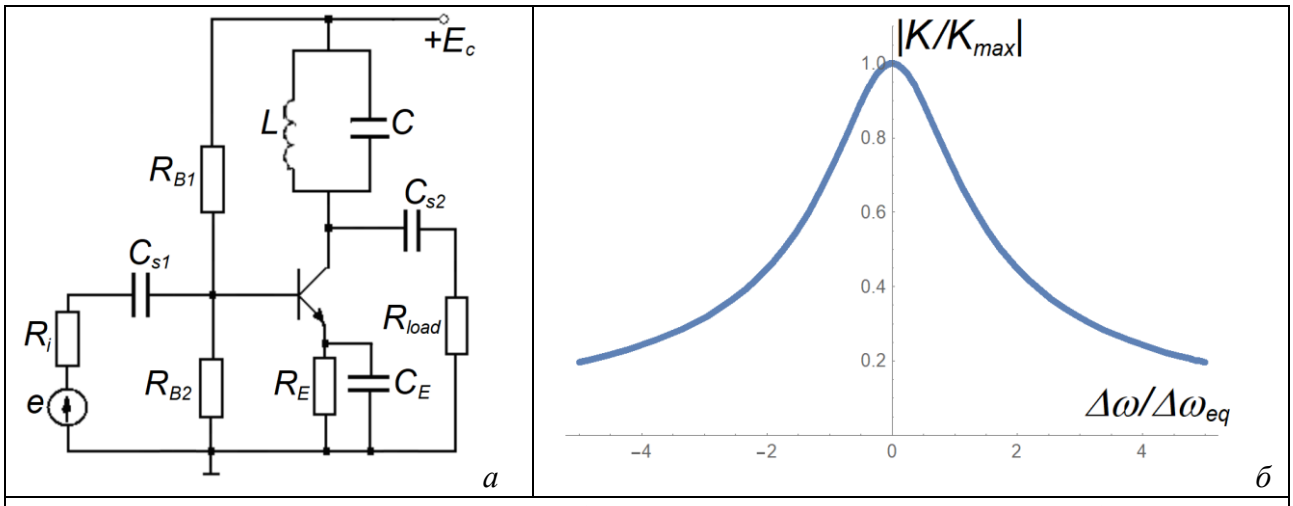


Рис. 4.6.4. Резонансний підсилювач на біполярному транзисторі: а – схема, б – типовий вигляд графіка його амплітудно-частотної характеристики

Принципова схема резонансного підсилювача на біполярному транзисторі показана на рис. 4.6.4 а (пор. з рис. 4.5.4). Вхідний сигнал від джерела e з внутрішнім опором R_i через розділовий конденсатор C_{s1} подається на базу транзистора. Необхідне зміщення на базі забезпечує подільник R_{B1}, R_{B2}. Навантаженням транзистора виступає LC-контур (його активний опір на схемі не показаний). R_E, C_E – ланцюжок термостабілізації, C_{s2} – вихідний розділовий конденсатор.

Амплітудно-частотну та фазово-частотну характеристику такого підсилювача можна розрахувати за формулою (4.5.3), куди замість R_C слід підставити комплексний опір $Z_{//}(\omega)$:

$$K_u\approx-\frac{h_{21E}}{h_{11E}}\frac{Z_{//}}{1+h_{22E}Z_{//}}\approx-\frac{h_{21E}}{h_{11E}}\frac{\dfrac{R_{eq}}{1-2iQ\Delta\omega/\omega_0}}{1+h_{22E}\dfrac{R_{eq}}{1-2iQ\Delta\omega/\omega_0}}=-\frac{h_{21E}}{h_{11E}}\frac{R_{eq}}{1-2iQ\Delta\omega/\omega_0+h_{22E}R_{eq}}= \tag{4.6.5}$$

$$=-\frac{h_{21E}R_{eq}}{h_{11E}\left(1+h_{22E}R_{eq}\right)}\frac{1}{1-2iQ\Delta\omega/\omega_0\left(1+h_{22E}R_{eq}\right)}\approx-\frac{K_{\max}}{1-i\,\Delta\omega/\Delta\omega_{eq}},$$

де

$$\Delta\omega_{eq}=\frac{\omega_0\left(1+h_{22E}R_{eq}\right)}{2Q}. \tag{4.6.6}$$

Відповідно

$$|K_u|=\frac{K_{\max}}{\sqrt{1+\left(\Delta\omega/\Delta\omega_{eq}\right)^2}}, \tag{4.6.7}$$

де

$$K_{\max}=\frac{h_{21E}R_{eq}}{h_{11E}\left(1+h_{22E}R_{eq}\right)}. \tag{4.6.8}$$

Графік АЧХ резонансного підсилювача поданий на рис. 4.6.4 б.

4.6.3. Емітерний повторювач

Емітерний повторювач – це пристрій, який підсилює потужність сигналу, але не підсилює його напругу. Основною перевагою емітерного підсилювача є великий вхідний та малий вихідний опір, що дозволяє використовувати його для узгодження джерела сигналу з великим внутрішнім опором та навантаження з невеликим опором. Це дозволяє уникнути небажаних втрат вхідної напруги на внутрішньому опорі джерела.

Схема емітерного повторювача подана на рис. 4.6.5 а-б. Навантаженням транзистора тут виступає резистор R_E в колі емітера, тобто транзистор увімкнений за схемою зі спільним колектором (за сигналом). Вихідна напруга знімається з емітера, а її фаза повторює фазу вхідного струму. Звідси й назва – емітерний повторювач.

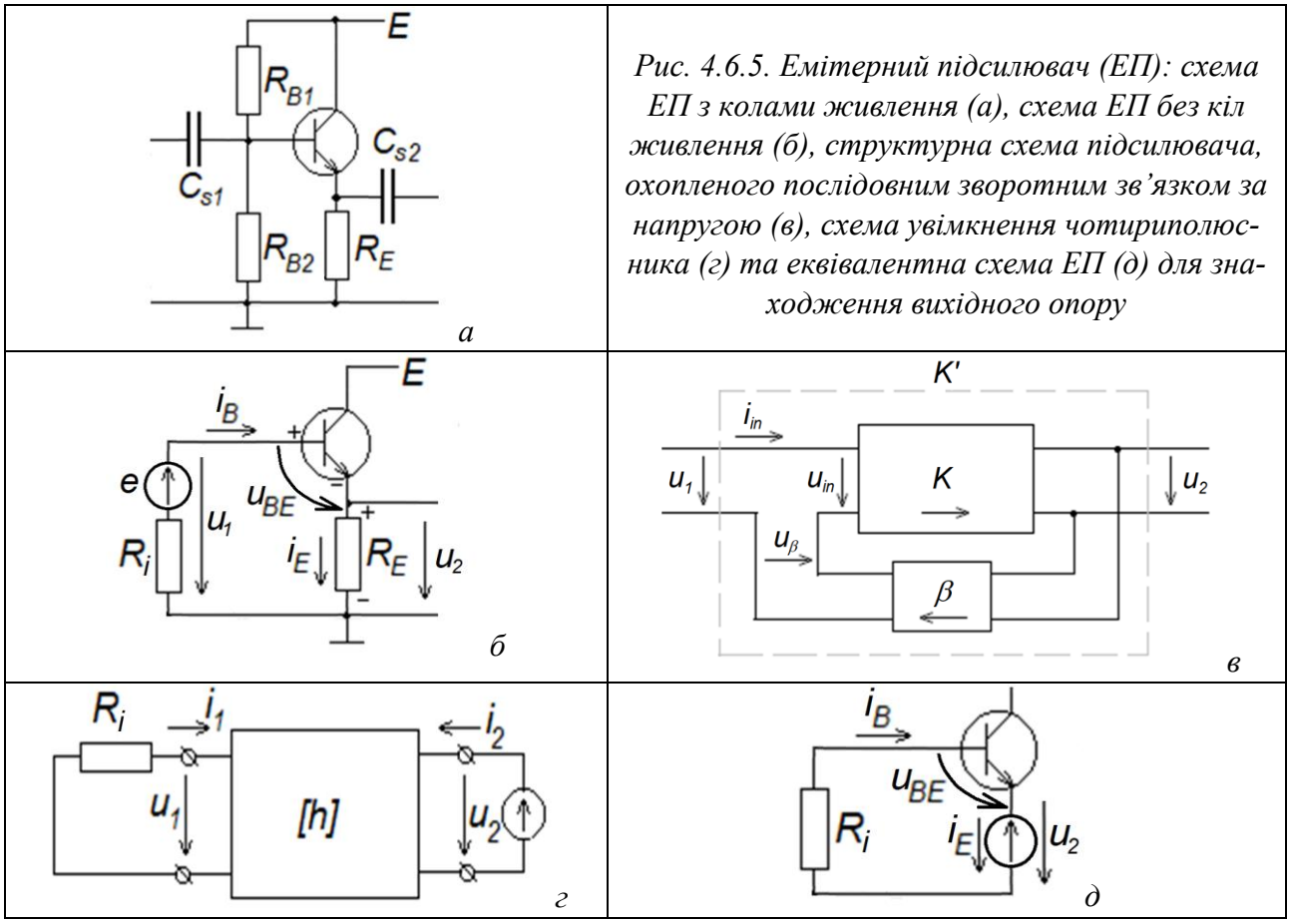


Рис. 4.6.5. Емітерний підсилювач (ЕП): схема ЕП з колами живлення (а), схема ЕП без кіл живлення (б), структурна схема підсилювача, охопленого послідовним зворотним зв'язком за напругою (в), схема увімкнення чотириполюсника (г) та еквівалентна схема ЕП (д) для знаходження вихідного опору

Порівняємо спрощену схему, подану на рис. 4.6.5 б, зі структурною схемою підсилювача, охопленого послідовним зворотним зв'язком за напругою (рис. 4.6.5 в, який повторює рис. 4.3.1). Вхідна u_1 та вихідна u_2 напруги на них позначені однаково. Роль напруги u_{in} у схемі емітерного повторювача відіграє напруга u_{BE}. В ролі ж напруги зворотного зв'язку виступає напруга u_2, яка віднімається від напруги u_1:

$$u_{BE} = u_1 - u_2. \qquad (4.6.9)$$

Таким чином, емітерний повторювач охоплений негативним зворотним зв'язком з коефіцієнтом $\beta = -1$, і його «зовнішній» коефіцієнт підсилення можна знайти за формулою (4.3.3):

$$K' = \frac{u_2}{u_1} = \frac{K}{1 - K\beta} = \frac{K}{1 + K} < 1. \qquad (4.6.10)$$

Очевидно, «внутрішній» коефіцієнт підсилення за напругою K в емітерному повторювачі аналогічний (з точністю до знаку) до отриманого раніше для підсилювача з навантаженням у колі колектора (див. вище формулу (4.5.5)):

$$K = \frac{u_2}{u_{BE}} \approx \frac{h_{21E} R_E}{h_{11E}\left(1 + h_{22E} R_E\right)} \approx h_{21E} R_E / h_{11E}$$

(за умови $h_{22E} R_E \ll 1$).

Оскільки $K \gg 1$, то, як впливає з формули (4.5.36), $K' \approx 1$.

Для знаходження вхідного опору емітерного повторювача скористаємося формулою (4.38 а) для підсилювача, охопленого негативним зворотним зв'язком. Вважаючи, що на середніх частотах «внутрішній» вхідний опір суто активний, $Z_{in} = R_{in}$, і, відповідно до формули (4.5.4), $R_{in} \approx h_{11E}$, отримуємо:

$$R'_{in} = R_{in}\left(1 + |K\beta|\right) \approx h_{11E}\left(1 + K\right) \approx h_{11E} + h_{21E} R_E. \qquad (4.6.11)$$

Для знаходження вихідного опору R'_{out} емітерного повторювача скористаємося схемою, поданою на рис. 4.6.5 г (вона повторює рис. 4.5.3, пояснення до неї подані в п. 4.5.2). Замінивши в ній довільний чотириполюсник на емітерний повторювач, приходимо до еквівалентної схеми, показаної на рис. 4.6.5 д.

Запишемо рівняння за другим правилом Кірхгофа для замкненого контуру R_i – u_{BE} – u_2 (рис. 4.6.5 д):

$$i_B R_i + u_{BE} = u_2.$$

Врахуємо, що $u_{BE} \approx i_B h_{11E}$. Тоді можна записати струм i_B через напругу u_2:

$$i_B = \frac{u_2}{R_i + h_{11E}}. \qquad (4.6.12)$$

В свою чергу, $i_E = i_B\left(1 + h_{21E}\right)$.

Вихідний опір визначається як $R'_{out} = u_2 / i_E$. Послідовно підставляючи формули для i_E через i_B та i_B через u_2, отримуємо:

$$R_{out}' = \frac{R_i + h_{11E}}{1 + h_{21E}}. \qquad (4.6.13)$$

Якщо $R_i \ll h_{11E}$, то

$$R'_{out} \approx h_{11E} / h_{21E}. \qquad (4.6.13\ а)$$

З формул (4.5.37) та (4.5.39) випливає, що

$$\frac{R'_{in}}{R_{out}'} \approx h_{21E} \frac{h_{11E} + h_{21E} R_E}{R_i + h_{11E}} \gg 1.$$

Знайдемо ще коефіцієнт підсилення емітерного підсилювача за потужністю. Нехай амплітуда сигналу на вході – u_{m1}, на виході – u_{m2}. Тоді середні потужності сигналу відповідно на вході та на виході можна записати як $P_{in} = u_{m1}^2 / R'_{in}$ та $P_{out} = u_{m2}^2 / R'_{out}$. Але $u_{m2} / u_{m1} = K' \approx 1$, тому коефіцієнт підсилення за потужністю має вигляд

$$K_P = \frac{P_{out}}{P_{in}} \approx \frac{R'_{in}}{R'_{out}} >> 1.$$

Для оцінки можливих значень всіх цих величин візьмемо як приклад типові значення h

-параметрів біполярного транзистора середньої чи малої потужності, увімкненого зі спільним емітером: $h_{11_E} = 1$ кОм, $h_{12_E} = 3 \cdot 10^{-4}$, $h_{21_E} = 40$ та $h_{22_E} = 5 \cdot 10^{-5}$ Сим (пор. з п. 4.5.2). Нехай опір емітерного резистора становить $R_C = 5$ кОм, а внутрішній опір джерела сигналу – $R_i = 200$ Ом. Підставивши ці значення до отриманих формул, одержимо: $K = 167$, $K' = 0.994$, $R'_{in} \approx 200$ кОм, $R_{out} = 150$ Ом, $K_P \approx 1.3 \cdot 10^3$.

Схеми, аналогічні емітерному повторювачу, можна реалізувати також на основі польових транзисторів (так званий стоковий повторювач).

Контрольні питання до підрозділу 4.6

1. В чому особливості підсилювача мінливої напруги? Як його можна реалізувати?
2. Для чого потрібен зсув рівня вихідної напруги в підсилювачі мінливої напруги?
3. Опишіть принцип роботи резонансного підсилювача.
4. Для чого використовується емітерний повторювач?

Задачі до підрозділу 4.6

4.6.1. Резонансний підсилювач налаштований на частоту 1 МГц, добротність коливного контура – 30. Параметри транзистора ті самі, що в задачі 4.5.3. Знайти максимальний коефіцієнт підсилення за напругою та ширину смуги пропускання.

4.6.2. Скориставшись даними задачі 4.5.3, знайти елементи схеми емітерного повторювача, поданої на рис. 4.5.15 а.

4.6.3. Скориставшись результатами задачі 4.6.2, знайти коефіцієнт передачі за напругою, коефіцієнт підсилення за потужністю, вхідний і вихідний опір емітерного повторювача. Дані про *h*-параметри транзистора взяти з задачі 4.5.3. Вважати, що $R_i = 0.5$ кОм.

4.7. Параметричні підсилювачі

Наостанок розглянемо підсилювачі, принципово відмінні від підсилювачів на транзисторах. В основі їхньої роботи лежить не керований подільник напруги (див. вище п. 4.2.1), а так звана параметрична взаємодія коливань.

За означенням, параметричні системи – це системи, у яких деякий параметр змінюється з часом під дією зовнішнього впливу незалежно від руху самої системи. З математичної точки зору такі системи описуються лінійними диференціальними рівняннями зі змінними коефіцієнтами. Зовнішній вплив, який породжує зміну параметричного елемента, називають накачуванням, або помпуванням (від англ. pumping – накачування).

Найбільш цікавий випадок, коли параметрична система є коливною й характеризується певною власною частотою, а накачування й спричинена ним зміна параметра періодично залежать від часу. Тоді за певних співвідношень між частотою накачування й частотою власних коливань спостерігається зростання амплітуди останніх – так званий параметричний резонанс. Найвідомішим прикладом параметричних систем описаного типу є маятник зі змінною довжиною підвісу (людина на гойдалці).

У параметричних підсилювачах в ролі елемента, що змінюється з часом, виступає деякий конденсатор (на практиці – варикап), до якої прикладено зовнішню змінну напругу – напругу накачування.

4.7.1. Параметричний підсилювач: принцип роботи

Перш ніж говорити про параметричний підсилювач, обговоримо спершу природу параметричного резонансу.

<u>Параметричний резонанс.</u>

Розглянемо коливання в контурі зі змінною ємністю (рис. 4.7.1, *а*). Нехай у моменти, коли напруга на ємності максимальна, пластини конденсатора розсуваються, тим самим зменшуючи його ємність. У моменти ж, коли напруга дорівнює нулю, пластини знову зсуваються, збільшуючи ємність (рис. 4.7.1, *б*, верхня епюра). Зазначимо, що період коливань ємності (період накачування) удвічі менший від періоду власних коливань контуру.

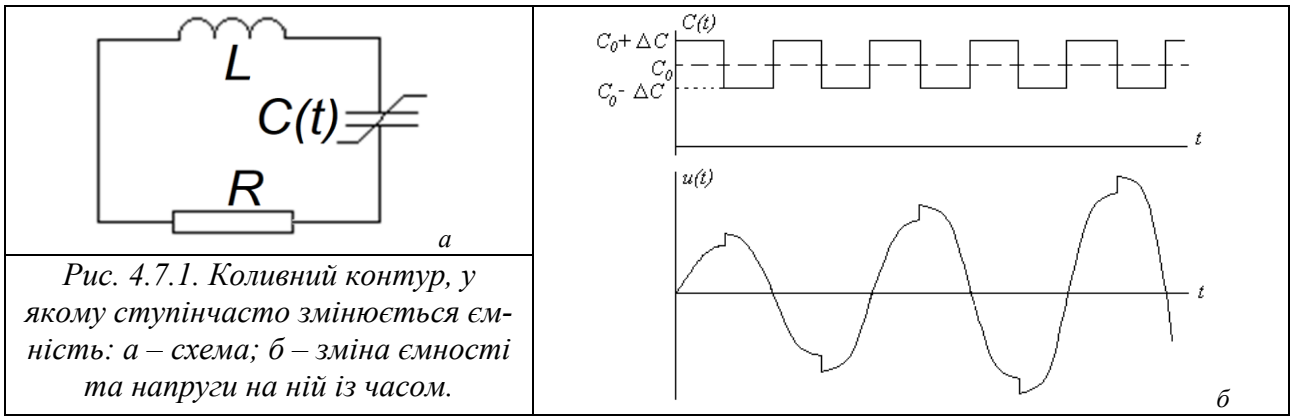


Рис. 4.7.1. Коливний контур, у якому ступінчасто змінюється ємність: а – схема; б – зміна ємності та напруги на ній із часом.

В моменти, коли ємність зменшується (тобто пластини конденсатора розсуваються), заряд на них не встигає змінитись (у ці моменти часу струм у колі дорівнює нулю). Отже, при цьому стрибком зростає напруга на конденсаторі – $U = q/C$ – і збільшується накопичена ним енергія – $W = q^2/2C$ (за рахунок роботи зовнішньої сили, що діє проти кулонівських сил). У моменти, коли ємність зростає (пластини зсуваються), вся енергія зосереджена в котушці індуктивності і, отже, залишається незмінною. Таким чином, амплітуда коливань у контурі з часом зростає (рис. 4.7.1, *б*, нижня епюра).

Проаналізуємо розглянутий процес кількісно. Вважатимемо глибину модуляції ємності $m = \Delta C/C_0$ малою ($m \ll 1$), а добротність контуру $Q = \sqrt{L/C_0}/R \gg 1$ – великою (R -

активний опір у контурі). Тоді відносна зміна амплітуди коливань за період також буде малою, тобто коливання наближено можна вважати гармонічними.

При зміні ємності на величину dC енергія конденсатора (при незмінному заряді) змінюється на величину

$$dW = d\left(\frac{q^2}{2C}\right) = -\frac{q^2}{2C}\frac{dC}{C}. \qquad (4.7.1)$$

У нашому випадку $dC = -2\Delta C$, розсування пластин здійснюється двічі за період коливань, тому енергія, що надходить у контур за період коливань від джерела накачування (унаслідок зміни ємності), буде

$$\Delta W_+ = 2\frac{q_m^2}{2C_0}\frac{2\Delta C}{C_0} = 2m\frac{q_m^2}{C_0}. \qquad (4.7.2)$$

Енергія втрат на активному опорі за період коливань, які наближено вважаються гармонічними, складає

$$\Delta W_- = \frac{1}{2}RI_m^2\frac{2\pi}{\omega_0} = \pi R q_m^2 \omega_0, \qquad (4.7.3)$$

($\omega_0 = 1/\sqrt{LC_0}$ – частота власних коливань) де враховано, що для гармонічних коливань $I_m = \omega_0 q_m$.

Амплітуда коливань зростає, коли енергія, що надходить у контур за період його коливань, перевищує втрати, тобто $\Delta W_+ > \Delta W_-$, або

$$m > m_{cr} \equiv \frac{\pi}{2}R\omega_0 C_0 = \frac{\pi}{2Q}. \qquad (4.7.4)$$

Отже, нерівність (4.7.4) є умовою параметричного резонансу. Бачимо, що він має поріг (за глибиною модуляції ємності), обумовлений дисипацією в коливній системі.

Явище параметричного резонансу лежить в основі роботи параметричного генератора. Оскільки в розглянутому прикладі частота накачування (точніше, основна гармоніка сигналу накачування) вдвічі більша за частоту збуджуваних коливань, то з практичної точки зору параметричний генератор (у найпростішому випадку) ділить частоту прикладеної напруги навпіл. Відповідно практична цінність такого пристрою обмежена.

<u>Дисипативний параметричний осцилятор як регенеративний підсилювач</u>**.**

Якщо $m < m_{cr}$, то енергія, яка надходить від джерела накачування, лише частково компенсує втрати на активному опорі. Це можна інтерпретувати як зменшення активного опору контуру і збільшення його еквівалентної добротності:

$$\begin{aligned}\Delta W_- \to \Delta W_- - \Delta W_+ &= \pi\omega_0 R q_m^2 - 2m q_m^2/C_0 = \\ &= \pi\omega_0 R q_m^2\left(1 - m/m_{cr}\right) = \pi\omega_0 R_{eff} q_m^2,\end{aligned} \qquad (4.7.5)$$

$$R_{eff} = R\left(1 - m/m_{cr}\right),\ Q_{eff} = \frac{1}{R_{eff}}\sqrt{\frac{L}{C_0}} = \frac{Q}{1 - m/m_{cr}}. \qquad (4.7.6)$$

У результаті падіння напруги на реактивних елементах зростає в

$$K = \frac{Q_{eff}}{Q} = \frac{1}{1 - m/m_{cr}} \qquad (4.7.7)$$

разів порівняно з випадком, коли накачування відсутнє. Структура формули (4.67) така сама, як і формули для коефіцієнта підсилення підсилювача, охопленого позитивним зворотним зв'язком з коефіцієнтом β (див. вище п. 4.3.2, формула (4.3.5))

$$K' = \frac{K}{1 - K\beta}.$$

Підвищення коефіцієнта підсилення підсилювача за рахунок його охоплення позитивним зворотним зв'язком називають, як уже вказувалося, регенеративним підсиленням. Отже,

у випадку, коли умова (4.7.4) не виконана, параметричний контур виступає в ролі підсилювача, що працює за регенеративним механізмом. Як і звичайний регенеративний підсилювач, при виконанні умови самозбудження він перетворюється на автогенератор.
Практична цінність параметричного підсилювача пов'язана з низьким рівнем його власних шумів, що дає змогу використати його для підсилення слабких сигналів. А низький рівень власних шумів спричинений відсутністю в його схемі транзисторів, які в звичайних підсилювачах є основним джерелом шумів (див. нижче п. 10.1).

Умови справедливості параметричного наближення.

Зазвичай на практиці ніхто не змінює ємність конденсатора шляхом механічної зміни відстані між пластинами. У ролі змінної ємності, як уже говорилося, виступає варикап, на який подається відповідна керувальна напруга накачування $e_P(t)$ – як правило, гармонічна (рис. 4.7.2). Стала напруга E_0 забезпечує вибір робочої точки на вольт-фарадній характеристиці варикапа (див. вище рис. 3.3.4).

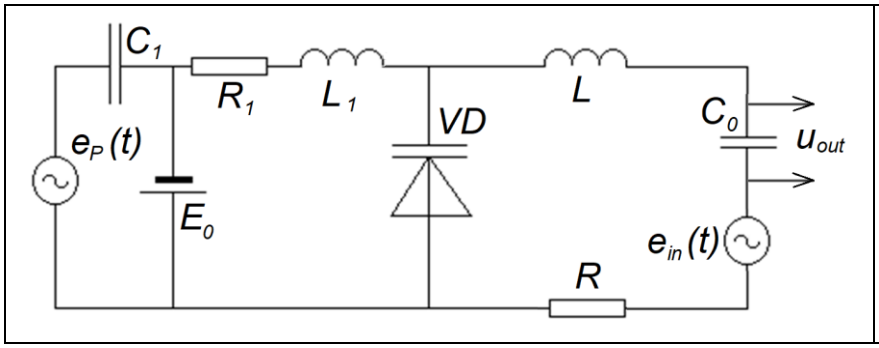


Рис.4.7.2. Схема реального параметричного підсилювача з варикапом. Правий контур налаштований на частоту сигналу, лівий – на частоту накачування

Отже, маємо нелінійний елемент, на який діє зовнішня сила (напруга $e_P(t)$), тобто права частина схеми на рис. 4.7.2 являє собою не параметричний, а нелінійний неавтономний осцилятор. Справді, ємність варикапа може змінюватися не лише під впливом напруги накачування, але й під впливом напруги сигналу в контурі. Останнім ефектом, очевидно, можна знехтувати в тому випадку, коли напруга сигналу за амплітудою значно менша від напруги накачування. Це і є умова справедливості параметричного наближення.

Оскільки в параметричному генераторі напруга сигналу з часом зростає, то параметричне наближення справджується лише в початкові моменти часу. А от для параметричних підсилювачів, де таке зростання відсутнє, модель параметричної системи добре працює (при помірних амплітудах вхідного сигналу).

4.7.2. Одноконтурний параметричний підсилювач

Для кількісного опису параметричного підсилювача з гармонічним накачуванням використовують так зване рівняння Матьє.

Виведення рівняння Матьє

Вважаючи виконаними умови параметричного наближення, розглядатимемо коливання в контурі з параметричною ємністю (рис.4.7.1). Ємність контуру змінюється за законом:

$$C(t) = C_0(1 - m\cos\omega_p t) \qquad (4.7.8)$$

Рівняння для коливань у цьому контурі можна записати за II правилом Кірхгофа:

$$L\frac{di}{dt} + Ri + \frac{1}{C(t)}\int idt = 0. \qquad (4.7.9)$$

Перейдемо від струму до заряду ємності, скориставшись співвідношенням $q = \int idt$.

Врахуємо, що $m << 1$; тоді

$$\frac{1}{C(t)} \approx \frac{1}{C_0}\left(1 + m\cos\omega_H t\right). \qquad (4.7.10)$$

Підставивши (4.7.10) до (4.7.9) і скориставшись звичайними позначеннями $\omega_0^2=1/LC_0$, $\delta=R/2L$, остаточно дістанемо:

$$\ddot{q} + 2\delta\dot{q} + \left(1 + m\cos\omega_p t\right)\omega_0^2 q = 0 \qquad (4.7.11)$$

Це – *рівняння Матьє*, що описує коливання параметричного осцилятора з гармонічним накачуванням. З математичної точки зору це лінійне диференціальне рівняння, один з коефіцієнтів якого залежить від часу.

Розв'язання неоднорідного рівняння Матьє

Проаналізуємо роботу одноконтурного параметричного підсилювача (рис. 4.7.3 а) детальніше. Для цього скористаємося неоднорідним рівнянням Матьє

$$\ddot{q} + 2\delta\dot{q} + \left(1 + m\cos\omega_p t\right)\omega_0^2 q = f_m \exp\left(i\omega t\right), \qquad (4.7.12)$$

права частина якого враховує наявність у параметричному контурі джерела зовнішньої гармонічної напруги. При записі рівняння ми скористалися методом комплексних амплітуд (див. вище п. 2.2.1).

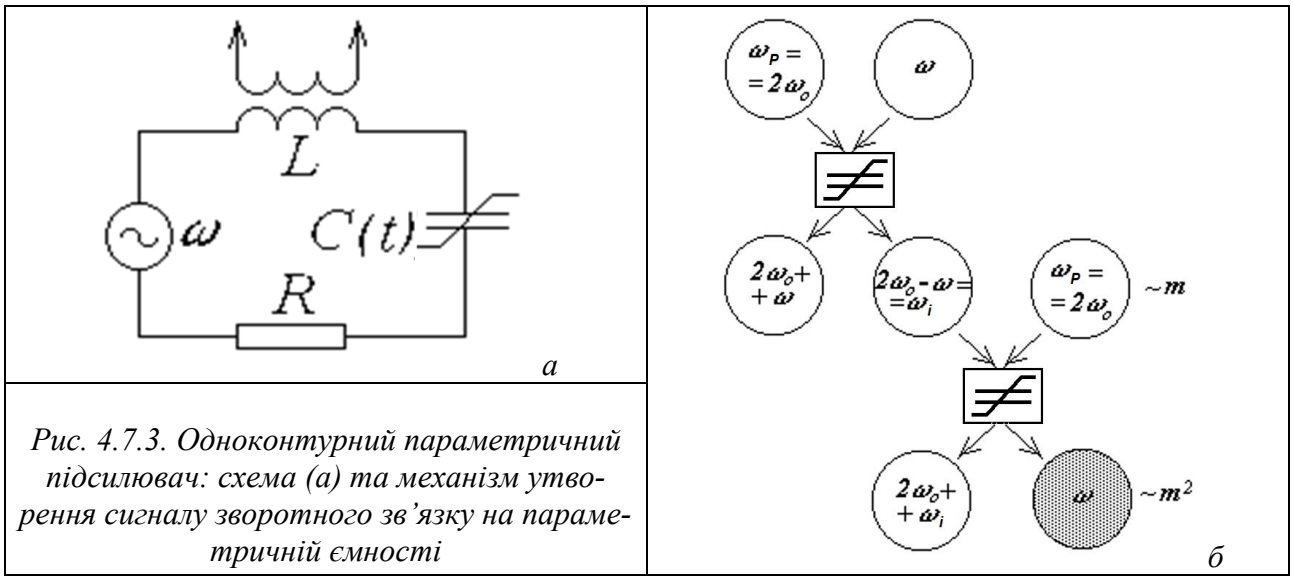


Рис. 4.7.3. Одноконтурний параметричний підсилювач: схема (а) та механізм утворення сигналу зворотного зв'язку на параметричній ємності

Вважатимемо, що:

- $\omega_p=2\omega_0$ (частота накачування рівно вдвічі більша від власної частоти контуру);
- амплітуда зовнішньої сили f_m мала (це умова застосовності параметричного наближення, див. вище п. 4.7.1);
- $\delta<<\omega_0$ (контур високодобротний);
- $|\Delta|<<\omega_0$, де $\Delta\equiv\omega_0-\omega$ (частота вхідного сигналу близька до власної частоти контуру);

Взаємодія сигналу з накачуванням на параметричній ємності буде породжувати різницеву частоту $\omega_i\equiv2\omega_0-\omega$ – так звану *холосту*, або ялову (англ. idler, неодружений) частоту. Оскільки вона також близька до резонансної ($\omega_0-\omega_i=-\Delta$), її слід обов'язково брати до уваги. Справді, саме взаємодія холостої частоти з накачуванням породжує на параметричній ємності "сигнал позитивного зворотного зв'язку" з частотою ω (рис. 4.7.3 б), наявність якого є необхідною умовою регенеративного підсилення сигнальної частоти.

Розглядатимемо усталені коливання. Тоді їхні амплітуди можна вважати сталими. Отже, розв'язок рівняння (4.7.12) можна шукати як суму коливань на сигнальній та холостій

частотах[9]:

$$q(t) = A\exp(i\omega t) + B\exp(-i\omega_i t). \qquad (4.7.13)$$

Підставимо розв'язок (4.7.13) до (4.7.12). Відкинемо усі доданки другого та більш високих порядків мализни, врахувавши, зокрема, що $\omega_0^2-\omega^2\approx 2\omega_0\Delta$ та $\omega_0^2-\omega_i^2\approx -2\omega_0\Delta$. Замінимо ω та ω_i на ω_0 всюди, крім аргументів експонент. Відкинемо коливання з частотами $2\omega_0+\omega$ та $2\omega_0+\omega_i$, далекими від резонансної частоти контуру. Отримаємо:

$$(2\omega_0\Delta + i2\omega_0\delta)A\exp(i\omega t) + (-2\omega_0\Delta - i2\omega_0\delta)B\exp(-i\omega_i t) +$$

$$+\frac{m\omega_0^2}{2}\left[A\exp(-i\omega_i t) + B\exp(i\omega t)\right] = f_m\exp(i\omega t). \qquad (4.7.14)$$

Прирівнюючи амплітуди однакових експонент в обох частинах, дістанемо систему лінійних алгебраїчних рівнянь щодо амплітуд коливань на сигнальній та холостій частотах:

$$\begin{cases}(\Delta + i\delta)A + (m\omega_0/4)B = f_m/2\omega_0;\\ -(m\omega_0/4)A + (\Delta + i\delta)B = 0.\end{cases} \qquad (4.7.15)$$

Звідси

$$A = \frac{(\Delta + i\delta)f_m/2\omega_0}{(\Delta + i\delta)^2 + (m\omega_0/4)^2}, \quad B = -\frac{mf_m/8}{(\Delta + i\delta)^2 + (m\omega_0/4)^2}. \qquad (4.7.16)$$

У випадку точного резонансу ($\Delta=0$) вирази (4.7.9) набувають вигляду

$$A = -i\frac{Qf_m/\omega_0^2}{1-(m/m_{cr})^2}, \; B = \frac{m}{m_{cr}}\frac{Qf_m/\omega_0^2}{1-(m/m_{cr})^2}, \qquad (4.7.17)$$

де m_{cr} визначається формулою

$$m_{cr} \equiv \frac{4\delta}{\omega_0} = \frac{2}{Q}. \qquad (4.7.18)$$

Графіки залежностей $|A|$ та $|B|$ від m подано на рис. 4.7.4 а. При $m>m_{cr}$ відбувається самозбудження.

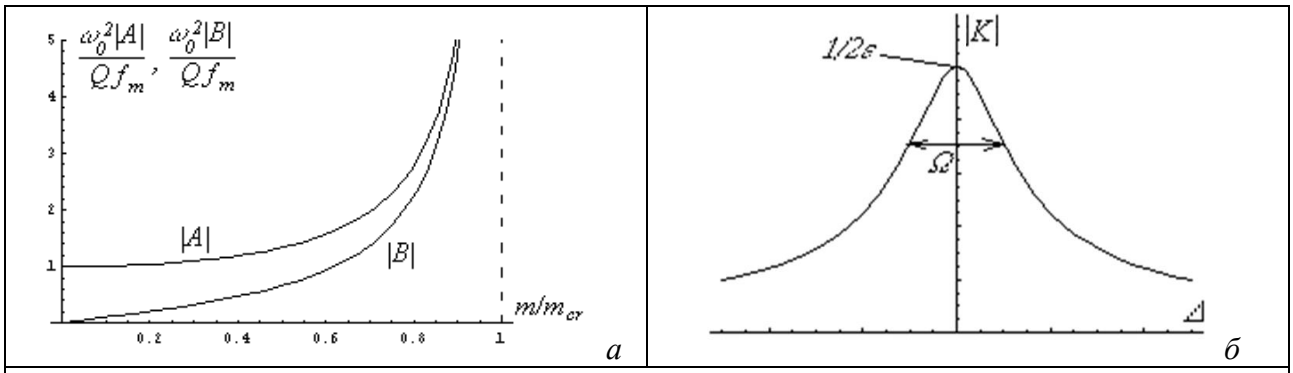


Рис.4.7.4. Одноконтурний параметричний підсилювач: а – амплітуди сигналу та холостої частоти в залежності від глибини модуляції параметричної ємності в режимі точного резонансу; б – амплітудно-частотна характеристика

4.7.3. Форма амплітудно-частотної характеристики

Коефіцієнтом підсилення параметричного підсилювача природно вважати відношення амплітуди сигналу до її значення за відсутності накачування:

$$K = \frac{A}{A|_{m=0}} = \frac{1}{1+\left[m\omega_0/4(\Delta + i\delta)\right]^2}. \qquad (4.7.19)$$

---

[9] Така форма розв'язку обрана тому, що в рівнянні Матьє добуток $exp(i\omega t)cos(2\omega_0 t)$ породжує коливання з холостою частотою у формі $exp(-i\omega_x t)$, і, навпаки, добуток $exp(-i\omega_x t)cos(2\omega_0 t)$породжує коливання з сигнальною частотою у формі $exp(i\omega t)$. Можна було б дописати в (1.6.13) комплексно спряжені доданки, але вони не дадуть нічого нового.

Вважатимемо, що

$$|\Delta| << \delta. \qquad (4.7.20)$$

Домножимо чисельник і знаменник другого доданку в знаменнику (4.7.19) на $\Delta - i\delta$ і знехтуємо Δ^2 поруч із δ^2. Отримаємо:

$$K = \frac{1}{1+\left[\frac{m\omega_0(\Delta - i\delta)}{4\delta^2}\right]^2} = \frac{1}{1-\left[\frac{m\omega_0}{4\delta}\left(1+\frac{i\Delta}{\delta}\right)\right]^2} = = \frac{1}{\left(1+\frac{m}{m_{cr}}+i\frac{\Delta}{\delta}\frac{m}{m_{cr}}\right)\left(1-\frac{m}{m_{cr}}-i\frac{\Delta}{\delta}\frac{m}{m_{cr}}\right)}. \qquad (4.7.21)$$

Як видно з рис. 4.7.4 а, помітний коефіцієнт підсилення досягається лише при $m \to m_{cr}$. Введемо параметр $\varepsilon \equiv 1 - m/m_{cr}$ і вважатимемо, що

$$0 < \varepsilon << 1. \qquad (4.7.22)$$

Тоді перша дужка в знаменнику (4.7.21) близька до 2, і остаточно отримаємо такі вирази для коефіцієнта підсилення та ширини смуги пропускання:

$$K = \frac{1}{2\varepsilon\left(1 - i\Delta/\varepsilon\delta\right)}, \quad \Omega = 2\varepsilon\delta. \qquad (4.7.23)$$

Графік амплітудно-частотної характеристики, що відповідає (4.7.23), поданий на рис.4.7.4 б. Видно, що при наближенні до порогу нестійкості коефіцієнт ($\varepsilon \to 0$) підсилення зростає, а ширина смуги пропускання зменшується. Така поведінка типова для регенеративних підсилювачів, тобто підсилювачів, охоплених позитивним зворотним зв'язком (див. вище п. 4.3.5).

Оскільки ширина смуги пропускання підсилювача Ω в силу (4.7.22) вийшла значно меншою, ніж для «холодного» контуру (без накачування), то зроблене припущення (4.7.20) добре виконується.

4.7.4. Пульсації коефіцієнту підсилення

Визначений вище (див. (4.7.19)) коефіцієнт підсилення відповідає випадку, коли на вихід підсилювача не надходить сигнал холостої частоти[10]. Насправді, як випливає з (4.7.16), у контурі параметричного підсилювача одночасно існують коливання з частотами сигналу та холостою, амплітуди яких при малих розстроюваннях та при $m \to m_{cr}$ близькі. Очевидно, суперпозиція цих коливань – биття – сприймається як пульсації коефіцієнту підсилення.

Різниця фаз між коливаннями на сигнальній та холостій частоті складає $\varphi_1 - \varphi_2 = 2\Delta t$ й монотонно зростає з часом. Амплітуда коливань у контурі q_m пульсує в межах від $|A|+|B|$ (коли $\varphi_1 - \varphi_2 = 2\pi n, \quad n = 1, 2, 3...$) до $|A|-|B|$ (коли $\varphi_1 - \varphi_2 = \pi(2n+1)$). Якщо вважати розстроювання значно меншим від ширини смуги пропускання (1.6.23),

$$\left(\frac{\Delta}{\varepsilon\delta}\right)^2 << 1, \qquad (4.7.24)$$

то для амплітуд A і B можна скористатись формулами (4.7.16). В результаті для мінімальної та максимальної амплітуд коливань у контурі дістанемо:

$$q_{\min,\max}(m) = |A| \mp |B| = \frac{Qf_m}{\omega_0^2}\frac{1}{1 \pm m/m_{cr}}. \qquad (4.7.25)$$

Часовий хід сигналу (4.7.13) показаний на рис.4.7.5. При цьому коефіцієнт підсилення

$$K = \frac{q_m(m)}{q_m(m=0)} \qquad (4.7.26)$$

пульсує в межах

$$\frac{1}{\left(1 - m/m_{cr}\right)} > K > \frac{1}{\left(1 + m/m_{cr}\right)}. \qquad (4.7.27)$$

Звичайно, в середньому за часом сигнал все-таки підсилюється, тому пульсації в принципі можна відфільтрувати за допомогою вузькосмугового фільтра. Але на практиці

[10] Для цього потрібно пропустити сигнал через смуговий фільтр із смугою пропускання $\Omega < \Delta$.

підсилюються немонохроматичні (модульовані) сигнали (з ненульовою шириною спектру), і пропускання їх через вузькосмуговий фільтр може призвести до втрати корисної інформації. Тому наявність пульсацій коефіцієнту підсилення, безумовно, є принциповим недоліком одноконтурного параметричного підсилювача.

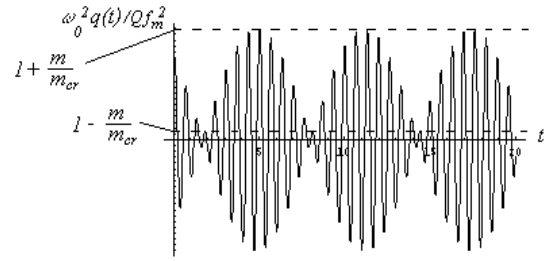


Рис.4.7.5. Пульсації сигналу в одноконтурному параметричному підсилювачі

4.7.5. Двоконтурний параметричний підсилювач

Для того, щоб позбутися пульсацій коефіцієнту підсилення, слід ускладнити схему, взявши замість одноконтурного параметричного підсилювача двоконтурний (рис. 4.7.6; коло накачування варикапа не показане).

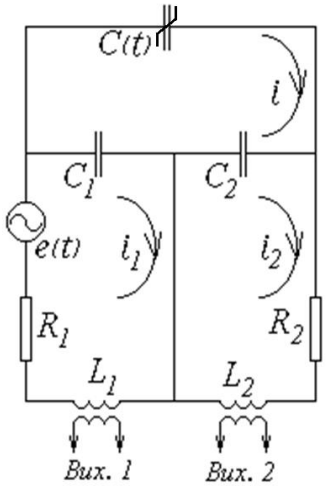


Рис. 4.7.6. Схема двоконтурного параметричного підсилювача.

<u>Принцип роботи двоконтурного параметричного підсилювача</u>

У двоконтурному параметричному підсилювачі, на відміну від одноконтурного, сигнальна та холоста частоти суттєво відрізняються одна від одної. Коливання на кожній із цих частот збуджуються в окремому коливному контурі, причому смуги пропускання контурів не перекриваються. Взаємодія між коливаннями на сигнальній та холостій частоті відбувається на параметричній ємності, яка одночасно забезпечує зв'язок між двома контурами. Таким чином, пульсації коефіцієнту підсилення в такій схемі не виникають.

<u>Вихідні рівняння</u>

Отримаємо рівняння, що описують роботу двоконтурного параметричного підсилювача (рис.4.7.6). Вважатимемо, що ємність варикапа змінюється з часом за законом

$$C(t) = C_0\left(1 + m\cos\omega_p t\right), \qquad m < 1. \tag{4.7.28}$$

Для контурів зі струмами i, i_1, i_2 запишемо рівняння за другим правилом Кірхгофа:

$$L_1\frac{di_1}{dt} + R_1 i_1 + \frac{1}{C_1}\int\left(i_1 - i\right)dt = e(t);$$

$$L_2\frac{di_2}{dt} + R_2 i_2 + \frac{1}{C_2}\int\left(i_2 - i\right)dt = 0; \tag{4.7.29}$$

$$\frac{1}{C(t)}\int idt+\frac{1}{C_1}\int\left(i-i_1\right)dt+\frac{1}{C_2}\int\left(i-i_2\right)dt=0.$$

Перейдемо від струмів до зарядів на конденсаторах:

$$q_1=\int i_1dt\,,\qquad q_2=\int i_2dt \tag{4.7.30}$$

Тоді з останнього рівняння системи (4.7.29) випливає, що

$$\left(\frac{1}{C(t)}+\frac{1}{C_1}+\frac{1}{C_2}\right)\int idt=\frac{q_1}{C_1}+\frac{q_2}{C_2}. \tag{4.7.31}$$

Вважатимемо середню ємність варикапа значно меншою від контурних ємностей,

$$C_0<<C_{1,2}. \tag{4.7.32}$$

Тоді

$$\frac{1}{C(t)}+\frac{1}{C_1}+\frac{1}{C_2}\approx\frac{1}{C(t)}, \tag{4.7.33}$$

і (4.7.23) з урахуванням (4.7.25) можна переписати у формі

$$\int idt=C(t)\left(\frac{q_1}{C_1}+\frac{q_2}{C_2}\right). \tag{4.7.34}$$

В результаті система (4.7.29) з урахуванням (4.7.30) та (4.7.34) набуває вигляду

$$L_1\frac{d^2q_1}{dt^2}+R_1\frac{dq_1}{dt}+\frac{q_1}{C_1}=\frac{C(t)}{C_1}\left(\frac{q_1}{C_1}+\frac{q_2}{C_2}\right)+e(t); \tag{4.7.35}$$

$$L_2\frac{d^2q_2}{dt^2}+R_2\frac{dq_2}{dt}+\frac{q_2}{C_2}=\frac{C(t)}{C_2}\left(\frac{q_1}{C_1}+\frac{q_2}{C_2}\right).$$

За відсутності накачування (при m=0) врахування в правій частині рівнянь (4.7.35) малих доданків, пропорційних $C_0/C_{1,2}<<1$, приводить до невеликого зсуву власних частот та появи слабкого зв'язку між контурами. Нехтуючи цими ефектами, перепишемо (4.7.35) у формі:

$$\frac{d^2q_1}{dt^2}+2\delta_1\frac{dq_1}{dt}+\omega_1^2q_1=\omega_1^2\left(m_1q_1+m_2q_2\right)\cos\omega_Pt+\omega_1^2C_1e(t); \tag{4.7.36}$$

$$\frac{d^2q_2}{dt^2}+2\delta_2\frac{dq_2}{dt}+\omega_2^2q_2=\omega_2^2\left(m_1q_1+m_2q_2\right)\cos\omega_Pt,$$

де використано позначення

$$\frac{1}{\sqrt{L_{1,2}C_{1,2}}}=\omega_{1,2};\qquad \frac{R_{1,2}}{2L_{1,2}}=\delta_{1,2};\qquad \frac{mC_0}{C_{1,2}}=m_{1,2}. \tag{4.7.37}$$

Будемо вважати контури високодобротними,

$$\delta_{1,2}<<\omega_{1,2}, \tag{4.7.38}$$

а їхні смуги пропускання – такими, що не перекриваються:

$$\delta_{1,2}<<\left|\omega_1-\omega_2\right|. \tag{4.7.39}$$

Напругу, що подається на вхід підсилювача, вважатимемо гармонічною:

$$e(t)=E_m\exp(i\omega t), \tag{4.7.40}$$

а її частоту – близькою до парціальної (а в силу зроблених припущень – власної) частоти першого контуру, такою, що потрапляє в смугу пропускання:

$$\omega=\omega_1-\Delta,\qquad |\Delta|<<\omega_1. \tag{4.7.41}$$

<u>Регенеративний підсилювач із високочастотним накачуванням</u>

Нехай частота накачування задовольняє умові

$$\omega_P=\omega_1+\omega_2 \tag{4.7.42}$$

(випадок високочастотного накачування). Тоді холоста частота (різницева частота, що виникне в результаті взаємодії сигналу та накачування на нелінійному елементі – варикапі)

$$\omega_i=\omega_P-\omega \tag{4.7.43}$$

буде близькою до частоти другого контуру:

$$\omega_i = \omega_2 + \Delta, \qquad |\Delta| << \omega_2 . \tag{4.7.44}$$

В силу умов (4.7.38)–(4.7.44) будемо шукати розв'язки системи (4.7.36) для зарядів $q_{1,2}$ в усталеному режимі у формі

$$q_1(t) = A\exp(i\omega t), \qquad q_2(t) = B\exp(-i\omega_X t) \tag{4.7.45}$$

(пор. із одноконтурним параметричним підсилювачем). Підставляючи (4.7.45) до системи (4.7.36), нехтуючи доданками другого та більш високих порядків мализни й залишаючи в першому рівнянні лише доданки з частотою ω, а в другому – з частотою ω_i, отримаємо систему лінійних алгебраїчних рівнянь щодо амплітуд A та B:

$$\left(2\omega_1\Delta + i2\omega_1\delta_1\right)A = \frac{m_2\omega_1^2}{2}B + \omega_1^2 C_1 E_m ; \tag{4.7.46}$$

$$\left(-2\omega_2\Delta - i2\omega_2\delta_2\right)B = \frac{m_1\omega_2^2}{2}A .$$

Розв'язуючи (4.7.46), можна отримати:

$$A = \frac{8\omega_1\left(\Delta + i\delta_2\right)C_1E_m}{16\left(\Delta + i\delta_1\right)\left(\Delta + i\delta_2\right) + m_1m_2\omega_1\omega_2} ;$$

$$B = -\frac{2m_1\omega_1\omega_2C_1E_m}{16\left(\Delta + i\delta_1\right)\left(\Delta + i\delta_2\right) + m_1m_2\omega_1\omega_2} . \tag{4.7.47}$$

У випадку точного резонансу (Δ=0) отримаємо:

$$A_{\max} = -i\frac{Q_1C_1E_m}{1 - \frac{m_1m_2Q_1Q_2}{4}} ; \qquad B_{\max} = \frac{1}{2}\frac{m_2Q_1Q_2C_1E_m}{1 - \frac{m_1m_2Q_1Q_2}{4}} , \tag{4.7.48}$$

$$Q_{1,2} = \frac{\omega_{1,2}}{2\delta_{1,2}} . \tag{4.7.49}$$

Як випливає з (4.7.48), самозбудження схеми відбудеться при

$$m_1m_2 > \frac{4}{Q_1Q_2}, \tag{4.7.50}$$

або, з урахуванням (4.7.37), при

$$m > m_{cr} \equiv \frac{2}{C_0}\sqrt{\frac{C_1C_2}{Q_1Q_2}} . \tag{4.7.51}$$

Залежності амплітуд (4.7.48) від m подані на рис.4.7.7 (пор. з рис. 4.7.4 а).

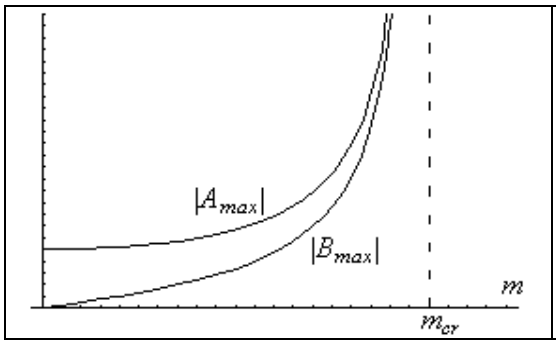


Рис. 4.7.7. Графіки залежності амплітуд сигналів у першому та другому контурах від глибини модуляції параметричної ємності в режимі точного резонансу в двоконтурному параметричному підсилювачі з високочастотним накачуванням.

Повернемося до аналізу загального випадку (4.7.47). Щоб знайти коефіцієнт підсилення та форму смуги пропускання, вважатимемо, що ми працюємо біля порогу генерації, тобто при

$$0 < \varepsilon \equiv 1 - \frac{m_1m_2Q_1Q_2}{4} << 1 \tag{4.7.52}$$

(пор. з одноконтурним підсилювачем). Будемо цікавитися малими розстроюваннями,

$$|\Delta| << \delta_{1,2} , \tag{4.7.53}$$

відкидаючи доданки порядку Δ^2. Тоді для амплітуди A (4.7.47) можна записати:

$$A=\frac{\dfrac{C_1E_m\omega_1}{2(\Delta+i\delta_2)}}{1-\dfrac{m_1m_2\omega_1\omega_2}{16(\Delta+i\delta_1)(\Delta+i\delta_2)}}\approx\frac{Q_1C_1E_m}{1-\dfrac{m_1m_2Q_1Q_2}{4(1-i\Delta/\delta_1)(1-i\Delta/\delta_2)}}\approx\frac{Q_1C_1E_m}{1-\dfrac{m_1m_2Q_1Q_2}{4}(1+i\Delta/\delta_1)(1+i\Delta/\delta_2)}\approx$$

$$\approx\frac{Q_1C_1E_m}{1-\dfrac{m_1m_2Q_1Q_2}{4}-i\Delta\left(\dfrac{1}{\delta_1}+\dfrac{1}{\delta_2}\right)\dfrac{m_1m_2Q_1Q_2}{4}}\approx\frac{\delta Q_1C_1E_m}{\Delta+i\varepsilon\delta},\qquad \delta=\frac{\delta_1\delta_2}{\delta_1+\delta_2}\qquad(4.7.54)$$

(в чисельнику відкинутий доданок порядку Δ і враховано, що $Q_{1,2}=\omega_{1,2}/2\delta_{1,2}$ та $m_1m_2Q_1Q_2/4\approx1$). Аналогічний розрахунок для B дає:

$$B\approx i\frac{m_1Q_2}{2}\frac{\delta Q_1C_1E_m}{\Delta+i\varepsilon\delta}.\qquad(4.7.55)$$

Як випливає з (4.7.54)-(4.7.55), смуга пропускання підсилювача складає $\Omega=2\varepsilon\delta$, а максимальний коефіцієнт підсилення – $K^s_{\max}=Q_1/\varepsilon$ (на частоті сигналу) та $K^i_{\max}=m_1Q_1Q_2/2\varepsilon$ (на холостій частоті).

У розглянутій схемі підсилення сигнальної та холостої частот відбувається в окремих контурах, смуги пропускання яких не перекриваються (див. умову (4.7.39)). Це означає, що биття між сигнальною та холостою частотами, які в одноконтурній схемі призводили до пульсацій коефіцієнту підсилення, тепер не виникатимуть.

Як і в одноконтурній схемі, в розглянутому випадку маємо справу з регенеративним механізмом підсилення, на що вказує можливість самозбудження схеми (за виконання умови (4.7.51)).

Контрольні питання до підрозділу 4.7

1. Чому при періодичній стрибкоподібній зміні ємності, увімкненої в коливний контур, у ньому може зростати амплітуда коливань? За яких умов це відбувається?
2. Чи можна в параметричному підсилювачі позбутися коливань на холостій частоті?
3. Чому в одноконтурному параметричному підсилювачі спостерігаються пульсації коефіцієнта підсилення?
4. Чому в двоконтурному параметричному підсилювачі пульсації коефіцієнта підсилення відсутні?

Задачі до підрозділу 4.7

4.7.1. Добротність контуру складає 50. Знайти критичну глибину модуляції параметричної ємності, необхідну для його самозбудження, якщо накачування
а) ступінчасте;
б) гармонічне.

4.7.2. Глибина модуляції параметричної ємності одноконтурного параметричного підсилювача складає 0.09, а відповідне критичне значення – 0.1. В яких межах змінюється коефіцієнт підсилення? Чому дорівнює період його зміни, якщо частота сигналу – 1.01 МГц, а частота контуру – 1 МГЦ?

4.7.3. Контури двоконтурного параметричного підсилювача мають ємності відповідно 81 пФ і 100 пФ та добротності 25 і 36. Параметрична ємність складає 1 нФ. При якій глибині модуляції цієї ємності схема самозбудиться?

4.7.4. Вхідний сигнал двоконтурного параметричного підсилювача характеризується частотою-носієм 5 МГц і шириною спектру 100 кГц. Якою має бути ширина смуги пропускання контурів, щоб запобігти лінійним спотворенням сигналу? Якою повинна бути частота холостого контуру, щоб коефіцієнт підсилення не зазнавав пульсацій? Якою при цьому має бути частота накачування?

Розділ 5. Автогенератори та інші нелінійні аналогові схеми

Розглянувши лінійні схеми аналогової радіоелектроніки, перейдемо тепер до аналізу нелінійних схем. До них належать автогенератори гармонічних (точніше, квазігармонічних) коливань, генератори імпульсних (релаксаційних) коливань, модулятори, детектори, перетворювачі частоти тощо. З цих схем найближчими до лінійних є автогенератори гармонічних коливань, але при аналізі навіть таких пристроїв врахування нелінійності обов'язкове, оскільки саме за рахунок неї відбувається обмеження амплітуди коливань, яка в початкові моменти часу зазвичай зростає за експоненціальним законом.

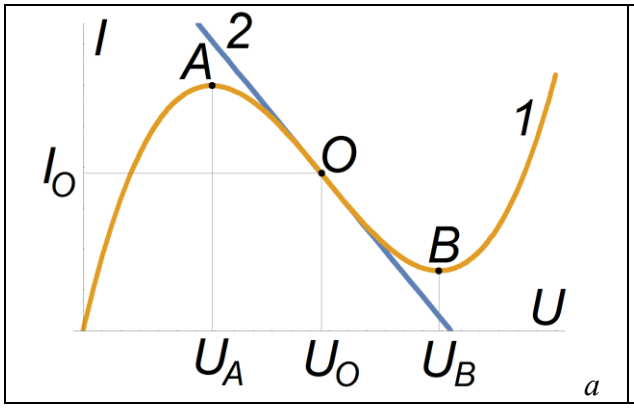
Спільною рисою усіх цих схем є те, що для адекватного опису їхньої роботи потрібно використовувати нелінійні диференціальні рівняння. Коефіцієнти таких рівнянь в загальному випадку залежать від самої функції – наприклад, струму або напруги.

В ролі нелінійних елементів виступають різноманітні діоди та (найчастіше) біполярні й польові транзистори. Таким чином, властивості останніх уже не можуть бути описані лінійними еквівалентними схемами (на зразок схеми Джіаколетто). Для опису нелінійних елементів потрібні інші підходи, з якими ми й ознайомимося в цій частині курсу.

5.1. Автогенератори на елементах із негативною диференціальною провідністю

Автогенератори, або автоколивні системи – це системи, в яких за відсутності зовнішнього періодичного впливу можуть виникати і встановлюватись періодичні коливання. Властивості цих коливань мало залежать від початкових умов і визначаються властивостями самої системи. Автогенератори – це різновид неавтономних[1] нелінійних дисипативних[2] систем. Саме за рахунок неавтономності забезпечується надходження енергії, необхідної для підтримання незгасаючих коливань.

Прикладами автогенераторів можуть служити механічні та електронні годинники, автогенератори періодичних коливань в радіотехніці (*LC*-автогенератори, *RC*-автогенератори, мультивібратори в автоколивному режимі, блокінг-генератори та ін.). До виникнення автоколивань зводиться й відоме в авіації явище флатера (вібрація крила, що виникає при перевищенні деякої критичної швидкості польоту й може призвести до руйнування літака).



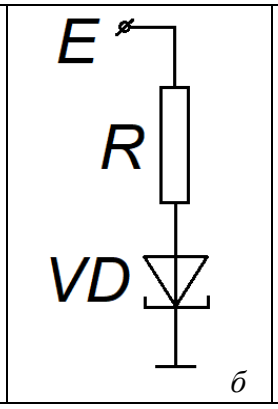


Рис. 5.1.1. Елемент із негативною диференціальною провідністю: а –графік вольт-амперної характеристики (1) та її апроксимація в околі точки О (2); б – схема увімкнення (нелінійним елементом служить тунельний діод)

5.1.1. Баланс потужності в колі з негативною диференціальною провідністю

Найпростіший автогенератор можна побудувати, скориставшись нелінійним елементом (двополюсником), на вольт-амперній характеристиці якого є ділянка з негативною диференціальною провідністю (рис. 5.1.1 а). Прикладом такого елемента може служити тунельний

[1] Неавтономною називають систему, яка може взаємодіяти з іншими системами (зокрема, відбирати від них енергію). Неавтономними є всі електричні кола, підключені до джерел живлення.

[2] Дисипативною називають систему, в якій відбувається перетворення різних видів енергії на тепло. Прикладом дисипативного елемента є активний опір.

діод (див. рис. 3.3.5 а). З рис. 5.1.1 а видно, що в інтервалі між точками А та В струм I зменшується зі зростанням напруги U, прикладеної до двополюсника (в реальності така ділянка виявляється нестійкою). В околі точки О (посередині між точками А та В), де вольт-амперну характеристику (лінія 1 на рис. 5.1.1) зручно апроксимувати відрізком прямої з від'ємним нахилом (лінія 2), можна записати:

$$\Delta I = -|G_d|\Delta U, \qquad G_d = \left.\frac{dI}{dU}\right|_{U=U_O} < 0. \tag{5.1.1}$$

Нехай до такого двополюсника одночасно прикладена стала напруга U_O, що відповідає робочій точці "О", та гармонічна напруга з амплітудою u_m (для спрощення вважатимемо, що $u_m << U_O$), тобто $u(t) = U_O + u_m \sin\omega t$. Тоді через двополюсник протікає струм

$$i(t) = I\left(U_O + u_m \sin\omega t\right) \approx I\left(U_O\right) + \left.\frac{dI}{dU}\right|_{U=U_O} u_m \sin\omega t = I_O - |G_d|\, u_m \sin\omega t \tag{5.1.2}$$

(враховано формулу (5.1.1)).

Відповідно миттєва потужність, що виділяється на нелінійному елементі, складає величину

$$\begin{aligned} P_1(t) &= U(t)I(t) = \left(U_O + u_m \sin\omega t\right)\left(I_O - |G_d| u_m \sin\omega t\right) = \\ &= U_O I_O + \left(I_O - U_O|G_d|\right)u_m \sin\omega t - |G_d| u_m^2 \sin^2\omega t, \end{aligned} \tag{5.1.3}$$

а середня за період коливань потужність –

$$\overline{P}_1 = U_O I_O - \frac{1}{2}|G_d| u_m^2. \tag{5.1.4}$$

Видно, що формально потужність сталого струму позитивна, а потужність сигналу негативна. Негативна потужність сигналу означає, що даний двополюсник не поглинає енергію коливань, а, навпаки, виділяє її. Він стає не споживачем, а джерелом енергії коливань з частотою ω.

Тепер врахуємо, що падіння змінної напруги на нелінійному елементі означає, що послідовно з ним до джерела живлення E увімкнене деяке навантаження R_l, яке для спрощення вважатимемо активним (див. рис.5.1.1 б; в цій ролі може, наприклад, виступати внутрішній опір джерела живлення та/або еквівалентний опір коливного контуру на резонансній частоті). Миттєва потужність, що виділяється на ньому, складає величину

$$\begin{aligned} P_2(t) &= \left[E - U(t)\right]I(t) = \left(E - U_O - u_m \sin\omega t\right)\left(I_O - |G_d| u_m \sin\omega t\right) = \\ &= \left(E - U_O\right)I_O - \left[I_O + \left(E - U_O\right)|G_d|\right]u_m \sin\omega t + |G_d| u_m^2 \sin^2\omega t, \end{aligned} \tag{5.1.5}$$

середня –

$$\overline{P}_2 = \left(E - U_O\right)I_O + \frac{1}{2}|G_d| u_m^2. \tag{5.1.6}$$

Видно, що сума $\overline{P}_1 + \overline{P}_2 = EI_O$ не залежить від сигнальної складової: потужність коливань, яка виділяється елементом з негативною диференціальною провідністю, поглинається опором навантаження.

В реальності подавати на нелінійний елемент змінну напругу не потрібно – вона виникає спонтанно. Питання про її амплітуду та частоту вимагають окремого обговорення.

5.1.2. Увімкнення негативної диференціальної провідності паралельно до коливного контуру

Для того, щоб отримати генератор квазігармонічних коливань на елементі з негативною диференціальною провідністю, треба якось задати частоту таких коливань. Для цього можна скористатися коливним контуром, увімкнувши його паралельно до елементі з негативною диференціальною провідністю (рис. 5.1.2 а). Замінимо нелінійний елемент його диференціальним опором R (рис. 5.1.2 б). Для врахування впливу опору R, увімкненого паралельно до коливного контуру, можна перерахувати паралельне сполучення опору R та ємності C у

послідовне (рис. 5.1.2 в). Для цього прирівняємо імпеданси паралельно та послідовно увімкнених елементів R і C:

$$\frac{R/i\omega C}{R+1/i\omega C}=\frac{R}{1+i\omega RC}=\frac{R-i\omega R^2C}{1+(\omega RC)^2}=\frac{R}{1+(\omega RC)^2}-\frac{i\omega R^2C}{1+(\omega RC)^2}=R'+1/i\omega C', \qquad (5.1.7)$$

звідки

$$R'=\frac{R}{1+(\omega RC)^2}, \qquad C'=\frac{1+(\omega RC)^2}{\omega^2R^2C}. \qquad (5.1.8)$$

При $\omega^2R^2C^2 >> 1$ ці формули спрощуються і набувають вигляду

$$R'=\frac{1}{\omega^2RC^2}, \qquad C'=C. \qquad (5.1.9)$$

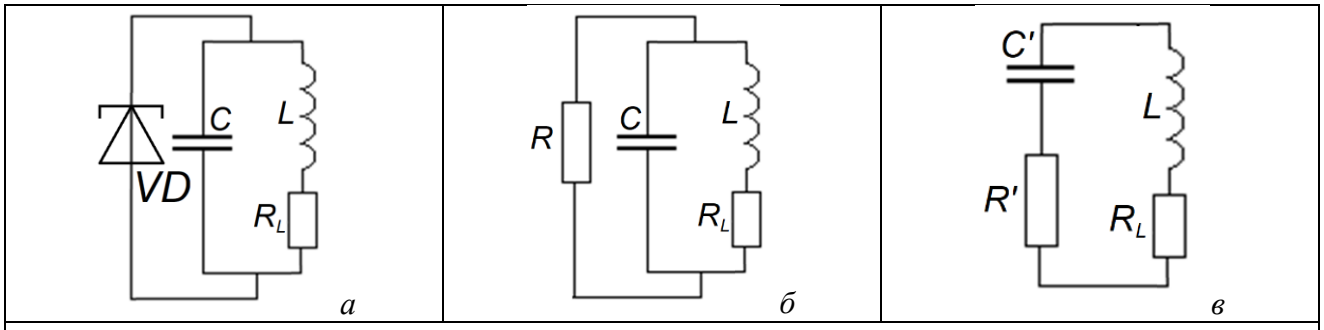


Рис. 5.1.2. Спрощена схема генератора на елементі з негативною диференціальною провідністю (а), заміна цього елементу еквівалентним лінійним опором (б) та перерахунок паралельного сполучення гілок у послідовне (в)

Оскільки опір R моделює елемент з негативною диференціальною провідністю G_d, можна написати:

$$R=-1/|G_d|<0, \qquad (5.1.10)$$

тобто опір R', що вноситься у контур, також виявляється негативним:

$$R'=-\frac{|G_d|}{\omega^2C^2}. \qquad (5.1.11)$$

В результаті при $|R'|>R_L$ повний опір контуру $R_\Sigma=R_L+R'$ стає негативним.

Вільні коливання струму у контурі описуються відомим диференціальним рівнянням

$$\frac{d^2i}{dt^2}+2\delta\frac{di}{dt}+\omega^2i=0, \qquad (5.1.12)$$

де

$$\omega=\sqrt{\omega_0^2-\delta^2}\approx\omega_0, \qquad \omega_0=1/\sqrt{LC}, \qquad \delta=R_\Sigma/2L$$

(прийнято, що $\delta \ll \omega_0$). Розв'язок рівняння (5.1.12) має вигляд

$$i(t)=i_0\sin\left(\omega_0t+\varphi_0\right)\exp\left(-\delta t\right).$$

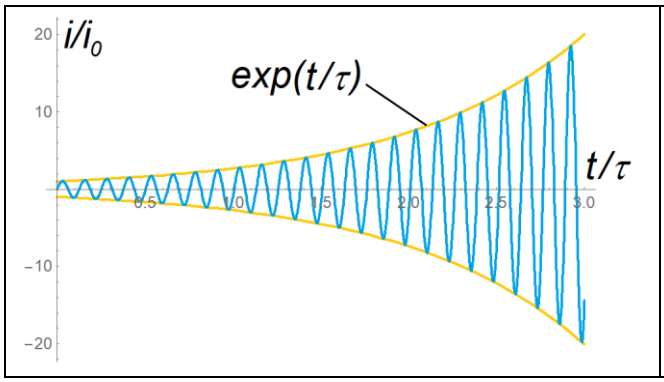


Рис. 5.1.3. Зростання амплітуди коливань у контурі з негативним еквівалентним опором

Якщо $|R'| > R_L$, загальний опір контуру та згасання δ, як уже вказувалося, стають негативними. В цьому випадку амплітуда коливань струму не згасає, як у звичайному контурі, а, навпаки, експоненціально зростає (рис. 5.1.3). Це зростання припиняється лише тоді, коли напруга на нелінійному елементі виходить за межі ділянки АВ вольт-амперної характеристики (див. рис. 5.1.1 а), де диференціальний опір нелінійного елемента від'ємний. Початкова амплітуда коливань i_0, від якої починається їхнє зростання, може бути зумовлена перехідними процесами при вмиканні напруги живлення або тепловими флуктуаціями струму, які завжди присутні в будь-якому електричному колі.

Таким чином, умова самозбудження коливань у контурі, до якого підключено негативну диференціальну провідність G_d, має вигляд

$$\frac{|G_d|}{\omega_0^2 C^2} > R_L, \tag{5.1.13}$$

або

$$|G_d| > \frac{1}{R_{eq}}, \qquad R_{eq} = \frac{1}{\omega_0^2 C^2 R_L} = \frac{L}{R_L C}, \tag{5.1.14}$$

де R_{eq} – еквівалентний опір контуру на резонансній частоті (див. вище п. 4.1.5).

5.1.3. Реальні елементи з негативною диференціальною провідністю

Серед напівпровідникових приладів зустрічаються такі, у яких на вольт-амперних характеристиках існують ділянки з негативною диференціальною провідністю (див. вище пп. 3.2-3.3). Прикладом може бути *p-n* перехід, до якого прикладено зворотну напругу. При тепловому пробої на його вольт-амперній характеристиці може виникнути ділянка з негативною диференціальною провідністю (рис. 5.1.4 а), яку можна використати для генерації високочастотних коливань.

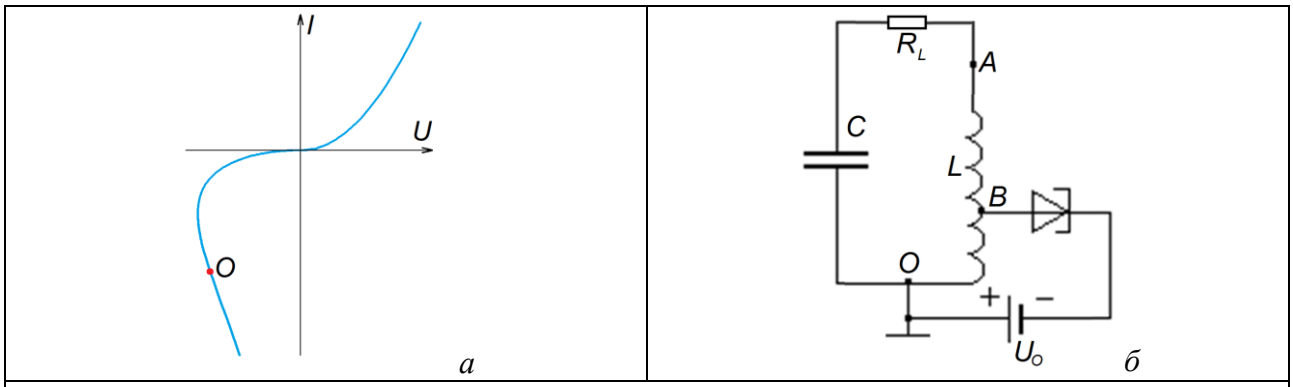


Рис. 5.1.4: Елементи з негативною диференціальною провідністю та їхнє використання: а – графік вольт-амперної характеристики напівпровідникового діода з ділянкою теплового пробою, б – схема автогенератора на тунельному діоді

Найпоширенішим приладом з негативною диференціальною провідністю є вже згаданий тунельний діод, який має вольт-амперну характеристику, подібну до зображеної на рис. 5.1.1 а. Оскільки негативна диференціальна провідність тунельного діода порівняно велика, він звичайно підключається до коливного контуру автотрансфоматорно (рис. 5.1.4 б), так щоб негативний опір, який він вносить до контуру, за модулем складав величину

$$R_{opt} = -(1.5 \div 3) R_L. \tag{5.1.15}$$

Якщо модуль внесеного негативного опору перевищує цю величину, то робоча точка, що відповідає генерованим коливанням, виходить далеко за межі ділянки АВ (рис. 5.1.1), і їхня форма стає суттєво несинусоїдальною, що небажано.

Знайдемо точку В оптимального увімкнення тунельного діода до контуру (рис. 5.1.4 б). Для цього перерахуємо паралельне сполучення довільної індуктивності L та активного опору R у послідовне, прирівнявши імпеданси обох ланцюжків:

$$R' + i\omega L' = \frac{R \cdot i\omega L}{R + i\omega L} = \frac{R}{1 + R/i\omega L} = \frac{R\left(1 + iR/\omega L\right)}{1 + \left(R/\omega L\right)^2},$$

звідки при $\left(R/\omega L\right)^2 \gg 1$

$$L' = L, \quad R' = \frac{\left(\omega L\right)^2}{R}. \tag{5.1.16}$$

В аналізованому випадку $L = L_{BO}$, $R = -G_d$, звідки

$$R' = -\left(\omega_0 L_{BO}\right)^2 G_d = -m^2 \frac{L}{C} G_d = R_{opt}, \tag{5.1.17}$$

де $m = L_{BO}/L$ – коефіцієнт автотрансформації. Таким чином,

$$m = \sqrt{\frac{\left|R_{opt}\right| C}{L G_d}} = \left(1.2 \div 1.7\right)\sqrt{\frac{R_L C}{L G_d}} = \frac{\left(1.2 \div 1.7\right)}{\sqrt{R_{eq} G_d}}. \tag{5.1.18}$$

Для оцінки вважаємо, що $1/G_d = 20$ Ом, звідки $G_d = 50$ мСим. Оберемо такі параметри контуру: $R_L = 5$ Ом, $L = 1$ мГн, $C = 0.1$ мкФ. Тоді частота коливань дорівнює $\omega_0 = 10^5$ с$^{-1}$, або $f_0 = 16$ кГц, хвильовий опір – $\rho = \sqrt{L/C} = 100$ Ом, еквівалентний опір – $R_{eq} = 2$ кОм. Оптимальний внесений опір – $R_{opt} = -\left(1.5 \div 3\right) R_L = 10$ Ом. Нарешті, коефіцієнт увімкнення індуктивності – $m = \left(1.2 \div 1.7\right)/\sqrt{R_{eq} G_d} = \left(1.2 \div 1.7\right)/\sqrt{2 \times 50} \approx 0.15$, індуктивність $L_{BO} = mL = 150$ мкГн.

Контрольні питання до підрозділу 5.1

1. Назвіть реальні електронні прилади, які мають негативну диференціальну провідність (НДП). Намалюйте їхні вольт-амперні характеристики та вкажіть на них області НДП.
2. Чи можуть бути негативним одночасно і омічний, і диференціальний опори електронного приладу?
3. Чи можлива ситуація, за якої коливання в контурі, зображеному на рис. 5.1.2 в, при виконанні амплітудної умови самозбудження не зростатимуть за браком початкової амплітуди коливань i_0?
4. Як залежить від часу амплітуда коливань в контурі якщо $1/|G_d|=R_{eq}$?
5. Чому на рис. 5.1.4 б тунельний діод підключений не до всього контуру, а тільки до його частини?

Задачі до підрозділу 5.1

5.1.1. Паралельно до коливного контуру з параметрами L = 1000 мкГ, C = 250 пФ; Q_0 = 40 увімкнений негативний диференціальний опір 15 кОм. Чи будуть самозбуджуватися коливання в контурі?

5.1.2. Для автогенератора з тунельним діодом, зображеного на рис. 5.1.4 б, обчисліть величину індуктивності L_{BO}, необхідну для успішної роботи автогенератора. Умови: частота генерації f_0 = 10 МГц; C = 500 пФ, добротність контуру Q_0 = 50. Негативний диференціальний опір діода $|R_d|$=50 Ом.

5.2. LC-автогенератори квазігармонічних коливань

Автогенератори на елементах із негативним диференціальним опором або провідністю на практиці використовуються нечасто. Значно частіше зустрічаються автогенератори, в основі яких лежить підсилювач, охоплений позитивним зворотним зв'язком (рис. 5.2.1), для якого виконані амплітудна та фазова умови самозбудження (див. вище п. 4.3.6). До цієї категорії належить, зокрема, автогенератор Ван-дер-Поля (інакше – LC-генератор чи генератор томсонівського типу), що являє собою резонансний підсилювач (підсилювач, де навантаженням служить паралельний коливний контур, див. вище п. 4.5.6) – рис. 5.2.2 а. Зворотний зв'язок у даному разі забезпечує трансформатор. Для спрощення вважатимемо, що активним елементом служить МДН-транзистор із вбудованим *n*-каналом (нагадаємо, що він не потребує сталого зсуву напруги на затвор).

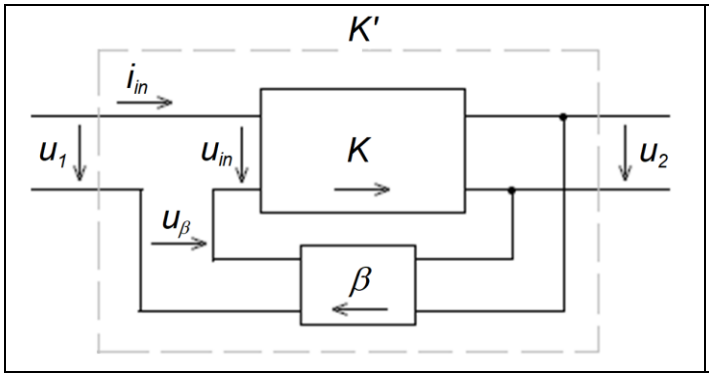
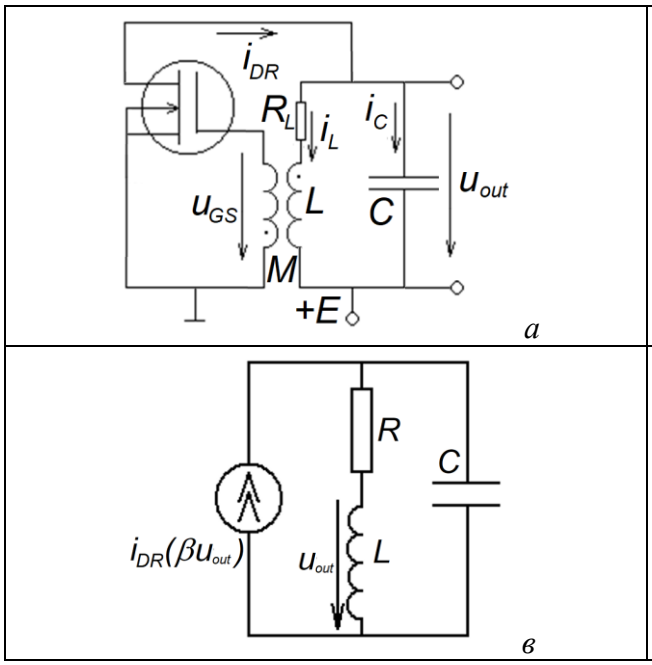


Рис. 5.2.1. Структурна схема підсилювача, охопленого послідовним зворотним зв'язком за напругою

5.2.1. Виведення нелінійного рівняння автогенератора Ван-дер-Поля

Схема автогенератора Ван-дер-Поля, досліджена в 1920-х – 30-х роках Б. Ван-дер-Полем та О.О. Андроновим, подана на рис. 5.2.2 а. Первісно активними елементами таких схем були електронні лампи (тріоди та ін.). МДН-транзистор із вбудованим каналом, використаний у схемі (рис. 5.2.2.а), за своїми властивостями подібний до тріода.



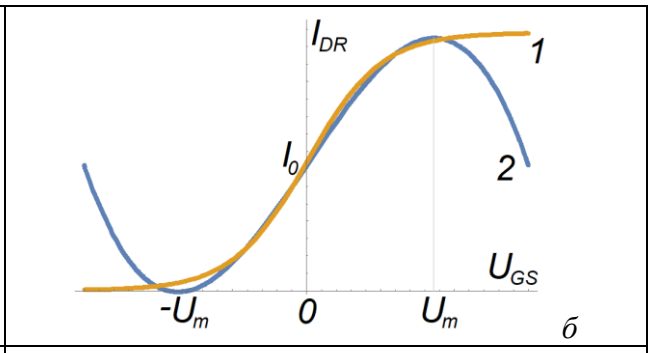


Рис. 5.2.2: Автогенератор Ван-дер-Поля: а – принципова схема; б – динамічна прохідна характеристика польового транзистора (1) та її апроксимація кубічною параболою (2); в – еквівалентна схема автогенератора за змінним струмом

Основна частина автогенератора – це резонансний підсилювач, де коливний контур служить навантаженням транзистора. З цього контуру знімається вихідна напруга. Підсилювач охоплений зворотним зв'язком через трансформатор із взаємною індуктивністю *M*. Нехтуючи падінням напруги на малому опорі R_L, можна записати коефіцієнт зворотного зв'язку

як відношення напруги на індуктивності L, $u_{out} = i\omega L i_L$, до напруги на вторинній обмотці трансформатора, тобто $u = i\omega M i_L$. Отже, $\beta = u/u_{out} = M/L$.

Еквівалентна схема автогенератора за сигналом подана на рис. 5.2.2 в.

Для адекватного опису роботи автогенератора слід врахувати нелінійність прохідної характеристики транзистора. Якщо робоча точка знаходиться на середині лінійної ділянки цієї характеристики (рис. 5.2.2 б), у розкладі в ряд Тейлора основну роль відіграватимуть непарні ступені аргументу. Обмежившись кубічною нелінійністю, запишемо апроксимацію прохідної характеристики у вигляді:

$$I_{DR}(U) = I_0 + SU_{GS} - \frac{SU_{GS}^3}{3U_m^2}. \qquad (5.2.1)$$

Апроксимація (5.2.1) якісно відповідає реальній характеристиці при $-U_m<U<U_m$ (рис. 5.2.2 б). Положення робочої точки (за відсутності затворної напруги) відповідає максимальній крутості прохідної характеристики (див. рис. 5.2.2 б).

Запишемо для змінних складових струму та напруги рівняння за другим законом Кірхгофа для коливного контуру автогенератора:

$$L\frac{di_L}{dt} + R_L i_L - \frac{1}{C}\int i_C dt = 0. \qquad (5.2.2)$$

Напруга на вторинній обмотці трансформатора, що прикладається між затвором та витоком польового транзистора, буде

$$u_{GS} = M\frac{di_L}{dt}. \qquad (5.2.3)$$

За I законом Кірхгофа

$$i_C + i_L = i_{DR}. \qquad (5.2.4)$$

Продиференціюємо (5.2.2), виключимо i_C за допомогою (5.2.4), а i_{DR} запишемо за допомогою співвідношень (5.2.1) (виключивши постійну складову) та (5.2.3). В результаті отримаємо:

$$L\frac{d^2 i_L}{dt^2} + R_L\frac{di_L}{dt} + \frac{i_L}{C} - \frac{1}{C}\left[MS\frac{di_L}{dt} - \frac{M^3 S}{3U_m^2}\left(\frac{di_L}{dt}\right)^3\right] = 0. \qquad (5.2.5)$$

Позначивши

$$\omega_0^2 = \frac{1}{LC}, \qquad \alpha = \frac{MS}{LC} - \frac{R_L}{L}, \qquad \gamma = \frac{M^3 S}{3U_m^2 LC}, \qquad (5.2.6)$$

запишемо остаточно, замінивши позначення i_L на x:

$$\ddot{x} - \left(\alpha - \gamma\dot{x}^2\right)\dot{x} + \omega_0^2 x = 0. \qquad (5.2.7)$$

Рівняння (5.2.7) відоме в літературі як рівняння Рейлі. Диференціюванням та замінами $\dot{x} \to x$, $3\gamma \to \gamma$ рівняння Рейлі зводиться до рівняння Ван-дер-Поля:

$$\ddot{x} - \left(\alpha - \gamma x^2\right)\dot{x} + \omega_0^2 x = 0. \qquad (5.2.8)$$

5.2.2. Умови самозбудження

Вважатимемо виконаною умову

$$\alpha > 0. \qquad (5.2.9)$$

З урахуванням явного вигляду α перепишемо її у вигляді

$$MS/R_L C > 1. \qquad (5.2.10)$$

Тепер врахуємо, що $M/L=\beta$ – це коефіцієнт зворотного зв'язку автогенератора (див. вище), $L/R_L C=R_{eq}$ – еквівалентний опір коливного контуру на резонансній частоті, $SR_{eq}=K$ – коефіцієнт підсилення резонансного підсилювача, який можна отримати з автогенератора Ван-дер-Поля, викинувши вторинну обмотку трансформатора й подавши вхідний сигнал на затвор транзистора. Отже, замість (5.2.10) отримаємо:

$$K\beta > 1. \qquad (5.2.11)$$

Це – амплітудна умова самозбудження підсилювача, охопленого зворотним зв'язком,

яка визначається критерієм Найквіста (див. формулу (4.3.15), п. 4.3.6).

Друга необхідна умова – фазова (див. формулу (4.3.14)) – вимагає, щоб сигнал, пройшовши через підсилювач та коло зворотного зв'язку, змінив свою фазу на величину, кратну 2π. Вона задовольняється завдяки тому, що вторинна обмотка трансформатора переполюсована (для цього виводи трансформатора треба поміняти місцями). Таким чином, транзистор і трансформатор кожен зсувають фазу сигналу на π (тобто величини K та β від'ємні).

Умова (5.2.11) – по суті, умова експоненціального зростання амплітуди малих коливань – може бути отримана в рамках лінійної теорії, коли нелінійним доданком у рівнянні (5.2.8) можна знехтувати. Ця теорія правильно описує умови самозбудження автогенератора та його поведінку у початкові моменти після увімкнення (експоненціальне зростання амплітуди коливань). Для коректного опису подальшої поведінки автогенератора, коли амплітуду коливань уже не можна вважати малою, слід користуватися повним нелінійним рівнянням (5.2.8).

Отже, рівняння Рейлі (5.2.7) (або рівняння Ван-дер-Поля (5.2.8)) за виконання умови (5.2.9) описує автогенератор, для якого виконано умови самозбудження.

5.2.3. Рівняння Рейлі: якісний аналіз розв'язку

Перш ніж розв'язувати рівняння Рейлі, спробуємо якісно схарактеризувати поведінку його розв'язків.

Нехай $|\alpha|<\omega_0$. Тоді при малих x, коли нелінійним доданком можна знехтувати, рівняння (5.2.7) відповідає осцилятору з малою від'ємною дисипацією. Його розв'язок – коливання, амплітуда яких експоненціально зростає з інкрементом $\alpha/2$ (при $|\alpha|>\omega_0$ відхилення від положення рівноваги при малих x має аперіодичний характер, див. нижче п. 5.4).

При не дуже малих x можна розглядати величину $\tilde{\alpha}=\alpha-\gamma\dot{x}^2$ як ефективний інкремент, що спадає із зростанням x. Коли він зменшиться до нуля, зростання коливань, очевидно, припиниться. Тоді амплітуду A коливань, що встановляться, можна оцінити із співвідношення $\gamma(\omega_0 A)^2\sim\alpha$ (враховано, що $|\dot{x}|\simeq\omega_0 A$), звідки

$$A\sim\frac{1}{\omega_0}\sqrt{\frac{\alpha}{\gamma}}\,. \tag{5.2.12}$$

Таким чином, з наведених міркувань випливає, що причиною, яка приводить до обмеження амплітуди автоколивань, є нелінійність прохідної характеристики транзистора.

5.2.4. Квазілінійна теорія автогенератора

Точніше визначити амплітуду коливань, що встановлюються в автогенераторі, дозволяє так звана *квазілінійна теорія*. В рамках цієї теорії зв'язок між затворною напругою та стоковим струмом польового транзистора вважається лінійним, $i=\bar{S}u$, але замість звичайної локальної крутості прохідної характеристики S вводиться поняття середньої крутості:

$$\bar{S}(u_m)=\frac{i_{m1}}{u_m}, \tag{5.2.13}$$

де u_m – амплітуда затворної напруги, яка вважається гармонічною, i_{m1} – амплітуда першої гармоніки струму стоку, який в силу нелінійності прохідної характеристики є ангармонічним, але залишається періодичним із періодом затворної напруги (рис. 5.2.7 а). Геометрично середня крутість відповідає тангенсу кута α нахилу січної на рис. 5.7.а. Залежність середньої крутості від амплітуди прикладеної напруги якісно зображена на рис. 5.2.7 б.

Для схеми, зображеної на рис.5.2.2 а, коефіцієнт підсилення, як уже вказувалося, дорівнює

$$K(u_m)=\bar{S}(u_m)L/R_L C\,, \tag{5.2.14}$$

коефіцієнт зворотного зв'язку –

$$\beta=\frac{M}{L}\,, \tag{5.2.15}$$

тому амплітудна умова самозбудження набуває вигляду

$$K(u_m)\beta = \frac{\bar{S}(u_m)M}{R_L C} \equiv \frac{\bar{S}(u_m)}{S_{cr}} > 1, \qquad (5.2.16)$$

де

$$S_{cr} \equiv \frac{R_L C}{M} \qquad (5.2.17)$$

– критичне значення крутості (пор. (5.2.16) із умовою (5.2.10)). Доки нерівність (5.2.16) виконуватиметься, сигнал зростає (при проходженні кільця підсилювач – коло зворотного зв'язку його амплітуда збільшується в $K\beta$ разів).

Ми обрали робочу точку А (рис. 5.2.3 а) там, де крутість прохідної характеристики максимальна. В цьому випадку зі збільшенням амплітуди середня крутість спадає (рис. 5.2.3 б), тому при деякому значенні амплітуди $u_{m\,st}$, для якого

$$\bar{S}(u_{m\,st}) = S_{cr}, \qquad (5.2.18)$$

зростання коливань припиниться.

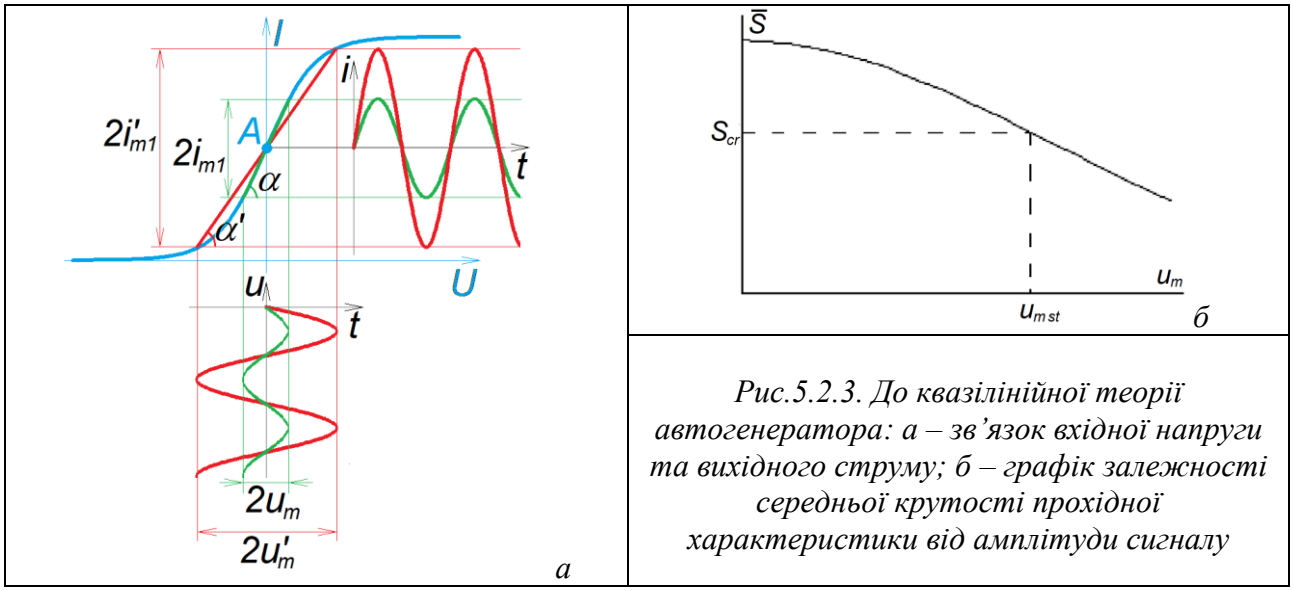


Рис.5.2.3. До квазілінійної теорії автогенератора: а – зв'язок вхідної напруги та вихідного струму; б – графік залежності середньої крутості прохідної характеристики від амплітуди сигналу

Квазілінійна теорія дозволяє, таким чином, знайти амплітуду коливань, що встановлюються в автогенераторі Ван-дер-Поля при $t \to +\infty$. Однак для коректного опису процесу встановлення коливань необхідно побудувати розв'язок рівняння Рейлі (5.2.7). Це рівняння вдається аналітично розв'язати для двох крайніх випадків – малих та великих α. Розглянемо їх послідовно.

5.2.5. Розв'язання рівняння Рейлі для режиму квазігармонічних коливань

Нехай параметри рівняння (5.2.7) задовольняють таким умовам:

– $\alpha << \omega_0$ (малі коливання є майже гармонічними);
– $\gamma \to 0$ (параметр нелінійності формально вважаємо малим).

Тоді вказане рівняння можна розв'язувати методом повільних амплітуд. Оскільки форма коливань залишається майже гармонічною, а нелінійність є слабкою, шукаємо розв'язок у вигляді коливань на частоті ω_0:

$$x(t) = A(t)\exp(i\omega_0 t) + \text{к.с.}^3, \qquad \omega_0^2|A| \gg \omega_0|\dot{A}| \gg |\ddot{A}| \qquad (5.2.19)$$

(подвійна нерівність означає, що протягом періоду коливань їхня амплітуда та її похідна за часом змінюються мало).

Підставляючи (5.2.19) до (5.2.7), залишаючи в рівнянні лише доданки першого порядку мализни за параметрами α та γ, відкидаючи третю гармоніку й прирівнюючи до нуля

[3] к.с. – скорочення від слів «комплексно спряжене».

амплітуду кожної з експонент у лівій частині рівняння, можна отримати таке вкорочене рівняння:

$$\dot{A}-\frac{\alpha}{2}A+\frac{3}{2}\gamma\omega_0^2A^2A^*=0\,. \qquad (5.2.20)$$

Перейдемо від комплексної амплітуди до дійсних амплітуди $a(t)$ й фази $\varphi(t)$ коливань:

$$A(t)=a(t)\exp\left[i\varphi(t)\right]. \qquad (5.2.21)$$

Підставивши (5.2.21) до (5.2.20), скоротивши на $\exp\left[i\varphi(t)\right]$ і прирівнявши до нуля окремо дійсну та уявну частини виразу, отримаємо:

$$\begin{cases} \dot{\varphi}=0; \\ \dot{a}-\dfrac{\alpha}{2}a+\dfrac{3}{2}\gamma\omega_0^2a^3=0 \end{cases} \qquad (5.2.22)$$

Інтегрування першого з рівнянь (5.2.22) дає: $\varphi=\varphi_0=const$, тобто коливання відбуваються з незмінною початковою фазою. В другому рівнянні вдається розділити змінні й подати його у формі:

$$\frac{da}{a\left(\dfrac{\alpha}{2}-\dfrac{3}{2}\gamma\omega_0^2a^2\right)}=dt\,. \qquad (5.2.23)$$

Скориставшись стандартним інтегралом

$$\int\frac{dx}{x\left(ax^2+c\right)}=\frac{1}{2c}\ln\frac{x^2}{ax^2+c}\,,$$

інтегруючи за часом від нуля до поточного моменту t, а за амплітудою – від початкового значення a_0 до поточного значення $a(t)$, можна, нарешті, отримати:

$$a(t)=\frac{a_\infty}{\sqrt{1+\left(a_\infty^2/a_0^2-1\right)\exp(-\alpha t)}}\,, \qquad (5.2.24)$$

$$a_\infty=\frac{1}{\omega_0}\sqrt{\frac{\alpha}{3\gamma}}\,. \qquad (5.2.25)$$

Проаналізуємо залежність (5.2.24). Перш за все відзначимо, що $a(t=0)=a_0$ (як ми й поклали) та $a(t\to\infty)=a_\infty$. Таким чином, a_∞ – це амплітуда стаціонарних коливань, що встановлюються в автогенераторі незалежно від їхньої початкової амплітуди. Звернемо увагу на те, що точне значення (5.2.25) мало відрізняється від оцінки (5.2.12).

Для малих початкових амплітуд та початкових моментів часу, коли виконано умову $(a_\infty/a_0)^2\exp(-\alpha t)>>1$, отримаємо $a(t)=a_0\exp(\alpha t/2)$: поки нелінійним доданком в (5.2.7) можна знехтувати, коливання експоненціально зростають з інкрементом $\alpha/2$.

Навпаки, якщо $a_\infty/a_0<1$, то, яке випливає з (5.2.24), функція $a(t)$ з часом спадає. Зокрема, при $a_\infty/a_0<<1$ матимемо $a(t)=a_\infty/(1-\exp(-\alpha t))^{1/2}$.

Хід залежності (5.2.24) для різних початкових амплітуд поданий на рис. 5.2.4.

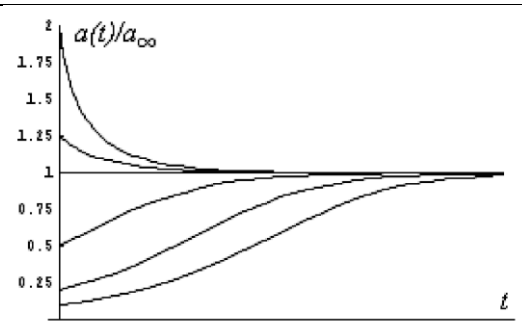


Рис. 5.2.4. Графік часової залежності амплітуди коливань автогенератора для різних початкових амплітуд

Як видно з формули (5.2.4) та рис. 5.2.4, амплітуда коливань a_∞ із часом встановлюється за як завгодно малої початкової амплітуди коливань a_0. Такий режим роботи автогенератора називається *м'яким*. Він реалізується, коли положення робочої точки відповідає максимальній крутості прохідної характеристики (див. вище п. 5.2.1).

5.2.6. Поняття про фазовий портрет

Для подальшого аналізу автогенератора Ван-дер-Поля необхідно ознайомитися з методом аналізу поведінки коливних систем за допомогою фазового портрету. Цим методом зручно користуватися для аналізу руху систем з одним ступенем вільності[4] (до яких належить і автогенератор Ван дер Поля).

Фазова площина – це площина, на якій по осі абсцис відкладається координата, а по осі ординат – швидкість (або імпульс) частинки, тоді в кожний момент часу стан системи відповідає деякій точці на фазовій площині, яку називають *зображувальною точкою*. Лінія, яку описує на фазовій площині зображувальна точка, називається *фазовою траєкторією*. Сукупність усіх можливих фазових траєкторій називається *фазовим портретом* системи.

Проаналізуємо, користуючись методом фазової площини, найпростіше рівняння лінійного консервативного осцилятора[5]:

$$\ddot{x} + \omega_0^2 x = 0. \tag{5.2.26}$$

Ввівши до розгляду швидкість

$$y\left(x(t)\right) = \dot{x} \tag{5.2.27}$$

і врахувавши, що

$$\ddot{x} = \frac{d}{dt}y = \frac{dy}{dx}\frac{dx}{dt} = y\frac{dy}{dx}, \tag{5.2.28}$$

можна переписати рівняння (5.2.26) у *фазових змінних* x, y, виключивши з нього час:

$$y\frac{dy}{dx} + \omega_0^2 x = 0. \tag{5.2.29}$$

Видно, що перехід до фазових змінних дозволив знизити порядок рівняння на одиницю, але замість лінійного рівняння отримали нелінійне. Однак у даному разі це рівняння легко проінтегрувати, розділивши змінні:

$$ydy = -\omega_0^2 xdx, \tag{5.2.30}$$

звідки

$$y^2 + \omega_0^2 x^2 = const. \tag{5.2.31}$$

Фазовий портрет лінійного консервативного осцилятора, як випливає з (5.2.31), являє собою набір вкладених еліпсів (рис. 5.2.5, *а*). Різні еліпси відповідають різним значенням константи, яка з точністю до сталого коефіцієнта визначає повну енергію коливань

Особлива (або стаціонарна чи нерухома) точка на фазовій площині – це точка $x = const$, яка задовольняє рівняння руху (очевидно, в ній $y = 0$). Для рівняння (5.2.26) існує єдина особлива точка $x = 0$, яка є ізольованою (через неї не проходить жодна фазова траєкторія). Така точка називається *центром* (*еліптичною точкою*).

Визначимо напрямок руху вздовж фазових траєкторій. Оскільки додатна швидкість відповідає зростанню координати, у верхній півплощині фазової площини, де швидкість

---

[4] Для системи, описуваної набором рівнянь вигляду $dx_i/dt = f_i\left(x_1, x_2, ..., x_n\right)$, $i = 1, 2, \ldots, n$, права частина яких не повинна явно залежати від часу, кількість ступенів вільності – це $n/2$. Відповідно однорідне диференціальне рівняння другого порядку, яке зводиться до двох рівнянь першого порядку, описує систему з одним ступенем вільності.

[5] Осцилятор – найпростіша коливна система з одним ступенем вільності (наприклад, пружинний маятник або коливний контур).

додатна, зображувальні точки завжди рухаються праворуч, відповідно в нижній – ліворуч.

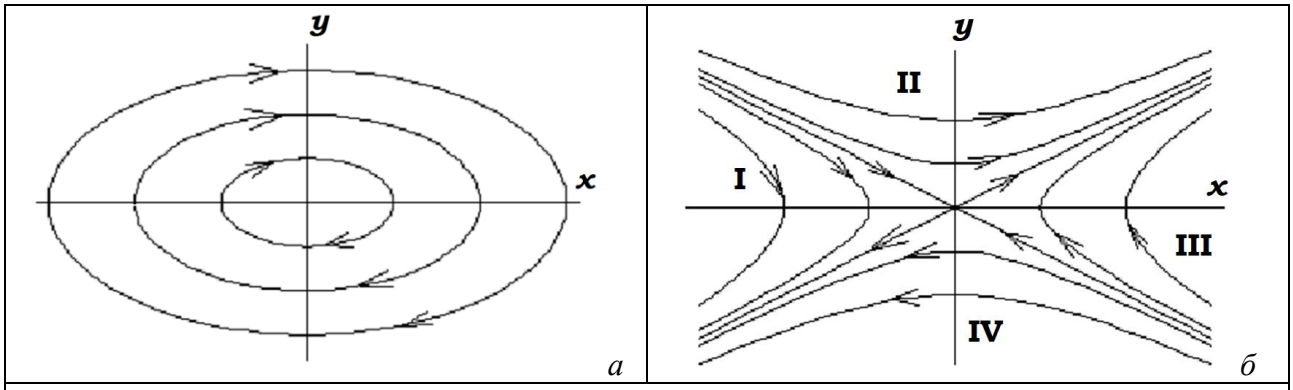


Рис. 5.2.5. Фазовий портрет лінійного консервативного осцилятора: а) $\omega_0^2 > 0$ *; б)* $\omega_0^2 < 0$

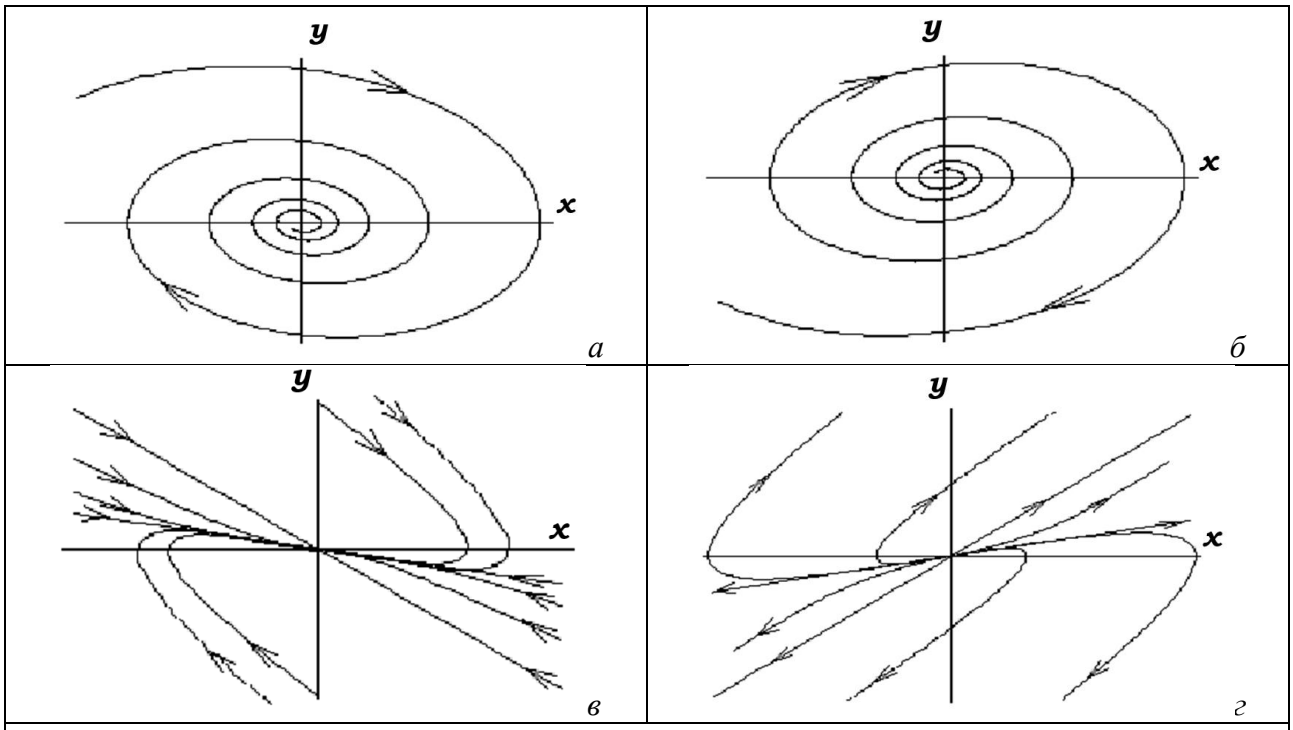


Рис. 5.2.6. Фазовий портрет лінійного дисипативного осцилятора для $\omega_0^2 > 0$ *:*
а – $\omega_0 > \delta > 0$ *; б –* $\omega_0 > -\delta > 0$ *; в –* $\delta > \omega_0 > 0$ *; г –* $-\delta > \omega_0 > 0$

Замінивши в рівнянні руху (5.2.26) ω_0^2 на $-\omega_0^2$, отримаємо перевернутий маятника (для малих відхилень; замість підвісу в моделі фігурує невагомий стрижень) або візок на гірці з параболічним профілем[6]. Виконавши ту саму заміну в (5.2.31), отримаємо рівняння, що описує сім'ю гіпербол:

$$y^2 - \omega_0^2 x^2 = \text{const} \qquad (5.2.32)$$

(рис. 5.2.5 *б*). Асимптоти $y = \pm\omega_0 x$ поділяють усі фазові траєкторії на чотири групи з відмінним характером руху, позначені на рис. 5.2.5, б римськими цифрами. Наприклад, траєкторії групи I відповідають випадку, коли візок, наїжджаючи на гірку з лівого боку, в деякій точці спиняється і скочується назад. Траєкторії групи II відповідають випадку, коли візок, також наїжджаючи на гірку з лівого боку, перекочується через неї та з'їжджає праворуч і т. ін.

Лінії на фазовій площині, що поділяють фазовий портрет на області з якісно відмінним

[6] За умови, коли нахил гірки з профілем $y(x)$ є малим, $|dy/dx| \ll 1$.

характером руху, називаються *сепаратрисами*. Отже, в даному випадку асимптоти $y = \pm\omega_0 x$ є сепаратрисами.

Особлива точка в цьому випадку є нестійкою, в ній починаються фазові траєкторії, які водночас є сепаратрисами. Така точка називається *сідловою*, або *гіперболічною*.

Лінійний дисипативний осцилятор описується рівнянням вигляду

$$\ddot{x} + 2\delta\dot{x} + \omega_0^2 x = 0 . \qquad (5.2.33)$$

Фазовий портрет такого осцилятора для випадку $\omega_0^2 > 0$ і різних значень параметра δ показаний на рис. 5.2.6. При $\delta > 0$ особливі точки на фазовому портреті є стійкими. При $\omega_0 > \delta$ говорять про *стійкий фокус*, коли амплітуда коливань з часом згасає, при $\delta > \omega_0$ – про *стійкий вузол*, коли відхилення від положення рівноваги згасає аперіодично (рис. 5.2.6 а, в). При $\delta < 0$ маємо, відповідно, *нестійкий фокус* (при $\omega_0 > -\delta$) і *нестійкий вузол* (при $-\delta > \omega_0$) – див. рис. 5.2.6 б, г.

5.2.7. Фазовий портрет автогенератора Ван-дер-Поля у м'якому режимі

Для побудови фазового портрета автогенератора, описуваного рівнянням Рейлі при $\alpha > 0$, скористаємося отриманим вище розв'язком (5.2.19), (5.2.24). При малих амплітудах коливань ($a < a_\infty$) амплітуда коливань із часом повільно зростає, тому відповідна фазова траєкторія матиме вигляд спіралі, що розкручується. Відповідно початок відліку на фазовій площині являтиме собою нестійкий фокус. При великих амплітудах ($a > a_\infty$), коли амплітуда коливань повільно спадає, фазова траєкторія являтиме собою спіраль, що скручується. Обидві ці спіралі асимптотично прямують до еліпса з півосями a_∞, $y_\infty = \omega_0 a_\infty$. Фазовий портрет автогенератора наведено на рис. 5.2.7.

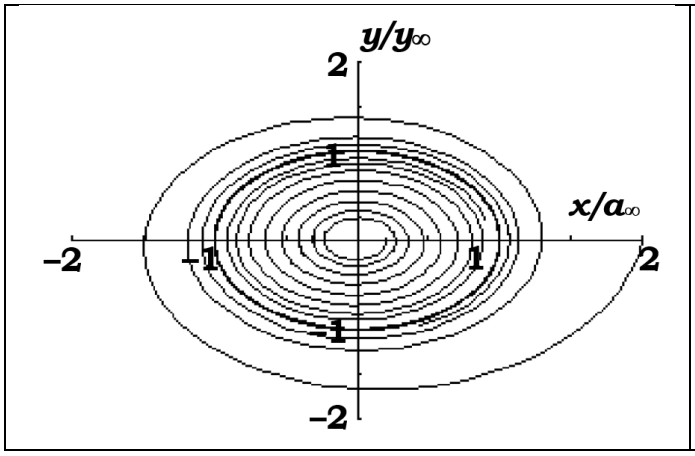


Рис. 5.2.7. Фазовий портрет автогенератора майже гармонічних коливань у м'якому режимі самозбудження

Еліпс

$$\left(\frac{x}{a_\infty}\right)^2 + \left(\frac{y}{y_\infty}\right)^2 = 1 , \qquad (5.2.34)$$

що розділяє області зростання та спадання коливань, виступає в ролі сепаратриси або, точніше, стійкого граничного циклу на фазовій площині.

Стійким граничним циклом на фазовій площині називають замкнену фазову траєкторію, до якої асимптотично прямують усі фазові траєкторії з деякого її околу при $t \to +\infty$. Аналогічну траєкторію, до якої прямують усі фазові траєкторії з деякого її околу при $t \to -\infty$, називають *нестійким граничним циклом*.

Існування стійкого граничного циклу означає, що в деякому діапазоні початкових умов незалежно від їх конкретного значення в системі з часом будуть установлюватися періодичні коливання. По суті, це й означає, що така система є автогенератором.

При $\alpha < 0$ фазовий портрет автогенератора якісно не відрізняється від фазового портрета лінійного слабкодисипативного осцилятора (рис. 5.2.6, *б*); відмінності виявляються лише при значних відхиленнях від положення рівноваги.

5.2.8. Жорсткий режим самозбудження автогенератора

Якщо робочу точку на прохідній характеристиці транзистора вибрати не на середині лінійної ділянки, а поблизу точки запирання, де стоковий струм прямує до нуля, то середня крутість спочатку зростатиме із зростанням затворної напруги, а тільки потім спадатиме (рис. 5.2.8-5.2.9). Можна чекати, що поблизу точки запирання, де крутість мала, умова самозбудження (5.2.10) не виконуватиметься. В той же час максимальне значення середньої крутості може бути достатнім для самозбудження. Тоді реалізується так званий *жорсткий режим самозбудження*: малі коливання згасають, і лише при перевищенні деякого критичного значення початкової амплітуди в системі встановлюються автоколивання (рис. 5.2.10).

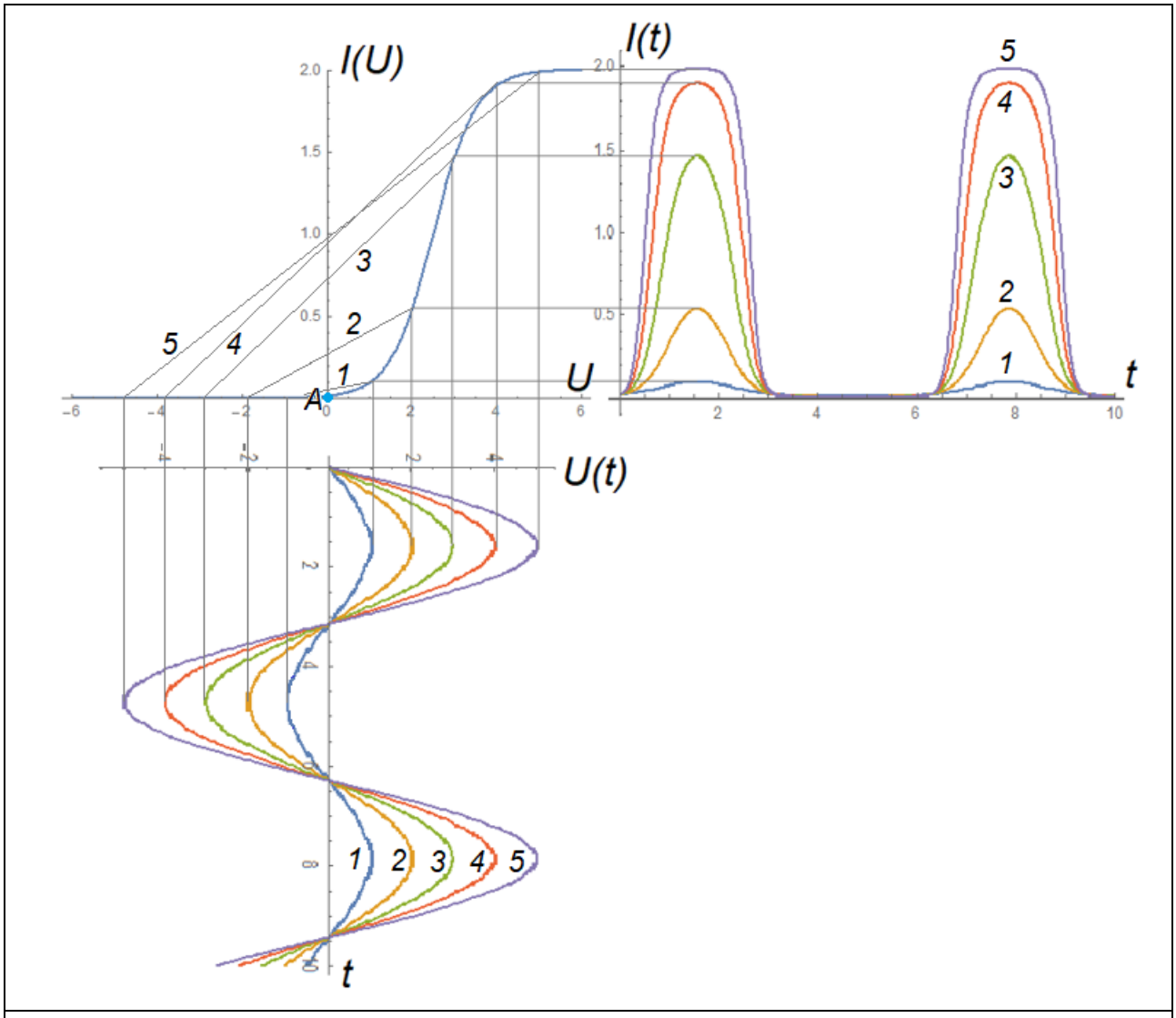


Рис. 5.2.8. Графік прохідної характеристики польового транзистора (ліворуч угорі) та часові епюри струму через нього (праворуч угорі) при поданні на затвор гармонічної напруги з різними амплітудами (внизу). Робоча точка А обрана на межі області запирання

Очевидно, в жорсткому режимі самозбудження, на відміну від м'якого режиму, поведінка автогенератора суттєво залежить від початкових умов. Така ситуація типова для нелінійних систем.

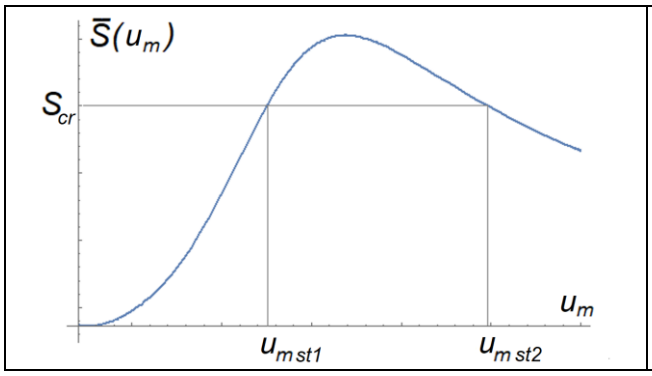


Рис. 5.2.9. Залежність середньої крутості від амплітуди вхідного сигналу при виборі робочої точки на межі області запирання

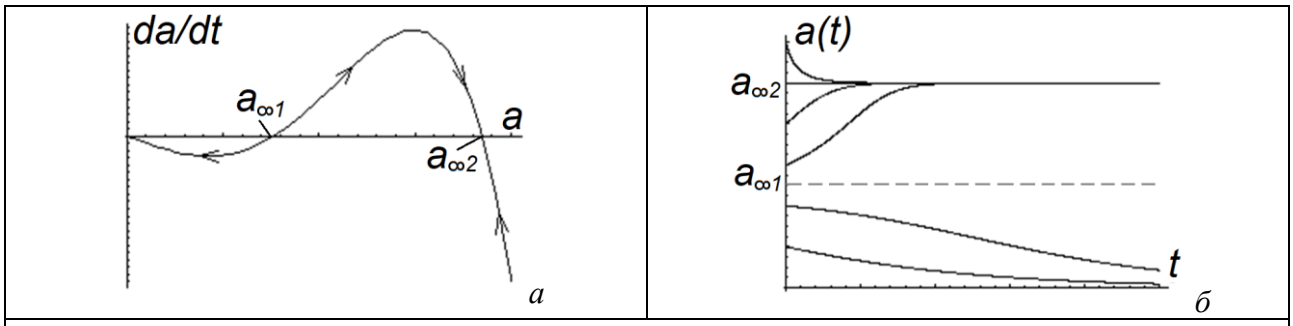


Рис. 5.2.10. Автогенератор із жорстким режимом самозбудження: а – швидкість зміни повільної амплітуди в залежності від її величини; б – часова залежність амплітуди для різних початкових умов

Фазовий портрет такого автогенератора має вигляд, поданий на рис. 5.2.11. Він має стійкий граничний цикл, до якого прямують усі фазові траєкторії з початковими амплітудами коливань, вищими за деяке порогове значення (пор. з рис. 5.2.10 б). Це порогове значення амплітуди на фазовому портреті відповідає нестійкому граничному циклу. Всередині нестійкого граничного циклу фазові траєкторії мають вигляд спіралей, що скручуються, а початок координат є стійким фокусом. У проміжку між нестійким і стійким граничними циклами маємо спіралі, що розкручуються, зовні стійкого граничного циклу – спіралі, що скручуються.

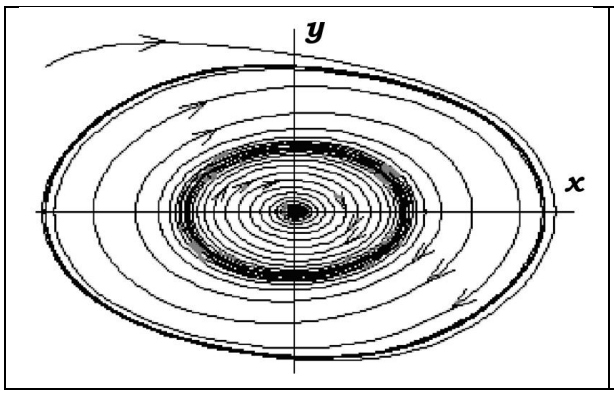


Рис. 5.2.11. Фазовий портрет автогенератора майже гармонічних коливань у жорсткому режимі самозбудження

5.2.9. Вимушена синхронізація автогенератора

Як правило, автогенератори, що забезпечують високу стабільність частоти, є малопотужними. Для отримання більш потужних сигналів у радіоелектроніці часто застосовуються так звані генератори із зовнішнім збудженням – потужні автогенератори, що керуються слабким вхідним сигналом. Для аналізу роботи таких пристроїв розглянемо поведінку автогенератора під дією зовнішньої сили.

Нехай у контур автогенератора увімкнене джерело змінної напруги (рис. 5.2.12). Обрана схема автогенератора відрізняється від зображеної на рис. 5.2.2 а тим, що коливний контур розташований не в колі стоку, а в колі затвору. Іншими словами, елементом, що задає частоту коливань, служить не навантаження, а коло зворотного зв’язку (на частотах, відмінних

від резонансної, паралельний контур шунтує вхід польового транзистора). За відсутності зовнішньої напруги робота цієї схеми описується рівнянням Ван-дер-Поля (5.2.8) для заряду на ємності коливного контуру.

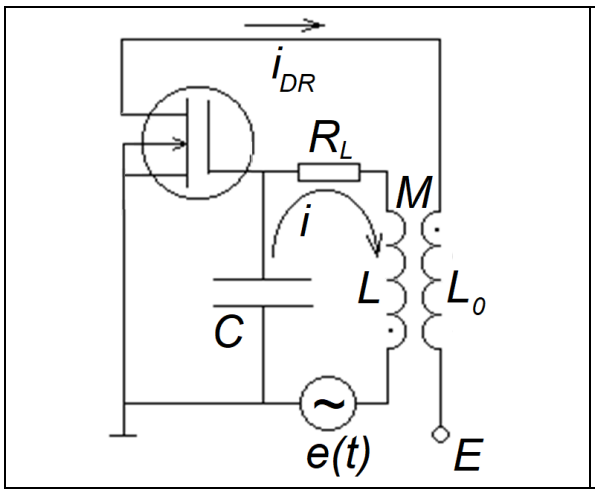


Рис. 5.2.12. Схема автогенератора із джерелом змінної напруги в контурі

Для того, щоб зрозуміти суть явищ, які відбуваються в автогенераторі під дією зовнішньої періодичної сили, скористаємося поняттями квазілінійної теорії автогенераторів (див. підрозділ 5.2.4). Як уже вказувалося, в рамках цієї теорії зростання коливань у контурі призводить до зменшення середньої крутості і, як наслідок, до порушення амплітудної умови самозбудження (у м'якому режимі).

За наявності джерела зовнішньої змінної напруги в контурі відбуваються коливання одночасно на двох частотах – на власній частоті контуру та на частоті зовнішньої напруги. Якщо коливання під дією зовнішньої напруги мають велику амплітуду, то відповідне значення середньої крутості може виявитися меншим за критичне значення (5.2.17). В цьому випадку автоколивання зриваються. Цей ефект носить назву *вимушеної синхронізації* автогенератора, або *нав'язування частоти*.

5.2.10. Реальні схеми LC-автогенераторів квазігармонічних коливань на біполярних транзисторах

Оскільки біполярні транзистори звичайно потужніші від польових, розглянемо тут реальні схеми автогенераторів саме на біполярних транзисторах.

Необхідний для самозбудження автогенератора зворотний зв'язок, що інвертував би фазу коливань, можна створити не лише за допомогою трансформатора, як це зображено на рис. 5.2.2 а, але й іншими шляхами.

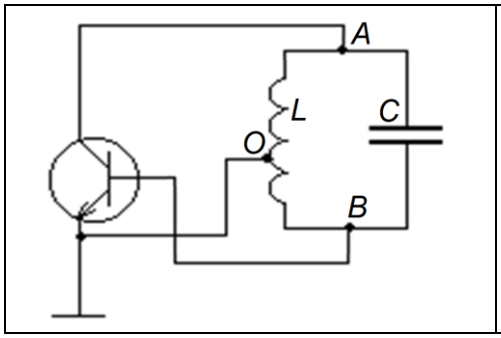


Рис. 5.2.13. Схема автогенератора з автотрансформаторним зворотним зв'язком (індуктивна триточка, або схема Хартлі)

Прикладом може бути зображена на рис. 5.2.13 схема з автотрансформаторним зворотним зв'язком (так звана індуктивна триточка, або схема Хартлі). В цій схемі середня точка О котушки індуктивності контуру заземлюється, а колектор і база приєднуються до протилежних кінців контуру А і В. Отже, протифазність напруг на базі і колекторі реалізується тут без жодних проблем, просто шляхом відповідного приєднання транзистора до контуру.

Коефіцієнт зворотного зв'язку складає $\beta = L_{BO}/L_{AO}$. На рис. 5.2.13 і далі джерело живлення схеми і кола, що створюють зміщення на базі, не показані, щоб не ускладнювати рисунок.

Іншим прикладом може бути ємнісна триточка (схема Колпітца), де зворотний зв'язок утворюється за допомогою конденсаторів $C_A C_B$, що утворюють подільник напруги (рис. 5.2.14). Обидва конденсатори складають ємність контуру $C = C_A C_B / (C_A + C_B)$.

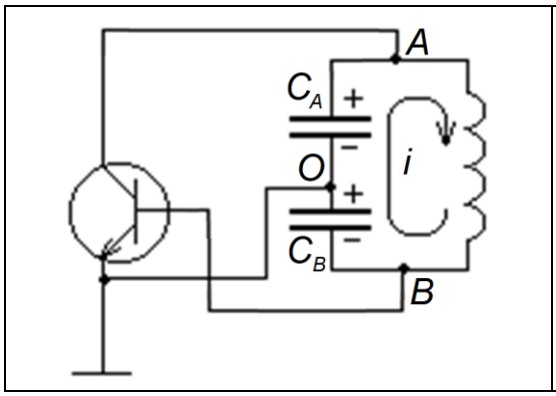


Рис. 5.2.14. Схема автогенератора з ємнісним зворотним зв'язком (ємнісна триточка, схема Колпітца)

Середня точка подільника О заземлюється. Оскільки через обидва конденсатора протікає той самий контурний струм i, напруги, що утворюються на них, протифазні щодо спільної точки О і дорівнюють

$$u_{AO}(t) = -\frac{i(t)}{i\omega C_A} \qquad u_{BO}(t) = \frac{i(t)}{i\omega C_B}.$$

Отже, коефіцієнт зворотного зв'язку u_{BO}/u_{AO} дорівнює $\beta = C_A/C_B$.

Певний інтерес являє схема, зображена на рис. 5.2.15 а. Це біполярний транзистор, до входу і виходу якого увімкнені два однакових контури, налаштовані на деяку частоту ω_0. Така ситуація складається, наприклад, у багатокаскадному резонансному підсилювачі. Подібна схема має схильність до самозбудження через прохідну ємність транзистора C_C.

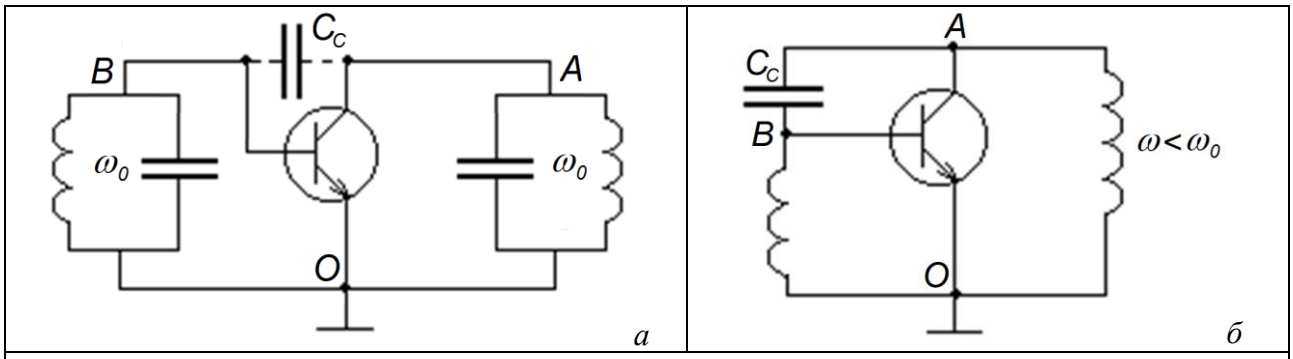


Рис. 5.2.15. До питання про самозбудження багатокаскадних резонансних підсилювачів

Справді, на частотах $\omega < \omega_0$ обидва контури мають індуктивну провідність і схема на рис. 5.2.15 а може бути зведена до еквівалентної схеми зображеної на рис. 5.2.15 б. А та, в свою чергу, зводиться до розглянутої вище індуктивної триточки (рис. 5.2.13). Таким чином, у схемі, зображеній на рис. 5.2.15 а, справді можливе самозбудження. Воно відбувається, якщо ємність C_C має достатню величину для забезпечення амплітудної умови самозбудження.

В усіх розглянутих вище схемах транзистор був увімкнений за схемою зі спільним емітером (СЕ). Але можна побудувати й автогенератор на транзисторі, увімкненому за схемою зі спільною базою (СБ). Така схема зображена на рис. 5.2.16.

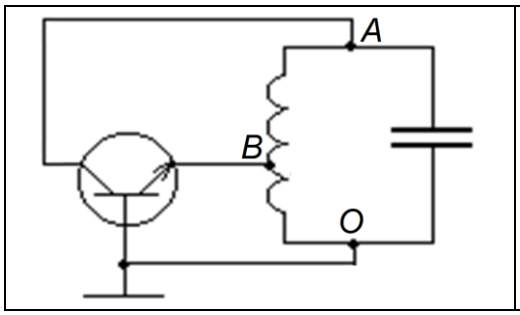


Рис. 5.2.16. Схема автогенератора на транзисторі, увімкненому зі спільною базою

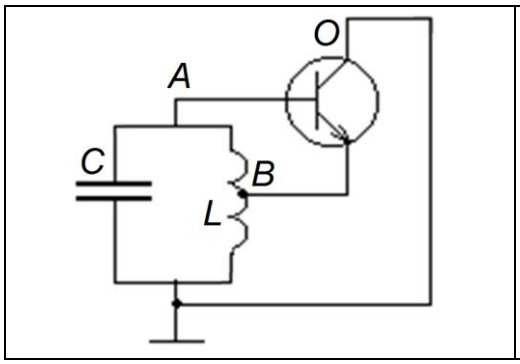


Рис. 5.2.17. Схема автогенератор на транзисторі, увімкненому зі спільним колектором

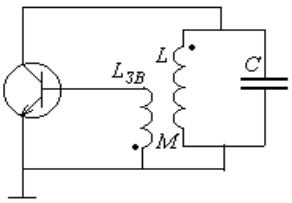
Особливість підсилювального каскаду зі спільною базою полягає у тому, що його вихідна напруга синфазна зі вхідною. Тому для створення позитивного зворотного зв'язку нема потреби інвертувати за фазою напругу, що подається з колектора на емітер. Зображена на рис. 5.2.16 схема є індуктивною триточкою. Аналогічно можна побудувати і ємнісну триточку та інші подібні схеми автогенераторів на транзисторі зі спільною базою. Можна також побудувати й проаналізувати індуктивну триточку зі спільним колектором (рис. 5.2.17).

Контрольні питання до підрозділу 5.2

1. Що обмежує амплітуду автоколивань у реальних автогенераторах?
2. Як на основі квазілінійної теорії можна знайти усталену амплітуду коливань автогенератора?
3. Які особливі точки може мати фазовий портрет системи, описуваної рівнянням лінійного дисипативного осцилятора?
4. Що таке стійкий граничний цикл?
5. Чим жорсткий режим самозбудження відрізняється від м'якого?
6. Який вигляд має фазовий портрет автогенератора в м'якому та в жорсткому режимі самозбудження?
7. Чи залежить усталена амплітуда коливань автогенератора від початкових умов?
8. Що таке індуктивна триточка та ємнісна триточка?
9. В чому полягає ефект вимушеної синхронізації автогенератора?

Приклади розв'язання задач до підрозділу 5.2

П 5.2.1. Контур з добротністю 50 складений з ємності 500 пФ та індуктивності 250 мкГн. Параметри транзистора: $h_{11E} = 0.5 кОм$, $h_{21E} = 40$, $h_{22E} = 100 мкСм$. При якій взаємній індуктивності M відбудеться самозбудження схеми?



Розв'язок

Умова самозбудження:

$\beta k_u > 1$, $\beta = \dfrac{M}{L}$, $k_u = \dfrac{R_{eq} h_{21E}}{h_{11E}\left(1 + h_{22E} R_{eq}\right)}$,

$R_{eq} = Q\rho = Q\sqrt{\dfrac{L}{C}} = 50\sqrt{\dfrac{250 \cdot 10^{-6}}{500 \cdot 10^{-12}}} = \dfrac{50}{\sqrt{2}} 10^3 \approx 35$ кОм. $k_u = \dfrac{35 \cdot 40}{0.5\left(1 + 35 \cdot 0.1\right)} = \dfrac{2800}{4.5} \approx 620$.

Отже, $M > \dfrac{L}{k} = \dfrac{250}{620} \approx 0.4$ мкГн.

П 5.2.2. Якщо прохідна характеристика транзистора $i(u)$ апроксимується поліномом $i(u) = a_0 + a_1 u + a_2 u^2 + a_3 u^3 + a_4 u^4 + \ldots$, то його середня крутість $\overline{s(u_m)}$ має вигляд $\overline{s(u_m)} = a_1 + \dfrac{3}{4} a_3 u_m^2 + \dfrac{5}{8} a_5 u_m^4 + \ldots$ Знайти усталену амплітуду коливань автогенератора, якщо $a_1 = 40$ мА/В, $a_3 = 200$ мА/В^3, $a_5 = -400$ мА/В^5, коефіцієнт зворотного зв'язку $\beta = 10^{-2}$, а еквівалентний опір навантаження – $R_{eq} = 1.67$ кОм.

Розв'язок

Усталену амплітуду знаходимо з умови $\beta k_u = 1$, або $\overline{s(u_m)} = 1/\beta R_{eq}$. Маємо з урахуванням знаку коефіцієнтів:

$$a_1 + \frac{3}{4} a_3 u_m^2 - \frac{5}{8} a_5 u_m^4 = \frac{1}{\beta R_{eq}}, \text{ або } \frac{5}{8} a_5 x^2 - \frac{3}{4} a_3 x + \left(\frac{1}{\beta R_{eq}} - a_1\right) = 0, \qquad x = u_m^2.$$

Дискримінант:

$$D = \left(\frac{3}{4} a_3\right)^2 - 4\frac{5}{8} a_5 \left(\frac{1}{\beta R_{eq}} - a_1\right) = (150)^2 - 2.5 \cdot 400\left(\frac{100}{5/3} - 40\right) =$$

$$= 22500 - 1000(60 - 40) = 22500 - 20000 = 2500.$$

Корені:

$$x_{1,2} = \frac{\frac{3}{4} a_3 \pm \sqrt{D}}{\frac{5}{4} a_5} = \frac{150 \pm 50}{500}.$$

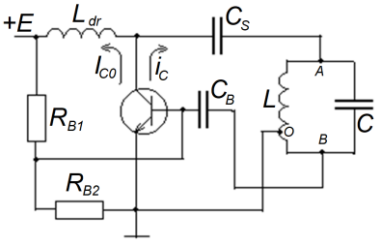
$$x_1 = \frac{200}{500} = 0.4, \; x_2 = \frac{100}{500} = 0.2.$$

Відповідно $u_{m1} = 0.63$ В, $u_{m2} = 0.45$ В.

Отже, маємо генератор із жорстким режимом самозбудження. Поріг самозбудження – $v_{m2} = 0.45B$, усталена амплітуда коливань – $v_{m1} = 0.63B$.

Задачі до підрозділу 5.2

5.2.1. Для схеми автогенератора, зображеного на рисунку, обчислити точку оптимального підключення контуру до транзистора, величину коефіцієнту зворотного зв'язку та індуктивність L_{bo}, необхідні для самозбудження, а також ємність розділового конденсатора та індуктивність дроселя. Прийняти L = 64 мкГ; C = 100 пФ; Q_0 =50; h_{11_E} = 1.2 кОМ; h_{21_E} = 60; h_{22_E} = 250 мкСим.



5.3. RC-автогенератори

На низьких частотах (нижче $10^4 - 10^5$ Гц) використовувати *LC*-контури в автогенераторах незручно, оскільки їхня індуктивність виходить надто великою, а добротність – низькою. Тому для частот звукового діапазону й нижче застосовують схеми без індуктивностей, тобто схеми, що містять лише резистори *R* і конденсатори *C* – так звані *RC*-автогенератори.

Принцип їхньої роботи ілюструє рис. 5.5.1. У підсилювальному каскаді має бути створений позитивний зворотний зв'язок з коефіцієнтом передачі $\beta(\omega)$ через чотириполюсник, що складається лише з резисторів і конденсаторів. На потрібній частоті ω_0 цей чотириполюсник повинен зсувати фазу сигналу на π. Таким чином забезпечується виконання фазової умови самозбудження саме на цій частоті. Крім того, на цій самій частоті має задовольнятися ще й амплітудна умова $K(\omega_0)\beta(\omega_0) > 1$, де $K(\omega_0)$ – коефіцієнт підсилення каскаду на частоті ω_0. Звичайно цю умову неважко виконати, за рахунок підбору величини K.

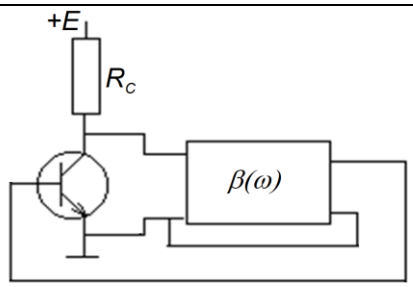


Рис. 5.3.1. Схема RC-генератора на біполярному транзисторі

5.3.1. Автогенератор з багатоланковим RC-фільтром

У п. 2.3.1 було отримано формулу (2.3.11) для комплексного коефіцієнту передачі диференціювального *RC*-ланцюжка (рис. 5.3.2 а) – він же *RC*-фільтр верхніх частот (див. п. 4.1.2):

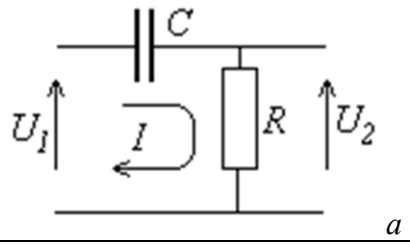
$$K(\omega) = \frac{R}{R + 1/i\omega C}.$$

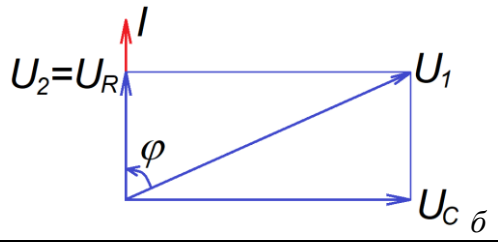
З неї випливає вираз для відповідного повороту фази:

$$\varphi(\omega) = arctg(i/\omega RC), \qquad (5.3.1)$$

а амплітуда вихідного сигналу на тій самій частоті дорівнює $U_2 = U_1 \cos\varphi$. (рис. 5.3.2 б-в).

Отже, такий ланцюжок повертати фазу вхідного сигналу на кут від 0 (при $\omega \to \infty$) до π/2 (при $\omega \to 0$), див. рис. 5.3.2 б-в.





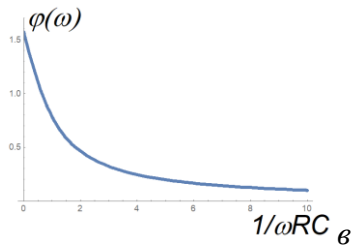


Рис. 5.3.2. RC-комірка: а – схема; б – векторна діаграма; в – фазово-частотна характеристика

Отже, для забезпечення зсуву фази сигналу на величину π на скінченній частоті потрібно щонайменше три *RC*-ланки. Відповідна схема автогенератора подана на рис. 5.3.3.

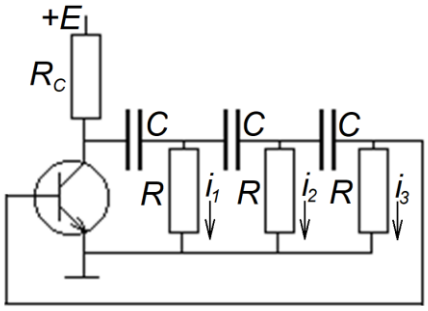


Рис. 5.3.3. Схема найпростішого RC - генератора з фазообертальним ланцюжком

5.3.2. Розрахунок триланкового RC-фільтру для автогенератора

Здавалося б, що обравши частоту, на якій диференціювальна комірка забезпечує зсув фази на $\pi/3$, та увімкнувши послідовно три таких комірки, можна отримати зсув фази на π. Справа, однак, ускладнюється тим, що при послідовному увімкненні кількох комірок кожна наступна комірка шунтує і навантажує вихід попередньої, так що зроблене вище припущення про ненавантажений режим уже не справджується. Потрібно розв'язувати задачу одразу для всіх комірок (наприклад, методом контурних струмів) і розраховувати відповідні залежності $\beta(\omega)$ й $\varphi(\omega)$.

Перемалюємо *RC*-ланцюжок автогенератора, показаного на рис. 5.3.3, врахувавши, що паралельно резистору останньої ланки (рис. 5.3.4) увімкнений вхідний опір транзистора h_{11E}. Коефіцієнт передачі такого ланцюжка в припущенні, що $R \ll h_{11E}$, складає величину

$$\beta = \frac{i_3 R}{u_{in}} = \frac{i_3 R}{i_1 R + (i_1 + i_2 + i_3)/i\omega C}. \qquad (5.3.2)$$

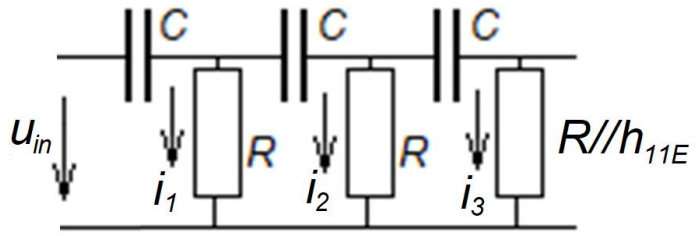


Рис. 5.3.4. Еквівалентна схема RC-ланцюжка автогенератора

Щоб пов'язати між собою струми i_1, i_2 та i_3, запишемо рівняння за другим законом Кірхгофа

$$i_3\left(R + \frac{1}{i\omega C}\right) = i_2 R; \qquad (5.3.3)$$

$$i_2 R + (i_2 + i_3)\frac{1}{i\omega C} = i_1 R. \qquad (5.3.4)$$

Тоді

$$i_2 = i_3\left(1 + \frac{1}{i\omega RC}\right), \qquad (5.3.5)$$

$$i_1 = i_2 + (i_2 + i_3)\frac{1}{i\omega CR} = \frac{i_3}{i\omega CR} + i_2\left(1 + \frac{1}{i\omega CR}\right) = i_3\left[\frac{1}{i\omega CR} + \left(1 + \frac{1}{i\omega CR}\right)^2\right]. \qquad (5.3.6)$$

Таким чином,

$$\beta = \frac{i_3 R}{i_1 R + (i_1 + i_2 + i_3)/i\omega C} = \frac{i_3}{i_1\left(1 + \frac{1}{i\omega RC}\right) + (i_2 + i_3)\frac{1}{i\omega RC}} =$$

$$=\frac{1}{\left[1+\left(1+\frac{1}{i\omega RC}\right)\right]\frac{1}{i\omega RC}+\left(1+\frac{1}{i\omega RC}\right)\left[\frac{1}{i\omega RC}+\left(1+\frac{1}{i\omega RC}\right)^2\right]}. \qquad (5.3.7)$$

Позначимо $x=1/\omega RC$. Маємо:

$$\beta=\frac{1}{-2ix-x^2-ix(1-ix)+(1-ix)^3}=$$
$$=\frac{1}{-i2x-x^2-ix-x^2+1-i3x-3x^2+ix^3}=\frac{1}{1-5x^2+ix\left(x^2-6\right)}. \qquad (5.3.8)$$

Видно, що β дійсне при $x=\sqrt{6}$, тобто на частоті

$$\omega_0=\frac{1}{\sqrt{6}RC}, \qquad (5.3.9)$$

при цьому

$$\beta(\omega_0)=-1/29. \qquad (5.3.10)$$

Тому для успішної роботи схеми, зображеної на рис. 5.3.3, модуль коефіцієнта підсилення каскаду має бути не меншим від 29.

Подібний результат можна отримати і з трьома послідовно увімкненими інтегрувальними *RC*-комірками, або фільтрами високих частот (рис. 4.1.3а). В цьому випадку

$$\omega_0=\frac{\sqrt{6}}{RC} \qquad \beta(\omega_0)=1/29 \qquad (5.3.12)$$

Можна, звичайно, побудувати ланцюжок з *RC* – комірок, кількість яких більше трьох. Вирази для ω_0 та $\beta(\omega_0)$ будуть тут трохи іншими. Але при цьому зменшується величина $\beta(\omega_0)$, що недоцільно, тому на практиці звичайно користуються ланцюжками з трьох комірок.

5.3.3. Особливості реальних схем RC-автогенераторів

До роботи схеми, зображеної на рис. 5.3.3, слід зробити декілька зауважень:

а) вхідний опір *RC*-ланцюжка шунтує вихід підсилювального каскаду. Це призводить до того, що коефіцієнт підсилення каскаду дещо знижується, а зсув фази підсилювальним каскадом виявляється відмінним від π. Тому при вибору елементів схеми бажано робити C якомога меншим, а R більшим (але так, щоб їхній добуток зберігав потрібне значення);

б) вхід транзистора шунтує вихідний резистор останньої комірки ланцюжка. Тому при $h_{11e}\le R$ бажано узгоджувати вихід фазообертального ланцюжка зі входом транзистора через емітерний повторювач (див. вище п. 4.5.7);

в) на відміну від *LC*-генераторів, де коливний контур ефективно виділяє на виході основну гармоніку навіть при великих нелінійних спотвореннях колекторного струму, в *RC*-генераторах такої частотно-вибіркової ланки (вузькосмугового фільтру) немає. Тому нелінійні спотворення колекторного струму повною мірою повторюються у вихідної напруги. З цієї причини не рекомендується виконувати амплітудну умову самозбудження $K(\omega_0)\beta(\omega_0)>1$ з великим запасом, бо значного виграшу у амплітуді при цьому не досягається, а нелінійні спотворення[7] суттєво зростають.

---

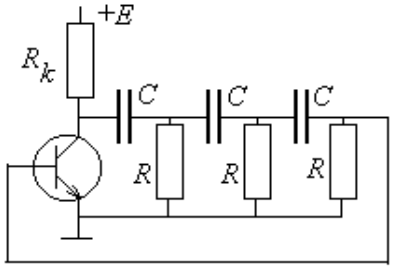
[7] Нелінійні спотворення гармонічного сигналу приводять до того, що його форма суттєво спотворюється. Відповідно в спектрі з'являються вищі гармоніки (див. нижче п. 5.7.1). Нелінійні спотворення з'являються, зокрема, тоді, коли робоча точка, рухаючись уздовж прохідної характеристики транзистора під дією вхідного сигналу з великою амплітудою, виходить за межі її лінійної ділянки.

Контрольні питання до підрозділу 4.3

1. Чому на низьких (звукових) частотах краще застосовувати *RC*-генератори, а не *LC*-генератори?
2. В чому полягає ідея створення *RC*-автогенератора?
3. Яким вимогам має задовольняти чотириполюсник зворотного зв'язку в *RC*-автогенераторі?
4. Одна диференціювальна *RC*-комірка обертає фазу гармонічного сигналу на кут φ . Чи можна при послідовному ввімкненні кількох таких комірок для підрахунку загального повороту фази просто помножити φ на кількість комірок?
5. Зобразіть схему *RC*-автогенератора подібну до наведеної на рис.5.3.3, але не з диференціювальними *RC*-комірками, а з інтегрувальними.
6. Як можна узгодити вихідний опір фазообертального *RC*-ланцюжка, зображеного на рис.5.3.3, зі вхідним опором транзистора?
7. Чому форма коливань, генерованих *RC* - автогенераторами, звичайно відрізняється від гармонічної? Що треба робити, аби наблизитись до гармонічної форми генерованих коливань?

Приклади розв'язання задач до підрозділу 5.3

П 5.3.1. R=3 кОм, C=0.01 мкФ, h_{11E}=1.2 кОм, h_{21E}=50, h_{22E}=0.1 мСм, R_k=10 кОм. Чи самозбудиться цей автогенератор? Якщо так, то на якій частоті?



Розв'язок

$K=\dfrac{h_{21E}R_k}{h_{11E}\left(1+h_{22E}R_k\right)}=\dfrac{50\cdot 10}{1.2\cdot\left(1+0.1\cdot 10\right)}\approx 210>>29$, тобто генератор самозбудиться – на частоті

$\dfrac{1}{2\pi\sqrt{6}RC}=\dfrac{10^3}{6.3\cdot 2.4\cdot 3\cdot 0.01}=2.2$ кГц.

Знайдемо вхідний опір RC-ланцюжка на частоті генерації:

$$Z_{eq}=\frac{U_{in}}{I_1+I_2+I_3}=\frac{I_1R+\left(I_1+I_2+I_3\right)/i\omega C}{I_1+I_2+I_3}=R\frac{I_1\left(1+1/i\omega RC\right)+\left(I_2+I_3\right)/i\omega RC}{I_1+I_2+I_3}=$$

$$=R\frac{\left[1+\left(1+\dfrac{1}{i\omega RC}\right)\right]\dfrac{1}{i\omega RC}+\left(1+\dfrac{1}{i\omega RC}\right)\left[\dfrac{1}{i\omega RC}+\left(1+\dfrac{1}{i\omega RC}\right)^2\right]}{\left(1+\dfrac{1}{i\omega RC}\right)+\left[\dfrac{1}{i\omega CR}+\left(1+\dfrac{1}{i\omega CR}\right)^2\right]}=$$

$$=R\frac{-ix\left(2-ix\right)+\left(1-ix\right)\left[-ix+\left(1-ix\right)^2\right]}{1-ix+\left[-ix+\left(1-ix\right)^2\right]}=R\frac{-i2x-x^2-ix-x^2+1-i3x-3x^2+ix^3}{1-i2x+1-i2x-x^2}=$$

$$=R\frac{ix\left(x^2-6\right)+1-5x^2}{-i4x+2-x^2}=\left\|x=\sqrt{6}\right\|=R\frac{-29}{-i24-4}=R\frac{29}{4\left(1+i6\right)}=R\frac{29}{4}\frac{1-i6}{37}$$

$$\left|Z_{eq}\right|=R\frac{29}{4\sqrt{37}}\approx 1.2\text{ кОм.}$$

Щоб шунтування не було, має виконуватися умова $R_k\le R$. Щоб вхід транзистора не впливав на RC-ланцюжок, має виконуватися умова $h_{11E}\gg R$.

Задачі до підрозділу 5.3.

5.3.1. Знайти поворот фази в колі зворотного зв'язку RC-автогенератора з урахуванням колекторного опору, вихідного та вхідного опору біполярного транзистора.

5.3.2. Знайти коефіцієнт зворотного зв'язку RC-автогенератора з урахуванням колекторного опору, вихідного та вхідного опору біполярного транзистора.

5.3.3. Знайти коефіцієнт підсилення RC-автогенератора з урахуванням колекторного опору, вихідного та вхідного опору біполярного транзистора в припущенні, що вхідна напруга подається на базу, а вихід ланцюжка, що обертає фазу, заземлений.

5.4. Автогенератор Ван-дер-Поля в режимі релаксаційних коливань

RC-генератори є прикладом генераторів без коливного контуру. Звичайно вони використовуються як джерела квазігармонічних коливань.

Коливний контур відсутній і в схемах генераторів негармонічних коливань, або імпульсних генераторів (релаксаційних генераторів). Можливо, найпростішим прикладом такого генератора є генератор Ван-дер-Поля (див. вище п. 5.2) в релаксаційному режимі. Хоч такий генератор і містить коливний контур, але його автоколивання не є власними коливаннями цього контуру.

5.4.1. Умови реалізації релаксаційного режиму

Отже, повернемося до генератора Ван-дер-Поля у м'якому режимі самозбудження, який описується рівнянням Рейлі (5.2.7):

$$\ddot{x}-\left(\alpha-\gamma\dot{x}^2\right)\dot{x}+\omega_0^2x=0\,.$$

Раніше ми детально проаналізували режим квазігармонічних коливань, що реалізується за умови $\alpha \ll \omega_0$. Розглянемо тепер інший крайній випадок його роботи, коли

$$\alpha \gg \omega_0\,. \tag{5.4.1}$$

Його називають режимом релаксаційних коливань. У цьому режимі при малих x розв'язок рівняння Рейлі являє собою зростаючі експоненти з дійсними показниками.

Замінимо змінні в рівнянні (5.2.7):

$$\xi=\frac{\omega_0^2\gamma^{1/2}}{\alpha^{3/2}}x\,,\qquad \tau=\frac{\omega_0^2}{\alpha}t\,. \tag{5.4.2}$$

В результаті воно набуде вигляду:

$$\mu\frac{d^2\xi}{d\tau^2}-\left[1-\left(\frac{d\xi}{d\tau}\right)^2\right]\frac{d\xi}{d\tau}+\xi=0,\qquad \mu\equiv\left(\frac{\omega_0}{\alpha}\right)^2<<1\,. \tag{5.4.3}$$

В рівнянні (5.4.3) малий параметр стоїть при старшій похідній. Такі рівняння можна розв'язувати шляхом <u>виділення ділянок швидкого та повільного руху</u>. На ділянках повільного руху старшою похідною взагалі можна знехтувати. На ділянках швидкого руху вона відіграє провідну роль.

5.4.2. Аналіз релаксаційних коливань

Перепишемо перенормоване рівняння Рейлі (5.4.3) у формі системи:

$$\varsigma=\frac{d\xi}{d\tau};\qquad \frac{d\varsigma}{d\tau}=\frac{\left(1-\varsigma^2\right)\varsigma-\xi}{\mu}. \tag{5.4.4}$$

У фазових змінних друге з рівнянь (5.4.4) набуде вигляду

$$\frac{d\varsigma}{d\xi}=\frac{\left(1-\varsigma^2\right)\varsigma-\xi}{\mu\varsigma}. \tag{5.4.5}$$

При μ=0 воно дає траєкторію повільного руху:

$$\xi=\left(1-\varsigma^2\right)\varsigma\,. \tag{5.4.6}$$

Рухаючись уздовж цієї траєкторії, зображувальна точка опиниться в одній з точок $\zeta=\pm 3^{-1/2}$, жодна з яких не є стаціонарною (швидкості там відмінні від нуля). Стаціонарна ж точка ξ=0, ζ=0 є нестійкою.

Щоб усунути отриману суперечність, врахуємо ненульову величину параметра μ. Тоді рівняння (5.4.5) за межами траєкторії повільного руху ($\xi\neq(1-\zeta^2)\zeta$) описує майже вертикальну пряму ($d\zeta/d\xi\to\infty$) – траєкторію швидкого руху. Напрямок руху та його швидкість задає друге з рівнянь (5.4.4). Траєкторії швидкого та повільного руху зображені на рис. 5.4.1 а.

Тепер зрозуміло, що зображувальна точка, яка потрапила в точку B з координатами $\xi_B=2\cdot 3^{-3/2}$, $\zeta_B=3^{-1/2}$ (рис. 5.4.1 б), опиниться на траєкторії швидкого руху й рухатиметься по ній до точки C ($\xi_C=2\cdot 3^{-3/2}$, $\zeta_C=-2\cdot 3^{-1/2}$). Потім по траєкторії повільного руху вона досягне точки D ($\xi_D=-2\cdot 3^{-3/2}$, $\zeta_D=-3^{-1/2}$), а з неї по траєкторії швидкого руху перестрибне в точку A ($\xi_A=-2\cdot 3^{-3/2}$,

$\zeta_A=2\cdot 3^{-1/2}$) і по траєкторії повільного руху знову досягне точки B, сформувавши тим самим граничний цикл ABCD.

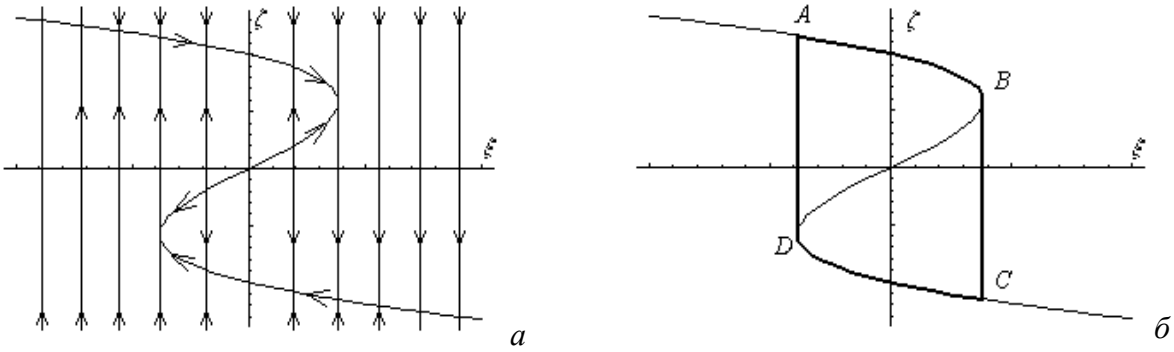


Рис. 5.4.1. Фазовий портрет автогенератора Ван-дер-Поля в релаксаційному режимі: а – траєкторії швидкого та повільного руху; б – граничний цикл.

Скориставшись фазовим портретом, можна побудувати епюри коливань. Вони подані на рис. 5.4.2. Видно, що координата ξ (нагадаємо, це струм через індуктивність) стрибків не зазнає. Координата ζ (падіння напруги на індуктивності) на ділянках швидкого руху (BC та DA) змінюється стрибком.

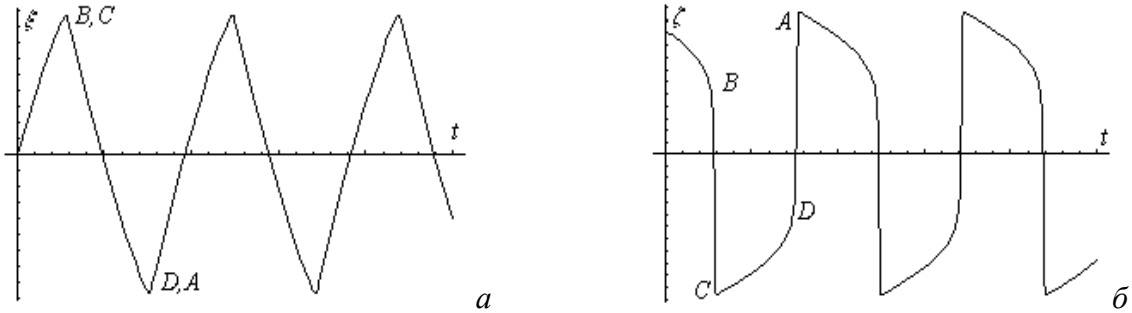


Рис. 5.4.2. Епюри коливань автогенератора Ван-дер-Поля в релаксаційному режимі, отримані шляхом числового розв'язку рівняння Рейлі (а – струм через індуктивність або напруга на конденсаторі, б – напруга на індуктивності, тобто на виході).

5.4.3. Інтерпретація релаксаційного режиму

З'ясуємо перш за все, коли може реалізуватися релаксаційний режим коливань. З урахуванням явного вигляду параметра α умова реалізації релаксаційного режиму (5.4.1) набуває вигляду

$$\frac{R_L}{\rho}\left(\frac{MS}{R_L C}-1\right)>>1. \qquad (5.4.7)$$

З іншого боку, використана нами кубічна апроксимація прохідної характеристики польового транзистора справедлива лише за умови $-U_m<U<U_m$ (див. вище п.5.2.1). В релаксаційному режимі максимальна затворна напруга U_{max}, яка відповідає значенню $\zeta_A=2\cdot 3^{-1/2}$, з урахуванням позначень (5.2.6), (5.4.2) задовольняє умові

$$U_{\max}=2U_{\mathrm{m}}\sqrt{1-\frac{R_L C}{MS}}<U_{\mathrm{m}}, \qquad (5.4.8)$$

звідки

$$0<\frac{MS}{R_L C}-1<\frac{1}{3}. \qquad (5.4.9)$$

Таким чином, з урахуванням (5.4.7) умова реалізації релаксаційного режиму набуває вигляду $R_L>>\rho$. Іншими словами, вільні коливання в контурі автогенератора за відсутності зворотного зв'язку повинні мати аперіодичний характер.

На ділянці повільного руху ми нехтуємо доданком із другою похідною в рівнянні Рейлі (5.4.3). Цьому доданку відповідає падіння напруги на індуктивності L. Отже, на ділянці повільного руху цим падінням напруги можна знехтувати – роль індуктивності полягає тільки в забезпеченні зворотного зв'язку через трансформатор, первинною обмоткою якого вона є.

Оскільки змінна ξ відповідає струму в індуктивній гілці контуру (див. п. 5.2.1 та формулу (5.4.2)), тобто на ділянках повільного руху – струму через активний опір, то напруга на конденсаторі (при нехтуванні падінням напруги на індуктивності) пропорційна до цього струму. Отже, епюра, подана на рис. 5.4.2 а, демонструє зміну напруги на активному опорі контуру, паралельно якому увімкнений контурний конденсатор. Епюра, подана на рис. 5.4.2 б, являє собою похідну за часом від струму через індуктивність, тобто, по суті, напругу на виході (якщо вихідний сигнал знімається з індуктивності) або напругу зворотного зв'язку, прикладену до затвору польового транзистора. На ділянках повільного руху ця напруга змінюється слабко, спадаючи за абсолютним значенням від U_{max} до значення $0.5U_{max}$, що відповідає точці B – див. рис. 5.4.3 (пор. з рис. 5.2.2 б). Таким чином, ділянка повільного руху AB відповідає спаданню $|U|$ від U_{max} до $0.5U_{max}$, а ділянка CD – зміні U від $-U_{max}$ до $-0.5U_{max}$.

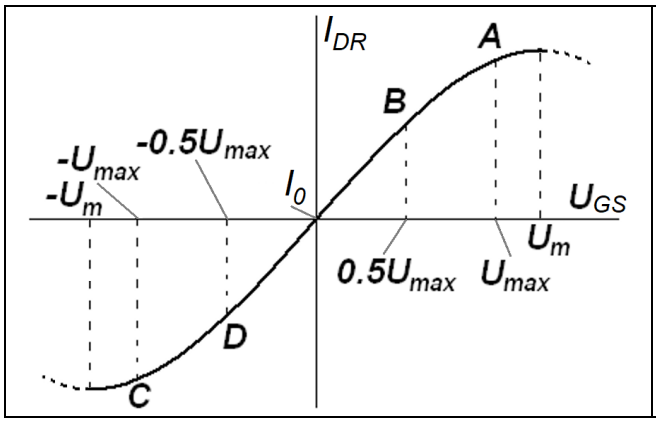


Рис. 5.4.3. Розташування характерних точок, що визначають релаксаційний режим автогенератора, на графіку прохідної характеристики польового транзистора (початок відліку відповідає струму I_0 у формулі (5.2.1))

В точці A транзистор майже повністю відкритий, у точці C – навпаки, закритий. Отже, якщо знехтувати стоковим струмом через закритий транзистор і падінням напруги на відкритому транзисторі, то можна сказати, що в точці A до контуру прикладена напруга живлення, а в точці C – нульова напруга. На ділянках швидкого руху ці стани змінюють один одного. Це й приводить до періодичної перезарядки конденсатора, якій відповідає епюра, наведена на рис. 5.4.2 а.

Режим транзистора, коли він перебуває або в стані насичення, або в стані запирання, називається ключовим. Саме ключовий режим транзистора й відповідає релаксаційному режиму автогенератора Ван-дер-Поля. Відповідна еквівалентна схема подана на рис. 5.4.4. Розімкнутий ключ відповідає стану запирання транзистора, замкнений – стану насичення.

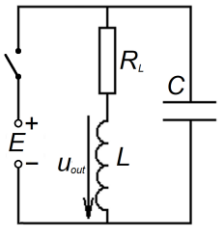


Рис. 5.4.4. Еквівалентна схема автогенератора Ван-дер-Поля в релаксаційному режимі

Як випливає з виразу (5.2.1), диференціальна крутість прохідної характеристики залежить від прикладеної до затвору напруги за законом

$$S_{diff} \equiv \frac{dI_{DR}}{dU_{GS}} = S\left(1 - \frac{U_{GS}^2}{U_m^2}\right). \qquad (5.4.10)$$

У точках B і D, в яких $U=\pm 0.5U_{max}$ вона набуває критичного значення S_{cr} (5.2.17).

Значить, на ділянках AB та CD (рис. 5.4.3) амплітудна умова самозбудження генератора не виконується, а на ділянці BD – виконується. Отже, перехід з ділянки AB через точку B на ділянку BD означає, що в системі починає розвиватися лавиноподібний процес її перекидання (швидкої зміни стану). Після його завершення транзистор, який був майже повністю відкритим, виявляється майже повністю закритим.

Оскільки процес перекидання відбувається дуже швидко, напруга на конденсаторі C за цей час змінитися не встигає. Так само не встигає змінитися струм через індуктивність, а, значить, і падіння напруги на контурному опорі (рис. 5.4.4). Отже, зміна напруги, обумовлена зміною стану транзистора і відповідною зміною струму i_{DR}, виявляється прикладеною до індуктивності. Це й приводить до стрибка напруги на індуктивності від $0.5U_{max}$ до $-U_{max}$. В результаті транзистор виявляється майже повністю закритим. Описаний процес перекидання схеми якраз і відповідає ділянці швидкого руху BC.

Інша ділянка швидкого руху відповідає переходу з точки D в точку A (рис. 5.4.3), коли транзистор, що був майже закритим, практично повністю відкривається.

Прикладом автогенератора релаксаційних коливань може служити мультивібратор з колекторно-базовими зв'язками в автоколивному режимі (див. нижче п. 5.5). Він являє собою двокаскадний підсилювач, у якого вихід замкнений на вхід. У такій схемі реалізується стопроцентний позитивний зворотний зв'язок (кожен каскад інвертує фазу сигналу), а коефіцієнт підсилення кожного каскаду значно більший від одиниці, в результаті чого умова (5.4.1) добре виконується.

5.4.4. Період релаксаційних коливань

Можна також підрахувати період коливань вздовж циклу ABCD (беручи до уваги тільки ділянки повільного руху). Отримаємо для «нового» (нормованого) часу:

$$\mathrm{T} = 3 - 2\ln 2 \approx 1.6. \qquad (5.4.11)$$

Для «старого» часу t отримаємо:

$$T \approx 1.6\,\alpha/\omega_0^2 = 1.6\left(MS/LC - R_L/L\right)LC = 1.6\left[\left(MS/R_L C\right) - 1\right]R_L C\,. \qquad (5.4.12)$$

З урахуванням умови (5.4.9) $T_{\max} \approx 0.53\,R_L C$.

При цьому

$$T \approx 1.6\,\alpha/\omega_0^2 = (1.6/2\pi)(\alpha/\omega_0)(2\pi/\omega_0) >> 2\pi/\omega_0$$

в силу умови (5.4.1). Таким чином, у релаксаційному режимі період автоколивань зростає в порівнянні з режимом майже гармонічних коливань.

Цьому результату можна дати просту якісну інтерпретацію. Формула (5.4.12) підтверджує, що період релаксаційних коливань за порядком величини можна оцінити як час заряджання конденсатора C через опір R_L, тобто $T \simeq R_L C$. Але в релаксаційному режимі $R_L \gg \rho$, тому

$$T >> C\rho = C\sqrt{L/C} = \sqrt{LC} = 1/\omega_0\,. \qquad (5.4.13)$$

Контрольні питання до розділу 5.4

1. За якої умови автогенератор Ван-дер-Поля працює в релаксаційному режимі?
2. Чим визначається форма коливань автогенератора Ван-дер-Поля в релаксаційному режимі?
3. Від чого залежить амплітуда коливань автогенератора Ван-дер-Поля в релаксаційному режимі?
4. Від чого залежить період коливань автогенератора Ван-дер-Поля в релаксаційному режимі?
5. Які реальні пристрої моделює релаксаційний режим автогенератора Ван-дер-Поля?

5.5. Мультивібратор

Мультивібратор – це генератор періодичних імпульсів, форма яких близька до прямокутної. Схема мультивібратора з тими чи іншими модифікаціями лежить в основі цілої сім'ї радіоелектронних пристроїв, що служать для створення імпульсів різноманітної форми, зокрема, трикутної.

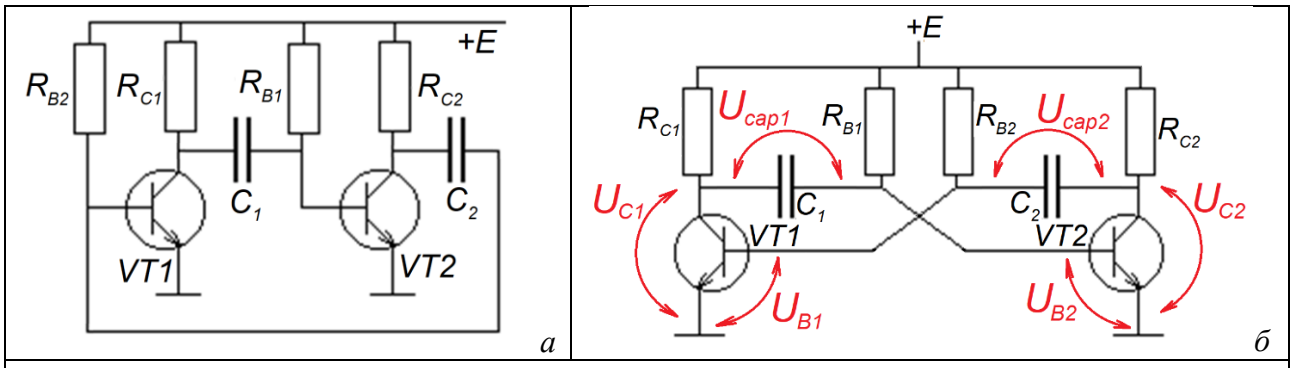


Рис. 5.5.1. Несиметричне (а) та симетричне (б) зображення схеми мультивібратора в автоколивному режимі

Мультивібратор – це двокаскадний підсилювач, охоплений стопроцентним позитивним зворотним зв'язком (рис. 5.5.1 а-б). Звичайно мультивібратор зображають у вигляді симетричної схеми (рис. 5.5.1 б). Однак навіть тоді, коли всі елементи схеми повністю симетричні, симетричний стан цієї схеми виявляється нестійким.

Кожен з каскадів інвертує сигнал, тому випадкова гармоніка напруги з малою амплітудою, спричинена флуктуацією, пройшовши через обидва каскади, повертається у вихідну точку з тією самою фазою, але з амплітудою, підсиленою у K^2 разів, де K – коефіцієнт підсилення кожного з каскадів.

Те саме буде за випадкової аперіодичної зміни напруги в симетричному стані схеми. Наприклад, якщо струм першого транзистора I_{C1} випадково зменшується, то це спричиняє збільшення напруги U_{C1} на його колекторі. Стрибок напруги проходить через ємність C_1 на базу другого транзистора і приводить до збільшення його колекторного струму I_{C2} і зменшення колекторної напруги U_{C2}. Підсилений стрибок напруги U_{C2} через ємність C_2 потрапляє на базу транзистора $VT1$ і приводить до подальшого зменшення струму I_{C1} [8]. Відхилення від положення рівноваги з часом експоненціально (лавиноподібно) зростає (пор. з ділянками швидкого руху в релаксаційному режимі генератора Ван-дер-Поля, див. вище п. 5.4), і врешті відбувається повне запирання транзистора $VT1$ та відкривання транзистора $VT2$ (зазвичай, до його повного насичення). В результаті обидва транзистори стають некерованими, й на цьому розвиток лавиноподібного процесу обривається. Встановлюється режим, який виявляється метастабільним (зберігає стійкість протягом певного часу).

5.5.1. Якісний розгляд роботи мультивібратора

Простежимо тепер за зміною напруги U_{B1} на базі транзистора $VT1$, який щойно закрився. Ця напруга дорівнює $U_{B1} = U_{C2} - U_{cap2}$ (див. рис. 5.5.1). Приймемо, що транзистор $VT2$ відкритий до насичення, тому напруга U_{C2} близька до нуля, і можна вважати, що $U_{B1} \approx -U_{cap2}$.

Нехай в початковий момент $t = 0$ напруга на C_2 дорівнює $-\left|U_{cap2}(0)\right|$. Тоді й

[8] Слід зауважити, що хоча режими транзисторів під час цих стрибків істотно змінюються, напруги на конденсаторах C_1 і C_2 (і заряди на них) не можуть зазнавати миттєвих змін і тому зберігають своє попереднє значення.

$U_{B1} = -|U_{cap2}(0)|$. З часом ємність C_2 перезаряджається від джерела живлення E через резистор R_{B2} та відкритий транзистор $VT2$. Напруга на C_2 поступово наближається до $+E$, змінюючись за законом

$$U_{cap2}(t) = E - \left[E - U_{cap2}(0)\right]\exp\left(-t/\tau_2\right), \qquad (5.5.1)$$

де $\tau_2 = R_{B2}C_2$ (внутрішнім опором джерела живлення та транзистора $VT2$ нехтуємо). Цей хід напруги $U_{B1}(t) \approx U_{cap2}(t)$ зображено на рис. 5.5.2.

Доки ця напруга залишається негативною, вона утримує транзистор $VT2$ в закритому стані. Але в момент $t = \tau_2$ напруга U_{B1} досягає нуля й змінює знак. Тоді напруга на базі $VT1$ транзистора стає позитивною, і він відкривається[9]. Це спричиняє лавиноподібний процес, в результаті якого $VT1$ відкривається, а $VT2$ закривається. Далі процес періодично повторюється.

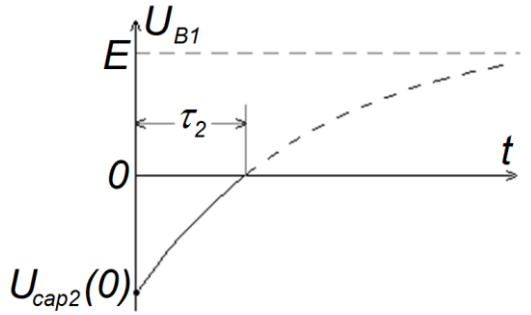


Рис. 5.5.2. Перезарядка конденсатора

Для того, щоб уявити собі хід процесів у всій схемі в цілому, розглянемо набір епюр, що характеризують одночасні зміни напруг на базах (рис.5.5.3 а, в) і на колекторах(рис.5.5.3 б, г) обох транзисторів.

Коли один з транзисторів відкривається, його колекторна напруга стрибком зменшується від E до напруги насичення U_{Csat}, яка зазвичай складає кілька десятих вольта (рис. 5.5.3 б, г). При запиранні транзистора колекторна напруга знову зростає до E. Таким чином, напруга на колекторі кожного з транзисторів має форму періодичної послідовності імпульсів майже прямокутної форми величиною $U_m \simeq E$. Напруга на базах визначається процесами перезарядки конденсаторів C_1 і C_2 через резистори R_{B1} та R_{B2} і має вигляд імпульсів, форма яких близька до трикутної (рис. 5.5.3 а, в).

Наведену картину треба дещо уточнити. Після запирання транзистора $VT2$ починається зарядка ємності C_2. Струм через неї тече в зворотному напрямку через R_{C2} і емітерний *p-n* перехід транзистора $VT1$, який щойно відкрився. Стала часу цього процесу складає $\tau_{C2} = R_{C2}C_2$. Він закінчується тим, що напруга на C_2 досягає значення E (напругою відкритого емітерного *p-n* переходу U_{B1} нехтуємо). Доки цей процес триває, через R_{C2} протікає струм, і напруга на колекторі $VT2$ залишається дещо меншою від E. Тому передній фронт імпульсу напруги U_{C2} виявляється заокругленим, ця напруга встановлюється на рівні E лише через час $(3 \div 4)\tau_{C2}$.

Водночас при протіканні струму, що заряджає ємність C_2 через емітерний *p-n* перехід транзистора $VT1$, на останньому створюється невеликий позитивний імпульс напруги ΔU_{B1},

[9] Насправді транзистор $VT1$ відкривається, коли напруга на його базі стає позитивною та перевищує потенціал відкривання транзистора U^*_{BE}. Але цей потенціал складає лише кілька десятих вольта, і ним можна у першому наближенні знехтувати (див. вище п. 3.5.2).

який викликає додаткове збільшення струму I_{C1} і короткочасне зменшення напруги U_{C1} [10].

Все викладене вище справедливо й для перезарядження ємності C_1, коли транзистор $VT1$ закритий.

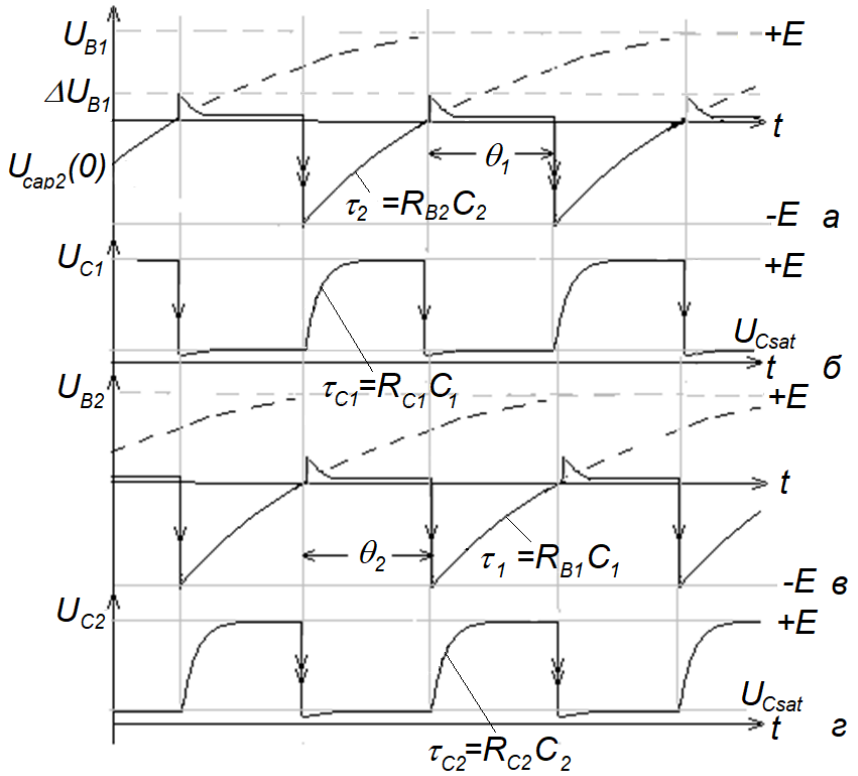


Рис. 5.5.3. Базові та колекторні напруги на транзисторах у схемі мультивібратора

5.5.2. Кількісні оцінки параметрів мультивібратора

– Амплітуда стрибка напруги на колекторі і, отже, амплітуда згенерованого імпульсу колекторної напруги буде

$U_m = E - U_{Csat} \approx E$.

Тривалість імпульсів визначається часом перебування відповідного транзистора в закритому стані (того, з колектора якого знімається вихідна напруга).

– Перший транзистор буде знаходитися у закритому стані протягом проміжку часу θ_2, доки напруга на його базі U_{B1}, яка практично дорівнює U_{cap2}, залишатиметься негативною. Для оцінки цього часу слід скористатися виразом (5.5.1). Можна вважати, що напруга $U_{cap2}(0)$ дорівнює $U_m \approx E$, бо за час перезарядження ємність C_2 встигає зарядитися до напруги джерела живлення E. Отже,

$$U_{B1}(t) = E\left[1 - 2\exp\left(-t/\tau_2\right)\right], \qquad (5.5.2)$$

і інтервал часу θ_2 може бути визначений з умови $U_{B1}\left(\theta_2\right) = 0$. Це дає

$$\theta_2 = \tau_2 \ln 2 \approx 0.7 R_{B2} C_2. \qquad (5.5.3)$$

Аналогічно $\theta_1 \approx 0.7 R_{B1} C_1$, так що повний період повторення імпульсів на виході мультивібратора дорівнює

$$\theta = \theta_1 + \theta_2 = 0.7(R_{B1}C_1 + R_{B2}C_2) \qquad (5.5.4)$$

– Зроблене вище припущення про те, що за час θ_1 перебування транзистора $VT2$ у закритому стані ємність C_2 встигає перезарядитися до напруги $U_m \approx E$, виконане лише тоді,

[10] Ці останні ефекти будуть майже непомітними, якщо транзистор $VT1$ буде відкритим до насичення.

коли $\tau_{C2} << \theta_1$ або ж

$$R_{C2}C_2 << R_{B1}C_1 \qquad (5.5.5)$$

(практично «запас» нерівності (5.5.5) може бути $3 \div 5$). Якщо нерівність (5.5.5) не виконано, напруга $U_{cap2}(0)$ буде меншою від E, і півперіод θ_2 буде коротшим, аніж це дає вираз (5.5.3).

Нерівність, аналогічну (5.5.3), має задовольнятися й для другого півперіоду:

$$R_{C1}C_1 << R_{B2}C_2 \qquad (5.5.5\ а)$$

З цих міркувань випливає, що неможливо зробити мультивібратор, у якого півперіоди θ_1 і θ_2 сильно відрізняються, бо відповідна ємність не встигатиме відновлювати свій заряд за коротший півперіод[11]. Найкраще працює мультивібратор, якщо півперіоди однакові (у так званому режимі меандру).

– Для того, щоб транзистор, який відкривається (наприклад $VT1$), знаходився в насиченому стані, необхідно виконати умову $I_{B1}\beta_1 > I_{C1}$ [12].Оскільки можна вважати, що $I_{B1} = E / R_{B2}$ і $I_{C1} = E / R_{C1}$, виконання цієї умови гарантується нерівністю

$$R_{C1}\beta_1 > R_{B2} \qquad (5.5.6)$$

і аналогічно

$$R_{C2}\beta_2 > R_{B1},$$

де запас рекомендується брати не більшим як $1.5 \div 2.0$. При більшому запасі відкритий транзистор буде знаходитися у глибокому насиченні і при закриванні буде надто повільно виходити з цього стану. А це призведе до затримки лавиноподібного процесу і затягне тривалість фронтів імпульсів.

– Нарешті, природно, мають бути виконані умови, за якими напруги живлення і струми через транзистори не повинні перевищувати допустимих значень

$$E < E_{C\max}, \quad I_{C1,2} = \frac{E}{R_{C1,2}} < I_{C\max} \qquad (5.5.7)$$

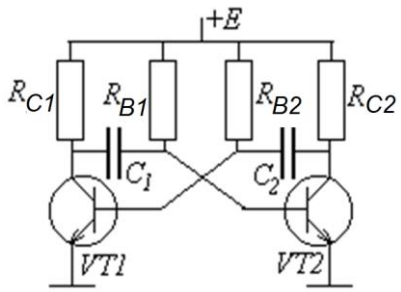
Використовуючи наведені нерівності, можна розрахувати мультивібратор і оцінити значення параметрів елементів, що його складають.

Контрольні питання до підрозділу 5.5

1. Опишіть принцип роботи мультивібратора з колекторно-базовими зв'язками.
2. Опишіть призначення елементів схеми мультивібратора в автоколивному режимі. Чим визначається тривалість та період повторення імпульсів на його виході?

Приклади розв'язання задач до підрозділу 5.5

П 5.5.1. У симетричному мультивібраторі R_B=40 кОм, R_C=5 кОм, C=1 нФ, h_{21E}=20. Чому дорівнює період коливань мультивібратора? Чи виконуються умови перезарядки ємності та насичення відкритого транзистора?



[11] Використовуючи умови (5.5.5) і (5.5.6) можна показати, що для успішної роботи несиметричного мультивібратора має бути виконана вимога $\theta_1 / \theta_2 << h_{21E}$.

[12] Величину β можна вважати близькою до відповідного значення h_{21E}.

Розв'язок

$\theta = 1.4 R_B C = 1.4 \cdot 40 \cdot 10^{-6} = 56$ мкс

Умова $R_B >> R_C$ виконана.

Умова насичення

$h_{21E} R_C = 20 \cdot 5 = 100$ кОм $> R_B = 40$ кОм

також виконана.

П 5.5.2. Розрахувати елементи схеми несиметричного мультивібратора, у якого θ_1/θ_2=1.2, $\theta_1+\theta_2$=600 мкс, U_m=12 В, $I_{C\,max}$=10 мА, h_{21E}=20.

Розв'язок

$\theta_1 = 1.2\theta_2$, $\theta_1 + \theta_2 = 2.2\theta_2 = 600$ мкс, $\theta_2 = \dfrac{600}{2.2} = 273$ мкс, $\theta_1 = 327$ мкс.

$R_{C1} = R_{C2} = \dfrac{E}{I_{C\,\max}} = \dfrac{12}{10} = 1.2$ кОм. $h_{21E} R_C = 20 \cdot 1.2 = 24$ кОм $> R_B >> R_C = 1.2$ кОм.

Оберемо $R_B = 12$ кОм. Тоді $C_1 = \dfrac{\theta_1}{0.7 R_B} = \dfrac{327 \cdot 10^{-6}}{0.7 \cdot 12 \cdot 10^3} \approx 39$ нФ, $C_2 = \dfrac{\theta_2}{0.7 R_B} = \dfrac{C_1}{1.2} \approx \dfrac{39}{1.2} = 32$ нФ.

Задачі до підрозділу 5.4

5.5.1. У симетричному мультивібраторі R_B=50 кОм, R_C=4 кОм, C=2 нФ, h_{21E}=30. Чому дорівнює тривалість імпульсу вихідної напруги мультивібратора? Чи виконуються умови перезарядки ємності та насичення відкритого транзистора?

5.5.2. Розрахувати елементи схеми несиметричного мультивібратора, у якого θ_1/θ_2=1.5, $\theta_1+\theta_2$=800 мкс, U_m=10 В, $I_{C\,max}$=12 мА, h_{21E}=40.

5.6. Тригери

Тригер (від англ. trigger – клямка, заскочка) – електронна логічна схема, яка має два стійкі стани і два входи. За відсутності сигналів на входи тригер зберігає стан, у якому він перебував раніше. Під дією вхідних сигналів тригер може змінити свій стан. При переході з одного стану в інший напруги й струми на виході тригера змінюються стрибкоподібно.

Завдяки можливості зберігати стан, заданий вхідними сигналами, тригер може служити елементом пам'яті комп'ютера.

5.6.1. Схема тригера та його стаціонарні стани

Схема тригера являє собою двокаскадний підсилювач мінливої напруги з двополярним живленням, у якого вихід приєднаний безпосередньо до входу (рис. 5.6.1 а). Звичайно каскади ППН, що утворюють тригер, ідентичні, тому, щоб підкреслити симетрію схеми, її подають у характерному симетричному вигляді (рис. 5.6.1 б). Входи тригера, сигнали на яких можуть змінити його стан, тут не показані.

Відмінності схеми тригера від розглянутої раніше схеми мультивібратора – це, по перше, відсутність розділових конденсаторів між каскадами, по друге – зумовлене цим двополярне живлення (див. вище п. 5.5).

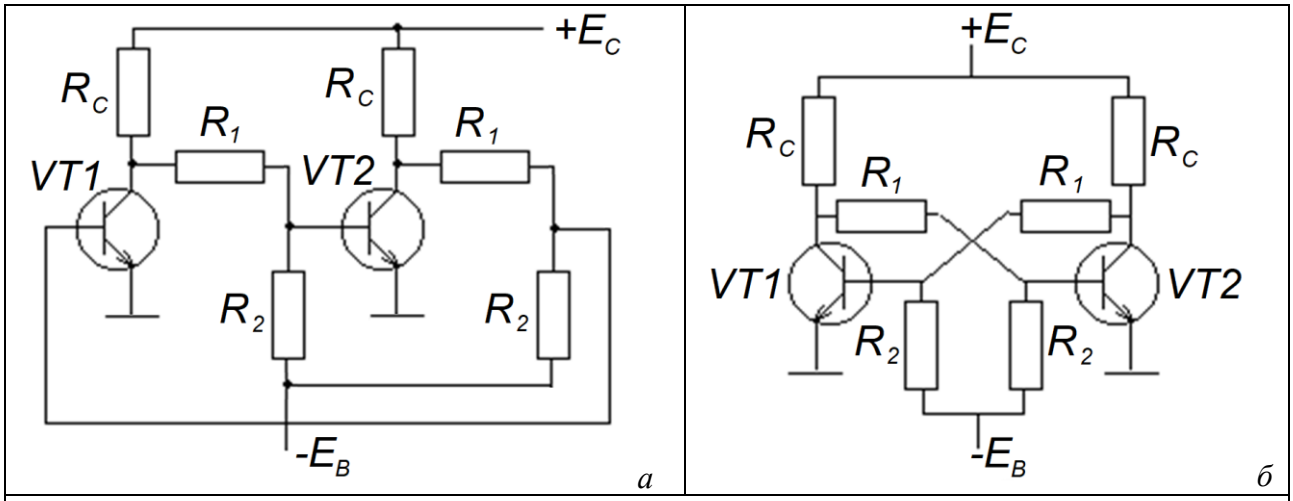


Рис. 5.6.1. Варіанти зображення принципової схеми тригера

Однак навіть при ідеальній симетрії така схема не може перебувати в симетричному стані. Як і аналогічний стан симетричного мультивібратора, такий стан тригера нестійкий, і найменша флуктуація виводить виводить тригер із симетричного стану.

Нехай, наприклад, струм першого транзистора I_{C1} дещо збільшився. Це приводить до зниження напруги U_{C1}, яке через подільник $R_1 R_2$ передається на базу другого транзистора. В результаті колекторний струм I_{C2} зменшується, а напруги U_{C2} і U_{B1} зростають, що приводить до подальшого збільшення струму I_{C1}. Якщо кожний з каскадів підсилює зміну напруги ($K > 1$), то випадкове відхилення від симетричного стану лавиноподібно (експоненціально) зростає. Це зростання триває доти, доки хоч один із транзисторів не втратить керувальної спроможності чи то через його запирання, чи то через перехід у режим насичення. На цьому перехідний процес закінчується, й тригер переходить у стійкий стан, коли один з транзисторів повністю закритий, а другий – відкритий і насичений. В такому режимі відкритий транзистор своїм низьким потенціалом на колекторі утримує інший транзистор у закритому стані, а той, у свою чергу, своїм високим колекторним потенціалом утримує перший транзистор у відкритому стані.

Для успішної роботи тригера мають виконуватися такі умови:

- транзистор підтримується у закритому стані, якщо $U_B < 0$ (точніше, якщо $U_B < U_{BE}^*$, де

U_{BE}^{*} – потенціал відкривання транзистора); для цього слід виконати умову

$$\frac{R_1}{R_2} > \frac{U_{Csat}}{|E_B|},$$

або

$$R_2 < R_1 \frac{|E_B|}{U_{Csat}} \qquad (5.6.1)$$

де U_{Csat} – напруга насичення на колекторі (вона звичайно складає кілька десятих вольта), а E_B – напруга джерела живлення, яке задає зміщення на базі (див. рис. 5.6.1; вважаємо, що струм подільника R_1, R_2 значно більший від струму, що відгалужується на базу закритого транзистора наступного каскаду);

- насичення транзистора досягається при $\beta I_B > I_C$; цій умові відповідає нерівність

$$R_1 < R_C \left(\frac{h_{21E} R_2 E_C}{h_{21E} R_C \,|\, E_B \,| + R_2 E_C} - 1 \right) \qquad (5.6.2)$$

(тут і далі β замінено близьким до неї значенням h_{21E}, а напруга на базі відкритого транзистора U_{BE} настільки мала, що нею можна нехтувати порівняно з E_C і E_B);

– коефіцієнт підсилення каскаду (з урахуванням того, що сигнал на його вхід подається через подільник R_1, R_2) повинен бути більшим за одиницю

$$K = \frac{h_{21E}}{h_{11E}} \frac{R_l}{1 + R_l h_{22E}} \frac{R_2}{R_1 + R_2} > 1, \qquad (5.6.3)$$

де опір $R_l = R_C \,||\, (R_1 + R_2)$ відіграє роль навантаження (як уже вказувалося, ця умова забезпечує розвиток лавиноподібного процесу);

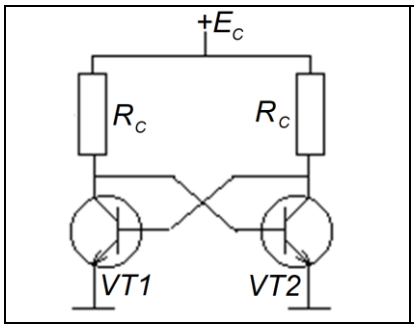
- колекторний струм відкритого транзистора не повинен перевищувати максимально допустиме значення, тобто

$$E_C / R_C < I_{C\max}; \qquad (5.6.4)$$

- напруга живлення E_C не повинна перевищувати максимально допустиму колекторну напругу для даного типу транзистора

$$E_C < E_{C\max}. \qquad (5.6.5)$$

Якщо напруга U_{Csat} на колекторі відкритого насиченого транзистора менша за напругу U_{BE}^{*}, яка відповідає відкриванню транзистора, схему тригера можна істотно спростити, поєднавши безпосередньо колектор одного транзистора з базою іншого (рис. 5.6.2 а). Ця схема особливо зручна при використанні МОН-транзисторів з індукованим каналом, для відкривання яких потрібна досить висока напруга на затворі (рис. 5.6.2 б). Тут транзистори $VT3$ та $VT4$ використовуються як навантажувальні резистори, величина опору яких визначається напругою на затворі. Схеми, зображені на рис. 5.6.2 звичайно і використовуються в інтегральних мікросхемах.



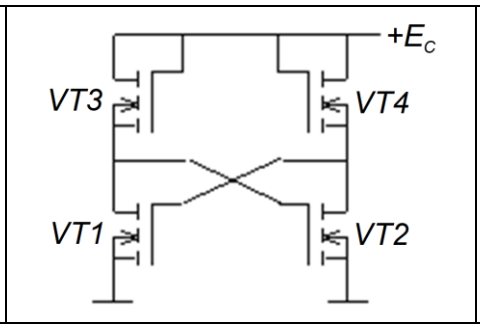


Рис. 5.6.2. Спрощені схеми тригерів на біполярних транзисторах (а) та на МОН-транзисторах з індукованим каналом (б)

5.6.2. Запуск тригера

Тригер, що перебуває в одному зі стаціонарних станів, можна перевести в інший стан, подавши короткий імпульс відкривання на базу закритого транзистора або імпульс запирання на базу відкритого. Для цього в схемі тригера передбачаються входи *R* і *S* (рис. 5.6.3). Пусковий імпульс ініціює розвиток лавиноподібного процесу(див. вище п. 5.6.1), в результаті якого транзистор, що був закритим, відкривається, а той, що був відкритим, навпаки, закривається. Режим схеми змінюється на протилежний (як то кажуть, відбувається перекидання тригера), і в такому стані схема залишається й після закінчення пускового імпульсу.

Описана схема відома як *RS*-тригер. Позначення його пускових входів походить від англійських слів Set – встановлювати та Reset – перевстановлювати. Можливі стани тригера в залежності від напруг на вході подані в табл. 5.6.1. При комбінаціях *B/H* або *H/B* (*B* – високий рівень напруги, а *H* – низький) тригер однозначно встановлюється в стан, коли на *Q*-виході рівень напруги високий або низький. При комбінації *H/H* тригер зберігає стан, у який він був попередньо встановлений вхідними сигналами. Комбінація *B/B* заборонена, оскільки при цьому обидва транзистори відкриваються, а при подальшому одночасному поданні на *S*- та *R*-входи напруги низького рівня не можна передбачити, в якому стані опиниться тригер (див. вище п. 5.6.1).

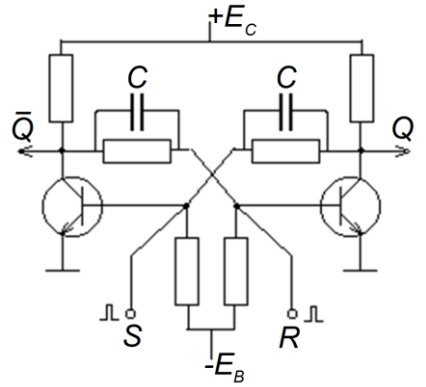


Рис. 5.6.3. Принципова схема RS-тригера

Таблиця 5.6.1

Стан тригера в залежності від вхідних напруг

Напруга на вході *S*	Напруга на вході *R*	Стан тригера	Напруга на виході *Q*
*B**	*H**	Встановлення високого рівня напруги на *Q*-виході	*B*
H	*B*	Встановлення низького рівня напруги на *Q*-виході	*H*
H	*H*	Тригер зберігає попередній стан	
B	*B*	Заборонена комбінація	

**B* та *H* – відповідно високий та низький рівень напруги.

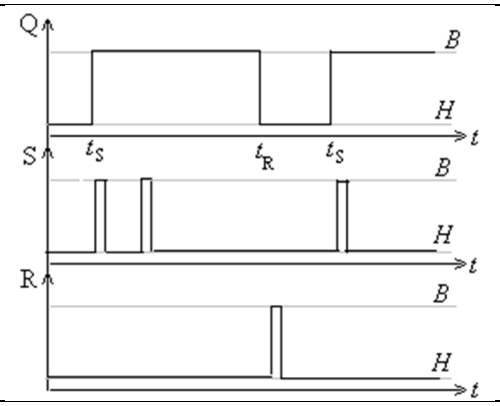


Рис. 5.6.4. Зміни стану тригера при різних комбінація напруг на його входах

На рис. 5.6.4 зображені епюри напруг на *Q*-виході та на *S* і *R* входах, які ілюструють

матеріал, поданий у таблиці 5.6.1.

Швидкість перекидання тригера визначається як схемотехнічними чинниками, так і інерційністю транзисторів. Слід мати на увазі, що транзистор, який знаходиться у стані насичення, потребує певного часу для виходу з цього стану. Щоб зменшити цей час, запас у нерівності (5.6.2), яка визначає глибину насичення, повинна бути невеликим (кілька одиниць). Для прискорення процесу перекидання рекомендується шунтувати резистори R_1 «прискорювальними» конденсаторами C (рис. 5.6.3). При цьому опір R_1 протягом перехідного процесу можна вважати начебто закороченим, а останній множник у виразі (5.6.3) – близьким до одиниці.

В інтегральних мікросхемах для запуску тригерів звичайно використовують паралельне увімкнення до транзисторів $VT1$ і $VT2$ допоміжних «пускових» транзисторів $VT3$ та $VT4$ (рис. 5.6.5). Якщо, наприклад, $VT1$ закритий, то при поданні на базу $VT3$ (S-вхід) позитивного пускового імпульсу транзистор $VT3$ відкривається. Напруга на його колекторі знижується, що призводить до закривання $VT2$ і перекидання тригера. Порівняно зі схемою, зображеною на рис. 5.6.3, схема з пусковими транзисторами має ту перевагу, що її S- і R-входи розв'язані щодо схеми тригера: зміна його режиму ніяк не відбивається на цих входах, чого не можна сказати про схему, зображену на рис. 5.6.3.

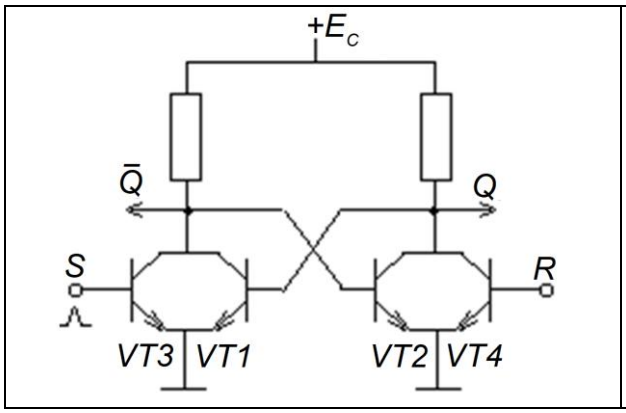


Рис. 5.6.5. Схема RS-тригера в інтегральному виконанні

Таким чином, тригер має здатність фіксувати та «запам'ятовувати» факт надходження на один з його входів (наприклад, на S-вхід) пускового імпульсу. Короткочасний, тривалістю лише в частки мікросекунди, пусковий імпульс здатен привести до того, що стан тригера зміниться, і в такому «інвертованому» стані тригер перебуватиме як завгодно довго, аж поки новий пусковий імпульс, поданий на протилежний вхід, не поверне тригер до його попереднього стану. Ця властивість тригера («електронна пам'ять»), широко використовується в цифровій радіоелектроніці для запам'ятовування та зберігання інформації, котра передається у вигляді коротких прямокутних імпульсів (див. нижче розділ 12). Тригер, що є елементарною коміркою в системах електронної пам'яті мікропроцесорів та комп'ютерів, фіксує факт надходження такого імпульсу, зберігає інформацію про це та може видавати її при відповідному запиті.

5.6.3. Тригер Шмідта

Окремим різновидом тригерів є так званий тригер Шмідта (рис. 5.6.6 а). В ньому керувальний сигнал U_{in} подається на базу одного з транзисторів через подільник напруги R_1, R_2.

При високому позитивному рівні U_{in} транзистор $VT1$ однозначно знаходиться у відкритому стані, а $VT2$ – у закритому. На виході схеми буде напруга високого рівня U_H (рис. 5.6.6 б). При достатній негативній напрузі на вході транзистор $VT1$ буде закритий, а $VT2$ – відкритий, і на виході буде напруга низького рівня U_L, близька до напруги насичення U_{Csat}.

Якщо ж вхідна напруга U_{in} близька до нуля, вона мало впливає на режим роботи схеми. В цьому випадку тригер може знаходитись як у стані, коли $VT1$ відкритий, а $VT2$ закритий, так і, навпаки, в стані, коли $VT1$ закритий, а $VT2$ відкритий. При зміні вхідної напруги від великого позитивного рівня до негативного при переході через $U_{in} \approx 0$ тригер зберігає стан, у

якому він був, тобто $U_{out} = U_H$. Вихідна напруга перейде у стан $U_{out} = U_H$ лише тоді, коли U_{in} перейде через порогове значення U_{in1}. При зворотній зміні U_{in} і проходженні її через нуль на виході зберігається напруга U_H. Перехід U_{out} до U_H відбудеться лише тоді, коли U_{in} перейде через поріг U_{in2}.

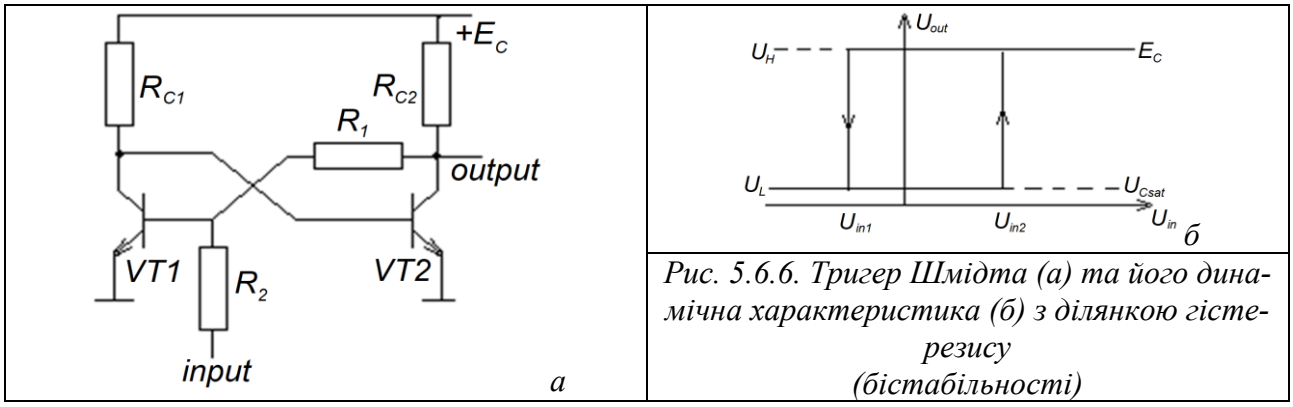


Рис. 5.6.6. Тригер Шмідта (а) та його динамічна характеристика (б) з ділянкою гістерезису (бістабільності)

Таким чином, на проміжку $U_{in1} < U_{in} < U_{in2}$ тригер Шмідта є бістабільним, тобто може перебувати в одному з двох стійких станів залежно від початкових умов. Його амплітудна характеристика утворює петлю гістерезису, в межах якої режим тригера залежить від його «передісторії», тобто від того, звідки схема перейшла в цей режим– чи то від високих, чи то від низьких значень U_{in}.

Тригер Шмідта можна використовувати для перетворення імпульсів з розмитими тривалими фронтами на імпульси з різкими фронтами. Так, наприклад, на рис. 5.6.7 зображено, як можна перетворити синусоїду на послідовність прямокутних імпульсів.

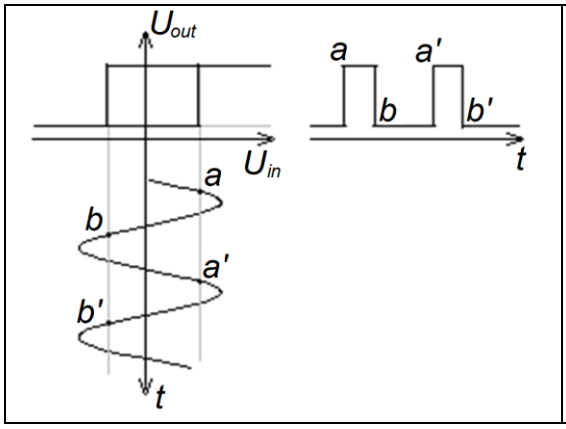


Рис. 5.6.7. Перетворення гармонічного сигналу на послідовність прямокутних імпульсів за допомогою тригера Шмідта

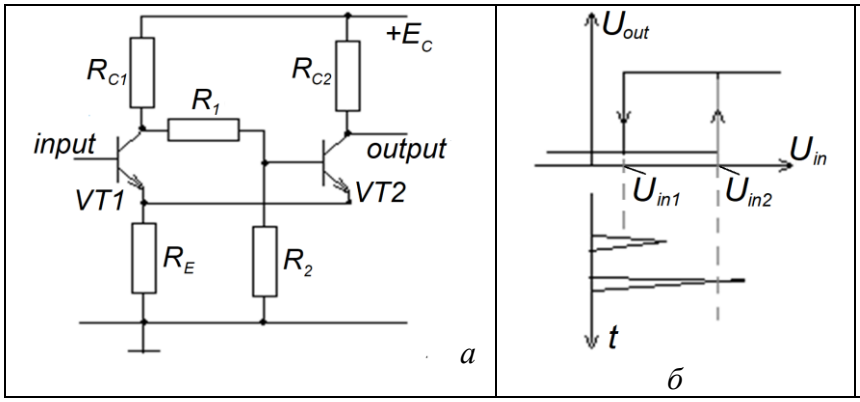


Рис. 5.6.8. Схема тригера Шмідта з емітерним зв'язком (а) та його динамічна характеристика (б)

На практиці часто використовують тригер Шмідта з емітерним зв'язком (рис. 5.6.8 а).

У цій схемі в обох стаціонарних станах через емітерний опір R_E протікає струм I_E, створюючи відповідне падіння напруги $I_E R_E$. Отже, напруга між базою і емітером транзистора VT1 у такій схемі складає величину $U_{in} - I_E R_E$. Тому для відкриття цього транзистора (якщо він закритий) і, відповідно, запирання транзистора VT2 на вхід тригера треба подати помітну додатну напругу $U_{in2} > I_E R_E$. Навпаки, якщо транзистор VT1 відкритий, то напруга, необхідна для його запирання, має задовольняти умові $0 < U_{in1} < I_E R_E$. Таким чином, тепер петля гістерезису лежить в області додатних вхідних напруг (рис. 5.6.8 б).

Тригер Шмідта з емітерним зв'язком може бути використаний, наприклад, для амплітудної дискримінації вхідних імпульсів: він змінює свій стан лише під дією імпульсів, більших від U_{in2}, і не реагує на імпульси меншої величини (рис. 5.6.8 б).

Контрольні питання до підрозділу 5.6

1. Чому схема симетричного тригера, зображена на рис. 4.5.1 *б*, не може перебувати у симетричному режимі?
2. Запропонуйте, як вдосконалити схему, зображену на рис. 4.5.1 *б*, щоб для живлення тригера можна було б обійтися однополярним (позитивним) джерелом.
3. Чому при виконанні нерівностей (5.6.1) та (5.6.2) не слід брати надто великий запас?
4. Чи може функціонувати тригер, якщо його відкритий транзистор не перебуватиме в режимі насичення?
5. Чому для побудови схеми тригера, зображеної на рис. 5.6.2 а, бажано застосувати кремнієві, а не германієві транзистори?
6. Чому в схемі, зображеній на рис. 5.6.2 *б*, бажано застосувати *МДН*-транзистори з індукованим, а не вбудованим каналом?
7. Для чого у тригері рекомендується шунтувати опір зв'язку R_1 невеликою ємністю (рис. 5.6.3)?
8. Чому для керування тригером краще подавати позитивний імпульс на базу закритого транзистора, а не негативний - на базу відкритого?
9. Які переваги має схема, зображена на рис. 5.6.5, порівняно з запуском прямим поданням пускових імпульсів на бази транзисторів (рис. 5.6.3)?
10. Запропонуйте спосіб, яким можна було б забезпечити, щоб при увімкненні живлення тригера, відкритим був би завжди певний транзистор.
11. Поясніть, чому тригер називають елементом електронної пам'яті.
12. Чому тригер Шмідта називають бістабільним?
13. З якою метою застосовують тригери Шмідта?

Задачі до підрозділу 5.6

5.6.1. У схемі тригері, зображеній на рис. 5.6.1, $R_1 = 20$ кОм, $R_2 = 10$ кОм, $R_C = 3$ кОм, $E_C = 20$ В, $E_B = -2$ В, $U_{Csat} = 0.5$ В, $h_{21E} = 40$. Перевірте виконання умов (5.6.1) та (5.6.2).

5.6.2. До даних, наведених в попередній задачі, додано $h_{11E} = 1$ кОм, $h_{22E} = 10^{-4}$ Сим. Перевірте виконання умови (5.6.5).

5.6.3. Використовуючи характеристики транзистора типу КТ201В, визначте, при якому R_C умова $U^*_{BE} > U_{Csat}$ (яка потрібна для успішної роботи схеми, зображеної на рис. 5.6.2 а) буде виконана? Напруга живлення $E_C = 10$ В.

5.7. Проходження детермінованих сигналів через нелінійні кола

Ми вже бачили, що автогенератори (як гармонічних коливань, так і імпульсні) є принципово нелінійними пристроями. У цьому розділі ми розглянемо інші види нелінійних пристроїв, що використовуються в радіоелектроніці – помножувачі частоти, модулятори, детектори, перетворювачі частоти тощо.

5.7.1. Спотворення гармонічного сигналу при проходженні через нелінійний елемент

З рис. 5.7.1 видно, що при поданні гармонічного сигналу на вхід нелінійного елементу (наприклад, підсилювача з обмеженим динамічним діапазоном[13]) зі зростанням амплітуди вхідного сигналу форма вихідного сигналу спотворюється. Такі спотворення називають нелінійними. Але при цьому вихідний сигнал залишається періодичним із тим самим періодом, що і вхідний сигнал. Тому його можна розкласти в ряд Фур'є (див. п. 1.2):

$$f(t) = c_0 + \sum_{n=1}^{\infty} |d_n| \cos\left(\frac{2\pi n}{T} t + \varphi_n\right). \qquad (5.7.1)$$

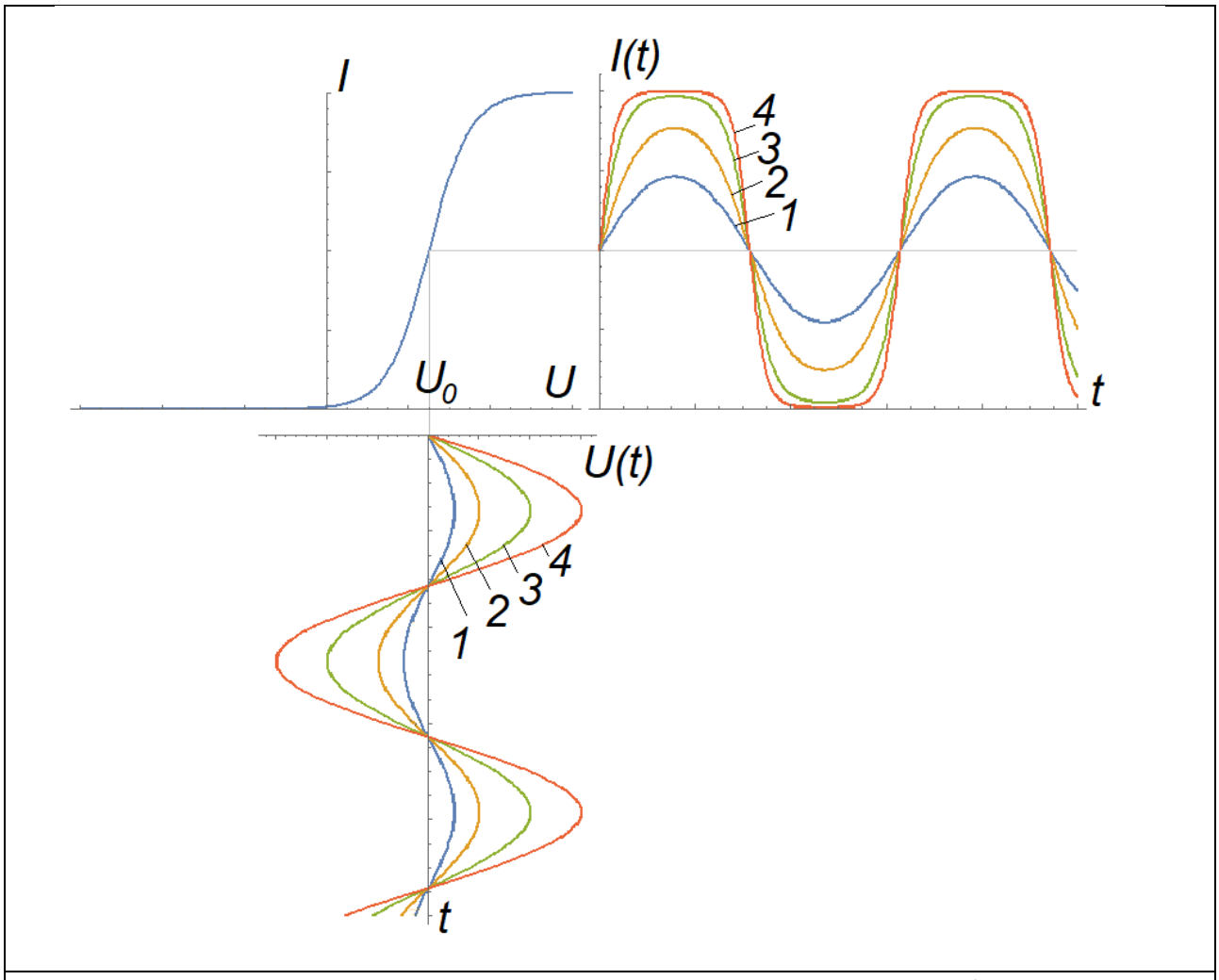


Рис. 5.7.1. Проходження гармонічного сигналу через нелінійний елемент (робоча точка обрана на середині лінійної ділянки прохідної характеристики

Кількісною мірою рівня нелінійних спотворень сигналу є коефіцієнт нелінійних спотворень, або клір-фактор:

[13] Динамічний діапазон підсилювача – це співвідношення між амплітудами найбільшого та найменшого сигналів, які можуть бути підсилені без спотворень і водночас помітні на фоні власних шумів.

$$k_f = \frac{\sqrt{\sum_{n=2}^{\infty} d_n^2}}{|d_1|}. \qquad (5.7.2)$$

Мова людини залишається зрозумілою при $k_f \le 20-30\%$, музика не дуже спотворюється при $k_f \le 7-10\%$, а гарна якість її відтворення досягається при $k_f \le 1-2\%$.

З рис. 5.7.1 очевидно, що коефіцієнт нелінійних спотворень збільшується зі збільшенням амплітуди вхідного сигналу.

5.7.2. Помножувач частоти

Історично радіоелектроніка розвивалася від нижчих частот до вищих. Пристроєм, що дозволяє підвищити частоту сигналу, є так званий помножувач частоти. Опишемо принцип його роботи.

Розглянемо проходження гармонічного сигналу через нелінійний елемент з квадратичною вольт-амперною характеристикою,

$$i = \alpha u^2. \qquad (5.7.3)$$

Робочу точку обираємо в мінімумі вольт-амперної характеристики (рис. 5.7.2). Якщо $u = u(t) = u_m \cos\omega t$, на виході маємо:

$$i(t) = \frac{1}{2}\alpha u_m^2 (1 + \cos 2\omega t). \qquad (5.7.4)$$

Навантаживши нелінійний елемент паралельним контуром, налаштованим на частоту 2ω, отримаємо на виході коливання з цією частотою (рис. 5.7.2).

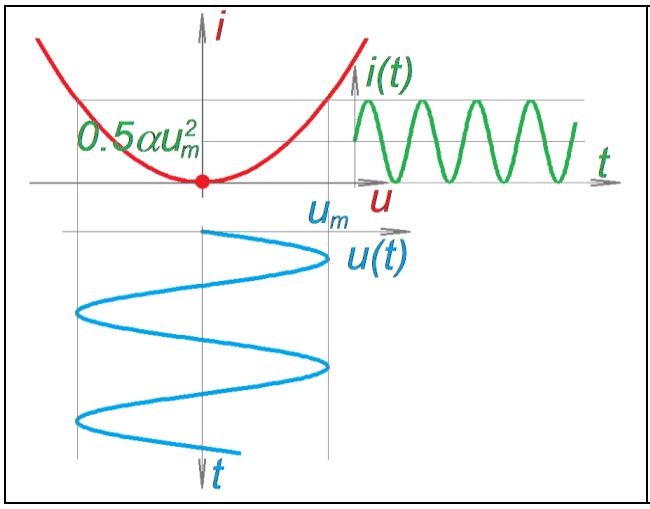


Рис. 5.7.2. До пояснення роботи квадратичного детектора (робоча точка обрана в мінімумі ВАХ)

Взявши елемент із кубічною нелінійністю, можна аналогічно отримати на виході третю гармоніку вхідного сигналу.

Насправді ВАХ довільного нелінійного елемента $i(u)$ можна розкласти в ряд Тейлора в околі обраної робочої точки,

$$i(u_0 + u_m \cos\omega t) = i(u_0) + \left.\frac{di}{du}\right|_{u_0} u_m \cos\omega t + \frac{1}{2}\left.\frac{d^2 i}{du^2}\right|_{u_0} (u_m \cos\omega t)^2 + \frac{1}{6}\left.\frac{d^3 i}{du^3}\right|_{u_0} (u_m \cos\omega t)^3 + \ldots \qquad (5.7.5)$$

і виділити потрібну гармоніку за допомогою відповідно налаштованого коливного контуру. Зазвичай коефіцієнти в ряду Тейлора (5.7.2) швидко спадають, тому таким способом вдається отримати лише другу або третю гармоніку.

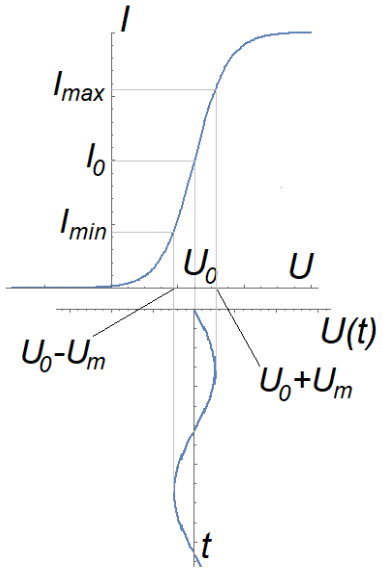


Рис. 5.7.3. До визначення крутості та кривини

Коефіцієнт

$$S = \left.\frac{di}{du}\right|_{u_0} \approx \frac{i_{\max} - i_{\min}}{2u_m} \tag{5.7.6}$$

називають (локальною) крутістю ВАХ[14], коефіцієнт

$$b = \left.\frac{1}{2}\frac{d^2 i}{du^2}\right|_{u_0} \approx \frac{i_{\max} + i_{\min} - 2i_0}{2u_m^2} \tag{5.7.7}$$

– кривиною (строго кажучи, це величина, пропорційна кривині) – див. рис. 5.7.3.

5.7.3. Детектування сигналу

Ми бачили, що до складу будь-якого радіоприймача (див. вступ, рис. В.3-В.4) входить детектор, який перетворює модульований сигнал на керувальний (сигнал модуляції). В найпростішому випадку амплітудної модуляції детектор разом з фільтром низьких частот дозволяє виділити обвідну високочастотного радіосигналу. Розглянемо принцип роботи цього пристрою.

<u>Квадратичний детектор</u>

Вигляд формули (5.7.1) зберігається й тоді, коли амплітуда вхідного сигналу є повільною функцією амплітуди, $u_m = u_m(t)$, $|du_m/dt| << \omega u_m$. Тоді перший доданок у цій формулі,

$$i(t) = \frac{1}{2}\alpha u_m^2(t),$$

відповідає детектованому сигналу обвідної на виході квадратичного детектора (рис. 5.7.2).

<u>Детектор із кусково-лінійною характеристикою</u>

Розглянемо тепер детектор із кусково-лінійною ВАХ вигляду

$$i(u) = \alpha u \;\; \theta(u) = \begin{cases} \alpha u, & u \ge 0; \\ 0, & u < 0. \end{cases} \tag{5.7.8}$$

(рис. 5.7.4). При надходженні на вхід такого детектора гармонічного сигналу $u(t) = u_m(t)\cos\omega t$ на виході отримаємо:

[14] Пор. з визначенням крутості прохідної характеристики у п. 3.4.1. Про локальну та середню крутість див. п. 5.1.4.

$i(t) = \alpha u_m(t)\cos\omega t \;\; \theta(\cos\omega t)$

(рис. 5.7.4), обвідна якого збігається із законом амплітудної модуляції $u_m(t)$ вхідного сигналу. Цю обвідну можна виділити за допомогою ФНЧ.

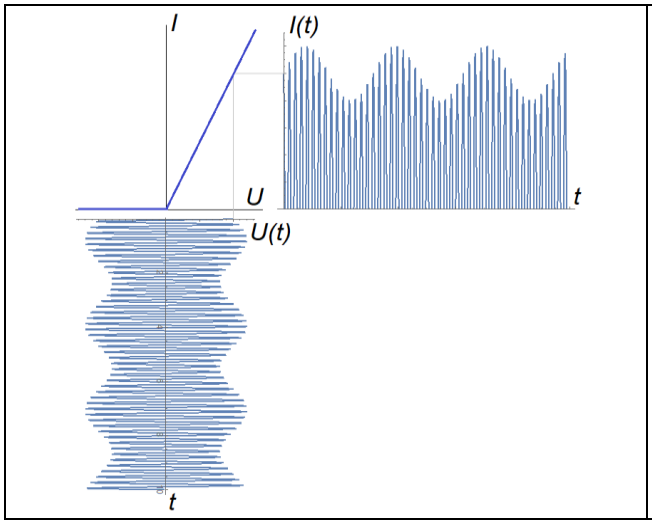


Рис. 5.7.4. Детектор з кусково-лінійною ВАХ: сигнали на вході та на виході

5.7.4. Інтермодуляційні спотворення

Вище (див. п. 5.7.1) вказувалося, що мірою нелінійного спотворення сигналу є коефіцієнт нелінійних спотворень (клір-фактор), який визначає частку вищих гармонік у вихідному сигналі при поданні на вхід гармонічного сигналу. На практиці вхідний сигнал зазвичай не є гармонічним, а складається з багатьох гармонік. При проходженні такого сигналу через нелінійні кола, крім вищих гармонік, виникають ще й комбінаційні гармоніки, або так звані інтермодуляційні спотворення. Розглянемо найпростіший приклад таких спотворень, коли на вхід подається лише дві гармоніки.

Знову розглянемо квадратичний детектор, але тепер робочу точку оберемо на схилі його вольт-амперної характеристики, для чого прикладену напругу подамо у вигляді суми постійної та змінної складових: $u = U_0 + u(t)$. Тоді струм можна записати у формі

$$i = \alpha u^2 = \alpha\left[U_0 + u(t)\right]^2 = \alpha U_0^2 + 2\alpha U_0 u(t) + \alpha u^2(t) = I_0 + Su(t) + \alpha u^2(t), \qquad (5.7.9)$$

де використані позначення

$$I_0 = \alpha U_0^2, \qquad S = 2\alpha U_0. \qquad (5.7.10)$$

Розглянемо спершу простіший випадок гармонічної змінної складової:

$u(t) = u_m \cos\omega t$.

Тоді

$$i(t) = I_0 + Su(t) + \alpha u^2(t) = I_0 + Su_m\cos\omega t + +\alpha u_m^2\cos^2\omega t =$$
$$= Su_m\cos\omega t + \left(I_0 + \frac{\alpha}{2}u_m^2\right) + \frac{\alpha}{2}u_m^2\cos 2\omega t \qquad (5.7.11)$$

(рис. 5.7.5).

Вихідний струм (5.7.11) можна розглядати як сигнал, подібний до вхідного (доданок $Su_m\cos\omega t$), до якого додана постійна складова $I_0 + \alpha u_m^2/2$ і друга гармоніка $\left(\alpha u_m^2/2\right)\cos 2\omega t$, яка в даному разі відповідає нелінійним спотворенням (див. вище п. 5.7.1). Порівнюючи формули (5.7.1) і (5.7.11), бачимо, що в даному разі відмінні від нуля лише коефіцієнти

$c_0 = I_0 + \alpha u_m^2/2$, $\qquad d_1 = Su_m\cos\omega t$, $\qquad d_2 = \alpha u_m^2/2$.

Відповідно до формули (5.7.2), клір-фактор в даному разі складає величину

$k_f = d_2/d_1 = \alpha u_m^2/2Su_m = \alpha u_m/2S = u_m/4U_0$

(враховано формулу (5.7.10)) і зростає зі збільшенням амплітуди вхідного сигналу. За умови $u_m \ll U_0$ справджується наближення малого сигналу, і нелінійні спотворення стають малопомітними, $k_f \ll 1$.

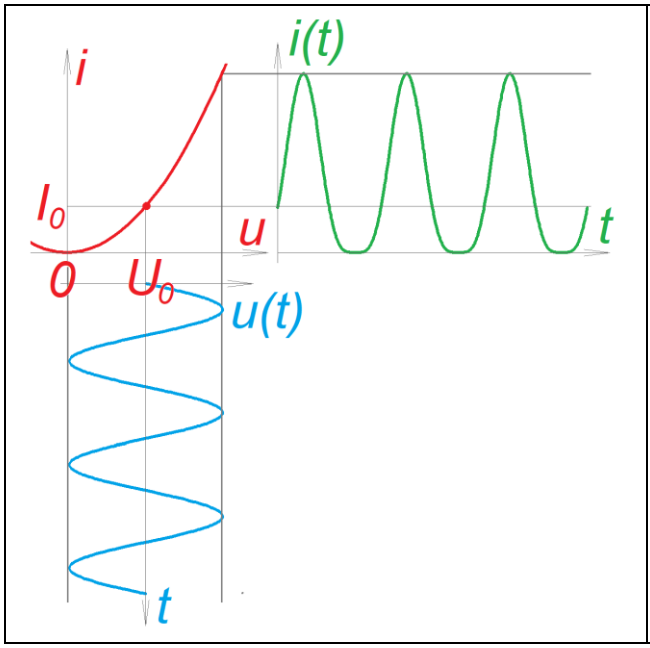


Рис. 5.7.5. . До пояснення роботи квадратичного детектора (робоча точка обрана на схилі ВАХ)

Нехай у ролі змінної складової виступає сума двох гармонічних сигналів із відмінними частотами та амплітудами:

$u(t) = u_{m1}\cos\omega_1 t + u_{m2}\cos\omega_2 t$.

Струм на виході:

$$\begin{aligned} i(t) &= I_0 + Su(t) + \alpha u^2(t) = I_0 + S\left(u_{m1}\cos\omega_1 t + u_{m2}\cos\omega_2 t\right) + \\ &+\alpha\left(u_{m1}^2\cos^2\omega_1 t + u_{m2}^2\cos^2\omega_2 t + 2u_{m1}u_{m2}\cos\omega_1 t\cos\omega_2 t\right) = \\ &= S\left(u_{m1}\cos\omega_1 t + u_{m2}\cos\omega_2 t\right) + \left[I_0 + \frac{\alpha}{2}\left(u_{m1}^2 + u_{m2}^2\right)\right] + \frac{\alpha}{2}\left(u_{m1}^2\cos 2\omega_1 t + u_{m2}^2\cos 2\omega_2 t\right) + \\ &+\alpha u_{m1}u_{m2}\left[\cos\left(\omega_1+\omega_2\right)t + \cos\left(\omega_1-\omega_2\right)t\right]. \end{aligned} \qquad (5.7.12)$$

Видно, що крім постійних складових та других гармонік, у складі вихідного сигналу з'явилися сигнали з сумарною та різницевою частотами.

Для довільного нелінійного елемента в загальному випадку на виході з'являються комбінаційні частоти вигляду $\omega_{nm} = n\omega_1 + m\omega_2, \quad n,m = \pm1,\pm2,...$ – так звані *інтермодуляційні спотворення* сигналу.

Однак виникнення комбінаційних частот можна використати для розв'язання ряду корисних із практичної точки зору задач – наприклад, амплітудної модуляції сигналу та зміни частоти-носія такого сигналу.

5.7.5. Амплітудна модуляція

Як уже відзначалося (див. вступ), для передачі аналогових сигналів в ефір використовують високочастотні модульовані сигнали, у яких амплітуда або фаза змінюється за законом керувального сигналу. Цю операцію здійснюю спеціальні пристрої – модулятори. Розглянемо принцип їхньої роботи.

Спектр амплітудно-модульованого сигналу з гармонічною модуляцією складається з частоти-носія ω_0 та двох однакових за амплітудою бічних ліній, віддалених від неї на частоту модуляції Ω (див. рис. 5.7.6, який повторює рис. 1.4.1.). Нехай у формулі (5.7.12) $\omega_1 = \omega_0$, $\omega_2 = \Omega$ ($\omega_0 \gg \Omega$). Тоді, виділивши за допомогою фільтру гармоніки з частотами ω_0 та $\omega_0 \pm \Omega$, отримаємо амплітудно-модульований сигнал. Більш наочно процес формування амплітудно-модульованого сигналу ілюструє рис. 5.7.7, а принципова схема відповідного пристрою (на біполярному транзисторі) подана на рис. 5.7.8.

Зрозуміло, що керувальний сигнал модулятора не обов'язково повинен бути гармонічним. Потрібно лише, щоб його спектральні складові мали частоти, значно менші за частоту-носій.

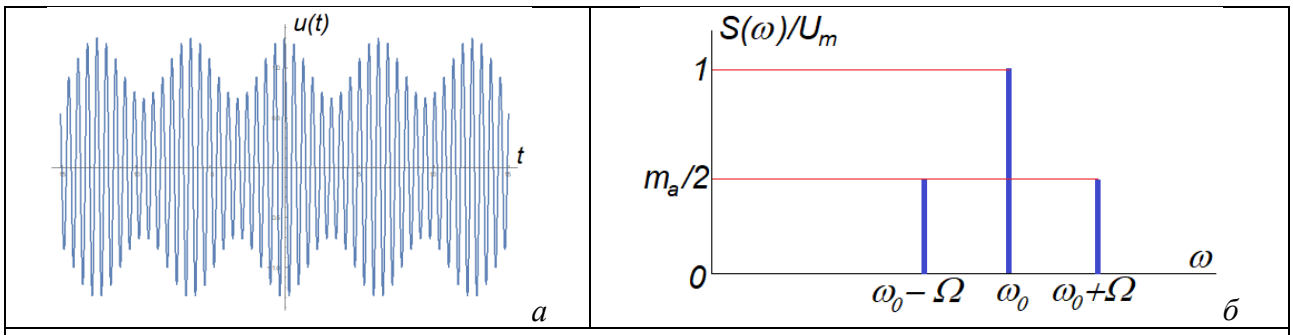


Рис. 5.7.6. Амплітудно-модульований сигнал (а) та його спектр (б)

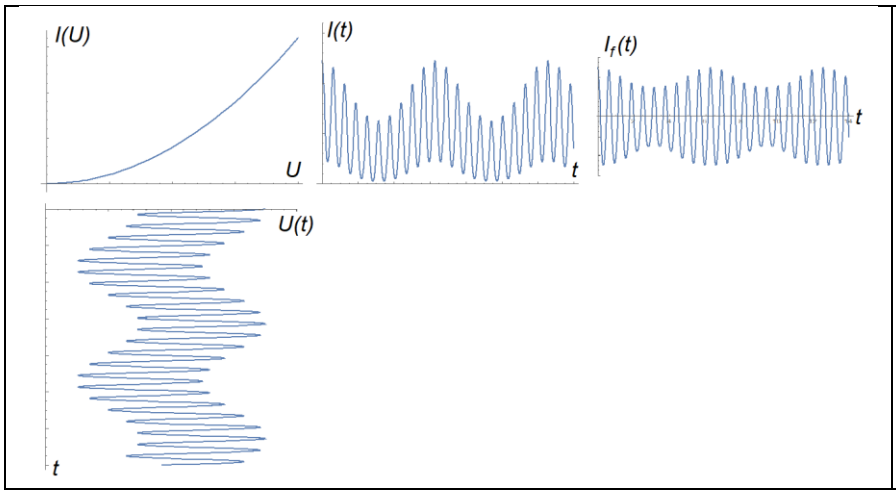


Рис. 5.7.7. Отримання амплітудно-модульованого сигналу ($I_f(t)$ – сигнал із відфільтрованими низькочастотними складовими).

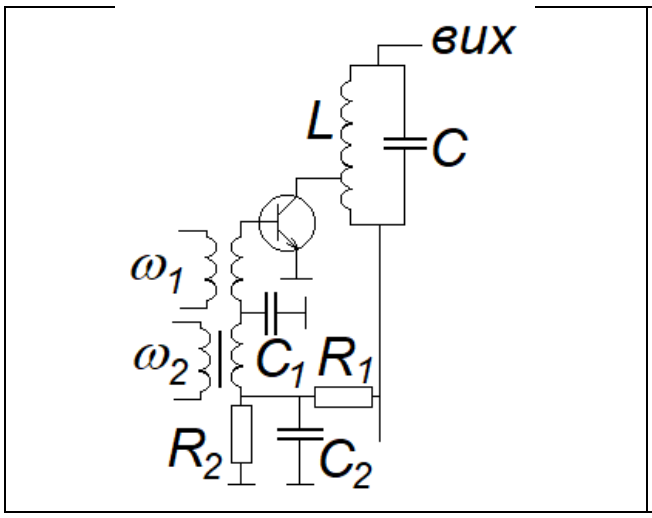


Рис. 5.7.8. Схема амплітудного модулятора на біполярному транзисторі

5.7.6. Отримання фазової модуляції

Маючи амплітудний модулятор, можна отримати й фазово-модульований сигнал (для малих індексів фазової модуляції. Принцип отримання фазово-модульованих сигналів ілюструє рис. 5.7.9. До сигналу на частоті-носії з постійною амплітудою U_1 додається амплітудно-модульований сигнал на цій самій частоті зі змінною амплітудою $U_2(t) > 0$, зсунутий за фазою на π/2 (рис. 5.7.6 а). Якщо виконана умова $U_1 >> U_2(t)$, то довжина сумарного вектора в першому наближенні залишається незмінною, $U(t) = \sqrt{U_1^2 + U_1^2(t)} \approx U_1$, а миттєвий зсув фази щодо сигнала від задаючого генератора визначається величиною $U_2(t)$. Таким способом можна отримати сигнали з невеликим індексом фазової модуляції $m_\varphi \ll 1$, в спектрі яких присутні лише дві бічні частоти (див. вище пп. 1.4.5-1.4.6).

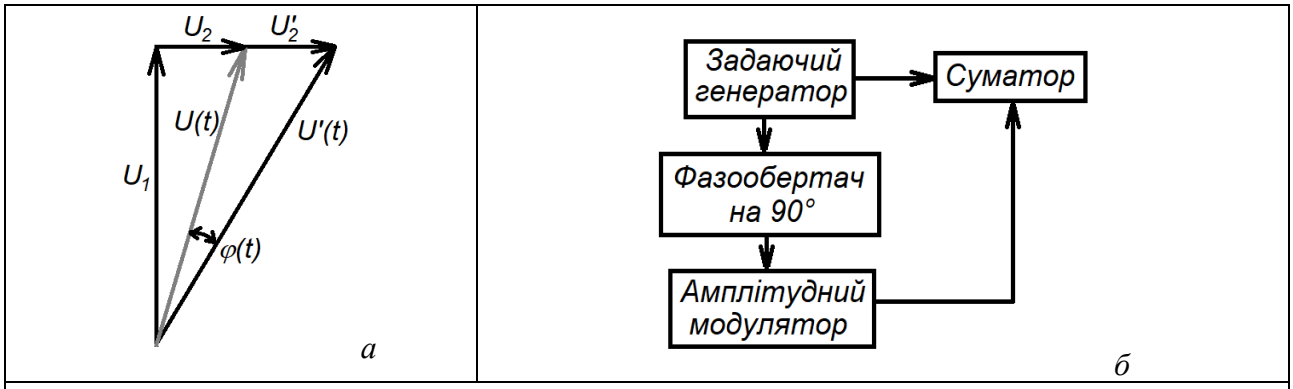


Рис. 5.7.9. Фазовий модулятор: а – векторна діаграма; б – блок-схема

5.7.7. Перетворювач частоти

Ми бачили (див. Вступ), що в супергетеродинному радіоприймачі присутній перетворювач частоти – пристрій, що переносить модуляцію сигналу з однієї частоти-носія на іншу.

Нехай у формулі (5.7.12) ω_1 – частота-носій, ω_2 – частота гетеродина, а амплітуда $u_{m1} = u_{m1}(t)$ є повільною функцією часу. Тоді доданки вигляду $\alpha u_{m1}(t) u_{m2} \cos(\omega_1 \pm \omega_2)t$ якраз і відповідають сигналу з перетвореною частотою-носієм і законом модуляції вхідного сигналу. Схема перетворювача частоти подібна до поданої на рис. 5.7.5, де вихідний контур налаштовується на потрібну (сумарну або різницеву) частоту.

5.7.8. Синхронний детектор

Нехай тепер $\omega_2 = \omega_1$. Тоді різницевій частоті відповідає сигнал вигляду $\alpha u_{m1}(t) u_{m2} \cos(\varphi_1 - \varphi_2)$, де $\varphi_{1,2}$ – фази відповідних сигналів. Звичайно намагаються підібрати $\varphi_1 - \varphi_2 = 0$. Тоді вихідний сигнал з точністю до сталого множника збігається із законом модуляції вхідного сигналу. Щоб виділити потрібний сигнал, використовують ФНЧ з вузькою смугою пропускання. Це дозволяє відфільтрувати значну частину шумів і тим самим виділити дуже слабкі корисні сигнали.

Щоб забезпечити сталість різниці фаз $\varphi_1 - \varphi_2$ в лабораторних експериментах, у ролі джерела корисного сигналу ω_1 та опорного сигналу ω_2 використовується один і той самий автогенератор.

Перевагою синхронного детектора є те, що сигнал на його виході лінійно залежить від слабкого вхідного сигналу, тоді як у звичайному детекторі ця залежність квадратична.

Блок схема пристрою, який використовує синхронний детектор, подана на рис. 5.7.7.

Областю проходження сигналу може служити, наприклад, експериментальна установка (для лабораторних експериментів) або приземна атмосфера (для радіолокаторів).

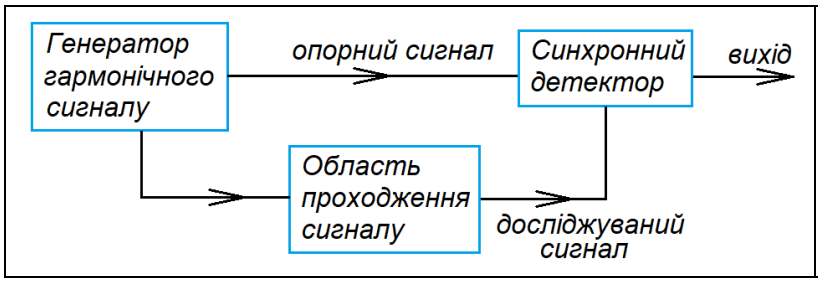


Рис. 5.7.7. Блок-схема виділення слабкого сигналу з шумів за допомогою синхронного детектора

5.7.9. Амплітудний обмежувач

Схема однобічного паралельного амплітудного обмежувача подана на рис. 5.7.8 а, епюри вхідної на вихідної напруги – на рис. 5.7.8 б. Коли до діода прикладена пряма напруга, він відкритий, і вихідна напруга збігається зі зміщенням *E*, оскільки при цьому опір діода близький до нуля. В протилежному випадку діод закритий, і вихідна напруга збігається зі вхідною.

Двобічний амплітудний обмежувач, по суті, являє собою два послідовних однобічних обмежувача, у яких полярність увімкнення джерел постійної наруги та діодів є протилежною (рис. 5.7.8 б, г). В результаті його динамічна характеристика (залежність вихідної напруги від вхідної) буде подібною до поданої на рис. 5.7.9 г.

Амплітудний обмежувач із такою кусково-лінійного характеристикою в деяких випадках дозволяє прибрати небажані зовнішні завади (вони можуть бути зумовлені, наприклад, блискавковими розрядами в атмосфері) – наприклад, для сигналів із фазовою модуляцією (рис. 5.7.8 а-в).

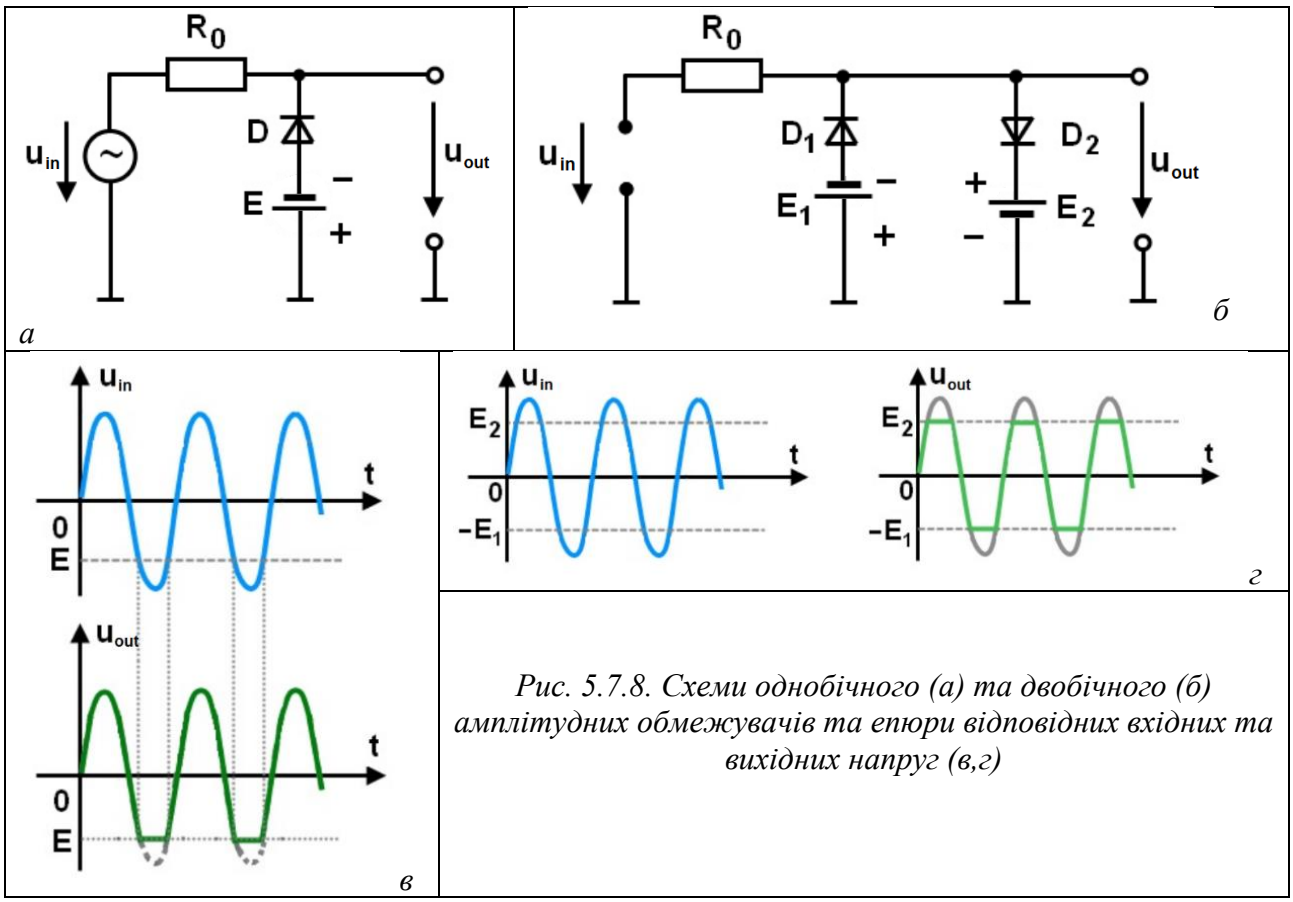


Рис. 5.7.8. Схеми однобічного (а) та двобічного (б) амплітудних обмежувачів та епюри відповідних вхідних та вихідних напруг (в,г)

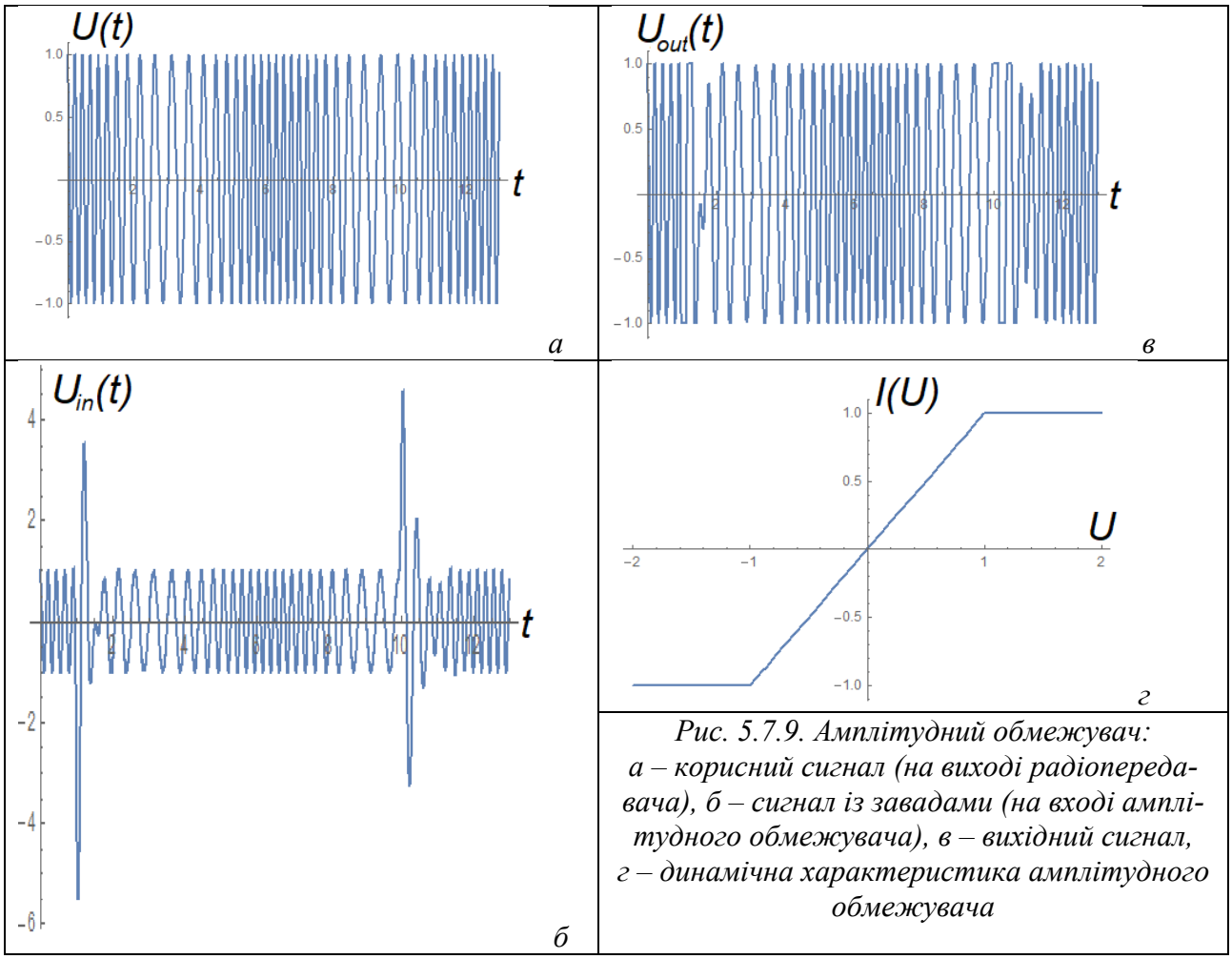


Рис. 5.7.9. Амплітудний обмежувач: а – корисний сигнал (на виході радіопередавача), б – сигнал із завадами (на вході амплітудного обмежувача), в – вихідний сигнал, г – динамічна характеристика амплітудного обмежувача

Контрольні питання до підрозділу 5.7

1. Що таке коефіцієнт нелінійних спотворень сигналу?
2. Що таке інтермодуляційні спотворення?
3. Опишіть принцип роботи помножувача частоти.
4. Які завдання вирішує детектор сигналу?
5. Опишіть принцип роботи модулятора сигналу за амплітудою.
6. Як можна отримати модуляцію сигналу за фазою?
7. Опишіть принцип роботи перетворювача частоти. Де він використовується?
8. Опишіть принцип роботи синхронного детектора. Де він використовується?
9. Для чого потрібен амплітудний обмежувач?
10. Опишіть, як працює схема амплітудного обмежувача.

Приклади розв'язання задач до підрозділу 5.7

П 5.7.1. В околі робочої точки прохідна характеристика нелінійного елемента апроксимується поліномом $i=a_0+a_1u+a_2u^2+a_3u^3+\ldots$, a_1=50 мА/В, a_2=8 мА/В^2, a_3=12 мА/В^3. Амплітуда вхідної гармонічної напруги – 1В. Знайти коефіцієнт нелінійних спотворень вихідного струму.

Розв'язок

Вихідний струм залежить від часу за законом

$$I_k(t)=a_0+a_1\left[\frac{u_m\exp(i\omega t)+u_m\exp(-i\omega t)}{2}\right]+a_2\left[\frac{u_m\exp(i\omega t)+u_m\exp(-i\omega t)}{2}\right]^2+$$

$$+a_3\left[\frac{u_m\exp(i\omega t)+u_m\exp(-i\omega t)}{2}\right]^3=a_0+a_1u_m\cos\omega t+$$

$$+\frac{1}{4}a_2u_m^2\left[\exp(i2\omega t)+2+\exp(-i2\omega t)\right]+\frac{1}{8}a_2u_m^3\left[\exp(i3\omega t)+3\exp(i\omega t)+3\exp(-i\omega t)+\exp(-i3\omega t)\right]=$$

$$=a_0+a_1u_m\cos\omega t+\frac{1}{2}a_2u_m^2+\frac{1}{2}a_2u_m^2\cos 2\omega t+\frac{1}{4}a_2u_m^3\cos 3\omega t+\frac{3}{4}a_2u_m^3\cos\omega t=$$

$$=\left(a_0+\frac{1}{2}a_2u_m^2\right)+\left(a_1u_m+\frac{3}{4}a_2u_m^3\right)\cos\omega t+\frac{1}{2}a_2u_m^2\cos 2\omega t+\frac{1}{4}a_2u_m^3\cos 3\omega t$$

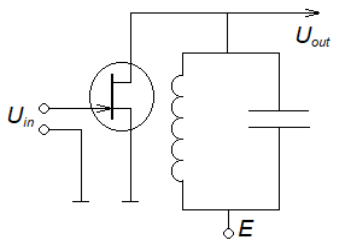
Таким чином, $I_0=a_0+\frac{1}{2}a_2u_m^2$, $I_{m1}=a_1u_m+\frac{3}{4}a_2u_m^3$, $I_{m2}=\frac{1}{2}a_2u_m^2$, $I_{m3}=\frac{1}{4}a_2u_m^3$.

Коефіцієнт нелінійних спотворень:

$$k_f=\frac{\sqrt{I_{m2}^2+I_{m3}^2}}{I_{m1}}=\frac{\sqrt{\left(\frac{1}{2}a_2u_m^2\right)^2+\left(\frac{1}{4}a_2u_m^3\right)^2}}{a_1u_m+\frac{3}{4}a_2u_m^3}=\frac{\sqrt{(a_2u_m)^2+\left(\frac{1}{2}a_2u_m^2\right)^2}}{2a_1+\frac{3}{2}a_2u_m^2}=$$

$$=\frac{\sqrt{8^2+6^2}}{2\cdot 50+1.5\cdot 8}=\frac{10}{112}\approx 0.09$$

П 5.7.2. Навантаженням польового транзистора є контур, налаштований на другу гармоніку вхідного сигналу з частотою 159 кГц і амплітудою 0.1 В Ємність конденсатора контуру – 2 нФ, добротність контуру – 50. Кривина прохідної характеристики в околі робочої точки – 20 мА/В^2. Вихідний опір польового транзистора вважати великим. Знайти амплітуду напруги другої гармоніки на виході схеми.



Розв'язок

Амплітуда другої гармоніки струму на виході польового транзистора – $i_{m2}=bu_m^2/2$, де b – кривина (див. формулу (5.7.7)). Тоді амплітуда напруги другої гармоніки на виході буде

$$u_{m2}=i_{m2}R_{eq}=\frac{1}{2}bu_m^2\frac{Q}{2\pi(2f_1)C}=\frac{1}{2}2\cdot 10^{-2}\cdot\left(10^{-1}\right)^2\frac{50}{6.3\cdot 2\cdot 1.59\cdot 10^5\cdot 2\cdot 10^{-9}}=\frac{5\cdot 10^{-3}}{4\cdot 10^{-3}}=1.25\text{ В}.$$

П 5.7.3. Вольт-амперна характеристика напівпровідникового діода має вигляд $I(U)=I_0[\exp(-\alpha U)-1]$, I_0=10 мкА, α=15 В$^{-1}$. Знайти крутість і кривину цієї характеристики в околі нуля. Яку амплітуду керувальної напруги треба прикласти до діода, щоб отримати на виході коефіцієнт амплітудної модуляції 30%?

Розв'язок

$$I(U)=I_0\left[\exp(\alpha U)-1\right].$$

Крутість:

$$S=\frac{dI}{dU}=\alpha I_0\exp(\alpha U),\ S(U=0)=\alpha I_0=0.15\text{ мА/В};$$

кривина:

$$b=\frac{1}{2}\frac{d^2I}{dU^2}=\frac{1}{2}\alpha^2 I_0\exp(\alpha U),\ b(U=0)=\frac{1}{2}\frac{d^2I}{dU^2}=\frac{1}{2}\alpha^2 I_0=0.5\cdot 225\cdot 0.01\approx 1.13\ \text{мА/В}^2.$$

Амплітудно-модульований сигнал:

$$I(t)=I_m(1+m\cos\Omega t)\cos\omega t=I_m\cos\omega t+mI_m\cos\Omega t\cos\omega t=$$

$$=I_m\cos\omega t+\frac{1}{2}mI_m\left[\cos(\omega+\Omega)t+\cos(\omega-\Omega)t\right]$$

Отже, $m=\dfrac{2I_m(\omega\pm\Omega)}{I_m(\omega)}$.

При поданні на вхід суми сигналів маємо:

$$I(U)\approx SU+bU^2=S(U_m\cos\omega t+U_{m1}\cos\Omega t)+b(U_m\cos\omega t+U_{m1}\cos\Omega t)^2\approx$$

$$\approx SU_m\cos\omega t+bU_mU_{m1}\left[\cos(\omega+\Omega)t+\cos(\omega-\Omega)t\right]$$

Таким чином,

$$m=\frac{2bU_mU_{m1}}{SU_m}=\frac{2b}{S}U_{m1},$$

звідки

$$U_{m1}=\frac{mS}{2b}=\frac{m\alpha I_0}{2\cdot\frac{1}{2}\alpha^2 I_0}=\frac{m}{\alpha}=\frac{0.3}{15}=0.02B.$$

Задачі до підрозділу 5.7

5.7.1. В околі робочої точки прохідна характеристика нелінійного елемента апроксимується поліномом $I=a_0+a_1u+a_2u^2+a_3u^3+\ldots$, a_1=40 мА/В, a_2=7 мА/В^2, a_3=15 мА/В^3. Амплітуда вхідної гармонічної напруги – 1.5 В. Знайти коефіцієнт нелінійних спотворень вихідного струму.

5.7.2. Навантаженням польового транзистора є контур, налаштований на третю гармоніку вхідного сигналу з частотою 159 кГц і амплітудою 0.2 В Ємність конденсатора контуру – 1 нФ, добротність контуру – 50. Кривина прохідної характеристики в околі робочої точки – 25 мА/В^2. Вихідний опір польового транзистора вважати великим. Знайти амплітуду третьої гармоніки напруги на виході схеми.

5.7.3. Вольт-амперна характеристика напівпровідникового діода має вигляд $I(U)=I_0[\exp(-\alpha U)-1]$, I_0=15 мкА, α=20 В$^{-1}$. Знайти крутість і кривину цієї характеристики в околі нуля. Яку амплітуду керувальної напруги треба прикласти до діода, щоб отримати на виході коефіцієнт амплітудної модуляції 20%?

Частина III. Передавання сигналів на віддаль

У цій частині курсу розглядається передавання радіосигналів від їхнього джерела (передавача) до споживача (приймача). Як відзначалось у вступі, це можна зробити через ефір або через лінію зв'язку. Відповідно розділ 6 цієї частини присвячений лініям зв'язку (переважно на прикладі двопровідної лінії), розділ 7 – передавальним та приймальним антенам, а розділ 8 – поширенню радіохвиль у приземному просторі.

Розділ 6. Лінії зв'язку

Сигнал від джерела до споживача в принципі може бути переданий двома шляхами.

Перший шлях – передача сигналу за допомогою деякої спеціально створеної лінії передачі, яка їх сполучає. В ролі такої лінії можуть виступати двопровідна лінія (в радіодіапазоні), хвилевід, смужкова лінія, коаксіальний кабель (мікрохвильовий діапазон), волоконно-оптична лінія зв'язку (в оптичному діапазоні) – рис. 6.1.1. Звичайно лінії зв'язку забезпечують високу якість передачі сигналу, але найчастіше використовуються, якщо віддаль між джерелом сигналу й споживачем невелика.

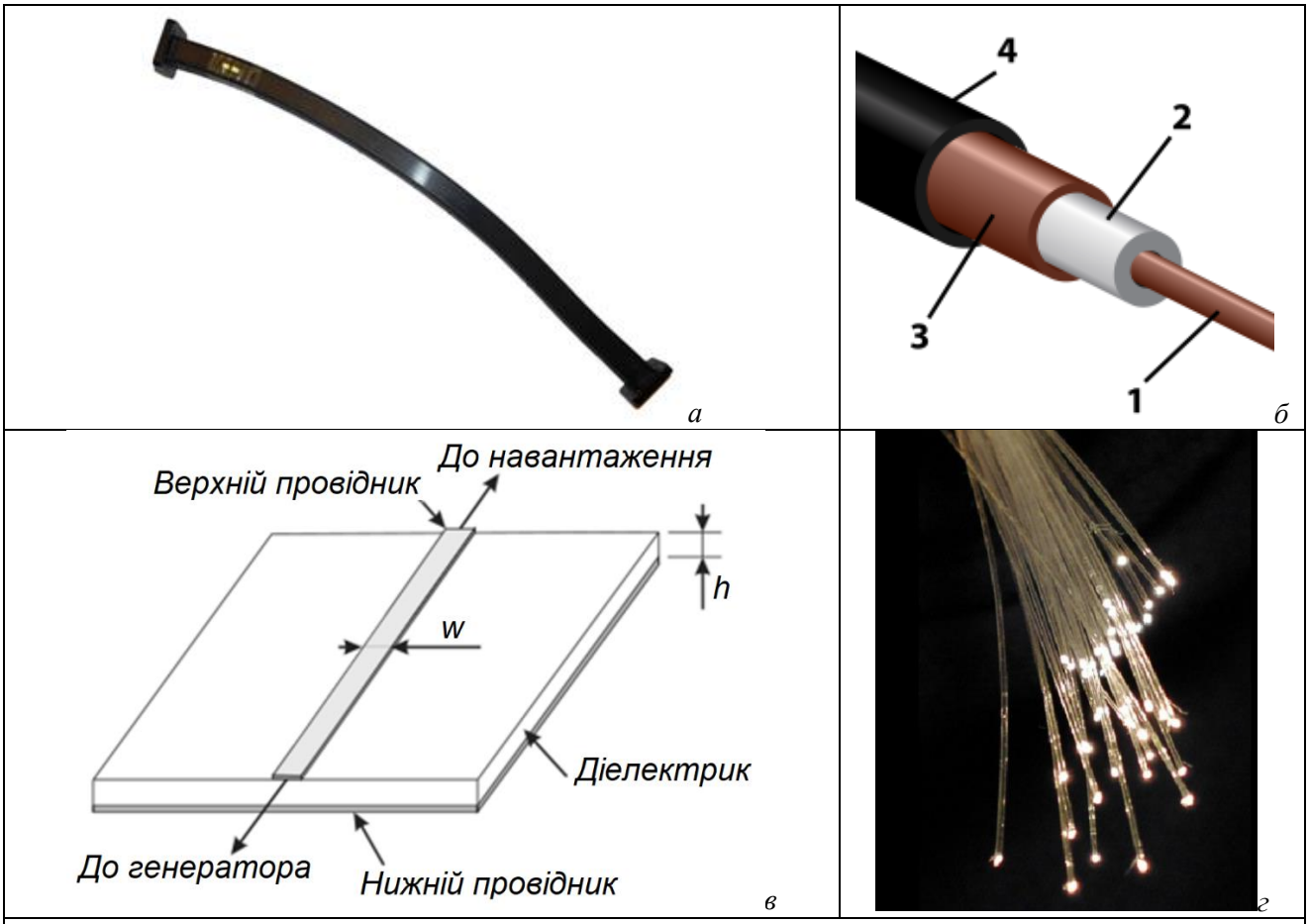


Рис. 6.1. Лінії передачі: а – відрізок прямокутного хвилеводу з фланцями; б – схема коаксіального кабелю (1 – внутрішній провідник, 2 – шар діелектрика, 2 – зовнішній провідник, 3 – оболонка); в – схема мікросмужкової лінії; г – пучок оптичних волокон

Другий шлях, про який була мова у Вступі– це випромінювання сигналу в ефір і подальше приймання та обробка його споживачем (рис. 6.1.2). Цей шлях не потребує прокладання спеціальних ліній зв'язку, але він значно менш захищений від можливих шумів та завад. Випромінюванню та поширенню радіохвиль будуть присвячені наступні розділи цього курсу. Втім, і в такій схемі використовуються довгі лінії – цього разу для сполучення між передавачем (приймачем) та передавальною (приймальною) антеною.

У цій частині курсу ми детально розглянемо процеси, які протікають у лініях зв'язку, на прикладі двопровідної лінії, й стисло обговоримо використання хвилеводів. Інші типи ліній зв'язку розглядатимуться в наступних навчальних курсах.

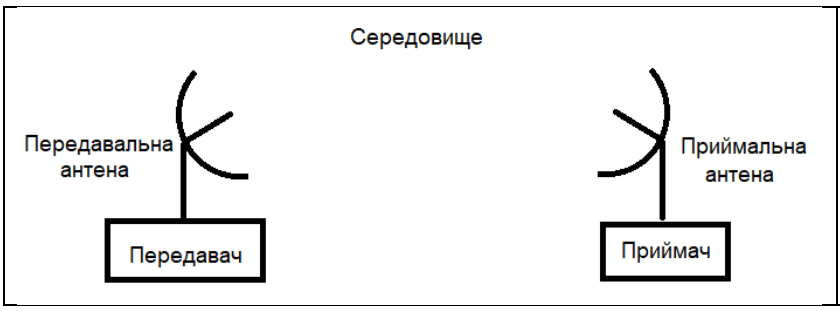


Рис. 6.1.2. Схема тракту передавання – приймання сигналу

6.1. Елементи теорії хвиль

На початку даного курсу йшлося, що системи з розмірами, набагато меншими від довжини електромагнітної хвилі на частоті сигналу, який нас цікавить, можуть бути описані рівняннями в повних похідних (наприклад, струм у довільному перерізі провідника можна вважати однаковим, тобто він залежить лише від часу, але не від координати).
Для ліній передачі маємо іншу картину: наприклад, у двопровідній лінії напруга між провідниками суттєво змінюється в просторі, тому вона є функцією як часу, так і координати. Отже, такі системи (їх називають системами з розподіленими параметрами) описуються рівняннями в частинних похідних.

6.1.1. Двопровідна лінія. Телеграфні рівняння

Розглянемо двопровідну лінію (рис. 6.1.3 а), яку можна вважати найпростішим прикладом лінії передачі. Для спрощення вважатимемо, що один з провідників (нижній) в усіх точках має потенціал землі. Оскільки при заряджанні (верхнього) провідника навколо нього виникає електричне поле, а при протіканні через нього струму – магнітне, то кожному відрізку лінії можна приписати деяку ємність та індуктивність. Крім того, провідники завжди мають ненульовий опір, а проміжок між ними - ненульову провідність. Тому ділянку провідника завдовжки Δx можна подати у вигляді еквівалентної схеми, зображеної на рис. 6.1.3 б. Запишемо для неї рівняння за першим та другим правилами Кірхгофа:

$$\Delta u = -L\frac{\partial i}{\partial t} - ri; \qquad (6.1.1)$$

$$\Delta i = -C\frac{\partial u}{\partial t} - gu,$$

$$\Delta u \equiv u(x+\Delta x) - u(x), \qquad \Delta i \equiv i(x+\Delta x) - i(x);$$

частинні похідні враховують, що напруга і струм залежать не тільки від часу, але й від координати.

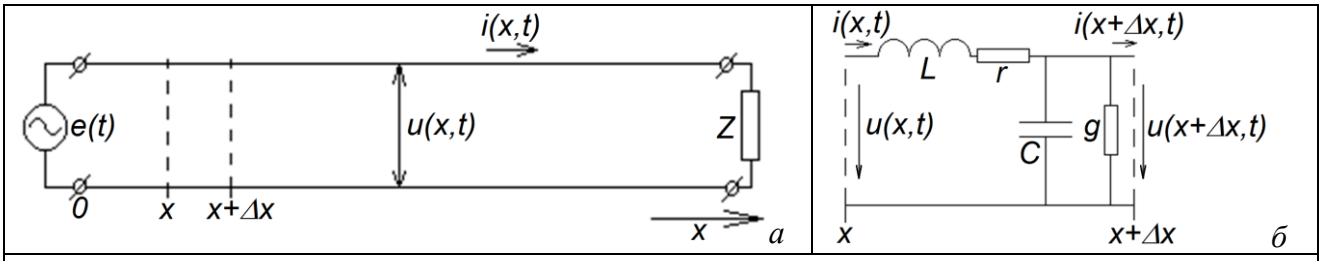


Рис. 6.1.3. Двопровідна лінія (а) та еквівалентна схема її відрізку (б)

Природно вважати, що еквівалентні параметри відрізка лінії пропорційні його довжині:

$$L = L_1\Delta x; \qquad C = C_1\Delta x; \qquad r = r_1\Delta x; \qquad g = g_1\Delta x \qquad (6.1.2)$$

(величини L_1, C_1, r_1, g_1 називаються *погонними параметрами* лінії - погонна індуктивність, погонна ємність та ін.). Вважаючи довжину відрізка Δx малою, можна записати:

$$\Delta u = \frac{\partial u}{\partial x}\Delta x, \qquad \Delta i = \frac{\partial i}{\partial x}\Delta x. \tag{6.1.3}$$

Підставляючи (6.1.2) та (6.1.3) до системи (6.1.1), дістанемо систему так званих *телеграфних рівнянь*:

$$\frac{\partial u}{\partial x} = -\left(L_1\frac{\partial}{\partial t} + r_1\right)i; \tag{6.1.4}$$

$$\frac{\partial i}{\partial x} = -\left(C_1\frac{\partial}{\partial t} + g_1\right)u.$$

Систему (6.1.4) легко звести до одного рівняння, виключивши, наприклад, струм:

$$\frac{\partial^2 u}{\partial x^2} = -\left(L_1\frac{\partial}{\partial t} + r_1\right)\frac{\partial i}{\partial x} = \left(L_1\frac{\partial}{\partial t} + r_1\right)\left(C_1\frac{\partial}{\partial t} + g_1\right)u,$$

звідки

$$\frac{\partial^2 u}{\partial x^2} - \left(L_1\frac{\partial}{\partial t} + r_1\right)\left(C_1\frac{\partial}{\partial t} + g_1\right)u = 0. \tag{6.1.5}$$

При $r_1=0$, $g_1=0$ рівняння (6.1.5) набуває вигляду

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c_0^2}\frac{\partial^2 u}{\partial t^2} = 0, \tag{6.1.6}$$

де $c_0=(L_1C_1)^{-1/2}$ (легко переконатися, що ця величина має розмірність швидкості). Рівняння (6.1.6) – це найпростіше хвильове рівняння для лінії без втрат.

6.1.2. Розв'язок хвильового рівняння. Дисперсія

Загальний розв'язок рівняння в частинних похідних записується через довільні функції, кількість яких має дорівнювати порядку рівняння. Для найпростішого хвильового рівняння (6.1.6) він, як можна переконатися шляхом безпосередньої підстановки, має вигляд

$$u(x,t) = f_1(x - c_0 t) + f_2(x + c_0 t), \tag{6.1.7}$$

де $f_{1,2}$ - довільні функції. Перша з них описує хвилю незмінної форми, що поширюється зі швидкістю c в додатному напрямку осі x, друга - подібну хвилю, що біжить у протилежному напрямку.

Однак на практиці для лінійних систем, до яких застосовні принцип суперпозиції і перетворення Фур'є, звичайно записують розв'язок хвильового рівняння у вигляді гармонічних хвиль:

$$u(x,t) = C\exp(i\omega t - ikx), \tag{6.1.8}$$

де ω – частота хвилі, k – її хвильове число (довжина хвилі – $\lambda=2\pi/k$), $\omega t - kx$ – поточна фаза. Підставивши розв'язок (6.1.8) до хвильового рівняння, отримуємо *дисперсійне співвідношення*, що пов'язує між собою частоту та хвильове число. Для найпростішого хвильового рівняння (6.1.6) воно має вигляд:

$$\omega^2 = c_0^2 k^2, \tag{6.1.9}$$

або $\omega = c_0 k$.

Знайдемо *фазову швидкість* гармонічної хвилі – швидкість поширення її *фронту*, тобто поверхні сталої фази. Взявши диференціал від фази, яку вважаємо сталою, отримуємо:

$$\omega dt - kdx = 0,$$

звідки

$$v_{ph} \equiv \frac{dx}{dt} = \frac{\omega}{k}. \tag{6.1.10}$$

Для найпростішого дисперсійного співвідношення (6.1.9) буде $v_{ph}=c_0$.

Складніші моделі систем із розподіленими параметрами приводять до складніших

хвильових рівнянь і, відповідно, складніших дисперсійних співвідношень. Наприклад, для хвиль у хвилеводі дисперсійне співвідношення має вигляд

$$\omega^2 = \omega_c^2 + c^2 k^2, \qquad (6.1.11)$$

(c – швидкість світла у вакуумі; див. нижче п. 6.7).

Як випливає з цього співвідношення, хвилі у хвилеводі можуть поширюватися лише на частотах, вищих від ω_c (так званої частоти відсічки).

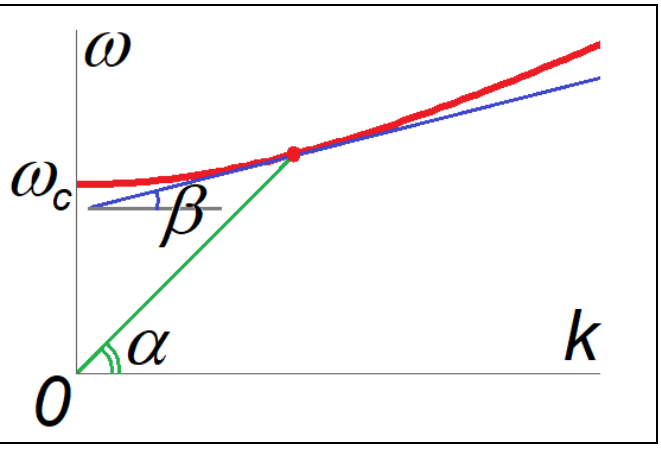
Тоді фазова швидкість такої хвилі може бути подана у вигляді

$$v_{ph} = c\sqrt{1 + \frac{\omega_c^2}{c^2 k^2}} = \frac{c}{\sqrt{1 - \omega_c^2/\omega^2}}. \qquad (6.1.12)$$

Якщо фазова швидкість хвилі залежить від її частоти або хвильового числа, говорять, що в системі має місце *дисперсія* хвиль.

Графік залежності $\omega(k)$ називають дисперсійною кривою. Геометрично фазова швидкість хвилі, що визначається формулою (6.1.10) - це тангенс кута нахилу січної, проведеної з початку координат у дану точку дисперсійної кривої (рис. 6.1.4).

Рис. 6.1.4. Дисперсійна крива, що відповідає дисперсійному співвідношенню (6.1.11); $v_{ph}=tg\alpha$, $v_g=tg\beta$



6.1.3. Поширення хвильового пакету в лінії з дисперсією

Строго гармонічна, тобто нескінчена в часі і просторі хвиля вигляду (6.1.8) - це, зрозуміло, ідеалізація. В реальності ми завжди маємо справу з хвильовими пакетами, що мають скінчену тривалість і протяжність. Оскільки різні спектральні складові хвильового пакету в середовищі з дисперсією матимуть відмінні швидкості, можна чекати, що форма хвильового пакету повинна змінюватися в процесі його поширення.

Нехай закон дисперсії довгої лінії (рис. 6.1.3 а) має вигляд $k=k(\omega)$, а сигнал на початку лінії має вигляд радіоімпульсу скінченої тривалості:

$$u(x=0,t) = u_m(t)\exp(i\omega_0 t), \qquad (6.1.13)$$

де $u_m(t)$ – повільна амплітуда ($|du_m/dt| << \omega_0|u_m|$, $u_m(t\to\pm\infty)=0$). Відбиття вважатимемо відсутнім.

Розрахуємо вигляд сигналу $u(x,t)$ в довільному перерізі лінії.

Якби замість пакету ми мали монохроматичну хвилю ($U_m=const$), можна було б відразу записати:

$$u(x,t) = u_m \exp(i\omega_0 t - ik_0 x), \qquad k_0 \equiv k(\omega_0).$$

Оскільки на початку лінії ми маємо справу з немонохроматичним сигналом, gjlfvj його у формі інтегралу Фур'є:

$$u(x=0,t) = \int_{-\infty}^{\infty} S(\omega)\exp[i(\omega_0+\omega)t]d\omega, \qquad (6.1.14)$$

$$S(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} u_m(t)\exp(-i\omega t)dt.$$

Для того, щоб отримати сигнал у довільному перерізі лінії, додамо до фаз експонент з частотами $\omega_0+\omega$ відповідні просторові набіги фаз:

$$u(x,t)=\int_{-\infty}^{\infty}S(\omega)\exp\left[i(\omega_0+\omega)t-ik(\omega_0+\omega)x\right]d\omega\,. \qquad (6.1.15)$$

Формально інтеграл (6.1.15) дає відповідь на поставлену задачу.

Врахуємо тепер, що в силу повільної зміни обвідної радіоімпульсу (6.1.13) її спектр $S(\omega)$ відмінний від нуля лише при $\omega<<\omega_0$. Отже, реально інтегрування в (6.1.15) виконується лише в деякому інтервалі частот, вузькому в масштабі ω_0. Це дозволяє нам розкласти функцію $k(\omega_0+\omega)$ в ряд Тейлора, обмежившись лінійним доданком за ω:

$$k(\omega_0+\omega)\approx k_0+\left.\frac{\partial k}{\partial\omega}\right|_{\omega=\omega_0}\omega\,. \qquad (6.1.16)$$

Очевидно, така операція буде законною при не дуже великих x, доки внесок доданків вищих порядків у фазу буде неістотним. Підставляючи (6.1.16) до (6.1.15), отримаємо:

$$u(x,t)=\int_{-\infty}^{\infty}S(\omega)\exp\left[i(\omega_0+\omega)t-i\left(k_0+\frac{\partial k}{\partial\omega}\omega\right)x\right]d\omega=$$

$$=\exp(i\omega_0t-ik_0x)\int_{-\infty}^{\infty}S(\omega)\exp\left[i\omega\left(t-\frac{x}{\partial\omega/\partial k}\right)\right]d\omega=$$

$$=u_m\left(t-\frac{x}{\partial\omega/\partial k}\right)\exp(i\omega_0t-ik_0x)\,. \qquad (6.1.17)$$

Видно, що в обраному наближенні хвильовий пакет поширюється без зміни форми, але його поширення характеризується двома швидкостями. Заповнення пакету, як і раніше, поширюється з фазовою швидкістю (6.1.10). Обвідна ж поширюється зі швидкістю

$$v_g=\frac{\partial\omega}{\partial k}\ - \qquad (6.1.18)$$

так званою груповою швидкістю. Очевидно, саме групова швидкість характеризує перенесення енергії хвильовим пакетом, тому вона не може перевищувати за величиною швидкість світла у вакуумі. Величина фазової швидкості може бути якою завгодно.

Геометрично групова швидкість хвилі - це тангенс кута нахилу дотичної до дисперсійної кривої у даній точці (рис. 6.1.4).

На закінчення відзначимо, що врахування відкинутих доданків вищих порядків у ряду Тейлора (6.1.16) буде призводити до розпливання хвильового пакету на великих віддалях від початку лінії.

Контрольні питання до підрозділу 6.1

1. Поясніть, як будують еквівалентну схему відрізку двопровідної лінії.
2. Чим визначаються втрати в двопровідній лінії?
3. Чому розв’язок хвильового рівняння зазвичай записують у формі гармонічної хвилі?
4. Що таке фазова і групова швидкість хвилі?
5. Що розуміють під дисперсією хвиль?
6. Опишіть, як поширюється хвильовий пакет у лінії з дисперсією.
7. З чим пов’язане розпливання хвильового пакету при його поширенні на далекі віддалі?

Приклади розв’язання задач до підрозділу 6.1

П 6.1.1. Закон дисперсії хвиль має вигляд $\omega^2=\omega_0^2+c^2k^2$ (хвилі в хвилеводі, електромагнітні хвилі в плазмі). Знайти фазову та групову швидкості хвиль в залежності від їхньої частоти.

Розв'язок

Фазова швидкість:

$$v_{ph}=\frac{\omega}{k}=\frac{\sqrt{\omega_0^2+c^2k^2}}{k}=\sqrt{\frac{\omega_0^2}{k^2}+c^2}\ .$$

Групова швидкість:

$$v_g=\frac{\partial\omega}{\partial k}=\frac{\partial\sqrt{\omega_0^2+c^2k^2}}{\partial k}=\frac{1}{2}\frac{2c^2k}{\sqrt{\omega_0^2+c^2k^2}}=\frac{c^2}{\sqrt{\omega_0^2/k^2+c^2}}\ .$$

Перепишемо закон дисперсії у формі

$$k^2=\frac{\omega^2-\omega_0^2}{c^2}$$

і з його допомогою виключимо k з раніше отриманих співвідношень. Остаточно отримаємо:

$$v_{ph}=\sqrt{\frac{\omega_0^2}{k^2}+c^2}=\sqrt{\frac{c^2\omega_0^2}{\omega^2-\omega_0^2}+c^2}=\frac{\omega c}{\sqrt{\omega^2-\omega_0^2}}=\frac{c}{\sqrt{1-\omega_0^2/\omega^2}}$$

$$v_g=c\sqrt{1-\omega_0^2/\omega^2}$$

(див. рис. 6.1.5).

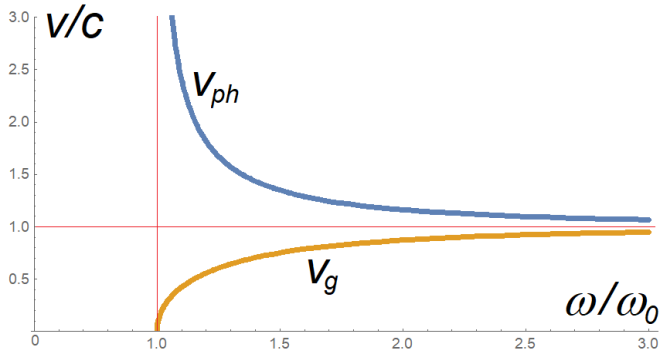


Рис. 5.1.6

П 6.1.2. Закон дисперсії хвиль має вигляд $\omega^2=\omega_0^2+c^2k^2$. На яких віддалях форма радіоімпульсу тривалості τ з частотою-носієм $\omega_1<\omega_0$ буде помітно спотворюватися?

Розв'язок

Як відзначалося вище, розпливання хвильового пакету спричинене врахуванням доданків вищого порядку в розкладанні в ряд Тейлора закону дисперсії $k(\omega)$ за малим відхиленням $\Delta\omega$ спектральних складових сигналу ω від частоти-носія ω_1 ($\Delta\omega=\omega-\omega_0$, $|\Delta\omega|\ll\omega_0$).

$$k(\omega)=k(\omega_1+\Delta\omega)\approx k(\omega_1)+\left.\frac{dk}{d\omega}\right|_{\omega=\omega_0}\Delta\omega+\frac{1}{2}\left.\frac{d^2k}{d\omega^2}\right|_{\omega=\omega_0}(\Delta\omega)^2+\ldots$$

$$\frac{dk}{d\omega}=\frac{d}{d\omega}\left(\frac{\sqrt{\omega^2-\omega_0^2}}{c}\right)=\frac{\omega}{c\sqrt{\omega^2-\omega_0^2}}=\frac{1}{c\sqrt{1-\omega_0^2/\omega^2}}$$

$$\frac{d^2k}{d\omega^2}=\frac{1}{c}\frac{d}{d\omega}\left(\frac{1}{\sqrt{1-\omega_0^2/\omega^2}}\right)=-\frac{1}{2c}\frac{2\,\omega_0^2/\omega^3}{\left(1-\omega_0^2/\omega^2\right)^{3/2}}$$

Додатковий набіг фази на віддалі x, зумовлений доданком другого порядку в ряду Тейлора, складає

$$\Delta\varphi = \frac{1}{2}\frac{d^2k}{d\omega^2}\left(\Delta\omega\right)^2 x = -\frac{\left(\Delta\omega\right)^2 x}{2c}\frac{\omega_0^2/\omega^3}{\left(1-\omega_0^2/\omega^2\right)^{3/2}}\Bigg|_{\omega=\omega_1} = -\frac{\left(\Delta\omega\right)^2 x}{2c}\frac{\omega_0^2/\omega_1^3}{\left(1-\omega_0^2/\omega_1^2\right)^{3/2}}.$$

Очевидно, цей набіг фази помітно вплине на форму імпульсу за умови $\Delta\varphi \geq \pi$.

Ширину спектру радіоімпульсу тривалістю τ можна оцінити як $\Delta\omega \approx 2\pi/\tau$. Тоді віддаль, на якій форма радіоімпульсу помітно спотвориться, можна оцінити як

$$x_{\min} = \frac{c\tau^2}{2\pi}\frac{\left(\omega_1^2-\omega_0^2\right)^{3/2}}{\omega_0^2}.$$

Задачі до підрозділу 6.1

6.1.1. Закон дисперсії хвиль має вигляд:

а) $\omega^2 = \omega_0^2\dfrac{k^2}{k_0^2+k^2}$ (акустичні хвилі в твердому тілі, іонно-звукові хвилі в плазмі);

б) $\omega = ak^2$ (т. зв. вістлери, або свистові хвилі в іоносфері Землі);

в) $\omega^2 = \omega_{01}^2 + \omega_{02}^2\sin^2 ka$ (хвилі в сповільнювальній структурі).

Знайти фазову та групову швидкості хвиль в залежності від їхньої частоти.

6.1.2. Для законів дисперсії, наведених у задачі 6.1.1, визначити, на яких віддалях форма радіоімпульсу тривалості τ з частотою-носієм ω_0 буде помітно спотворюватися.

6.2. Ланцюжкові системи

У багатьох випадках лінії передачі можна подати у вигляді еквівалентних схем, які мають вигляд ланцюжків, тобто послідовного сполучення великої кількості однакових ланок. Прикладами ланцюжкових систем можуть служити ланцюжок зв'язаних маятників (рис. 6.2.1 а), ланцюжок зв'язаних коливних контурів (рис. 6.2.1 б), молекула полімеру (довгий ланцюжок, складений з однакових ланок – мономерів). Модель ланцюжкових систем добре описує, зокрема, так звані сповільнювальні структури (структури, в яких можуть існувати повільні електромагнітні хвилі), які звичайно характеризуються просторовою періодичністю (рис. 6.2.2).

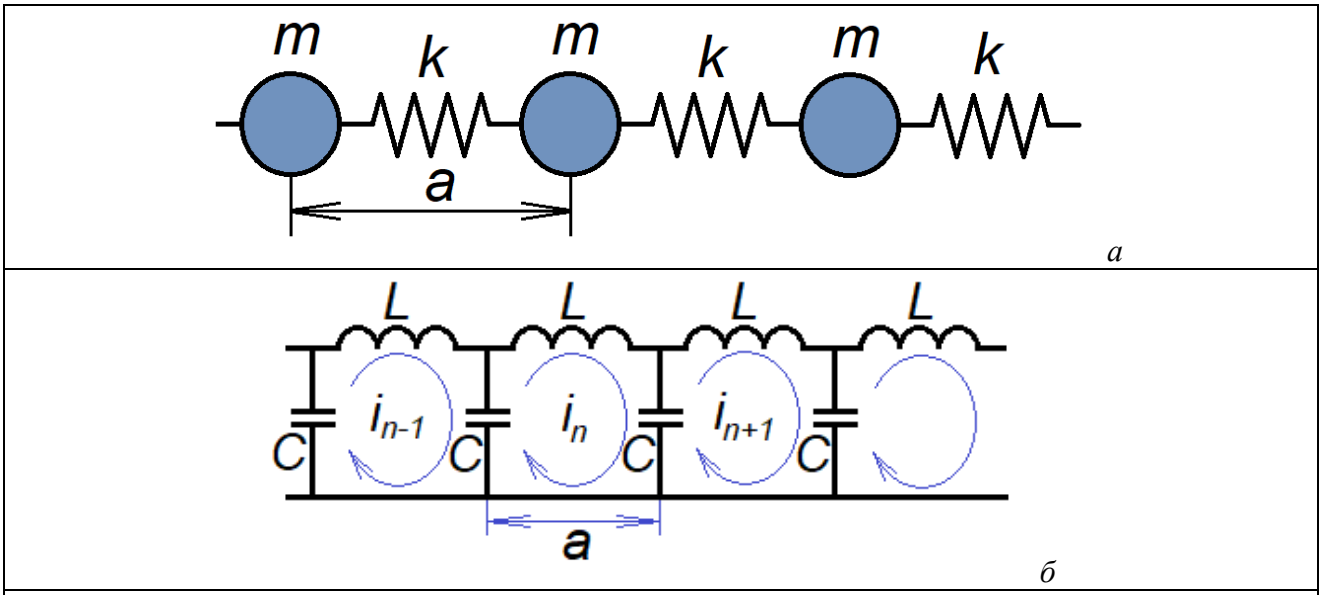


Рис. 6.2.1. Приклади нескінчених ланцюжків: а - зв'язані пружинні маятники; б - коливні контури з ємнісним зв'язком

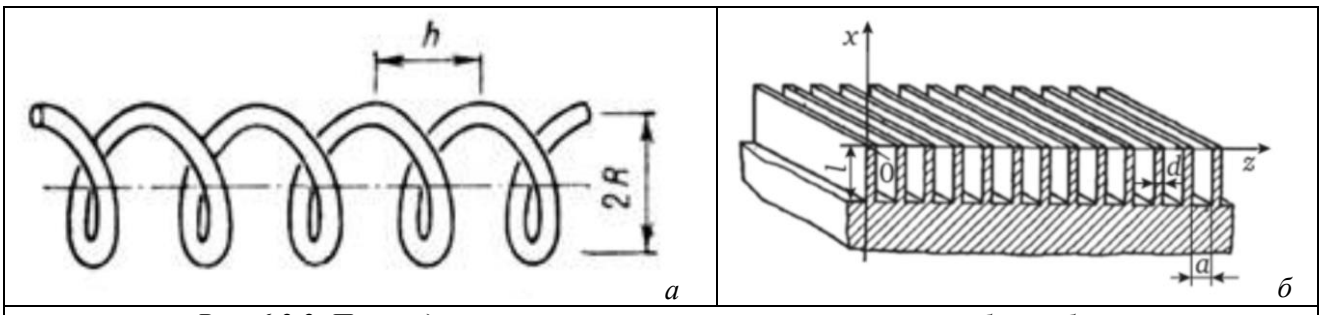


Рис. 6.2.2. Приклади сповільнювальних структур: а – спіраль, б – гребінка

6.2.1. Ланцюжкова система як частотний фільтр

Розглянемо коливання в ланцюжку, складеному з коливних контурів із ємнісним зв'язком (рис.6.2.1 б). Рівняння для *n*-го контуру, записане за другим правилом Кірхгофа, має вигляд

$$L\frac{di_n}{dt}+\frac{1}{C}\int(i_n-i_{n+1})\,dt+\frac{1}{C}\int(i_n-i_{n-1})\,dt=0\,. \qquad (6.2.1)$$

Перейшовши від струмів до зарядів,

$$q_n=\int i_n dt\,,$$

і ввівши позначення

$$\omega_0^2=\frac{2}{LC} \qquad (6.2.2)$$

(враховано, що кожен контур містить дві послідовно увімкнені ємності *C*), можна переписати рівняння (6.2.1) у формі

$$\ddot{q}_n + \omega_0^2 q_n = \frac{1}{2}\omega_0^2\left(q_{n+1} + q_{n-1}\right). \qquad (6.2.3)$$

Оскільки всі контури зв'язані між собою, а дисипація в системі відсутня, природно вважати, що коливання в сусідніх контурах мають однакову частоту і амплітуду, відрізняючись лише фазами, причому набіг фази φ на одну ланку однаковий уздовж всього ланцюжка. Тому розв'язки рівняння (6.2.3) будемо шукати у формі:

$$q_n(t) = A\exp(i\omega t - in\varphi). \qquad (6.2.4)$$

Підстановка розв'язку (6.2.4) до рівняння (6.2.3) дозволяє отримати співвідношення між набігом фази на одну ланку і частотою коливань:

$$\omega^2 = \omega_0^2(1 - \cos\varphi). \qquad (6.2.5)$$

Як випливає з (6.2.5), при зміні $\cos\varphi$ від $+1$ до -1 частота коливань змінюватиметься в межах від

$$\omega_{гр1} = 0$$

до

$$\omega_{гр2} = \sqrt{2}\omega_0 .$$

На частотах, які перевищують $\omega_{гр2}$, формально буде $|\cos\varphi|>1$. Це означає, що параметр φ стає комплексним:

$$\varphi=\alpha+i\beta, \qquad \cos\varphi = \cos(\alpha + i\beta) = \cos\alpha ch\beta - i\sin\alpha sh\beta . \qquad (6.2.6)$$

Підставивши (6.2.6) до (6.2.5) і прирівнявши окремо дійсні та уявні частини рівняння, можна отримати таку систему:

$$\cos\alpha ch\beta = 1 - \frac{\omega^2}{\omega_0^2}; \qquad (6.2.7)$$

$$\sin\alpha sh\beta = 0 .$$

Щоб задовольнити обидва рівняння (права частина першого з них при $\omega>\omega_{гр2}$ буде від'ємною), треба обрати

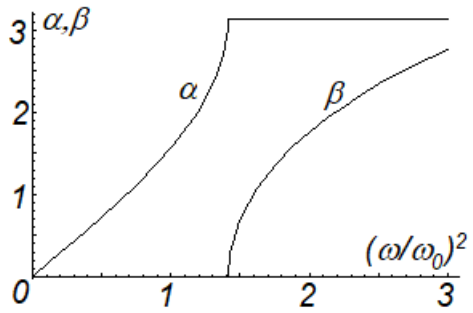
$$\alpha=\pi. \qquad (6.2.8)$$

Тоді

$$\beta_{1,2} = \ln\left[\frac{\omega^2}{\omega_0^2} - 1 \pm \sqrt{\frac{\omega^2}{\omega_0^2}\left(\frac{\omega^2}{\omega_0^2} - 2\right)}\right]. \qquad (6.2.9)$$

Легко показати, що значення коренів (6.2.9) відрізняються лише знаком, тому надалі оберемо позитивний корінь[1].

Рис.6.2.3. Залежність набігу фази між ланками α та параметра дисипації β від частоти



Частотні залежності параметрів α та β подані на рис.6.2.3. З нього видно, що аналізована схема являє собою фільтр низьких частот, смуга пропускання якого простягається від $\omega_{гр1}$ до $\omega_{гр2}$. За межами смуги пропускання сигнал у лінії екпоненційно спадатиме при відда-

[1] Наявність двох коренів вказує на те, що амплітуда коливань спадає в обох напрямках при віддаленні від джерела сигналу.

ленні від джерела.

Наявність смуг прозорості та непрозорості - це універсальна властивість ланцюжкових систем. Звичайно, вони не обов'язково є при цьому саме фільтрами низьких частот (наприклад, ланцюжок, зображений на рис.6.2.1б, перетворюється на фільтр високих частот, якщо поміняти місцями ємності та індуктивності).

6.2.2. Хвильовий процес. Непрозорість

Повернемося до розгляду ланцюжка зв'язаних контурів (рис. 6.2.1 б).

На рис.6.2.4 а, б подано миттєві розподіли струмів по ланках для відмінних моментів часу для сигналів, частоти яких потрапляють у смугу прозорості.

В першому випадку наявність постійного зсуву фаз між сусідніми ланками приводить до того, що миттєвий розподіл струму вздовж лінії є гармонічним, причому з плином часу ця гармоніка зсувається в просторі. Отже, маємо хвильовий процес. Якщо прийняти, що довжина однієї ланки дорівнює a, то номери ланок можна пов'язати з координатою x уздовж лінії співвідношенням

$$na = x\,. \tag{6.2.10}$$

Тоді, позначивши

$$\varphi = ka, \tag{6.2.11}$$

розподіл сигналу вздовж лінії (6.2.4) можна подати у вигляді хвилі:

$$q(x,t) = q_m \exp(i\omega t - ikx), \tag{6.2.12}$$

де хвильове число k пов'язане з частотою хвилі співвідношенням

$$\omega = \frac{\omega_0}{\sqrt{2}}\left|\sin\frac{ka}{2}\right| \tag{6.2.13}$$

(його можна отримати, підставивши (6.2.11) до (6.2.5)).

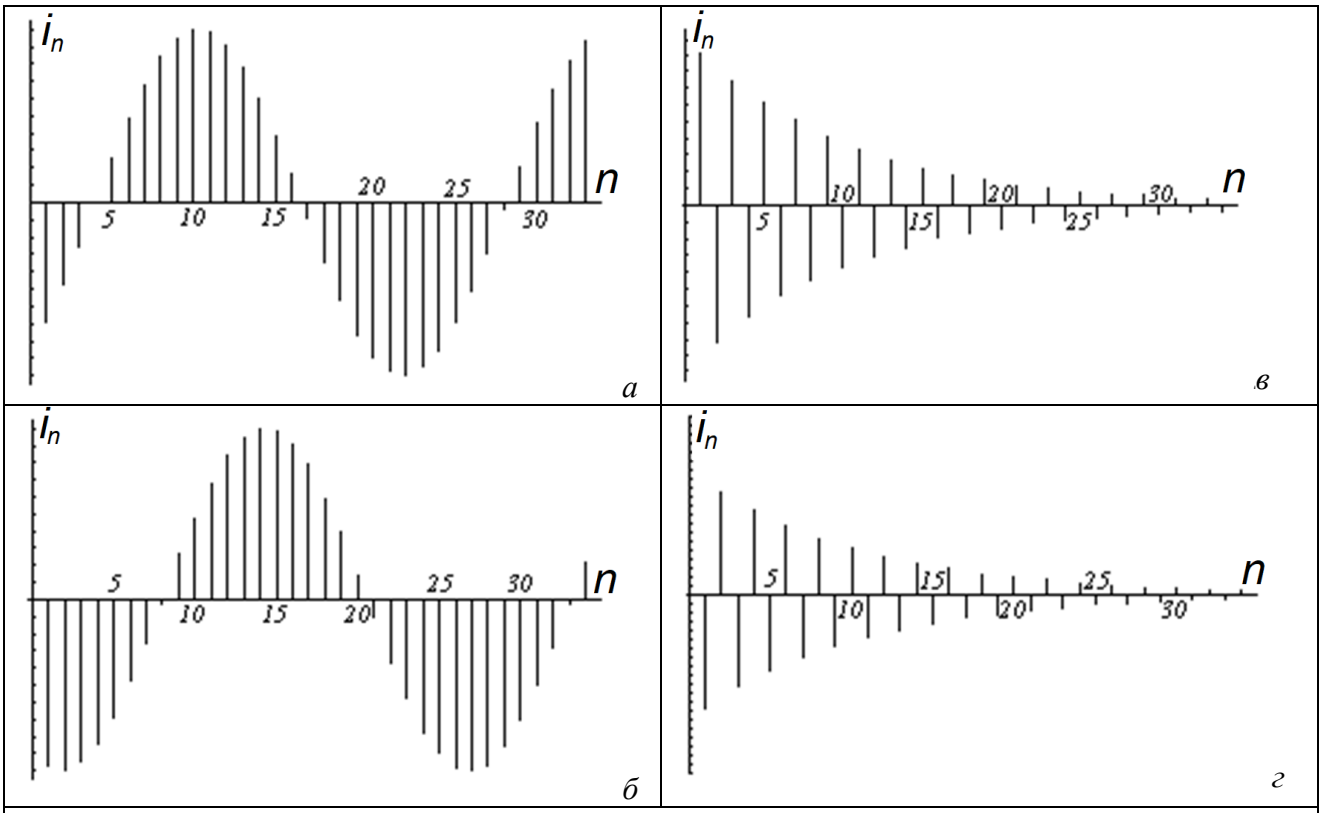


Рис.6.2.4. Миттєвий розподіл струмів по ланках для двох відмінних моментів часу в смугах прозорості (а, б) та непрозорості (в, г)

Нехай тепер частота сигналу потрапляє в смугу непрозорості. Підставивши (6.2.8)-(6.2.9) до (6.2.4), для цього випадку можна отримати:

$$q_n(t) = A(-1)^n \exp(i\omega t)\exp(-\beta n). \qquad (6.2.14)$$

Тепер коливання в усіх ланках відбуваються з однаковими фазами (якщо не брати до уваги протифазність коливань у сусідніх ланках), а амплітуди коливань експоненційно спадають при віддаленні від джерела сигналу. Відповідні миттєві розподіли струмів уздовж лінії подані на рис.6.2.4 в, г (для відмінних моментів часу).

Знову перейшовши від номерів ланок до координати згідно співвідношення (6.2.10) і врахувавши, що фізичний зміст має лише дійсна частина комплексної змінної, можна переписати (6.2.14) у формі:

$$\mathrm{Re}\left[q_n(t)\right] = A(-1)^n \exp(-\beta n)\cos\omega t. \qquad (6.2.15)$$

Непрозорість консервативних середовищ для хвиль певної частоти та напрямку - поширене явище. Для прикладу досить згадати повне внутрішнє відбиття електромагнітної хвилі, яка падає під певним кутом з оптично менш щільного в оптично більш щільне середовище. Характерно, що в експериментах Мандельштама в оптично більш щільному середовищі при цьому спостерігалися синфазні коливання електромагнітного поля, амплітуда яких експоненційно спадала при віддаленні від межі двох середовищ (пор. з формулою (6.2.15)).

6.2.3. Ланцюжкові системи як сповільнювальні структури

Знову повернемося до ланцюжкової системи, показаної на рис. 6.2.1 б. Ми бачили, що її дисперсійне рівняння в смузі прозорості має вигляд (6.2.13),

$$\omega = \frac{\omega_0}{\sqrt{2}}\left|\sin\frac{ka}{2}\right|.$$

Скориставшись цим рівнянням, побудуємо графік дисперсійної залежності $\omega(k)$ (див. рис. 6.2.5).

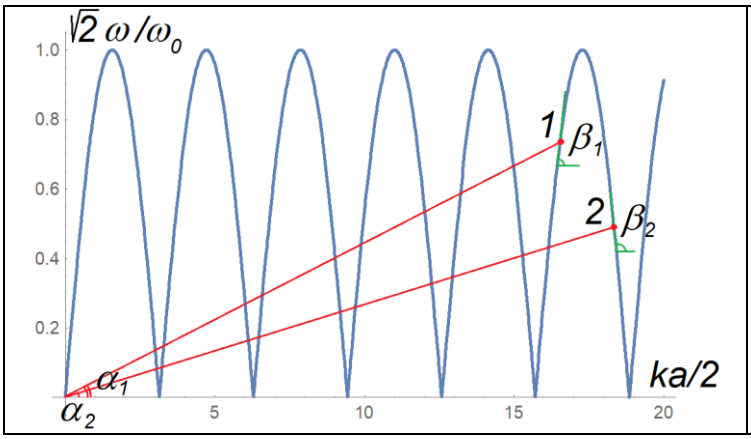


Рис. 6.2.5. Дисперсійна залежність для ланцюжка зв'язаних контурів, поданих на рис. 6.2.1 б

Перш за все зазначимо, що залежність (6.2.13) є періодичною функцією хвильового числа k. Нагадаємо, що $ka = \varphi$ – це набіг фази на одну ланку. Очевидно, зміна набігу φ на величину $2\pi n,\ n = \pm1, \pm2, \ldots$ ніяк не вплине на частоту поширюваного сигналу.

З іншого боку, як уже вказувалося (див. формулу (6.1.10))

$$v_{ph} = \frac{\omega}{k}$$

– це фазова швидкість хвилі. На графіку дисперсійної залежності їй відповідає тангенс кута нахилу січної, проведеної з початку координат у точку з координатами (ω, k). При великих k можна одержати малі значення фазової швидкості (рис. 6.2.5), де січні, проведені в точки 1 і 2, показані червоним. Таким чином, ланцюжкові системи можуть служити сповільнювальними структурами. Приклади реальних сповільнювальних структур, як уже зазначалося, показані на рис. 6.2.2.

Звернемо увагу ще на одну обставину. Груповій швидкості (6.1.18)

$$v_g = \frac{\partial \omega}{\partial k}$$

відповідає тангенс кута нахилу дотичної до дисперсійної кривої в точці (ω, k). На графіку (рис. 6.2.5) видно, що групова швидкість (дотичні показані зеленим) в точці 1 є додатною, а в точці 1 – від'ємною. Системи, в яких фазова й групова швидкості мають різний знак (як у точці 2), називаються системами з від'ємною (або зворотною) дисперсією. Хвилі в таких системах називають зворотними. Таким чином, точка 2 на дисперсійній кривій відповідає повільній зворотній хвилі (рис. 6.2.6).

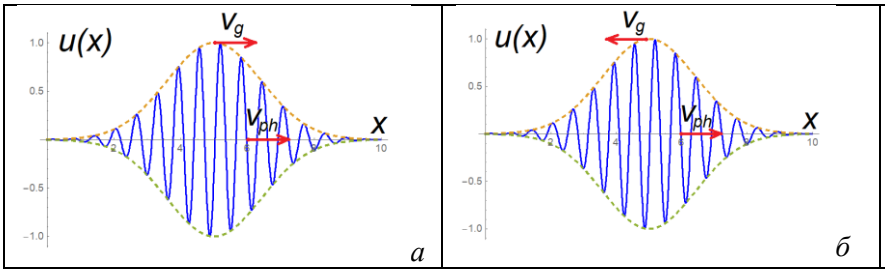


Рис. 6.2.6. Поширення радіоімпульсу в лінії з додатною (а) та від'ємною (б) дисперсією

З рис. 6.2.5 видно, що від'ємна дисперсія існує в областях, де величина φ лежить у проміжках $2\pi n - \pi < \varphi < 2\pi n$, $n = \pm 1, \pm 2, \ldots$ Якщо приводити набіг фаз φ до інтервалу $-\pi < \varphi < \pi$, то на цих проміжках φ набуває від'ємних значень, тобто можна сказати, що фронт хвилі поширюється до початку лінії. Енергія ж сигналу завжди передається від початку лінії, де розташований генератор. Тому в цих областях фазова й групова швидкості справді мають різні знаки.

6.2.4. Причини виникнення дисперсії

Розглянемо тепер складніший ланцюжок, показаний на рис. 6.2.7. Рівняння за другим правилом Кірхгофа для *n*-ї ланки має вигляд:

$$L\frac{d^2 i_n}{dt^2} + \frac{1}{C_1} i_n + \frac{2}{C_2} i_n = \frac{1}{C_2}\left(i_{n-1} + i_{n+1}\right),$$

або

$$\frac{d^2 i_n}{dt^2} + \omega_1^2 i_n + \omega_2^2 i_n = \frac{1}{2}\omega_2^2\left(i_{n-1} + i_{n+1}\right), \qquad (6.2.16)$$

де використано позначення

$$\omega_1^2 = \frac{1}{LC_1}, \qquad \omega_2^2 = \frac{2}{LC_2}.$$

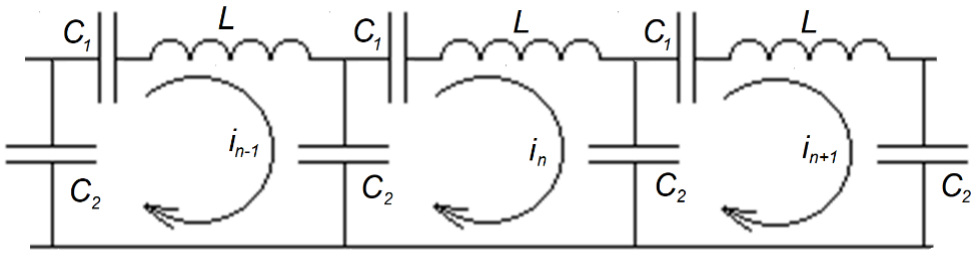


Рис. 6.2.7. Схема ланцюжкового смугового фільтру

Підставивши розв'язок у формі типу (6.2.4)

$$i_n = A\exp\left(i\omega t - in\varphi\right),$$

приходимо до дисперсійного рівняння вигляду

$$\omega^2 = \omega_1^2 + \omega_2^2\left(1-\cos\varphi\right),$$

або

$$\omega^2 = \omega_1^2 + 2\omega_2^2 \sin^2\frac{ka}{2}. \qquad (6.2.17)$$

Відповідна дисперсійна залежність подана на рис. 6.2.8 а.

а) Спростимо модель, поклавши $\omega_1 = 0$. Тоді дисперсійне рівняння набуде вигляду

$$\omega = \sqrt{2}\omega_2\left|\sin\frac{ka}{2}\right| \qquad (6.2.18)$$

При $ka \ll 1$, або $a \ll \lambda$

$$\omega = \frac{1}{\sqrt{2}}\omega_2 ka, \qquad (6.2.19)$$

тобто дисперсія зникає. Отже, в даному разі причиною виникнення дисперсії виступає існування характерного розміру в системі.

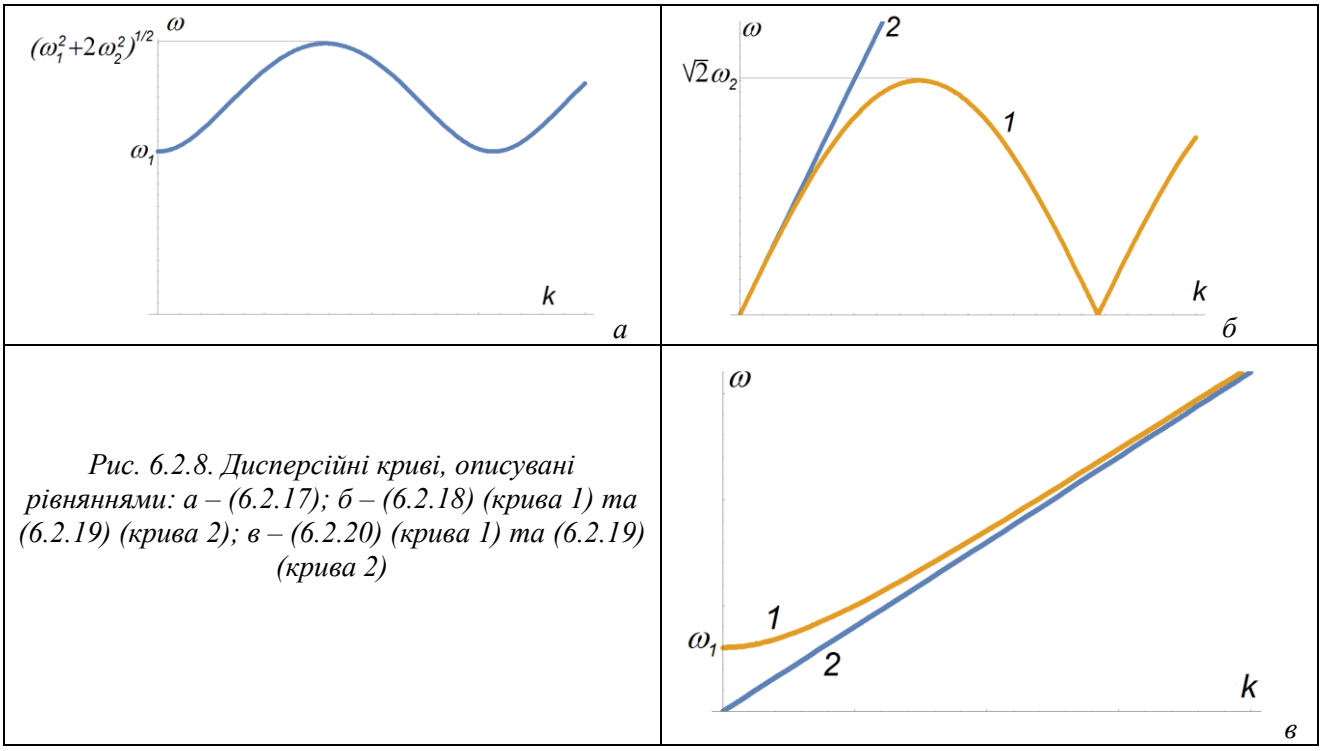


Рис. 6.2.8. Дисперсійні криві, описувані рівняннями: а – (6.2.17); б – (6.2.18) (крива 1) та (6.2.19) (крива 2); в – (6.2.20) (крива 1) та (6.2.19) (крива 2)

б) Повернемося до загального випадку. При $ka \ll 1$, або $a \ll \lambda$ рівняння (6.2.17) набуває вигляду

$$\omega^2 = \omega_1^2 + \frac{1}{2}\omega_2^2 k^2 a^2. \qquad (6.2.20)$$

При $\omega^2 \gg \omega_1^2$

знову отримуємо рівняння (6.2.19):

$$\omega = \frac{1}{\sqrt{2}}\omega_0 ka.$$

Тепер причиною дисперсії є наявність власної частоти.

6.2.5. Механізм формування смуг прозорості та непрозорості

Розглянемо тепер ланцюжок, складений з k ланок, вигляд яких зображений на рис. 6.2.1 б. скінчена довжина ланцюжка приведе до того, що для його власних коливань набіг фази φ на одну ланку може набувати лише дискретного набору значень.

Нехай коливання, що виникло в першій ланці, поширюється уздовж ланцюжка. В останній ланці воно буде зсунуте за фазою на *(k-1)φ*. Від останньої ланки сигнал відіб'ється і побіжить уздовж ланцюжка назад. Сигнал, що повернувся в першу ланку, буде, очевидно, зсунутий на *2(k-1)φ* щодо початкового коливання. Аналізована мода буде власною модою ланцюжка тоді, коли обидва коливання виявляться синфазними, тобто при

$$2(k-1)\varphi = 2\pi n, \qquad n = 0, \pm 1, \pm 2, \pm 3... \tag{6.2.21}$$

Підставляючи отримані значення

$$\varphi_n = \frac{n\pi}{k-1}$$

до рівняння (6.2.5), можна отримати дискретний спектр частот:

$$\omega_n = \frac{\omega_0}{\sqrt{2}}\left|\sin\frac{\varphi_n}{2}\right|. \tag{6.2.22}$$

Оскільки

$$\sin(\pi - z) = \sin z, \qquad \sin(\pi + z) = -\sin z,$$

значення частот, які не повторюються, будуть виникати лише при зміні φ_n в межах від *0* до π, тобто при зміні індексу *n* від *0* до *k*. Таким чином, ми отримаємо *k* відмінних значень власних частот.

Спрямуємо *k* – кількість ланок у ланцюжку – до нескінченності. Видно, що при зростанні *k* віддаль між сусідніми частотами буде поступово зменшуватись (рис. 6.2.9). В результаті формується неперервна смуга прозорості. Підсумовуючи, можна сказати, що вона з'являється в результаті взаємодії великої кількості ідентичних осциляторів як наслідок зв'язку між ними.

Якщо окремі ланки, з яких складений ланцюжок, мають декілька власних частот, може виникнути кілька зон прозорості та непрозорості.

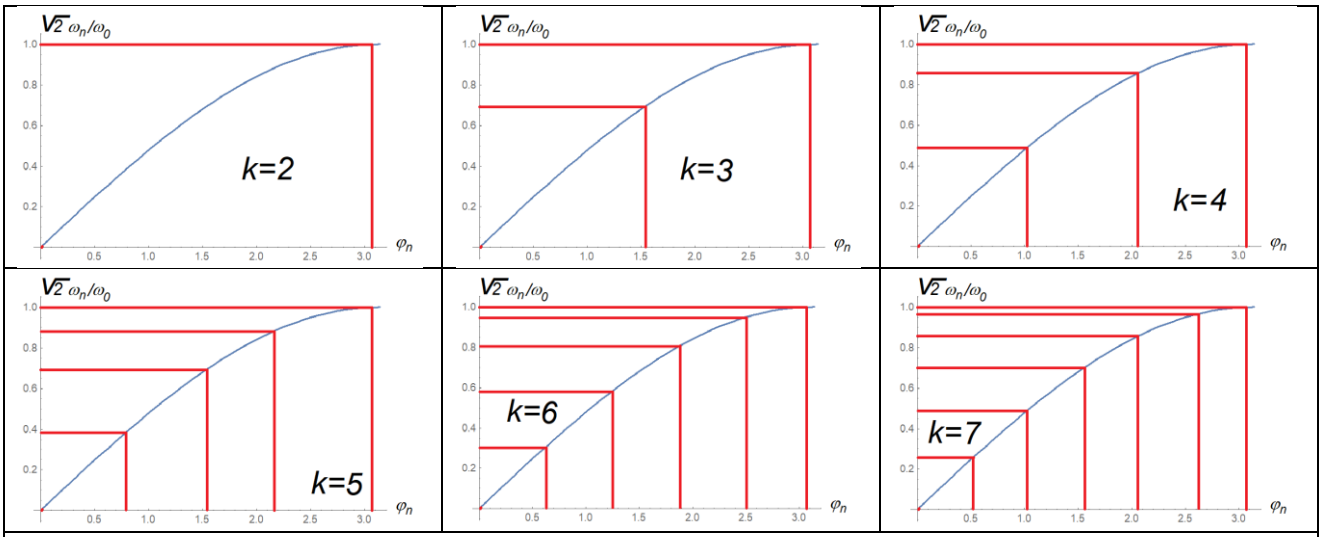


Рис. 6.2.9. Набіги фаз на одну ланку та відповідні власні частоти для ланцюжків з різною кількістю ланок

Контрольні питання до розділу 6.2

1. Що таке ланцюжкова система? Наведіть приклади.
2. Коли поширення коливання в ланцюжковій системі можна розглядати як хвильовий процес?
3. Чому ланцюжкові системи можуть служити частотними фільтрами?
4. Чому в ланцюжкових системах можуть існувати повільні хвилі?
5. Як у ланцюжку зв'язаних осциляторів виникають смуги прозорості?
6. Які механізми виникнення дисперсії вам відомі?

Приклади розв'язання задач до підрозділу 6.2

П 6.2.1. Знайдіть граничні частоти та побудуйте дисперсійні криві для нескінченної ланцюжкової системи з однакових зв'язаних контурів, показаної на рис. 6.2.10.

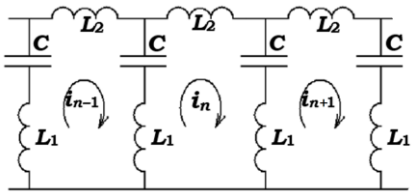


Рис. 6.2.10

Розв'язок

Запишемо рівняння за другим правилом Кірхгофа для ланки з номером *n*:

$$\left(2L_1+L_2\right)\frac{di_n}{dt}+\frac{2}{C}\int i_n dt-L_1\left(\frac{di_{n-1}}{dt}+\frac{di_{n+1}}{dt}\right)-\frac{1}{C}\left(\int i_{n-1}dt+\int i_{n+1}dt\right)=0,$$

або

$$\left(2L_1+L_2\right)\frac{d^2i_n}{dt^2}+\frac{2}{C}i_n t=L_1\left(\frac{d^2i_{n-1}}{dt^2}+\frac{d^2i_{n+1}}{dt^2}\right)+\frac{1}{C}\left(i_{n-1}+i_{n+1}\right).$$

Вважатимемо, що за відсутності дисипації струми в сусідніх ланках відрізняються лише сталим набігом фази, тобто

$i_k=I_m\exp\left(i\omega t-ik\varphi\right)$.

Підставивши цей розв'язок до попереднього рівняння і скоротивши, отримаємо:

$$-\omega^2\left(2L_1+L_2\right)+\frac{2}{C}=-\omega^2L_1\left[\exp\left(i\varphi\right)+\exp\left(-i\varphi\right)\right]+\frac{1}{C}\left[\exp\left(i\varphi\right)+\exp\left(-i\varphi\right)\right],$$

або

$$-\omega^2\left(1+\lambda\right)+\omega_0^2=\left(-\omega^2+\omega_0^2\right)\cos\varphi,\quad \frac{L_2}{2L_1}=\lambda,\quad \frac{1}{L_1C}=\omega_0^2.$$

Звідси

$$\omega^2=\frac{\omega_0^2\left(1-\cos\varphi\right)}{1+\lambda-\cos\varphi}=\omega_0^2\frac{\sin^2\frac{\varphi}{2}}{\frac{\lambda}{2}+\sin^2\frac{\varphi}{2}}.$$

Отже,

$$\omega_{\min}=0,\quad \omega_{\max}=\frac{2\omega_0^2}{\lambda+2}.$$

Нехай довжина однієї ланки – *a*. Тоді від набігу фази на одну ланку можна перейти до хвильового числа k $ka=\varphi$, і закон дисперсії хвиль у ланцюжку набуває вигляду

$$\omega^2==\omega_0^2\frac{\sin^2\frac{ka}{2}}{\frac{\lambda}{2}+\sin^2\frac{ka}{2}}.$$

Графік цієї залежності поданий на рис. 6.2.11.

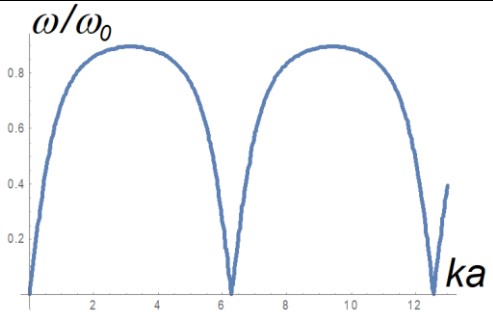


Рис. 6.2.11

Задачі до підрозділу 6.2

6.2.1. Знайдіть граничні частоти та побудуйте дисперсійні криві для нескінченної ланцюжкової системи з однакових зв'язаних контурів, показаної на рис. 6.2.12:

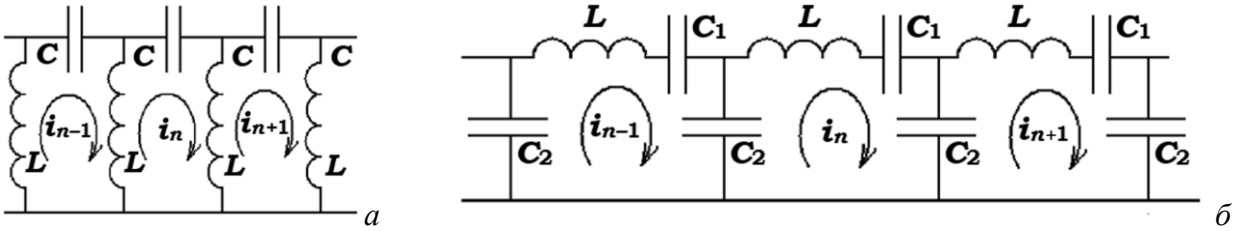


Рис. 6.2.12

6.2.2. Розрахуйте дисперсію хвиль у ланцюжку, складеному з однакових пружин і мас (рис. 6.2.1 а).

6.3. Розв'язок телеграфних рівнянь для двопровідної лінії

Розглянемо властивості хвиль у двопровідній лінії більш детально, маючи на увазі, що така лінія в багатьох випадках може розглядатися як найпростіша модель згаданих раніше ліній передач мікрохвильового та оптичного діапазонів (хвилевод, коаксіальний кабель, смужкова або мікросмужкова лінія, оптоволокно тощо).

6.3.1. Розв'язок телеграфних рівнянь для лінії без втрат

Як було показано в п. 6.1.1, процеси в двопровідній лінії описуються системою телеграфних рівнянь (6.1.4)

$$\frac{\partial u}{\partial x} = -\left(L_1 \frac{\partial}{\partial t} + r_1 \right) i \,;$$

$$\frac{\partial i}{\partial x} = -\left(C_1 \frac{\partial}{\partial t} + g_1 \right) u \,,$$

де L_1, C_1, r_1, g_1 – погонні параметри лінії – погонна індуктивність, погонна ємність та ін.). Телеграфні рівняння можуть бути зведені до одного хвильового рівняння вигляду (6.1.5)

$$\frac{\partial^2 u}{\partial x^2} - \left(L_1 \frac{\partial}{\partial t} + r_1 \right)\left(C_1 \frac{\partial}{\partial t} + g_1 \right) u = 0 \,,$$

або, якщо знехтувати дисипацією, до рівняння (6.1.6)

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c_0^2} \frac{\partial^2 u}{\partial t^2} = 0 \,,$$

де $c_0=(L_1C_1)^{-1/2}$. Рівняння, аналогічні (6.1.5)-(6.1.6), можна записати й для струму I .

Розв'язок рівняння у формі гармонічних хвиль (6.1.8) має вигляд суми прямої та зворотної хвиль (як і раніше, користуємося методом комплексних амплітуд):

$$u(x,t) = u_{m1} \exp(i\omega t - ikx) + u_{m2} \exp(i\omega t + ikx), \qquad (6.3.1)$$

де u_{m1} та u_{m2} – амплітуди прямої та зворотної хвиль. Частота ω і хвильове число k пов'язані співвідношенням $\omega = c_0 k$, яке випливає з дисперсійного рівняння (6.1.9)

$$\omega^2 = c_0^2 k^2 \,,$$

що його отримуємо, підставивши розв'язок (6.3.1) до рівняння (6.1.6).

Підставивши (6.3.1) до другого з телеграфних рівнянь (без дисипативного доданку), отримуємо:

$$\frac{\partial i}{\partial x} = -C_1 \frac{\partial u}{\partial t} = -i\omega C_1 \left[u_{m1} \exp(i\omega t - ikx) + u_{m2} \exp(i\omega t + ikx) \right],$$

звідки

$$i(x,t) = \frac{\omega}{k} C_1 \left[u_{m1} \exp(i\omega t - ikx) - u_{m2} \exp(i\omega t + ikx) \right] =$$

$$= \sqrt{\frac{C_1}{L_1}} \left[u_{m1} \exp(i\omega t - ikx) - u_{m2} \exp(i\omega t + ikx) \right] \equiv \qquad (6.3.2)$$

$$\equiv i_{m1} \exp(i\omega t - ikx) + i_{m2} \exp(i\omega t + ikx),$$

де

$$i_{m1,m2} = \pm \frac{u_{m1,m2}}{Z_0} \,, \qquad Z_0 = \sqrt{\frac{L_1}{C_1}} \,, \qquad (6.3.3)$$

Z_0 – хвильовий опір лінії.

6.3.2. Розв'язок телеграфних рівнянь для лінії з втратами

З урахуванням втрат дисперсійне рівняння, що випливає з (6.1.5), набуває вигляду

$$k^2 - (\omega L_1 - i r_1)(\omega C_1 - i g_1) = 0 \,. \qquad (6.3.4)$$

Рівняння (6.3.4) можна розв'язувати двома способами.

Якщо вважати хвильове число k заданою дійсною величиною, з рівняння (6.3.4) можна отримати комплексну частоту ω. Це відповідає так званій задачі з початковими умовами, коли в початковий момент часу в лінії існує гармонічне збурення з просторовим періодом $2\pi/k$, а комплексна частота описує подальшу еволюцію цього збурення в часі.

Якщо, навпаки, вважати частоту ω заданою дійсною величиною, з рівняння (6.3.4) можна отримати комплексне хвильове число k. Такий підхід відповідає задачі з граничними умовами, коли на початку лінії передачі стоїть генератор гармонічного сигналу з частотою ω, а комплексне хвильове число описує зміну цього сигналу в просторі. Задача з граничними умовами є більш реалістичною, тому саме її ми будемо надалі розглядати.

Отже, з дисперсійного рівняння маємо:

$$k = \pm\omega\sqrt{L_1C_1}\sqrt{\left(1-\frac{ir_1}{\omega L_1}\right)\left(1-\frac{ig_1}{\omega C_1}\right)} \approx$$

$$\approx \pm\omega\sqrt{L_1C_1}\left[1-\frac{i}{2}\left(\frac{r_1}{\omega L_1}+\frac{g_1}{\omega C_1}\right)\right] = \pm\left(k'-ik''\right),$$

де

$$k' = \frac{\omega}{c}, \qquad k'' = \frac{\omega}{2k'}\left(L_1g_1+C_1r_1\right) = \frac{\sqrt{L_1C_1}}{2}\left(\frac{g_1}{C_1}+\frac{r_1}{L_1}\right) \tag{6.3.5}$$

(дисипацію в лінії вважаємо слабкою, $r_1 \ll \omega L_1$, $g_1 \ll \omega C_1$, тому $|k'| \gg |k''|$).

Таким чином, розв'язок хвильового рівняння для слабкодисипативної лінії набуває вигляду

$$u(x,t) = u_{m1}\exp\left(i\omega t - ik'x\right)\exp\left(-k''x\right) + u_{m2}\exp\left(i\omega t + ik'x\right)\exp\left(k''x\right). \tag{6.3.6}$$

Бачимо, що пряма й зворотна хвилі згасають у напрямку свого поширення.

Аналогічно до (6.3.6) можна записати розв'язок для струму:

$$i(x,t) = i_{m1}\exp\left(i\omega t - ik'x\right)\exp\left(-k''x\right) + i_{m2}\exp\left(i\omega t + ik'x\right)\exp\left(k''x\right). \tag{6.3.7}$$

Підставивши розв'язки (6.3.6)-(6.3.7) до одного з телеграфних рівнянь, отримуємо зв'язок між амплітудами прямої та зворотної хвиль у формі:

$$\mp iku_{m1,m2} = -\left(i\omega L_1 + r_1\right)i_{m1,m2},$$

або

$$\frac{u_{m1,m2}}{i_{m1,m2}} = \pm\frac{\omega L_1 - ir_1}{k} = \pm\frac{\omega L_1 - ir_1}{\sqrt{\left(\omega L_1 - ir_1\right)\left(\omega C_1 - ig_1\right)}} = \pm Z_0\sqrt{\frac{1-ir_1/\omega L_1}{1-ig_1/\omega C_1}}. \tag{6.3.8}$$

Величина

$$Z \equiv Z_0\sqrt{\frac{1-ir_1/\omega L_1}{1-ig_1/\omega C_1}} = |Z|\exp\left(i\varphi\right) \approx Z_0\left[1-\frac{i}{2\omega}\left(\frac{r_1}{L_1}-\frac{g_1}{C_1}\right)\right], \tag{6.3.9}$$

де

$$|Z| = Z_0\sqrt[4]{\frac{1+\left(r_1/\omega L_1\right)^2}{1+\left(g_1/\omega C_1\right)^2}}, \qquad \varphi = \frac{1}{2}\left[arctg\left(\frac{g_1}{\omega C_1}\right) - arctg\left(\frac{r_1}{\omega L_1}\right)\right], \tag{6.3.10}$$

– це хвильовий опір лінії з дисипацією. Величина φ – це зсув фаз між хвилями напруги й струму, зумовлений дисипацією.

6.3.3. Погонні параметри двопровідної лінії та її фізичні властивості

Найпростіша модель двопровідної лінії – це два паралельні провідники циліндричної форми. Переріз такої лінії поданий на рис. 6.3.9 а, а розподіл електричного та магнітного полів при протіканні струму – на рис. 6.3.9 б.

Користуючись рівняннями електродинаміки, для лінії, поданої на рис. 6.3.9 а, можна

розрахувати погонні ємність та індуктивність:

$$C_1 = \frac{\pi\varepsilon}{\ln\left(\frac{d}{2a} + \sqrt{\left(\frac{d}{2a}\right)^2 - 1}\right)}, \tag{6.3.11}$$

$$L_1 = \frac{\mu}{\pi}\ln\left(\frac{d}{2a} + \sqrt{\left(\frac{d}{2a}\right)^2 - 1}\right), \tag{6.3.12}$$

де ε та μ – відповідно діелектрична та магнітна провідність середовища (в системі одиниць СІ). Видно, що єдиним параметром лінії, від якого залежать ці величини, є відношення d/a. Зі зростанням цього відношення погонна ємність зменшується, а погонна індуктивність – зростає.

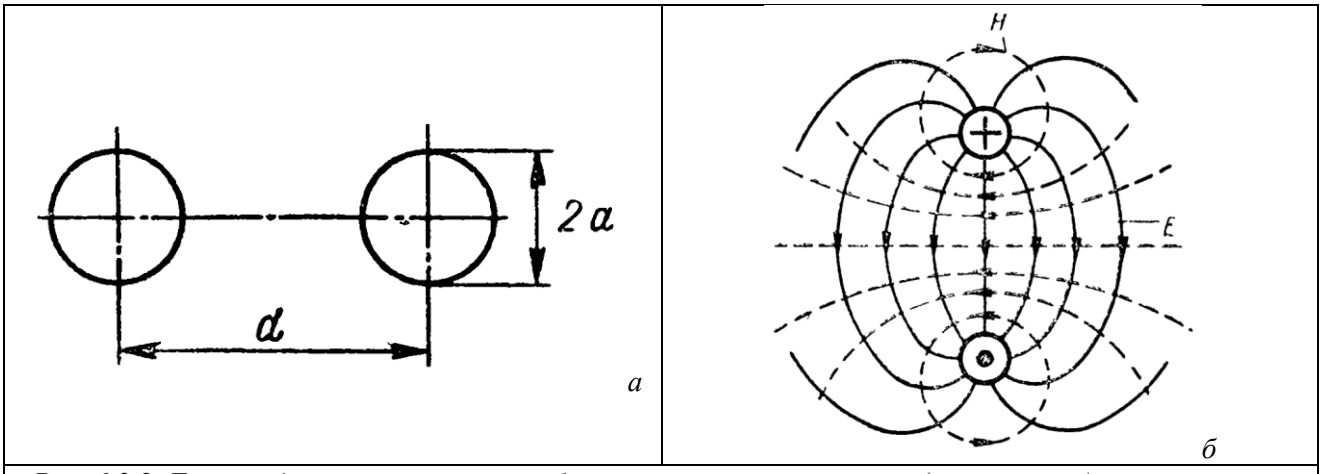


Рис. 6.3.9. Двопровідна лінія: а – переріз, б – силові лінії електричного (суцільні лінії) та магнітного (штрихові лінії) полів

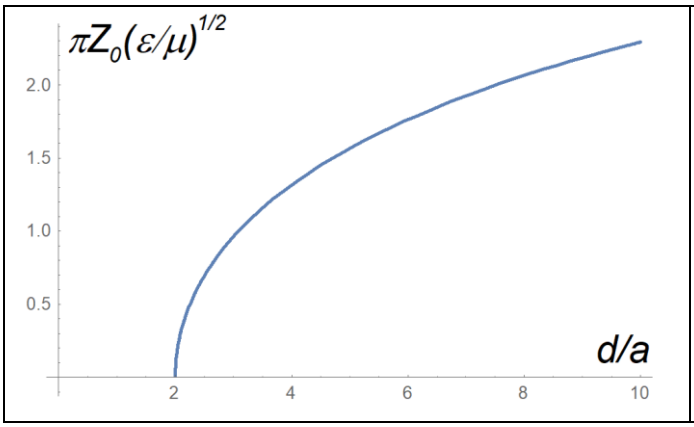


Рис. 6.3.10. Залежність нормованого хвильового опору двопровідної лінії від її розмірів

Знаючи погонні індуктивність та ємність, неважко розрахувати хвильовий опір та фазову швидкість хвиль у двопровідній лінії:

$$Z_0 = \sqrt{\frac{L_1}{C_1}} = \frac{1}{\pi}\sqrt{\frac{\mu}{\varepsilon}}\ln\left(\frac{d}{2a} + \sqrt{\left(\frac{d}{2a}\right)^2 - 1}\right), \tag{6.3.13}$$

$$c_0 = \frac{1}{\sqrt{L_1 C_1}} = \frac{1}{\sqrt{\varepsilon\mu}}. \tag{6.3.14}$$

Залежність хвильового опору лінії від її розмірів показана на рис. 6.3.10.

Якщо двопровідна лінія знаходиться у вакуумі, тобто $\varepsilon = \varepsilon_0$, $\mu = \mu_0$, то c_0 збігається

зі швидкістю електромагнітних хвиль у вакуумі, а $\sqrt{\mu/\varepsilon} = 120\pi\left[Ом\right]$. Тоді, за умови, що $d/a \gg 1$, справедлива наближена числова формула

$$Z_0\left[Ом\right] \approx 120\ln\left(\frac{d}{a}\right) = 276\lg\left(\frac{d}{a}\right). \qquad (6.3.15)$$

Для двопровідної лінії у вакуумі можна вважати, що $g_1 \approx 0$, і втрати в лінії визначаються її погонним опором. Останній можна розрахувати за формулою

$$r_1 \approx \frac{1}{\pi a}\sqrt{\frac{\omega\mu_0}{2\sigma}}, \qquad (6.3.16)$$

де σ – питома електропровідність матеріалу лінії.

Контрольні питання до підрозділу 6.3

1. Як втрати в лінії вливають на поширення сигналу в ній?
2. Що таке задача з початковими умовами й задача з граничними умовами в теорії хвиль?
3. Що таке хвильовий опір двопровідної лінії?
4. Від чого залежать погонні параметри двопровідної лінії?
5. Який вид втрат відіграє основну роль у двопровідній лінії?

Задачі до підрозділу 6.3

6.3.1. Двопровідна лінія складається з двох паралельних дротів радіусом 1мм, розташованих на віддалі 1см. Знайти погонні ємність та індуктивність такої лінії.

6.3.2. Знайдіть хвильовий опір лінії, описаної в задачі 6.3.1. Порівняйте його з погонним опором лінії на частоті 100 МГц. Дроти вважати мідними. Питомий опір міді – 0.0175 Ом·мм^2/м.

6.3.3. Знайдіть швидкість поширення хвиль у лінії, описаній в задачі 6.3.1. Якою буде довжина хвилі в цій лінії на частоті 50 МГц?

6.3.4. На якій відстані від генератора з частотою 50 МГц амплітуда хвилі в лінії, описаній у задачах 6.3.1-6.3.2, зменшиться вдвічі?

6.4. Режими навантаженої лінії передачі

Характер хвиль у лінії передачі залежить від опору навантаження, який увімкнений на її віддаленому кінці, а також від джерела сигналу, увімкненого на її початку. Обговоримо ці питання більш детально.

6.4.1. Відбиття хвиль від навантаження

Розв'язки телеграфних рівнянь (6.3.1)-(6.3.2) або (6.3.6)-(6.3.7) з урахуванням відповідно співвідношень (6.3.3) або (6.3.8) містять невідомі амплітуди прямої та зворотної хвиль $u_{m1,m2}$. Ці величини знаходяться з граничних умов.

Вважатимемо, що на вході лінії з хвильовим опором Z (див. формулу (6.3.9)) довжиною l увімкнене джерело змінної напруги з частотою ω, а на виході – комплексний опір Z_2, який відіграє роль навантаження (рис. 6.4.1). Тоді граничні умови при $x = l$ мають вигляд:

$$u(x=l,t) = Z_2 i(x=l,t), \tag{6.4.1}$$

або

$$u_{m1}\exp(i\omega t - ik'l)\exp(-k''l) + u_{m2}\exp(i\omega t + ik'l)\exp(k''l) =$$

$$= \frac{Z_2}{Z}\left[u_{m1}\exp(i\omega t - ik'l)\exp(-k''l) - u_{m2}\exp(i\omega t + ik'l)\exp(k''l)\right]$$

(враховано формули (6.3.6)-(6.3.10)). Звідси знаходимо амплітуду відбитої хвилі через амплітуду падаючої:

$$R \equiv \frac{u_{m2}}{u_{m1}} = \frac{Z_2 - Z}{Z_2 + Z}\exp(-i2k'l)\exp(-2k''l), \tag{6.4.2}$$

де R (від англ. reflection – відбиття) – коефіцієнт відбиття (за напругою). Цей коефіцієнт обертається на нуль за виконання умови узгодження $Z_2 = Z$. За її виконання відбиття від навантаження немає, і в лінії присутня лише пряма хвиля, що біжить від генератора до навантаження.

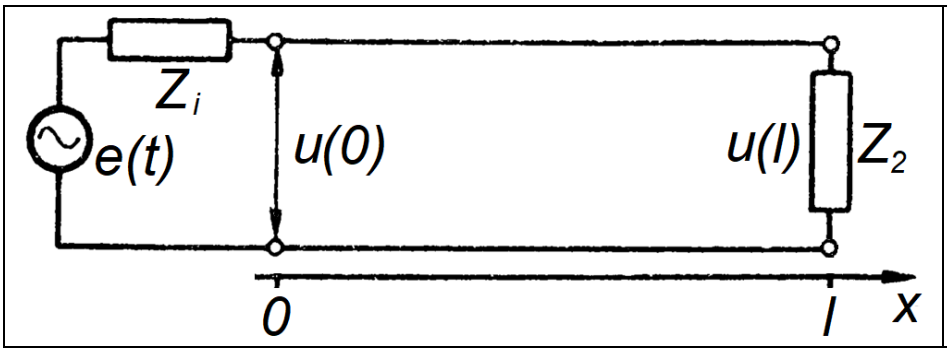


Рис. 6.4.1. Лінія передачі з джерелом сигналу та навантаженням

У режимі короткого замикання на виході ($Z_2 = 0$)

$$R = -\exp(-i2k'l)\exp(-2k''l),$$

у ненавантаженому режимі ($Z_2 \to \infty$) –

$$R = \exp(-i2k'l)\exp(-2k''l).$$

Для лінії без втрат формула (6.4.2) спрощується:

$$R = \frac{Z_2 - Z_0}{Z_2 + Z_0}\exp(-i2kl). \tag{6.4.3}$$

Відповідно в режимі короткого замкнення на виході та в ненавантаженому режимі

$$R = \mp\exp(-i2kl),$$

тобто в обох цих випадках $|R| = 1$. Це означає, що в лінії без втрат амплітуди прямої та відбитої хвиль однакові, а їх суперпозиція дає стоячу хвилю, що характеризується стаціонарним

розподілом амплітуди коливань у просторі (при цьому сама амплітуда змінюється в часі).

6.4.2. Режим біжучих хвиль у лінії

В загальному випадку напруга і струм у лінії з втратами можуть бути подані у вигляді:

$$u(x,t)=u_{m1}\exp(i\omega t)\left[\exp(-ik'x)\exp(-k''x)+R\exp(ik'x)\exp(k''x)\right]=$$

$$=u_{m1}\exp(i\omega t)\left[\exp(-ik'x)\exp(-k''x)+\frac{Z_2-Z}{Z_2+Z}\exp(-i2k'l)\exp(-2k''l)\exp(ik'x)\exp(k''x)\right]=$$

$$=u_{m1}\exp(i\omega t-ik'l-k''l)\left\{\exp\left[-i(k'-ik'')(x-l)\right]+\frac{Z_2-Z}{Z_2+Z}\exp i(k'-ik'')(x-l)\right\},$$

(6.4.4)

$$i(x,t)=\frac{u_{m1}}{Z}\exp(i\omega t-ik'l-k''l)\left\{\exp\left[-i(k'-ik'')(x-l)\right]-\frac{Z_2-Z}{Z_2+Z}\exp\left\{i(k'-ik'')(x-l)\right\}\right\}.$$

(6.4.5)

Для лінії без втрат $Z=Z_0$, і формули (6.4.4)-(6.4.5) набувають вигляду

$$u(x,t)=u_{m1}\exp(i\omega t-ikl)\left\{\exp\left[-ik(x-l)\right]+\frac{Z_2-Z_0}{Z_2+Z_0}\exp\left[ik(x-l)\right]\right\},\qquad(6.4.6)$$

$$i(x,t)=\frac{u_{m1}}{Z_0}\exp(i\omega t-ikl)\left\{\exp\left[-ik(x-l)\right]-\frac{Z_2-Z_0}{Z_2+Z_0}\exp\left[ik(x-l)\right]\right\}\qquad(6.4.7)$$

В узгодженому режимі ($Z_2=Z_0$) маємо:

$$u(x,t)=u_{m1}\exp(i\omega t-ikx),\qquad(6.4.8)$$

$$i(x,t)=\frac{u_{m1}}{Z_0}\exp(i\omega t-ikx),\qquad(6.4.9)$$

тобто в лінії, як уже вказувалося, існує тільки пряма хвиля, що біжить уздовж додатного напрямку осі x (рис. 6.4.2).

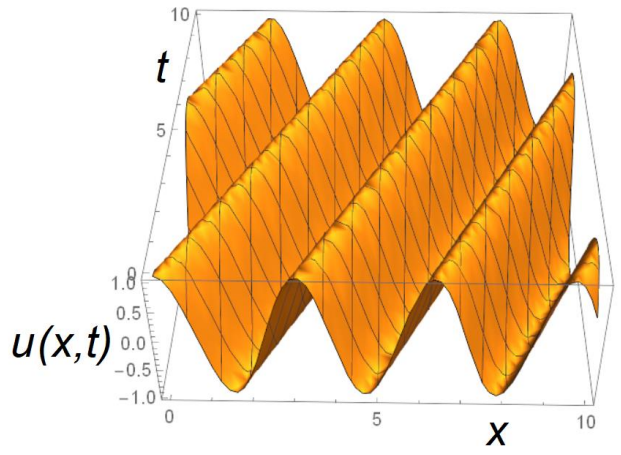


Рис. 6.4.2. Біжуча хвиля в лінії без втрат

Врахування втрат у лінії приводить до того, що амплітуда хвилі згасає в напрямку її поширення, однак за виконання умови узгодження режим біжучих хвиль у лінії зберігається. Справді, для цього випадку формули (6.4.4)-(6.4.5) набувають вигляду:

$$u(x,t)=u_{m1}\exp(i\omega t-ik'x)\exp(-k''x),\qquad(6.4.10)$$

$$i(x,t)=\frac{u_{m1}}{Z}\exp(i\omega t-ik'x)\exp(-k''x).\qquad(6.4.11)$$

6.4.3. Режими стоячих хвиль

Обмежимося розглядом лінії без втрат, хвилі в якій описуються формулами (6.4.6)-(6.4.7).

Короткозамкнений вихід. Розглянемо спершу випадок, коли вихід лінії замкнений накоротко, тобто $Z_2 = 0$, $R = -\exp(-i2kl)$. Маємо:

$$u(x,t) = u_{m1}\exp(i\omega t - ikl)\left\{\exp\left[-ik(x-l)\right] - \exp\left[ik(x-l)\right]\right\} = \\ = 2iu_{m1}\exp(i\omega t - ikl)\sin\left[ik(x-l)\right], \tag{6.4.12}$$

$$i(x,t) = \frac{u_{m1}}{Z_0}\exp(i\omega t - ikl)\left\{\exp\left[-ik(x-l)\right] + \exp\left[ik(x-l)\right]\right\} = \\ = \frac{2u_{m1}}{Z_0}\exp(i\omega t - ikl)\cos\left[k(x-l)\right]. \tag{6.4.13}$$

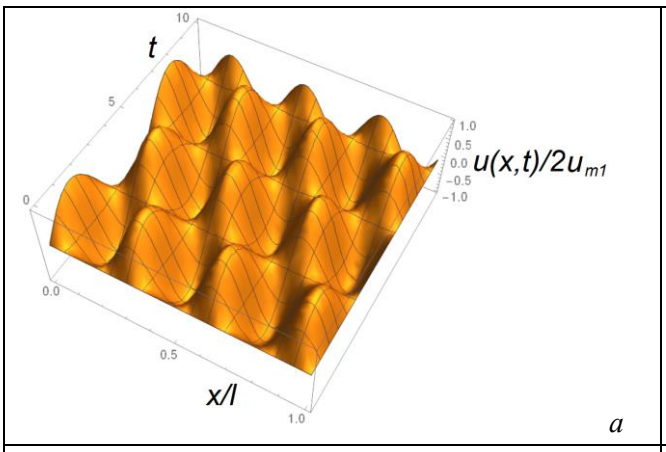
Відповідно до методу комплексних амплітуд для отримання вимірюваних величин слід узяти дійсні частини комплексних функцій. Вважаючи амплітуду U_1 дійсною, отримаємо:

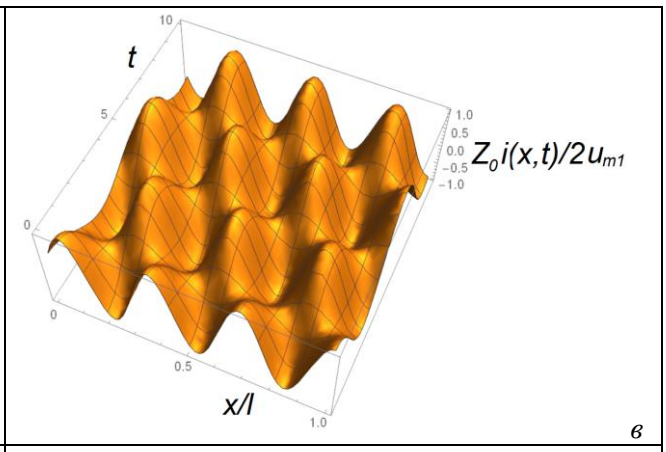
$$\mathrm{Re}\left\{u(x,t)\right\} = -2u_{m1}\sin(\omega t - kl)\sin\left[k(x-l)\right], \tag{6.4.14}$$

$$\mathrm{Re}\left\{i(x,t)\right\} = \frac{2u_{m1}}{Z_0}\cos(\omega t - kl)\cos\left[k(x-l)\right]. \tag{6.4.15}$$

Розв'язки (6.4.12)-(6.4.13) або (6.4.14)-(6.4.15) відповідають стоячим хвилям.

Графіки залежностей (6.4.14)-(6.4.15) подані на рис. 6.4.4 а, б. Видно, що положення максимумів (пучностей) і нулів (вузлів) стоячої хвилі в просторі не змінюється з часом, а коливання в кожній точці простору є гармонічними, причому часові коливання напруги та струму зсунуті на чверть періоду. На виході лінії (при $x = l$) амплітуда коливань напруги дорівнює нулеві, а амплітуда коливань струму, навпаки, досягає максимуму.





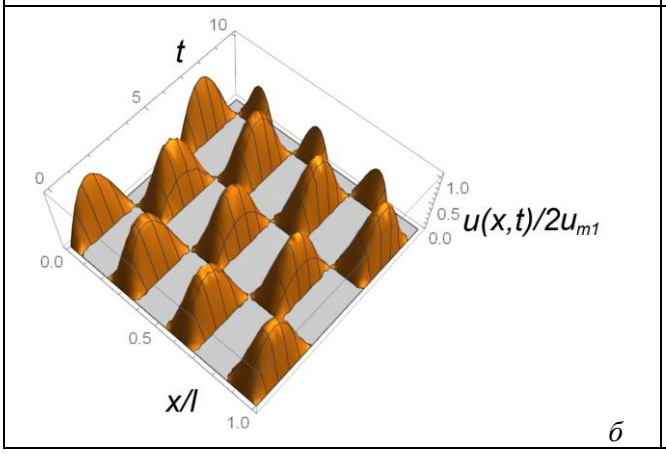


Рис. 6.4.4. Стоячі хвилі напруги (а, б) та струму (в) в лінії з короткозамкненим виходом. Для кращого розуміння показано весь діапазон значень напруги (а) та його додатну частину (б)

Ненавантажений режим. Розглянемо тепер випадок, коли вихід лінії розімкнений. У

цьому випадку $Z_2 \to \infty$, $R = \exp(-i2kl)$. Маємо:

$$u(x,t) = u_{m1}\exp(i\omega t - ikl)\{\exp[-ik(x-l)] + \exp[ik(x-l)]\} = \\ = 2u_{m1}\exp(i\omega t - ikl)\cos[ik(x-l)], \tag{6.4.16}$$

$$i(x,t) = \frac{u_{m1}}{Z_0}\exp(i\omega t - ikl)\{\exp[-ik(x-l)] - \exp[ik(x-l)]\} = \\ = i\frac{2u_{m1}}{Z_0}\exp(i\omega t - ikl)\sin[k(x-l)]. \tag{6.4.17}$$

Відповідні вимірювані величини:

$$u(x,t) = 2u_{m1}\cos(i\omega t - ikl)\cos[ik(x-l)], \tag{6.4.18}$$

$$i(x,t) = -\frac{2u_{m1}}{Z_0}\sin(i\omega t - ikl)\sin[k(x-l)]. \tag{6.4.19}$$

Розв'язки (6.4.16)-(6.4.17) або (6.4.18)-(6.4.19) також відповідають стоячим хвилям. Просторово-часовому розподілу напруги тепер відповідає рис. 6.4.4 б, струму – рис. 6.4.4 а. Іншими словами, при $x = l$ струм завжди дорівнює нулю, а напруга коливається з максимальною амплітудою.

<u>Режим з реактивним навантаженням</u>. Нарешті, розглянемо випадок, коли навантаженням лінії без втрат служить реактивний опір, тобто $Z_2 = iX_2$. У цьому випадку

$$R = \frac{iX_2 - Z_0}{iX_2 + Z_0}\exp(-i2kl) = -\exp(-i2kl - i2\Phi), \qquad \Phi = arctg\frac{X_2}{Z_0}, \tag{6.4.20}$$

тобто $|R| = 1$. Отже, в цьому випадку також реалізується режим стоячих хвиль. Маємо:

$$u(x,t) = u_{m1}\exp(i\omega t)[\exp(-ikx) + R\exp(ikx)] = \\ = u_{m1}\exp(i\omega t)[\exp(-ikx) - \exp(-i2kl - i2\Phi)\exp(ikx)] = \\ = u_{m1}\exp(i\omega t - ikl - i\Phi)\{\exp[-ik(x-l) + i\Phi] - \exp[ik(x-l) - i\Phi]\} = \\ = -2iu_{m1}\exp(i\omega t - ikl - i\Phi)\sin[k(x-l) - \Phi], \tag{6.4.21}$$

$$i(x,t) = \frac{u_{m1}}{Z}\exp(i\omega t)[\exp(-ikx) - R\exp(ikx)] = \\ = \frac{u_{m1}}{Z}\exp(i\omega t)[\exp(-ikx) + \exp(-i2kl - i2\Phi)\exp(ikx)] = \\ \frac{u_{m1}}{Z}\exp(i\omega t - ikl - i\Phi)\{\exp[-ik(x-l) + i\Phi] + \exp[ik(x-l) - i\Phi]\} \\ = \frac{2u_{m1}}{Z}\exp(i\omega t - ikl - i\Phi)\cos[k(x-l) - \Phi]. \tag{6.4.22}$$

Вимірювані величини:

$$\mathrm{Re}\{u(x,t)\} = 2u_{m1}\sin(\omega t - kl - \Phi)\sin[k(x-l) - \Phi], \tag{6.4.23}$$

$$\mathrm{Re}\{i(x,t)\} = \frac{2u_{m1}}{Z}\cos(\omega t - kl - \Phi)\cos[k(x-l) - \Phi]. \tag{6.4.24}$$

І в цьому випадку коливання напруги та струму як в часі, так і в просторі зсунуті на чверть періоду. На відміну від попередніх випадків, в точці $x = l$ немає ні пучностей, ні вузлів напруги чи струму.

Якщо обмежитися розглядом лінії без втрат і ввести величину

$$l_{eff} = l + \frac{\Phi}{k} = l + \frac{1}{k}arctg\frac{X_2}{Z_0} = l + \frac{\lambda}{2\pi}arctg\frac{X_2}{Z_0}, \tag{6.4.25}$$

то формули (6.23)-(6.24) можна для цього випадку переписати у формі

$$\mathrm{Re}\{u(x,t)\} = 2u_{m1}\sin(\omega t - kl_{eff})\sin[k(x - l_{eff})], \qquad (6.4.23\ \text{а})$$

$$\mathrm{Re}\{i(x,t)\} = \frac{2u_{m1}}{Z_0}\cos(\omega t - kl_{eff})\cos[k(x - l_{eff})]. \qquad (6.4.24\ \text{а})$$

Якщо у формулах (6.4.14)-(6.4.15), які описують лінію з короткозамкненим кінцем, формально замінити l на l_{eff}, то вони збігатимуться з формулами (6.4.23 а)-(6.4.24 а). Іншими словами, наявність реактивного навантаження лінії можна інтерпретувати як зміну її ефективної довжини порівняно з лінією, кінець якої замкнений накоротко. Ефективна довжина лінії зростає, якщо навантаження носить індуктивний характер ($X_2 > 0$), і зменшується, якщо характер навантаження ємнісний ($X_2 < 0$). З такої точки зору розглянуту вище лінію з розімкнутим кінцем ($X_2 \to \infty$) можна розглядати як короткозамкнену з ефективною довжиною $l_{eff} = l + \lambda/4$.

Наостанок зазначимо, що режим стоячих хвиль, строго кажучи, можливий лише в лінії без втрат. Наявність втрат, як уже зазначалося, приводить до згасання хвиль вздовж напрямку їхнього поширення (п. 6.3.2). Таким чином, навіть за умови $|R| = 1$ однакові амплітуди прямої та зворотної хвиль, необхідні для утворення стоячої хвилі, будуть лише поблизу точки відбиття (рис. 6.4.5-6.4.6). Якщо на довжині лінії хвиля помітно згасає, то біля її початку режим буде близький до біжучої хвилі.

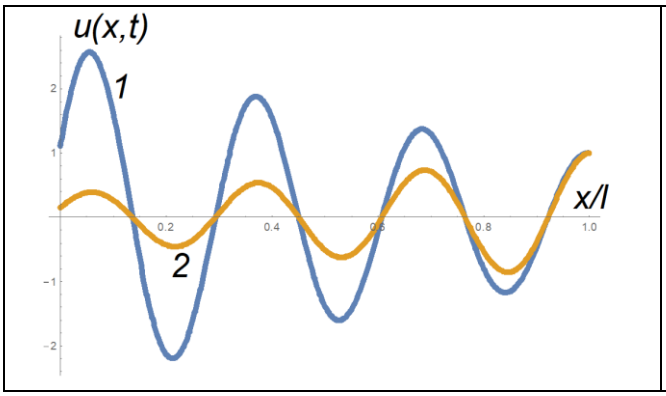
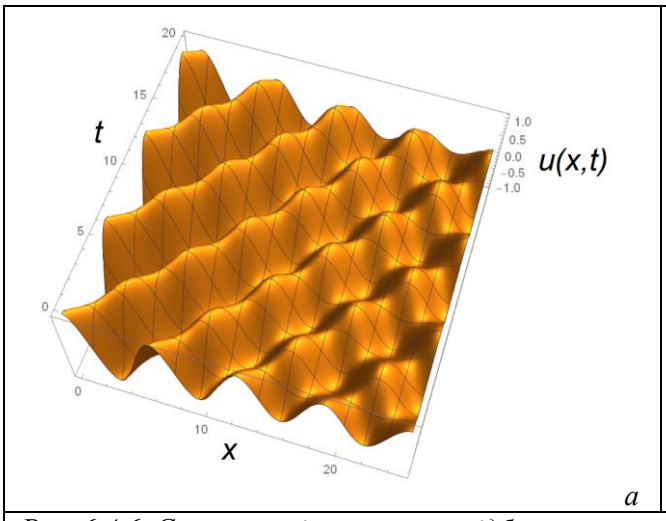


Рис. 6.4.5. Пряма (1) та зворотна (2) хвилі в ненавантаженій лінії з втратами



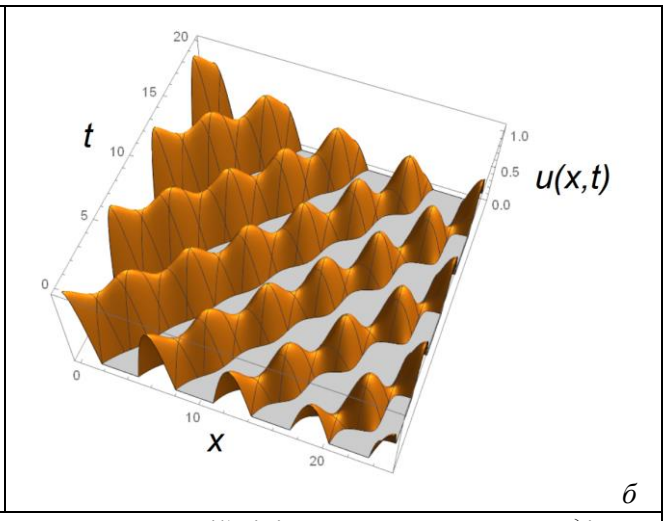


Рис. 6.4.6. Суперпозиція прямої та відбитої хвиль у ненавантаженій лінії з втратами: а – весь діапазон значень напруги, б – область додатних значень напруги. Напруга нормована на амплітуду падаючої хвилі в точці x=0; довжина лінії – kl=25

6.4.4. Мішаний режим

В загальному випадку в лінії одночасно існують пряма та зворотна хвилі з неоднаковими амплітудами. Наприклад, розподіл напруги вздовж лінії може бути поданий у формі

(6.3.1),

$$u(x,t) = u_{m1}\exp(i\omega t - ikx) + u_{m2}\exp(i\omega t + ikx),$$

де u_{m1} та u_{m2} – комплексні амплітуди прямої та зустрічної хвиль (між ними може бути зсув за фазою). Нехай

$$u_{m2} = |u_{m2}|\exp(i\varphi_0).$$

Тоді вимірювана напруга вздовж лінії може бути подана у вигляді

$$\mathrm{Re}\{u(x,t)\} = u_{m1}\cos(\omega t - kx) + |u_{m2}|\cos(\omega t + kx + \varphi_0) \qquad (6.4.26)$$

(рис. 6.4.7). Розподіл дещо нагадує стоячу хвилю, оскільки положення максимумів і мінімумів напруги в просторі залишаються незмінними. Але з рис. 6.4.7 б видно, що в мішаному режимі максимальна напруга не спадає до нуля (на відміну від того, як це має місце в режимі стоячих хвиль). Справді, максимальна амплітуда коливань $u_{\max} = u_{m1} + |u_{m2}|$ досягається, коли пряма й зворотна напруги додаються у фазі, а мінімальна $u_{\min} = u_{m1} - |u_{m2}|$ – коли вони додаються в протифазі (припускаємо для визначеності, що $u_{m1} > |u_{m2}|$).

Для кількісної характеристики співвідношення між прямою і зустрічною хвилями вводять поняття коефіцієнта біжучої хвилі (к.б.х.) K_t (індекс від англ. travelling wave – біжуча хвиля):

$$K_t = \frac{u_{\min}}{u_{\max}} = \frac{u_{m1} - |u_{m2}|}{u_{m1} + |u_{m2}|}. \qquad (6.4.27)$$

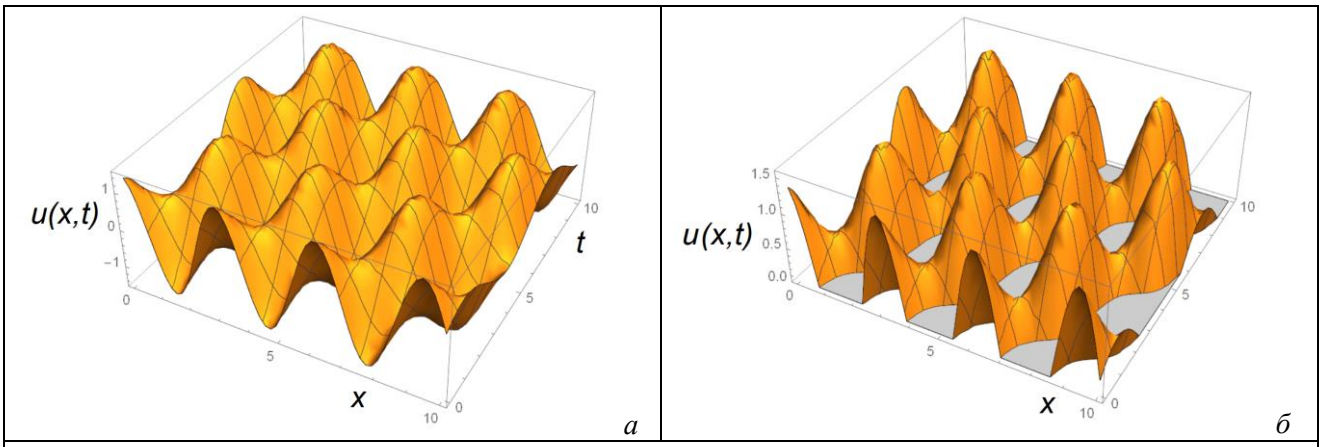


Рис. 6.4.7. Просторово-часовий розподіл напруги в мішаному режимі ($U_2 = 0.5U_1$, $\varphi_0 = \pi/2$):
а – весь діапазон значень напруги, б – область додатних значень напруги. Напруга нормована на амплітуду падаючої хвилі в точці x=0; довжина лінії – kl=15

Видно, що $0 \le K_t \le 1$. Очевидно, для біжучої хвилі $K_t = 1$, для стоячої – $K_t = 0$.

Користуються також поняттям коефіцієнта стоячої хвилі (к.с.х.) K_s (індекс від англ. standing wave – стояча хвиля):

$$K_s = \frac{1}{K_t}. \qquad (6.4.28)$$

Контрольні питання до підрозділу 6.4

1. Що таке коефіцієнт відбиття хвилі в лінії? Від чого він залежить?
2. Що таке режим біжучих хвиль у лінії? Коли він реалізується?
3. За яких умов у лінії може реалізуватися режим стоячих хвиль?
4. Чи можливий режим стоячих хвиль з втратами? Відповідь обґрунтуйте.

5. Який режим реалізується в лінії з реактивним навантаженням?
6. Який режим у лінії називається мішаним?
7. Що таке коефіцієнт біжучої хвилі? Що він характеризує?
8. Яке значення к.с.х. відповідає стоячій хвилі?

Задачі до підрозділу 6.4

6.4.1. Двопровідна лінія без втрат довжиною 100 м з хвильовим опором 200 Ом навантажена на активний опір 100 Ом. Знайдіть коефіцієнт відбиття та його модуль.

6.4.2. Знайдіть коефіцієнти біжучої та стоячої хвилі для лінії, описаній у задачі 6.4.1.

6.4.3. Двопровідна лінія без втрат довжиною 50 м з хвильовим опором 100 Ом навантажена на активний опір 200 Ом. Чому дорівнює її ефективна довжина для хвиль з частотою 30 МГц?

6.5. Вхідний опір ділянки лінії

При аналізі ліній зв'язку часто користуються поняттям вхідного опору ділянки лінії. З'ясуємо його зміст.

6.5.1. Вхідний опір ділянки лінії з навантаженням

Вхідний опір лінії з втратами Z_{in} можна подати у формі:

$$Z_{in}=\frac{u(x=0,t)}{i(x=0,t)}=Z\frac{\exp\left[(ik'-k'')l\right]+\frac{Z_2-Z}{Z_2+Z}\exp\left[-(ik'-k'')l\right]}{\exp\left[(ik'-k'')l\right]-\frac{Z_2-Z}{Z_2+Z}\exp\left[-(ik'+k'')l\right]}=$$

$$=Z\frac{(Z_2+Z)\exp\left[(ik'-k'')l\right]+(Z_2-Z)\exp\left[-(ik'-k'')l\right]}{(Z_2+Z)\exp\left[(ik'-k'')l\right]-(Z_2-Z)\exp\left[-(ik'+k'')l\right]}. \qquad (6.5.1)$$

Нагадаємо, що тут Z_2 – імпеданс навантаження, Z – хвильовий опір лінії з урахуванням втрат.

Для лінії без втрат вираз для вхідного опору набуває вигляду (Z_0 – хвильовий опір лінії без втрат)

$$Z_{in}=Z_0\frac{(Z_2+Z_0)\exp(ikl)+(Z_2-Z_0)\exp(-ikl)}{(Z_2+Z_0)\exp(ikl)-(Z_2-Z_0)\exp(-ikl)}=Z_0\frac{Z_2\cos kl+iZ_0\sin kl}{iZ_2\sin kl+Z_0\cos kl}=Z_0\frac{Z_2+iZ_0tg\ kl}{iZ_2tg\ kl+Z_0}. \qquad (6.5.2)$$

Видно, що залежність $Z_{in}(l)$ є періодичною з періодом, що дорівнює половині довжині хвилі в лінії. Перепишемо її у формі

$$Z_{in}=Z_0\frac{(Z_2+iZ_0tg\ kl)(-iZ_2tg\ kl+Z_0)}{Z_2^2tg^2kl+Z_0^2}=\frac{Z_0}{Z_2^2tg^2kl+Z_0^2}\left[Z_0Z_2\left(1+tg^2kl\right)+i\left(Z_0^2-Z_2^2\right)tg\ kl\right]. \qquad (6.5.3)$$

З формул (6.5.2)-(6.5.3) видно, що за умови <u>активного навантаження</u> (коли Z_2 – дійсна величина) вхідний опір лінії виявляється суто активним у таких випадках:

$tg\ kl=0,\quad kl=n\pi,\ n=1,2,3,...\quad Z_{in}=Z_2;$ (6.5.4)

$tg\ kl\to\pm\infty,\quad kl=\pi/2+n\pi,\ n=0,1,2,3,...\quad Z_{in}=Z_0^2/Z_2$ (6.5.5)

$Z_2=Z_0,\quad Z_{in}=Z_0$ (6.5.6)

(узгоджений режим).

Зробити вхідний опір суто реактивним можна за умов:

$Z_2=0,\quad Z_{in}=iZ_0tg\ kl$ (6.5.7)

(режим короткого замикання на виході)

$Z_2\to\infty,\quad Z_{in}=-iZ_0ctg\ kl$ (6.5.8)

(лінія не навантажена).

Формула (6.5.5) для Z_{in} справедлива й для комплексного навантаження. В цьому випадку його реактивна частина змінює знак. Така лінія, зокрема, трансформує нульове навантаження в нескінченне і навпаки. Звичайно для цього використовують чвертьхвильову лінію, яку через це називають *чвертьхвильовим трансформатором*.

Якщо <u>навантаження комплексне</u>, $Z_2=R_2+iX_2$, формула (6.5.2) набуває вигляду

$$Z_{in}=Z_0\frac{R_2+iX_2+iZ_0tg\ kl}{iR_2tg\ kl-X_2tg\ kl+Z_0}=Z_0\frac{\left[R_2+i\left(X_2+Z_0tg\ kl\right)\right]\left[\left(Z_0-X_2tg\ kl\right)-iR_2tg\ kl\right]}{\left(R_2tg\ kl\right)^2+\left(Z_0-X_2tg\ kl\right)^2}=$$

$$=\frac{Z_0\left\{R_2\left[\left(Z_0-X_2tg\ kl\right)+\left(X_2+Z_0tg\ kl\right)tg\ kl\right]+i\left[\left(X_2+Z_0tg\ kl\right)\left(Z_0-X_2tg\ kl\right)-R_2^2tg\ kl\right]\right\}}{\left(R_2tg\ kl\right)^2+\left(Z_0-X_2tg\ kl\right)^2}=$$

$$= \frac{Z_0\left\{R_2 Z_0\left(tg^2 kl+1\right)-i\left[Z_0 X_2 tg^2 kl+\left(X_2^2+R_2^2-Z_0^2\right)tg\ kl-Z_0 X_2\right]\right\}}{\left(R_2 tg\ kl\right)^2+\left(Z_0-X_2 tg\ kl\right)^2}. \qquad (6.5.9)$$

Видно, що в цьому випадку вхідний опір лінії не може бути суто реактивним.

Опір лінії буде суто активним за умови

$$Z_0 X_2 tg^2 kl+\left(X_2^2+R_2^2-Z_0^2\right)tg\ kl-Z_0 X_2=0. \qquad (6.5.10)$$

Це квадратне рівняння щодо $tg\ kl$, яке завжди має дійсні корені. Отже, підбираючи відповідну довжину лінії, її опір можна зробити суто активним за довільного комплексного навантаження.

6.5.2. Фізична інтерпретація вхідного опору лінії передачі

Дамо інтерпретацію отриманим результатам.

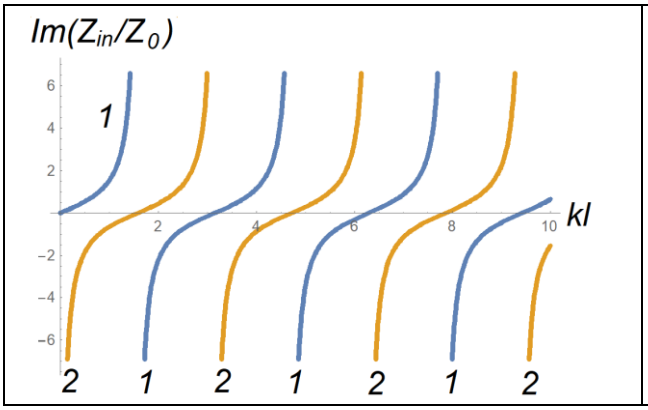
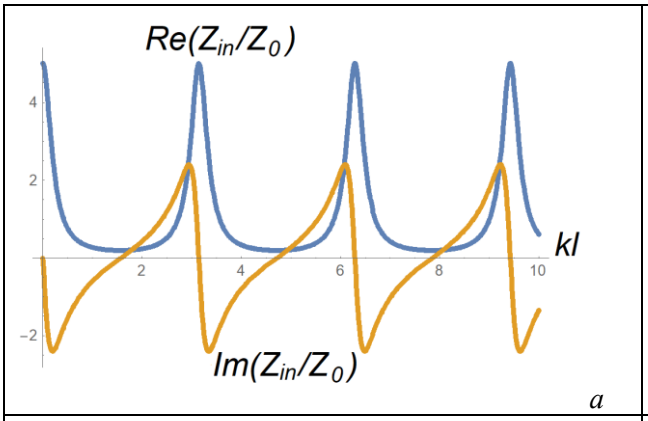
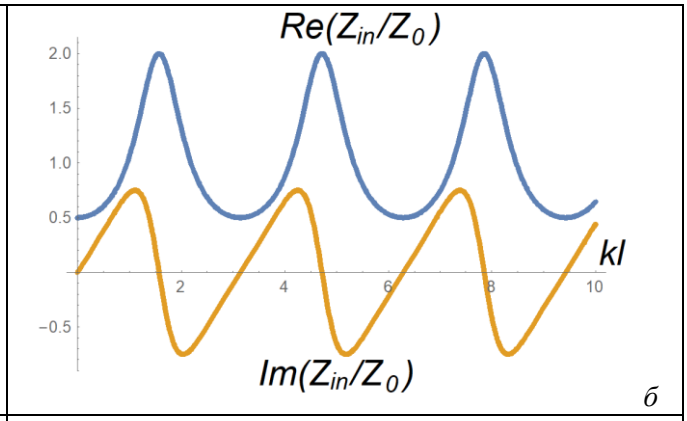


Рис. 6.5.1. Графік залежності уявного вхідного опору лінії передачі без втрат від її довжини для режиму короткого замикання (криві 1) та ненавантаженого режиму (криві 2) на виході





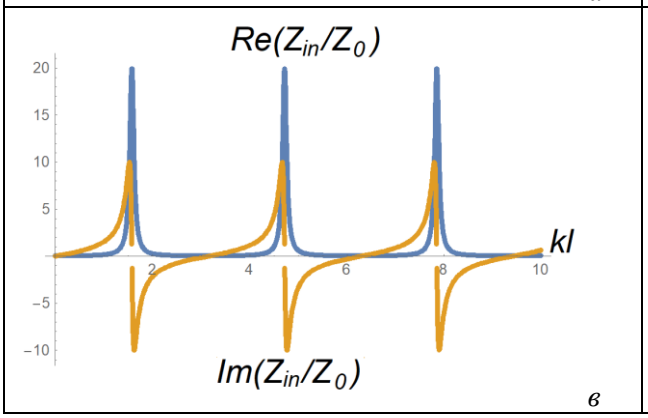


Рис. 6.5.2. Графіки залежності дійсної та уявної частини вхідного опору лінії передачі без втрат від її довжини при $Z_0=0.2Z_2$ (а), $Z_0=2Z_2$ (б), $Z_0=20Z_2$ (в) (опір навантаження вважається суто активним)

Якщо лінія <u>узгоджена з навантаженням</u>, $Z_2=Z_0$, то в ній існує лише пряма хвиля. Відповідно амплітуди напруги й струму на вході лінії не залежать від її довжини. Отже, й відношення цих величин, тобто вхідний опір лінії, не залежить від її довжини (див. формулу

(6.5.6)).

В лінії, вихід якої замкнений накоротко, $Z_2 = 0$, встановлюється стояча хвиля, причому просторові розподіли напруги й струму зсунуті на чверть довжини хвилі один щодо одного. Якщо на початку лінії виникає вузол напруги, вхідний опір обертається на нуль, якщо вузол струму – на нескінченність (див. формулу (6.5.7) та рис. 6.5.1, криві 1). Видно, що вхідний опір в залежності від довжини лінії змінюється з періодом, що дорівнює половині довжини хвилі.

У ненавантаженій лінії ($Z_2 \to \infty$) також виникає стояча хвиля, яка зсунута на чверть довжини хвилі щодо випадку лінії, замкненої накоротко (див. формулу (6.5.8) та рис. 6.5.1, криві 2).

Довільне навантаження (формула (6.5.2)) відповідає мішаному режиму в лінії. В цьому випадку ніде в просторі ні амплітуда струму, ні амплітуда напруги не обертаються на нуль. Тому опір лінії незалежно від її довжини не обертається ні на нуль, ні на нескінченність (рис. 6.5.2), але періодичний характер залежності $Z_{in}(l)$ з періодом $\lambda/2$ зберігається. При $Z_0 \gg Z_2$ режим лінії наближається до короткого замкнення на виході, при $Z_0 \ll Z_2$ – до ненавантаженого режиму.

У точках, де $\mathrm{Re}\{Z_{in}\}$ досягає максимального значення, а $\mathrm{Im}\{Z_{in}\} = 0$ (рис. 6.5.2), поведінка відрізку лінії нагадує поведінку паралельного коливного контуру. Навпаки, в точках, де $\mathrm{Re}\{Z_{in}\}$ досягає мінімального значення, а $\mathrm{Im}\{Z_{in}\} = 0$, відрізок лінії можна уподібнити послідовному коливному контуру.

Нарешті, дамо інтерпретацію ефективній довжині (6.4.25) лінії без втрат

$$l_{eff} = l + \frac{\lambda}{2\pi} arctg \frac{X_2}{Z_0} = l + \Delta l,$$

навантаженій довільним реактивним імпедансом X_2. Ввівши ефективну довжину, ми тим самим замінюємо імпеданс X_2 короткозамкненим відрізком лінії завдовжки Δl, вхідний імпеданс якої має ту саму величину X_2.

6.5.3. Зв'язок хвилі в лінії з параметрами генератора на її вході

Ми записували загальний розв'язок телеграфних рівнянь (6.31) через дві невідомі амплітуди – u_{m1} та u_{m2}:

$$u(x,t) = u_{m1} \exp(i\omega t - ikx) + u_{m2} \exp(i\omega t + ikx).$$

Розглядаючи характерні режими лінії з навантаженням, ми виражали амплітуду u_{m2} через u_{m1}, використавши для цього граничну умову в точці увімкнення навантаження.

Тепер ми використаємо граничну умову на початку лінії для знаходження амплітуди u_{m2} через параметри джерела сигналу (див. рис. 6.5.3, що повторює рис. 6.3.1):

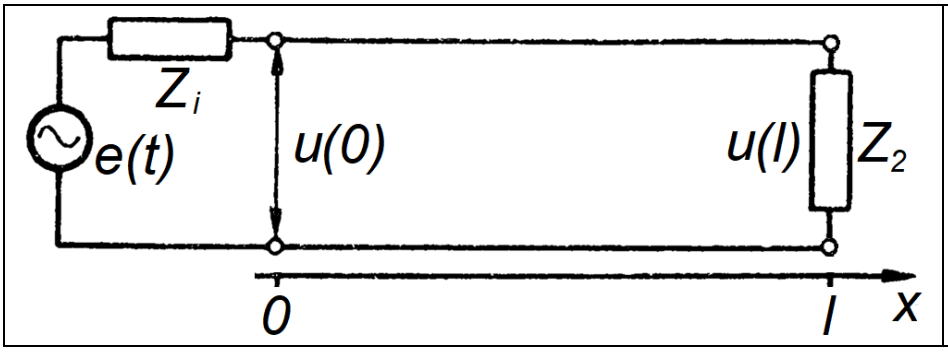


Рис. 6.5.3. Лінія передачі з джерелом сигналу та навантаженням

Знаючи вхідний опір лінії, можна пов'язати напругу на її початку з е.р.с. та внутріш-

нім опором джерела сигналу (див. рис. 6.4.1):

$$u(0,t) = i(0,t)Z_{in} = \frac{e(t)}{Z_{in} + Z_i} Z_{in}, \qquad e(t) = e_m \exp(i\omega t). \tag{6.5.11}$$

З іншого боку,

$$u(0,t) = u_{m1} \exp(i\omega t)(1+R), \tag{6.5.12}$$

звідки отримуємо вираз для амплітуди прямої хвилі через параметри джерела сигналу:

$$u_{m1} = \frac{e_m}{(1+R)(1+Z_i/Z_{in})}, \tag{6.5.13}$$

де R і Z_{in}, в свою чергу, виражаються через частоту сигналу, параметри лінії та навантаження.

Контрольні питання до підрозділу 6.5

1. Від чого залежить вхідний опір ділянки з навантаженням?
2. Що таке чвертьхвильовий трансформатор?
3. За яких умов опір лінії не залежить від її довжини? Чому так відбувається?
4. В яких межах в залежності від її довжини може змінюватися опір лінії, в якій існує режим стоячих хвиль?
5. Порівняйте резонансні властивості коливного контуру та відрізку лінії передачі.

Задачі до підрозділу 6.5

6.5.1. Двопровідна лінія, навантажена на послідовне сполучення опору 100 Ом і ємності 200 нФ, має довжину 100 м і хвильовий опір 200 Ом. Швидкість хвиль у лінії – 10^8 м/с. Частота хвилі – 3 ГГц. Розрахувати вхідний опір лінії.

6.5.2. Двопровідна лінія з хвильовим опором 150 Ом навантажена індуктивністю 100 мГн. Швидкість хвиль у лінії – $1.5 \cdot 10^8$ м/с. Частота хвилі 400 МГц. За якої мінімальної довжини лінії її вхідний опір буде

а) нескінченним;

б) нульовим?

6.5.3. Двопровідна лінія має довжину 50 м і хвильовий опір 250 Ом. Довжина хвилі – 30 см. Яким повинен бути опір навантаження, щоб вхідний опір лінії був удвічі більший від її хвильового опору?

6.5.4. Резонатор являє собою ділянку двопровідної лінії завдовжки 2 м. Швидкість хвиль у лінії – $1.2 \cdot 10^8$ м/с. Знайдіть три найнижчі власні частоти такого резонатора.

6.6. Узгодження лінії з навантаженням та джерелом сигналу

Розглянемо лінію з джерелом сигналу на вході й навантаженням на виході (рис. 6.6.1, який повторює рис. 6.4.1). В цьому випадку лінію передачі називають фідером (від англ. to feed – живити). В ролі навантаження може виступати який-небудь пристрій (наприклад, антена). Очевидно, при фіксованій потужності джерела максимальна потужність виділятиметься на навантаженні за відсутності відбиття, тобто коли в лінії реалізується режим біжучих хвиль.

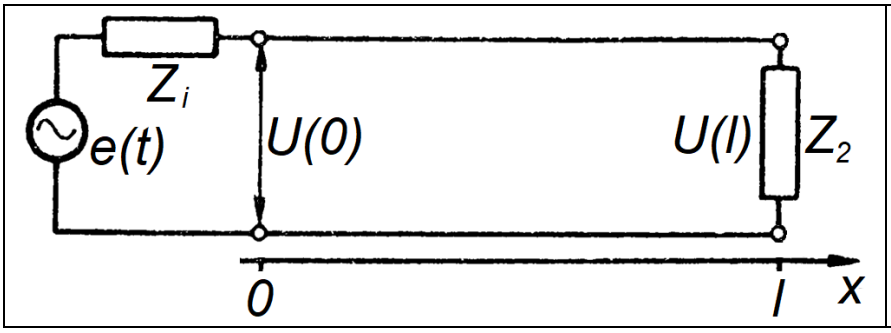


Рис. 6.6.1. Лінія з джерелом сигналу та навантаженням

Варто також зазначити, що в лінії без втрат максимальна напруга (яка визначає поріг пробою) в режимі біжучих хвиль буде меншою, ніж в мішаному режимі або в режимі стоячих хвиль.

Розглянемо кілька способів узгодження лінії з навантаженням.

6.6.1. Узгодження лінії з навантаженням за допомогою чвертьхвильового трансформатора

Випадок активного навантаження. Нехай спершу навантаження лінії є суто активним, $Z_2 = R_2 \neq Z_0$. В цьому випадку для узгодження можна використати чвертьхвильовий трансформатор з хвильовим опором Z_0', що задовольняє умові узгодження між точками A і B, $Z_0 = (Z_0')^2 / R_2$, тобто

$$Z_0' = \sqrt{R_2 Z_0} \qquad (6.6.1)$$

(рис. 6.6.2 а). Для відрізку довгої лінії найпростіше підібрати потрібний хвильовий опір Z_0', регулюючи віддаль між провідниками (див. формулу (6.3.13)). Фізично виконання умови узгодження означає, що хвилі, відбиті від навантаження та від неоднорідності лінії, є однаковими за амплітудою й протилежними за фазою.

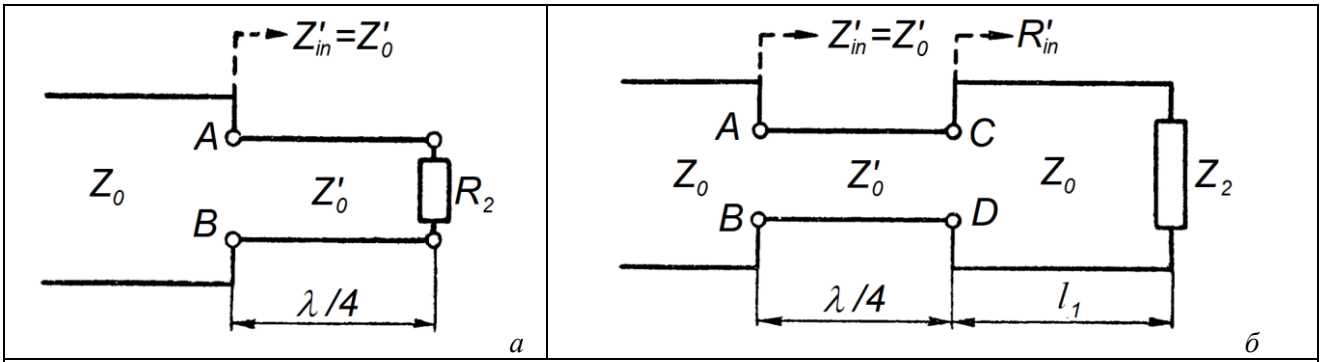


Рис. 6.6.2. Узгодження лінії з навантаженням за допомогою чвертьхвильового трансформатора: а – суто активне навантаження; б – комплексне навантаження

Загальний випадок. Нехай тепер навантаження лінії є комплексним. Як указувалося вище, можна так підібрати довжину l_1 ділянки лінії, навантаженої на комплексний опір, що вхідний опір цієї ділянки між точками C і D стане суто дійсним. Тим самим задача зводиться

до попередньої: перед цією ділянкою слід увімкнути чвертьхвильовий трансформатор для узгодження з основною частиною лінії (рис. 6.6.2 б).

Використання чвертьхвильового трансформатора дозволяє забезпечити узгодження більшої частини довгої лінії з навантаженням. Однак інколи для цього бувають потрібні чвертьхвильові трансформатори з дуже великим або дуже малим хвильовим опором, що на практиці незручно. Тому розроблені й інші методи узгодження ліній із навантаженням.

6.6.2. Узгодження лінії з навантаженням за допомогою шлейфа Татаринова

Шлейф Татаринова, або паралельний реактивний шлейф – це ділянка лінії, кінець якої замкнений накоротко. Відповідно його вхідний опір є реактивним і може мати довільну величину в залежності від його довжини (див. формулу (6.5.22)).

Нехай провідність навантаження є комплексною, $Y_2 = Y_2' + iY_2''$. Можна так підібрати довжину відрізку l_0 (рис. 6.6.3), щоб дійсна частина вхідної провідності цього відрізку (між точками А і В) з навантаженням Y_2 дорівнювала хвильовій провідності лінії $Y_0 = 1/Z_0$. Як випливає з формули (6.5.9), для цього має виконуватись умова

$$\mathrm{Re}\left\{\frac{1}{Z_0}\frac{\left(Z_0 - X_2 tg\ kl\right) + iR_2 tg\ kl}{R_2 + i\left(X_2 + Z_0 tg\ kl\right)}\right\} = \frac{1}{Z_0},$$

або

$$\frac{R_2\left(Z_0 - X_2 tg\ kl\right) + R_2 tg\ kl\left(X_2 + Z_0 tg\ kl\right)}{R_2^2 + \left(X_2 + Z_0 tg\ kl\right)^2} = 1,$$

або

$$Z_0\left(Z_0 - R_2\right) tg^2 kl + 2Z_0 X_2 tg\ kl + \left(R_2^2 + X_2^2 - Z_0 R_2\right) = 0. \qquad (6.6.2)$$

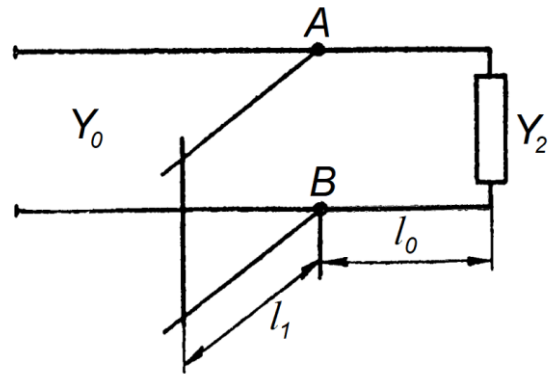


Рис. 6.6.3. Узгодження лінії з навантаженням за допомогою шлейфа Татаринова

Рівняння (6.6. 2) є квадратним щодо $tg\ kl$, а його корені відповідають шуканій довжині l_0 відрізку лінії. Зазначимо, що на практиці цю довжину визначають експериментально.

Далі на відстані l_0 від навантаження паралельно до навантаженого відрізку лінії підключають шлейф Татаринова, за допомогою якого компенсують реактивну складову вхідної провідності навантаженого відрізку. Очевидно, довжина шлейфу l_1 повинна задовольняти умові

$$\mathrm{Im}\left\{\frac{1}{Z_{in}\left(x = l - l_0\right)}\right\} = \frac{1}{Z_0} ctg\ kl_1. \qquad (6.6.3)$$

Таким чином, сумарна вхідна провідність ділянки між точками А і В (з урахуванням шлейфу Татаринова) дорівнює Y_0, що й вимагається для узгодження.

Фізично умова узгодження означає, що хвилі, відбиті від навантаження та від короткозамкненого кінця шлейфу Татаринова, виявляються однаковими за амплітудою й протилеж-

ними за фазою.

Узгодження за допомогою одного шлейфу Татаринова вимагає регулювання довжини відрізку l_0, що на практиці буває незручно. Тому був запропонований спосіб узгодження за допомогою двох шлейфів Татаринова, який позбавлений цього недоліку.

6.6.3. Узгодження лінії з навантаженням за допомогою двох шлейфів Татаринова

Схема такого узгодження подана на рис. 6.6.4. Видно, що один із шлейфів увімкнений безпосередньо до навантаження, а інший – на віддалі $l_0 = (3/8)\lambda$ від нього. Регулюючи довжину l_1, добиваються того, щоб активна провідність ділянки l_0 і шлейфа l_1 між точками А та В дорівнювала Y_0, а реактивну складову цієї провідності компенсують за допомогою шлейфу l_2.

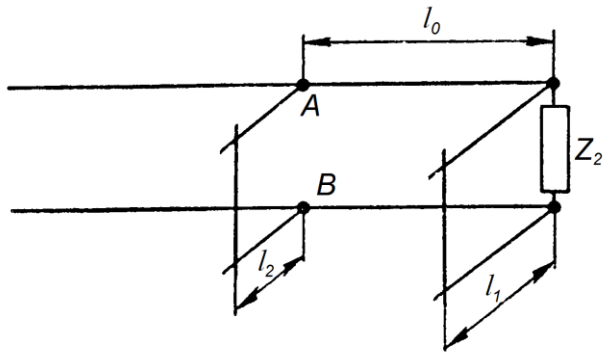


Рис. 6.6.4. Узгодження лінії з навантаженням за допомогою двох шлейфів Татаринова

Для знаходження довжини l_1 треба задовольнити рівняння (6.6.2), куди внаслідок умови $l_0 = (3/8)\lambda$ слід підставити $kl_0 = 3\pi/4$, $tg\ kl_0 = -1$, а під $1/X_2$ слід розуміти суму реактивної провідності навантаження $-X_2/(R_2^2 + X_2^2)$ і реактивної провідності шлейфа l_1, що дорівнює $-1/(Z_0 tg\ kl_1)$. Розв'язавши рівняння (6.6.2) щодо X_2, знаходимо потрібне значення X_{20}, а далі з умови

$$\frac{1}{X_{20}} = -\frac{X_2}{R_2^2 + X_2^2} - \frac{1}{Z_0 tg\ kl_1} \tag{6.6.4}$$

знаходимо l_1. Нарешті, з умови

$$\mathrm{Im}\left\{\frac{1}{Z_{in}\left(x = l - \frac{3}{4}\lambda\right)}\right\} = \frac{1}{Z_0} ctg\ kl_2, \tag{6.6.5}$$

аналогічної (6.6.3), знаходимо довжину l_2.

6.6.4. Узгодження джерела сигналу з лінією

Збільшити потужність сигналу, що виділяється на (комплексному) опорі навантаження, можна не тільки шляхом узгодження навантаження з лінією передачі, а й шляхом узгодження джерела сигналу з лінією. Вважатимемо, що лінія узгоджена з навантаженням, тобто в ній досягнутий режим біжучої хвилі. Втратами в лінії знехтуємо. Тоді її вхідний опір дорівнює хвильовому опору Z_0. Еквівалентна схема джерела сигналу, навантаженого такою лінією, подана на рис. 6.6.5.

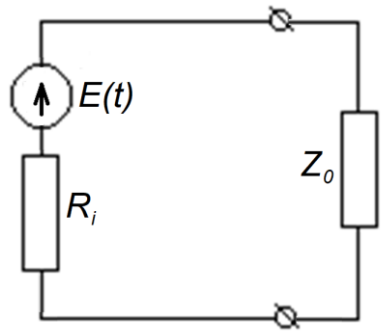


Рис. 6.6.5. Еквівалентна схема джерела сигналу, навантаженого лінією зв'язку в режимі біжучої хвилі

Подібна схема розглядалася в п. 6.4.2. Там було показано, що максимальна потужність на опорі навантаження виділяється за виконання умови узгодження $Z_0 = R_i$. В загальному випадку ця умова не виконується. Тоді для забезпечення умови узгодження джерела сигналу з лінією передачі можна скористатися трансформатором. Відповідна еквівалентна схема подана на рис. 6.6.6.

Відразу зазначимо, що ми вважатимемо трансформатор ідеальним у тому сенсі, що магнітний потік, створений кожною з обмоток, повністю пронизує й іншу обмотку, а самі обмотки відрізняються лише кількістю витків. Тоді взаємна індуктивність обмоток пов'язана з їхніми власними індуктивностями співвідношенням

$$M = \sqrt{L_1 L_2}\ . \tag{6.6.6}$$

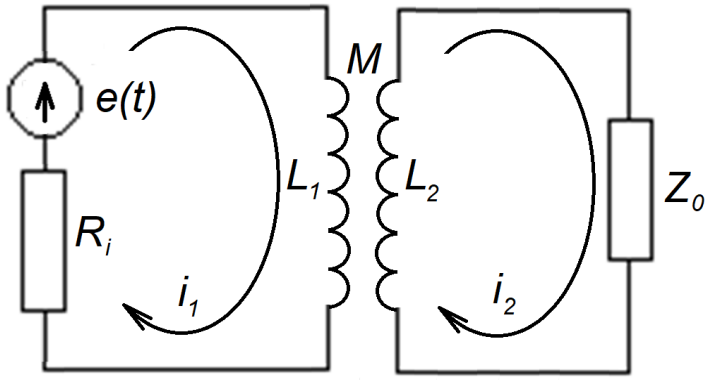


Рис. 6.6.6. Еквівалентна схема узгодження джерела сигналу з навантаженням за допомогою трансформатора

Вважатимемо сигнал джерела гармонічним, $e(t) = e_m \exp(i\omega t)$. Тоді рівняння за другим правилом Кірхгофа для контурів схеми, поданої на рис. 6.6.6, мають вигляд:

$$\begin{cases} (i\omega L_1 + R_i) i_{m1} + i\omega M i_{m2} = e_m; \\ i\omega M i_{m1} + (i\omega L_2 + Z_0) i_{m2} = 0. \end{cases} \tag{6.6.7}$$

Із системи (6.6.7) неважко знайти струм i_{m2}. З урахуванням умови (6.6.6) його можна записати у формі

$$i_{m2} = \frac{-i\omega\sqrt{L_1 L_2}\, e_m}{i\omega(L_1 Z_0 + L_2 R_i) + Z_0 R_i}, \tag{6.6.8}$$

а середню потужність, що виділяється на опорі Z_0 –

$$\overline{P} = \frac{1}{2} i_{m2} i_{m2}^* Z_0 = \frac{1}{2} \frac{\omega^2 L_1 L_2 e_m^2}{\omega^2 (L_1 Z_0 + L_2 R_i)^2 + (Z_0 R_i)^2}. \tag{6.6.9}$$

Зафіксуємо індуктивність L_1 і знайдемо, при якому значенні L_2 потужність (6.6.9) досягає максимуму. Маємо:

$$\frac{d\overline{P}}{dL_2} \sim \left[\omega^2\left(L_1Z_0 + L_2R_i\right)^2 + \left(Z_0R_i\right)^2\right] - L_2\left[2\omega^2\left(L_1Z_0 + L_2R_i\right)R_i\right] = 0,$$

або

$$\left(\omega L_1 Z_0\right)^2 + \left(Z_0R_i\right)^2 = \omega^2 L_2^2 R_i^2, \qquad (6.6.10)$$

звідки

$$L_2^2 = L_1^2\left[\left(\frac{Z_0}{R_i}\right)^2 + \left(\frac{Z_0}{\omega L_1}\right)^2\right]. \qquad (6.6.11)$$

Зазвичай $Z_0 \ll \omega L_1$, і остаточно

$$\frac{L_2}{L_1} \approx \frac{Z_0}{R_i}. \qquad (6.6.12)$$

За виконання умови опір R_i в колі первинної обмотки трансформується в опір Z_0 у колі вторинної обмотки, за рахунок чого й досягається узгодження джерела сигналу з лінією передачі.

Контрольні питання до розділу 6.6

1. Для чого потрібне узгодження фідера з навантаженням?
2. У чому відмінність випадків суто активного й комплексного навантаження при узгодженні з ним лінії за допомогою чвертьхвильового трансформатора?
3. Який фізичний зміст узгодження лінії з навантаженням за допомогою шлейфу Татаринова? За рахунок чого досягається таке узгодження?
4. В чому перевага схеми узгодження з використанням двох шлейфів Татаринова?
5. Що дає узгодження джерела сигналу з лінією? Як його можна забезпечити?

Приклади розв'язання задач до підрозділу 6.6

П 6.6.1. Довга лінія з хвильовим опором Z_{01}=200 Ом навантажена резистором R_l=50 Ом і конденсатором ємністю C_l=0.1 нФ, увімкненими паралельно. Швидкість поширення хвиль у лінії – c=2·10^8 м/с, частота хвилі – $\omega/2\pi$=1 ГГц. Знайти хвильовий опір Z_{02} чвертьхвильового трансформатора та довжини l_1 і $\lambda/4$ (рис. 6.6.1 б), які забезпечують узгодження лінії з навантаженням.

Розв'язок

Спершу знайдемо комплексний опір навантаження:

$$Z_l = \left(\frac{1}{R_l} + \frac{1}{i\omega C_l}\right)^{-1} = \frac{i\omega C_l R_l}{R_l + i\omega C_l} = \frac{i\omega C_l R_l\left(R_l - i\omega C_l\right)}{R_l^2 + \left(\omega C_l\right)^2} = R_l\frac{\left(\omega C_l\right)^2}{R_l^2 + \left(\omega C_l\right)^2} + i\omega C_l\frac{R_l^2}{R_l^2 + \left(\omega C_l\right)^2};$$

$$\omega C_l = 2\pi\cdot 10^9\cdot 10^{-10} \approx 62.8\ Ом;$$

$$Z_l = R_l\frac{\left(\omega C_l\right)^2}{R_l^2 + \left(\omega C_l\right)^2} + i\omega C_l\frac{R_l^2}{R_l^2 + \left(\omega C_l\right)^2} = 50\frac{\left(62.8\right)^2}{\left(50\right)^2 + \left(62.8\right)^2} + i62.8\frac{\left(50\right)^2}{\left(50\right)^2 + \left(62.8\right)^2} =$$

$$= 50\frac{3944}{2500 + 3944} + i62.8\frac{\left(50\right)^2}{\left(50\right)^2 + \left(62.8\right)^2} = 50\cdot 0.61 + i62.8\cdot 0.39 = \left(31 + i24\right)\text{Ом} = R_l' + iX_l'$$

Відповідно до формули (6.5.9) лінія довжиною l_1 з хвильовим опором Z_{01}, навантажена на комплексний опір $R_l' + iX_l'$, має вхідний опір

$$Z_l' = Z_{01}\frac{R_l'\left[Z_{01}tg^2kl_1 + Z_{01}\right] - i\left\{Z_{01}X_l'tg^2kl_1 + \left[\left(X'\right)_l^2 + \left(R'\right)_l^2 - Z_{01}^2\right]tg\ kl_1 - Z_{01}X_l'\right\}}{\left(R_l'tg\ kl_1\right)^2 + \left(Z_{01} - X_l'tg\ kl_1\right)^2}.$$

Для узгодження цей опір має бути дійсним, тобто довжину l_1 можна знайти з умови

$$Z_{01}X_l'tg^2kl_1 + \left[\left(X'\right)_l^2 + \left(R'\right)_l^2 - Z_{01}^2\right]tg\ kl_1 - Z_{01}X' = 0,$$

або

$$tg^2kl_1 + a\ tg\ kl_1 - 1 = 0,$$

$$a = \frac{X_l'}{Z_{01}} + \frac{\left(R_l'\right)^2}{Z_{01}X_l'} - \frac{Z_{01}}{X_l'}Z_{01}^2 = -\frac{31}{200} + \frac{\left(24\right)^2}{31\cdot 200} - \frac{200}{31} = -0.16 + 0.10 - 6.45 \approx 6.5$$

$$tg\ kl_1 = \frac{1}{2}\left[-6.5 \pm \sqrt{\left(6.5\right)^2 + 4}\right] = \frac{6.8 - 6.5}{2} = 0.15 << 1$$

$$kl_1 = arctg\ 0.15 \approx 0.15$$

$$k = c/\omega = 2\pi\cdot 2\cdot 10^8/10^{10} = 1.26\ \text{м}^{-1};\ l_1 = 0.15/1.26 = 0.12\ \text{м};$$

$$\lambda/4 = \pi/2k = 6.28/1.26 = 5.0\ \text{м}.$$

Знайдемо тепер вхідний опір R_l'' ділянки завдовжки l_1 з навантаженням $R_l' + iX_l'$:

$$R_l'' = Z_{01}\frac{R_l'\left[Z_{01}tg^2kl_1 + \left(Z_{01} + X_l'\right)\right]}{\left(R_l'tg\ kl_1\right)^2 + \left(Z_{01} - X_l'tg\ kl_1\right)^2} = Z_{01}\frac{R_l'tg^2kl_1/Z_{01} + 1 + X_l'/Z_{01}}{\left(R_l'tg^2kl_1/Z_{01}\right)^2 + \left(1 - X_l'tg\ kl_1/Z_{01}\right)^2} \approx$$

$$\approx Z_{01}\frac{1 + X_l'/Z_{01}}{1 - 2\,X_l'tg\ kl_1/Z_{01}} = 200\frac{1 + 24/200}{1 - 24\cdot 0.3/200} = 200\frac{1.12}{0.96} = 233\ \text{Ом}.$$

(ми знехтували малими доданками, пропорційними tg^2kl_1).

Хвильовий опір чвертьхвильового трансформатора, який забезпечує узгодження лінії з навантаженням R_l'', визначається формулою (6.6.1):

$$Z_{02} = \sqrt{R_l''Z_{01}} = \sqrt{233\cdot 200} = 216\ \text{Ом}.$$

Задачі до підрозділу 6.6

6.6.1. Лінія з хвильовим опором 100 Ом навантажена активним опором 10 Ом. Яким має бути хвильовий опір чвертьхвильового трансформатора, що забезпечує їхнє узгодження?

6.6.2. Довга лінія з хвильовим опором Z_0=200 Ом навантажена резистором R_l=50 Ом і конденсатором ємністю C_l=0.1 нФ, увімкненими паралельно. Швидкість поширення хвиль у лінії – c=2·10^8 м/с, частота хвилі – $\omega/2\pi$=1 ГГц. Лінія узгоджена з навантаженням за допомогою шлейфу Татаринова (рис. 6.6.2). Знайти довжини l_0 та l_1.

6.6.3. Розв'язати попередню задачу для випадку, коли для узгодження лінії з навантаженням використовується два шлейфи Татаринова (рис. 6.6.3). Знайти довжини l_1 та l_2, вважаючи, що l_0=1м.

6.6.4. Хвильовий опір довгої лінії складає Z_0=100 Ом, внутрішній опір джерела сигналу – R_i=25 Ом. Частота сигналу – $\omega/2\pi$=100 МГц. Узгодження джерела сигналу з лінією забезпечується за допомогою трнсформатора. Знайти індуктивності первинної L_1 та вторинної обмоток і їхню взаємну індуктивність M.

6.7. Відрізок довгої лінії як резонатор

Досі (див. вище пп. 6.3-6.6) ми розглядали довгу лінію, до початку якої підключене джерело сигналу, а до кінця – деяке навантаження (див. рис. 6.7.1 а, який повторює рис. 6.4.1). Але виявляється, що відрізок довгої лінії обмеженої довжини може також служити резонатором (зазвичай у мікрохвильовому діапазоні).

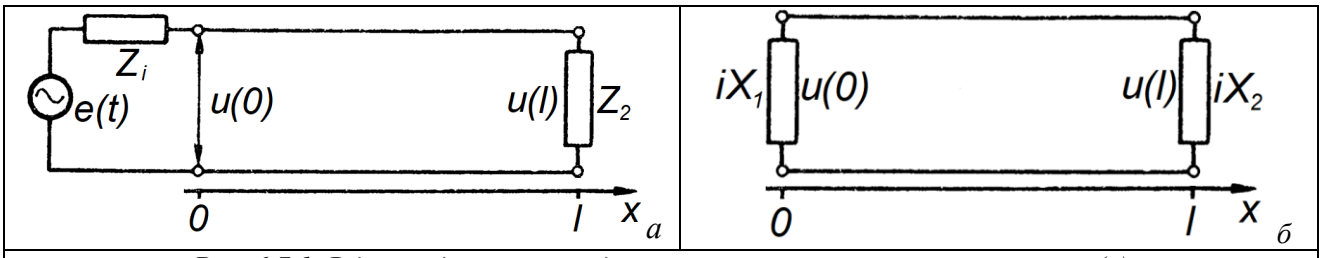


Рис. 6.7.1. Відрізки довгої лінії з джерелом сигналу та навантаженням (а) та з реактивними елементами на кінцях (б)

Річ у тім, що на високих частотах (починаючи з діапазону ультракоротких (метрових) хвиль), використання коливних контурів як частотновибіркових елементів стає недоцільним. Ємність контуру не може бути меншою від характерної паразитної ємності, тому просування в бік менших частот можливе лише за рахунок зменшення індуктивності. Але при цьому зменшується хвильовий опір і, як наслідок, добротність коливного контуру. Саме тому на високих частотах замість коливних контурів використовують відрізки передавальних ліній.

6.7.1. Власні частоти і власні моди в загальному випадку

Розглядатимемо спочатку систему без втрат. Щоб знайти власні частоти відрізку довгої лінії, ми повинні розглянути випадок, коли на обох кінцях цього відрізку увімкнені деякі реактивні елементи, імпеданси яких ми вважаємо суто уявними (рис. 6.7.1 б). Тоді потрібно відповідним чином записати граничні умови на обох кінцях цього відрізку. В такій постановці ми розглядаємо власні коливання, залишаючи осторонь питання про їхнє збудження (точніше, вважатимемо, що вони виникають за рахунок ненульовим початкових умов).

Загальний розв'язок телеграфних рівнянь ми записували у формі (6.3.1)-(6.3.2):

$$u(x,t)=u_{m1}\exp(i\omega t-ikx)+u_{m2}\exp(i\omega t+ikx).$$

$$i(x,t)=(1/Z_0)\left[u_{m1}\exp(i\omega t-ikx)-u_{m2}\exp(i\omega t+ikx)\right],$$

$$Z_0=\sqrt{L_1/C_1}\ .$$

Граничні умови:

$$u(x=0,t)=iX_1 i(x=0,t);\qquad u(x=l,t)=iX_2 i(x=l,t), \tag{6.7.1}$$

або

$$\begin{cases} u_{m1}\exp(i\omega t)+u_{m2}\exp(i\omega t)=(iX_1/Z_0)\left[u_{m1}\exp(i\omega t)-u_{m2}\exp(i\omega t)\right]; \\ u_{m1}\exp(i\omega t-ikl)+u_{m2}\exp(i\omega t+ikl)=(iX_2/Z_0)\left[u_{m1}\exp(i\omega t-ikl)-u_{m2}\exp(i\omega t+ikl)\right], \end{cases}$$

або

$$\begin{cases} (1-iX_1/Z_0)u_{m1}+(1+iX_1/Z_0)u_{m2}=0; \\ (1-iX_2/Z_0)u_{m1}\exp(-ikl)+(1+iX_2/Z_0)u_{m2}\exp(ikl)=0. \end{cases} \tag{6.7.1 а}$$

Щоб система лінійних однорідних алгебраїчних рівнянь (6.7.1 а) щодо амплітуд u_{m1}, u_{m2} мала нетривіальні розв'язки, її визначник має дорівнювати нулю:

$$\begin{vmatrix} 1-iX_1/Z_0 & 1+iX_1/Z_0 \\ (1-iX_2/Z_0)\exp(-ikl) & (1+iX_2/Z_0)\exp(ikl) \end{vmatrix}=0\,. \tag{6.7.2}$$

Поділимо перший рядок на $1-iX_1/Z_0$, другий – $(1-iX_2/Z_0)\exp(-ikl)$ і позначимо

$$\frac{1+iX_{1,2}/Z_0}{1-iX_{1,2}/Z_0}=\exp\left(i2\Phi_{1,2}\right). \qquad (6.7.3)$$

(пор. з формулою (6.4.20)). Тоді умова (6.7.2) набуває вигляду

$$\begin{vmatrix} 1 & \exp\left(i2\Phi_1\right) \\ 1 & \exp\left(i2\Phi_2\right)\exp\left(i2kl\right) \end{vmatrix}=0, \qquad (6.7.2\ \text{а})$$

або

$$\exp\left\{i\left[2kl+2\left(\Phi_2-\Phi_1\right)\right]\right\}=1,$$

або

$$kl+\Phi_2-\Phi_1=\pi n, \qquad n=1,2,3,... \qquad (6.7.4)$$

Рівняння (6.7.4) – це характеристичне рівняння, з якого можна знайти власні значення хвильових чисел, довжин хвиль та частот відрізку довгої лінії:

$$k_n=\frac{1}{l}\left(\pi n+\Phi_1-\Phi_2\right); \qquad (6.7.5)$$

$$\lambda_n=\frac{2\pi}{k_n}=\frac{2l}{n+\left(\Phi_1-\Phi_2\right)/\pi}; \qquad (6.7.6)$$

$$\omega_n=k_nc_0=\frac{c_0}{l}\left(\pi n+\Phi_1-\Phi_2\right) \qquad (6.7.7)$$

(c_0, як і раніше – фазова швидкість хвилі в лінії).

Після того, як ми прирівняли до нуля визначник системи (6.7.1 а), відповідні рівняння стали еквівалентними. З першого з них можна знайти зв'язок між амплітудами u_{m1} і u_{m2}:

$$u_{m2}=-\frac{1-iX_1/Z_0}{1+iX_1/Z_0}u_{m1}=-u_{m1}\exp\left(-i2\Phi_1\right). \qquad (6.7.8)$$

Тоді

$$u\left(x,t\right)=u_{m1}\exp\left(i\omega_nt-ik_nx\right)-u_{m1}\exp\left(-i2\Phi_1\right)\exp\left(i\omega_nt+ik_nx\right)=$$
$$=u_{m1}\exp\left(i\omega_nt-i\Phi_1\right)\left[\exp\left(-ik_nx+i\Phi_1\right)-\exp\left(ik_nx-i\Phi_1\right)\right]= \quad .$$
$$=-2iu_{m1}\exp\left(i\omega_nt-i\Phi_1\right)\sin\left(k_nx-\Phi_1\right);$$
$$i\left(x,t\right)=\left(1/Z_0\right)\left[u_{m1}\exp\left(i\omega_nt-ik_nx\right)+u_{m1}\exp\left(-i2\Phi_1\right)\exp\left(i\omega_nt+ik_nx\right)\right]=$$
$$=\left(u_{m1}/Z_0\right)\exp\left(i\omega_nt-i\Phi_1\right)\left[\exp\left(-ik_nx+i\Phi_1\right)+\exp\left(ik_nx-i\Phi_1\right)\right]= \quad ,$$
$$=\left(2u_{m1}/Z_0\right)\exp\left(i\omega_nt-i\Phi_1\right)\cos\left(k_nx-\Phi_1\right).$$

Вимірювані величини:

$$\mathrm{Re}\left\{u\left(x,t\right)\right\}=2u_{m1}\sin\left(i\omega_nt-i\Phi_1\right)\sin\left(k_nx-\Phi_1\right); \qquad (6.7.9)$$

$$\mathrm{Re}\left\{i\left(x,t\right)\right\}=\left(2u_{m1}/Z_0\right)\cos\left(i\omega_nt-i\Phi_1\right)\cos\left(k_nx-\Phi_1\right), \qquad (6.7.10)$$

З формул (6.7.9)-(6.7.10) видно, що вздовж відрізку лінії встановлюються стоячі хвилі. Просторові розподіли струму й напруги зсунуті на чверть довжини хвилі, часові коливання в кожній фіксованій точці – на чверть періоду. Кожній власній частоті ω_n (або довжині хвилі λ_n) відповідають свої просторові розподіли напруги й струму.

Тип коливання (в даному разі – резонатора, що має вигляд відрізку лінії), який характеризується власною частотою та просторовим розподілом напруги й струму, називається (власною) модою.

Ми розв'язали задачу про власні моди відрізку довгої лінії в загальному випадку. Тепер розглянемо деякі частинні випадки.

6.7.2. Відрізок лінії з короткозамкненими кінцями

Для відрізку лінії з короткозамкненими кінцями $X_1 = X_2 = 0$ і, відповідно, $\Phi_1 = \Phi_2 = 0$. Тоді формули для власних частот і відповідних довжин хвиль набувають найпростішого вигляду:

$$\lambda_n = \frac{2l}{n}; \tag{6.7.11}$$

$$f_n = \frac{nc_0}{2l}. \tag{6.7.12}$$

Перепишемо формулу (6.7.11) у формі $n\lambda_n/2 = l$. Як бачимо, в цьому випадку на відрізку лінії вкладається ціле число півхвиль. Найменша власна частота – $f_{min} = f_1 = c_0/2l$, всі інші власні частоти кратні їй, утворюючи еквідистантний спектр (рис. 6.7.2).

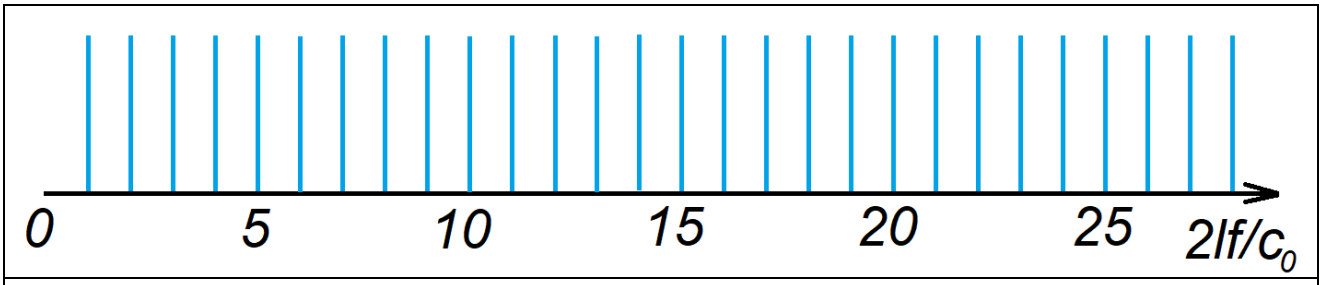


Рис. 6.7.2. Власні частоти відрізка лінії з короткозамкненим кінцем

Розподіли напруги й струму тепер визначаються формулами

$$\mathrm{Re}\{u(x,t)\} = 2u_{m1}\sin(i\omega_n t)\sin(k_n x); \tag{6.7.13}$$

$$\mathrm{Re}\{i(x,t)\} = (2u_{m1}/Z_0)\cos(i\omega_n t)\cos(k_n x) \tag{6.7.14}$$

(рис. 6.7.3 а).

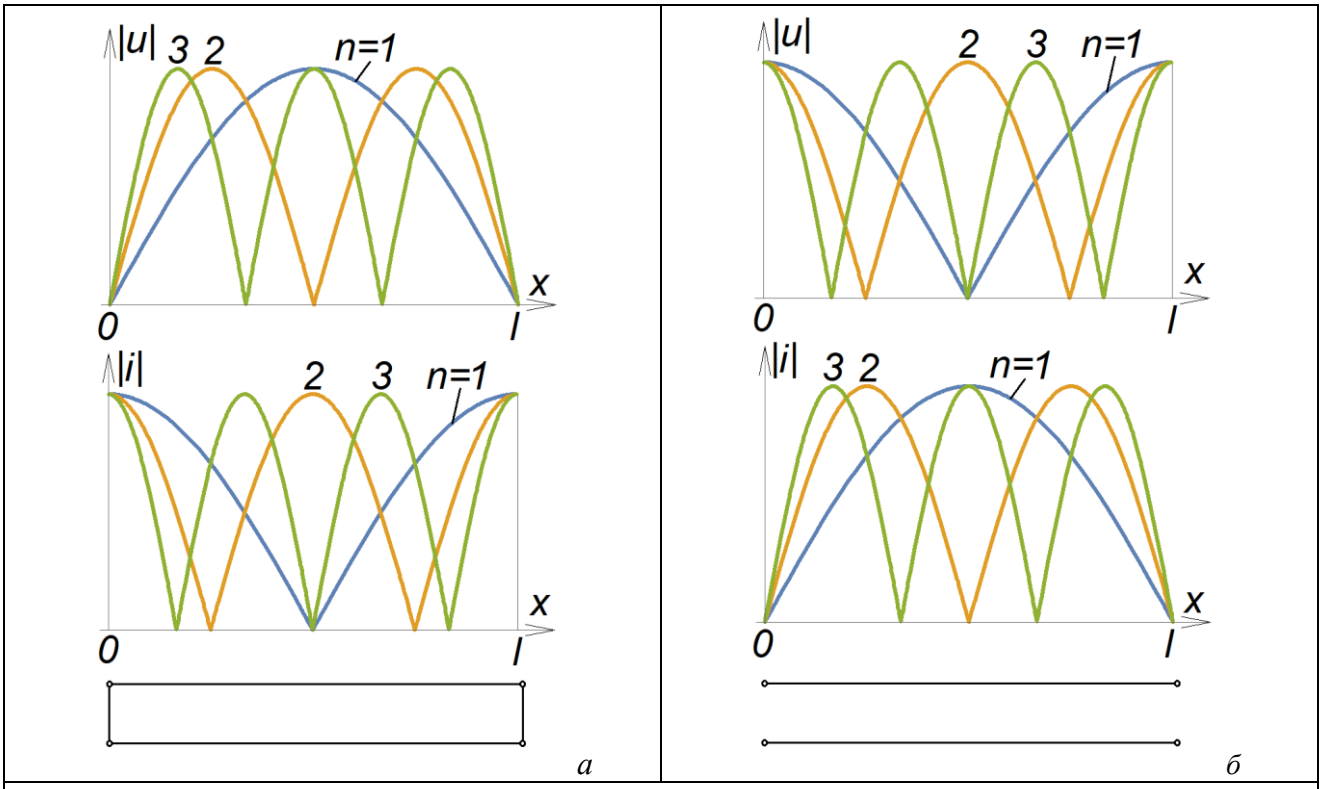


Рис. 6.7.3. Розподіли напруги й струму вздовж відрізку лінії із короткозамкненими (а) та розімкненими (б) кінцями

Як і на короткозамкненому кінці лінії з джерелом сигналу, в аналізованому випадку на кінцях відрізку довгої лінії бачимо пучності струму й нулі напруги.

На рис. 6.7.3 а подані модулі розподілу амплітуд напруги й струму. Але безпосередньо з формул (6.7.13)-(6.7.14) випливає, що різких просторових змін цих амплітуд не відбувається. Тому в мінімумах на рис. 6.7.3 а насправді відбувається зміна знаку комплексної амплітуди. Це, свою чергу, означає, що моди з парними номерами характеризуються симетричними щодо центра відрізку розподілами струму й напруги, а моди з непарними номерами – відповідно антисиметричними розподілами. Таким чином, усі моди в даному разі можна розділити на симетричні (з парними n) і антисиметричні (відповідно з непарними n)[2].

6.7.3. Відрізок лінії з розімкнутими кінцями

Для відрізку лінії з розімкнутими кінцями $X_1 = X_2 = \infty$ і, відповідно, $\Phi_1 = \Phi_2 = \pi/2$. Оскільки в цьому випадку $\Phi_1 - \Phi_2 = 0$ (як і в лінії з короткозамкненими кінцями), формули (6.7.11)-(6.7.12) для власних частот та довжин хвиль системи зберігаються.

Просторові розподіли напруги та струму вздовж відрізка ліній в порівнянні з попереднім випадком міняються місцями: на кінцях лінії тепер спостерігаються вузли струму й пучності напруги (рис. 3.7.3 б). Як і раніше, парним значенням n відповідають симетричні моди, непарним – антисиметричні.

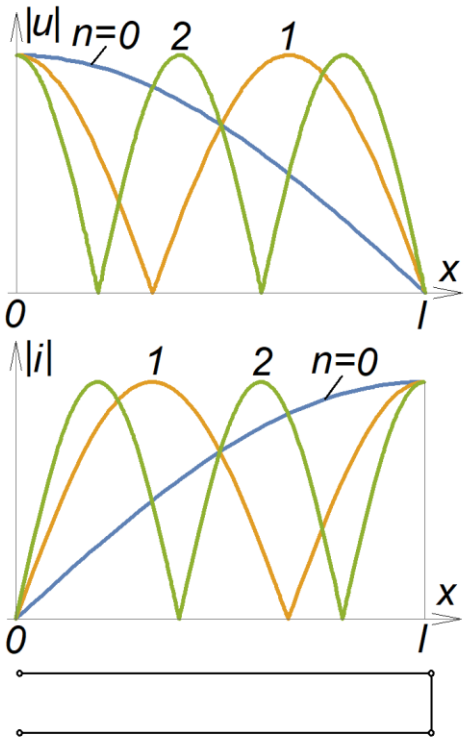


Рис. 6.7.4. Розподіли напруги й струму вздовж відрізку лінії, у якої один кінець розімкнений, а інший замкнений накоротко

6.7.4. Відрізок лінії, у якого один кінець короткозамкнений, а другий розімкнутий

Для повноти картини розглянемо ще відрізок лінії, у якого один кінець замкнений накоротко, а інший розімкнутий. У цьому випадку $\Phi_1 = \pi/2$, $\Phi_2 = 0$. Формула для власних значень хвильових чисел (6.7.4) набуває вигляду:

$$k_n = (\pi/l)(n + 1/2). \qquad (6.7.15)$$

Відповідно

$$\lambda_n = 2l/(n + 1/2), \qquad (6.7.16)$$

[2] Цей висновок важливий при дослідженні вимушених коливань (зокрема, резонансу) в такому резонаторі. Скажемо, джерелом напруги, розміщеним у центрі відрізку лінії, неможливо збудити антисиметричні моди.

$$f_n = c_0 k_n / 2\pi = \left(c_0 / l\right)\left(n/2 + 1/4\right), \tag{6.7.17}$$

де тепер $n = 0,1,2,3,...$ Формулу (6.7.16) можна перетворити: $l = \lambda_n \left(n/2 + 1/4\right)$, тобто тепер на довжині відрізку вкладається ціле число півхвиль і ще чверть хвилі. Мінімальна частота досягається при $n = 0$: $f_{min} = f_0 = c_0 / 4l$.

Очевидно, на розімкненому кінці спостерігається вузол струму й пучність напруги, на короткозамкненому – навпаки, вузол напруги й пучність струму (рис. 6.7.4).

Оскільки аналізована модель несиметрична, говорити про симетричні чи антисиметричні моди тут не доводиться.

6.7.5. Відрізок лінії з реактивними елементами на обох кінцях

Повернемося знов до загального випадку, який розглядався в п. 6.7.1. Перш за все відзначимо, що для такої системи можна ввести ефективну довжину відрізку лінії, як це було зроблено в п. 6.4.3 (див. формулу (6.4.25):

$$l_{eff} = l + \frac{\Phi_2 - \Phi_1}{k} = l + \frac{1}{k}\left(arctg\frac{X_2}{Z_0} - arctg\frac{X_1}{Z_0}\right) = l + \frac{\lambda}{2\pi}\left(arctg\frac{X_2}{Z_0} - arctg\frac{X_1}{Z_0}\right). \tag{6.7.18}$$

Як відзначалося наприкінці п. 6.5.2, зміна ефективної довжини лінії зумовлена тим, що ми заміняємо імпеданси реактивних елементів на її кінцях короткозамкненими відрізками такої самої лінії. Тоді власні частоти й довжини хвиль для такого відрізку можна знайти за формулами (6.7.11)-(6.7.12), де треба замінити l на l_{eff}. Очевидно, на кінцях такого відрізку лінії не буде ні пучностей, ні вузлів стоячої хвилі.

Варто ще відзначити, що коли реактивні елементи на обох кінцях лінії однакові, то $\Phi_1 = \Phi_2$ і $l_{eff} = l$. В цьому частинному випадку знову можна говорити про симетричні та антисиметричні моди.

6.7.6. Вплив слабкої дисипації

Якщо дисипація в лінії слабка, то для власних частот можна користуватися формулами, отриманими без її врахування. Кількісною мірою впливу дисипації може бути, зокрема, добротність.

Для розрахунку добротності відрізку лінії скористаємося загальною формулою, справедливою для довільних коливних систем із малими втратами:

$$Q = \omega W / P, \tag{6.7.19}$$

де W – накопичена в коливній системі енергія, P – середня потужність втрат.

В елементі лінії завдовжки dx накопичена енергія може бути подана у вигляді

$$dW = \frac{1}{2}\left(L_1 dx\right) I_m^2\left(x\right), \tag{6.7.20}$$

а середня потужність втрат – у вигляді

$$dP = \frac{1}{2}\left(r_1 dx\right) I_m^2\left(x\right). \tag{6.7.21}$$

Інтегруючи співвідношення (6.7.20)-(6.7.21) по x від 0 до l, маємо:

$$W = \int_0^l dW\left(x\right) = \frac{1}{2} L_1 \int_0^l dx I_m^2\left(x\right), \tag{6.7.22}$$

$$P = \int_0^l dP\left(x\right) = \frac{1}{2} r_1 \int_0^l dx I_m^2\left(x\right). \tag{6.7.23}$$

Підставляючи (6.7.22)-(6.7.23) до (6.7.19) і враховуючи, що під ω слід розуміти одну з власних частот ω_n, отримуємо добротність моди з номером *n*:

$$Q_n = \frac{\omega_n L_1}{r_1}. \tag{6.7.24}$$

Взявши до уваги формули (6.3.16)

$$r_1 \approx \frac{1}{\pi a}\sqrt{\frac{\omega\mu_0}{2\sigma}},$$

остаточно отримуємо (в системі СІ):

$$Q_n = \frac{\omega_n L_1}{r_1} = \pi a L_1 \sqrt{\frac{2\sigma\omega_n}{\mu_0}} = \pi a L_1 \sqrt{\frac{n\pi\sigma}{\mu_0 l}\frac{1}{\sqrt{L_1 C_1}}} = \pi^{3/2} a \sqrt{\frac{n\sigma L_1 Z_0}{\mu_0 l}} \sim \sqrt{\frac{n}{l}}. \qquad (6.7.32)$$

Видно, що добротність зменшується зі зростанням довжини лінії й зростає зі збільшенням номеру моди (або її частоти). Типова добротність відрізку двопровідної лінії складає величину порядку 10^2.

Контрольні питання до підрозділу 6.7

1. Порівняйте резонансні властивості коливного контуру та відрізку лінії передачі.
2. Поясніть фізичний зміст ефективної довжини лінії.
3. Чому при однакових реактивних елементах, приєднаних до кінців відрізку лінії, спектр власних частот такий самий, як у лінії з короткозамкненими кінцями?

Задачі до підрозділу 6.7

6.7.1. Резонатор являє собою ділянку двопровідної лінії завдовжки 2 м. Швидкість хвиль у лінії – $1.2\cdot10^8$ м/с. Знайдіть три найнижчі власні частоти такого резонатора.

6.7.2. У відрізка довгої лінії, описаної в задачі 6.3.1, один кінець розімкнений, а інший замкнений накоротко. Знайти мінімальну власну частоту цього відрізка. Дисипацією знехтувати.

6.8. Хвилеводи

Досі в цій частині курсу йшлося про двопровідні лінії передач. Для кращого розуміння матеріалу наступної частини курсу, присвяченої антенам, бажано також мати уявлення про хвилеводи, які використовуються для спрямованої передачі сигналів у дециметровому та сантиметровому діапазонах довжин хвиль. Конструктивно хвилевід являє собою трубку з провідними стінками, від яких відбиваються хвилі. За рахунок відбиття від стінок енергія хвилі, що передається, зосереджена всередині хвилеводу. Ми приділимо основну увагу прямокутним хвилеводам (рис. 6.8.1 а), але коротко обговоримо також циліндричні хвилеводи (рис. 6.8.1 б).

Крім хвилеводів з провідними стінками, використовуються також діелектричні хвилеводи, виготовлені з матеріалу з великою діелектричною проникністю. Хвилі, що поширюються в таких хвилеводах, на їхніх стінках зазнають повного внутрішнього відбиття. До цього класу можна зарахувати, зокрема, оптичні волокна. В цьому курсі діелектричні хвилеводи ми не розглядатимемо.

Перш ніж безпосередньо розглядати хвилеводи з провідними стінками, слід згадати деякі поняття з електродинаміки та теорії електромагнітних хвиль.

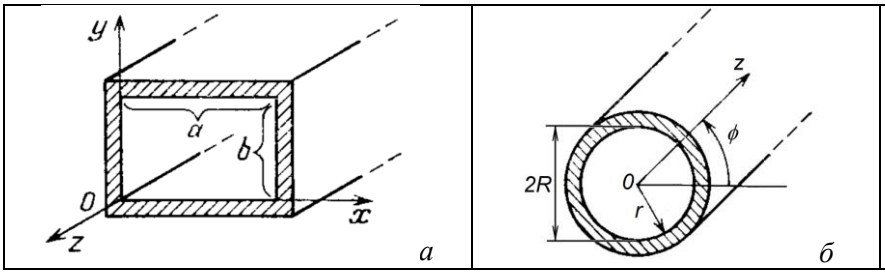


Рис. 6.8.1. Прямокутний (а) та циліндричний (б) хвилеводи

6.8.1. Диференціальні оператори

Перш за все, нагадаємо, що таке диференціальні оператори, які надалі використовуються для опису електромагнітного поля.

Оператор Гамільтона ∇ (читається «набла», від давньогрецького νάβλα – давньоєгипетська трикутна арфа) в декартових координатах має вигляд

$$\nabla = \vec{e}_x \frac{\partial}{\partial x} + \vec{e}_y \frac{\partial}{\partial y} + \vec{e}_z \frac{\partial}{\partial z}, \tag{6.8.1}$$

де $\vec{e}_x, \vec{e}_y, \vec{e}_z$ – одиничні вектори, паралельні відповідним осям.

Результат дії оператора Гамільтона на скалярну функцію $f(x,y,z)$

$$\nabla f = grad\ f = \vec{e}_x \frac{\partial f}{\partial x} + \vec{e}_y \frac{\partial f}{\partial y} + \vec{e}_z \frac{\partial f}{\partial z} \tag{6.8.2}$$

називають градієнтом (від лат. gradiens – той, що крокує, або той, що зростає) цієї функції. Градієнт функції $f(x,y,z)$ – це вектор, напрямок якого збігається з напрямком її найшвидшого зростання, а його величина визначається швидкістю зростання цієї функції в просторі. Зокрема, якщо функція $f(x)$ залежить тільки від координати x, вектор ∇f паралельний до осі x.

Скалярний добуток оператора ∇ на векторну функцію $\vec{f}(x,y,z) = \vec{e}_x f_x + \vec{e}_y f_y + \vec{e}_z f_z$

$$\left(\nabla \cdot \vec{f}\right) = div\ f = \frac{\partial f_x}{\partial x} + \frac{\partial f_y}{\partial y} + \frac{\partial f_z}{\partial z} \tag{6.8.3}$$

називають дивергенцією ((від лат. divergere – розходження, розбіжність) цієї функції.

Фізичний зміст дивергенції зручно пояснити на прикладі течії нестисливої рідини, вважаючи, що $\vec{f}(x,y,z)$ – поле течії цієї рідини (іншими словами, $\vec{f}(x,y,z)$ – це вектор швидкості рідини в точці з координатами (x,y,z)). Якщо в деякій області витоки (джерела рідини) та стоки (області, де рідина зникає) відсутні, там $\left(\nabla \cdot \vec{f}\right) = 0$. В області існування ви-

токів $\left(\nabla \cdot \vec{f}\right) > 0$, в області існування стоків, навпаки, $\left(\nabla \cdot \vec{f}\right) < 0$.

Векторний добуток оператора ∇ на векторну функцію $\vec{f}(x,y,z)$

$$\left[\nabla \times \vec{f}\right] = rot\ f = \begin{bmatrix} \vec{e}_x & \vec{e}_y & \vec{e}_z \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ f_x & f_y & f_z \end{bmatrix} = \begin{pmatrix} \partial f_z/\partial y - \partial f_y/\partial z \\ \partial f_x/\partial z - \partial f_z/\partial x \\ \partial f_y/\partial x - \partial f_x/\partial y \end{pmatrix} =$$

$$= \vec{e}_x\left(\frac{\partial f_z}{\partial y} - \frac{\partial f_y}{\partial z}\right) + \vec{e}_x\left(\frac{\partial f_x}{\partial z} - \frac{\partial f_z}{\partial x}\right) + \left(\frac{\partial f_y}{\partial x} - \frac{\partial f_x}{\partial y}\right) \qquad (6.8.4)$$

називають ротором (від лат. roto – обертатися, rotor – вихор) цієї функції.

Для пояснення змісту ротору знову звернемося до течії нестисливої рідини. Якщо в деякій області лінії течії не утворюють замкнених кривих (вихорів), то там $\left[\nabla \times \vec{f}\right] = 0$, в протилежному випадку $\left[\nabla \times \vec{f}\right] \neq 0$. В частинному випадку, коли циліндричний об'єм рідини обертається як єдине ціле навколо осі, то $\left[\nabla \times \vec{f}\right] = 2\vec{\Omega}$, де величина вектора $\vec{\Omega}$ – це кутова швидкість обертання, а його напрямок збігається з віссю обертання (якщо дивитися з кінця цієї осі, обертання відбувається за годинниковою стрілкою). Очевидно, лінії течії в цьому випадку мають форму кіл, площини розташування яких перпендикулярні до осі обертання, а центри лежать на цій осі.

Можна показати, що

$$rot(grad\ f) = \left[\nabla \times \nabla f\right] = 0, \qquad (6.8.5)$$

$$div\left(rot\ \vec{f}\right) = \left(\nabla \cdot \left[\nabla \times \vec{f}\right]\right) = 0, \qquad (6.8.6)$$

$$div(grad\ f) = (\nabla \cdot \nabla f) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \equiv \Delta f, \qquad (6.8.7)$$

де

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \qquad (6.8.8)$$

– оператор Лапласа. Цей оператор може діяти як на скалярні, так і на векторні функції. Зокрема, виконується співвідношення

$$rot(rot\ \vec{f}) = \nabla \times \nabla \times \vec{f} = \nabla\left(\nabla \cdot \vec{f}\right) - \Delta\vec{f}. \qquad (6.8.9)$$

6.8.2. Характеристики електромагнітного поля та рівняння Максвелла

З курсу вищої математики відомо, що будь-яке поле, зокрема, електричне поле $\vec{E}(\vec{r},t)$ ($\vec{E}$ – напруженість електричного поля), може бути подане як сума потенціальної і вихрової складових

$$\vec{E}(\vec{r},t) = \vec{E}_{pot}(\vec{r},t) + \vec{E}_{vor}(\vec{r},t) \qquad (6.8.10)$$

(індекси – від англ. potential та vortex – відповідно потенціальний і вихровий).

Силові лінії потенціального електричного поля починаються й закінчуються на електричних зарядах. Його завжди можна представити у формі

$$\vec{E}_{pot} = -\nabla\varphi \quad , \qquad (6.8.11)$$

де φ – електростатичний потенціал. Для таких полів, відповідно до формули (6.8.5),

$$\nabla \times \vec{E}_{pot} = 0. \qquad (6.8.12)$$

Силові лінії вихрових полів є замкненими. Вихрові поля (зокрема, й магнітне поле $\vec{H}(\vec{r},t)$, яке є суто вихровим; $\vec{H}$ – напруженість магнітного поля) завжди можуть бути подані у вигляді

$$\vec{H}(\vec{r},t) = \nabla \times \vec{A}. \qquad (6.8.13)$$

де $\vec{A}$ – векторний потенціал. Для таких полів, відповідно до формули (6.8.6),

$$\nabla \cdot \vec{H}(\vec{r},t) = 0 \cdot \qquad (6.8.14)$$

Як електричне, так і магнітне поле характеризується не тільки напруженістю, а й індукцією.

Напруженість електричного поля $\vec{E}$ характеризує його силову дію:

$$\vec{F} = q\vec{E} \cdot \qquad (6.8.15)$$

Індукція електричного поля $\vec{D}$ враховує вплив середовища: $\vec{D} = \varepsilon\vec{E}$, ε – (відносна) діелектрична проникність середовища.

Силову дію магнітного поля визначає його індукція $\vec{B}$:

$$\vec{F} = \frac{q}{c}\left[\vec{\upsilon} \times \vec{B}\right]. \qquad (6.8.16)$$

Згадана раніше напруженість магнітного поля $\vec{H}$ визначає вплив середовища: $\vec{B} = \mu\vec{H}$, де μ – (відносна) магнітна проникність середовища.

Саме величини ε і μ характеризують електромагнітні властивості середовища. Строго кажучи, їх можна ввести лише тоді, коли електрична та магнітне поля мають вигляд гармонічних хвиль. У деяких середовищах ε і μ можуть бути не скалярами, а тензорами. В цьому випадку напруженість відрізняється від індукції не тільки величиною, а й напрямком. Нас будуть цікавити високочастотні діелектрична та магнітна проникності (для змінних полів), які можуть відрізнятись від проникностей для постійних полів.

Система рівнянь Максвелла, яка дозволяє розрахувати електромагнітне поле за відомим розподілом густин заряду $\rho(\vec{r},t)$ і струму $\vec{j}(\vec{r},t)$, складається з чотирьох диференціальних рівнянь першого порядку в частинних похідних (тут і далі ми користуємося системою одиниць Гаусса):

$$\nabla \times \vec{H} = \frac{4\pi}{c}\vec{j} + \frac{1}{c}\frac{\partial \vec{D}}{\partial t} \qquad (6.8.17)$$

(закон Біо-Савара-Лапласа з урахуванням струму зміщення; він показує, що джерелом магнітного поля є густина струму провідності $\vec{j}$ та/або струму зміщення $\partial\vec{D}/\partial t$)

$$\nabla \times \vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t} \qquad (6.8.18)$$

(закон електромагнітної індукції; він показує, що джерелом вихрового електричного поля є змінне магнітне поле);

$$\nabla \cdot \vec{D} = 4\pi\rho \qquad (6.8.19)$$

(теорема Гаусса для електричного поля; вона показує, що джерелом потенціального електричного поля є електричні заряди);

$$\nabla \cdot \vec{B} = 0 \qquad (6.8.20)$$

(теорема Гаусса для магнітного поля; вона показує, що магнітне поле не має потенціальної компоненти).

У цьому пункті нас цікавлять електромагнітні поля у вакуумі. Тому надалі вважатимемо, що $\varepsilon = 1$ і $\mu = 1$, тобто $\vec{D} = \vec{E}$ і $\vec{B} = \vec{H}$. Через те надалі ми будемо користуватися лише напруженостями полів – електричного $\vec{E}$ та магнітного $\vec{H}$.

6.8.3. Хвильове та дисперсійне рівняння для електромагнітних хвиль у вакуумі

У вакуумі за відсутності струмів провідності та з урахуванням сказаного наприкінці п. 6.8.2 рівняння (6.8.17) набуває вигляду:

$$\nabla \times \vec{H} = \frac{1}{c}\frac{\partial \vec{E}}{\partial t}. \qquad (6.8.21)$$

Візьмемо ротор обох частин рівняння (6.8.17) і виключимо з нього магнітне поле за допомогою (6.8.20). Маємо:

$$\nabla\times\nabla\times\vec{E}=-\frac{1}{c}\frac{\partial}{\partial t}\nabla\times\vec{H}=-\frac{1}{c}\frac{\partial}{\partial t}\left(\frac{1}{c}\frac{\partial\vec{E}}{\partial t}\right)=-\frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2}. \qquad (6.8.22)$$

Врахуємо співвідношення (6.8.9):

$$\nabla\times\nabla\times\vec{f}=\nabla\left(\nabla\cdot\vec{f}\right)-\Delta\vec{f}.$$

Візьмемо також до уваги, що у вакуумі за відсутності об'ємних зарядів теорема Гаусса для електричного поля (6.8.19) набуває вигляду

$$\nabla\cdot\vec{E}=0 \qquad (6.8.23)$$

(електричне поле не має потенціальної компоненти і є суто вихровим). Тому рівняння (6.8.22) набуває вигляду

$$\Delta\vec{E}-\frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2}=0. \qquad (6.8.24)$$

Надалі при розгляді прямокутних хвилеводів ми будемо використовувати систему декартових координат, показану на рис. 6.8.1. У цій системі рівняння (6.8.24) можна записати у формі (див. вище формулу (6.8.8))

$$\frac{\partial^2\vec{E}}{\partial x^2}+\frac{\partial^2\vec{E}}{\partial y^2}+\frac{\partial^2\vec{E}}{\partial z^2}-\frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2}=0. \qquad (6.8.25)$$

Рівняння, аналогічні (6.8.24)-(6.8.25), легко отримати й для магнітного поля.

Будемо шукати розв'язок рівняння (6.8.24) (або (6.8.25)) у вигляді плоскої гармонічної хвилі:

$$\vec{E}(\vec{r},t)=\vec{E}_m\exp\left(i\omega t-i\vec{k}\vec{r}\right)=\vec{E}_m\exp\left(i\omega t-ik_x x-ik_y y-ik_z z\right), \qquad (6.8.26)$$

$$k=\left|\vec{k}\right|=\sqrt{k_x^2+k_y^2+k_z^2}. \qquad (6.8.27)$$

Підставивши (6.8.26) до (6.8.25), отримуємо дисперсійне рівняння для електромагнітних хвиль у вакуумі (пор. з п. 6.1):

$$k^2=k_0^2, \qquad k_0\equiv\frac{\omega}{c}. \qquad (6.8.28)$$

Бачимо, що дисперсія для електромагнітних хвиль у вакуумі відсутня: їхня фазова й групова швидкості незалежно від частоти дорівнюють c – швидкості світла у вакуумі (пор. з формулою (6.1.9)).

6.8.4. Поляризація плоских електромагнітних хвиль у вакуумі

Зупинимося детальніше на властивостях плоскої гармонічної електромагнітної хвилі вигляду (6.8.25), що поширюється у вакуумі. Нехай така хвиля поширюється паралельно осі z:

$$\vec{E}(\vec{r},t)=\vec{E}_m\exp\left(i\omega t-ikz\right). \qquad (6.8.29)$$

Підстановка розв'язку (6.8.29) до теореми Гаусса (6.8.23) дає, що $kE_{mz}=0$, або $E_{mz}=0$, тобто плоска електромагнітна хвиля не має поздовжньої компоненти електричного поля і, таким чином, є поперечною. Компоненти поперечного електричного поля E_{mx} та E_{my} між собою не пов'язані.

Підстановка цього самого розв'язку до рівняння (6.8.18) дозволяє знайти магнітне поле хвилі:

$$-\frac{i\omega}{c}\vec{H}=-i\vec{k}\times\vec{E}=-i\begin{vmatrix}\vec{e}_x & \vec{e}_y & \vec{e}_z\\ 0 & 0 & k\\ E_x & E_y & 0\end{vmatrix}=-ik\begin{pmatrix}-E_y\\ E_x\\ 0\end{pmatrix}, \qquad (6.8.30)$$

або, з урахуванням формул (6.8.28), $H_{mx}=-E_{my}$, $H_{my}=-E_{mx}$.

Хвилі, у яких $E_{mx}\neq 0, E_{my}=0$ та $E_{mx}=0, E_{my}\neq 0$ утворюють пару нормальних хвиль, які поширюються незалежно одна від одної. Будь-яку іншу хвилю, яка поширюється вздовж осі z, можна подати у вигляді суперпозиції таких нормальних хвиль.

На рис. 6.8.2 а показано миттєвий розподіл компонент електричного поля для хвилі, у якої $E_{mx} = 0, E_{my} \neq 0$.

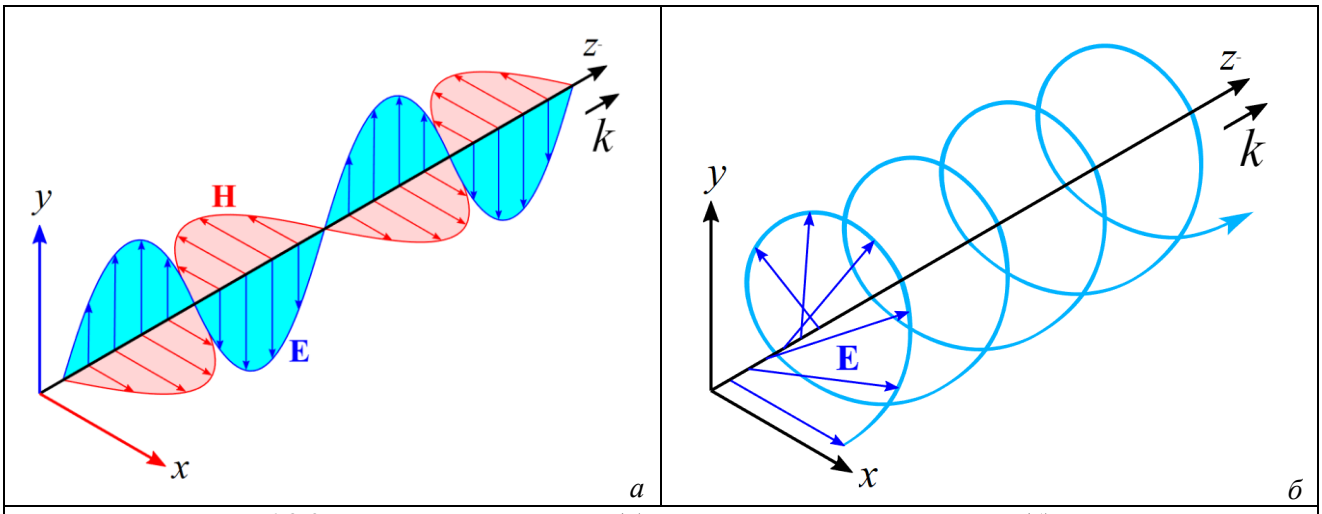


Рис. 6.8.2. Лінійно поляризована (а) та циркулярно поляризована (б) хвилі (остання – з лівою поляризацією)

Поляризація електромагнітної хвилі визначається законом зміни напрямку вектора напруженості електричного поля в даній точці простору з часом. Якщо згаданий вектор залишається паралельним (чи антипаралельним) до свого початкового напрямку, говорять про лінійну (плоску) поляризацію. Площина, яка вміщує вектор напруженості електричного поля і хвильовий вектор, називається площиною поляризації такої хвилі (рис. 6.8.2 а). Очевидно, обидві розглянуті раніше хвилі ($E_{mx} \neq 0, E_{my} = 0$ та $E_{mx} = 0, E_{my} \neq 0$) є плоскополяризованими.

Розглянемо дві нормальні лінійно поляризовані хвилі, зсунуті за фазою на $\pi/2$. Їхнє сумарне вимірюване електричне поле в деякій фіксованій точці простору можна представити у вигляді

$$E_x(t) = E_{mx}\cos\omega t, \qquad E_y(t) = E_{my}\sin\omega t. \tag{6.8.31}$$

В площині $\{E_x, E_y\}$ цьому полю відповідає деякий вектор змінної довжини, кінець якого обертається вздовж замкненої кривої. Рівняння цієї кривої можна легко отримати зі співвідношень (6.8.30):

$$\left(\frac{E_x}{E_{mx}}\right)^2 + \left(\frac{E_y}{E_{my}}\right)^2 = 1. \tag{6.8.32}$$

Ця крива, очевидно, є еліпсом із півосями E_{mx}, E_{my} (рис. 6.8.3).

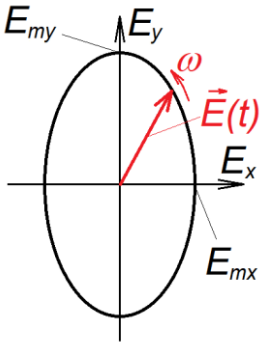


Рис. 6.8.3. Зміна електричного поля з часом для еліптично поляризованої хвилі, що поширюється вздовж осі z

Якщо $E_{mx} \neq E_{my}$, говорять, що така хвиля є еліптично поляризованою.

Якщо ж $E_{mx} = E_{my}$, говорять про хвилю з циркулярною (круговою) поляризацією (рис. 6.8.2 б).

В залежності від напрямку обертання вектора $\vec{E}$ говорять про праву і ліву поляризацію хвилі. Якщо дивитися вздовж напрямку хвильового вектора, в першому випадку вектор напруженості електричного поля обертається праворуч (за годинниковою стрілкою), в другому – ліворуч (проти годинникової стрілки).

Лінійну та циркулярну поляризацію можна розглядати як окремі випадки еліптичної. Як уже вказувалося, при $E_{mx} = E_{my}$ маємо циркулярно поляризовану хвилю, якщо ж $E_{mx} = 0$ (або $E_{my} = 0$) – лінійно поляризовану.

6.8.5. Граничні умови для електромагнітного поля на поверхні ідеального провідника

З курсу електродинаміки відомо, що на різкій межі поділу двох однорідних середовищ 1 і 2 для нормальної й тангенціальної компонент електромагнітного поля (рис. 6.8.4) справджуються такі граничні умови, що безпосередньо випливають з рівнянь Максвелла:

$$\vec{E}_{\tau 1} - \vec{E}_{\tau 2} = 0, \quad D_{n1} - D_{n2} = 4\pi\sigma, \quad H_{n1} - H_{n2} = 0, \quad \vec{n} \times \left(\vec{H}_{\tau 1} - \vec{H}_{\tau 2}\right) = 4\pi \vec{j}_{surf}, \qquad (6.8.33)$$

де індекси n і τ позначають нормальні та тангенціальні складові[3] (рис. 6.8.4) відповідних векторів, σ та $\vec{j}_{surf}$ – густини поверхневих зарядів і струмів на межі поділу. Тут вектор нормалі $\vec{n}$ спрямований у бік середовища 1 (рис. 6.8.4).

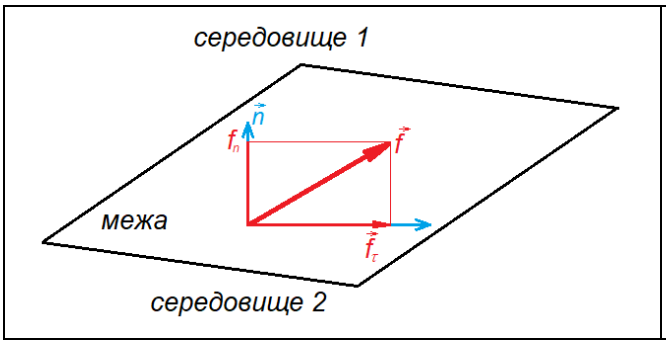


Рис. 6.8.4. Нормальна й тангенціальна складові вектора на межі двох середовищ

Нехай тепер середовище 1 – це вакуум, а середовище 2 – ідеальний провідник. Як відомо, електричне поле в ідеальному провіднику має дорівнювати нулеві, оскільки інакше воно породжувало б нескінченно великий струм. Те саме стосується змінного магнітного поля, оскільки воно є джерелом вихрового електричного поля. Тоді в аналізованому випадку граничні умови для змінного електромагнітного поля на поверхні ідеального провідника набувають вигляду (індекс, що вказує на номер середовища, опускаємо, оскільки поле є лише у вакуумі):

$$\vec{E}_\tau = 0, \qquad E_n = 4\pi\sigma, \qquad H_n = 0, \qquad \vec{n} \times \vec{H}_\tau = 4\pi \vec{j}_{surf}. \qquad (6.8.34)$$

Таким чином, на поверхні ідеального провідника тангенціальна компонента електричного поля та нормальна компонента магнітного поля відсутні, а нормальна компонента електричного поля та тангенціальна компонента магнітного поля породжуються відповідно поверхневими зарядами з густиною σ та поверхневими струмами з густиною $\vec{j}_{surf}$ у провіднику.

В реальних провідниках заряди та струми зосереджені у приповерхневому шарі скінченної товщини. Протікання струму породжує в цьому випадку омічні втрати, що призводить до дисипації електромагнітного поля. Для зниження цих втрат хвилеводи виготовляють зазвичай з міді, яка є гарним провідником, а їхню внутрішню поверхню інколи вкривають тонким ша-

[3] В кожній точці криволінійної поверхні можна побудувати дотичну до неї площину. Тангенціальна складова вектора лежить у цій площині, нормальна – перпендикулярна до неї.

ром срібла чи навіть золота.

6.8.6. TE- та TM-хвилі

Хвиля в хвилеводі (рис. 6.8.1 а) поширюється в напрямку осі z. Тому шукатимемо розв'язки рівнянь Максвелла для хвилеводу у формі гармонічної біжучої хвилі в напрямку z:

$$\vec{E}(\vec{r},t)=\vec{E}_m(x,y)\exp(i\omega t-ik_z z), \qquad \vec{H}(\vec{r},t)=\vec{H}_m(x,y)\exp(i\omega t-ik_z z). \tag{6.8.35}$$

Підставимо ці розв'язки до рівнянь (6.8.18) і (6.8.21) з урахуванням того, що $\varepsilon=1$ і $\mu=1$, і спроєктуємо отримані рівняння на осі координат. Маємо:

$$\begin{aligned}
&\frac{\partial}{\partial y}E_z+ik_zE_y=-ik_0H_x\\
&-ik_zE_x-\frac{\partial}{\partial x}E_z=-ik_0H_y\\
&\frac{\partial}{\partial x}E_y-\frac{\partial}{\partial y}E_x=-ik_0H_z
\end{aligned} \tag{6.8.36}$$

$$\begin{aligned}
&\frac{\partial}{\partial y}H_z+ik_zH_y=ik_0E_x\\
&-ik_zH_x-\frac{\partial}{\partial x}H_z=ik_0E_y\\
&\frac{\partial}{\partial x}H_y-\frac{\partial}{\partial y}H_x=ik_0E_z
\end{aligned} \tag{6.8.37}$$

($k_0=\omega/c$). Перенесемо в рівняннях (6.8.36)-(6.8.37) поздовжні компоненти поля праворуч, а поперечні – ліворуч:

$$\begin{aligned}
&+ik_zE_y+ik_0H_x=-\frac{\partial}{\partial y}E_z\\
&-ik_zE_x+ik_0H_y=\frac{\partial}{\partial x}E_z\\
&-\frac{\partial}{\partial y}H_x+\frac{\partial}{\partial x}H_y=ik_0E_z\\
&-ik_0E_x+ik_zH_y=-\frac{\partial}{\partial y}H_z\\
&-ik_0E_y-ik_zH_x=\frac{\partial}{\partial x}H_z\\
&-\frac{\partial}{\partial y}E_x+\frac{\partial}{\partial x}E_y=-ik_0H_z
\end{aligned} \tag{6.8.38}$$

Перше, друге, четверте й п'яте рівняння цієї системи утворюють набір лінійних неоднорідних алгебраїчних рівнянь щодо поперечних компонент електромагнітного поля, а в правій частині стоять похідні від поздовжніх компонент. З неї можна отримати:

$$\begin{aligned}
&E_x=\frac{i}{k_z^2-k_0^2}\left(k_z\frac{\partial E_z}{\partial x}+k_0\frac{\partial H_z}{\partial y}\right);\\
&E_y=\frac{i}{k_z^2-k_0^2}\left(k_z\frac{\partial E_z}{\partial y}-k_0\frac{\partial H_z}{\partial x}\right);\\
&H_x=-\frac{i}{k_z^2-k_0^2}\left(k_0\frac{\partial E_z}{\partial y}-k_z\frac{\partial H_z}{\partial x}\right);\\
&H_y=\frac{i}{k_z^2-k_0^2}\left(k_0\frac{\partial E_z}{\partial x}+k_z\frac{\partial H_z}{\partial y}\right).
\end{aligned} \tag{6.8.39}$$

Підстановка співвідношень (6.8.39) до третього й шостого рівнянь системи (6.8.38) дає хвильові рівняння для поздовжніх полів у формі

$$\frac{\partial^2 E_z}{\partial y^2}+\frac{\partial^2 E_z}{\partial x^2}-k_z^2 E_z+k_0^2 E_z=0\,; \tag{6.8.40}$$

$$\frac{\partial^2 H_z}{\partial^2 y}+\frac{\partial^2 H_z}{\partial^2 x}-k_z^2 H_z+k_0^2 H_z=0\,. \tag{6.8.41}$$

Перше з цих співвідношень можна отримати безпосередньо з рівняння (6.8.25), підставивши розв'язок у формі (6.8.35).

Рівняння (6.8.40) і (6.8.41) незалежні, тому їх можна розв'язувати окремо.

Якщо $E_z \neq 0$, а $H_z = 0$, отримаємо хвилю типу E, або TM (трансверсально-магнітну[4]).

Якщо, навпаки, $E_z = 0$, а $H_z \neq 0$, отримаємо хвилю типу H, або TE (трансверсально-електричну).

З незалежності рівнянь (6.8.40) і (6.8.41) випливає, що TE- і TM-хвилі можуть поширюватися в системі незалежно одна від одної.

Бачимо, що, на відміну від гармонічних хвиль у вакуумі, хвилі в хвилеводі не є суто поперечними, а мають також поздовжню компоненту поля – електричного або магнітного.

6.8.7. TM-моди прямокутного хвилеводу

Почнемо з рівняння (6.8.40), яке описує TM-хвилі.

Припустимо, що функція $E_z(x,y)$ може бути подана у вигляді двох функцій, одна з яких залежить лише від x, а інша – лише від y:

$$E_z(x,y)=f_1(x)f_2(y)\,. \tag{6.8.42}$$

Підставимо цей розв'язок до рівняння (6.8.40), після чого поділимо всі доданки на $f_1(x)f_2(y)$. Отримаємо:

$$\frac{1}{f_1}\frac{d^2 f_1}{dx^2}+\frac{1}{f_2}\frac{d^2 f_2}{dy^2}=k_z^2-k_0^2\,. \tag{6.8.43}$$

Перший доданок у лівій частині може залежати лише від x, другий – лише від y, права частина рівняння – стала. Очевидно, рівняння можна задовольнити лише тоді, коли обидва доданки в лівій частині також є сталими:

$$\frac{1}{f_1}\frac{d^2 f_1}{dx^2}=-k_x^2, \qquad \frac{1}{f_2}\frac{d^2 f_2}{dy^2}=-k_y^2 \tag{6.8.44}$$

(далі ми побачимо, що для додатних констант у правих частинах неможливо задовольнити граничні умови). Розв'язуючи ці рівняння і підставляючи результат до (6.8.42), врешті отримаємо:

$$E_z(x,y)=\left[A_1\sin(k_x x)+A_2\cos(k_x x)\right]\left[B_1\sin(k_y y)+B_2\cos(k_y y)\right]. \tag{6.8.45}$$

Підставивши (6.8.45) до (6.8.35), маємо остаточний вираз для поздовжнього поля TM-хвилі:

$$E_z(\vec{r},t)=\left[A_1\sin(k_x x)+A_2\cos(k_x x)\right]\left[B_1\sin(k_y y)+B_2\cos(k_y y)\right]\exp(i\omega t-ik_z z)\,. \tag{6.8.46}$$

Підстановка (6.8.46) до (6.8.40) дає дисперсійне співвідношення у формі:

$$k_x^2+k_y^2+k_z^2=k_0^2\,, \tag{6.8.47}$$

де, нагадаємо, $k_0=\omega/c$. Іншими словами, дисперсійне співвідношення пов'язує хвильове число k_z (у напрямку поширення хвилі) та її частоту ω.

Для знаходження констант $A_{1,2}, B_{1,2}$ слід скористатися граничними умовами. Поле E_z є тангенціальним для обох пар стінок хвилеводу (рис. 6.8.1 а) і, отже, має дорівнювати нулеві на цих стінках (див. першу з формул (6.8.34)), тому для нього можна записати такі умови:

---

[4]Від англ. transversal – поперечний.

$$E_z(x=0)=0, \qquad E_z(x=a)=0, \qquad E_z(y=0)=0, \qquad E_z(y=b)=0. \qquad (6.8.48)$$

З першої та третьої умов (6.8.48) відразу випливає, що $A_2 = B_2 = 0$. Відповідно друга й четверта умови зводяться до вигляду

$$\sin(k_x a)=0, \qquad \sin(k_y b)=0, \qquad (6.8.49)$$

звідки

$$k_x = \frac{n\pi}{a}, \qquad k_y = \frac{m\pi}{b}, \qquad n,m = 1,2,3... \qquad (6.8.50)$$

Таким чином, остаточно можна записати:

$$E_z(\vec{r},t) = E_{0z}\sin\left(n\pi\frac{x}{a}\right)\sin\left(m\pi\frac{y}{b}\right)\exp(i\omega t - ik_z z), \qquad E_{0z} = A_1 B_1. \qquad (6.8.51)$$

Решту компонент електромагнітного поля можна знайти з формул (6.8.39).

Видно, що поле в хвилеводі являє собою стоячі хвилі в напрямках x та y (пор. з п. 6.4) і біжучу хвилю в напрямку z. Вздовж сторони a вкладається n півхвиль, вздовж сторони b – m півхвиль.

Хвиля фіксованого типу (в даному разі – ТМ-хвиля, або Е-хвиля) із заданими значеннями n і m називається (власною) модою хвилеводу (позначення: E_{nm}, або TM_{nm})[5]. Оскільки кожну стоячу хвилю можна подати як суму двох біжучих, поле (6.8.51) може бути подане як сума чотирьох хвиль:

$$E_z(\vec{r},t) = \frac{E_{0z}}{4}\exp(i\omega t - ik_z z)\times\left[-\exp\left(in\pi\frac{x}{a}+im\pi\frac{y}{b}\right)+\right.$$
$$\left.+\exp\left(-in\pi\frac{x}{a}+im\pi\frac{y}{b}\right)+\exp\left(in\pi\frac{x}{a}-im\pi\frac{y}{b}\right)-\exp\left(-in\pi\frac{x}{a}-im\pi\frac{y}{b}\right)\right]. \qquad (6.8.52)$$

Ці хвилі називають парціальними.

На рис. 6.8.5 показані силові лінії електричного та магнітного полів для мод E_{11} та E_{32} для фіксованого моменту часу. З часом картинки зсуваються в додатному напрямку осі z. Видно, що частина електричних силових ліній закінчується на стінках. Відповідно до другої з граничних умов (6.8.26) це означає, що там існують поверхневі заряди і протікають поверхневі струми (оскільки, як уже вказувалося, з часом картинка зсувається).

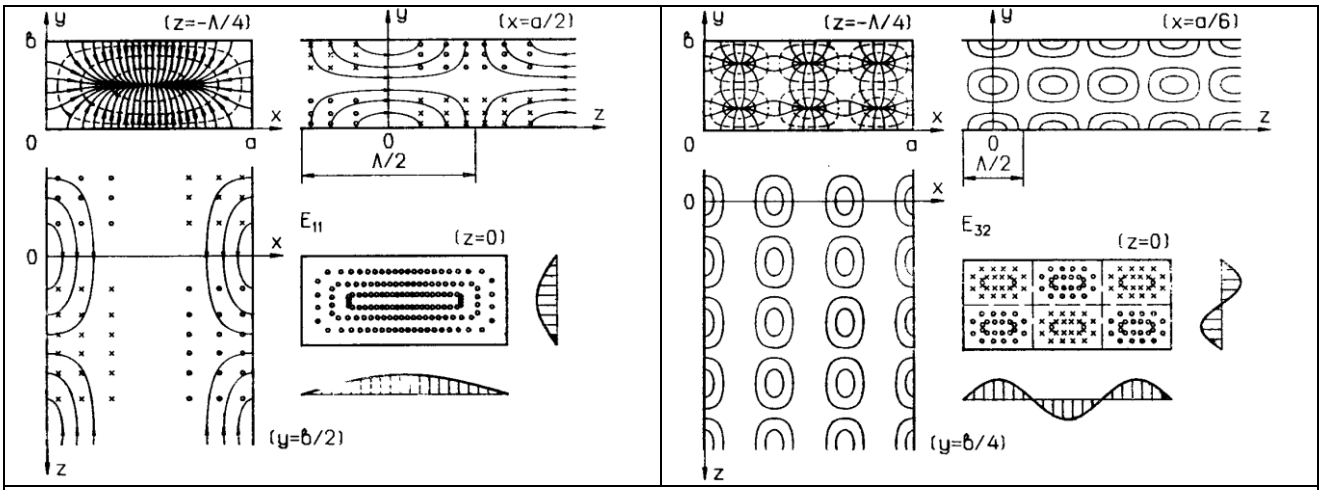


Рис. 6.8.5. Силові лінії електричного (суцільні лінії) та магнітного (штрихові лінії) полів мод E_{11} (а) та E_{32} (б) прямокутного хвилеводу. Використано позначення $\Lambda = 2\pi/k_z$

[5] В п. 6.7.1 йшлося про власні моди відрізка довгої лінії. Там теж фігурувала стояча хвиля, але вона встановлювалася вздовж лінії, а не в поперечному напрямку, як у хвилеводі. Відповідно в п. 6.7.1 йшлося про моди розподілених коливань зі своїми власними частотами. Для хвилеводів мова йде про моди хвиль, які там можуть поширюватися.

Тепер підставимо (6.8.51) до дисперсійного рівняння (6.8.47) з урахуванням (6.8.50). Отримаємо:

$$\left(\frac{n\pi}{a}\right)^2+\left(\frac{m\pi}{b}\right)^2+k_z^2=\left(\frac{\omega}{c}\right)^2, \qquad (6.8.53)$$

звідки

$$\omega^2=\omega_{cr\ nm}^2+c^2k_z^2, \qquad (6.8.54)$$

де

$$\omega_{cr\ nm}^2=c^2\left[\left(\frac{n\pi}{a}\right)^2+\left(\frac{m\pi}{b}\right)^2\right] \qquad (6.8.55)$$

– частота відтину для моди E_{nm}. Графіки дисперсійних залежностей (6.8.54) подані на рис. 6.8.6 а, залежність фазової та групової швидкості від частоти – на рис. 6.8.6 б. При великих ω всі дисперсійні залежності прямують до асимптоти $\omega=kc$, яка відповідає хвилям у необмеженому просторі (див. формулу (6.8.27)).

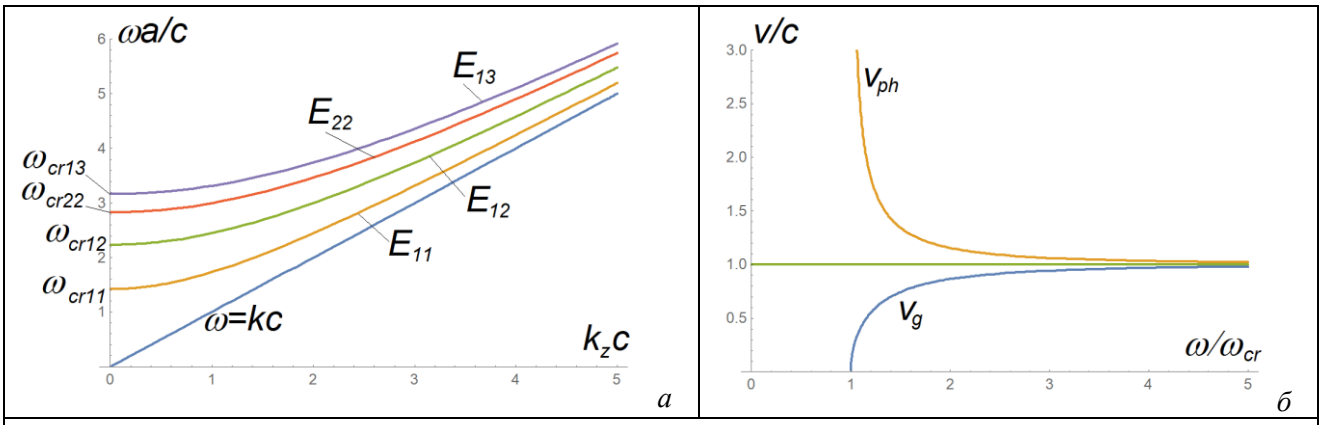


Рис. 6.8.6. Дисперсія хвилеводних мод: а – дисперсійні криві для різних мод (для випадку a=b); б – залежності фазової та групової швидкостей від частоти

Перш за все зазначимо, що в хвилеводі можуть поширюватися лише хвилі з частотами вище деякої критичної (так звана частота відтину). Значення цієї частоти (6.8.55) своє для кожної моди. Наявність такої частоти приводить до виникнення дисперсії (див. вище розділ 6.2). Фактично причиною непрозорості на низьких частотах є неможливість задоволення граничних умов для надто довгих хвиль.

Мода з найменшою частотою відтину називається основною модою хвилеводу. Серед ТМ-мод найнижчу частоту відтину має мода E_{11}.

6.8.8. ТЕ-моди прямокутного хвилеводу

Рівняння (6.8.41) для поздовжнього магнітного поля ТЕ-мод збігається з аналогічним рівнянням (6.8.40) для поздовжнього електричного поля ТМ-мод з точністю до позначень. Тому для нього можна скористатися розв'язком (6.8.46) і відразу записати:

$$H_z(\vec{r},t)=\left[C_1\sin(k_xx)+C_2\cos(k_xx)\right]\left[D_1\sin(k_yy)+D_2\cos(k_yy)\right]\exp(i\omega t-ik_zz). \qquad (6.8.56)$$

Гранична умова для тангенціального магнітного поля (четверта з умов (6.8.34)) містить невідомий поверхневий струм, тому скористатися нею безпосередньо не вдається. Натомість на основі формул (6.8.39) знайдемо компоненти електричного поля:

$$E_x=\frac{ik_0}{k_z^2-k_0^2}\frac{\partial H_z}{\partial y}=$$

$$=\frac{ik_0}{k_z^2-k_0^2}\frac{\partial}{\partial y}\left[C_1\sin(k_xx)+C_2\cos(k_xx)\right]\left[D_1\sin(k_yy)+D_2\cos(k_yy)\right]\exp(i\omega t-ik_zz)=$$

$$= \frac{ik_0 k_y}{k_z^2 - k_0^2}\left[C_1 \sin(k_x x) + C_2 \cos(k_x x)\right]\left[D_1 \cos(k_y y) - D_2 \sin(k_y y)\right]\exp(i\omega t - ik_z z); \quad (6.8.57)$$

$$E_y = -\frac{ik_0}{k_z^2 - k_0^2}\frac{\partial H_z}{\partial x} =$$

$$-\frac{ik_0}{k_z^2 - k_0^2}\frac{\partial}{\partial x}\left[C_1 \sin(k_x x) + C_2 \cos(k_x x)\right]\left[D_1 \sin(k_y y) + D_2 \cos(k_y y)\right]\exp(i\omega t - ik_z z) =$$

$$= \frac{ik_0 k_x}{k_z^2 - k_0^2}\left[C_1 \cos(k_x x) - C_2 \sin(k_x x)\right]\left[D_1 \sin(k_y y) + D_2 \cos(k_y y)\right]\exp(i\omega t - ik_z z). \quad (6.8.58)$$

Поле E_x є тангенціальним на стінках $y = 0$ та $y = b$, а поле E_y – на стінках $x = 0$ та $x = a$. Відповідно до першої з умов (6.8.34) граничні умови для цих компонент можна записати у формі

$$E_x(y = 0) = 0, \qquad E_x(y = b) = 0, \qquad E_y(x = 0) = 0, \qquad E_y(x = a) = 0. \qquad (6.8.59)$$

З першої та третьої умов (6.8.59) випливає, що $C_1 = D_1 = 0$, з другої та четвертої – що компоненти k_x та k_y задовольняють співвідношенням (6.8.50). Але тепер поле (6.8.56) відмінне від нуля, якщо один з індексів n, m дорівнює нулеві[6].

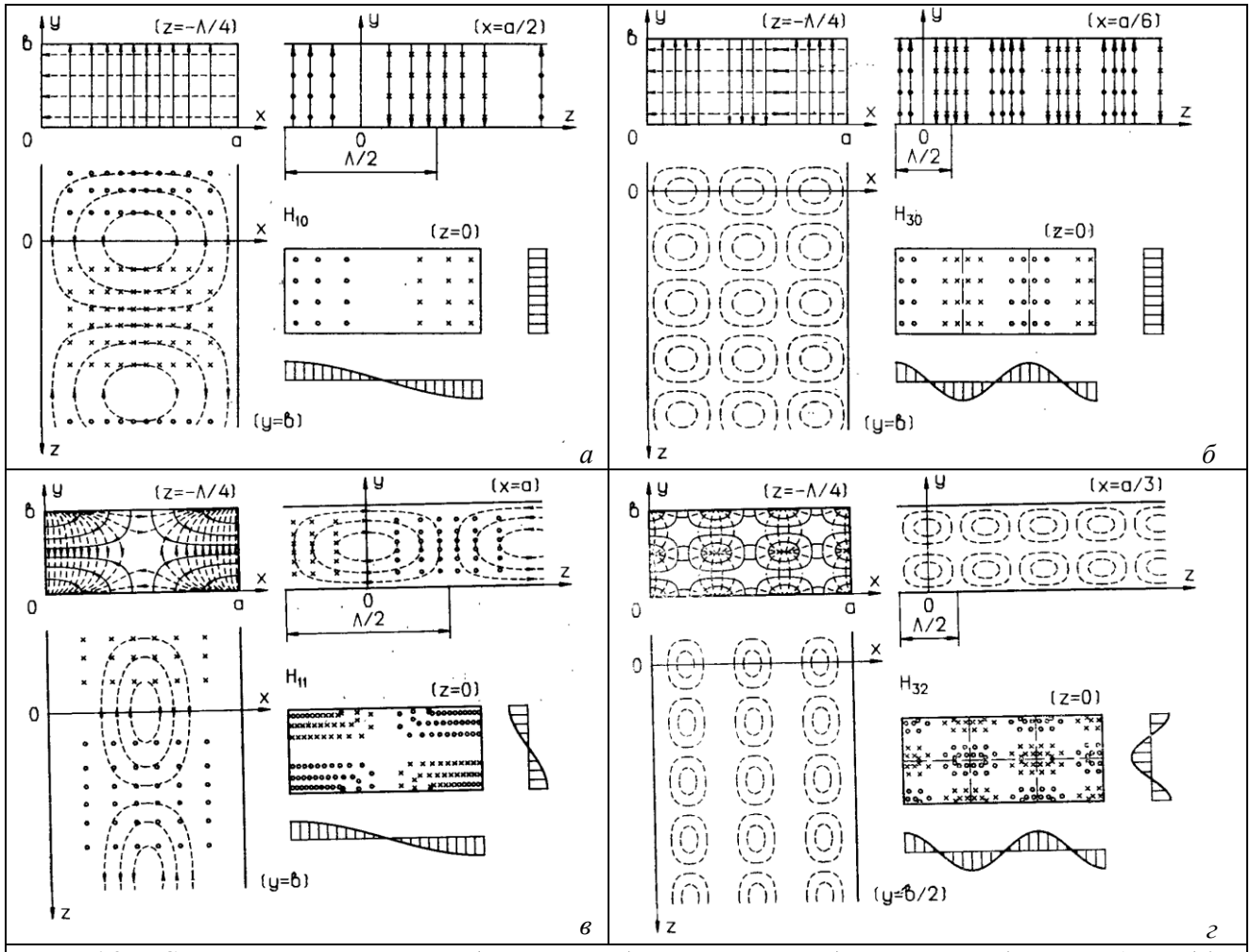


Рис. 6.8.7. Силові лінії електричного (суцільні лінії) та магнітного (штрихові лінії) полів мод H_{10} (а), H_{30} (б), H_{11} (в) та H_{32} (г) прямокутного хвилеводу. Використано позначення $\Lambda = 2\pi/k_z$

Формули для закону дисперсії і частоти відтину у мод H_{nm} такі самі, як у мод E_{nm} із

[6] При *m=n=0* у хвилі буде відмінною від нуля лише поздовжнє магнітне поле, чого не може бути.

тими самими індексами (див. формули (6.8.54)-(6.8.55)). Але серед ТЕ-мод існують моди H_{10} та H_{01}. Якщо $a > b$, то основною модою[7] хвилеводу виступає мода H_{10} з частотою відтину

$$\omega_{cr10} = \pi c/a. \qquad (6.8.60)$$

На рис. 6.8.7 показані силові лінії електричного та магнітного полів кількох ТЕ-мод для фіксованого моменту часу. Видно, що для основної моди пов'язані з нею поверхневі заряди наводяться лише на широких стінках хвилеводу.

6.8.9. Хвильовий опір хвилеводу. Узгодження хвилеводу з навантаженням

Для фідерів на основі хвилеводів, як і для двопровідних ліній, узгодження з навантаженням (тобто забезпечення режиму біжучих хвиль) так само відіграє важливу роль. Для опису узгодженого режиму вводиться поняття хвильового опору хвилеводу – деякого аналога хвильового опору довгої лінії.

Хвильовий опір хвилеводу (для даної моди – ТМ або ТЕ типу) вводиться як відношення амплітуди поперечного електричного поля до амплітуди поперечного магнітного поля, тобто в нашому випадку

$$\left|\frac{E_\perp}{H_\perp}\right| = \frac{\sqrt{E_x^2 + E_y^2}}{\sqrt{H_x^2 + H_y^2}}. \qquad (6.8.61)$$

Для фіксованої моди ця величина не залежить від координат.

У гауссівській системі одиниць електричне й магнітне поле має однакову розмірність. Але в системі одиниць СІ, яка зазвичай використовується в інженерних розрахунках, їхнє відношення має розмірність опору.

Можна показати, що в цій системі одиниць для моди TM_{mn}

$$\left|\frac{E_\perp^{TM}}{H_\perp^{TM}}\right| \equiv Z_0^{TM} = \frac{Z_0}{\sqrt{1-\left(\omega_{cr\,nm}/\omega\right)^2}}, \qquad (6.8.62)$$

де $\omega_{cr\,nm}$ – частота відтину даної моди (див. вище формулу (6.8.55)),

$$Z_0 = \sqrt{\frac{\mu\mu_0}{\varepsilon\varepsilon_0}} = \sqrt{\frac{\mu}{\varepsilon}} \cdot 376{,}7\ Ом \qquad (6.8.63)$$

– характеристичний опір необмеженого середовища (хвильовий опір вакууму, див. нижче п. 7.2.3), ε_0 та μ_0 – діелектрична та магнітна проникності вакууму в системі одиниць СІ.

Відповідно для моди TE_{mn}

$$\left|E_\perp^{TE}/H_\perp^{TE}\right| \equiv Z_0^{TE} = Z_0\sqrt{1-\left(\omega_{cr\,nm}/\omega\right)^2}. \qquad (6.8.64)$$

Нагадаємо, що частота відтину прямокутного хвилеводу відповідно до формули (6.8.54) однозначно задається його поперечними розмірами і прямує до нуля при збільшенні цих розмірів. В такому випадку величини Z_0^{TM} і Z_0^{TE} одночасно прямують до Z_0 – хвильового опору вакууму, оскільки при великих поперечних розмірах хвилеводу вплив його стінок зменшується.

Аналогічно до того, як у відрізка навантаженої двопровідної лінії вводиться вхідний опір, можна ввести поняття вхідного опору для хвилеводу певної довжини з навантаженням. Відрізок хвилевода з провідним торцем можна розглядати як аналог шлейфу Татаринова (див. вище п. 6.6). Тоді узгодження хвилеводу з навантаженням (тобто режим біжучої хвилі) можна забезпечити аналогічно до того, як забезпечується узгодження двопровідних ліній з навантаженням за допомогою шлейфів Татаринова.

6.8.10. Моди циліндричного хвилеводу

У циліндричному хвилеводі (рис. 6.8.1 б) також можуть існувати хвилі типу ТЕ та ТМ. Для опису полів цих хвиль зручно скористатися циліндричною системою координат. Як ра-

[7] Основна мода хвилеводу – це мода з найнижчою частотою відтину.

ніше, залежність від часу та координати z має вигляд біжучої гармонічної хвилі, $\exp(i\omega t - ik_z z)$.

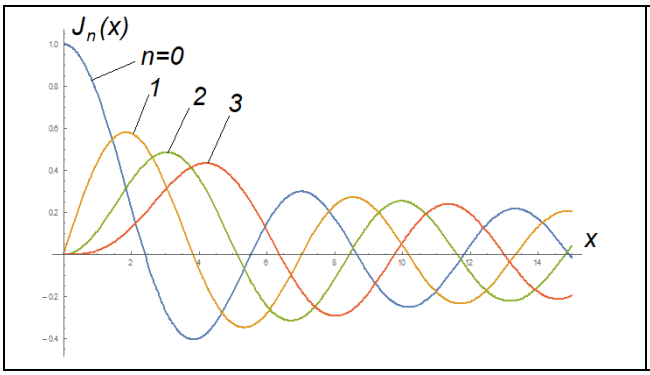


Рис. 6.8.8. Графіки функцій Бесселя дійсного аргументу

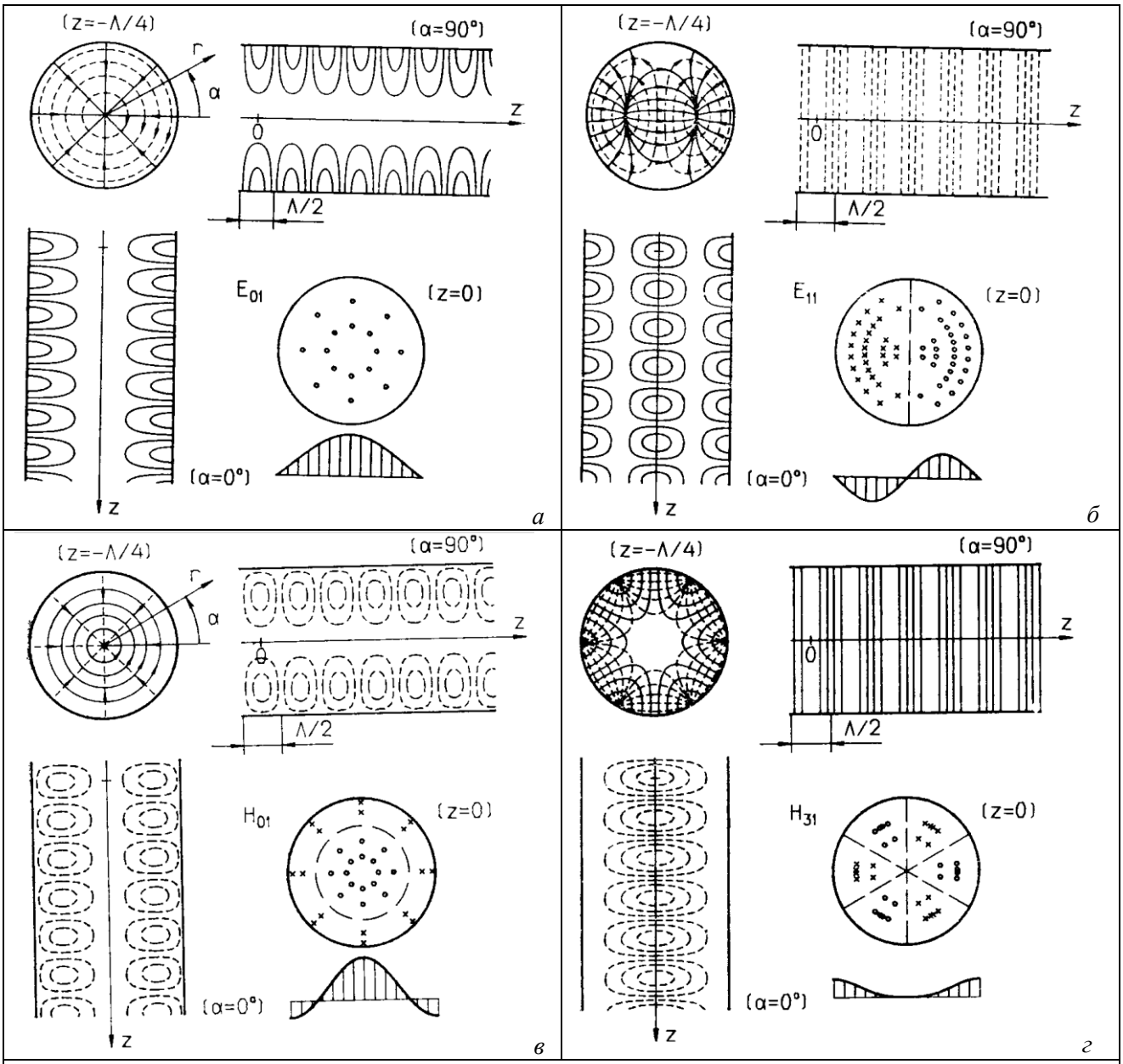


Рис. 6.8.9. Силові лінії електричного (суцільні лінії) та магнітного (штрихові лінії) полів мод E_{01} (а), E_{11} (б), H_{01} (в) та H_{31} (г) циліндричного хвилеводу. Використано позначення $\Lambda = 2\pi/k_z$

Залежність від азимутального кута φ має вигляд $\exp(in\varphi)$ (за змістом задачі вона має

бути періодичною з періодом 2π, щоб забезпечити однозначність полів). Нарешті, радіальна залежність описується функцією Бесселя порядку n (див. вище п. 1.4.5 та рис. 6.8.8) і має вигляд $J_n\left(k_{r\,nm}r\right)$. Дискретність компоненти $k_{r\,nm}$ забезпечується умовою, що на стінці хвилеводу функція $J_n\left(k_{r\,nm}r\right)$ або похідна від неї (в залежності від типу хвилі) повинна обертатися на нуль.

Дисперсійна залежність мод циліндричного хвилеводу виражається формулою (6.8.40), але вираз для частоти відтину буде не таким, як у прямокутному хвилеводі.

Просторовий розподіл полів для різних мод циліндричного хвилеводу показано на рис. 6.8.9.

Контрольні питання до розділу 6.8

1. У чому відмінність між потенціальним і вихровим полем?
2. Які характеристики електричного та магнітного полів визначають їхню силову дію?
3. Запишіть систему рівнянь Максвелла та охарактеризуйте фізичний зміст кожного з рівнянь.
4. Які переваги має хвилевід порівняно з ефіром при передачі сигналу?
5. Що таке ТЕ- і ТМ-хвилі хвилеводу?
6. Який зміст вкладають у поняття моди хвилеводу?
7. Чому поперечні хвильові числа в прямокутному хвилеводі виявляються дискретними?
8. Чому в хвилеводі існує частота відтину?
9. Опишіть залежність фазової та групової швидкостей хвилеводної моди від частоти.
10. Яку моду хвилеводу називають основною?
11. Чому в стінках хвилеводу протікають струми?

Задачі до підрозділу 6.8

6.8.1. Прямокутний хвилевід має розміри 6×10 см. Знайти частоту відтину такого хвилеводу.

6.8.2. Частота відтину хвилеводної моди складає $\omega_0/2\pi$=100 МГц, частота хвилі даної моди – $\omega/2\pi$=120 МГц. Знайти відповідні фазову й групову швидкості та довжину хвилі.

Розділ 7. Антени

В багатьох випадках сигнал від передавача до приймача передається через ефір. Тоді невід'ємною частиною тракту передавання сигналу є передавальна та приймальна антени (рис. В.1 а). Перша з них (передавальна) перетворює сигнал, що надходить від передавача, в електромагнітні хвилі, які випромінюються в ефір. Друга (приймальна), навпаки, приймає електромагнітні хвилі, що надходять від передавальної антени, перетворює їх на коливання напруги та струму й передає на вхід приймача.

В цій частині курсу ми розглянемо роботу передавальних і приймальних антен, а в наступній – поширення радіохвиль у навколоземному просторі.

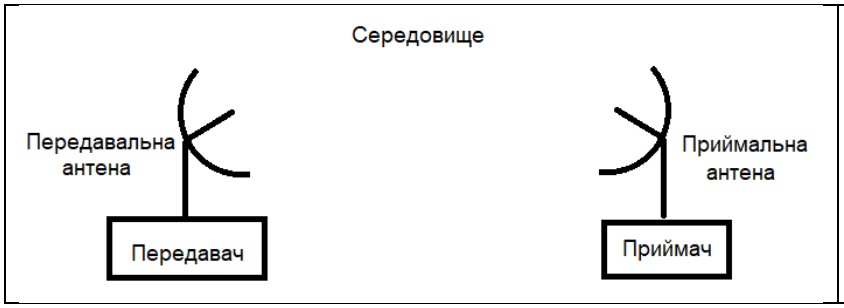


Рис. В.1 а. Схема тракту передавання – приймання радіосигналу

7.1. Випромінювання електромагнітних хвиль системою зарядів, що рухаються

Перш ніж безпосередньо говорити про антени, нагадаємо деякі матеріали з курсу електродинаміки, які стосуються випромінювання електромагнітних хвиль.

7.1.1. Отримання хвильового рівняння для електромагнітних хвиль у вакуумі за наявності джерел

Ми вже отримували хвильове рівняння для електромагнітних хвиль у вакуумі (див. п. 6.7.3). Але зараз нас цікавлять джерела, здатні збудити ці хвилі, а саме струми (або, що те саме, заряди, які рухаються).

Отже, розглянемо систему зарядів, що рухаються у вакуумі, породжуючи розподіл густини струму $\vec{j}(\vec{r},t)$. Нас цікавитимуть електромагнітні хвилі, що виникають у такій системі.

Для отримання хвильового рівняння для електромагнітних хвиль скористаємося першою парою рівнянь Максвелла (див. п. 6.7.2), які у вакуумі набувають вигляду

$$\nabla\times\vec{H}=\frac{4\pi}{c}\vec{j}+\frac{1}{c}\frac{\partial\vec{E}}{\partial t},\qquad \nabla\times\vec{E}=-\frac{1}{c}\frac{\partial\vec{H}}{\partial t}. \tag{7.1.1}$$

Шукатимемо розв'язок цих рівнянь у формі

$$\vec{H}=\nabla\times\vec{A},\qquad \vec{E}=-\nabla\cdot\varphi-\frac{1}{c}\frac{\partial\vec{A}}{\partial t}. \tag{7.1.2}$$

Тоді друге рівняння задовольняється тотожно, а перше набуває вигляду:

$$\nabla\times\nabla\times\vec{A}=\frac{4\pi}{c}\vec{j}+\frac{1}{c}\frac{\partial}{\partial t}\left(-\frac{1}{c}\frac{\partial\vec{A}}{\partial t}\right), \tag{7.1.3}$$

або (див. формулу (6.7.8))

$$\nabla\left(\nabla\cdot\vec{A}\right)-\Delta\vec{A}=\frac{4\pi}{c}\vec{j}-\frac{1}{c^2}\frac{\partial^2\vec{A}}{\partial t^2}. \tag{7.1.4}$$

Нагадаємо, що вектор-потенціал визначається з точністю до градієнту довільної функції. Тому для його однозначного визначення треба використати деяку додаткову умову (так звану умову калібрування), яку зручно обрати у формі $\nabla\cdot\vec{A}=0$ (тобто поле вектора-потенціалу є суто вихровим). В результаті хвильове рівняння набуває вигляду

$$\frac{1}{c^2}\frac{\partial^2\vec{A}}{\partial t^2}-\Delta\vec{A}=\frac{4\pi}{c}\vec{j}. \tag{7.1.5}$$

Таким чином, в аналізованому випадку джерелом електромагнітних хвиль є протікання струму, тобто рух електричних зарядів. Як випливає з цього рівняння, саме напрямок вектора густини струму визначає напрямок вектор-потенціалу, а, отже, і полів випромінюваних хвиль, тобто їх поляризацію.

Обговоримо на якісному рівні питання про те, чому для випромінювання електромагнітних хвиль обов'язково потрібен струм, а, наприклад, точковий заряд, що змінюється з часом, джерелом електромагнітних хвиль бути не може.

Електричне поле точкового заряду в квазістаціонарному наближенні описується законом Кулона (рис. 7.1.1 а). Але зміна такого поля з часом не може створити магнітного поля, необхідного для поширення електромагнітної хвилі. Найзручніше пояснити це, врахувавши, що на далеких віддалях від заряду силові лінії електричного поля можна вважати паралельними (рис. 7.1.1 б). Кожна силова лінія такого поля, подібно до провідника зі струмом, створює магнітне поле, силові лінії якого – це концентричні кільця, надіті на силову лінію електричного поля і перпендикулярні до неї. Це безпосередньо випливає з першого рівняння (7.1.1), де струм провідності $\vec{j}$ і струм зміщення $\partial\vec{E}/\partial t$, що стоять у правій частині, цілком рівноправні. Але в кожній точці простору магнітне поле, створене певною силовою лінією електричного поля, компенсується магнітним полем, створеним іншими силовими лініями електричного поля.

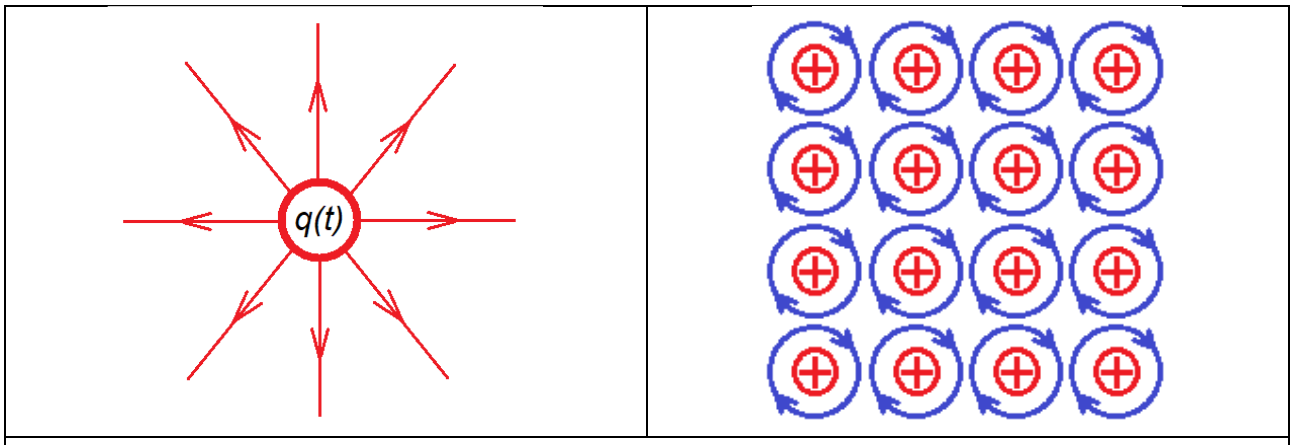


Рис. 7.1.1. До пояснення неможливості випромінювання електромагнітних хвиль змінним точковим зарядом: а – квазістаціонарне електричне поле точкового заряду; б – силові лінії магнітного поля (показані синім), породженого змінним однорідним електричним полем (силові лінії, перпендикулярні до площини рисунка, показані червоним)

Аналогічні міркування приводять до висновку, що й на малих віддалях від змінного точкового заряду в силу геометрії задачі магнітне поле не виникає. Отже, сумарне магнітне поле такого заряду дорівнює нулеві в усіх точках простору.

До цього результату можна прийти й з інших міркувань. На сферичній поверхні, що оточує змінний точковий заряд, немає виділеного напрямку, з яким можна пов'язати напрямок магнітного поля, перпендикулярного електричному.

Нарешті, той самий результат суто формально випливає з другого рівняння (7.1.1). Очевидно, що для точкового заряду $q(t)$, який змінюється з часом, електричне поле можна подати у формі $\vec{E}(\vec{r},t)=\vec{e}_r E(r,t)$, де r – радіус у сферичній системі координат, початок відліку якої суміщений з точкою розташування заряду. Тоді в сферичних координатах

$$\left[\nabla\times\vec{E}(\vec{r},t)\right]=\frac{\vec{e}_r}{r\sin\theta}\left[\frac{\partial}{\partial\theta}\left(E_\varphi\sin\theta\right)-\frac{\partial E_\theta}{\partial\varphi}\right]+\frac{\vec{e}_\theta}{r}\left[\frac{1}{\sin\theta}\frac{\partial E_r}{\partial\varphi}-\frac{\partial}{\partial r}\left(E_\varphi r\right)\right]+\frac{\vec{e}_\varphi}{r}\left[\frac{\partial}{\partial r}\left(rE_\theta\right)-\frac{\partial E_r}{\partial\theta}\right]=$$

$$=\frac{\vec{e}_\theta}{r\sin\theta}\frac{\partial E_r}{\partial\varphi}-\frac{\vec{e}_\varphi}{r}\frac{\partial E_r}{\partial\theta}=0.$$

Отже, й $\partial\vec{B}/\partial t=0$, тобто змінне магнітне поле в цій моделі відсутнє.

Таким чином, нерухомий точковий заряд, навіть якщо його величина змінюється з

часом, нездатний збудили електромагнітну хвилю. Для цього обов'язково потрібен рух заряджених частинок, що й підтверджує рівняння (7.1.5).

Тепер розглянемо довільний (незмінний у просторі) розподіл густини змінного заряду вигляду $\rho(\vec{r},t)=\rho_m(\vec{r})f(t)$. Розіб'ємо його на окремі невеликі області, які можна розглядати як змінні точкові заряди. Тоді в силу принципу суперпозиції висновок про відсутність змінного магнітного поля справедливий і для такої моделі.

Далі ми будемо розв'язувати рівняння (7.1.5) у наближенні заданого струму, тобто знехтуємо зворотними впливом електромагнітного поля на його джерела.

7.1.2. Поле випромінювання системи зарядів, що рухаються

Вважатимемо, що в деякій області простору, позначеній на рис. 7.1.2 сірим кольором, протікають струми, які є джерелом випромінювання електромагнітних хвиль. На великій віддалі від цієї області (порівняно з її розмірами) знаходиться точка спостереження *P*.

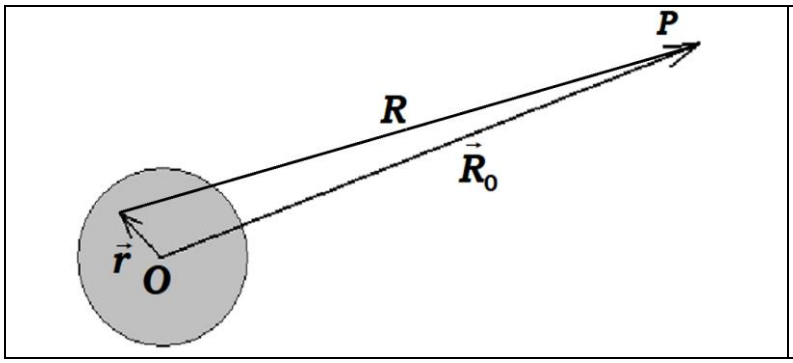


Рис. 7.1.2. Область джерел (позначена сірим), початок відліку O й точка спостереження P; вектор $\vec{r}$ закінчується в точці, де знаходиться елемент об'єму області джерел dV

Можна вважати, що сумарний заряд області джерел дорівнює нулеві; як було показано раніше, такий заряд, навіть якщо він змінюється з часом, не може випромінювати електромагнітні хвилі. Тому поле випромінювання в точці спостереження можна вважати вихровим. Зокрема, в околі точки спостереження (далеко від джерела) це поле можна розглядати як плоску поперечну електромагнітну хвилю. Тому там, виходячи з другого рівняння (7.1.2) та другого рівняння (7.1.1), можна записати такі співвідношення:

$$\vec{E}=-\frac{1}{c}\frac{d\vec{A}}{dt}, \qquad \vec{H}=\frac{1}{c}\left[\frac{d\vec{A}}{dt}\times\vec{n}\right]. \tag{7.1.6}$$

Тут $\vec{n}$ – одиничний вектор, що характеризує напрямок від області розташування зарядів на точку спостереження (враховано, що на великих відстанях від джерела можна вважати, що хвиля поширюється вздовж цього напрямку)[1]. Видно, що необхідною умовою появи поля випромінювання на великих відстанях від джерела є залежність вектор-потенціалу від часу.

Вектор-потенціал, створюваний системою зарядів, що рухаються, в довільній точці простору можна записати у формі так званого загаяного потенціалу (потенціалу, що запізнюється):

$$\vec{A}=\frac{1}{c}\int\frac{\vec{j}(\vec{r},t-R/c)}{R}dV, \tag{7.1.7}$$

де R – відстань від точки розташування заряду до точки спостереження, а інтегрування ведеться по області джерел, тобто по області, де густина струму відмінна від нуля (рис. 7.1.2). Із формули (7.1.7) видно, що значення поля в точці спостереження в момент часу t визначається рухом заряду в попередній момент часу $t-R/c$. Тут R/c – це час запізнення, протягом якого електромагнітна хвиля рухається від точки випромінювання до точки спостереження. Подібна ситуація типова для будь-якого випромінювача хвиль довільної природи.

Як видно із формули (7.1.7), випромінювання від джерела великих розмірів

[1] Виведення цієї та ряду наступних формул можна знайти в книзі: Ландау Л.Д., Лифшиц Е.М. Теория поля. – М.: Наука, 1988.

визначається шляхом підсумовування випромінювання з різних частин цього джерела. Це результат лінійності хвильового рівняння (7.1.5).

Очевидно, аналогом точкового джерела для електромагнітних хвиль може бути деяка область, зайнята зарядами, що рухаються, з розмірами, значно меншими від довжини випромінюваних хвиль. Випромінювання з такої області можна розглядати як деякий аналог функції Гріна (див. вище п. 2.3.3). Тоді випромінювання довільної області можна знайти як суму внесків від таких елементарних випромінювачів, подібно до того, як це можна зробити в задачі про випромінювання поздовжніх хвиль.

Розглянемо хвильове поле, збуджуване таким елементарним випромінювачем на великих відстанях (фактично – на відстанях, значно більших від довжини випромінюваних хвиль). Виберемо в області руху зарядів деяку точку, яку умовно вважатимемо початком відліку. Радіус-вектор, проведений від неї до точки спостереження, позначимо через $\vec{R}_0$, а радіус-вектор, проведений від неї до точки розташування деякого заряду – через $\vec{r}$ (рис. 7.1.1). Мализна розмірів області випромінювання означає, зокрема, що $r \ll R_0$, тоді відстань R від заряду до точки спостереження, що входить до формули (7.1.7), визначається співвідношенням:

$$R = \left|\vec{R}_0 - \vec{r}\right| = \sqrt{R_0^2 + r^2 - 2\left(\vec{R}_0 \cdot \vec{r}\right)} \approx R_0 - \left(\vec{n} \cdot \vec{r}\right), \qquad \vec{n} = \vec{R}_0 / R_0 . \tag{7.1.8}$$

Підставимо (7.1.8) до (7.1.7). При цьому в знаменнику замінимо R на R_0, отримаємо

$$\vec{A}\left(\vec{R}_0, t\right) = \frac{1}{cR_0} \int \vec{j}\left(\vec{r}, t' - \frac{\left(\vec{n} \cdot \vec{r}\right)}{c}\right) dV , \qquad t' = t - \frac{R_0}{c} \tag{7.1.9}$$

(інтегрування ведеться по області руху зарядів).

Функцію $\vec{j}\left(t' - \left(\vec{n} \cdot \vec{r}\right)/c\right)$ розкладемо в ряд за малим параметром $\left(\vec{n} \cdot \vec{r}\right)$, обмежившись лінійним доданком[2], отримаємо

$$\vec{A}\left(\vec{R}_0, t\right) = \frac{1}{cR_0} \int \vec{j}\left(\vec{r}, t'\right) dV + \frac{1}{c^2 R_0} \int \left(\vec{n} \cdot \vec{r}\left(t'\right)\right) \frac{\partial \vec{j}\left(\vec{r}, t'\right)}{\partial t'} dV . \tag{7.1.10}$$

7.1.3. Дипольне, магнітодипольне та квадрупольне випромінювання

Перейдемо від неперервного розподілу густини зарядів у просторі до системи точкових зарядів за допомогою співвідношення

$$\vec{j}\left(\vec{r}, t\right) = \rho\left(\vec{r}, t\right) \vec{v}\left(\vec{r}, t\right) = \sum_i q_i \delta\left(\vec{r} - \vec{r}_i\left(t\right)\right) \vec{v}_i\left(t\right), \qquad \vec{v}_i = \frac{d\vec{r}_i}{dt} , \tag{7.1.11}$$

тоді формула (7.1.10) набуде вигляду

$$\vec{A}\left(\vec{R}_0, t\right) = \frac{1}{cR_0} \frac{d}{dt'} \sum_i q_i \vec{r}_i\left(t'\right) + \frac{1}{c^2 R_0} \frac{d}{dt'} \sum_i q_i \vec{v}_i\left(t'\right) \left(\vec{n} \cdot \vec{r}_i\left(t'\right)\right). \tag{7.1.12}$$

У свою чергу, формула (7.1.12) після деяких перетворень може бути переписана у формі

$$\vec{A}\left(\vec{R}_0, t\right) = \frac{1}{cR_0} \frac{d\vec{d}}{dt'} + \frac{1}{cR_0} \left[\frac{d\vec{m}}{dt'} \times \vec{n}\right] + \frac{1}{6c^2 R_0} \frac{d^2 \vec{D}}{dt'^2} , \tag{7.1.13}$$

де фігурують інтегральні характеристики системи зарядів:

$$\vec{d} = \sum_i q_i \vec{r}_i \tag{7.1.14}$$

– електричний дипольний момент (див. рис. 7.1.3 а);

$$\vec{m} = \frac{1}{2c} \sum_i q_i \left[\vec{r}_i \times \vec{v}_i\right] \tag{7.1.15}$$

– магнітний дипольний момент (див. рис. 7.1.3 б);

вектор $\vec{D} = \hat{D}\vec{n}$ виражається через тензор квадрупольного моменту (див. рис. 7.1.3 в-г)

---

[2] Строго кажучи, малим параметром задачі є безрозмірна величина $\left(\vec{n} \cdot \vec{r}\right)/R_0$.

$$D_{\alpha\beta} = \sum_i q_i \left(3x_{i\alpha}x_{i\beta} - \delta_{\alpha\beta}r_i^2\right), \qquad \vec{r}_i = \{x_{i1}, x_{i2}, x_{i3}\}. \tag{7.1.16}$$

Перший доданок у правій частині (7.1.13), породжений першим доданком у формулі (7.1.10), має нульовий порядок мализни за параметром $(\vec{n}\cdot\vec{r})$, а другий і третій, породжені другим доданком у (7.10) – перший порядок. Підставивши (7.1.13) до (7.1.7), легко знайти компоненти електромагнітного поля в точці спостереження. Видно, що амплітуда поля залежить від R_0 як $1/R_0$, що характерно для будь-яких точкових джерел хвиль у консервативних середовищах.

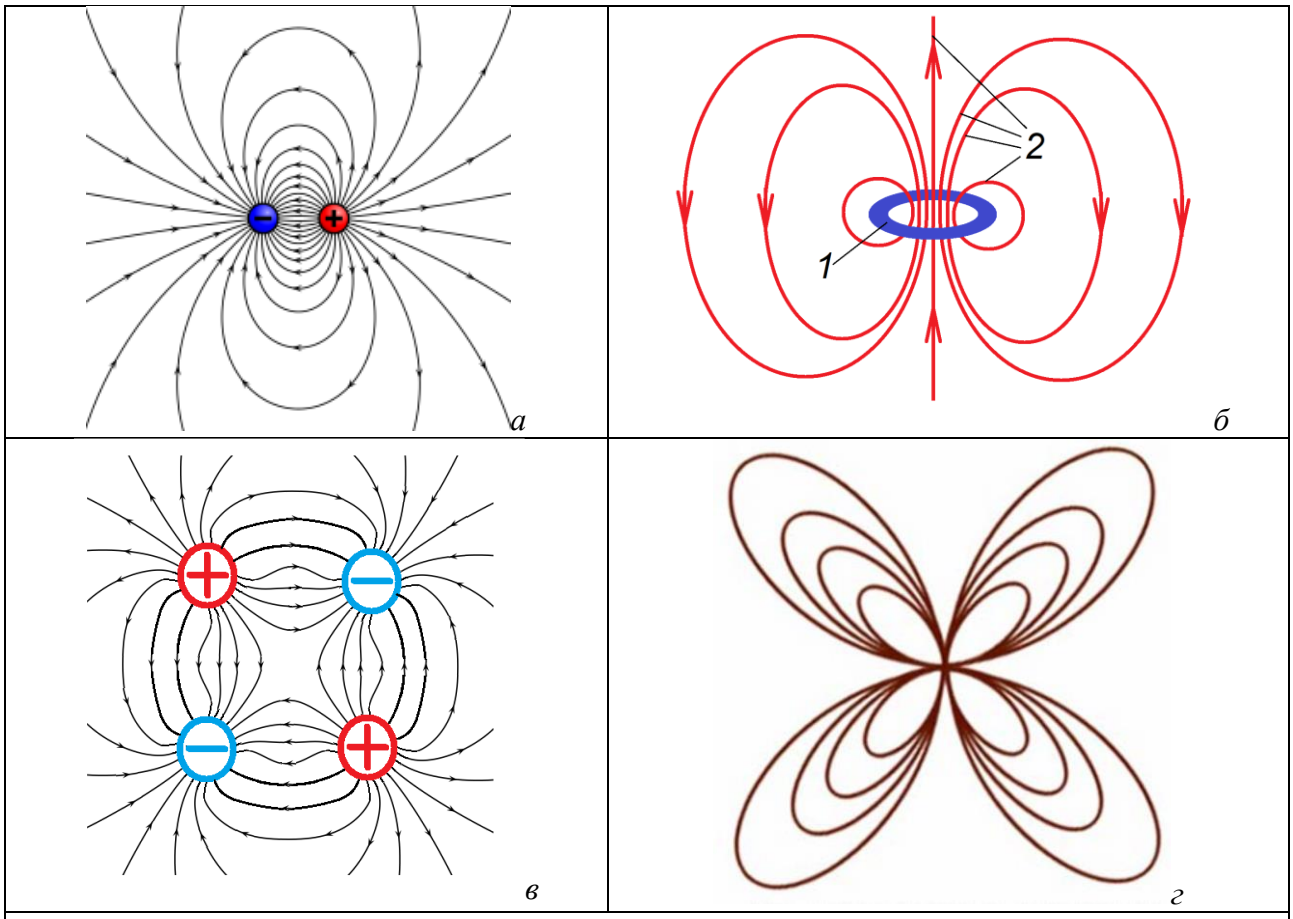


Рис. 7.1.3. Просторові розподіли електричного (а, в-г) та магнітного (б) поля стаціонарних мультиполів: а – поле електричного диполя; б – поле магнітного диполя (1 – кільце зі струмом, 2 – силові лінії); в-г – поле електричного квадруполя в близькій (в) та далекій (г) зоні (рисунок г повернутий щодо рисунку в на 45°)

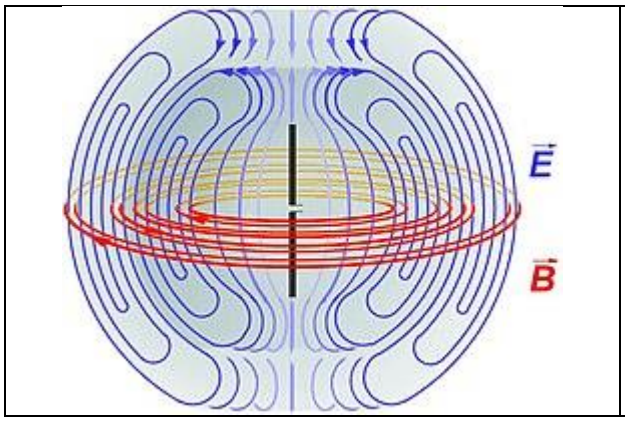


Рис. 7.1.4. Просторові розподіли електричного (сині лінії) та магнітного (червоні лінії) поля електричного диполя, момент якого гармонічно змінюється з часом

Поле випромінювання в даному наближенні можна розділити на три частини. Перша з

них, пропорційна другій похідній від дипольного моменту, відома в літературі як *дипольне випромінювання*. У загальному випадку воно відіграє основну роль. З формул (7.1.7) та (7.1.13) випливає, що поле випромінювання диполя пропорційне другій похідній за часом від дипольного моменту. Отже, змінний дипольний момент служить джерелом електромагнітних хвиль, якщо друга похідна за часом від нього відмінна від нуля.

Якщо припустити, що дипольний момент змінюється з часом за гармонічним законом, то приходимо до відомої з курсів електрики і магнетизму й електродинаміки моделі диполя Герца (див. нижче пп. 7.2.1-7.2.4). Нагадаємо, що такий диполь можна уявити собі як малий у масштабі довжини випромінюваних хвиль відрізок провідника зі змінним струмом. Поле випромінювання в далекій зоні (на відстані, значно більшій за довжину хвилі) містить азимутальну компоненту магнітного поля та компоненту електричного поля, спрямовану вздовж орта аксіального кута (вісь z спрямована вздовж провідника зі струмом) – див. рис. 7.1.4.

Доданок у полі випромінювання, пропорційний другій похідній від магнітного моменту, називають *магнітодипольним випромінюванням*. Якщо магнітний момент змінюється в часі за гармонічним законом, то цьому випромінюванню відповідає так званий магнітний диполь Герца (див. нижче п. 7.2.5). Його реалізацію можна уявити собі як кільце зі змінним струмом, причому радіус кільця знову-таки значно менший за довжину випромінюваних хвиль. Поле випромінювання в далекій зоні містить азимутальну компоненту електричного поля та компоненту магнітного поля, спрямовану вздовж орта аксіального кута (вісь z збігається з віссю кільця зі струмом).

Частина поля випромінювання, пропорційна до третьої похідної за часом від тензора квадрупольного моменту, відома як *квадрупольне випромінювання*. Квадрупольний випромінювач можна уявити собі як два однакові за величиною протилежно спрямовані змінні диполі, розташовані поруч, тому дипольний момент усієї системи дорівнює нулю. На відміну від електричного та магнітного диполів Герца, поле квадрупольного випромінювача вже не буде аксіально-симетричним.

Врахування наступних доданків у розкладі функції $\vec{j}\left(t' - (\vec{n}\cdot\vec{r})/c\right)$ у ряд дає додаткові типи випромінювання, пов‘язані з мультипольними моментами вищих порядків. Однак зазвичай вони виявляються несуттєвими.

Контрольні питання до підрозділу 7.1

1. Навіщо вводять скалярний і векторний потенціал електромагнітного поля?
2. Що таке умова калібрування для вектор-потенціалу?
3. Що таке загаяний потенціал?
4. Чому дипольне випромінювання системи зарядів, що рухаються, звичайно більше від магнітодипольного випромінювання?
5. Що так квадрупольне випромінювання?

Приклади розв’язання задач до підрозділу 7.1

П 7.1.1. Два однакових за величиною і протилежних за знаком заряди рівномірно рухаються вздовж паралельних прямих назустріч один одному, як показано на рис. 7.1.5. Як змінюватимуться з часом електричний та магнітний дипольні моменти такої системи? Чи буде така система випромінювати електромагнітні хвилі?

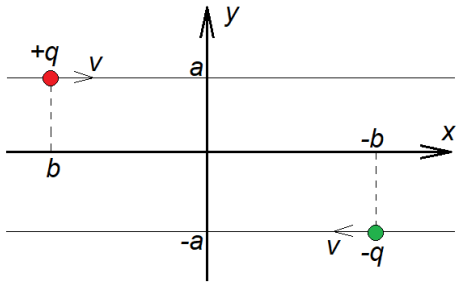


Рис. 7.1.5

Розв’язок

Запишемо координати x_1, y_1 та x_2, y_2 зарядів $+q$ та $-q$, врахувавши, що $b = \upsilon t$ (це означає, що в момент

часу $t = 0$ обидва заряди знаходяться на осі y):

$x_1 = \upsilon t,\ y_1 = a, \qquad x_2 = -\upsilon t,\ y_2 = -a$.

Електричний дипольний момент:

$$\vec{d} = \sum_i q_i \vec{r}_i = \{q(x_1 - x_2); q(y_1 - y_2)\} = \{2q\upsilon t; 2qa\};$$

$$\frac{d\vec{d}}{dt} = \{2q\upsilon; 0\}; \quad \frac{d^2\vec{d}}{dt^2} = 0.$$

Магнітний дипольний момент:

$$\vec{m} = \frac{1}{2c}\sum_i q_i \left[\vec{r}_i \times \vec{\upsilon}_i\right].$$

Оскільки швидкості зарядів паралельні осі *x*, то внесок дають лише у-компоненти радіус-векторів, а сам момент паралельний осі *z*:

$$m_z = \frac{1}{2c} q \begin{vmatrix} \upsilon t & a \\ \upsilon & 0 \end{vmatrix} = -\frac{q\upsilon a}{2c}; \quad \frac{dm_z}{dt} = 0.$$

Отже, в аналізованій системі електродипольне та магнітодипольне випромінювання відсутнє.

Задачі до підрозділу 7.1

7.1.1. Два однакових за величиною і протилежних за знаком заряди рівномірно обертаються по колу в протилежних напрямках з однаковою швидкістю. Лінія, що сполучає заряди, проходить через центр кола. Як змінюватимуться з часом електричний та магнітний дипольні моменти такої системи? Чи буде така система випромінювати електромагнітні хвилі?

7.1.2. Заряди розміщені у вершинах квадрата зі стороною *a*, як показано на рис. 7.1.6. Знайти тензор квадрупольного моменту такої системи. Вважати, що початок координат знаходиться в центрі квадрата, осі координат *x* та *y* паралельні його сторонам, а вісь *z* перпендикулярна площині малюнку.

7.1.3. Розв'язати попередню задачу в припущенні, що квадрат рівномірно обертається навколо осі *z*. Чи буде така система випромінювати електромагнітні хвилі?

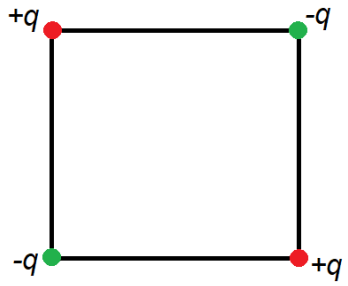


Рис. 7.1.6

7.2. Диполь Герца

Сучасні антени є конструктивно складними пристроями, тому розрахунок їхнього поля випромінювання можливий лише числовими методами. Тим не менше деякі основні поняття, що стосуються теорії антен, можна розглянути вже на прикладі найпростішої моделі – так званого (електричного) диполя Герца.

7.2.1. Елементарний електричний випромінювач: опис моделі

У п. 7.1.3 було показано, що випромінювання електромагнітних хвиль системою зарядів, що рухаються, зазвичай визначає (електричний) дипольний момент такої системи. Відповідно найпростіша модель випромінювача електромагнітних хвиль – це елементарний електричний випромінювач, або диполь Герца. Як уже згадувалося, диполь Герца можна уявити собі як малий у масштабі довжини випромінюваних хвиль відрізок провідника зі змінним струмом (рис. 7.2.1). Такий випромінювач був уперше реалізований у відомих експериментах Г. Герца (1886-1888)[3].

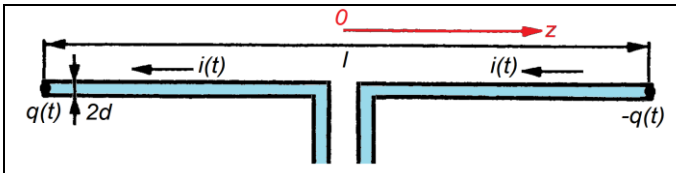


Рис. 7.2.1. Конструкція диполя Герца

Довжина диполя Герца l вважається значно меншою за випромінювану ним довжину хвилі, $l \ll \lambda$. Відповідно до зауваження, висловленого в п. 7.1.2, ця умова дозволяє вважати диполь Герца точковим джерелом хвиль. З неї також випливає, що струм

$$i(t) = i_m \exp i\omega t \tag{7.2.1}$$

через будь-який переріз провідника однаковий (див. вступ до розділу 1). Протікання змінного струму приводить до періодичного накопичення заряду

$$q(t) = \int i(t)\,dt = -\frac{i}{\omega} i_m \exp(i\omega t) \tag{7.2.2}$$

на кінцях провідника, тому його дипольний момент має величину

$$\vec{d}(t) = \vec{e}_z l q(t) = -i\vec{e}_z \frac{i_m l}{\omega} \exp(i\omega t). \tag{7.2.3}$$

Очевидно, друга похідна за часом від дипольного моменту відмінна від нуля, тому така система буде випромінювати електромагнітні хвилі.

7.2.2. Поле випромінювання диполя Герца

Формули, наведені наприкінці п. 7.1.2, описують поле випромінювання на великих віддалях від його джерела. Нас цікавить поле диполя Герца на різних віддалях від нього. Для його розрахунку скористаємося формулою (7.1.7), яка справедлива для будь-яких значень R. З урахуванням позначень, показаних на рис. 7.1.1, перепишемо її у формі

$$\vec{A}(\vec{R}_0) = \frac{1}{c}\int \frac{\vec{j}\left(\vec{r}, t - \left|\vec{R}_0 - \vec{r}\right|/c\right)}{\left|\vec{R}_0 - \vec{r}\right|} dV. \tag{7.2.4}$$

Нас цікавить випадок гармонічного струму (див. формулу (7.2.1). Тоді формула (7.1.7) набуває вигляду

$$\vec{A}(\vec{R}_0) = \frac{1}{c}\int \frac{\vec{j}(\vec{r})}{\left|\vec{R}_0 - \vec{r}\right|} \exp\left(i\omega t - ik_0\left|\vec{R}_0 - \vec{r}\right|\right) dV, \;\; k_0 = \frac{\omega}{c}. \tag{7.2.5}$$

Магнітне поле випромінювання набуває вигляду:

---

[3] Ці експерименти вперше продемонстрували існування електромагнітних хвиль, теоретично передбачених Максвеллом

$$\vec{H}\left(\vec{R}_0\right)=\frac{1}{c}\int\frac{\left[\nabla\times\vec{j}\left(\vec{r}\right)\right]}{\left|\vec{R}_0-\vec{r}\right|}\exp\left(i\omega t-ik_0\left|\vec{R}_0-\vec{r}\right|\right)dV\,. \qquad (7.2.6)$$

Однак останньою формулою незручно користуватися, оскільки треба брати ротор від густини струму, яка відмінна від нуля лише в провіднику. Тому перетворимо її, скориставшись тотожністю

$$\left[\nabla\times\psi\vec{F}\right]=\psi\left[\nabla\times\vec{F}\right]+\left[\nabla\psi\times\vec{F}\right]. \qquad (7.2.7)$$

Отримаємо:

$$\vec{H}\left(\vec{R}_0\right)=\frac{1}{c}\int\left[\nabla\times\left\{\frac{\vec{j}\left(\vec{r}\right)\exp\left(i\omega t-ik_0\left|\vec{R}_0-\vec{r}\right|\right)}{\left|\vec{R}_0-\vec{r}\right|}\right\}\right]dV-\frac{1}{c}\int\left[\nabla\frac{\exp\left(i\omega t-ik_0\left|\vec{R}_0-\vec{r}\right|\right)}{\left|\vec{R}_0-\vec{r}\right|}\times\vec{j}\left(\vec{r}\right)\right]dV\,. \qquad (7.2.8)$$

Застосуємо до першого з інтегралів формулу

$$\int\limits_V\left[\nabla\times\vec{F}\right]dV=\oint\limits_S d\vec{S}\times\vec{F}=-\oint\limits_S\vec{F}\times d\vec{S}\,, \qquad (7.2.9)$$

де S – замкнена поверхня, що охоплює об'єм V. Маємо:

$$\int\limits_V\left[\nabla\times\left\{\frac{\vec{j}\left(\vec{r}\right)\exp\left(i\omega t-ik_0\left|\vec{R}_0-\vec{r}\right|\right)}{\left|\vec{R}_0-\vec{r}\right|}\right\}\right]dV=-\oint\limits_S\frac{\exp\left(i\omega t-ik_0\left|\vec{R}_0-\vec{r}\right|\right)}{\left|\vec{R}_0-\vec{r}\right|}\left[\vec{j}\left(\vec{r}\right)\times d\vec{S}\right]. \qquad (7.2.10)$$

Поверхню інтегрування S можна обрати на деякій віддалі від поверхні, яка охоплює область джерел. Тоді на ній $\vec{j}\left(\vec{r}\right)=0$, тобто цей інтеграл обертається на нуль.

Врахуємо, що

$$\nabla\frac{\exp\left(i\omega t-ik_0\left|\vec{R}_0-\vec{r}\right|\right)}{\left|\vec{R}_0-\vec{r}\right|}=\vec{n}\exp\left(i\omega t-ik_0\left|\vec{R}_0-\vec{r}\right|\right)\left(\frac{1}{\left|\vec{R}_0-\vec{r}\right|^2}+\frac{ik_0}{\left|\vec{R}_0-\vec{r}\right|}\right). \qquad (7.2.11)$$

Таким чином, остаточно маємо:

$$\vec{H}\left(\vec{R}_0\right)=\frac{1}{c}\int\left(\frac{1}{\left|\vec{R}_0-\vec{r}\right|^2}+\frac{ik_0}{\left|\vec{R}_0-\vec{r}\right|}\right)\left[\vec{j}\left(\vec{r}\right)\times\frac{\vec{R}_0}{R_0}\right]\exp\left(i\omega t-ik_0\left|\vec{R}_0-\vec{r}\right|\right)dV\,. \qquad (7.2.12)$$

Тепер конкретизуємо цю формулу для диполя Герца, вважаючи, що початок відліку суміщений з його серединою. Тоді можна вважати, що

$$\vec{j}\left(\vec{r}\right)=\begin{cases}\dfrac{i_m}{\pi d^2}, & -\dfrac{l}{2}\le z\le\dfrac{l}{2}\cap\rho\le d;\\ 0, & |z|>\dfrac{l}{2}\cup\rho>d,\end{cases} \qquad (7.2.13)$$

де ρ, φ, z – циліндричні координати. В силу умови $l\ll\lambda$ густину струму в усіх точках антени вважаємо однаковою.

Спочатку проінтегруємо (7.2.12) по площині, перпендикулярній осі z. Цей інтеграл відмінний від нуля лише при $-l/2<z<l/2$. Отримаємо:

$$\vec{H}\left(\vec{R}_0\right)=\frac{1}{c}\int\limits_{-l/2}^{l/2}\left(\frac{1}{\left|\vec{R}_0-\vec{r}\right|^2}+\frac{ik_0}{\left|\vec{R}_0-\vec{r}\right|}\right)i_m\left[\vec{e}_z\times\frac{\vec{R}_0}{R_0}\right]\exp\left(i\omega t-ik_0\left|\vec{R}_0-\vec{r}\right|\right)dz\,. \qquad (7.2.14)$$

Надалі, крім умови $l\ll\lambda$, вважатимемо також виконаною умову $l\ll R_0$ (довжина диполя набагато менша за віддаль до точки спостереження). В силу цієї умови також і $r\ll R_0$, тому можна вважати, що $\left|\vec{R}_0-\vec{r}\right|\approx R_0$. Перепозначимо $R_0\to r$ і перейдемо до сферичної системи координат. Тоді після інтегрування по z магнітне поле можна подати у формі:

$$H_\varphi = \frac{i_m l}{rc}\left(\frac{1}{r} + ik_0\right)\exp\left(i\omega t - ik_0 r\right)\sin\theta \qquad (7.2.15)$$

(враховано, що вектор $\vec{r}$ лежить у площині ρ, z, тому векторний добуток $\left[\vec{e}_z \times \left(\vec{R}_0/R_0\right)\right] = \vec{e}_\varphi \sin\theta$ має лише азимутальну компоненту; тут θ – кут між віссю z та напрямком на точку спостереження).

Для знаходження компонент електричного поля врахуємо, що за межами провідника зі струмом виконується співвідношення

$$\vec{E} = -i\left(c/\omega\right)\left[\nabla \times \vec{H}\right] = -\left(i/k_0\right)\left[\nabla \times \vec{H}\right], \qquad (7.2.16)$$

яке випливає з рівняння (7.1.1). Тут зручно перейти до сферичних координат r, θ, φ. З урахуванням аксіальної симетрії задачі та відсутності компонент H_r та H_z маємо:

$$\vec{E} = -\frac{i}{\omega}\left[\nabla \times \vec{H}\right] = -\frac{i}{\omega}\left[\frac{\vec{e}_r}{r\sin\theta}\frac{\partial}{\partial\theta}\left(H_\varphi \sin\theta\right) - \frac{\vec{e}_r}{r}\frac{\partial}{\partial r}\left(rH_\varphi\right)\right] =$$

$$= -i\frac{i_m l}{\omega}\exp\left(i\omega t\right)\left\{\frac{\vec{e}_r}{r\sin\theta}\left(\frac{1}{r^2} + \frac{ik_0}{r}\right)\exp\left(-ik_0 r\right)\frac{\partial}{\partial\theta}\left(\sin^2\theta\right) - \frac{\vec{e}_\theta}{r}\sin\theta\frac{\partial}{\partial r}\left[\left(\frac{1}{r} + ik_0\right)\exp\left(-ik_0 r\right)\right]\right\} =$$

$$= \frac{i_m l}{\omega}\exp\left(i\omega t - ik_0 r\right)\left[\frac{2\vec{e}_r}{r^2}\left(-\frac{i}{r} + k_0\right)\cos\theta + \frac{\vec{e}_\theta}{r}\left(-\frac{i}{r^2} + \frac{k_0}{r} + ik_0^2\right)\sin\theta\right], \qquad (7.2.17)$$

або

$$E_\theta = \frac{i_m l}{\omega r}\left(-\frac{i}{r^2} + \frac{k_0}{r} + ik_0^2\right)\exp\left(i\omega t - ik_0 r\right)\sin\theta; \qquad (7.2.18)$$

$$E_r = \frac{2i_m l}{\omega r^2}\left(-\frac{i}{r} + k_0\right)\exp\left(i\omega t - ik_0 r\right)\cos\theta. \qquad (7.2.19)$$

Видно, що силові лінії електричного поля лежать у площинах, які вміщують вісь диполя, а лінії магнітного поля – в площинах, перпендикулярних до цієї осі (див. рис. 7.2.2, що повторює рис. 7.1.4).

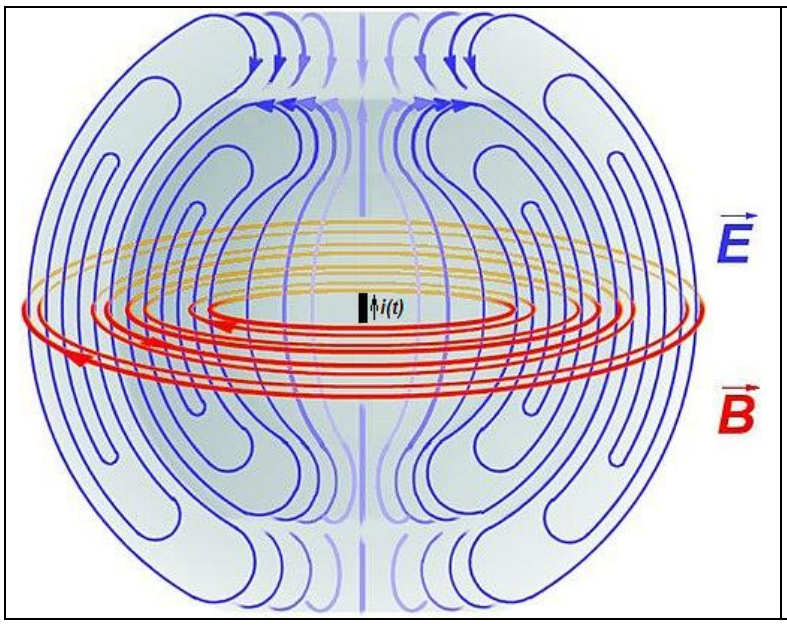


Рис. 7.2.2. Просторовий розподіл електричного (сині лінії) та магнітного (червоні лінії) поля електричного диполя, момент якого гармонічно змінюється з часом

7.2.3. Близька та далека зони випромінювання

Розглянемо так звану близьку зону випромінювання, яка задовольняє умові $r \ll \lambda$, або $k_0 r \ll 1$. У ній основну роль у формулах (7.2.15), (7.2.18), (7.2.19) відіграють доданки з максимальним показником степені r. Маємо:

$$H_\varphi = \left(i_m l/cr^2\right)\exp\left(i\omega t\right)\sin\theta, \qquad \left|H_\varphi\right| \ll \left|E_r\right|, \left|E_\theta\right|; \qquad (7.2.20)$$

$$E_\theta = -i\left(i_m l/\omega r^3\right)\exp\left(i\omega t\right)\sin\theta\,; \tag{7.2.21}$$

$$E_r = -i\left(2i_m l/\omega r^3\right)\exp\left(i\omega t\right)\cos\theta \tag{7.2.22}$$

(див. середню частину рис. 7.2.2).

Нагадаємо, що ці формули справедливі за умови $l \ll r \ll \lambda$, тому $\exp\left(-ik_0 r\right) \approx 1$. Для отримання вимірюваних значень треба взяти дійсну частину формул (7.2.20)-(7.2.22).

Видно, що форма силових ліній електричного поля в близькій зоні є незмінною (див. рис. 7.2.3), а напруженість цього поля змінюється з часом однаково в усіх точках простору. Основну роль відіграє електричне поле, структура якого відповідає відомому полю електричного диполя (пор. з рис. 7.1.3 а), а величина пропорційна миттєвому дипольному моменту $\left(i_m l \sin\omega t\right)/\omega$ (див. формулу (7.2.3)).

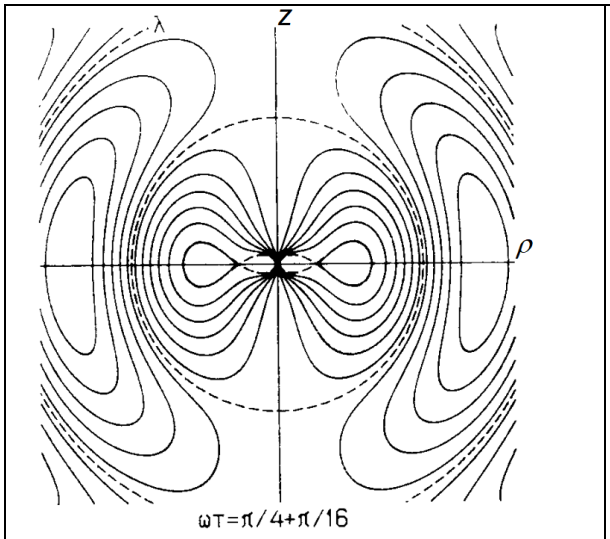


Рис. 7.2.3. Просторовий розподіл силових ліній електричного поля на невеликих віддалях від джерела (в циліндричній системі координат ρ, φ, z; $\rho^2 + z^2 = r^2$)

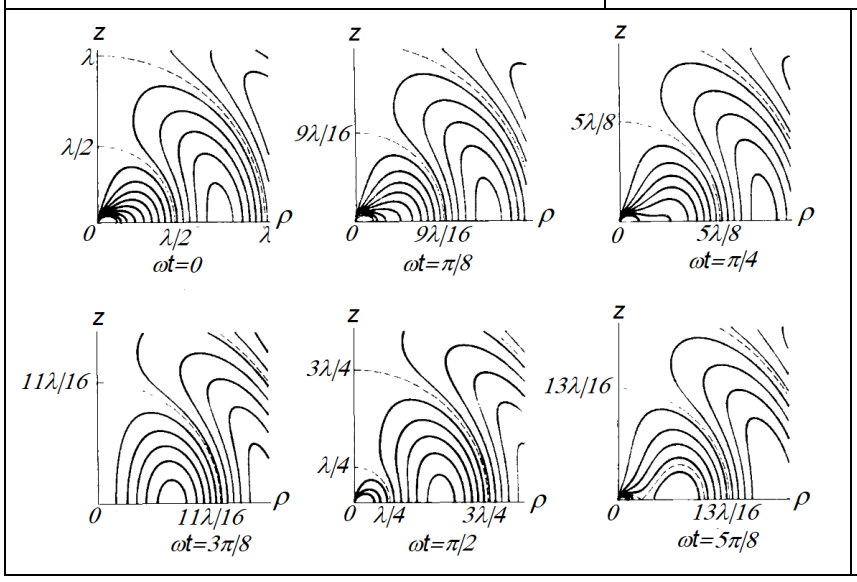


Рис. 7.2.4. Зміна силових ліній електричного поля в просторі з часом на невеликих віддалях від джерела (у близькій зоні та в перехідній області в циліндричній системі координат ρ, φ, z; $\rho^2 + z^2 = r^2$)

Вигляд силових ліній електричного поля в ближній зоні та в перехідній області між ближньою та далекою зонами та його зміна з часом подана на рис. 7.2.3-7.2.4.

У далекій зоні випромінювання, тобто при $k_0 r \gg 1$, основну роль у формулах (7.2.15), (7.2.18-(7.2.19) відіграють доданки, які найповільніше спадають із відстанню:

$$H_\varphi = i\frac{i_m k_0 l}{cr}\exp\left(i\omega t - ik_0 r\right)\sin\theta\,; \tag{7.2.23}$$

$$E_\theta = i\frac{i_m k_0 l}{cr}\exp\left(i\omega t - ik_0 r\right)\sin\theta\,; \tag{7.2.24}$$

$$E_r = \frac{2i_m l}{cr^2}\exp\left(i\omega t - ik_0 r\right)\cos\theta, \qquad |E_r| \ll |H_\varphi|, |E_\theta|. \tag{7.2.25}$$

Як бачимо, поля в далекій зоні мають структуру сферичних хвиль і стають поперечними, а електричне поле виявляється перпендикулярним до магнітного (див. рис. 7.2.5). Електричне й магнітне поля коливаються з однаковою фазою.

Формули для E_θ і H_φ в системі одиниць Гаусса вийшли однаковими. Як уже згадувалося, в системі СІ відношення цих величин має розмірність опору (див. вище п. 6.7.7):

$$\frac{E_\theta}{H_\varphi} = \sqrt{\frac{\mu_0}{\varepsilon_0}} \equiv Z_{0\upsilon} \approx 377\ Ом\ . \tag{7.2.26}$$

Як уже вказувалося, його називають хвильовим опором вакууму, або опором вільного простору.

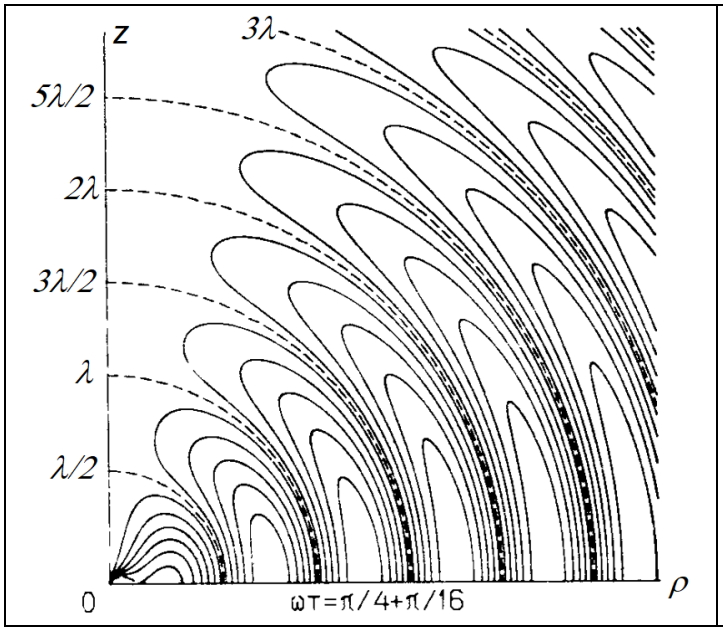


Рис. 7.2.5. Просторовий розподіл силових ліній електричного поля на помірних віддалях від джерела (в циліндричній системі координат ρ, φ, z*;* $\rho^2 + z^2 = r^2$*)*

7.2.4. Диполь Герца як антена

Знаючи розподіл електричного та магнітного полів у далекій зоні випромінювання, неважко знайти радіальну компоненту вектора Пойнтінга (густини потоку енергії, випромінюваної антеною):

$$\Pi_r = \frac{c}{4\pi}E_\theta H_\varphi^* = \frac{i_m^2 l^2 k_0^2}{4\pi c^2 r^2}\sin^2\theta\,. \tag{7.2.27}$$

Величину

$$F(\theta,\varphi) = \sqrt{\frac{\Pi_r}{\Pi_{r\,\max}}} = |\sin\theta| \tag{7.2.28}$$

звичайно називають діаграмою спрямованості[4]. Діаграма спрямованості диполя Герца подана на рис. 7.2.6. Вона аксіально-симетрична, має провал уздовж осі диполя і максимум у напрямку, перпендикулярному до осі.

Інтегруючи радіальну компоненту вектора Пойнтінга по кутах, отримаємо повну потужність випромінювання:

$$P_{rad} = \int_0^{2\pi} d\varphi \int_0^{\pi} d\theta r^2 \Pi_r \sin\theta = \frac{i_m^2 l^2 k_0^2}{4\pi c^2}\int_0^{2\pi} d\varphi \int_0^{\pi} d\theta \sin^3\theta = \frac{i_m^2 l^2 k_0^2}{4\pi c^2} 2\pi \frac{4}{3} = \frac{2}{3}\frac{i_m^2 l^2 k_0^2}{c^2}\,. \tag{7.2.29}$$

У системі СІ ця формула має вигляд

[4] В літературі трапляються й інші означення діаграми спрямованості (див. нижче розділ 7.4)

$$P_{rad} = \frac{\pi}{3} Z_{0v} i_m^2 \left(\frac{l}{\lambda}\right)^2 = \frac{1}{2} i_m^2 R_{rad}. \qquad (7.2.30)$$

Введена таким чином величина

$$R_{rad} = \frac{2\pi}{3} Z_{0v} \left(\frac{l}{\lambda}\right)^2 \qquad (7.2.31)$$

називається опором випромінюванню, або еквівалентним опором антени. На відміну від хвильового опору вакууму, еквівалентний опір антени визначається не через відношення електричного та магнітного полів, а виключно як міра втрат енергії (див. нижче пп. 7.3.2, 7.4.3).

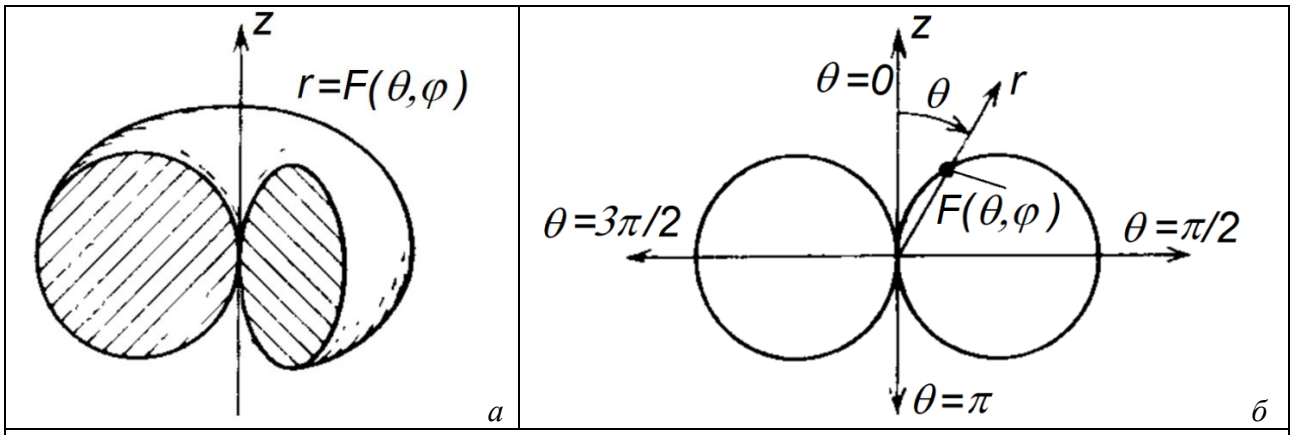


Рис. 7.2.6. Діаграма спрямованості диполя Герца: а – тривимірне зображення; б – переріз

7.2.5. Магнітний диполь Герца

Магнітний диполь Герца, або елементарний магнітний випромінювач, являє собою кільце, через яке протікає змінний струм (рис. 7.2.6). Радіус цього кільця a значно менший за довжину випромінюваних хвиль, тобто струм у будь-якому перерізі кільця в даний момент часу однаковий і змінюється в часі за гармонічним законом (пор. з електричним диполем Герца). Цей струм створює магнітне поле, яке на невеликих відстанях від кільця близьке до поля магнітного диполя (рис. 7.2.7). Магнітний момент цього диполя (в системі одиниць Гаусса) може бути поданий у формі

$$\vec{m} = \frac{1}{c} \vec{e}_z i(t) S = \vec{e}_z \frac{\pi a^2}{c} i(t), \qquad (7.2.32)$$

де вісь z збігається з віссю кільця зі струмом.

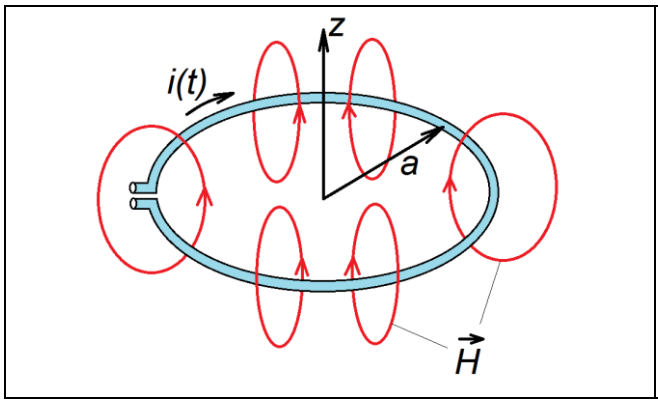


Рис. 7.2.7. Схематичне зображення магнітного диполя Герца та його магнітного поля в ближній зоні

Порівняємо магнітний та електричний моменти відповідних випромінювачів (7.2.32) та (7.2.3), вважаючи, що їхні розміри приблизно однакові, $a \approx l$, і амплітуда струмів однакові. Тоді

$$\left|\frac{\vec{m}_m}{\vec{d}_m}\right| = \frac{\pi a^2 i_m \omega}{c i_m l} \approx \pi k_0 a = 2\pi^2 \frac{a}{\lambda} \ll 1. \qquad (7.2.33)$$

Опускаючи розрахунок, наведемо формули для електромагнітного поля, створеного магнітним диполем Герца на віддалі $r >> a$:

$$E_\varphi = -i\frac{i_m S}{c^2}\left(\frac{1}{r^2}+\frac{ik_0}{r}\right)\exp(i\omega t - ik_0 r)\sin\theta; \qquad (7.2.34)$$

$$H_\theta = i\frac{i_m S}{c^2}\frac{1}{r}\left(-\frac{i}{r^2}+\frac{k_0}{r}+ik_0^2\right)\exp(i\omega t - ik_0 r)\sin\theta; \qquad (7.2.35)$$

$$H_r = i\frac{i_m S}{c^2}\frac{2}{r^2}\left(-\frac{i}{r}+k_0\right)\exp(i\omega t - ik_0 r)\cos\theta \qquad (7.2.36)$$

(рис. 7.2.8).

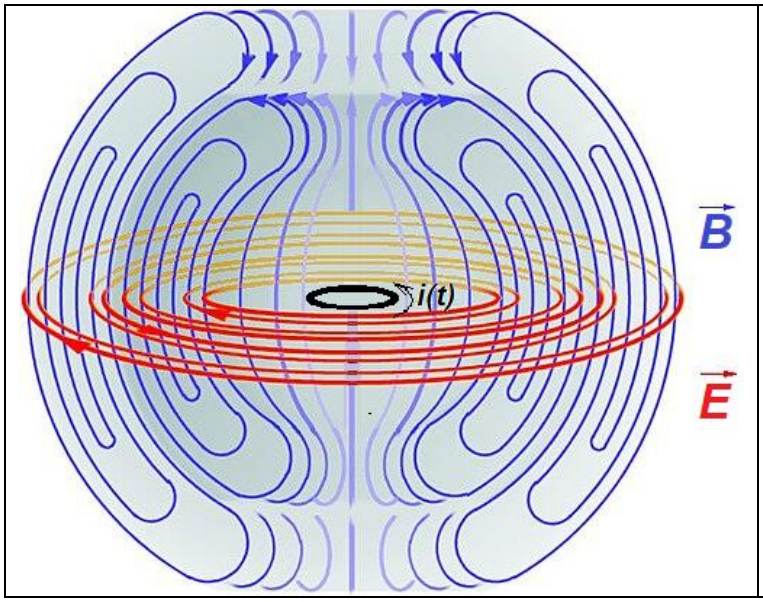


Рис. 7.2.8. Просторовий розподіл магнітного (сині лінії) та електричного (червоні лінії) поля магнітного диполя, момент якого гармонічно змінюється з часом

Для магнітного диполя Герца так само можна ввести поняття близької й далекої зон випромінювання. В близькій зоні магнітне поле буде значно більшим за електричне. В далекій зоні поляризація електромагнітної хвилі буде відрізнятися від звичайного диполя Герца: тепер лінії магнітного поля лежать у площинах, що вміщують вісь диполя, а лінії електричного – в площинах, перпендикулярних до цієї осі.

Радіальна компонента вектора Пойнтінга має вигляд:

$$\Pi_r = \frac{k_0^3}{4\pi c^3 r^2}(i_m S)^2 \sin^2\theta. \qquad (7.2.37)$$

Вона характеризується таким самим кутовим розподілом, як і для електричного диполя Герца (рис. 7.2.9).

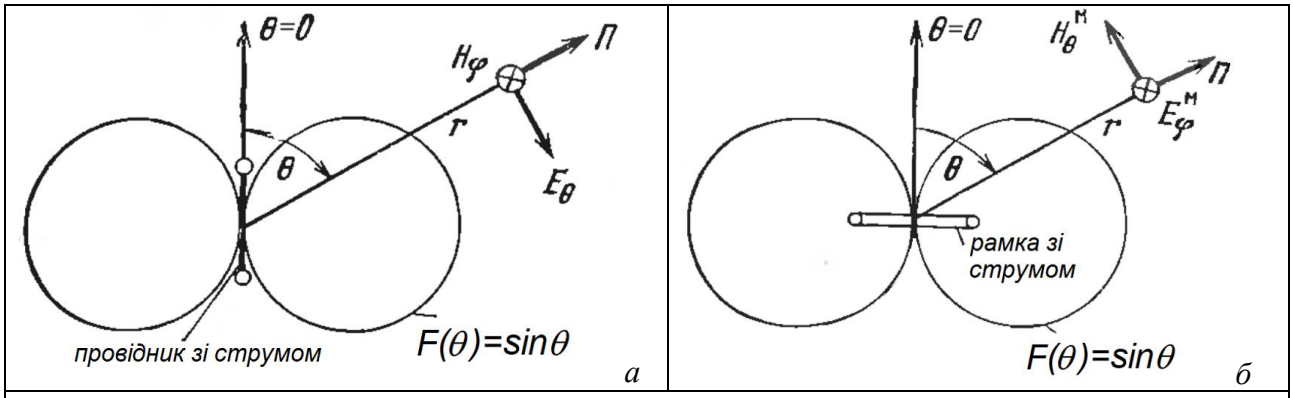


Рис. 7.2.9. Порівняння діаграм спрямованості та поляризації випромінювання елементарного електричного (а) та магнітного (б) випромінювачів

Інтегруючи (7.2.37), можна отримати повну потужність випромінювання:

$$P_{rad} = \frac{k_0^3}{3c^3}(i_m S)^2 . \qquad (7.2.38)$$

Остання виявляється приблизно в $(\lambda/a)^2$ разів меншою, ніж в електричного диполя Герца, що відповідає висловленим раніше міркуванням про співвідношення між величиною дипольного та магнітодипольного випромінювання (див. вище п. 7.1.3). В таку саму кількість разів зменшується еквівалентний опір магнітного диполя Герца.

Слід, однак, підкреслити, що умова $\lambda/a \gg 1$, що визначає точковий випромінювач, потрібна для спрощення розрахунків і в багатьох реальних ситуаціях не справджується. Тому магнітні випромінювачі застосовуються на практиці (див. нижче п. 7.5).

Контрольні питання до підрозділу 7.2

1. Який випромінювач називають (електричним) диполем Герца?
2. Що розуміють під близькою та далекою зоною випромінювання диполя Герца? Який вигляд мають електричне та магнітне поля в цих областях?
3. Чому випромінювання магнітного диполя Герца при тих самих характерних параметрах менше, ніж у електричного диполя Герца?
4. Поясніть характер поляризації випромінювання електричного та магнітного диполя Герца.
5. Який вигляд має діаграма спрямованості диполя Герца?

Задачі до підрозділу 7.2

7.2.1. Оцініть розміри ближньої зони випромінювання диполя Герца для частот 30 кГц та 0.3 ГГц.

7.2.2. Оцініть ширину діаграми спрямованості диполя Герца за аксіальним кутом (на рівні 0.5 за потужністю).

7.2.3. Довжина диполя Герца – 1 м, частота випромінювання – 30 МГц. Знайдіть еквівалентний опір такої антени.

7.2.4. Для умов задачі 7.2.3 знайдіть середню потужність випромінювання диполя Герца, збуджуваного струмом з амплітудою 10 А.

7.2.5. Для умов задачі 7.2.4 знайдіть середню потужність випромінювання магнітного диполя Герца.

7.3. Півхвильовий вібратор

Звичайний диполь Герца (елементарний електричний випромінювач) має, як указувалося раніше, довжину l, яка значно менша від довжини λ випромінюваних ним хвиль, $l \ll \lambda$. Ця обставина спрощує розрахунок, оскільки випромінювання з усіх точок такої антени надходить у точку спостереження з однаковими фазами. Однак з точки зору ефективності кращим є аналог диполя Герца завдовжки $l = \lambda/2$ – так званий півхвильовий вібратор, який може служити прикладом найпростішої (і водночас найпоширенішої) реальної антени. Вища ефективність півхвильового вібратора порівняно з диполем Герца пояснюється використанням явища резонансу, оскільки відрізок провідника, як і відрізок лінії передачі, можна розглядати як резонатор (див. вище п. 6.7).

Півхвильові вібратори використовуються також як окремі випромінювачі, з яких зібрані складні антени (див. нижче п. 7.5).

7.3.1. Півхвильовий вібратор як резонатор

Конструкція півхвильового вібратора така сама, як у диполя Герца (рис. 7.2.1). Як уже вказувалося, відмінність лише в тому, що у півхвильового вібратора $l = \lambda/2$. Ця відмінність приводить до того, що в різних точках вібратора струм і напруга будуть неоднаковими.

При розгляді хвиль у двопровідних лініях відзначалося, що відрізки таких ліній, у яких встановлюється стояча хвиля, можуть служити резонаторами (див. вище п. 6.7). Відповідно й відрізок провідника, в середній точці якого увімкнене джерело змінного струму, також може служити резонатором. Очевидно, на його кінцях струм має дорівнювати нулеві, відповідно там виникають вузли струму й пучності напруги (вважаємо, що напруга відраховується від середньої точки провідника) – див. рис. 7.3.1. В цьому сенсі півхвильовий вібратор аналогічний відрізку двопровідної лінії з розімкнутими кінцями (див. вище п. 6.7.3). Основна мода такого відрізку якраз і характеризується довжиною хвилі $\lambda = 2l$. Оскільки вона має пучність струму в центрі антени, то саме там доцільно розміщувати джерело змінного струму, яке її збуджує.

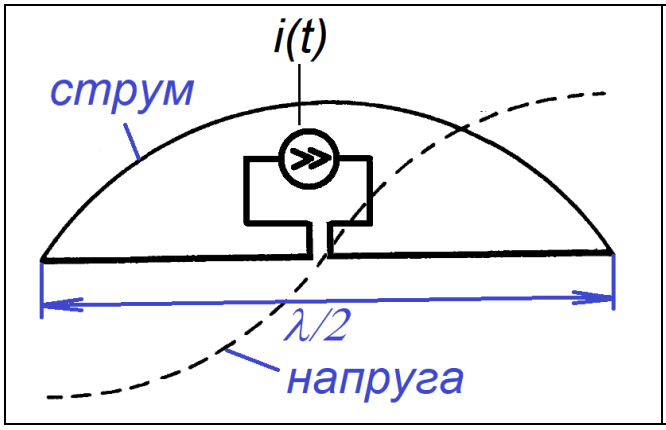


Рис. 7.3.1. Розподіл напруги та струму в ідеальному півхвильовому вібраторі

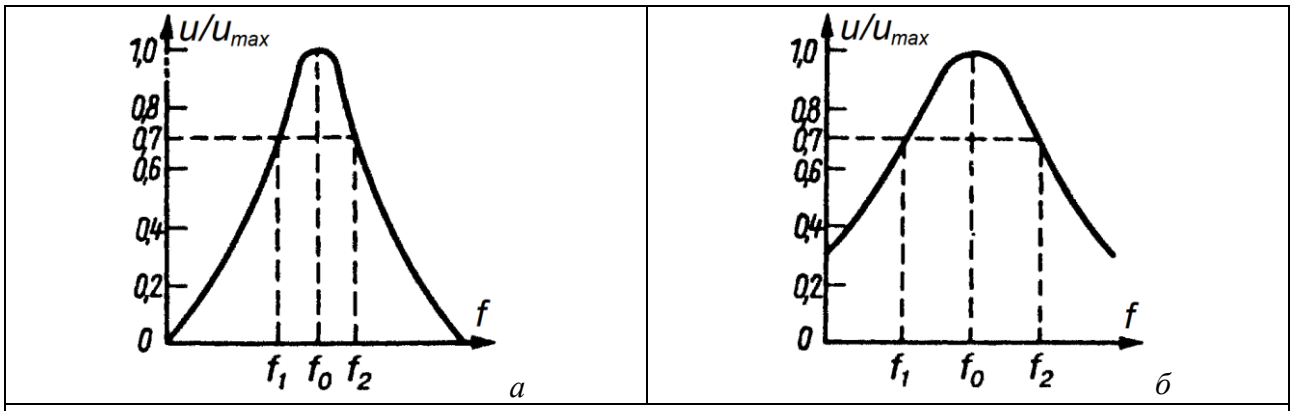


Рис. 7.3.2. Графіки резонансних кривих півхвильового вібратора при більшій (а) та меншій (б) добротності

Порівняно з відрізком двопровідної лінії (див. вище п. 6.7.6) півхвильвий вібратор характеризується додатковим джерелом втрат – втратами на випромінювання. Тому його добротність Q загалом виявляється значно меншою (рис. 7.3.2), ніж у відрізку лінії, але все ще набагато більшою від одиниці. Як і для відрізку лінії, резонанс виявлений тим більше, чим більше відношення L/C (за однакових інших умов).

За рахунок резонансних властивостей півхвильового вібратора при однаковій потужності живлення амплітуда коливань струму й напруги в ньому буде в Q разів більше, ніж у диполі Герца. Відповідно випромінювана потужність буде більшою в Q^2 разів.

7.3.2. Параметри півхвильового вібратора

В узгодженому півхвильовому вібраторі максимальні струм i_m і напруга u_m пов'язані через хвильовий опір вібратора Z_0:

$$u_m = Z_0 i_m, \qquad Z_0\,[Ом] = 120\ln\left(0.575\frac{l}{d}\right), \tag{7.3.1}$$

де l – довжина вібратора, d – діаметр провідника, з якого він виготовлений.

Опір випромінюванню визначається так само, як для диполя Герца (див. формулу (7.2.30)):

$$R_{rad} = \frac{2P_{rad}}{i_m^2} \tag{7.3.2}$$

(індекс – від англ. radiation, випромінювання), де P_{rad} – середня потужність, випромінювана вібратором.

Коректний розрахунок опору випромінюванню для півхвильового вібратора виявляється значно складнішим, ніж для диполя Герца. Тим не менше для його оцінки можна скористатися формулою (7.2.31)

$$R_{rad} = \frac{2\pi}{3} Z_{0v}\left(\frac{l}{\lambda}\right)^2,$$

підставивши туди $l = \lambda/2$. Отримуємо, що опір випромінювання півхвильового вібратора приблизно в $(\lambda/l)^2$ разів більший, ніж у диполя Герца. Нагадаємо, що для останнього $l \ll \lambda$.

Аналогічно до опору випромінювання вводиться опір втрат

$$R_{loss} = \frac{2P_{loss}}{i_m^2} \tag{7.3.3}$$

(індекс – від англ. losses, втрати), де $P_{loss} = P - P_{rad}$ – потужність втрат у вібраторі, P – повна потужність, що підводиться до антени від фідера.

З (7.3.2)-(7.3.3) випливає формула для коефіцієнту корисної дії η півхвильового вібратора. Це відношення потужності випромінювання P_r до підведеної до вібратора потужності:

$$\eta = \frac{P_{rad}}{P} = \frac{R_{rad}}{R_{rad} + R_{loss}}. \tag{7.3.4}$$

Аналогічно визначається к.к.д. будь-якої антени.

Опір втрат R_{loss} у півхвильовому вібраторі звичайно зумовлений поверхневим опором провідника, з якого він виготовлений, та діелектричними втратами в ізоляторах[5]. Для півхвильового вібратора зазвичай $R_{rad} \ll R_{loss}$ і, відповідно, $\eta \ll 1$. В реальних ситуаціях величини R_{rad} та R_{loss} залежать від того, на якій висоті над землею розміщений вібратор, які об'єкти розташовані поруч із ним, тощо.

Амплітуда струму вібратора визначається формулою

[5] В раніше розглянутій моделі диполя Герца ми нехтували розмірами вібратора, тому в цій моделі опір втрат взагалі відсутній.

$$i_m = \sqrt{\frac{2P}{R_{rad} + R_{loss}}}, \qquad (7.3.5)$$

Вхідний імпеданс Z_{in} антени (підкреслимо, що він відрізняється від введеного раніше хвильового опору, як це має місце й у двопровідних лініях) звичайно визначають між точками, до яких умикається живлення (фідер). Для півхвильового вібратора це середина провідника (рис. 7.3.1). В режимі точного резонансу вхідний імпеданс є суто активним і мінімальним за величиною, оскільки на середині вібратора виникає вузол напруги й пучність струму (див. рис. 7.3.1). Для нескінченно тонкого провідника розрахований імпеданс складає 73 Ом. Насправді при ненульовому значенні d ця величина виявляється меншою (рис. 7.3.3).

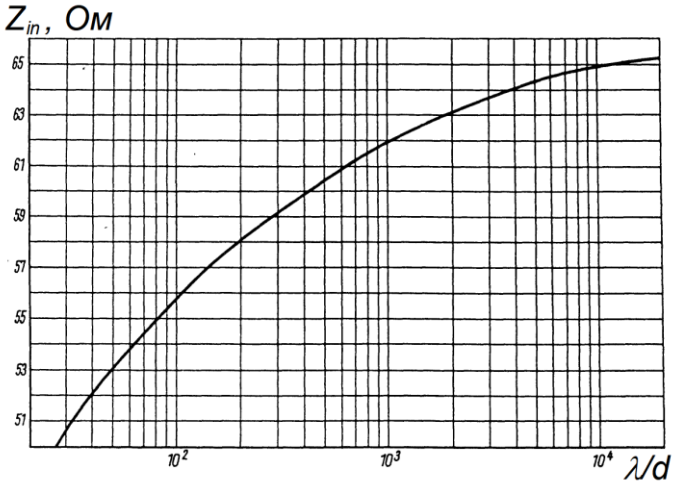


Рис. 7.3.3. Графік залежності вхідного опору півхвильового вібратора від його діаметру

На практиці розрізняють механічну та електричну довжини антени. Механічна довжина – це власне довжина провідника, електрична довжина – це половина довжини хвилі на резонансній частоті. Ці величини збігаються лише для нескінченно тонких провідників у вакуумі.

В реальності електрична довжина завжди більша за механічну. Це пов'язано, по-перше, з ненульовим діаметром провідника, по-друге –із впливом його кінців. Річ у тім, що внаслідок ненульової провідності повітря й неідеальності ізоляторів струм на кінцях провідника відмінний від нуля. Тому на його довжині вкладається менше півхвилі струму. Ми стикалися зі схожою ситуацією, коли вводили ефективну довжину відрізку довгої лінії, зумовлену реактивними навантаженнями, увімкненими на її кінцях (див. вище п. 6.7.5, формула (6.7.18)).

Зазвичай відношення механічної довжини до електричної (його називають коефіцієнтом укорочення) лежить у межах від 0.8 до 0.95.

Форма діаграми спрямованості півхвильового вібратора в порожньому просторі аксіально симетрична й близька до діаграми спрямованості диполя Герца (рис. 7.2.5). Це пов'язано з тим, що в будь-якому напрямку різниця фаз випромінювання, збудженого його окремими ділянками, не перевищує π. Справді, найгірший випадок відповідає випромінюванню вздовж осі вібратора ($\theta = 0$), коли різниця ходу між променями від його кінців складає величину $\Delta r = \lambda/2$, що й відповідає різниці фаз $k\Delta r = (2\pi/\lambda)(\lambda/2) = \pi$. Але тут варто нагадати, що якраз у цьому напрямку вібратор не випромінює (див. формули (7.2.27)-(7.2.28) та рис. 7.2.5). Таким чином, у тій області значень аксіального кута θ, де густина потоку енергії Π_r найбільша, вплив ненульової довжини вібратора буде мінімальним.

7.3.3. Вплив провідної поверхні на діаграму спрямованості

На практиці антени звичайно встановлюють на порівняно невеликій висоті над землею, дахом споруди, корпусом корабля або літака, тощо. В радіодіапазоні землю (особливо вологу землю) можна вважати гарним провідником (див. нижче пп. 8.3-8.4). Тому найпростішою моделлю, що описує вплив поверхні на діаграму спрямованості півхвильового вібратора, є розташування такого вібратора над ідеально провідною поверхнею (рис. 7.3.4, ліві частини

рисунків).

З урахуванням граничних умов на ідеально провідній поверхні картина електричного поля виглядає так, ніби по інший бік цієї поверхні існує електростатичне відображення антени (рис. 7.3.4, праві частини рисунків). В результаті поле являє собою інтерференційну картину випромінювання антени та її відображення.

Нехай півхвильовий вібратор розташований горизонтально на висоті h над провідною поверхнею (рис. 7.3.4 б). Перерізи його діаграми спрямованості вертикальною площиною, перпендикулярною до осі вібратора, що проходить через його середину, подані на рис. 7.3.5 для різних значень параметра h/λ .

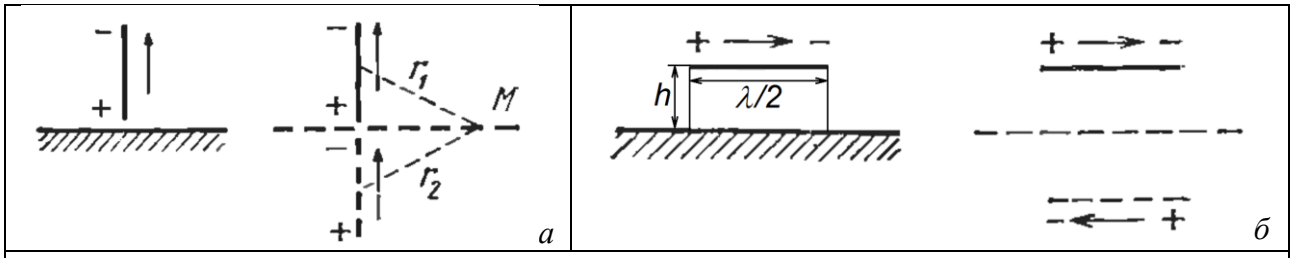


Рис. 7.3.4. Вертикальний (а) та горизонтальний (б) вібратор над провідною поверхнею (ліворуч) та створене ним електростатичне відображення (праворуч)

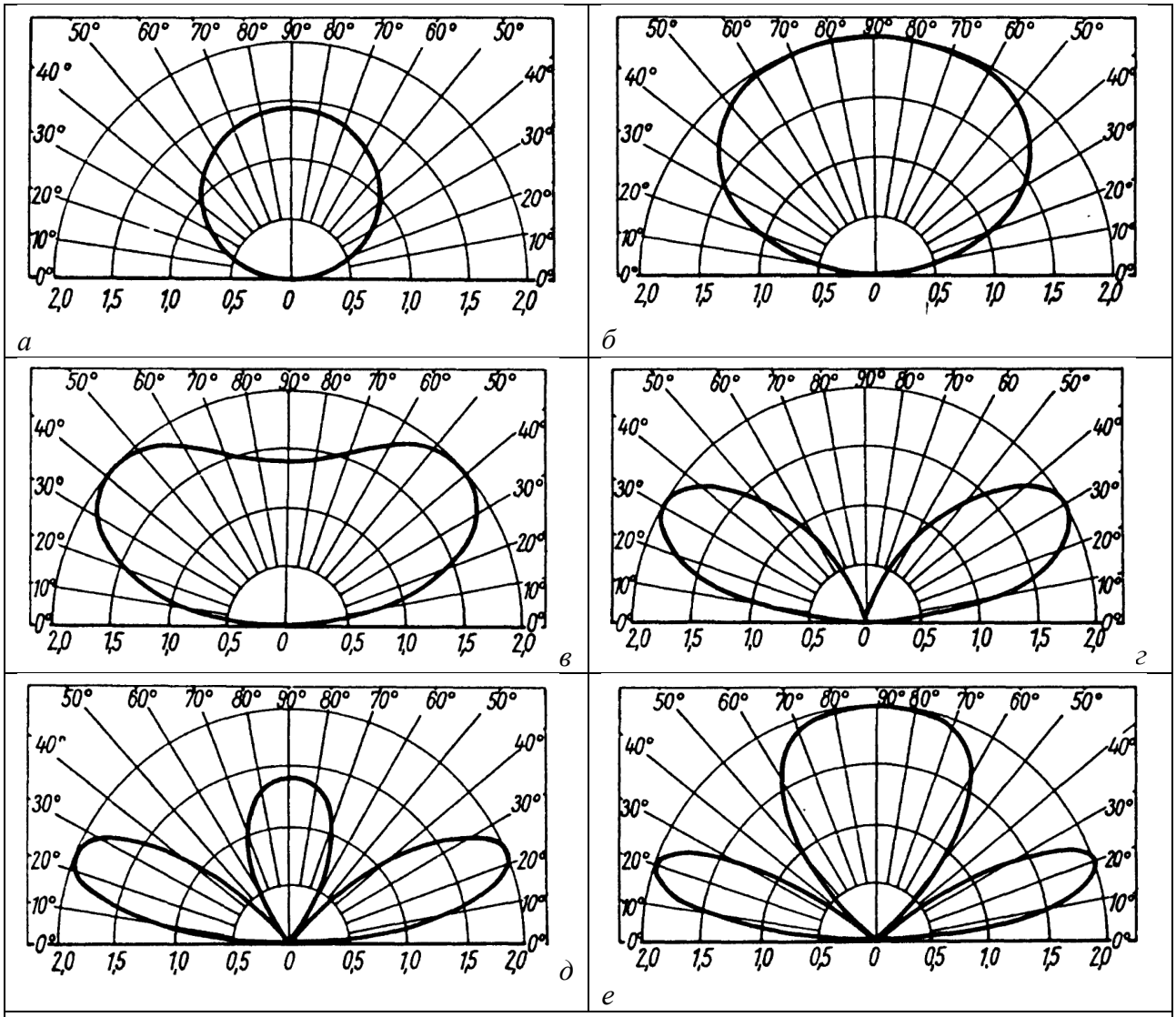
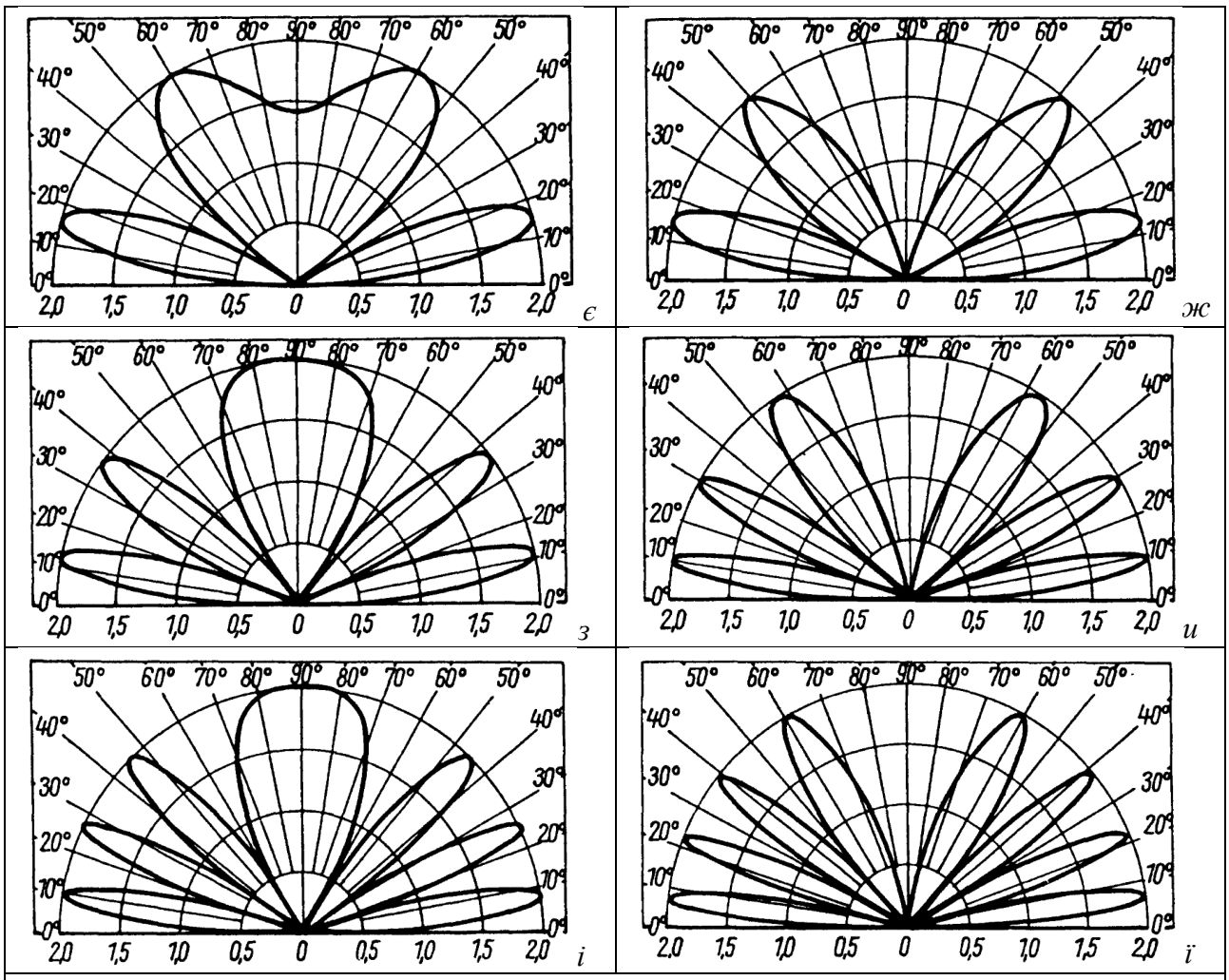


Рис. 7.3.5. Діаграми спрямованості горизонтально розташованого півхвильового вібратора в залежності від висоти над провідною поверхнею: а – h/λ =1/8; б – 1/4; в – 3/8; г – 1/2; д –5/8; е – 3/4



Продовження рис. 7.3.5: є – 7/8; ж – 1; з – 5/4; и – 3/2; і – 7/4; ї – 2

При малих значеннях h/λ випромінювання антени та її відображення значною мірою компенсуються (це помітно на рис. 7.3.5 а). Випадок рис. 7.3.5 б є аналогом квадрупольного випромінювача ($2h = \lambda/2$), коли умовні змінні заряди розташовані на вершинах квадрата (див. вище п. 7.1.4).

Зі зростанням віддалі між випромінювачем та його відображенням діаграма спрямованості стає все більш порізаною (рис. 7.3.5 в-ї). В цьому випадку антену та її відображення можна розглядати як точкові випромінювачі (рис. 7.3.6).

Знайдемо різницю ходу променів, що йдуть у точку спостереження від антени та її відображення. Для цього знайдемо спершу віддалі від них до точки спостереження:

$$R_{1,2} = \sqrt{(R_0\cos\alpha)^2 + (R_0\sin\alpha \mp h)^2} \approx \sqrt{R_0^2 \mp 2R_0 h\sin\alpha} \approx R_0 \mp h\sin\alpha \qquad (7.3.6)$$

(результат отримано в припущенні, що $R_0 \gg h$). Тоді

$$\Delta R(\alpha) \equiv R_2 - R_1 = 2h\sin\alpha\,. \qquad (7.3.7)$$

Нехай кути α_i та α_{i+1} відповідають напрямкам на сусідні максимуми діаграми спрямованості. Тоді можна записати, що

$$\Delta R(\alpha_{i+1}) - \Delta R(\alpha_i) = \lambda\,. \qquad (7.3.8)$$

Якщо $h \gg \lambda$, то можна вважати, що $|\Delta\alpha| \ll \alpha_i$, де $\Delta\alpha \equiv \alpha_{i+1} - \alpha_i$. Тоді, підставивши (7.3.7) до (7.3.8) і опускаючи індекс у α_i, маємо з точністю до першого порядку за $\Delta\alpha$:

$$\begin{aligned} &2h\sin(\alpha+\Delta\alpha) - 2h\sin\alpha = 2h(\sin\alpha\cos\Delta\alpha + \sin\Delta\alpha\cos\alpha - \sin\alpha) \approx \\ &\approx 2h(\sin\alpha + \Delta\alpha - \sin\alpha) = 2h\Delta\alpha = \lambda, \end{aligned} \qquad (7.3.9)$$

звідки

$$\Delta\alpha = \frac{\lambda}{2h}. \qquad (7.3.10)$$

Як бачимо, кутова віддаль між сусідніми пелюстками діаграми спрямованості справді зменшується зі зростанням висоти антени над провідною площиною (пор. з рис. 7.3.5 ж-ї).

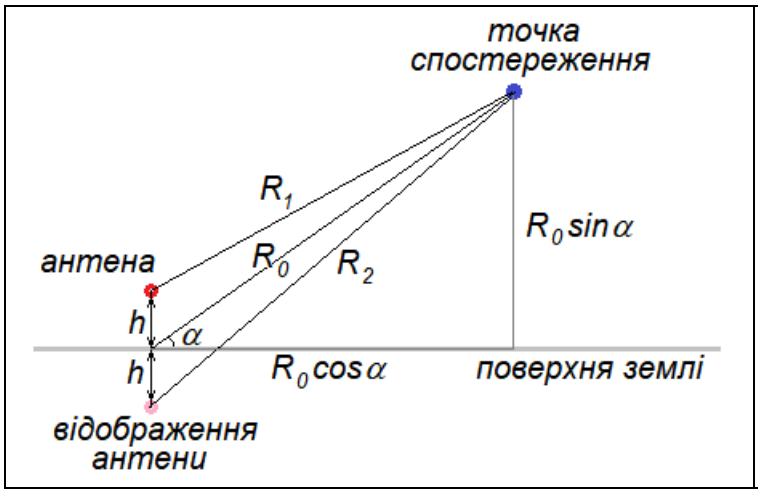


Рис. 7.3.6. Переріз через середню точку півхвильового вібратора вертикальною площиною, перпендикулярною до його осі

Контрольні питання до підрозділу 7.3

1. Порівняйте між собою диполь Герца і півхвильовий вібратор. Що у них спільне і що – відмінне?
2. Чим відрізняється електрична довжина півхвильового вібратора від механічної?
3. Як впливає поверхня Землі на поле випромінювання горизонтального півхвильового вібратора, розташованого на деякій висоті?
4. Чому при зростанні висоти розміщення горизонтального півхвильового вібратора його діаграма спрямованості стає все більш порізаною?

Задачі до підрозділу 7.3

7.3.1. Довжина півхвильового вібратора – 5 м, діаметр провідника, з якого він зроблений – 5 мм. Знайдіть його хвильовий опір.

7.3.2. Для півхвильового вібратора, описаного в задачі 7.3.1, знайдіть вхідний опір. Скористайтеся рис. 7.3.3.

7.3.3. Горизонтально розташований над поверхнею Землі півхвильовий вібратор має довжину 1м, частота випромінювання – 3 МГц. При яких висотах антени випромінювання вгору буде відсутнім? Поверхню Землі вважати ідеально провідною площиною.

7.4. Основні параметри та характеристики антен

Після того, як ми проаналізували властивості електричного й магнітного диполів Герца та півхвильового вібратора, можна тепер детальніше розглянути загальні характеристики та параметри антен різних типів.

Перш за все, нагадаємо, що антени використовуються для випромінювання та приймання електромагнітних хвиль. У теорії антен доводиться так звана теорема взаємності (див. нижче п. 7.4.5). З неї, зокрема, випливає, що діаграми спрямованості антени на передавання сигналів та на їх приймання однакові. Це дозволяє користуватися однією й тією самою антеною для випромінювання та приймання хвиль, що й використовується на практиці (наприклад, у радіостанціях, в активних радіолокаторах, тощо). Ця обставина спрощує теоретичний аналіз антен і зменшує кількість їхніх характеристик і параметрів.

До основних характеристик та параметрів передавальних антен належать:

- характеристика спрямованості;
- фазова характеристика;
- поляризаційна характеристика;
- коефіцієнт спрямованої дії;
- коефіцієнт корисної дії;
- коефіцієнт підсилення;
- робочий діапазон частот (смуга пропускання);
- ефективна площа чи ефективна довжина антени;
- опір випромінювання;
- вхідний опір.

Перейдемо до визначення та детального обговорення цих величин.

7.4.1. Характеристика спрямованості передавальної антени

Ізотропний випромінювач електромагнітних хвиль, який випромінює в усіх напрямках однаково – це спрощена модель, яку для поперечних хвиль не можна реалізувати принципово (див. вище п. 7.1.1). Будь-яка антена є спрямованою, тобто випромінює по-різному в різних напрямках. Цю властивість описують характеристикою спрямованості. Розрізняють характеристики спрямованості за полем і за потужністю.

Характеристикою спрямованості передавальної антени за полем називають залежність амплітуди поля від напрямку випромінювання (тобто від полярного кута θ та азимутального кута φ в сферичній системі координат), нормовану на її максимальне значення. Точніше, йдеться про величину:

$$F(\theta,\varphi) = \mathrm{Re}\left\{E_{\mathrm{m}}(\theta,\varphi)/E_{\max}\right\} = \mathrm{Re}\left\{H_{\mathrm{m}}(\theta,\varphi)/H_{\max}\right\}, \qquad (7.4.1)$$

де $E_{\mathrm{m}}(\theta,\varphi)$ та $H_{\mathrm{m}}(\theta,\varphi)$ – комплексні амплітуди полів у напрямку (θ,φ) (в сферичній системі координат), $E_{\max}$ та $H_{\max}$ – їхні максимальні значення. Амплітуди полів беруться на сферичній поверхні, центром якої служить антена, а радіус значно більший за довжину хвилі (тобто ця поверхня розташована в далекій зоні випромінювання).

Функцію $F(\theta,\varphi)$ інколи також називають нормованою характеристикою спрямованості. За означенням $|F(\theta,\varphi)| \le 1$.

Формулу (7.4.1) можна переписати у формі

$$F(\theta,\varphi) = \left|E_{\mathrm{m}}(\theta,\varphi)/E_{\max}\right| \mathrm{Re}\exp\left[i\psi(\theta,\varphi)\right] = \left|E_{\mathrm{m}}(\theta,\varphi)/E_{\max}\right| \cos\left[\psi(\theta,\varphi)\right], \qquad (7.4.2)$$

де $\psi(\theta,\varphi)$ - різниця фаз між променем, що відповідає кутам (θ,φ) і променем, орієнтованим на максимальне значення $|F(\theta,\varphi)|$ (див. нижче п. 7.4.2).

Проілюструємо поняття характеристика спрямованості на прикладі електричного диполя Герца (див. вище п. 7.2). Відповідно до формул (7.2.23)-(7.2.24), амплітуда його поля в далекій зоні –

$$E_{m\theta} = H_{m\varphi} = \left(i_m k_0 l/cr\right)\sin\theta .$$

Отже, характеристика спрямованості за полем – $F(\theta,\varphi)=\sin\theta$ (пор. з формулою (7.2.28)).

Якщо випромінювач можна вважати точковим, тобто його характерний розмір l значно менший за довжину випромінюваних хвиль, $l \ll \lambda$, то фаза поля, збуджувана в усіх точках згаданої сфери, однакова, $\psi(\theta,\varphi)=0$, а характеристика спрямованості невід'ємна, $F(\theta,\varphi)\geq 0$. Саме така картина має місце, як було щойно показано, для електричного диполя Герца.

Якщо ж розміри випромінювача сумірні з довжиною хвилі або перевищують її, $l \geq \lambda$, то хвилі в різних точках сфери можуть мати відмінну фазу. Як видно з формули (7.4.2), при переході $\psi(\theta,\varphi)$ через значення $\pi/2+\pi n,\ n=\pm1,\pm2,\pm3...$ функція $F(\theta,\varphi)$ змінює знак.

У п. 7.3.3 розглядалася задача про випромінювання точкової антени, розміщеної на висоті h над провідною поверхнею. Там було показано, що випромінювання від самої точкової антени та її електростатичного відображення може додаватися як у фазі, так і в протифазі (див. формулу (7.3.6)), формуючи максимуми й мінімуми характеристики спрямованості.

У сферичних координатах функція $F(\theta,\varphi)$ задає деяку поверхню (див. вище рис. 7.2.5). В залежності від форми цієї поверхні розрізняють тороїдальні, голчасті, віялоподібні, багатопелюсткові[6] та інші характеристики спрямованості (рис. 7.4.1).

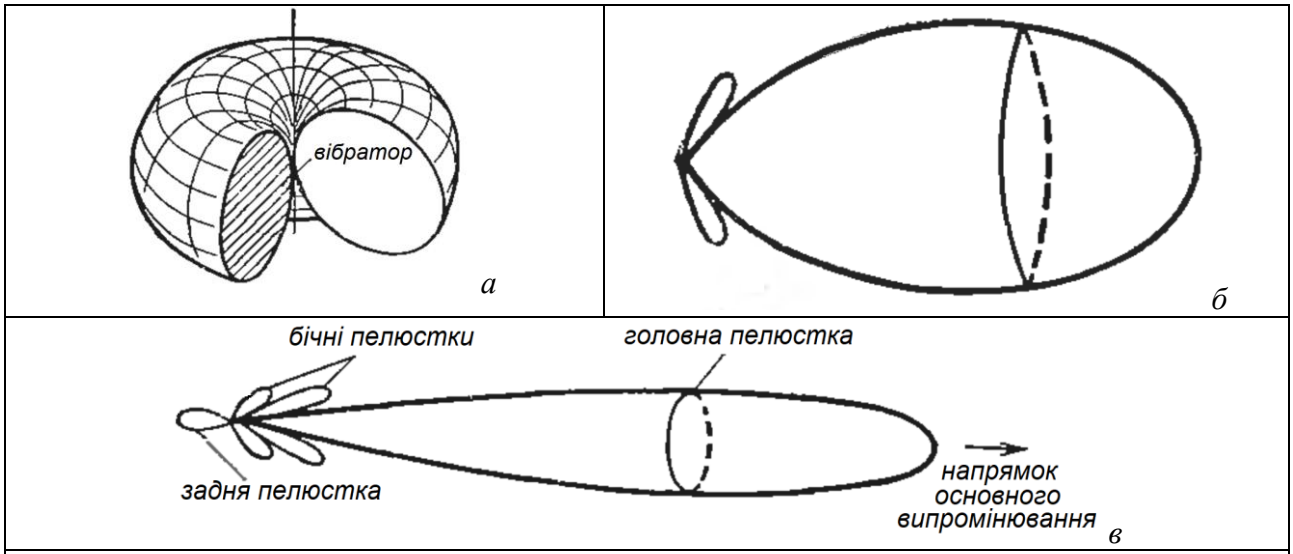


Рис. 7.4.1. Тривимірні (а) та площинні (б, в) характеристики спрямованості: а – тороїдальна (для електричного вібратора); б – віялоподібна; в – голчаста з бічними пелюстками

Як правило, на практиці будують перерізи характеристики спрямованості в горизонтальній та вертикальній площинах (рис. 7.4.2). Їх інколи називають площинними характеристиками спрямованості. Для їхньої побудови можна використовувати й декартові координати (рис. 7.4.3.). Параметром характеристики спрямованості є її півширина за кутом (рис. 7.4.4). Вона може визначатися як у горизонтальній, так і у вертикальній площині.

Характеристикою спрямованості передавальної антени за потужністю називають нормовану на своє максимальне значення кутову залежність радіальної компоненти вектора Пойнтінга від напрямку випромінювання. Оскільки вектор Пойнтінга пропорційний квадрату амплітуди поля випромінювання, то характеристика спрямованості за потужністю є квадратом характеристики спрямованості за полем, тобто $F^2(\theta,\varphi)$. Таким чином, справедливе

[6] Характеристику спрямованості називають багатопелюстковою, якщо вона має кілька напрямків максимального випромінювання і, відповідно, кілька напрямків нульового випромінювання (див., наприклад, рис. 7.3.5 ж-ї). Якщо одна пелюстка більша за інші, вона називається головною. Багатопелюсткові характеристики спрямованості притаманні антенам, розміри яких більші за довжину випромінюваних хвиль, і зумовлені інтерференцією випромінювання від різних частин антени (пор. з п. 7.3.3).

співвідношення

$$\Pi_r(r,\theta,\varphi) = \Pi_{r\ \max}(r) F^2(\theta,\varphi). \qquad (7.4.3)$$

В технічній літературі всі названі тут різновиди характеристик спрямованості зазвичай називають діаграмами спрямованості (на випромінювання) – див. вище п. 7.2.4.

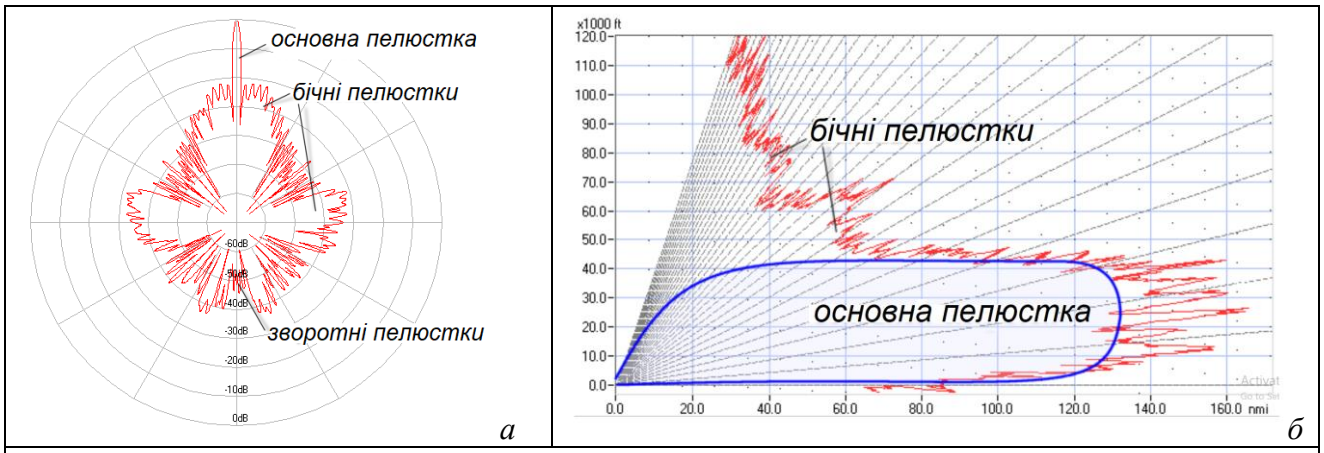


Рис. 7.4.2. Переріз віялоподібної характеристики спрямованості в горизонтальній (а) та вертикальній (б) площинах

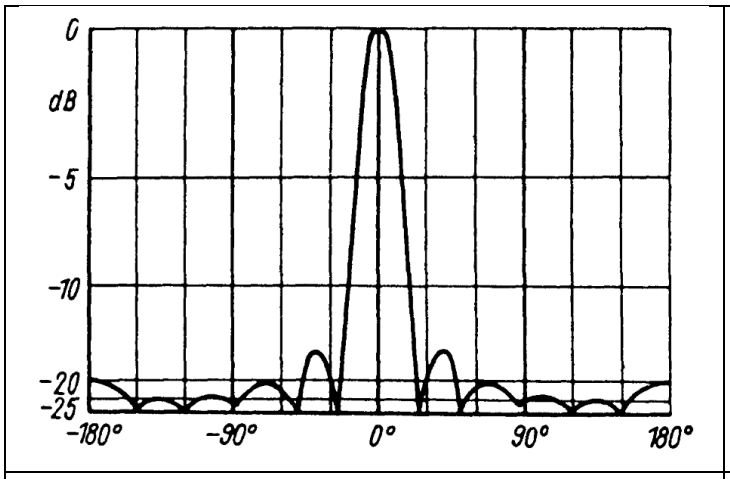


Рис. 7.4.3. Аксіально-симетрична характеристика спрямованості в декартових координатах

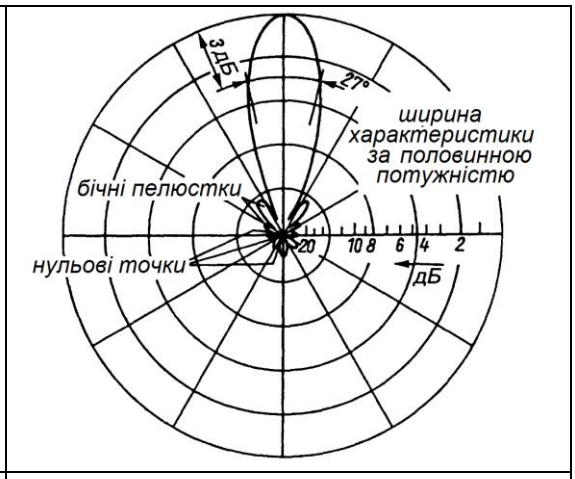


Рис. 7.4.4. Визначення ширини діаграми спрямованості за кутом (у горизонтальній площині)

7.4.2. Фазова та поляризаційна характеристики

Фазова характеристика антени – це залежність фази хвилі від напрямку випромінювання (на фіксованій віддалі від антени). Фактично йдеться про раніше введену функцію $\psi(\theta,\varphi)$.

При переході функції $F(\theta,\varphi)$ через нуль, як уже вказувалося, вона змінює свій знак, тобто при цьому значення $\cos[\psi(\theta,\varphi)]$ стрибком змінюється від 1 до -1 або навпаки.

Фазова характеристика антени $\psi(\theta,\varphi)$ визначається як залежність початкової фази хвилі від напрямку її поширення. Таким чином, у далекій зоні випромінювання будь-яка (комплексна) компонента електромагнітного поля може бути записана у формі

$$E(\vec{r},t) = \frac{Ai_m}{r} F(\theta,\varphi) \exp[i\omega t - ik_0 r + i\psi(\theta,\varphi)], \qquad (7.4.4)$$

де i_m – (комплексна) амплітуда струму антени, A – амплітудний коефіцієнт, величина якого залежить від типу антени.

Як уже вказувалося (див. вище п. 6.8.4), поляризація електромагнітної хвилі

визначається законом зміни напрямку вектора напруженості електричного поля в даній точці простору з часом. Якщо згаданий вектор залишається паралельним (чи антипаралельним) до свого початкового напрямку, говорять про лінійну поляризацію. Площина, яка вміщує вектор напруженості електричного поля і хвильовий вектор, називається площиною поляризації такої хвилі (рис. 7.4.5 а).

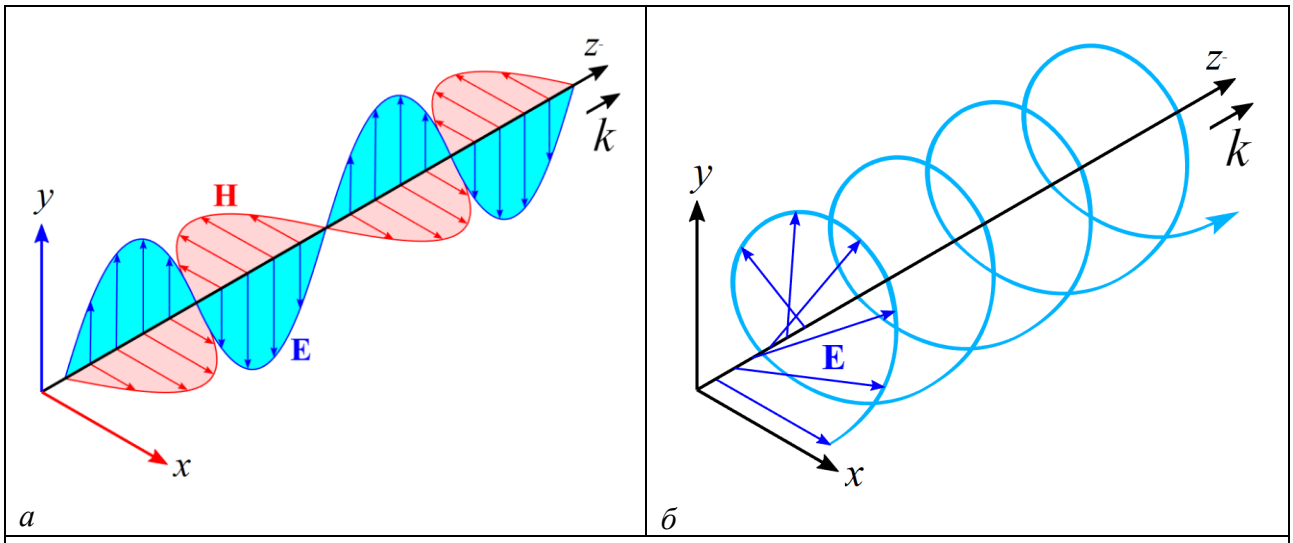


Рис. 7.4.5. Лінійно поляризована (а) та циркулярно поляризована (б) хвилі (остання – з лівою поляризацією)

Якщо кінець вектора напруженості електричного поля в даній точці простору з часом рівномірно обертається, а його довжина зберігається, говорять про циркулярну (кругову) поляризацію хвилі (рис. 7.4.5 б). Таке поле може бути подане як сума двох лінійно поляризованих хвиль, фази яких зсунуті на 90°. В залежності від напрямку обертання говорять про праву і ліву поляризацію. Якщо дивитися вздовж напрямку хвильового вектора, в першому випадку вектор напруженості електричного поля обертається праворуч (за годинниковою стрілкою), в другому – ліворуч (проти годинникової стрілки).

У найбільш загальному випадку, коли нормальні компоненти вектора напруженості електричного поля зсунуті на 90° і мають відмінні амплітуди, говорять, що хвиля є еліптично поляризованою (рис. 7.4.6):

$$\vec{E}(t) = \vec{e}_x E_{m1} \cos\omega t + \vec{e}_y E_{m2} \sin\omega t . \qquad (7.4.5)$$

Якщо $E_{m1} = 0$ (або $E_{m2} = 0$), маємо лінійно поляризовану хвилю, якщо $E_{m1} = E_{m2}$ – циркулярно поляризовану.

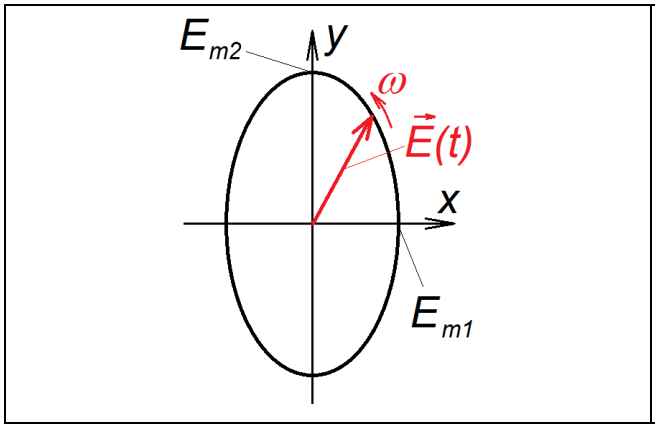


Рис. 7.4.6. Зміна електричного поля з часом для еліптично поляризованої хвилі, що поширюється вздовж осі z

Нехай $E_{m1} \le E_{m2}$. Тоді відношення $p = E_{m1}/E_{m2}$ $(0 \le p \le 1)$ називають коефіцієнтом поляризації, або коефіцієнтом еліптичності. Очевидно, для лінійної поляризації $p = 0$, для

циркулярної – $p=1$, для еліптичної – $0<p<1$. Залежність коефіцієнту поляризації від напрямку випромінювання $p(\theta,\varphi)$ в далекій зоні випромінювання називають *поляризаційною характеристикою* антени.

Випромінювання як електричного, так і магнітного диполів Герца (див. вище п. 7.2) є плоскополяризованим, тому для них $p(\theta,\varphi)=0$. Те саме стосується й півхвильового вібратора.

7.4.3. Коефіцієнт спрямованої дії та коефіцієнт підсилення

Якби антена мала ізотропну діаграму спрямованості, $F(\theta,\varphi)=1$, (що, як уже вказувалося, неможливо), то густину потоку енергії її випромінювання на віддалі r можна було б записати, як $\Pi_{avr}(r)=P_{rad}/4\pi r^2$ (індекс – від англ. average, середній), де, як і раніше, P_{rad} – повна потужність випромінювання. Величина $\Pi_{avr}(r)$ визначає густину потоку енергії ізотропного випромінювача заданої потужності P_{rad}. Тоді величина

$$K_{dir}(\theta,\varphi)=\frac{\Pi_r(r,\theta,\varphi)}{\Pi_{avr}(r)}=\frac{4\pi r^2\Pi_r(r,\theta,\varphi)}{P_{rad}} \tag{7.4.6}$$

називається *коефіцієнтом спрямованої дії* (к.с.д.) антени (індекс – від англ. direction, напрямок). Врахувавши співвідношення (7.4.3), можна пов'язати к.с.д. та характеристику спрямованості:

$$K_{dir}(\theta,\varphi)=\frac{\Pi_{r\,\max}(r)}{\Pi_{avr}(r)}F^2(\theta,\varphi). \tag{7.4.7}$$

Коефіцієнт спрямованої дії показує, у скільки разів густина потоку енергії реальної антени в заданому напрямку відрізняється від такої величини для ізотропного випромінювача тієї самої потужності.

З формули (7.4.7) видно, що максимальне значення к.с.д. –

$$K_{dir\,\max}=\frac{\Pi_{r\,\max}(r)}{\Pi_{avr}(r)}. \tag{7.4.8}$$

Для диполя Герца, як випливає з формул (7.4.6), (7.2.27) та (7.2.29),

$$K_{dir}(\theta,\varphi)=(3/2)\sin^2\theta. \tag{7.4.6 а}$$

Для сучасних антен ця величина може досягати 10^6 і більше.

Запишемо амплітуду електричного поля випромінювання в далекій зоні через $K_{dir\,\max}$. Для цього скористаємося формулою для вектора Пойнтінга (в системі СІ):

$$\Pi_r(r,\theta,\varphi)=\frac{1}{2}E_m(r,\theta,\varphi)H_m^*(r,\theta,\varphi)=\frac{E_m^2}{2Z_{0v}(r,\theta,\varphi)} \tag{7.4.9}$$

(враховано, що $E_m/H_m=Z_{0v}$, де, як і раніше, Z_{0v} – хвильовий опір вакууму; див. формулу (7.2.26)). З іншого боку, з формул (7.4.3), (7.4.6) та (7.4.8) випливає, що

$$\Pi_r(r,\theta,\varphi)=\Pi_{r\,\max}(r)F^2(\theta,\varphi)=\frac{\Pi_{r\,\max}(r)}{\Pi_{avr}(r)}F^2(\theta,\varphi)=K_{dir\,\max}F^2(\theta,\varphi). \tag{7.4.10}$$

Прирівнявши праві частини (7.4.9) і (7.4.10), маємо:

$$E(r,\theta,\varphi)=\frac{1}{r}\sqrt{\frac{Z_{0v}}{2\pi}P_{rad}K_{dir\,\max}}F(\theta,\varphi), \tag{7.4.11}$$

Коефіцієнт корисної дії антени в загальному випадку визначається так само, як це було зроблено при аналізі півхвильового вібратора (див. формулу (7.3.4)):

$$\eta=\frac{P_{rad}}{P}. \tag{7.4.12}$$

Тут P_{rad} – середня потужність, випромінювана антеною, P – повна потужність, що підводиться до антени від фідера, $P_{loss}=P-P_{rad}$ – потужність втрат в антені

Для так званих простих (дротяних) випромінювачів (див. нижче п. 7.5.2), для яких

можна говорити про струм антени з амплітудою i_m, вводяться поняття опору випромінювання антени $R_{rad} = 2P_{rad}/i_m^2$ та опору втрат $R_{loss} = 2P_{loss}/i_m^2$. Тоді можна записати, що $\eta = \dfrac{R_{rad}}{R_{rad} + R_{loss}}$.

К.к.д. довгохвильових антен звичайно малий, оскільки для них $R_{rad} \ll R_{loss}$ (ми бачили це на прикладі півхвильового вібратора, див. вище п.7.3.2), але зі зменшенням довжини хвилі він зростає і в діапазоні ультракоротких хвиль може досягати 0.9.

Коефіцієнт підсилення антени – де добуток її к.к.д. на к.с.д.:

$$G(\theta,\varphi) = \eta K_{dir}(\theta,\varphi). \qquad (7.4.13)$$

Зазвичай під коефіцієнтом підсилення розуміють його максимальне значення.

Коефіцієнтом підсилення користуються, якщо потрібно розрахувати поле випромінювання за відомою вхідною потужністю антени та її к.к.д.

7.4.4. Інші параметри передавальних антен

Якщо напруженість електричного поля на поверхні металу перевищує деяке порогове значення, то у відповідній області починається автоелектронна емісія – під дією електричного поля електрони починають вириватися з поверхні металу. Якщо це відбувається в повітрі, то запалюється так званий коронний розряд – під дією потоку електронів газ іонізується. Говорять, що відбувається електричний пробій. Такий пробій може відбутися в антені, якщо до неї підвести сигнал надмірної потужності.

Припустима потужність випромінювання – це найбільша потужність, випромінювана антеною, за якої електричний пробій ще не відбувається.

Поняттям вхідного опору (імпедансу) антени ми користувалися, коли розглядали півхвильовий вібратор. Для дротяних антен *вхідний опір* Z_{in} – це відношення комплексних амплітуд вхідної напруги та вхідного струму:

$$Z_{in} = \frac{u_m}{i_m}. \qquad (7.4.14).$$

Отже, в загальному випадку це комплексна величина: $Z_{in} = R_{in} + iX_{in}$.

Якщо розміри антени більші за довжину хвилі або сумірні з нею, то співвідношення між струмами й напругами в різних точках провідників, із яких вона складається, будуть різними. Тому вхідний опір такої антени залежить від точки увімкнення джерела сигналу (фідера).

Якщо антена узгоджена з фідером, то вся підведена потужність витрачається в антені, а її вхідний опір є суто активним, тобто

$$P = \frac{1}{2} R_{in} i_m^2. \qquad (7.4.15)$$

З іншого боку, з формул (7.3.2)-(7.3.4) випливає, шо

$$P = \frac{1}{2} i_m^2 \left(R_{rad} + R_{loss}\right), \qquad (7.4.16)$$

Таким чином, при узгодженні антени з фідером

$$R_{in} = R_{rad} + R_{loss}. \qquad (7.4.17)$$

Як уже згадувалося, для антен, розміри яких сумірні з довжиною хвилі, вхідний опір антени залежить від точки увімкнення джерела сигналу (фідера). В силу формули (7.4.17) те саме можна сказати й про опір випромінювання, що робить його визначення неоднозначним.

Для однозначного визначення опору випромінювання його прийнято визначати в пучності струму в антені незалежно від точки увімкнення фідера. Отже, можна сказати, що опір випромінювання – це такий ефективний опір, який при увімкненні в пучності струму в антені поглинає потужність, що дорівнює потужності випромінювання.

Як випливає з формули (7.4.11),

$$E_{\max}(r) = \frac{1}{r}\sqrt{\frac{Z_{0\upsilon}}{2\pi}P_{rad}K_{dir\ \max}} = \frac{i_m}{r}\sqrt{\frac{Z_{0\upsilon}}{4\pi}R_{rad}K_{dir\ \max}} \qquad (7.4.18)$$

Для диполя Герца, як випливає з формули (7.2.31) та (7.4.6 а)

$$E_{\max}(r) = \frac{Z_{0\upsilon}i_m l}{2r\lambda}. \qquad (7.4.19)$$

Застосування формули (7.4.19) до антени з розмірами порядку довжини хвилі дозволяє (з урахуванням (7.4.18)) визначити *діючу довжину* антени (використовують також термін «ефективна висота» для вертикальних антен) з умови

$$l_{eff} = \lambda\sqrt{\frac{R_{rad}K_{dir\ \max}}{\pi Z_{0\upsilon}}}. \qquad (7.4.20)$$

Використовуючи поняття діючої довжини, ми заміняємо реальну антену з нерівномірним розподілом струму диполем Герца з рівномірним розподілом струму.

Поняття діючої довжини застосовне лише для дротяних антен не дуже великих розмірів, у яких розподіл струму по довжині не змінює знаку.

Робочим діапазоном частот, або смугою пропускання антени називають смугу частот, у якій параметри та характеристики антени змінюються в допустимих межах. Річ у тім, що при зміні частоти сигналу змінюється її вхідний опір, умови узгодження, к.с.д., поляризація, тощо. Частотною характеристикою антени в різних джерелах називають частотні залежності вхідного опору, коефіцієнта біжучої хвилі у фідері, к.с.д., амплітуди вхідного струму (при фіксованій амплітуді вхідної напруги). Відповідно вводять смугу пропускання за струмом, к.с.д. та іншими параметрами.

Коефіцієнтом перекриття діапазону називають величину $\kappa = f_{\max}/f_{\min}$, де $f_{\max}$ та $f_{\min}$ – верхня та нижня межі смуги пропускання. За означенням $\kappa > 1$. Для вузькосмугових антен $\kappa < 1.1$, для надширокосмугових κ може досягати кількох десятків.

Говорячи про частотні властивості антен, слід також зазначити, що узгодити фідер з антеною можна тільки на фіксованій частоті (або у вузькій смузі частот). Ця обставина додатково звужує смугу пропускання антени.

Як відомо, для коливних контурів добротність визначається як відношення накопиченої (реактивної) енергії до енергії втрат у ньому за період власних коливань. Її можна знайти як відношення резонансної частоти до ширини амплітудно-частотної характеристики на рівні $1/\sqrt{2}$ за амплітудою або $1/2$ за потужністю коливань.

Добротність антени також визначається як відношення реактивної енергії, накопиченої антеною, до енергії, що витрачається на випромінювання та на активні втрати за період коливань.

Порівнюючи півхвильовий вібратор з диполем Герца, ми вказували, що більша ефективність першої з цих антен пояснюється використанням явища резонансу. З цієї причини розміри дротяних антен намагаються робити порядку довжини випромінюваних хвиль (саме ця обставина диктує необхідність використання модульованих радіосигналів, див. Вступ).

Якщо ж розміри антени зменшують настільки, що вони стають значно меншими за довжину випромінюваних нею хвиль, то її опір випромінювання так само різко зменшується (див. формулу (7.2.31)). Тому для випромінювання помітної потужності в антені повинні протікати великі струми. Це призводить до відповідного зростання електричного та магнітного полів у близькій зоні випромінювання. Саме поле в близькій зоні визначає реактивну енергію, накопичену антеною. Чим менше відношення l/λ (l – розмір антени), тим більшою виявляється її добротність, і, відповідно, меншою – її смуга пропускання (при малих l справедливе співвідношення $Q \sim l^{-3}$). Таким чином, сконструювати широкосмугову антену (тобто антену з невеликою добротністю) й водночас малих розмірів неможливо.

7.4.5. Теорема взаємності для антен

Розглянувши характеристики й параметри передавальних антен, перейдемо тепер до приймальних антен. Але спочатку більш детально розглянемо теорему взаємності, про яку вже

згадувалося раніше і яка відіграє важливу роль у теорії антен.

В теорії електричних кіл доводиться твердження, відоме як найпростішій варіант теореми взаємності.

Розглянемо лінійний чотириполюсник. Нехай на його вході увімкнене джерело напруги з е.р.с. e_1 та внутрішнім опором Z_{l1}, а на виході – навантаження Z_{l2} (рис. 7.4.7 а), через яке протікає струм i_{l2}. Тепер увімкнемо на виході е.р.с. e_2, а е.р.с. на вході приберемо (рис. 7.4.7 б). Нехай при цьому струм вхідного кола складає величину i_{l1}.

Теорема взаємності стверджує, що

$$\frac{e_1}{i_{l2}} = \frac{e_2}{i_{l1}}. \qquad (7.4.21)$$

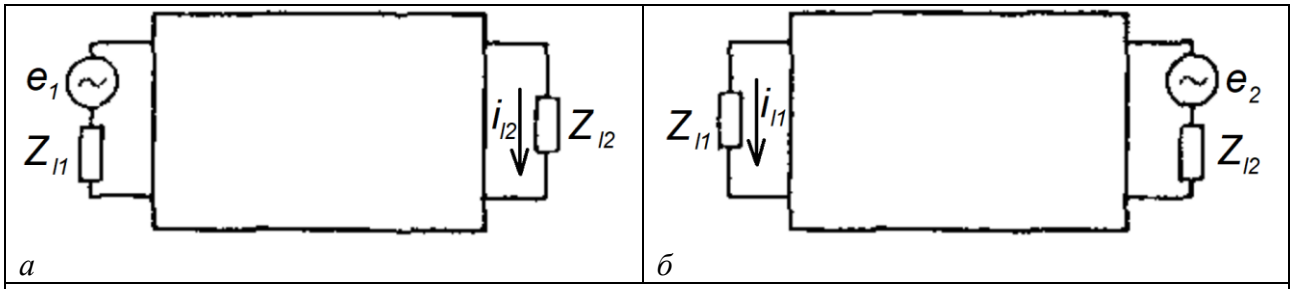


Рис. 7.4.7. До пояснення теореми взаємності

Тепер ускладнимо аналізовану модель (рис. 7.4.8). Нехай ми маємо дві антени 1 і 2, розташовані в ізотропному середовищі на деякій віддалі r одна від одної (у далекій зоні випромінювання). Для сигналів помірної амплітуди як антени, так і середовище, в якому поширюється електромагнітна хвиля, можна вважати лінійними. Тому систему, що включає антени й простір між ними, можна розглядати як лінійний чотириполюсник, зводячи цю модель до попередньої.

Увімкнемо на вхід антени 1 джерело гармонічної е.р.с. $e_1 = e_{m1}\exp(i\omega t)$ з внутрішнім опором Z_1, а до виходу антени 2 – опір навантаження Z_{l2}, тобто антена 1 є передавальною, а антена 2 – приймальною (рис. 7.4.8 а). При цьому в передавальній антені наводиться струм i_1, такий, що

$$e_1 = i_1\left(Z_1 + Z_{in1}\right), \qquad (7.4.22)$$

де Z_{in1} – вхідний опір антени 1 у передавальному режимі. В приймальній антені виникає струм i_{l2}, що протікає через опір Z_{l2}.

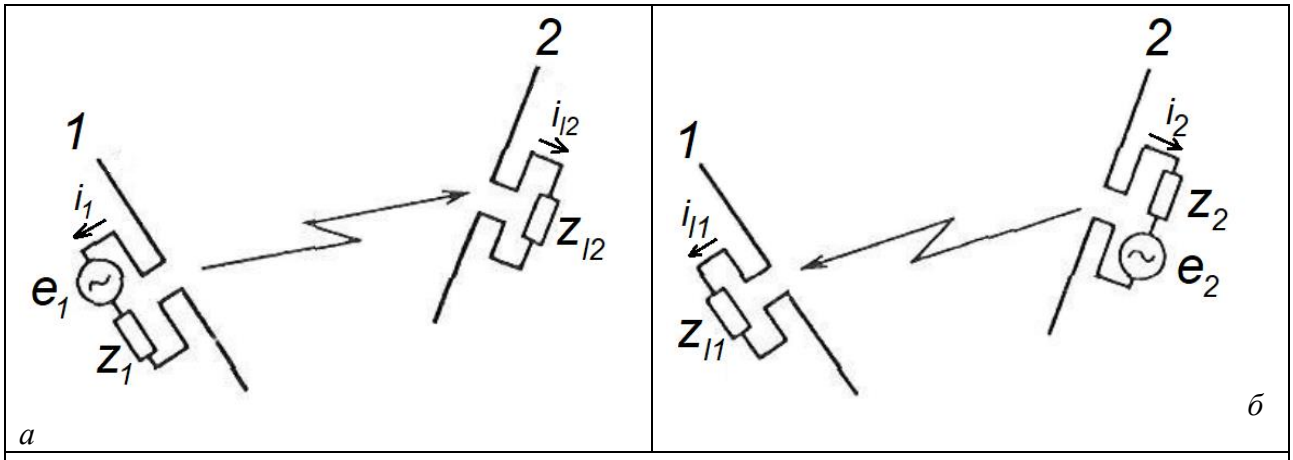


Рис. 7.4.8. До пояснення теореми взаємності

Поле, що створюється антеною 1 в області розташування антени 2, може бути відповідно до формули (7.4.4) записане у формі

$$\vec{E}_2(\vec{r},t)=\frac{i_1}{r}\exp(-ik_0r)\left\{\vec{e}_{\theta1}A_{\theta1}f_{\theta1}(\theta_1,\varphi_1)\exp\left[i\psi_{\theta1}(\theta_1,\varphi_1)\right]+\vec{e}_{\varphi1}A_{\varphi1}f_{\varphi1}(\theta_1,\varphi_1)\exp\left[i\psi_{\varphi1}(\theta_1,\varphi_1)\right]\right\},$$

(7.4.23)

(хвиля вважається сферичною, поперечною, плоскополяризованою, електричне поле лежить у площині $r=const$ і зорієнтоване довільним чином). Формула (7.4.23) записана в сферичній системі координат, пов'язаній з антеною 1, кути θ_1,φ_1 характеризують напрямок на антену 2. Функції $f_{\theta1}(\theta_1,\varphi_1)$ та $f_{\varphi1}(\theta_1,\varphi_1)$ характеризують діаграму спрямованості антени 1 на передавання сигналу.

Виключивши струм i_1 з формули (7.4.22) за допомогою (7.4.23), отримаємо:

$$e_1=\frac{\vec{E}_2(\vec{r},t)r(Z_1+Z_{in1})}{\left\{\vec{e}_{\theta1}A_{\theta1}f_{\theta1}(\theta_1,\varphi_1)\exp\left[i\psi_{\theta1}(\theta_1,\varphi_1)\right]+\vec{e}_{\varphi1}A_{\varphi1}f_{\varphi1}(\theta_1,\varphi_1)\exp\left[i\psi_{\varphi1}(\theta_1,\varphi_1)\right]\right\}\exp(-ik_0r)}$$

(7.4.24)

Тепер зробимо антену 2 передавальною, а антену 1 – приймальною, не змінюючи їхнього взаємного розташування (рис. 7.4.8 б). Тоді в антені 2 протікатиме струм i_2, такий, що

$$e_2=i_2(Z_2+Z_{in2}). \qquad (7.4.25)$$

Для цього випадку можна записати співвідношення, аналогічне до (7.4.24):

$$e_2=\frac{\vec{E}_1(\vec{r},t)r(Z_2+Z_{in2})}{\left\{\vec{e}_{\theta2}A_{\theta2}f_{\theta2}(\theta_2,\varphi_2)\exp\left[i\psi_{\theta2}(\theta_2,\varphi_2)\right]+\vec{e}_{\varphi2}A_{\varphi2}f_{\varphi2}(\theta_2,\varphi_2)\exp\left[i\psi_{\varphi2}(\theta_2,\varphi_2)\right]\right\}\exp(-ik_0r)}$$

(7.4.26)

Конкретизуємо модель, прийнявши, відповідно до теореми взаємності (див. рис. 7.4.7), що

$$Z_{l1}=Z_1, \qquad Z_{l2}=Z_2. \qquad (7.4.27)$$

Іншими словами, на рис. 4.7 а-б імпеданси на вході/виході антени 1 та антени 2 однакові (пор. з рис. 4.6 а-б).

Тоді сукупність антени 1, ефіру, через який поширюється сигнал, та антени 2 можна, як уже вказувалося, уподібнити чотириполюснику, що фігурує в теоремі взаємності. Отже, відповідно до формули (7.4.21) з урахуванням (7.4.24) та (7.4.26)-(7.4.27) можна записати, що

$$\frac{e_1}{i_{l2}}=\frac{(Z_1+Z_{in1})\vec{E}_2(\vec{r},t)r}{i_{l2}\left\{\vec{e}_{\theta1}A_{\theta1}f_{\theta1}(\theta_1,\varphi_1)\exp\left[i\psi_{\theta1}(\theta_1,\varphi_1)\right]+\vec{e}_{\varphi1}A_{\varphi1}f_{\varphi1}(\theta_1,\varphi_1)\exp\left[i\psi_{\varphi1}(\theta_1,\varphi_1)\right]\right\}\exp(-ik_0r)}=$$

$$=\frac{e_2}{i_{l1}}=\frac{(Z_2+Z_{in2})\vec{E}_1(\vec{r},t)r}{i_{l1}\left\{\vec{e}_{\theta2}A_{\theta2}f_{\theta2}(\theta_2,\varphi_2)\exp\left[i\psi_{\theta2}(\theta_2,\varphi_2)\right]+\vec{e}_{\varphi2}A_{\varphi2}f_{\varphi2}(\theta_2,\varphi_2)\exp\left[i\psi_{\varphi2}(\theta_2,\varphi_2)\right]\right\}\exp(-ik_0r)},$$

(7.4.28)

або

$$\frac{(Z_1+Z_{in1})i_{l1}}{\vec{E}_1(\vec{r},t)\left\{\vec{e}_{\theta1}A_{\theta1}f_{\theta1}(\theta_1,\varphi_1)\exp\left[i\psi_{\theta1}(\theta_1,\varphi_1)\right]+\vec{e}_{\varphi1}A_{\varphi1}f_{\varphi1}(\theta_1,\varphi_1)\exp\left[i\psi_{\varphi1}(\theta_1,\varphi_1)\right]\right\}}=$$

$$=\frac{(Z_2+Z_{in2})i_{l2}}{\vec{E}_2(\vec{r},t)\left\{\vec{e}_{\theta2}A_{\theta2}f_{\theta2}(\theta_2,\varphi_2)\exp\left[i\psi_{\theta2}(\theta_2,\varphi_2)\right]+\vec{e}_{\varphi2}A_{\varphi2}f_{\varphi2}(\theta_2,\varphi_2)\exp\left[i\psi_{\varphi2}(\theta_2,\varphi_2)\right]\right\}}. \qquad (7.4.29)$$

Ліва частина рівняння (7.4.29) визначається лише антеною 1, права – лише антеною 2. Оскільки властивості антен не залежать одна від одної, обидві частини (7.4.29) можна вважати сталою величиною (позначимо її C). Тоді з формули (7.4.29) випливає співвідношення, справедливе для будь-якої приймальної антени:

$$\frac{(Z_l+Z_{in})i_l}{E_\theta A_\theta f_\theta(\theta,\varphi)\exp\left[i\psi_\theta(\theta,\varphi)\right]+E_\varphi A_\varphi f_\varphi(\theta,\varphi)\exp\left[i\psi_\varphi(\theta,\varphi)\right]}=C, \qquad (7.4.30).$$

(введені позначення $E_{\theta,\varphi}=\left(\vec{E}(\vec{r},t)\cdot\vec{e}_{\theta,\varphi}\right)$, замість $Z_{1,2}$, $Z_{in1,2}$ та $i_{l1,2}$ пишемо відповідно Z_l,

Z_{in} та i_l).

Чисельник у формулі (7.4.30) має розмірність напруги. Тому, якщо ввести поняття еквівалентної е.р.с. приймальної антени

$$e_{eq} = C\left\{E_\theta A_\theta f_\theta(\theta,\varphi)\exp\left[i\psi_\theta(\theta,\varphi)\right] + \vec{e}_\varphi A_\varphi f_\varphi(\theta,\varphi)\exp\left[i\psi_\varphi(\theta,\varphi)\right]\right\}, \quad (7.4.31)$$

то можна записати, що

$$e_{eq} = (Z_l + Z_{in})i_l. \quad (7.4.32)$$

Таким чином, будь-яка приймальна антена може розглядатись як еквівалентний генератор з е.р.с. e_{eq} (7.4.31), навантажений опором Z_l. Внутрішній опір цього еквівалентного генератора дорівнює вхідному опору Z_{in} цього самого генератора в передавальному режимі (рис. 7.4.9).

Для знаходження константи C розглянемо вібратор Герца, який опромінюється джерелом вертикально поляризованих хвиль, напрямок на яке утворює кут θ з віссю вібратора (рис. 7.4.10). Це джерело створює плоскополяризовану хвилю з електричним полем $\vec{E} = \vec{e}_\theta E_\theta$, яка наводить в антені напругу $e_{eq} = E_\theta l \sin\theta$ (l – довжина антени).

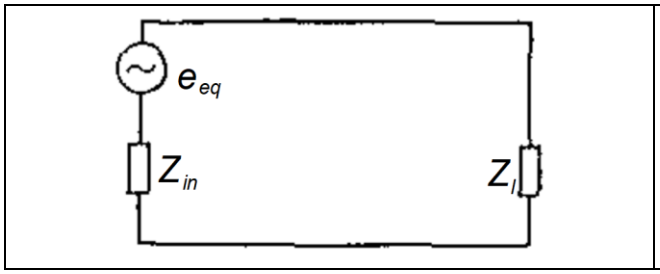


Рис. 7.4.9. Приймальна антена як еквівалентний генератор

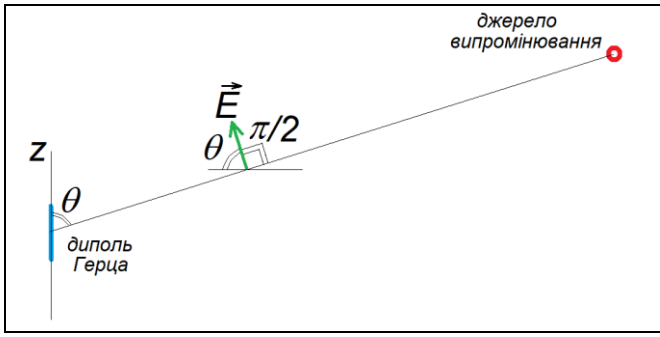


Рис. 7.4.10. Вібратор Герца, який опромінюється джерелом випромінювання

З іншого боку, для аналізованого випадку з формули (7.4.31) випливає, що

$$e_{eq} = CE_\theta A_\theta f_\theta(\theta,\varphi)\exp\left[i\psi_\theta(\theta,\varphi)\right] = i\frac{Z_{0\upsilon}l}{2\lambda}CE_\theta \sin\theta = iC\frac{60\pi l}{\lambda}E_\theta \sin\theta, \quad (7.4.33)$$

де, як і раніше, $Z_{0\upsilon}$ – хвильовий опір вакууму Тут враховано також, що для диполя Герца в системі СІ

$$A_\theta f_\theta(\theta,\varphi)\exp\left[i\psi_\theta(\theta,\varphi)\right] = i\frac{Z_{0\upsilon}l}{2\lambda}\sin\theta = i60\pi\frac{l}{\lambda}\sin\theta. \quad (7.4.34)$$

Порівнюючи дві записані вище формули для e_{eq}, бачимо, що

$$C = i\lambda/60\pi. \quad (7.4.35)$$

Отже, остаточно маємо:

$$e_{eq} = \frac{i\lambda}{60\pi}\left\{E_\theta A_\theta f_\theta(\theta,\varphi)\exp\left[i\psi_\theta(\theta,\varphi)\right] + E_\varphi A_\varphi f_\varphi(\theta,\varphi)\exp\left[i\psi_\varphi(\theta,\varphi)\right]\right\}. \quad (7.4.36)$$

З формул (7.4.31) та (7.4.36) випливає, що діаграми спрямованості антени на передавання й приймання сигналу пов'язані між собою. Якщо антена випромінює плоскополяризовані хвилі, то, прирівнюючи по черзі до нуля компоненти E_θ та E_φ, отримуємо, що діаграма приймальної антени для кожної з компонент тотожна відповідній діаграмі передавальної

антени. При довільній поляризації хвиль зв'язок виходить складнішим.

Конкретизуємо ці міркування для дротяної антени помірних розмірів.

Для диполя Герца е.р.с. e_{eq} може бути, як уже вказувалося, представлене у вигляді $e_{eq} = E_\theta l \sin\theta$, де E_θ – напруженість електромагнітного поля хвилі, створеної передавальною антеною, в точці розташування приймальної антени, θ – кут між напрямком цього поля і площиною, перпендикулярною до осі диполя (таким самим буде кут між напрямком на випромінювач та віссю приймальної антени; при $\theta = 0$ електричне поле перпендикулярне до антени, відповідно е.р.с. у ній не виникає), l – довжина диполя (рис. 7.4.9).

Тепер узагальнимо цю формулу на дротяну антену помірних розмірів, замінивши l на діючу довжину антени l_{eff} (7.4.20), а характеристику спрямованості диполя Герца $\sin\theta$ – на характеристику спрямованості довільної антени $f(\theta,\varphi)$:

$$e_{eq} = \lambda E_\theta \sqrt{\frac{R_r K_{dir\,\max}}{\pi Z_{0v}}} f(\theta,\varphi), \tag{7.4.37}$$

де, як і раніше, Z_{0v} – хвильовий опір вакууму.

Нарешті, підставивши (7.4.37) до (7.4.32), остаточно отримаємо:

$$i_l = \frac{\lambda E_\theta}{Z_{in} + Z_l} \sqrt{\frac{R_r K_{dir\,\max}}{\pi Z_{0v}}} f(\theta,\varphi). \tag{7.4.38}$$

Звідси випливає, що основні характеристики та параметри антени – характеристика спрямованості, к.к.д., вхідний опір, опір випромінювання, діюча довжина, к.с.д. та інші – однакові при роботі на передавання та приймання сигналу.

7.4.6. Характеристики та параметри приймальних антен

Визначення характеристик і параметрів приймальної антени дещо відрізняються від аналогічних визначень для передавальної антени. Тому нижче наведемо деякі з них.

Характеристика спрямованості (за полем) – це залежність величини е.р.с., яка наводиться в приймальній антені, від напрямку надходження плоскої електромагнітної хвилі, напруженість якої в точці приймання вважається незмінною.

Аналогічно визначається характеристика спрямованості за потужністю.

Коефіцієнт спрямованої дії – це відношення потужності, прийнятої антеною, до потужності, прийнятої ізотропною (неспрямованою) антеною, розташованою в тій самій точці. Використання приймальних антен спрямованої дії дозволяє зменшити завади від сторонніх джерел (наприклад, від інших передавачів) – за умови, що корисний сигнал і завади надходять з різних напрямків. Такі антени використовуються також для пеленгування (визначення напрямку на передавач), у радіолокації (зокрема, в пасивній), для приймання слабких сигналів (загоризонтна радіолокація, радіоастрономія) тощо.

Коефіцієнт корисної дії – це відношення потужності, яка надходить до приймача, до повної потужності, прийнятої антеною.

Вся енергія, прийнята антеною, витрачається на активному опорі самої антени та активному опорі навантаження. В свою чергу, активний вхідний опір антени складається з опору втрат та опору випромінювання. Наявність опору випромінювання означає, що під час приймання сигналу антена водночас випромінює в ефір частину прийнятої нею енергії (так зване вторинне випромінювання).

Потужність P_{rec} (індекс – від англ. receiver, приймач), що виділяється на навантаженні приймальної антени (корисна потужність) може бути подана у формі

$$P_{rec} = \frac{1}{2} i_m^2 R_{rec}. \tag{7.4.39}$$

Ця величина досягає максимального значення, коли виконана умова узгодження, а хвиля від передавача надходить уздовж напрямку максимального приймання, тобто $f(\theta,\varphi) = 1$. Цю максимальну потужність можна визначити зі співвідношення

$$P = \frac{e_{\max}^2(t)}{4R_{in}} = \frac{\lambda^2 E^2}{4\pi R_{in}} \frac{R_{rad}}{Z_{0\upsilon}} K_{dir\,\max} = \frac{\lambda^2 E^2}{4\pi Z_{0\upsilon}} \frac{R_{rad}}{R_{in}} K_{dir\,\max} = \frac{\lambda^2 E^2 G_{\max}}{4\pi Z_{0\upsilon}}, \qquad (7.4.40)$$

де, як і раніше, $G_{\max} = \eta K_{dir\,\max}$ (див. формулу (7.4.8)), $\eta = R_{rad}/R_{in}$.

Поняття діючої довжини, як уже вказувалося, застосовне лише до лінійних антен з розмірами, які не перевищують $\lambda/2$. Більш універсальною характеристикою здатності антени приймати енергію електромагнітної хвилі, яка на неї падає, є ефективна (діюча) площа приймальної антени. *Ефективна площа приймальної антени* – це відношення прийнятої антеною P потужності до густини потоку енергії електромагнітної хвилі Π в точці її розташування: $S_{eff} = P/\Pi$.

Для лінійних антен помірного розміру

$$e_{\max} = El_{eff}, \qquad P = e_{\max}^2/4R_{in} = \left(El_{eff}\right)^2/4R_{in}, \qquad \Pi = E^2/Z_{0\upsilon},$$

Тому з урахуванням формули (7.4.20)

$$S_{eff} = \frac{Z_{0\upsilon} l_{eff}^2}{4R_{in}} = \frac{\lambda^2}{4\pi} \frac{R_{rad}}{R_{in}} K_{dir\,\max} = \frac{\lambda^2}{4\pi} \eta K_{dir\,\max} = \frac{\lambda^2}{4\pi} G_{\max}. \qquad (7.4.41)$$

Користуючись останньою формулою, можна записати максимальний к.с.д. та коефіцієнт підсилення антени через її ефективну площу:

$$K_{dir\,\max} = \frac{4\pi}{\eta\lambda^2} S_{eff}, \qquad G_{\max} = \frac{4\pi}{\lambda^2} S_{eff}. \qquad (7.4.38)$$

Контрольні питання до підрозділу 7.4

1. Що таке характеристика спрямованості антени? Чим відрізняються характеристики спрямованості за полем і за потужністю?
2. Які різновиди характеристик спрямованості ви знаєте?
3. Що таке фазова та поляризаційна характеристики антени?
4. Що таке коефіцієнт спрямованої дії, коефіцієнт корисної дії та коефіцієнт підсилення антени?
5. Що таке робочий діапазон (смуга пропускання) антени та її добротність?
6. Чому основні характеристики та параметри антени однакові при роботі на передавання та приймання сигналу?
7. Дайте означення діаграми спрямованості та коефіцієнта спрямованої дії приймальної антени.

Задачі до підрозділу 7.4

7.4.1. Запишіть коефіцієнт спрямованості диполя Герца. Яким буде його максимальне значення?

7.4.2. Довжина диполя Герца, збуджуваного струмом з амплітудою 10 А – 1 м, частота випромінювання – 30 МГц. Знайти максимальну амплітуду електричного поля на віддалі 1 км.

7.4.3. Потужність випромінювання антени – 10 кВт, коефіцієнт спрямованої дії – 10^4. Знайти максимальну амплітуду електричного поля на віддалі 10 км.

7.5. Основні типи антен

Досі ми обговорювали лише найпростіші типи лінійних антен – електричний та магнітний диполь Герца і півхвильовий вібратор. У цьому розділі ми стисло розглянемо основні типи антен, які використовуються сьогодні в системах зв'язку, радіолокації тощо.

7.5.1. Принципи класифікації антен

В основу класифікації антен можуть бути покладені різні ознаки. Відповідно до цього можна отримати різні варіанти класифікації.

1. За характером поля, яке первісно створює антена, розрізняють електричні та магнітні антени. Як уже говорилося, зміна цього поля в часі породжує вихрове поле іншого характеру, і т.д. Прикладом електричної та магнітної антен можуть служити електричний та магнітний диполі Герца.
2. За характером поляризації випромінювання умовно розрізняють антени вертикальної, горизонтальної та циркулярної (кругової) поляризації. Вертикально розташований електричний диполь Герца є прикладом антени вертикальної поляризації, магнітний диполь Герца, вісь якого розташована вертикально – прикладом антени горизонтальної поляризації. Циркулярну поляризацію випромінювання забезпечують спіральні антени.
3. За частотними властивостями (точніше, за смугою пропускання) розрізняють резонансні (вузькосмугові) та нерезонансні (широкосмугові) антени. Прикладом перших може служити півхвильовий вібратор, прикладом других – вібратор Герца.
4. За довжиною випромінюваних хвиль можна умовно виділити антени довгих та середніх хвиль (від 100 м до 10 км й більше), антени коротких хвиль (1-100 м) та антени ультракоротких хвиль (від 1 мм до 1 м). В різних діапазонах радіохвиль використовують різні типи антен (див. нижче підрозділ 7.6).
5. За спрямованістю випромінювання розрізняють спрямовані та неспрямовані антени. Прикладом перших можуть служити антени з голкоподібною характеристикою спрямованості, прикладом других – диполь Герца та півхвильовий вібратор.
6. Залежно від відсутності або наявності вбудованого підсилювача сигналу розрізняють пасивні та активні антени. Активні антени звичайно застосовують для приймання сигналу. На відміну від пасивних антен, активні антени можна вважати лінійними системами лише для електромагнітних хвиль з порівняно невеликими амплітудами (це випливає з властивостей підсилювачів). Крім того, для таких антен не справджується теорема взаємності. Надалі ми будемо розглядати лише пасивні антени.
7. За способом збудження та підсилення сигналу антени умовно можна розділити на такі групи:
- прості випромінювачі;
- групові випромінювачі;
- випромінювальні структури;
- апертурні випромінювачі.

Прості випромінювачі можуть використовуватися безпосередньо, а можуть бути елементами, що входять до складу групових випромінювачів. Особливість випромінювальних структур та апертурних випромінювачів полягає в тому, що джерелом електромагнітних хвиль служить не відрізок провідника чи петля зі струмом, а деякий провідник складної форми або провідна поверхня (апертура).

Розглянемо останній варіант класифікації антен детальніше.

7.5.2. Прості випромінювачі

До простих випромінювачів належать ті, які не мають у своїй конструкції окремих випромінювальних елементів. Їх, у свою чергу, можна розбити на елементарні випромінювачі та прості антени.

До *елементарних випромінювачів* зараховують (електричний) диполь Герца, так званий диполь Фітцджеральда (магнітний диполь Герца), а також випромінювач Гюйгенса.

В оптиці відомий принцип Гюйгенса - Френеля, відповідно до якого кожна точка

фронту хвилі в поточний момент часу виступає джерелом вторинних хвиль. Випромінювач Гюйгенса (його можна назвати елементарним електромагнітним випромінювачем) і являє собою малий елемент фронту хвилі. Його можна уявити як прямокутник, з яким пов'язані змінні електричне та магнітне поля, паралельні до його сторін (рис. 7.5.1 а). Розміри цього прямокутника, як і розміри електричного та магнітного диполів Герца, мають бути значно меншими за довжину хвилі.

В реальності випромінювачем Гюйгенса називають сукупність провідника зі струмом та рамки зі струмом, які створюють взаємно перпендикулярні електричне та магнітне поля. Такі антени застосовують у пристроях для радіопеленгації. Діаграма спрямованості випромінювача Гюйгенса подана на рис. 7.5.1 б.

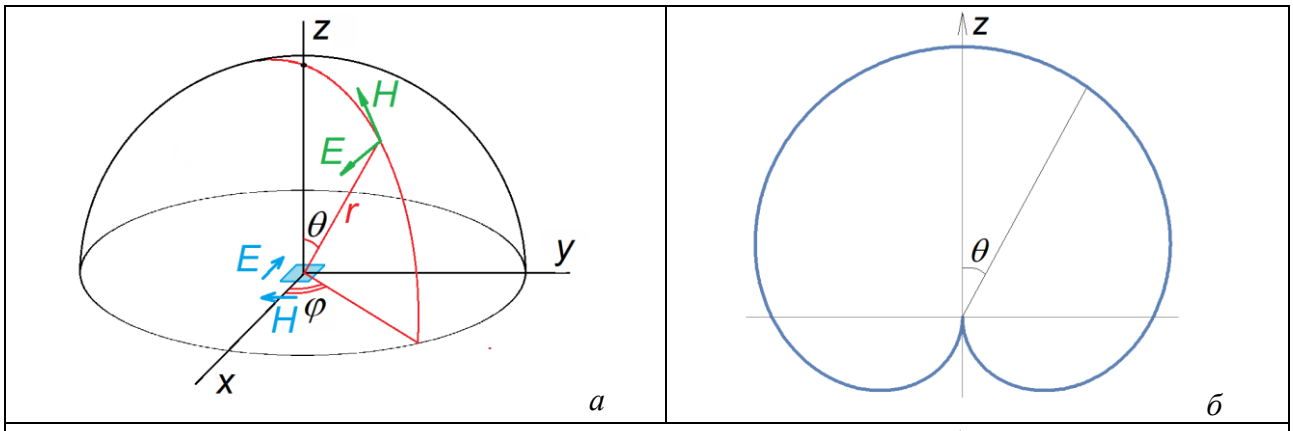


Рис. 7.5.1. Випромінювач Гюйгенса: а – схематичне зображення (блакитним показана область джерел та поля в ній, зеленим – поля в точці спостереження в далекій зоні); б – переріз діаграми спрямованості вертикальною площиною

До простих антен належать:
- лінійні;
- фігурні;
- рамкові;
- щілинні.

Оглянемо їх по черзі.

Лінійні антени – це відкриті (електричні) антени з прямолінійними елементами. Прикладом такої антени може бути півхвильовий вібратор.

До цього класу належить також монополь (уніполь, півдиполь, антена Марконі). Це найпростіший приклад антени з несиметричним живленням. Вона являє собою вертикальний провідник, нижній край якого сполучений із провідною поверхнею (рис. 7.5.2 а). Довжина провідника звичайно дорівнює $\lambda/4$, в окремих випадках – $\lambda/2$ або $5\lambda/8$ (див. нижче п. 7.6.1). До широкосмугових монополів зараховують, зокрема, конічні антени (рис. 7.5.2 б).

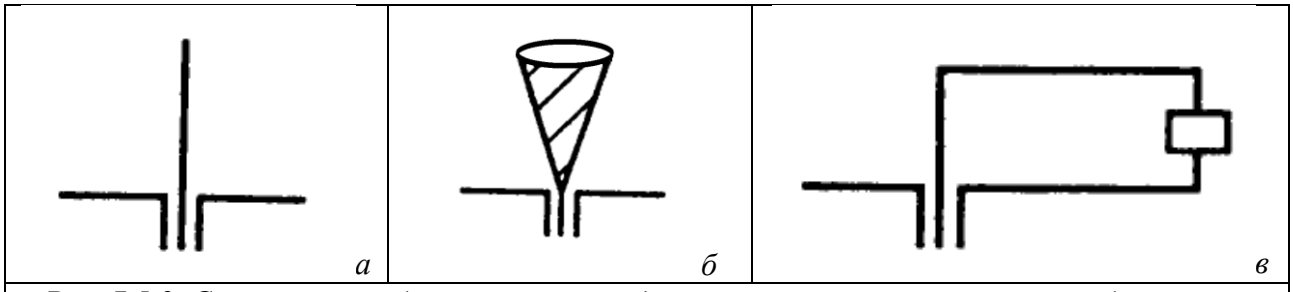


Рис. 7.5.2. Схематичні зображення різновидів лінійних антен: а – монополь, б – конічна антена, в – однопровідна антена біжучої хвилі (антена Бевереджа)

До монополів відносять і антени типу «довгий дріт», довжина яких перевищує довжину випромінюваних хвиль. Такі антени можуть бути симетричними чи несиметричними,

резонансними або аперіодичними, в них може встановлюватися біжуча або стояча хвиля. Прикладом таких антен може слугувати антена Бевереджа (рис. 7.5.2 в).

Фігурні антени подібні до лінійних. Це також відкриті (електричні) антени з криволінійними елементами. До групи фігурних антен належать вироджені диполі, вироджені монополі, вигнуті вібратори та вигнуті монополі. Основна особливість фігурних антен – це широка смуга пропускання.

Вироджені вібратори можуть додатково забезпечити формування кругової діаграми спрямованості (зокрема, кутовий вібратор – рис. 7.5.3 а).

Прикладом виродженого монополя може бути згадана раніше конічна антена (рис. 7.5.2 б) або подвійна конічна антена (рис. 7.5.3 б). Обидві вони характеризуються аксіальною симетрією.

Прикладом вигнутого вібратора може бути антена типу «помах крила» (рис. 7.5.3 в), прикладом вигнутого монополя – чашковий випромінювач (рис. 7.5.3 г).

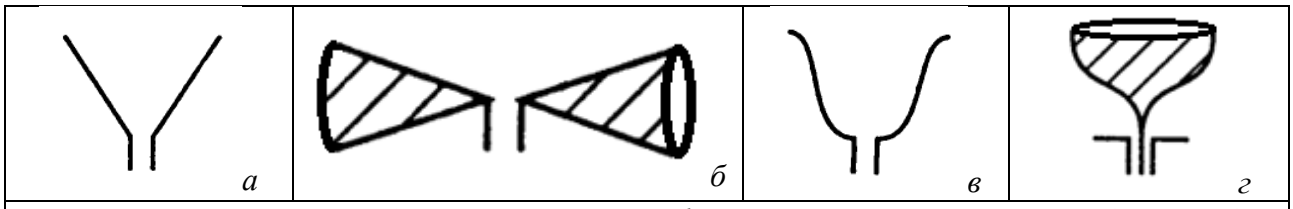


Рис. 7.5.3. Схематичні зображення різновидів фігурних антен: а – кутовий вібратор; б – подвійний конічний вібратор; в – антена типу «помах крила»; г – чашковий випромінювач

Рамкові антени – це замкнені антени з елементами у формі рамок. Очевидно, вони належать до класу магнітних антен. Рамкові антени розрізняють, у першу чергу, за їхніми розмірами. Мала рамка, або магнітна антена – це не що інше, як магнітний диполь Герца. Її типовий периметр складає $\lambda/10$. Великі рамки зазвичай мають периметр порядку λ. До цієї групи належать дискова і квадратна рамкові антени, а також магнітний вібратор (рис. 7.5.4, рис. 7.5.5).

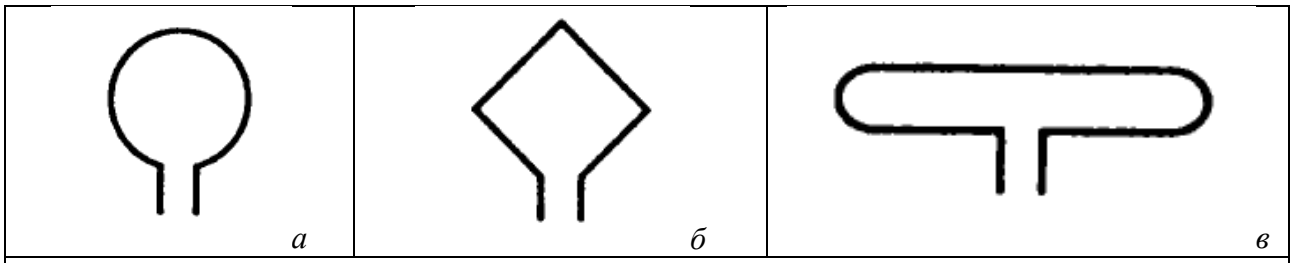


Рис. 7.5.4. Схематичні зображення різновидів рамкових антен: а – дискова рамкова антена; б – квадратна рамкова антена; в – магнітний вібратор

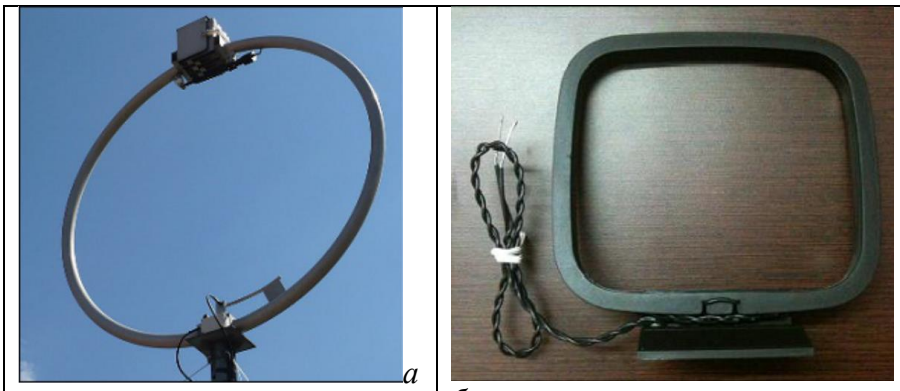
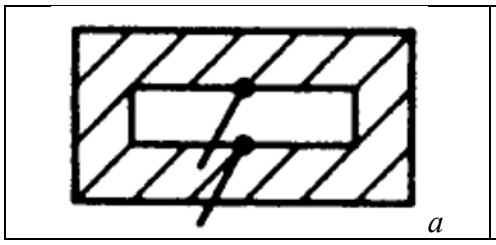


Рис. 7.5.5. Дискова (а) та прямокутна (б) рамкові антени: фото

Щілинні антени – це замкнені антени зі щілинними отворами на провідній поверхні.

Довжина щілин лежить у межах від $\lambda/2$ до λ. Зазвичай вони бувають лінійними (протяжні щілини на площині або на циліндрі) або хрестоподібними – зокрема, на дисковій щілинній антені (див. рис. 7.5.6 а-б, рис. 7.5.7).



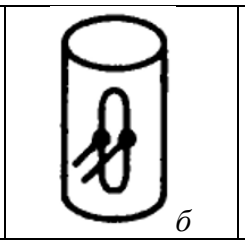


Рис. 7.5.6. Схематичні зображення щілинної (а) та трубчасто-щілинної (б) антен

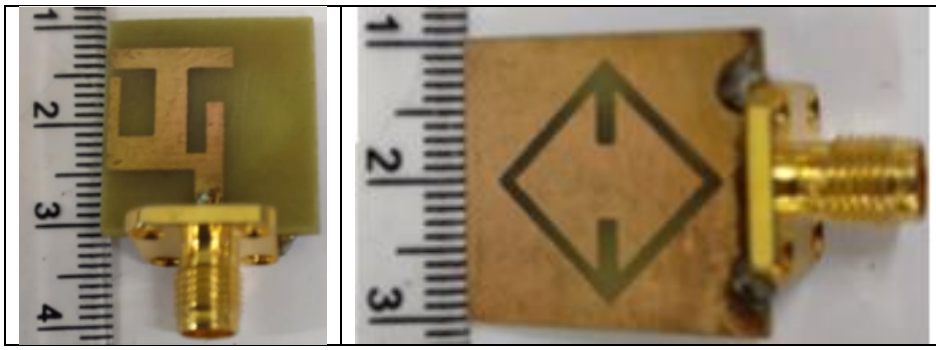


Рис. 7.5.7. Плоскі щілинні антени: фото

7.5.3. Групові випромінювачі

Групові випромінювачі складаються з певного набору простих випромінювачів, розглянутих у попередньому пункті. В цьому випадку властивості сумарного поля випромінювання визначають розташуванням простих випромінювачів та амплітудами й фазами їхнього живлення. Завдяки керуванню фазою сигналу з'являється можливість електронного сканування головним променем (головною пелюсткою діаграми спрямованості).

Проілюструємо цю можливість на простому прикладі. Нехай прості випромінювачі, розміри яких значно менші за довжину хвилі, розташовані у вершинах квадратів, які рівномірно покривають деяку площину (рис. 7.5.8 а). Кожен простий випромінювач є джерелом сферичних хвиль.

Якщо всі випромінювачі збуджуються синфазно, то радіуси фронтів хвиль, випромінюваних ними в деякий момент часу, є однаковими, і додавання хвиль від окремих випромінювачів формує плоский фронт, паралельний площині, в якій розміщені випромінювачі (рис. 7.5.8 б).

Нехай тепер у напрямку x між сусідніми випромінювачами існує постійний зсув фази φ, а в напрямку y сусідні випромінювачі й далі збуджуються синфазно. В такому випадку радіуси фронтів хвиль сусідніх випромінювачів у напрямку x відрізняються на величину

$$\Delta r = cT\frac{\varphi}{2\pi} = \frac{\lambda\varphi}{2\pi}, \qquad (7.5.1)$$

і вони додаватимуться в фазі на площині, нормаль до якої утворює кут

$$\alpha = \arcsin\frac{\Delta r}{a} = \arcsin\frac{\lambda\varphi}{2\pi a} \qquad (7.5.2)$$

з віссю z (тут a – віддаль між сусідніми випромінювачами).

Випромінювачі описаного типу відомі як фазовані антенні ґратки (див. нижче).

Груповий випромінювач може містити довільну кількість простих випромінювачів. Підбір кількості таких випромінювачів та їхнього розміщення дозволяє отримувати потрібне поле випромінювання.

Серед групових випромінювачів можна виділити такі різновиди, що відрізняються принципом розміщення простих випромінювачів:

- лінійні;

- площинні;
- просторові.

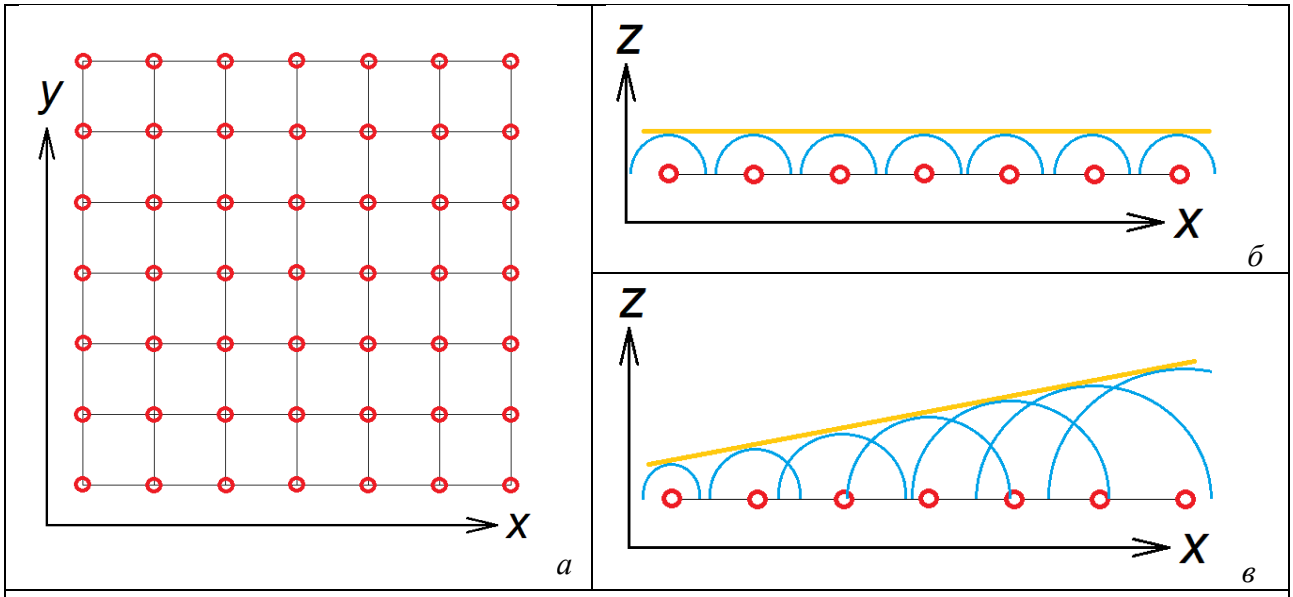


Рис. 7.5.8. Схема управління напрямком випромінювання: а – розташування простих випромінювачів (показані червоними кружечками) на площині; б – формування плоскої хвилі (її фронт показаний помаранчевим) при синфазному збудженні простих випромінювачів (фронти хвиль окремих випромінювачів показані блакитним); в – формування плоскої хвилі за наявності постійного зсуву фази між сусідніми випромінювачами в напрямку x

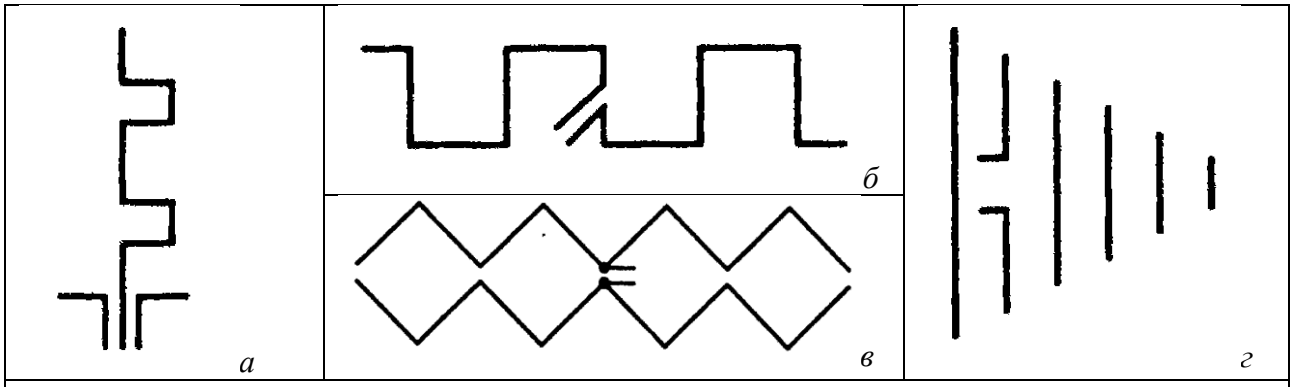


Рис. 7.5.9. Схематичні зображення лінійних (а-в) та площинної (г) групових антен: а – Франкліна; б – Брюса; в – Шире - Мезні; г – Уда - Ягі (хвильовий канал)

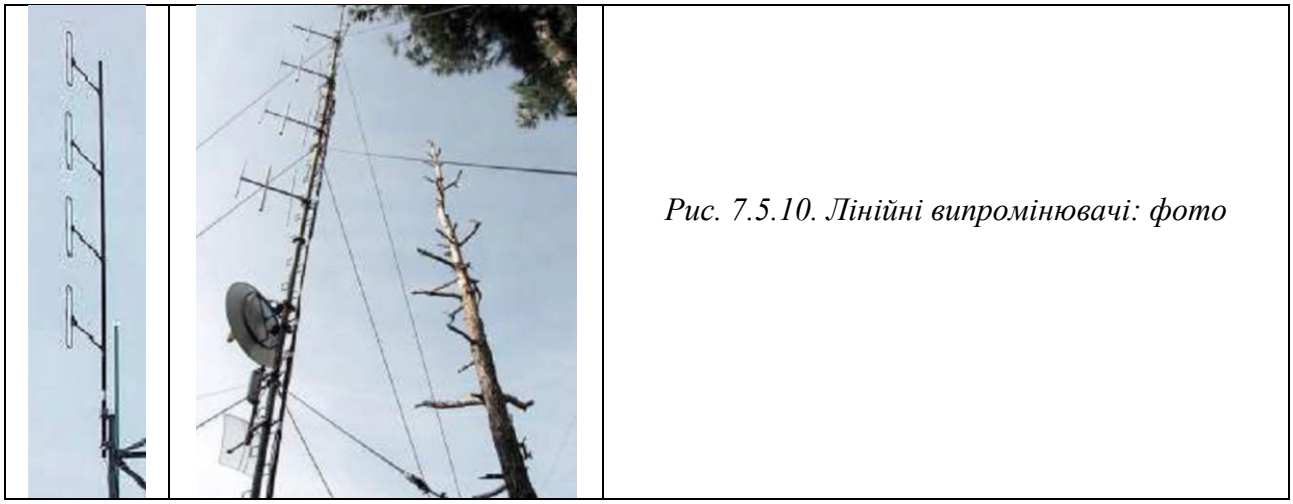


Рис. 7.5.10. Лінійні випромінювачі: фото

Лінійні випромінювачі складаються з простих випромінювачів, розміщених у лінію (нерідко ламану). Вони поділяються на паралельні та поздовжні.

Паралельний випромінювач – це лінійна послідовність простих випромінювачів (наприклад, вібраторів), які збуджуються синфазно і випромінюють по нормалі до площини, в якій вони лежать. Прикладами можуть бути антени Франкліна (Franklin), Брюса (Bruce) та Шире - Мезні (Chireix - Mesny), див. рис. 7.5.9 а-в та рис. 7.5.10.

Поздовжні антени також складаються з простих випромінювачів, розміщених у лінію, але збуджуваних у протифазі. Прикладом може служити відома антена Уда - Ягі (Uda – Yagi), або «хвильовий канал» (рис. 7.5.9 г, рис. 7.5.11; див. нижче п. 7.5.2).

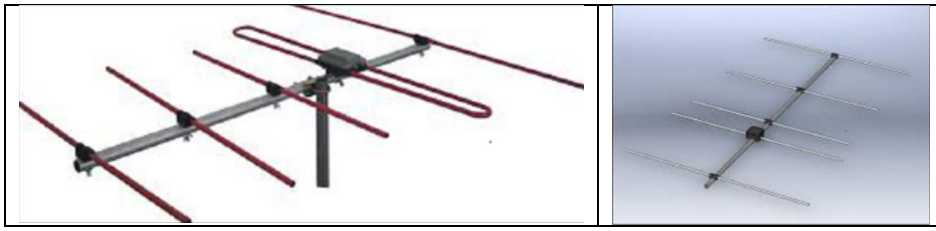


Рис. 7.5.11. Антена Уда - Ягі (хвильовий канал): фото

У *площинних випромінювачах* прості випромінювачі, що входять до їхнього складу, розміщуються в одній площині (двовимірний розподіл) – як правило, перед відбивачем. Група може складатися з кількох підгруп (2×2, 4×4, тощо). Цю групу випромінювачів умовно ділять на полотна, ґратки, площинні антени та підсумовувальні випромінювачі.

Антенні полотна з хвильових чи півхвильових вібраторів використовуються в діапазоні коротких хвиль (рис. 7.5.12а, 7.5.13).

Антенні ґратки застосовуються в діапазоні ультракоротких хвиль та в телебаченні. Вони також складаються з хвильових чи півхвильових вібраторів, інколи об’єднаних у матриці 4×4 або 8×8. Ідеалізована антенна ґратка подана на рис. 7.5.8 а.

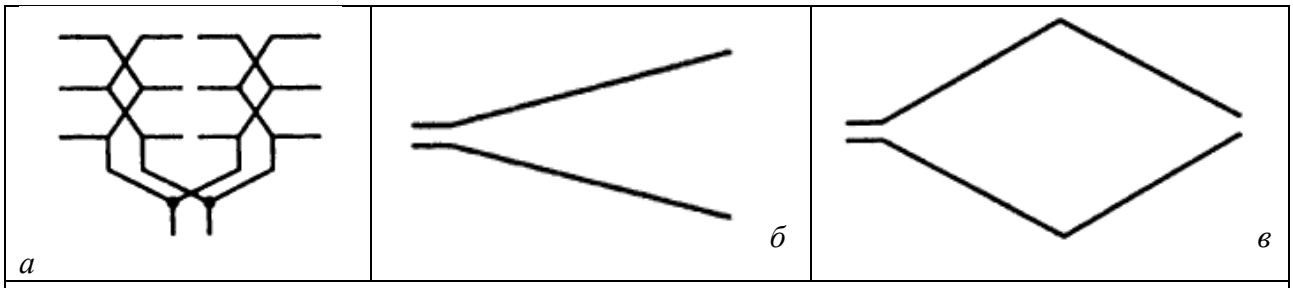


Рис. 7.5.12. Схематичні зображення площинних антен: а – полотно вібраторів; б – V-подібна антена; в – ромбічна антена

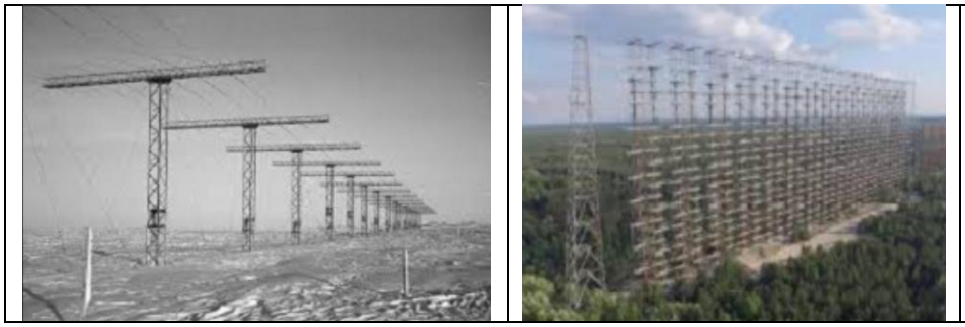


Рис. 7.5.13. Антенні полотна: фото

Площинні антени використовуються в мікрохвильовому діапазоні як супутникові. В ролі простих випромінювачів використовуються вібратори, випромінювачі Гюйгенса або щілинні випромінювачі.

Підсумовувальні випромінювачі складаються з вібраторів або антен типу «довгий дріт», розміщених на площині в різних конфігураціях. Загальне випромінювання є сумою внесків окремих випромінювачів. Прикладами таких випромінювачів можуть бути V-подібна та

ромбічна[7] (див. нижче п. 7.6.3) антени (рис. 7.5.12 б-в, 7.5.14).

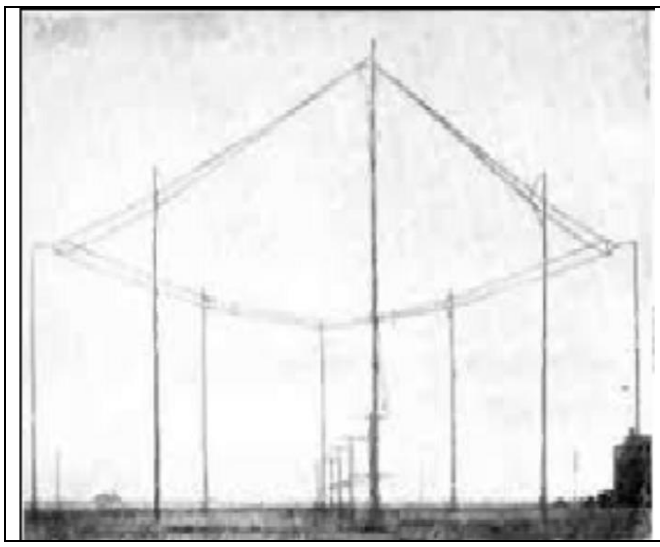


Рис. 7.5.14. Ромбічна антена: фото

Просторові випромінювачі характеризуються тривимірним розміщенням простих випромінювачів. Антени такого типу можуть забезпечувати як характеристику спрямованості, близьку до ізотропної (наприклад, антена типу «листок конюшини» – рис. 7.5.15 а), так і великий коефіцієнт спрямованої дії (антенні ґратки, розташовані на поверхні сфери, циліндра, конуса тощо – рис. 7.5.15 б-г, рис. 7.5.16).

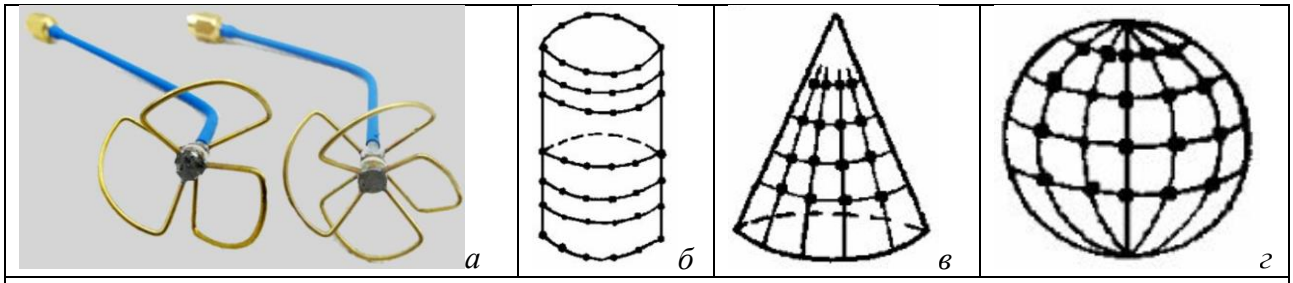


Рис. 7.5.15. Просторові групові випромінювачі: а – антени типу «листок конюшини»: фото; б-г – антенні ґратки, розміщені на поверхні циліндра (а), конуса (б) і сфери (г): схематичні зображення

Рис. 7.5.16. Циліндрична антенна ґратка: фото

7.5.4. Випромінювальні структури

Випромінювальні структури можна визначити як випромінювачі складної форми, які не є набором простих випромінювачів. Їх можна розділити на плоскі та просторові.

Серед *плоских випромінювальних структур* виділяють кілька груп.

У *прямолінійних випромінювальних структурах* окремі елементи розташовуються уздовж прямої. Прикладом може бути логоперіодична вібраторна антена (рис. 7.5.17 а,

[7] На відміну від квадратної рамкової антени, про яку йшлося раніше, ромбічна антена незамкнена, і кільцевий струм через неї не протікає.

рис. 7.5.18).

Прикладом *ламаних випромінювальних структур* може служити логоперіодична V-подібна антена (рис. 7.5.17 б).

Прикладами *складчастих випромінювальних структур* можуть бути антени у формі меандру та зигзаґу (рис. 7.5.19 а-б).

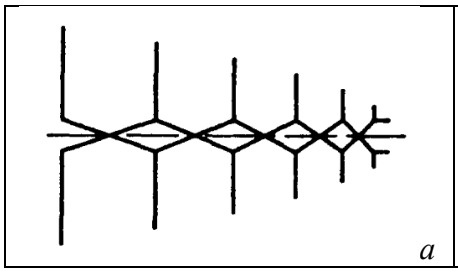


Рис. 7.5.17. Схематичні зображення прямолінійних (а) та ламаних (б) випромінювальні структур: вібраторна (а) та V-подібна (б) логоперіодичні антени

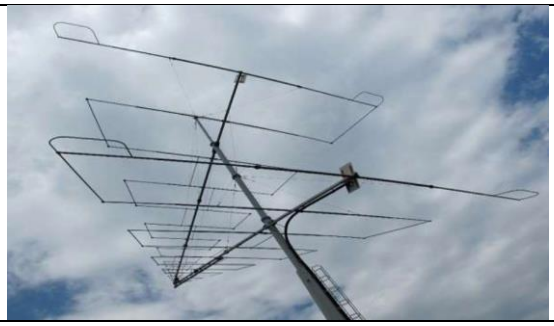
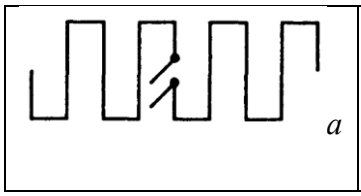


Рис. 7.5.18. Логоперіодична вібраторна антена: фото



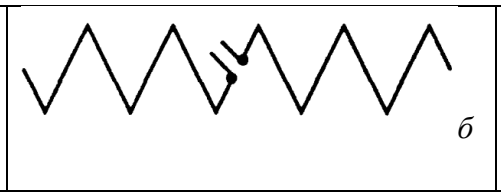
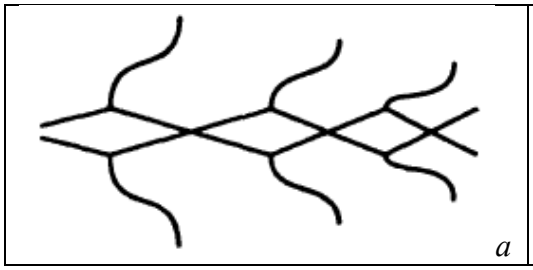


Рис. 7.5.19. Схематичні зображення складчастих випромінювальних структур – антен у формі меандру (а) та зигзаґу (б)

Прикладом *антени з вигинами* може служити логоперіодичний вібраторний випромінювач Ландсторфера (рис. 7.5.20 а), прикладом *площинної антени* – логоперіодична планарна антена (рис. 7.5.20 б).

Нарешті, прикладом *плоских закручених антен* (плоских спіральних структур) є архімедова спіраль та логарифмічна спіраль (рис. 7.5.21 а-б).



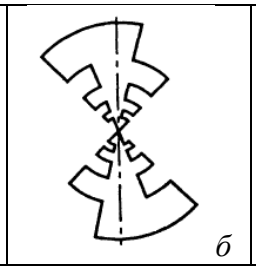
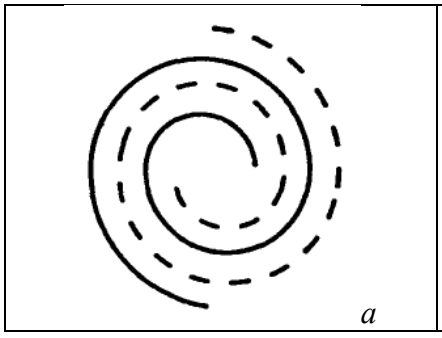


Рис. 7.5.20. Схематичні зображення випромінювальних структур з вигинами (а) та площинних структур (б): а – логоперіодичний вібраторний випромінювач Ландсторфера; б – логоперіодична планарна антена



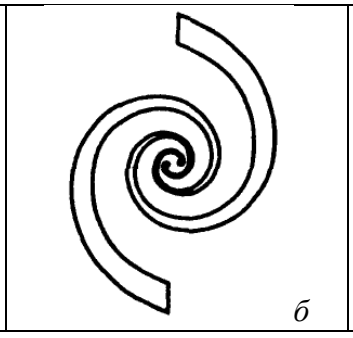


Рис. 7.5.21. Схематичні зображення плоских закручених антен: а – архімедова спіраль; б – логарифмічна спіраль

Просторові випромінювальні структури є принципово тривимірними. Їх також можна розділити на кілька груп.

Прикладом *структур зі зламом* можуть бути дві логоперіодичні антени, які дотикаються передніми кінцями (рис. 7.5.22 а).

Циліндричні спіральні антени мають форму гвинтових ліній (рис. 7.5.22 б, рис. 7.5.23). Залежно від їхнього радіусу розрізняють два характерні режими роботи. Нормальний режим (інакше його називають режимом всеспрямованого, або радіального випромінювання) реалізується, коли радіус спіралі значно менший за довжину випромінюваних хвиль. У цьому випадку випромінювання поширюється нормально (перпендикулярно) до осі спіралі, тобто випромінювач уподібнюється до диполя Герца. Аксіальний режим реалізується, коли радіус спіралі наближається до довжини хвилі. В цьому випадку випромінювання поширюється в аксіальному напрямку (вздовж осі спіралі). Строго на осі воно поляризоване по колу, під кутами до осі – по еліпсу.

Конічні спіральні антени мають форму гвинтової лінії зі змінним радіусом, яка лежить на бічній поверхні конуса (рис. 7.5.22 в). Живлення може подаватися до вершини цього конуса або до його основи. Конічні антени, як і циліндричні, можуть працювати в нормальному та аксіальному режимах.

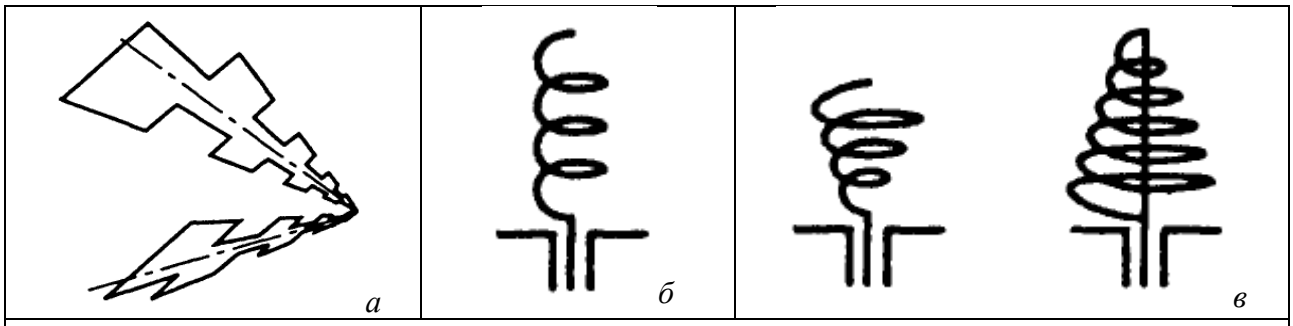


Рис. 7.5.22. Схематичні зображення просторових випромінювальних структур: а – логоперіодичні випромінювачі, передні кінці яких дотикаються; б – циліндрична спіральна структура; в – конічні спіральні структури

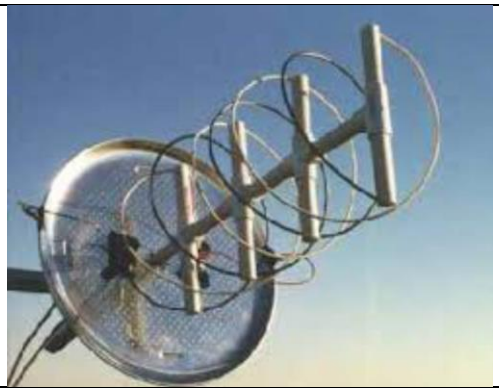


Рис. 7.5.23. Циліндрична спіральна антена: фото

7.5.5. Апертурні випромінювачі

Апертурні випромінювачі характеризуються тим, що струми, які безпосередньо породжують випромінювання, протікають по деякій провідній поверхні (апертурі). Зазвичай розміри апертурних антен більші (інколи значно більші) за довжину випромінюваних хвиль. Очевидно, діаграма спрямованості таких антен визначається, по-перше, формою та розмірами апертури, а по-друге – просторовим розподілом струмів, які по ній протікають. До цієї категорії зараховують і антени, що випромінюють вищі гармоніки.

Серед апертурних антен виділяють рефлекторні (дзеркальні), антени на основі відкритого кінця хвилеводу (зокрема, рупорні), лінзові, діелектричні та антени на розсіяних хвилях.

Антени рефлекторного типу складаються з випромінювача й відбивача (дзеркала) того

чи іншого типу (рис. 7.5.24 а-в, рис. 7.5.25-7.5.26). Вплив плоского дзеркала ми, по суті, вже розглянули на прикладі впливу земної поверхні на випромінення півхвильового вібратора (див. вище п. 7.3.3).

На практиці найчастіше застосовують параболічні дзеркала (рис. 7.5.24 в, рис. 7.5.25). Це пов'язано з тим, що параболоїд обертання фокусує пучок паралельних променів, паралельних його осі, в точку. І навпаки, якщо в фокусі такого дзеркала розмістити випромінювач, то в результаті відбиття від дзеркала формується пучок паралельних променів. Зазначимо, що ці результати отримано в наближенні геометричної оптики (див. нижче рис. 7.6.19). Для застосовності даного наближення потрібно, щоб розміри дзеркала були значно більшими за довжину хвилі.

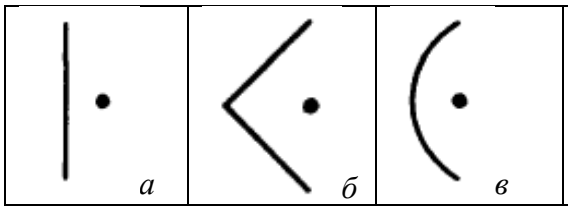


Рис. 7.5.24. Схематичні зображення апертурних випромінювачів рефлекторного типу (перерізи): а – з плоским дзеркалом; б – з кутовим відбивачем; в – з параболічним дзеркалом

Рис. 7.5.25. Антени з параболічним дзеркалом: фото

Рис. 7.5.26. Антена з кутовим відбивачем: фото

Антени на основі відкритого кінця хвилеводу можна розбити на дві групи. Одна з них включає відкритий кінець циліндричного або прямокутного хвилеводу. Хвиля, що біжить у такому хвилеводі в напрямку до відкритого кінця, буде частково відбиватися від нього, а частково випромінюватися в навколишній простір.

Модифікацією таких випромінювачів є *рупорні антени*, у яких відкрита частина хвилеводу розширюється і набуває форми піраміди або конуса (рис. 7.5.27 а-г, рис. 7.5.28). Розширення кінця хвилеводу збільшує випромінювальну спроможність системи, забезпечуючи узгодження опору випромінювання антени з хвильовим опором вільного простору (див. вище п. 6.6 та п. 6.7.8), а плавна зміна розмірів хвилеводу різко зменшує коефіцієнт відбиття від такої неоднорідності. Підбір форми рупора дозволяє керувати поляризацією випромінювання.

Гофрований рупор (рис. 7.5.29) – це рупорна антена з паралельними прорізами або канавками, розміри яких значно менші за довжину хвилі. Канавки вкривають внутрішню поверхню рупора й розташовані поперечно щодо оси. Гофровані рупори мають ширшу смугу

пропускання, у них менші бічні пелюстки в діаграмі спрямованості. Вони широко використовуються як рупори для супутникових антен і радіотелескопів.

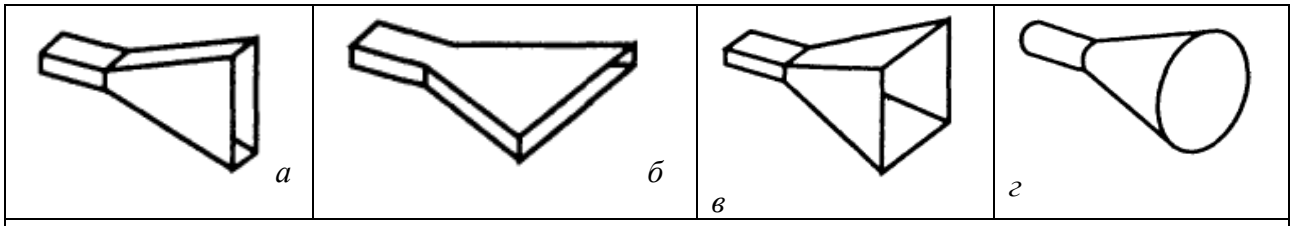


Рис. 7.5.27. Схематичні зображення рупорних антен: а – Е-площинний секторіальний рупор; б – Н-площинний секторіальний рупор; в – пірамідальний рупор; г – конічний рупор

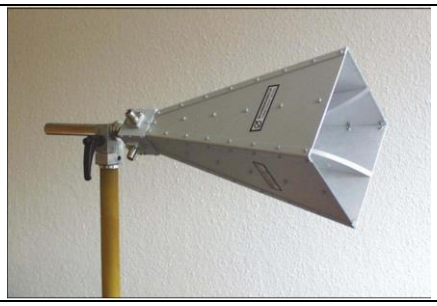


Рис. 7.5.28. Пірамідальна рупорна антена (фото)

Рис. 7.5.29. Гофрована конічна рупорна антена (фото)

Лінзові антени використовуються для покращення форми викривленого фронту хвилі. Вони можуть бути прискорювальними або сповільнюваними (перші збільшують фазову швидкість хвилі, другі – навпаки, зменшують). Лінзи можуть, зокрема, виготовлятися з діелектрика (рис. 7.5.30 а, б, рис. 7.5.31). Робота таких лінз аналогічна до звичайних оптичних лінз, відповідно їхні розміри мають бути значно більшими за довжину хвилі. Лінза ближнього поля (рис. 7.5.30 в) видозмінює поле випромінювача безпосередньо в близькій зоні (нагадаємо, що в просторово-неоднорідному середовищі електричне поле концентрується в областях із більшою діелектричною проникністю).

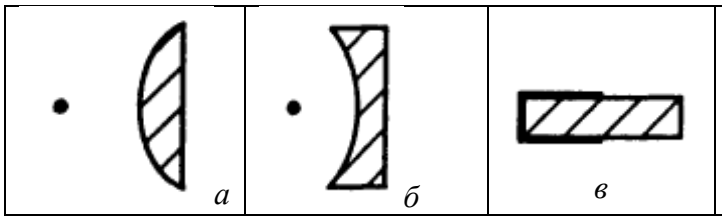


Рис. 7.5.30. Схематичні зображення лінзових антен: а – прискорювальна лінза; б – сповільнювальна лінза; в – лінза ближнього поля

Рис. 7.5.31. Лінзова антена: фото

Діелектричні антени, як зрозуміло з їхньої назви, виготовляють із діелектрика.

Розрізняють короткі та довгі діелектричні антени. Коротка діелектрична антена – це те саме, що лінза ближнього поля (рис. 7.5.30 в). Найпростіша довга антена – це видовжений діелектричний стрижень (рис. 7.5.32). Такий стрижень у певному діапазоні частот може виступати як хвилевід. Крім того, в ньому можуть існувати так звані поверхневі хвилі, поле яких максимальне на бічній поверхні, а при віддаленні від неї спадає за законом, близьким до експоненціального. Ці хвилі й використовуються для збудження електромагнітного випромінювання.

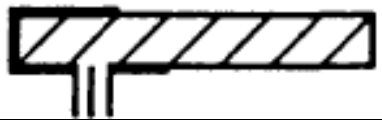
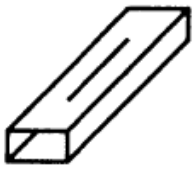
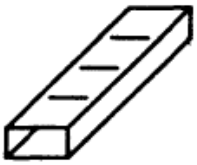


Рис. 7.5.32. Схематичне зображення полістиролової стрижневої антени – приклад діелектричної антени

а



б



в

Рис. 7.5.33. Схематичні зображення антен на розсіяних хвилях: а – щілинна антена на коаксіальній лінії; б-в – хвилеводні щілинні антени

Антена на розсіяних хвилях – це ділянка прямокутного (рис. 7.5.33 б, в, рис. 7.5.34) чи циліндричного хвилеводу або коаксіального кабелю (рис. 7.5.33 а) з однією або кількома щілинами, які й служать випромінювачами. Щілини можуть бути поздовжніми (рис. 7.5.33 а, б) або поперечними (рис. 7.5.33 в). Окреме місце займають гофровані антени на поверхневих хвилях (рис. 7.5.35).

Рис. 7.5.34. Щілинна антена: фото

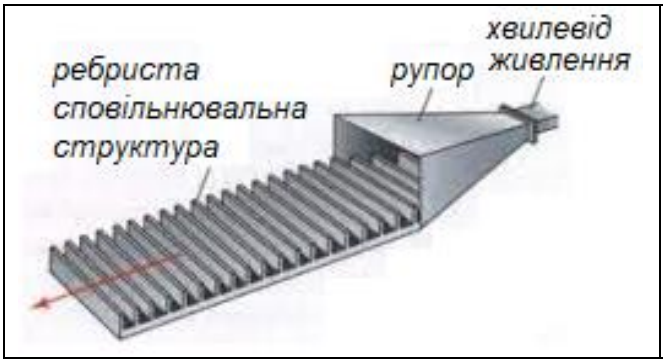


Рис. 7.5.35. Схематичне зображення антени на поверхневих хвилях на основі ребристої сповільнювальної структури

Контрольні питання до підрозділу 7.5

1. Які ознаки можуть бути покладені в основу класифікації антен?
2. Які елементарні випромінювачі Вам відомі?
3. Що таке прості антени? Які типи простих антен Ви знаєте?
4. В чому полягає особливість фігурних антен порівняно з лінійними?
5. Опишіть принцип електронного управління діаграмою спрямованості у фазованій антенній ґратці.

6. В чому специфіка випромінювальних структур у порівнянні з іншими типами антен?
7. Для чого використовуються спіральні антени?
8. Що таке апертурні випромінювачі?
9. Які типи квазіоптичних антен Ви знаєте?
10. В чому перевага параболічних дзеркальних антен над іншими типами дзеркальних антен?
11. Які типи антен на розсіяних хвилях Вам відомі?

Задачі до підрозділу 7.5

7.5.1. Фазована антенна ґратка показана на рис. 7.5.8 а. Віддаль між сусідніми випромінювачами – 1 м, частота випромінюваних хвиль – 10 МГц. Яким має бути набіг фази між сусідніми випромінювачами, щоб хвильовий вектор випромінюваних хвиль був спрямований під кутом 45º до поверхні антени? Яка при цьому буде фазова швидкість хвилі вздовж поверхні антени?

7.5.2. Переріз параболічної антени описується формулою $y=x^2$. Знайти положення фокусу цієї антени.

7.6. Антени різних частотних діапазонів

Поговоримо тепер про застосування антен у різних частотних діапазонах.

7.6.1. Випромінювальні антени довгих та середніх хвиль

Почнемо з випромінювальних антен.

Наддовгі (понад 10 км), довгі (1-10 км) та середні (0.1-1 км) хвилі (див. вище п. В.2) зазвичай поширюються вздовж поверхні Землі. Через те для випромінювання таких хвиль використовують антени, у яких максимум характеристики спрямованості лежить у горизонтальній площині.

Такою антеною є, зокрема, вертикально орієнтований несиметричний вібратор (монополь, уніполь, півдиполь, антена Марконі), який належить до класу лінійних антен (див. вище п. 7.5.2). Горизонтальні вібратори слабко випромінюють уздовж земної поверхні, тому вони зазвичай не застосовуються.

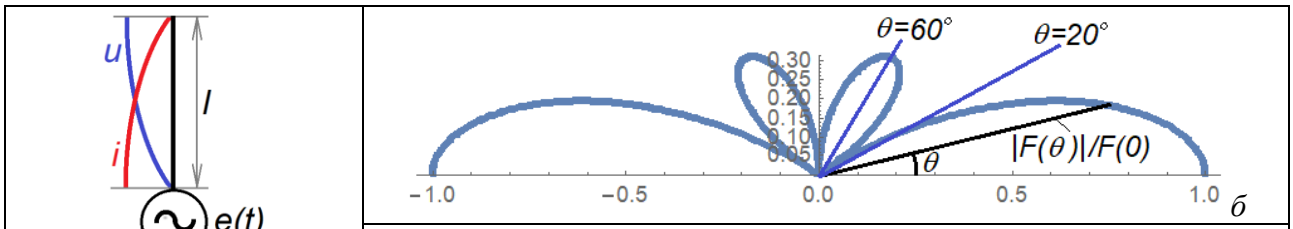


Рис. 7.6.1. Вертикальний несиметричний вібратор із заземленням: а – схематичне зображення й розподіл струму та напруги по його висоті; б – графік його діаграми спрямованості у вертикальній площині для випадку $l = 0.63\lambda$ $(kl = 4)$

Схема несиметричного вертикального вібратора, розподіл струму та напруги по його висоті та його діаграма спрямованості у вертикальній площині подані на рис. 7.6.1 а-в. Ці розподіли нагадують відрізок довгої лінії, у якого один кінець розімкнутий, а другий – замкнений накоротко (див. вище п. 6.7.4). Оскільки вісь такої антени розташована вертикально, її діаграма спрямованості є аксіально-симетричною.

В загальному випадку характеристика спрямованості несиметричного вертикального вібратора залежить від відношення l/λ й описується формулою

$$F(\theta) = \frac{\cos(kl\sin\theta) - \cos(kl)}{\left[1 - \cos(kl)\right]\cos\theta}, \qquad k = 2\pi/\lambda. \tag{7.6.1}$$

При $kl \le 4.4$, або $l/\lambda \le 0.7$ головна пелюстка діаграми спрямованості орієнтована горизонтально (рис. 7.6.1 б). Зі зростанням kl характеристика спрямованості стає більш порізаною (рис. 7.6.2).

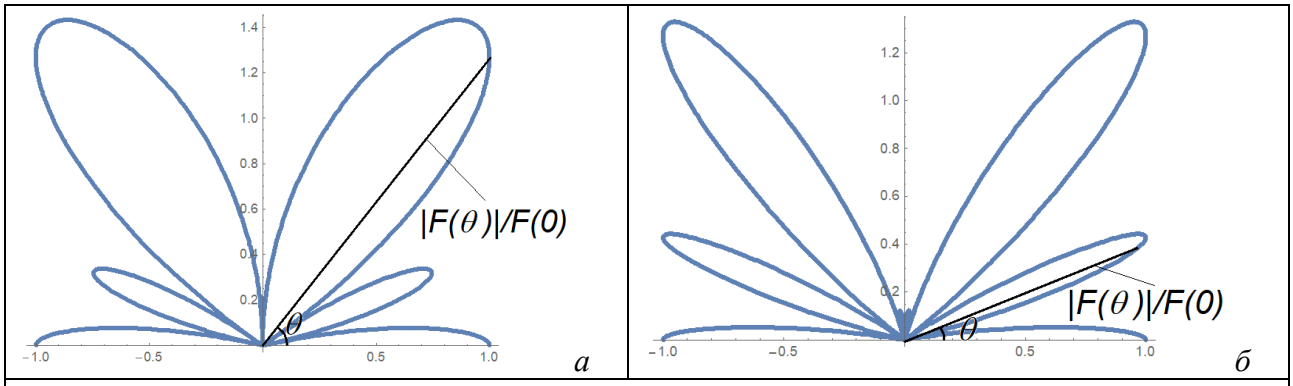


Рис. 7.6.2. Графіки характеристики спрямованості несиметричного вертикального вібратора при $kl = 8$ *(а) та* $kl = 16$ *(б)*

Формула (7.6.1) записана в припущенні, що поверхня Землі є ідеально провідною. Це припущення добре виконується для довгих та наддовгих хвиль (див. нижче п. 8.4.2). Для середніх та коротких хвиль дається взнаки скінченна провідність земної поверхні, і головна пелюстка виявляється спрямованою під ненульовим кутом θ_{max} до горизонту.

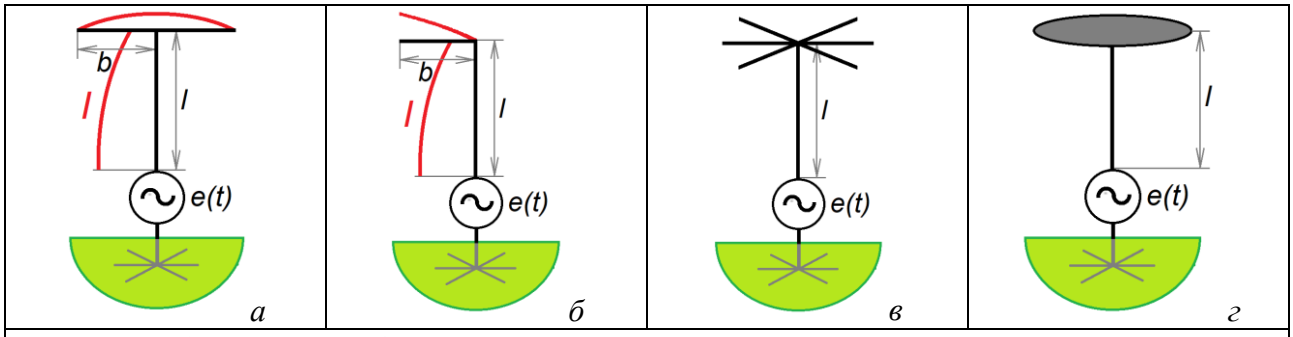


Рис. 7.6.3. Схематичні зображення вертикальних асиметричних антен з горизонтальними частинами: а – Т-подібна; б – Г-подібна; в – парасолькова; г – дискова

Оскільки вертикальний несиметричний вібратор випромінює лише у верхню півсферу, його максимальний к.к.д. виявляється вдвічі більшим, ніж у симетричного. В той же час при однакових струмах і довжинах плеч потужність випромінювання (а, отже, й опір випромінювання) несиметричного вібратора виявляється вдвічі меншим, ніж у симетричного. При $l/\lambda \le 0.1$ можна вважати, що опір випромінювання $R_r \sim (l/\lambda)^2$ (див. вище п. 7.4.3), тому для його збільшення треба збільшувати довжину антени. Такого ж результату можна досягти, додавши до антени горизонтальні частини, які збільшують її ефективну довжину. Будова деяких типів таких антен показана на рис. 7.6.3 а-г, на рис. 7.6.3 а-б показаний також розподіл амплітуди струму в різних частинах антени.

В реальності висота антени визначається довжиною хвилі та економічними міркуваннями. Для наддовгих та довгих хвиль (довжина понад 1 км) умова $l/\lambda \ll 1$ добре виконується. При цьому формула (7.6.1) набуває вигляду $f(\theta) \approx \cos\theta$. Відповідні антени звичайно мають висоту 150-250 м і більше. В діапазоні середніх хвиль (100м – 1км) розмір антени сумірний з довжиною хвилі – звичайно її висоту обирають у межах $(0.16 - 0.63)\lambda$.

Типовий вигляд антени довгохвильового діапазону показаний на рис. 7.6.4 а, діапазону середніх хвиль – на рис. 7.6.4 б.

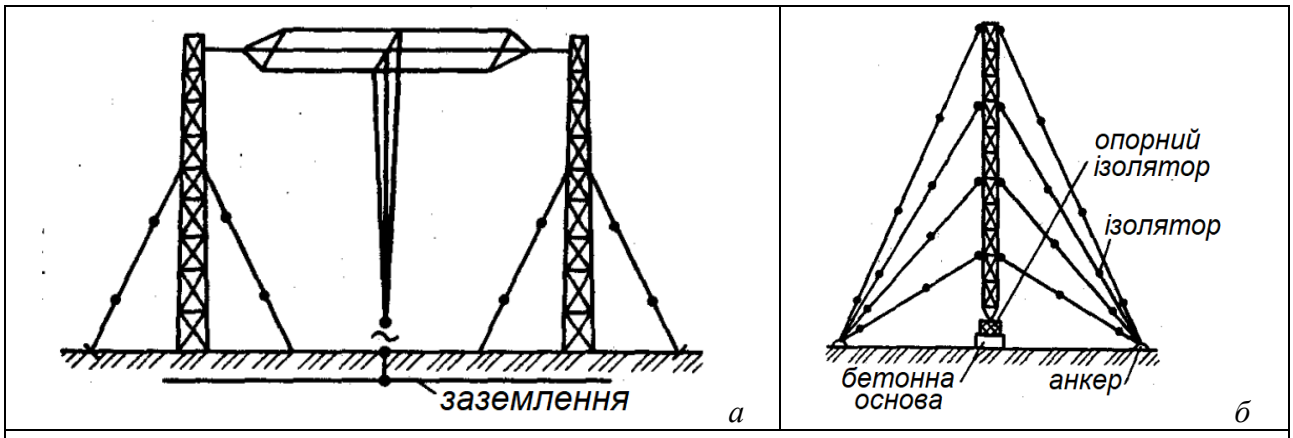


Рис. 7.6.4. Схематичні зображення типових антен діапазону довгих (а) і середніх (б) хвиль

У діапазоні середніх хвиль спостерігається явище завмирання радіосигналу (федінг), зумовлене інтерференцією двох хвиль, що дістаються точки приймання по різних траєкторіях (див. нижче п. 8.7). Для боротьби з цим явищем діаграма спрямованості антени має бути притиснутою до земної поверхні (так звані антифедінгові антени). Це забезпечується

вертикальними антенами довжиною $(0.53-0.58)\lambda$. Діаграма спрямованості такої антени подана на рис. 7.6.5.

Для підвищення потужності випромінювання використовують також *складні антени з кількох пов'язаних вібраторів* (антени з багатьма зниженнями). Приклад такої антени, складеної з трьох суміжних Т-подібних антен, показаний на рис. 7.6.6.

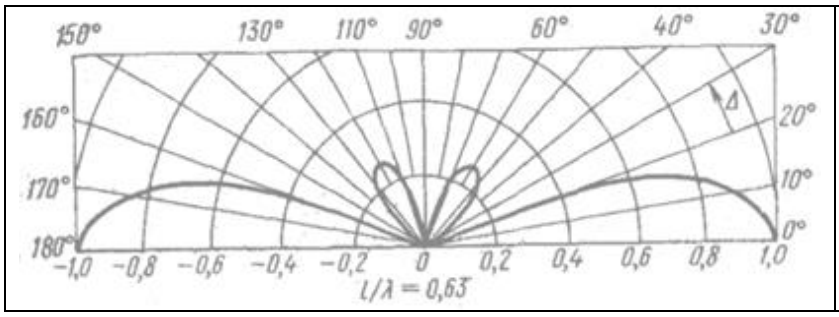


Рис. 7.6.5. Діаграма спрямованості антифедінгової антени

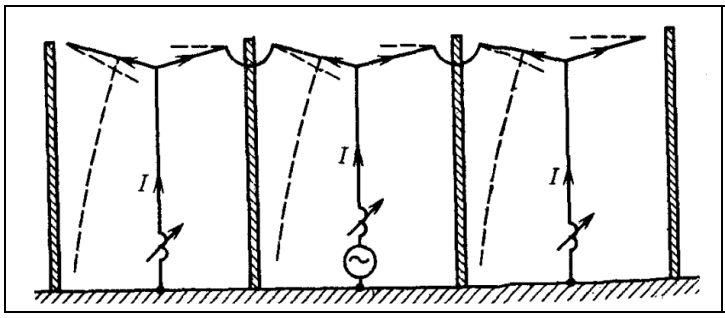


Рис. 7.6.6. Схематичне зображення складної антени з трьох суміжних Т-подібних антен (штриховими лініями показані розподіли струмів у гілках)

В таких антенах всі зниження (окремі вібратори) налаштовують шляхом підбору змінних котушок індуктивності так, що їхні струми виявляються синфазними й приблизно однаковими за амплітудою. Через синфазність струмів і мализну віддалі між зниженнями порівняно з довжиною хвилі випромінювання від окремих знижень також виявляється синфазним. В результаті амплітуда випромінювання зростає в n разів (тут n – загальна кількість знижень), а потужність – в n^2 разів порівняно з випромінюванням одиночної антени.

7.6.2. Приймальні антени довгих та середніх хвиль

Як уже вказувалося, в силу теореми взаємності передавальні антени можуть служити і для приймання сигналу (див. вище пп. 7.4.5-7.4.6). Однак такий підхід не завжди зручний. Справді, для приймальних антен відсутні проблеми, пов'язані з необхідністю використовувати велику потужність сигналу живлення і, відповідно, великі струми в антені. Крім того, щоб підвищити відношення сигнал-шум, бажано використовувати антени з великим коефіцієнтом спрямованої дії.

Основні типи приймальних антен, що використовуються в діапазоні довгих та середніх хвиль – це вертикальні несиметричні вібратори (відрізняються від аналогічних передавальних антен лише меншими розмірами та конструктивними особливостями), рамкові антени та антени біжучої хвилі.

Рамкова антена складається з одного або кількох послідовно сполучених витків дроту. Ці витки можуть мати форму кола, квадрата, трикутника тощо (рис. 7.5.4-7.5.5). Площина витків перпендикулярна поверхні Землі (вертикальна рамка). Розміри рамкових антен варіюються від сантиметрів (в цьому випадку витки намотуються на магнітодіелектричне осердя, див. нижче) до сотень метрів. Реально вони завжди набагато менші, ніж довжина хвилі.

Розглянемо модель прямокутної рамкової антени, показану на рис. 7.6.7. Вісь z спрямована вертикально (перпендикулярно до поверхні Землі. Рамка розміром $h \times b$ $(h, b \ll \lambda)$ лежить у вертикальній площині xz. На антену падає вертикально поляризована електромагнітна хвиля. Її хвильовий вектор $\vec{k} = \vec{n}\omega/c$ лежить у горизонтальній площині xy, утворюючи кут φ

з віссю x. Електричне поле хвилі паралельне z. Це поле перпендикулярне горизонтальним ділянкам рамки AC і BD, тому е.р.с. буде наводитися лише у вертикальних ділянках AB і CD.

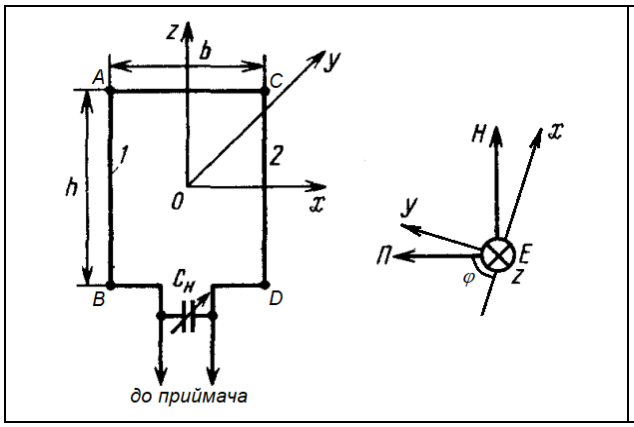


Рис. 7.6.7. До розрахунку е.р.с. у приймальній рамковій антені

Нехай комплексна амплітуда електричного поля в точці O (в центрі рамки) складає величину E_m. На ділянці AB поле зсунуте за фазою щодо точки O, тобто його комплексна амплітуда дорівнює

$$E_1 = E_m \exp\left[-i(kb\cos\varphi)/2\right], \tag{7.6.2}$$

а на ділянці CD – відповідно

$$E_2 = E_m \exp\left[i(kb\cos\varphi)/2\right]. \tag{7.6.3}$$

З урахуванням сказано раніше амплітуда е.р.с., наведена в рамці, складає величину

$$\begin{aligned} e_m &= (E_2 - E_1)h = E_m h\left\{\exp\left[i(kb\cos\varphi)/2\right] - \exp\left[-i(kb\cos\varphi)/2\right]\right\} = \\ &= 2iE_m h\sin\left[(kb\cos\varphi)/2\right] \approx iE_m kbh\cos\varphi = iE_m kS\cos\varphi, \qquad S = bh \end{aligned} \tag{7.6.4}$$

(враховано, що $kb \ll 1$).

Визначимо ефективну довжину антени як

$$l_{eff} = \frac{|e_m|_{\max}}{E_m}. \tag{7.6.5}$$

Тоді для аналізованої моделі

$$l_{eff} = kS = 2\pi\frac{hb}{\lambda} \ll \lambda. \tag{7.6.6}$$

Мализна ефективної довжини приводить до малого опору випромінювання таких антен (зазвичай він не перевищує 0.1 Ом і виявляється значно меншим від опору втрат). Відповідно й к.к.д виявляється дуже малим. Тому рамкові антени, як правило, не використовують для передавання сигналів.

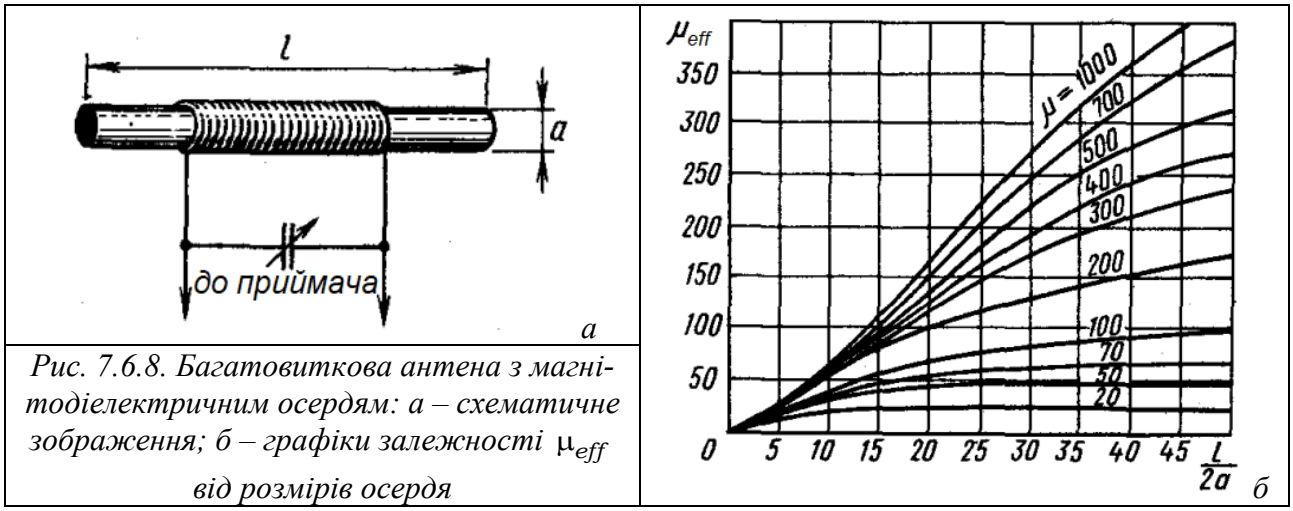


Рис. 7.6.8. Багатовиткова антена з магнітодіелектричним осердям: а – схематичне зображення; б – графіки залежності μ_{eff} від розмірів осердя

Для підвищення опору випромінювання рамкові антени роблять багатовитковими і використовують магнітодіелектричні осердя (рис. 7.6.8 а). Ефективна довжина такої антени визначається співвідношенням

$$l_{eff} = \mu_{eff} nkS, \qquad (7.6.7)$$

де μ_{eff} визначається відносною магнітною проникністю осердя μ та відношенням довжини осердя до його діаметра (рис. 7.6.8 б), n – кількість витків. Малі габарити таких антен дозволяють розміщувати їх безпосередньо в корпусі радіоприймача.

Антена біжучої хвилі[8] складена з багатьох однакових симетричних вібраторів, розміщених на однакових віддалях один від одного і приєднаних через деякі елементи зв'язку Z_c до збиральної лінії (рис. 7.6.9)[9]. До початку збиральної лінії приєднується приймач, до кінця – узгоджений опір навантаження R_l, величина якого дорівнює хвильовому опору лінії (рис. 7.6.9 а). Антена розміщується горизонтально на висоті 16-40 м над землею (рис. 7.6.9 б).

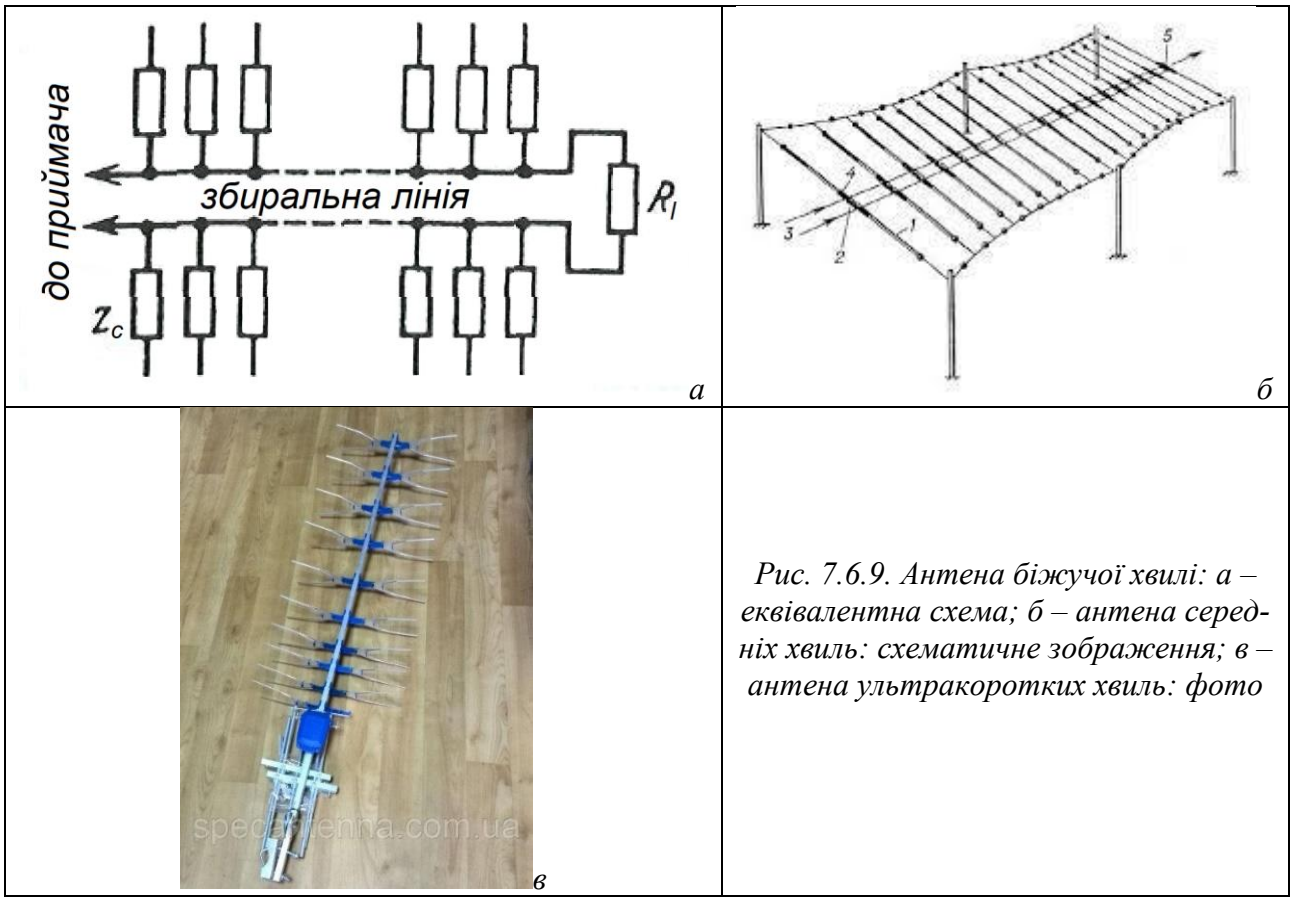


Рис. 7.6.9. Антена біжучої хвилі: а – еквівалентна схема; б – антена середніх хвиль: схематичне зображення; в – антена ультракоротких хвиль: фото

Антена біжучої хвилі – це широкодіапазонна антена, коефіцієнт перекриття діапазону $\kappa = f_{\max}/f_{\min}$ може досягати 2.5-3.0. Як правило, такі антени використовуються для приймання сигналу.

Частинним випадком антен біжучої хвилі є антена Уда – Ягі (хвильовий канал) – див. вище п. 7.5.3, рис. 7.5.9 г та рис. 7.5.11.

Якщо віддаль між сусідніми вібраторами значно менша за довжину хвилі, таку антену можна розглядати як еквівалентну довгу лінію (див. вище п. 6.2.3), погонні параметри якої залежать від вхідного імпедансу окремого вібратора Z_{in} та імпедансу зв'язку Z_c. Тоді

[8] Інші варіанти антени біжучої хвилі згадувався в п. 7.5.2 (антена Бевереджа, див. рис. 7.5.2 в), а також у п. 7.5.4 (циліндрична спіральна антена в аксіальному режимі, див. рис. 7.5.22 б та рис. 7.5.23).

[9] Відмінність такої антени від розглянутого в п. 7.5.3 паралельного випромінювача полягає в тому, що прості випромінювачі, з яких він складений, збуджуються синфазно.

додаткова погонна провідність збиральної лінії, зумовлена впливом вібраторів, визначається співвідношенням

$$Y_1 = \frac{1}{(Z_c + 2Z_{in})d}, \qquad (7.6.8)$$

де d – віддаль між сусідніми вібраторами.

Зміна погонної провідності лінії приводить до зміни її хвильового опору та фазової швидкості υ_{ph} хвиль у ній.

Якщо $|Y_1| \ll |Y_0|$, де Y_0 – власна провідність збиральної лінії, то остання зберігає високий к.б.х. в широкому діапазоні довжин хвиль. Мализна Y_1 забезпечується відповідним підбором імпедансу зв'язку Z_c.

Якщо Z_c має активний або ємнісний характер, реалізується так званий режим сповільнення, тобто υ_{ph} зменшується ($\upsilon_{ph} < c$). При цьому максимум випромінювання (якщо антена працює як передавальна) спрямований уздовж її осі в бік навантаження. Так само виглядає й діаграма спрямованості на приймання сигналу.

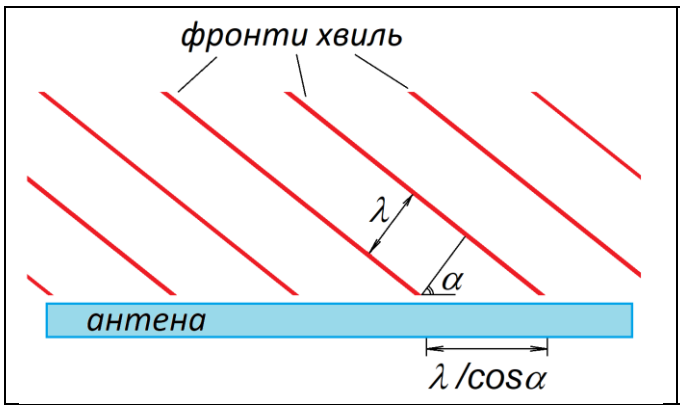


Рис. 7.6.10. Синхронізм між швидкою хвилею в антені та випромінюваною електромагнітною хвилею

Якщо Z_c має індуктивний характер, то $\upsilon_{ph} > c$, то випромінювання (і, відповідно, приймання сигналу) відбувається під кутом

$$\alpha = \arccos\frac{c}{\upsilon_{ph}} \qquad (7.6.9)$$

до осі антени (режим синхронізму між випромінюванням та хвилею в антені, рис. 7.6.10).

В режимі пришвидчення ($\upsilon_{ph} > c$) амплітуда хвилі в антені через втрати на випромінювання експоненціально спадає по довжині антени, тому на практиці такий режим не застосовується.

В режимі сповільнення діаграма спрямованості антени біжучої хвилі у вертикальній площині в припущенні ідеально провідної землі визначається формулою

$$F(\theta) = \frac{\sin\left[\frac{nkd}{2}\left(\frac{c}{\upsilon_{ph}} - \cos\theta\right)\right]}{\sin\left[\frac{kd}{2}\left(\frac{c}{\upsilon_{ph}} - \cos\theta\right)\right]}\sin(kH\sin\theta), \qquad (7.6.10)$$

де $k = 2\pi/\lambda$, n – загальна кількість вібраторів, H – висота над поверхнею Землі, на якій розміщена антена (рис. 7.6.11). Звичайно обирають n від 20 до 40.

Для описаного типу антени біжучої хвилі типові значення коефіцієнту спрямованої дії лежать у межах від 35 до 200, к.к.д. – (15-40)%

Для підвищення к.с.д. інколи використовують два полотна антени біжучої хвилі, які розташовують поруч і живлять від одного й того самого джерела (рис. 7.6.12)

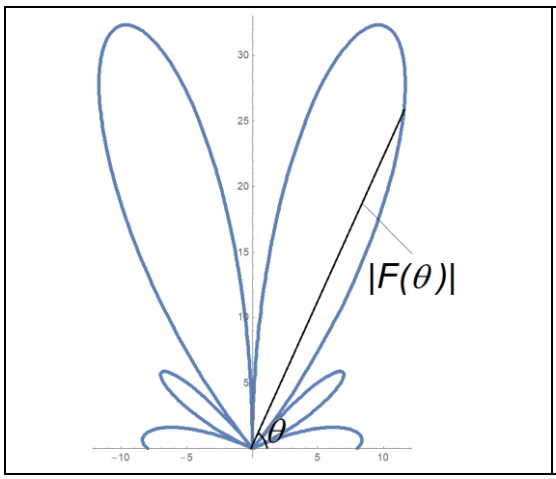


Рис. 7.6.11. Діаграма спрямованості антени біжучої хвилі: $kd = 0.2$, $n = 40$, $c/v_{ph} = 1.2$, $kH = 5$

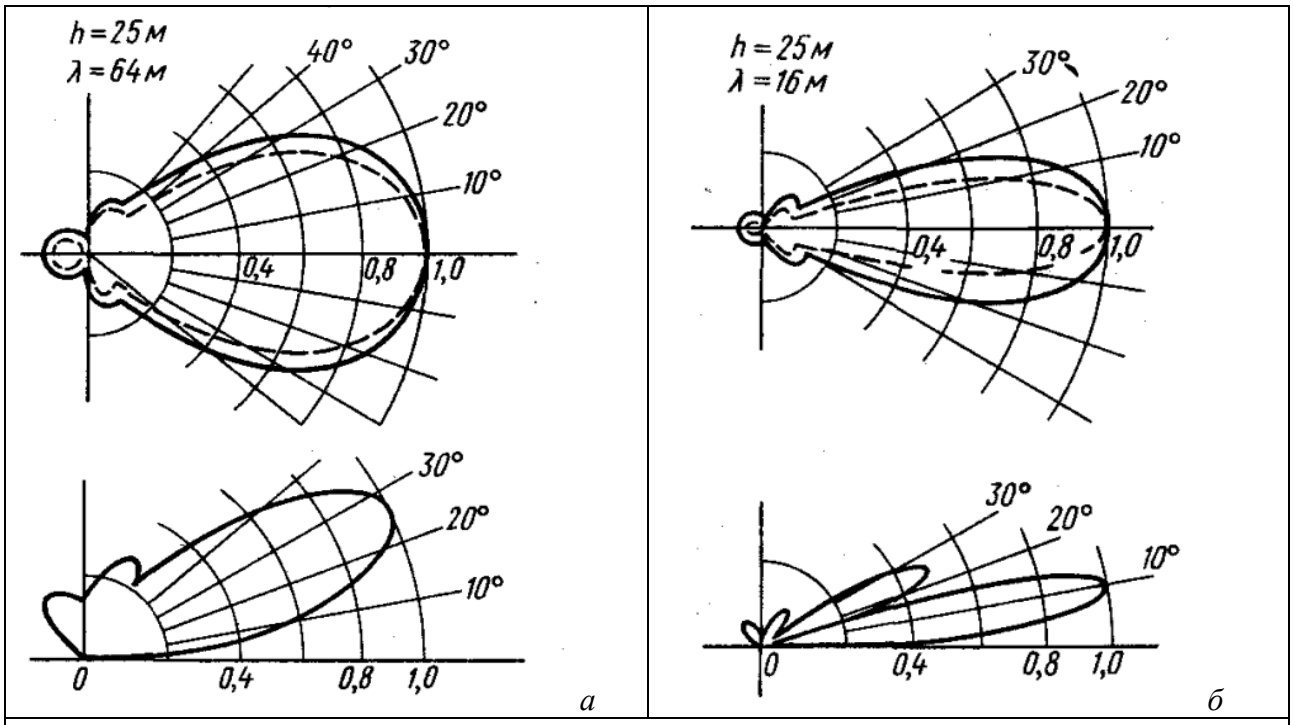


Рис.7.6.12. Перерізи діаграми спрямованості одиночної (а) та здвоєної (б) антени біжучої хвилі: верхні рисунки – в горизонтальній площині, нижні – у вертикальній

7.6.3. Особливості антен короткохвильового діапазону

У короткохвильовому діапазоні ($\lambda = 10 \div 100 м$) вдається побудувати антени з розмірами порядку довжини хвилі. У таких антен опір випромінювання перевищує опір втрат, а к.к.д., відповідно, наближається до одиниці (див. вище п. 7.4.3).

Умови поширення коротких радіохвиль визначаються станом іоносфери Землі. Останній суттєво відрізняється вдень і вночі, залежить також від пори року, сонячної активності тощо (див. нижче розділ 8). Тому для забезпечення стабільного зв'язку бажано мати можливість працювати з кількома суттєво відмінними довжинами хвиль. Для цього антена повинна бути широкодіапазонною.

Для зниження впливу завад к.с.д. антени повинен бути великим, а рівень бічних пелюсток – малим.

Найпростіші антени, що використовуються в діапазоні коротких хвиль – це *горизонтальні симетричні вібратори* (див. вище п. 7.3.3, рис. 7.3.4 б) та їхні модифікації. Вони використовуються в діапазоні довжин хвиль $\lambda = (1.5 \div 4)l$, де l – довжина плеча вібратора.

Для збільшення допустимої вхідної потужності використовують вібратори з

пониженим хвильовим опором, зокрема диполь Надененка та діапазонний шунтовий вібратор (рис. 7.6.13).

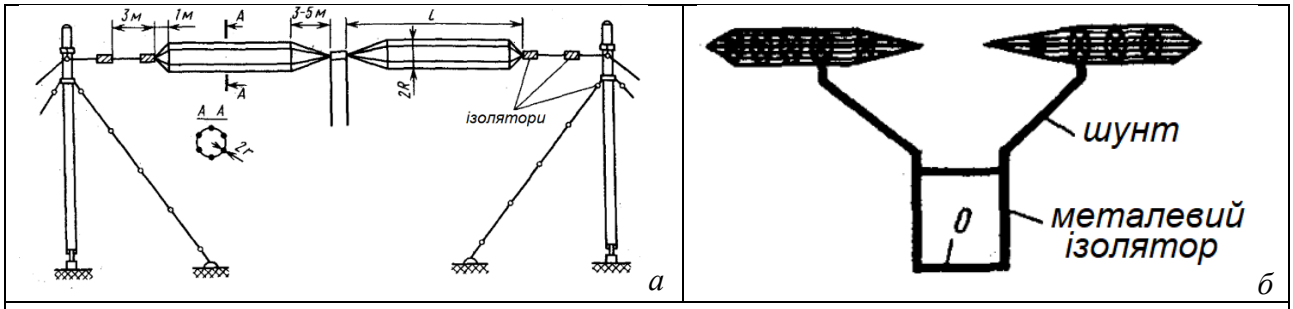


Рис. 7.6.13. Схематичні зображення вібраторів зі зниженим хвильовим опором: а – диполь Надененка; б – діапазонний шунтовий вібратор

Диполь Надененка має хвильовий опір у межах 250-600 Ом, що дозволяє забезпечити його узгодження з двопровідним фідером, який має хвильовий опір 600 Ом. Робочий діапазон такої антени оцінюють як $\Delta\lambda = (1.7 \div 3.3)l$.

Синфазні антени – це ґратки, складені з симетричних вібраторів (рис. 7.6.14) – див. вище п. 7.5.3, рис. 7.5.12 а. Полотно синфазної антени складається з кількох рядів (поверхів), кожен з яких включає ту саму кількість вібраторів. Віддаль між поверхами складає половину довжини хвилі. Оскільки всі вібратори збуджуються синфазно, зміна частоти не призводить до помітного погіршення властивостей антени. Для забезпечення випромінювання лише в один бік антена забезпечується пасивними (налаштовуваними) рефлекторами.

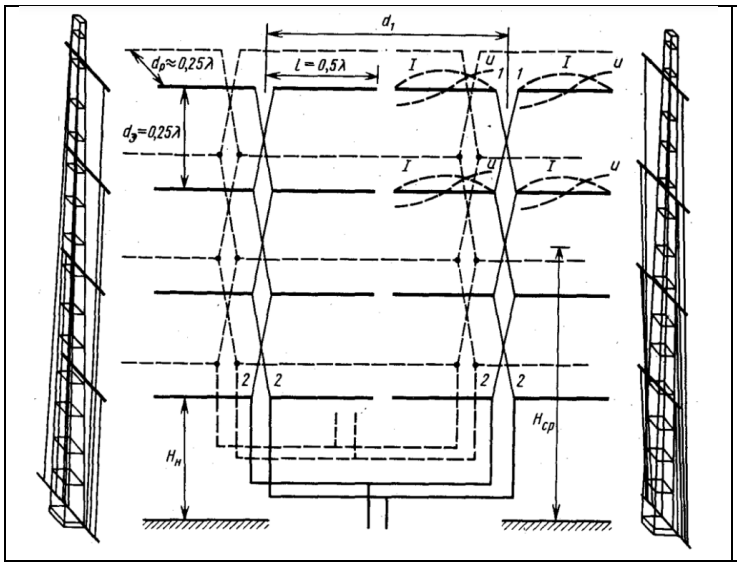


Рис. 7.6.14. Схематичне зображення синфазної ґратки, складеної з симетричних вібраторів

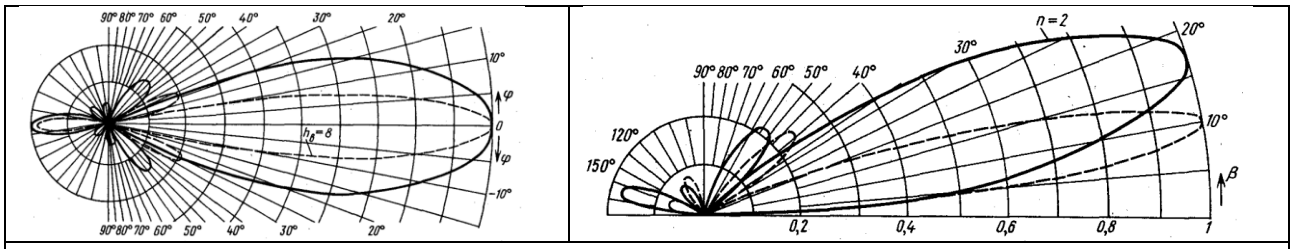


Рис. 7.6.15. Діаграма спрямованості синфазної ґратки: а – в горизонтальній площині для чотирьох (суцільна лінія) та восьми (штрихова лінія) випромінювачів на поверх; б – у вертикальній площині для двоповерхової (суцільна лінія) та чотириповерхової (штрихова лінія) антени, розміщеної на висоті в половину довжини хвилі над землею

Налаштовувані рефлектори складають полотно, аналогічне основному полотну антени.

Їхнє живлення забезпечується завдяки електромагнітному зв'язку між полотнами. Перемкнувши живлення з одного полотна на інше, можна змінити напрямок випромінювання антени на протилежний.

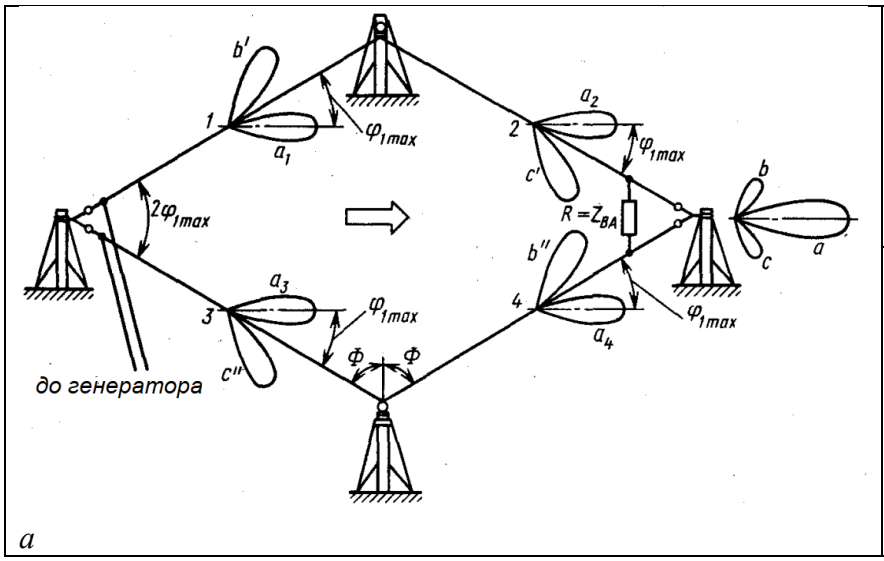
Аперіодичний рефлектор – це набір горизонтальних провідників, віддаль між якими значно менша за довжину хвилі, що встановлюється позаду антени.

На рис. 7.6.15 подано горизонтальний та вертикальний перерізи діаграми спрямованості такої антени. Видно, що зі збільшенням кількості випромінювачів кутові розміри головної пелюстки зменшуються.

Коефіцієнт спрямованої дії (к.с.д., див. вище п. 7.4.3) такої антени без врахування впливу земної поверхні можна оцінити за формулою

$$K_{dir} = \frac{16}{\pi} nm \approx 5.1\, nm \qquad (7.6.11)$$

(n – кількість поверхів антени, m – кількість випромінювачів на один поверх). Для антени з 4 поверхів по 8 випромінювачів на поверсі ця формула дає $K_{dir} \approx 160$.



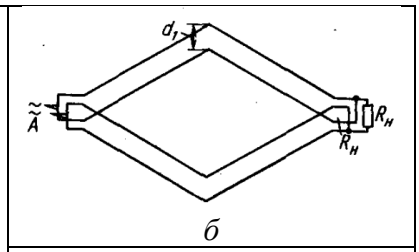


Рис. 7.6.16. Схематичні зображення ромбічної горизонтальної антени: а – одиночної, б – подвійної

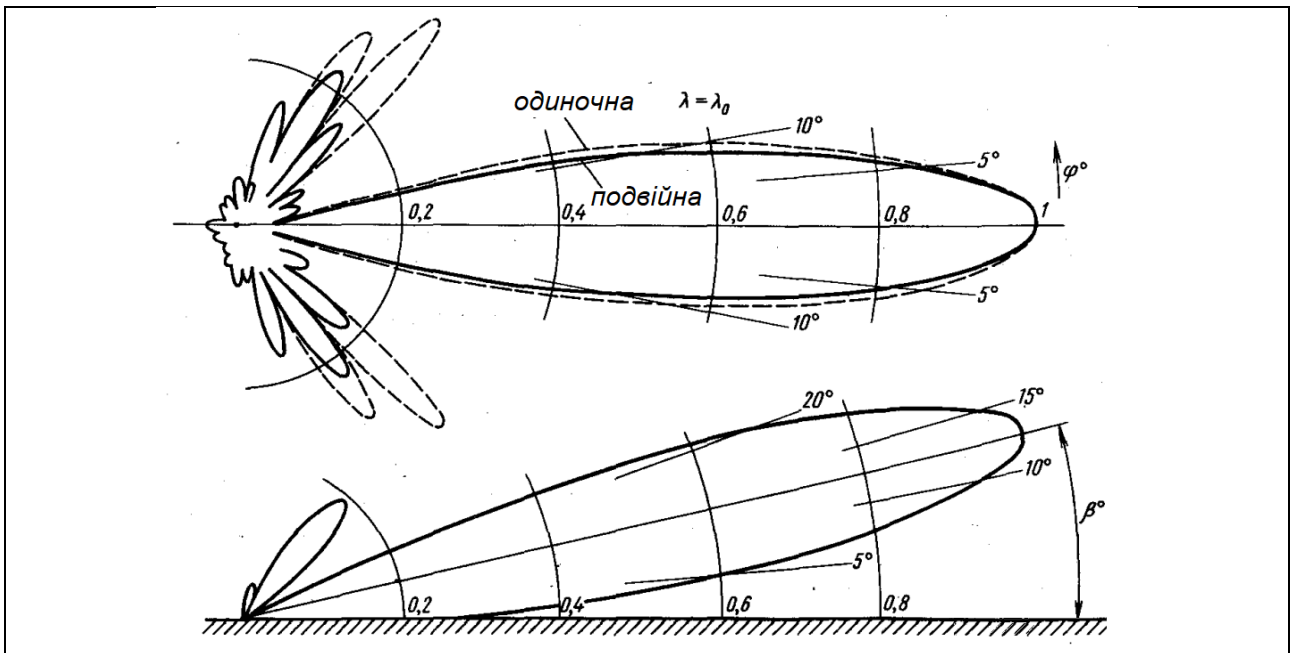


Рис. 7.6.17. Горизонтальний (вгорі) та вертикальний перерізи діаграми спрямованості горизонтальної ромбічної антени

Ромбічна горизонтальна антена (див. вище п. 7.5.3, рис. 7.5.12 в та рис. 7.5.14) являє

собою, по суті, двопровідну лінію, виконану у формі ромба (рис. 7.6.16 а) і підняту над земною поверхнею. До одного з гострих кутів приєднується джерело сигналу, до протилежного – опір навантаження, що дорівнює хвильовому опору лінії. В результаті в провідниках забезпечується режим біжучої хвилі. Завдяки режиму біжучої хвилі такі антени можуть працювати в широкому діапазоні частот.

Недоліками ромбічних антен є порівняно малий к.с.д. і високий рівень бічних пелюсток, низький к.к.д та велика площа, зайнята такою антеною. Менший рівень бічних пелюсток забезпечує подвійна ромбічна антена (рис. 7.6.16 б). Порівняльний вигляд діаграми спрямованості одиночної та подвійної ромбічних антен представлений на рис. 7.6.17.

Найпростіша *логоперіодична антена* (див. вище п. 7.5.4, рис. 7.5.17 а-б, рис. 7.5.20 а-в) показана на рис. 7.6.18. Вона характеризується широким робочим діапазоном ($\kappa = f_{\max}/f_{\min} = 10$) за рахунок того, що на кожній довжині хвилі приймання сигналу забезпечується лише частиною антени (робочою областю, або активною зоною). При зміні довжини хвилі робоча область зсувається вздовж антени, не змінюючи своїх відносних розмірів (виражених у довжинах хвилі). Робоча область складається з трьох елементів (вібраторів), які можна розглядати як спрощений варіант антени типу «хвильовий канал». Поля всіх трьох вібраторів складаються у фазі в напрямку найкоротшого з них, формуючи відповідну діаграму спрямованості (рис. 7.6.19).

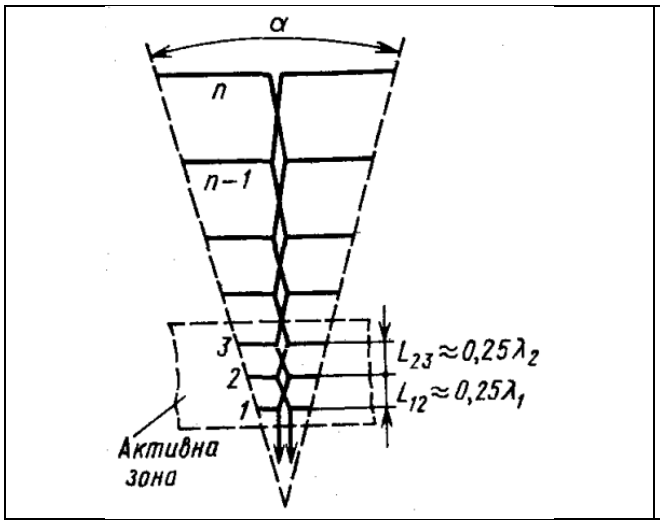


Рис. 7.6.18. Схематичне зображення логоперіодичної антени

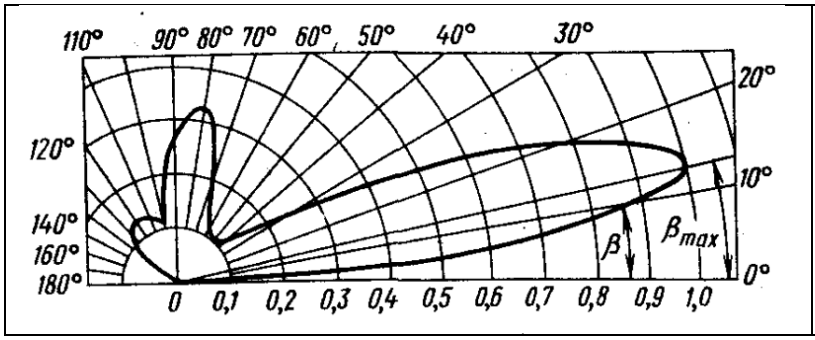


Рис. 7.6.19. Діаграма спрямованості логоперіодичної антени (вертикальний переріз)

Для логоперіодичних антен к.к.д. близький до 100%.

Недоліком логоперіодичних антен (якщо хочете, прямим наслідком їхніх переваг) є те, що при роботі на заданій довжині хвилі використовується лише незначна частина антени. Внаслідок цього к.с.д. такої антени порівняно невеликий.

В діапазоні коротких хвиль використовуються також антени біжучої хвилі, детально розглянуті в попередньому пункті.

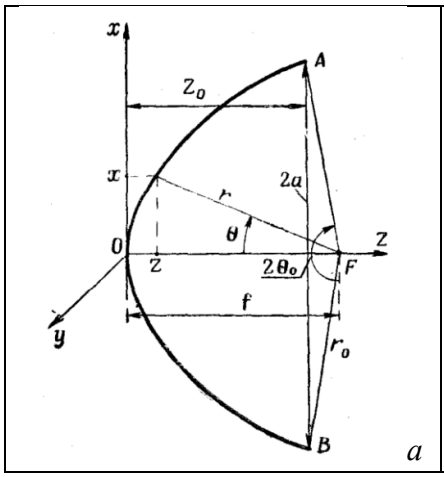
7.6.4. Квазіоптичні антени діапазону ультракоротких хвиль

До ультракороткохвильового діапазону відносять хвилі з довжинами від 1 мм до 10 м,

тобто міліметрові, сантиметрові, дециметрові та метрові хвилі (див. вище п. В.2). В цьому діапазоні розміри антен виявляються (значно) меншими за довжину хвилі, тому для приймання та передавання сигналів використовують апертурні антени (див. вище п. 7.5.5) – зокрема, антени квазіоптичного типу (дзеркальні й лінзові), а також рупорні. Використовуються також щілинні антени (див. вище п. 7.5.2), спіральні антени (п. 7.5.4), інколи – антенні ґратки (як у короткохвильовому діапазоні; див. вище пп. 7.5.3, 6.5.3).

Ідея роботи *дзеркальних антен* полягає в тому, що діаграма спрямованості майже ізотропного випромінювача покращується за рахунок відбиття від дзеркала спеціальної форми. Серед дзеркальних антен найбільш поширені *параболічні антени* (точніше, антени з відбивачем у формі параболоїда обертання – рис. 7.5.24 в та 7.5.25, рис. 7.6.19 а). Принцип їхньої роботи такий самий, як у світлового прожектора: при розміщенні випромінювача у фокусі параболоїда отримаємо пучок паралельних променів (рис. 7.6.19 б).

Типова форма діаграми спрямованості параболічної антени подана на рис. 7.6.20.



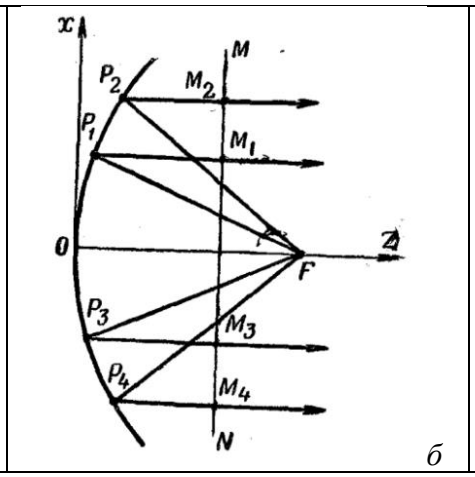


Рис. 7.6.19. Схематичне зображення параболічного дзеркала: а – параметри параболоїда обертання; б – відбиття променів від параболічного дзеркала

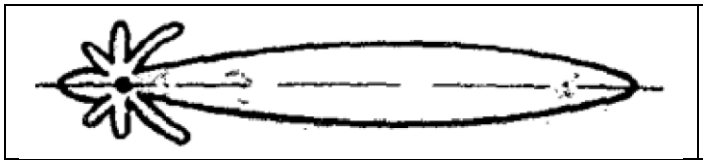


Рис. 7.6.20. Діаграма спрямованості параболічної антени

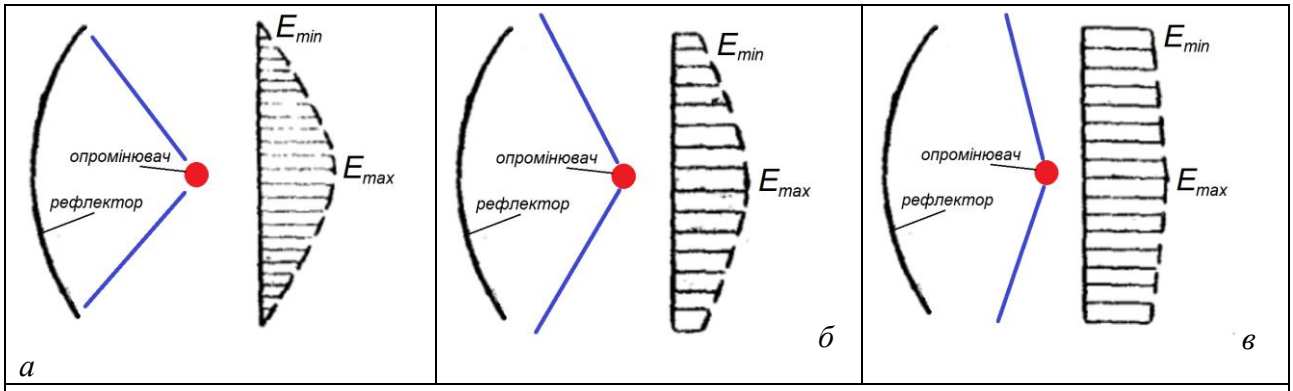


Рис. 7.6.21. Залежність розподілу поля в антені (рисунки праворуч) від розкриву діаграми спрямованості опромінювача (рисунки ліворуч): а – малий розкрив; б – оптимальний розкрив; в – надмірний розкрив

Форма діаграми спрямованості параболічної антени залежить від її розміру, спрямованих властивостей та розташування опромінювача, геометричних параметрів дзеркала, розподілу амплітуд і фаз по поверхні дзеркала, точності виконання профілю дзеркала. Опромінювач розміщується у фокусі (рис. 7.6.19 б) або у фокальній площині дзеркала. Відхилення від ідеальної схеми (зокрема, скінченні розміри відбивача, ненульові розміри опромінювача, інтерференційні ефекти тощо) призводять до деякого розширення головної пелюстки та зростання

рівня бічних пелюсток. Опромінювач створює область тіні для променів, відбитих від антени, частково ці промені на ньому розсіюються.

Для заданих розмірів дзеркала існує оптимальна схема його опромінення. Справді, при великому розкриві діаграми спрямованості опромінювача значна частка енергії не потрапляє на дзеркало, в результаті чого к.к.д. антени зменшується (рис. 7.6.21 в). При малому розкриві частина відбивача не бере участі у формуванні випромінювання антени (рис. 7.6.21 а). В оптимальному випадку, як показують розрахунки, напруженість поля на краю антени складає 30% від її максимального значення (рис. 7.6.21 б).

Ширина смуги пропускання параболічної антени визначається частотними властивостями опромінювача. При зміні довжини хвилі змінюється вхідний опір і діаграма спрямованості опромінювача, а через це змінюються властивості антени загалом – діаграма спрямованості та к.с.д.

Якщо опромінювач знаходиться у фокусі або близько до нього, він опиняється в максимумі потоку відбитої від антени потужності сигналу. Частину цієї потужності він приймає, в результаті чого у фідері порушується режим біжучої хвилі, або, іншими словами, порушуються умови узгодження фідера з антеною. Цей ефект призводить до зміни вхідного опору випромінювача та зменшення робочого діапазону частот.

Ідея роботи *лінзових антен* та сама, що й у дзеркальних антен – поліпшення діаграми спрямованості майже ізотропного випромінювача за допомогою деякого додаткового пристрою. Але тепер роль такого пристрою відіграє не дзеркало, а лінза (рисунки 7.5.30 а-б, 7.5.31 та 7.6.22).

Принцип дії лінз радіодіапазону такий самий, як у оптичних лінз. Вони виготовляються з діелектрика з відносною діелектричною проникністю ε (підкреслимо, що йдеться при високочастотну діелектричну проникність). Відповідно показник заломлення матеріалу лінзи $n = \sqrt{\varepsilon}$ (враховано, що у діелектрика зазвичай $\mu = 1$). В результаті за рахунок заломлення пучок променів, що розходяться, перетворюється на паралельний пучок (рис. 7.6.22). Сферична лінза перетворює хвилі зі сферичними фронтами на плоскі, циліндрична лінза робить таке саме перетворення хвиль із циліндричними фронтами.

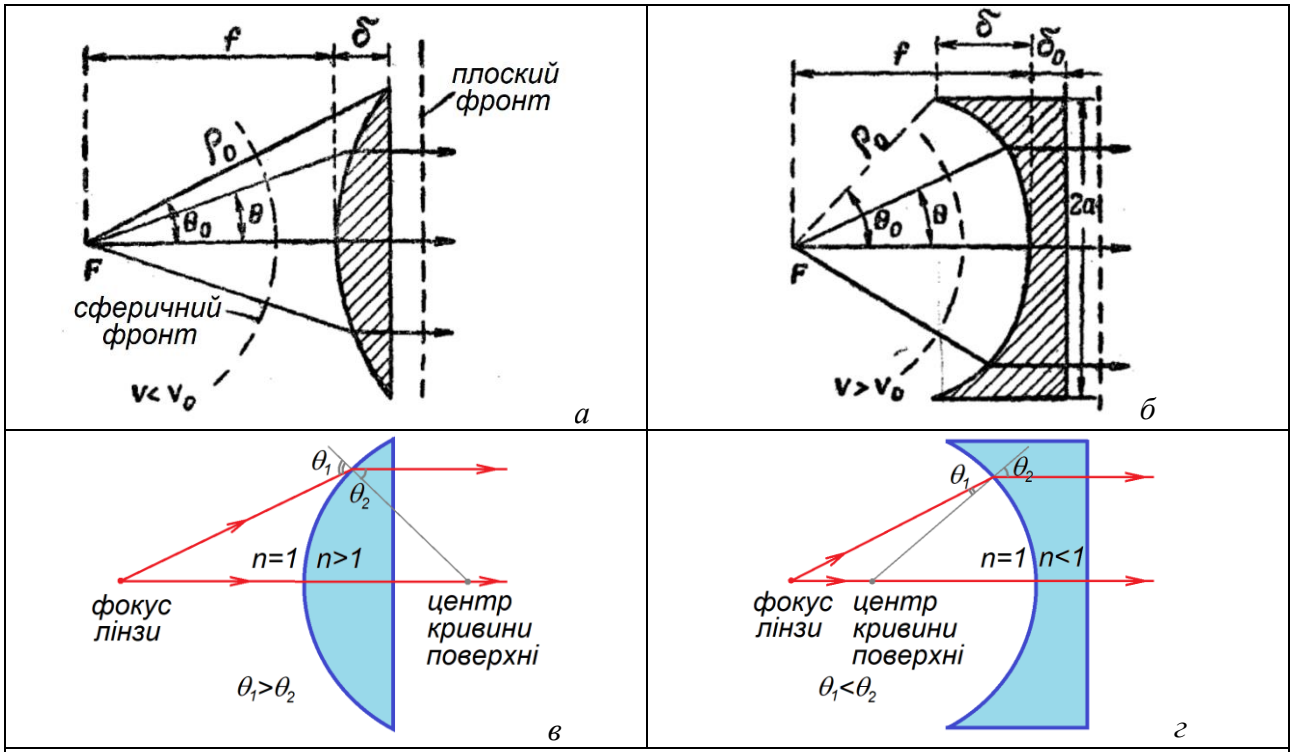


Рис. 7.6.22. Схематичні зображення сповільнювальної (а) та прискорювальної (б) лінз та схеми формування ними паралельного пучка променів (в-г)

В залежності від значення показника заломлення лінзи поділяють на сповільнювальні з $n > 1$ та прискорювальні з $n < 1$ (нагадаємо, що $n = c/\upsilon_{ph}$, де υ_{ph} – фазова швидкість хвиль у відповідному середовищі). Таким чином, у сповільнювальних лінзах фазова швидкість електромагнітних хвиль порівняно з вакуумом зменшується, а в прискорювальних – зростає. В результаті обидві лінзи є збиральними.

Для сповільнювальної ($n > 1$) лінзи з точковим опромінювачем її освітлена поверхня повинна мати форму гіперболоїда обертанння, а опромінювач розміщується у фокусі або на фокальній поверхні. Якщо опромінювач лінійний, потрібна лінза з поверхнею у формі гіперболічного циліндра (рис. 7.6.23).

Відповідно для прискорювальної ($n < 1$) лінзи її освітлена поверхня має бути еліпсоїдом обертання (для точкового опромінювача) або еліптичним циліндром (для лінійного). Опромінювач розміщується у фокусі або у фокальній площині.

Для зменшення розмірів лінзи показник заломлення її матеріалу повинен бути якомога більш відмінним від одиниці. Для сповільнювальних лінз звичайно $n = 1.3 \div 1.6$, для прискорювальних – $n = 0.5 \div 0.7$.

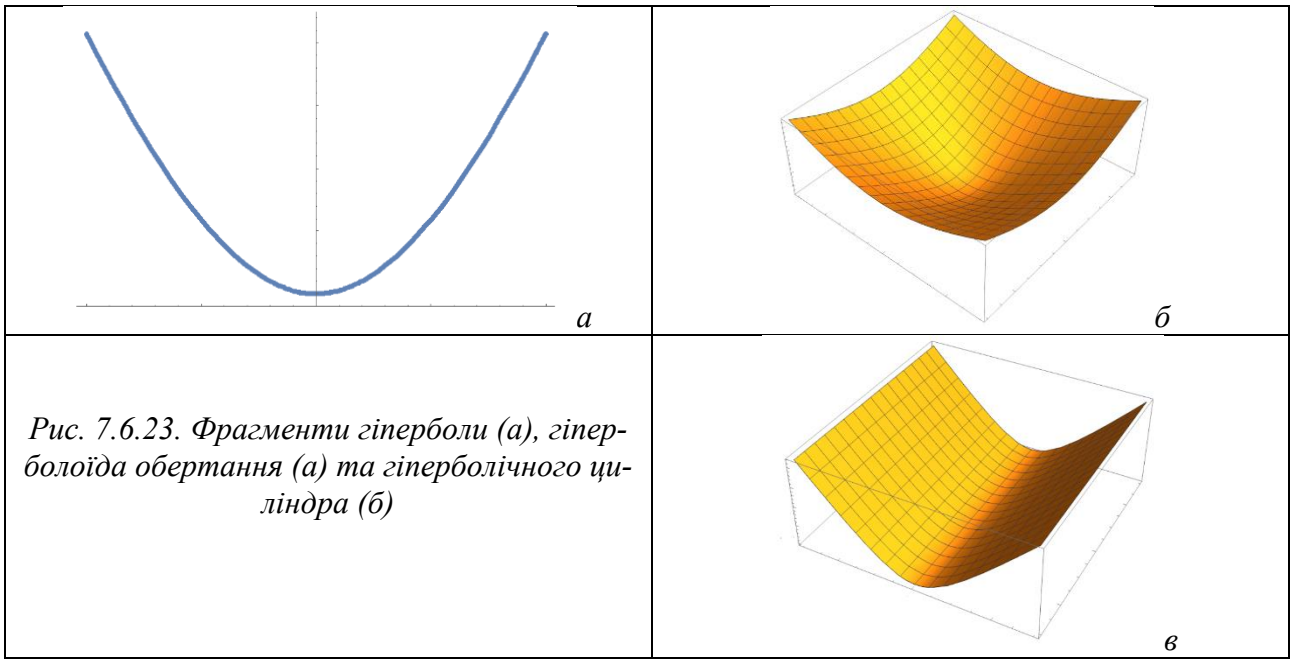


Рис. 7.6.23. Фрагменти гіперболи (а), гіперболоїда обертання (а) та гіперболічного циліндра (б)

Сповільнювальні лінзи невеликих розмірів звичайно виготовляють із високочастотних діелектриків (полістирол, фторопласт, тефлон), для лінз великих розмірів використовується металодіелектрик (рис. 7.6.24). Металодіелектрик – це метаматеріал[10], що являє собою просторову ґратку з ізольованих металевих частинок (їхні розміри мають бути значно менші за довжину хвилі), уміщену в легкий діелектрик з $\varepsilon \approx 1$ (найчастіше використовують пінистий полістирол). Лінзи такого типу відомі як метало-діелектричні лінзи Мікаеляна.

Металодіелектрик можна розглядати як модель звичайного діелектрика, де металеві частинки відіграють роль молекул, що поляризуються. Справді, в зовнішньому електричному полі металеві частинки поляризуються, породжуючи електричне поле, протилежне за напрямком до зовнішнього. Це й приводить до відповідного збільшення діелектричної проникності ($\vec{D} = \varepsilon\vec{E}$, тому при фіксованому $\vec{D}$, яке визначається зовнішнім полем, зменшення $\vec{E}$ в матеріалі означає відповідне збільшення ε).

[10] Метаматеріали – композити, які мають властивості, що не зустрічаються у природі (наприклад, водночас від'ємні діелектричну й магнітну проникності). Це штучно сформовані і особливим чином структуровані середовища, електромагнітні властивості яких відмінні від властивостей компонентів, з яких вони складаються.

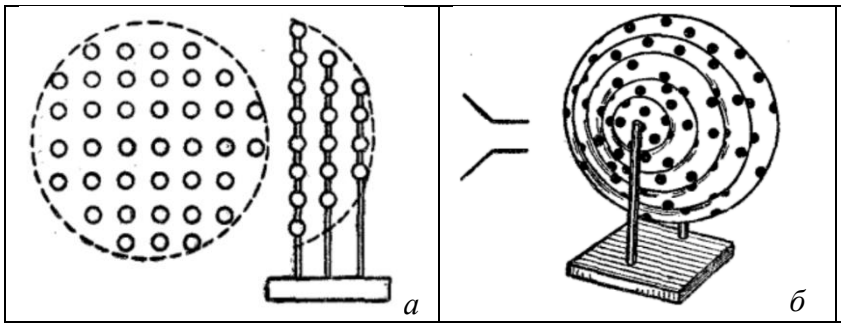


Рис. 7.6.24. Макети лінз із металодіелектрика: а – кулькова лінза; б – лінза з діелектричних пластин з дисками, нанесеними провідною фарбою

Прискорювальні лінзи повинні мати показник заломлення $n < 1$. Такою властивістю володіє хіба що газова плазма (див. нижче розділ 8), але використати її для створення лінзи дуже складно. Тому проблема розв'язується обхідним шляхом.

З електродинаміки відомо, що фазова швидкість хвиль у хвилеводі завжди більша за швидкість світла (див. вище п. 6.7). Зокрема, для основної моди прямокутного хвилеводу

$$\upsilon_{ph} = \frac{c}{\sqrt{1-(\lambda/2a)^2}}, \qquad (7.6.12)$$

де a – більший поперечний розмір хвилеводу. Таким чином, прискорювальну лінзу можна сконструювати з ділянок хвилеводів (рис. 7.25). Така металопластинчаста лінза являє собою набір профільованих металевих пластин, розташованих перпендикулярно до вектора електричного поля падаючої хвилі на віддалі *a* одна від одної. Кожна пара сусідніх пластин утворює ділянку прямокутного хвилеводу, в якому поширюється хвиля основного типу.

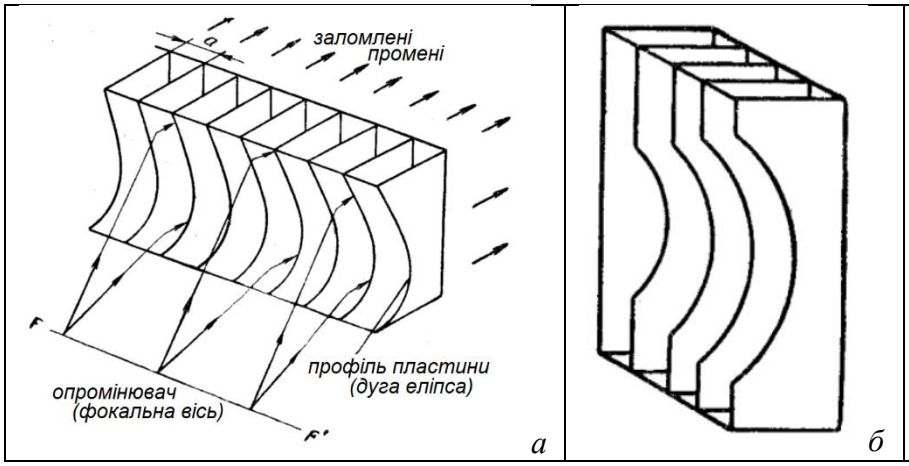


Рис. 7.6.25. Схематичні зображення циліндричної (а) та сферичної (б) прискорювальних лінз

7.6.5. Хвилеводні та рупорні випромінювачі

У дециметровому та сантиметровому діапазоні для передачі сигналів широко використовують прямокутні та циліндричні хвилеводи. Відкритий кінець такого хвилеводу може служити випромінювачем радіохвиль (*хвилеводний випромінювач*). Точніше, хвиля, що поширюється в такому хвилеводі, частково випромінюється, а частково відбивається від відкритого кінця, причому коефіцієнт біжучої хвилі в хвилеводі не перевищує 0.5. Схема такого випромінювача, перерізи діаграми спрямованості та розподіл електричного поля вздовж стінки прямокутного хвилеводу показані на рис. 7.6.26. Максимальний к.с.д. такого випромінювача звичайно складає кілька одиниць.

Таким чином, загалом параметри відкритого хвилеводу як випромінювальної антени далекі від бажаних.

Недоліки хвилеводних випромінювачів значною мірою подолані в *рупорних антенах* (див. рисунки 7.5.27 – 7.5.29, 7.6.27). Такі антени утворюються шляхом плавного розширення відповідних хвилеводів, чим і забезпечується їхнє узгодження з вільним простором.

Розширення хвилеводу в загальному випадку породжує відбиту хвилю та моди вищих порядків. Але якщо розширення плавне (в масштабі довжини хвилі), то ці ефекти незначні.

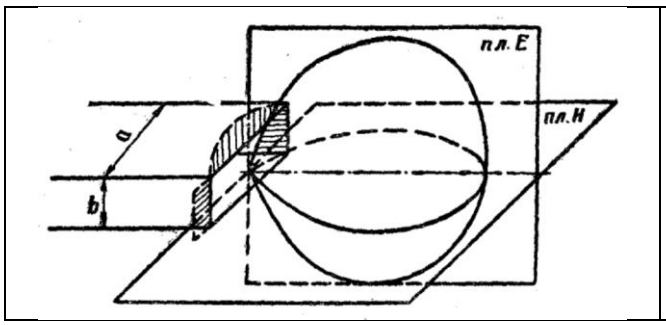


Рис. 7.6.26. Хвилеводний випромінювач на основі прямокутного хвилеводу та його діаграма спрямованості в головних площинах (показано також розподіл електричного поля вздовж стінок хвилеводу)

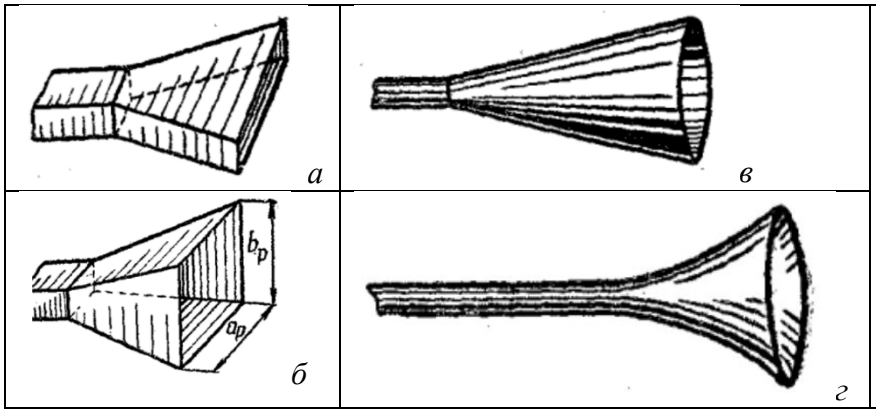


Рис. 7.6.27. Рупорні антени: а – секторіальна; б – пірамідальна; в – конічна; г – експоненціальна

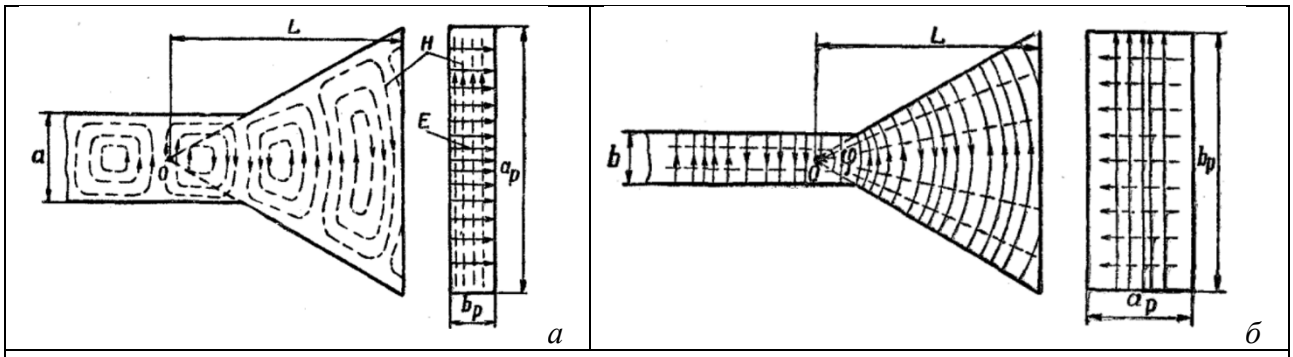


Рис. 7.6.28. Поля в секторіальних рупорних антенах: а – Н-площинна; б – Е-площинна (ліворуч показаний горизонтальний переріз антени, праворуч – вертикальний)

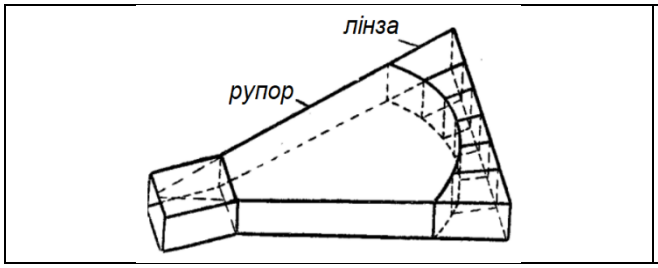


Рис. 7.6.29. Рупорно-лінзова антена

У *секторіальних рупорних антенах* (рис. 7.6.27 а) розширюється тільки одна грань хвилеводу: в Н-площинних – широка, в Е-площинних – вузька (рис. 7.6.28). В результаті отримуємо випромінювання відповідно з вертикальною та горизонтальною площиною поляризації. Порівняно з хвилеводними випромінювачами секторіальні рупори мають звужені діаграми спрямованості в горизонтальній площині. Фронти хвиль, випромінюваних секторіальними антенами, мають форму, наближену до циліндричної. Для отримання на виході плоскої хвилі використовують рупорно-лінзові антени (рис. 7.6.29). Це дозволяє зменшити розмір рупора в десятки разів при тій самій діаграмі спрямованості.

Пірамідальна рупорна антена (рис. 7.6.27 б) звужує діаграму спрямованості як у вертикальній, так і в горизонтальній площинах. Форма фронту на невеликих віддалях від антени

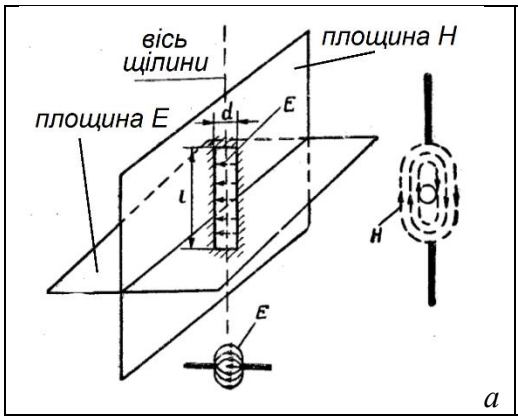
близька до еліпсоїду, її також можна поліпшити за допомогою рупорно-лінзової антени.

Конічні рупорні антени (рис. 7.6.27 в) використовуються для випромінювання циркулярно поляризованих хвиль із циліндричних хвилеводів. *Експоненціальні антени* (рис. 7.6.27 г) порівняно з конічним мають менший коефіцієнт відбиття.

Перевагами рупорних антен є простота конструкції та широкий частотний діапазон. Порівняно з хвилеводними випромінювачами вони мають більший к.с.д. і, як уже зазначалося, краще узгодження з фідером.

7.6.6. Щілинні випромінювачі

Щілинний випромінювач являє собою вузьку щілину, прорізану в металевій поверхні (рис. 7.6.30 а). Якщо на цій поверхні перпендикулярно до щілини протікає змінний поверхневий струм, він замикається у формі струму зміщення, тобто в щілині (й навколо неї) виникає змінне електричне поле (його силові лінії показані в нижній частині рис. 7.6.30 а). Це змінне електричне поле, в свою чергу, породжує змінне магнітне поле (його силові лінії показані в правій частині рис. 7.6.30 а). В результаті формується поле випромінювання.



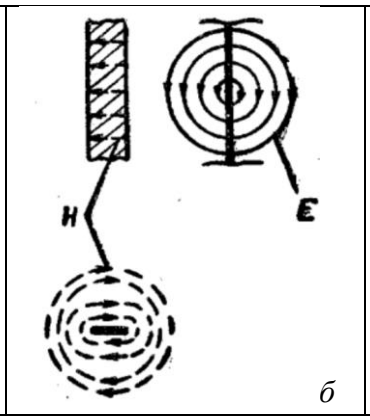


Рис. 7.6.30. Схематичне зображення елементарного щілинного випромінювача (а) та відповідного доповнювального вібратора (б)

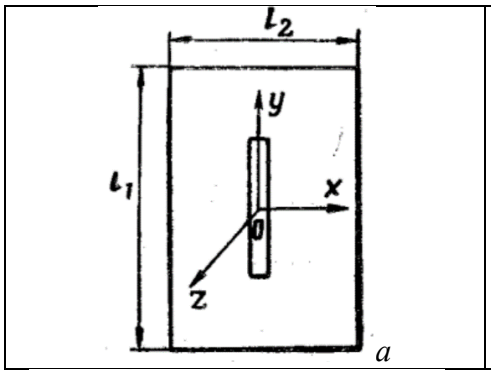
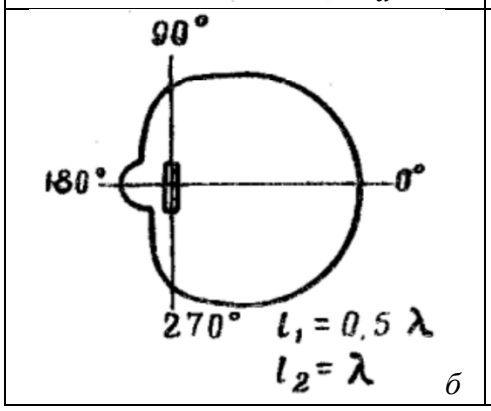
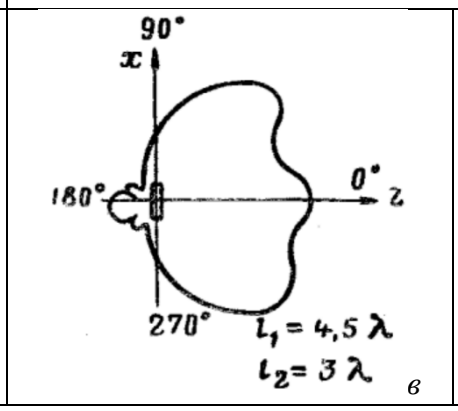
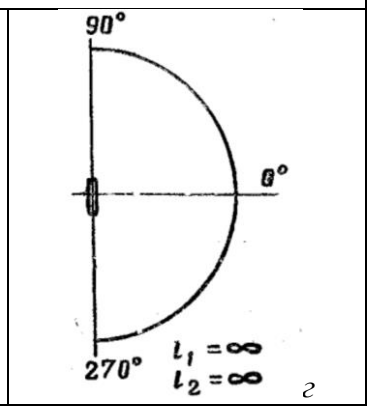


Рис. 7.6.31. Вплив розмірів екрану на діаграму спрямованості півхвильового щілинного випромінювача в площині E: а – схема випромінювача, б-г – зміна діаграми спрямованості при зміні розмірів екрану







Відповідно до принципу двоїстості, щілина в нескінченному екрані формує таке саме поле випромінювання, як електричний вібратор, що має форму стрічки з розмірами щілини й знаходиться у вільному просторі (рис. 7.6.30 б). Такий вібратор називається доповнювальним

до даної щілини. Електричне поле доповнювального вібратора має таку саму форму, як магнітне поле щілини, а магнітне поле доповнювального вібратора подібне до електричного поля щілини, тобто головні площини щілини й доповнювального вібратора міняються місцями.

Діаграма спрямованості щілини така сама, як у доповнювального вібратора.

На практиці довжина щілини звичайно сумірна з довжиною хвилі. Найчастіше використовуються резонансні щілини, довжина яких дорівнює половині довжини хвилі. Поле випромінювання півхвильової щілини подібне до поля півхвильового вібратора.

Щілини часто вирізають у зовнішньому провіднику коаксіального фідера, стінці хвилеводу або резонатора тощо. Іншими словами, реально щілина знаходиться не в нескінченній провідній площині, а в металевій поверхні скінченних розмірів. Таким чином, на випромінювальні властивості щілини впливає не тільки її довжина (й ширина), але також розміри та форма екрану, в якому вона прорізана (рис. 7.6.31). При збільшенні ширини щілини її резонансна довжина зменшується, а активна складова вхідного опору дещо зростає.

7.6.7. Спіральні антени

Ми вже згадували, що джерелом циркулярно поляризованих хвиль може бути конічний рупор, підключений до циліндричного хвилеводу. Якщо ж в ролі фідера виступає коаксіальний кабель, циркулярно поляризоване випромінювання можна збудити за допомогою спіральної антени (рис. 7.5.22 б).

Схематичне зображення циліндричної спіральної антени подане на рис. 7.6.32 а. Провідник антени має форму гвинтової лінії, а довжина одного витка складає λ. З урахуванням відстані між сусідніми витками діаметр циліндра дорівнює 0.31λ. Кількість витків повинна бути не менше трьох, зі збільшенням їхньої кількості еліптична поляризація випромінювання все більше наближається до циркулярної. Проста дротяна обмотка зі вказаними параметрами випромінює вперед та назад уздовж її поздовжньої осі. За наявності відбивача (рис. 7.6.32 б, в) випромінювання відходить лише в один бік і, відповідно, підсилюється.

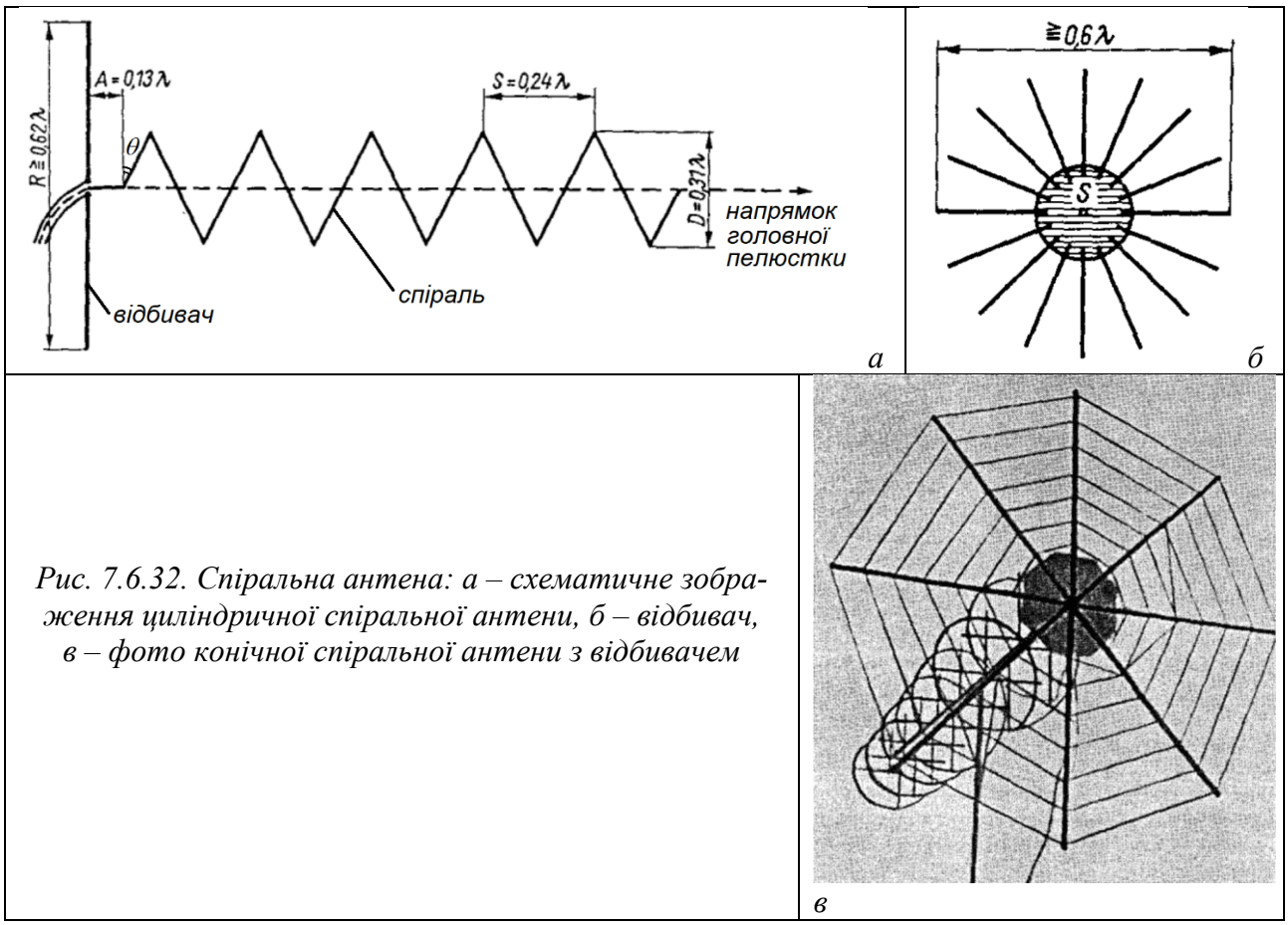


Рис. 7.6.32. Спіральна антена: а – схематичне зображення циліндричної спіральної антени, б – відбивач, в – фото конічної спіральної антени з відбивачем

Кут підняття спіралі θ – це кут між дотичною до спіралі й площиною, перпендикулярною до її осі. Звичайно він лежить у діапазоні від 6º до 24º, але найчастіше обирається значення 14º. Цей кут визначає крок спіралі:

$$S = \pi D \mathrm{tg}\theta = \lambda \sin\theta \qquad (7.6.13)$$

(позначення зрозумілі з рис. 7.6.32 а). Зазначимо, що співвідношення $S = 0.24\lambda$, подане на рис. 7.6.32 а, відповідає значенню $\theta = 14°$.

Підсилення та фокусування променю спіральної антени залежать від кількості витків, кроку витків та довжини провідника. Підсилення приблизно пропорційне кількості витків спіралі.

Контрольні питання до розділу 7.6

1. Які передавальні антени використовуються в діапазоні довгих та середніх хвиль?
2. Чому конструкція передавальних та приймальних антен може відрізнятися?
3. Які типи приймальних антен використовуються в діапазоні довгих та середніх хвиль?
4. Опишіть роботу антени біжучої хвилі.
5. Назвіть основні типи антен, які застосовуються в діапазоні коротких хвиль.
6. Опишіть конструкцію синфазної антени. Як забезпечується її випромінювання тільки в один бік?
7. За рахунок чого логоперіодична антена працює в широкому діапазоні довжин хвиль?
8. В чому полягає принцип роботи квазіоптичних антен короткохвильового діапазону?
9. Опишіть, як виготовляються сповільнювальні та прискорювальні лінзи.
10. Чому рупорні антени мають кращі параметри, ніж хвилеводні випромінювачі?
11. Опишіть принцип роботи щілинної антени.
12. Навіщо на кінці спіральної антени встановлюється відбивач?

Задачі до підрозділу 7.6

7.6.1. Під яким кутом до горизонту розташований максимум характеристики спрямованості несиметричного вертикального вібратора? Розрахувати значення цього кута, якщо довжина антени – 5 м, а частота випромінювання – 10 МГц.

7.6.2. Антена біжучої хвилі складається з 10 випромінювачів. Під яким кутом до горизонту розташований максимум її характеристики спрямованості? Висота антени над поверхнею Землі – 20 м, віддаль між сусідніми випромінювачами – 1 м, довжина хвилі випромінювання – 2.5 м, c/v_{ph}=0.3. Поверхню Землі вважати ідеально провідною.

Розділ 8. Поширення радіохвиль у приземному просторі

Якщо радіосигнал випромінюється передавальною антеною в ефір, а потім приймається приймальною антеною, то він має пройти певний шлях у приземному просторі (рис. 8.1.1). На цьому шляху він зазнає впливу середовища (нейтрального або частково іонізованого газу, який до того ж є просторово неоднорідним), а також впливу земної поверхні (зокрема, її кривини та рельєфу, наявності водоймищ, рослинного покрову тощо). В різних частотних діапазонах поширення радіохвиль у приземному просторі може бути суттєво відмінним.

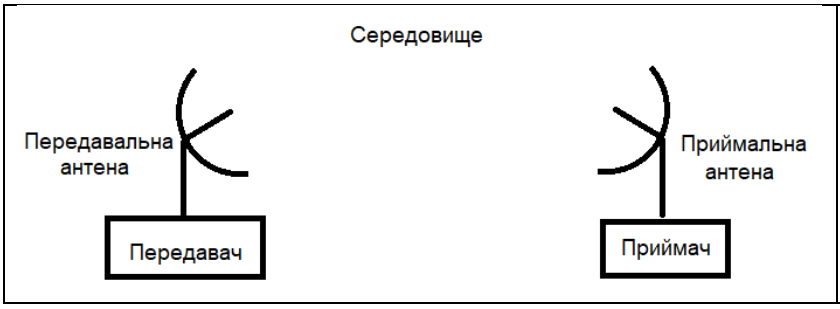


Рис. 8.1.1. Схема тракту передавання – приймання сигналу

8.1. Деякі питання теорії електромагнітних хвиль у середовищах

Перш ніж безпосередньо розглядати поширення радіохвиль у приземному просторі, розглянемо спершу деякі більш загальні питання: отримаємо дисперсійне рівняння для електромагнітних хвиль у середовищі (в загальному випадку – провідному), отримаємо найпростіші формули для діелектричної проникності іонізованого та нейтрального газу, а також оцінимо розміри й форму області, яка впливає на поширення сигналу вздовж заданої траси між передавальною і приймальною антенами.

8.1.1. Дисперсійне рівняння для електромагнітних хвиль у середовищі

Ми вже отримували хвильове рівняння для електромагнітних хвиль у вакуумі (див. вище пп. 6.8.3 та 7.1.1).

При поширенні електромагнітних хвиль у приземному просторі принциповим є врахування (відносної) діелектричної проникності нейтрального та іонізованого газу (останній ми називатимемо газовою плазмою, або просто плазмою). Тому ми спершу отримаємо рівняння для електромагнітних хвиль у середовищі. Можна вважати, що магнітна проникність газу і плазми дорівнює одиниці, тому нам буде тепер зручніше працювати з рівнянням для електричного поля.

Отже, нагадаємо, що перші два рівняння Максвелла в системі одиниць Гаусса мають вигляд (6.8.17) - (6.8.18):

$$\nabla \times \vec{H} = \frac{4\pi}{c}\vec{j} + \frac{1}{c}\frac{\partial \vec{D}}{\partial t}, \qquad \nabla \times \vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t}.$$

Їх слід доповнити матеріальними рівняннями, які пов’язують між собою напруженості $\vec{E}$ і $\vec{H}$ та індукції $\vec{D}$ і $\vec{B}$:

$$\vec{D} = \varepsilon \vec{E}, \qquad \vec{B} = \mu \vec{H}, \tag{8.1.1}$$

де ε і μ – відповідно (відносні) діелектрична та магнітна проникності середовища.

Ми обмежимося розглядом так званих ізотропних середовищ, у яких ε і μ можна вважати скалярами. В більш загальному випадку це будуть тензори.

Поняття діелектричної та магнітної проникностей можна строго ввести лише для полів, що мають форму гармонічних функцій часу й координат. Тому надалі вважатимемо, що електричне та магнітне поля залежать від часу та координат за законом $\exp\left(i\omega t - i\vec{k}\vec{r}\right)$ і врахуємо, що в загальному випадку

$$\varepsilon = \varepsilon\left(\omega, \vec{k}\right), \quad \mu = \mu\left(\omega, \vec{k}\right). \tag{8.1.2}$$

Як ми з'ясували раніше (див. п. 6.2.4), причиною виникнення дисперсії в середовищах є наявність власних частот та характерних розмірів. В електродинаміці наявність власних частот означає, що середовище має пам'ять (індукція в даний момент часу залежить від напруженості не тільки в цей самий, але й у попередні моменти часу). В результаті ε і μ виявляються функціями частоти. Відповідно наявність власного розміру означає, що індукція в даній точці простору залежить від напруженості не тільки в даній, але й у сусідніх точках простору (в певному сенсі характерний розмір середовища визначає просторовий масштаб такого нелокального впливу). В результаті ε і μ виявляються функціями хвильового числа.

Підставивши залежності для полів у формі $\exp\left(i\omega t - i\vec{k}\vec{r}\right)$ до рівнянь (6.8.17) - (6.8.18), перепишемо їх у формі

$$-i\left[\vec{k} \times \vec{H}\right] = \frac{4\pi}{c}\vec{j} + \frac{i\omega\varepsilon}{c}\vec{E}, \qquad -i\left[\vec{k} \times \vec{E}\right] = -\frac{i\omega\mu}{c}\vec{H}. \tag{8.1.3}$$

Оскільки нас зараз цікавить задача про поширення електромагнітних хвиль у середовищі, вважатимемо, що струм $\vec{j}$ пов'язаний з провідністю середовища та зумовлений полем електромагнітної хвилі. Для спрощення вважатимемо справедливим закон Ома в диференціальній формі:

$$\vec{j} = \sigma\vec{E} \tag{8.1.4}$$

(тут $\sigma = \sigma' - i\sigma''$ – високочастотна провідність середовища, яка може мати як активну σ', так і реактивну σ'' складову). З урахуванням (8.1.4) права частина першого з рівнянь (8.1.3) набуває вигляду

$$\frac{4\pi}{c}\vec{j} + \frac{i\omega\varepsilon}{c}\vec{E} = \frac{4\pi\sigma}{c}\vec{E} + \frac{i\omega\varepsilon}{c}\vec{E} = \frac{i\omega}{c}\left(\varepsilon - i\frac{4\pi\sigma}{\omega}\right)\vec{E} = \frac{i\omega\tilde{\varepsilon}}{c}\vec{E}. \tag{8.1.5}$$

Тим самим ми ввели комплексну діелектричну проникність

$$\tilde{\varepsilon} = \varepsilon - i\frac{4\pi\sigma}{\omega} = \varepsilon' - i\varepsilon''. \tag{8.1.6}$$

Тепер замінимо праву частину першого з рівнянь (8.1.3) за допомогою співвідношення (8.1.5) та виключимо з нього магнітне поле за допомогою другого з рівнянь (8.1.3). Маємо:

$$-\left[\vec{k} \times \left[\vec{k} \times \vec{E}\right]\right] = \frac{\omega^2\tilde{\varepsilon}\mu}{c^2}\vec{E}. \tag{8.1.7}$$

Розпишемо подвійний векторний добуток у лівій частині:

$$\vec{k}\left(\vec{k}\vec{E}\right) - k^2\vec{E} = -\frac{\omega^2\tilde{\varepsilon}\mu}{c^2}\vec{E}. \tag{8.1.8}$$

Нарешті, врахуємо, що в газі та в плазмі $\mu = 1$, і обмежимося розглядом поперечних хвиль, у яких $\left(\vec{k}\vec{E}\right) = 0$, після чого скорочуємо обидві частини на $\vec{E}$ і отримуємо шукане дисперсійне рівняння у формі

$$k^2 = \frac{\omega^2}{c^2}\tilde{\varepsilon} = k_0^2\tilde{\varepsilon}, \qquad k_0 \equiv \frac{\omega}{c}. \tag{8.1.9}$$

Дисперсійне рівняння можна розв'язувати щодо ω, вважаючи k заданою дійсною величиною (це відповідає так званій задачі з початковими умовами, коли ми задаємо початкове збурення й слідкуємо за його еволюцією в часі), а можна, навпаки, вважати заданою дійсною величиною ω і шукати з дисперсійного рівняння k (так звана задача з граничними умовами, коли на межі середовища задане джерело гармонічного сигналу із заданою частотою). Саме граничну задачу ми надалі й розглядатимемо. Тоді з дисперсійного рівняння (8.1.9) випливає, що

$k = \pm k_0 \sqrt{\tilde{\varepsilon}}$. (8.1.10).

Величина k виходить суто дійсною за виконання умов $\varepsilon' > 0$, $\varepsilon'' = 0$ (або $\sigma' = 0$). Порушення першої з цих умов означає, що середовище є непрозорим для хвиль із даною частотою (див. вище п. 6.2.2). Порушення другої умови за виконання першої означає, що хвильове число є комплексним,

$$k = \pm\left(k' - ik''\right). \qquad (8.1.12)$$

Вважаючи, що напрямок вектора $\vec{k}$ збігається з віссю x, отримаємо, що поля хвилі змінюються в просторі й часі за законом

$$\exp\left(i\omega t - ikx\right) = \exp\left[\left(i\omega t \mp ik'x\right) \mp k''x\right], \qquad (8.1.13)$$

тобто уявна частина хвильового числа описує згасання хвилі в напрямку її поширення. Це згасання обумовлене дисипацією (активною провідністю середовища).

Тепер ускладнимо модель. Будемо розглядати півпростір $x \geq 0$, заповнений розглянутим раніше середовищем, що межує з вакуумом (рис. 8.1.2). З вакууму (область $x < 0$) на межу середовища падає плоска гармонічна хвиля вигляду $\exp\left(i\omega_0 t - ik_{x0}x - ik_{z0}z\right)$, хвильовий вектор якої лежить у площині xz. З'ясуємо, коли ця хвиля буде породжувати заломлену хвилю в середовищі. Вважаємо, що остання має вигляд $\exp\left(i\omega t - ik_x x - ik_z z\right)$.

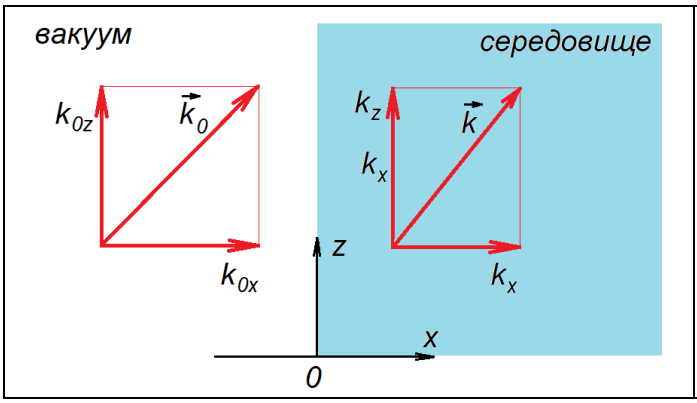


Рис. 8.1.2. Заломлення хвилі на межі вакуум-середовище

На межі двох середовищ (тобто при переході через площину $x = 0$) повинні зберігатися нормальна компонента індукції і тангенціальна компонента напруженості електричного поля. Цю умову можна задовольнити лише тоді, коли $\omega_0 = \omega$ і $k_{z0} = k_z$. Тоді компоненту хвильового вектора в середовищі знаходимо з дисперсійного рівняння (8.1.9), яке в даному разі набуває вигляду

$$k_x^2 + k_{z0}^2 = \frac{\omega_0^2}{c^2}\tilde{\varepsilon}, \qquad (8.1.14)$$

звідки

$$k_x = \pm\sqrt{\frac{\omega_0^2}{c^2}\tilde{\varepsilon} - k_{z0}^2}. \qquad (8.1.15)$$

Вважатимемо діелектричну проникність середовища $\tilde{\varepsilon}$ дійсною величиною. Тоді умова існування заломленої хвилі в середовищі має вигляд $k_x^2 > 0$, або $\omega_0^2\tilde{\varepsilon}/c^2 > k_{z0}^2$. За порушення цієї умови хвиля повністю відбиватиметься від середовища, яке буде для неї непрозорим (ефект повного внутрішнього відбиття). Ця умова більш жорстка, ніж записана раніше умова $\varepsilon' > 0$.

8.1.2. Діелектрична проникність плазми

Газова плазма – це повністю або частково іонізований газ, тобто в цілому електроней-

тральна суміш нейтральних атомів, електронів та іонів. Щоб іонізований газ вважався плазмою, він повинен існувати достатньо тривалий час, а розміри області, яку він займає, повинні бути достатньо великими.

Будемо вважати, що в першому наближенні нейтральні атоми не дають внеску в діелектричну проникність плазми, яка, відповідно до формули (8.1.6), зумовлена лише її провідністю. Ця провідність створюється електронами та іонами, які в зовнішньому змінному електричному полі здійснюють вимушені коливання. Але маса іонів принаймні на три порядки більша за масу електронів, тому й амплітуда їхніх коливань буде принаймні на три порядки меншою за амплітуду коливань іонів. Це означає, що ми з достатньою точністю можемо вважати іони нерухомими і брати до уваги лише рух електронів. Ми не враховуватимемо тепловий рух електронів, що відповідає моделі так званої холодної плазми. В цій моделі враховується тільки часова дисперсія, тобто $\tilde{\varepsilon}$ залежить тільки від частоти.

Електрони можуть зазнавати пружних зіткнень з важкими частинками – іонами та нейтральними атомами. Оскільки різниця їхніх мас дуже велика, то при таких зіткненнях енергія електронів практично не змінюється, а от напрямок їхньої швидкості змінюється випадковим чином. Результат можна інтерпретувати як втрату енергії спрямованого руху й перетворення її на теплову енергію відповідно до закону Джоуля - Ленца.

Таким чином, для знаходження діелектричної проникності плазми ми повинні знайти її провідність, зумовлену рухом електронів (та їхніми зіткненнями з важкими частинками). Для цього, в свою чергу, необхідно знайти густину струму електронів у зовнішньому змінному електричному полі, яка виражається через швидкість електронів. Останню знайдемо шляхом розв'язання рівняння руху.

Рівняння руху електрона в зовнішньому електричному полі може бути подане у вигляді

$$m\frac{d\vec{\upsilon}}{dt} = -e\vec{E} - m\vec{\upsilon}\nu, \qquad \vec{E}(t) = \vec{E}_m \exp(i\omega t), \tag{8.1.16}$$

m – маса електрона, e – елементарний заряд, ν – (середня) частота зіткнень електронів з важкими частинками (нейтральними атомами та іонами).

Вважаємо, що електричне поле паралельне осі x. Тоді, відповідно до рівняння (8.1.16), швидкість електронів також паралельна осі x. Таким чином, спроєктувавши рівняння (8.1.16) на вісь x, перейдемо від векторних величин $\vec{E}$ і $\vec{\upsilon}$ до скалярних. Надалі задачу вважаємо одновимірною.

Оскільки положення електрона з часом змінюється, то повна похідна за часом набуває вигляду

$$\frac{d\upsilon[t.x(t)]}{dt} = \frac{\partial\upsilon}{\partial t} + \frac{\partial\upsilon}{\partial x}\frac{\partial x}{\partial t} = \frac{\partial\upsilon}{\partial t} + \upsilon\frac{\partial\upsilon}{\partial x}. \tag{8.1.17}$$

Але при невеликих амплітудах електричного поля другим доданком, квадратичним за υ, можна знехтувати поруч із першим, тому рівняння (8.1.16) можна записати у формі

$$m\frac{\partial\upsilon}{\partial t} + m\upsilon\nu = -eE. \tag{8.1.18}$$

У правій частині рівняння (8.1.18) стоїть гармонічна зовнішня сила. Будемо шукати його розв'язок у вигляді вимушених коливань з такою самою залежністю від часу. Тоді з рівняння (8.1.18) отримаємо комплексну амплітуду швидкості:

$$\upsilon_m = -\frac{eE_m}{m(i\omega + \nu)}. \tag{8.1.19}$$

Відповідно густина струму електронів складає величину

$$j = -en_e\upsilon = \frac{e^2 n_e E}{m(i\omega + \nu)} \tag{8.1.20}$$

(тут n_e – концентрація електронів), а провідність плазми –

$$\sigma = \frac{e^2 n_e}{m(i\omega+\nu)} = \frac{e^2 n_e(\nu - i\omega)}{m\left(\omega^2+\nu^2\right)} = \frac{e^2 n_e \nu}{m\left(\omega^2+\nu^2\right)} - i\frac{e^2 n_e \omega}{m\left(\omega^2+\nu^2\right)} \equiv \sigma' - i\sigma''. \qquad (8.1.21)$$

Як бачимо, активна провідність плазми визначається зіткненнями електронів з важкими частинками. При $\omega \gg \nu$ вплив зіткнень незначний, і реактивна частина провідності набуває вигляду

$$\sigma'' \approx \frac{e^2 n_e}{m\omega} \approx \sigma. \qquad (8.1.22)$$

Відповідно до аналізованої моделі, внесок до діелектричної проникності плазми дає лише її електронна провідність, тобто у формулі (8.1.6) можна вважати, що $\varepsilon = 1$, і після підстановки σ з (8.1.21) отримуємо, перепозначивши $\tilde{\varepsilon} \to \varepsilon$:

$$\varepsilon = 1 - i\frac{4\pi\sigma}{\omega} = 1 - i\frac{4\pi e^2 n_e}{m\omega(i\omega+\nu)} = 1 - \frac{\omega_p^2}{\omega(\omega - i\nu)}, \qquad (8.1.23)$$

де

$$\omega_p = \sqrt{\frac{4\pi e^2 n_e}{m}} \qquad (8.1.24)$$

– (електронна) плазмова, або (електронна) ленгмюрівська частота. Це частота коливань, які виникають у плазмі при локальному збуренні концентрації електронів. Формула (8.1.24) може бути подана у вигляді:

$$f_p\left[Гц\right] \equiv \frac{\omega_p}{2\pi} = 8978.66\sqrt{n_e\left[см^{-3}\right]} \approx 10^4\sqrt{n_e\left[см^{-3}\right]}. \qquad (8.1.25)$$

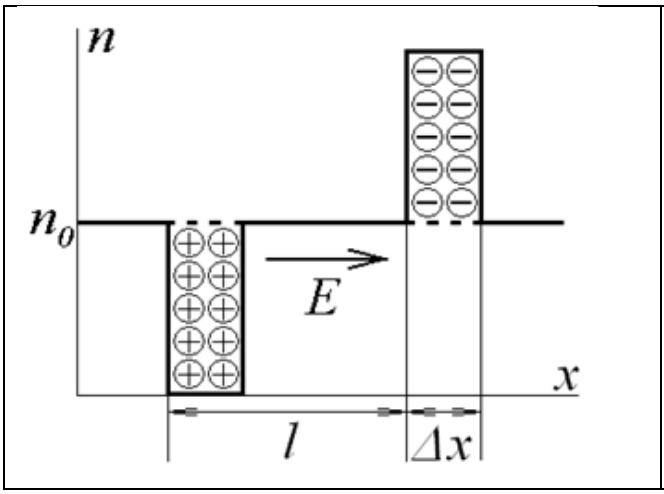


Рис. 8.1.3. До пояснення коливань електронів у плазмі

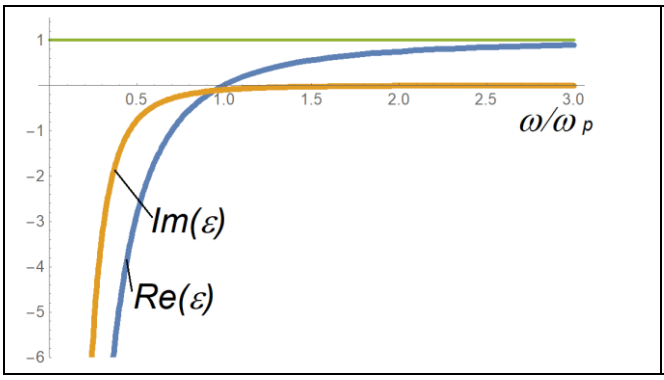


Рис. 8.1.4. Графік частотної залежності діелектричної проникності плазми ($\nu/\omega_p = 0.1$)

Характер ленгмюрівських коливань зрозумілий з рис. 8.1.3.

Припустимо, що ми відсунули шар електронів плазми завтовшки Δx на віддаль l. Це приводить до локального порушення електронейтральності і виникнення електричного поля, яке прагне повернути відхилені електрони на місце. Але в силу ненульової маси вони проскакують положення рівноваги і відхиляються в інший бік. Далі процес повторюється. Так і виникають плазмові, або ленгмюрівські коливання з частотою ω_p.

Графіки залежності дійсної та уявної частин діелектричної проникності від частоти подані на рис. 8.1.4.

За умови $\omega \gg \nu$ формулу (8.1.24) можна спростити:

$$\varepsilon \approx 1 - \frac{\omega_p^2(\omega + i\nu)}{\omega^3} = 1 - \frac{\omega_p^2}{\omega^2} - i\frac{\omega_p^2}{\omega^2}\frac{\nu}{\omega} \approx 1 - \frac{\omega_p^2}{\omega^2}. \qquad (8.1.26)$$

Бачимо, що при нехтуванні зіткненнями діелектрична проникність плазми завжди менша за одиницю, а на частотах $\omega < \omega_p$ стає від'ємною. Такі незвичайні властивості діелектричної проникності плазми пов'язані з тим, що вона є не діелектриком, а провідником. Якщо згадати рівняння Максвелла

$$\nabla \times \vec{H} = \frac{4\pi}{c}\vec{j} + \frac{1}{c}\frac{\partial \vec{E}}{\partial t},$$

то можна сказати, що перший доданок (одиниця) у виразі для ε відповідає вакуумному струму зміщення $(1/c)\partial\vec{E}/\partial t$, а другий ($-\omega_p^2/\omega^2$) – струму провідності $4\pi\vec{j}/c$. Один із цих струмів випереджає на чверть періоду коливання електричного поля, другий – навпаки, відстає від них на чверть періоду. В результаті струм зміщення й струм провідності виявляються протифазними. На високих частотах переважає струм зміщення, пропорційний частоті сигналу, на низьких – струм провідності, обернено пропорційний цій частоті (див. формулу (8.1.22)).

Таким чином, на частотах $\omega < \omega_p$ плазма має від'ємну діелектричну проникність і принципово непрозора для електромагнітних хвиль. Але при похилому падінні електромагнітної хвилі на шар плазми повне внутрішнє відбиття можливе й на частотах $\omega > \omega_p$ (див. формулу (8.1.15)).

Надалі ми користуватимемося поняттям дебаївського радіусу плазми.

Розглянемо пробний заряд, занурений у плазму. Для визначеності вважатимемо його позитивним. Електричний потенціал цього заряду створює потенціальну яму для електронів (рис. 8.1.5 а).

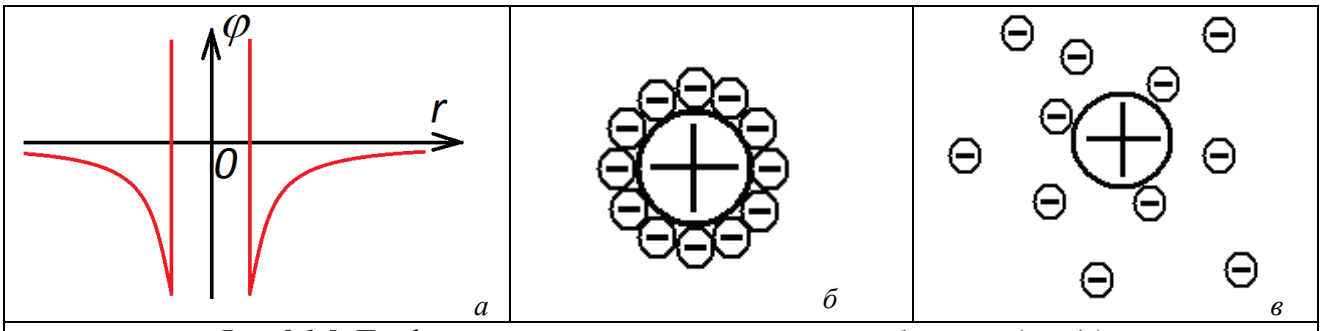


Рис. 8.1.5. Профіль потенціальної ями, створеної пробним зарядом (а) та екранування пробного заряду в холодній (б) і теплій (в) плазмі

Якби електрони плазми не мали кінетичної енергії (що відповідає так званій моделі холодної плазми), вони впритул наблизилися б до пробного заряду ("обліпили" б його) й повністю екранували б його кулонівське поле (рис. 8.1.5 б).

Оскільки в реальності електрони плазми мають ненульову кінетичну енергію (ненульову температуру), вони не опускаються на дно потенціальної ями, створеної пробним заря-

дом, а рухаються в ній. В результаті в околі пробного заряду утворюється підвищена концентрація електронів, які й екранують поле цього заряду, але тільки на помітних віддалях (рис. 8.1.5 в). Очевидно, в області екранування електронейтральність плазми порушується. Характерна віддаль, на якій поле малого пробного заряду буде екрануватися, називається дебаївським радіусом екранування, або просто дебаївським радіусом.

Розрахунок показує, що величина дебаївського радіусу визначається формулою

$r_D = \upsilon_{Te}/\omega_p$, $\qquad \upsilon_{Te} = \sqrt{k_B T_e/m}$,

де T_e та υ_{Te} – відповідно (абсолютна) температура та (середня) теплова швидкість електронів плазми. Оскільки, відповідно до формули (8.1.24), $\omega_p \sim \sqrt{n_e}$, то дебаївськи1 радіус зростає зі збільшенням електронної температури плазми T_e або зі зменшенням її концентрації n_e .

8.1.3. Діелектрична проникність газу

На відміну від плазми, в нейтральному газі немає вільних носіїв заряду (або їх настільки мало, що вони не відіграють помітної ролі). Електричні властивості газу визначаються зв'язаними зарядами, які входять до складу його молекул (або атомів, якщо газ одноатомний). Як відомо з курсу загальної фізики, ці заряди обумовлюють явище поляризації діелектрика в зовнішньому електричному полі.

Існує два основних механізми поляризації діелектрика. Перший з них реалізується, коли молекули мають відмінний від нуля дипольний момент (рис. 8.1.6 а). В цьому випадку зовнішнє електричне поле породжує обертальний момент, під дією якого молекула орієнтується вздовж цього поля (рис. 8.1.6 в).

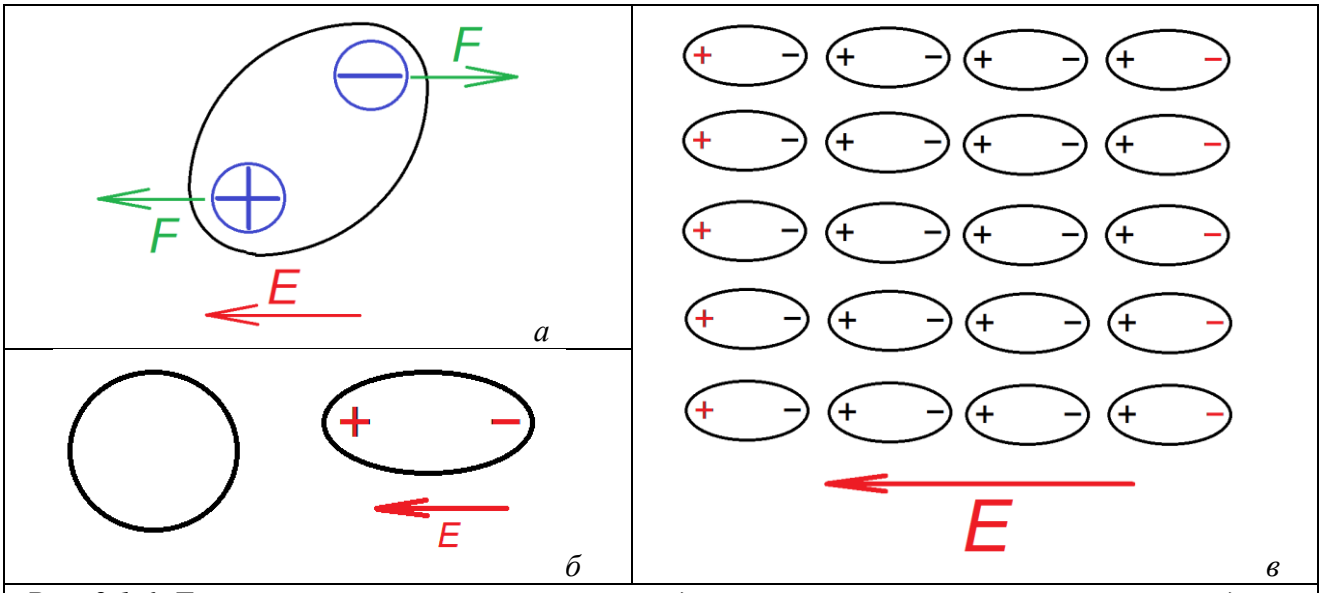


Рис. 8.1.6. До пояснення механізмів поляризації діелектрика: а – молекула з ненульовим дипольним моментом; б – поляризація молекули без дипольного моменту в зовнішньому електричному полі; в – поляризація плоского шару молекул і утворення поверхневих зарядів (показані червоним)

Для спрощення розглянемо шар таких орієнтованих молекул, перпендикулярний до зовнішнього електричного поля (рис. 8.1.6 в). Видно, що всередині шару заряди сусідніх молекул компенсуються, але на поверхнях утворюється шар заряду. Поверхневі заряди протилежних знаків породжують (як у плоскому конденсаторі) електричне поле, протилежне за напрямком до зовнішнього.

Другий механізм реалізується, коли за відсутності зовнішнього поля молекули діелектрика не мають дипольного моменту. Тоді такий момент з'являється в зовнішньому електричному полі (рис. 8.1.6 б). В однорідному полі всі молекули поляризуються однаково, тому знову приходимо до ситуації, показаної на рис. 8.1.6 в.

Насправді для молекул, що мають ненульовий дипольний момент, працюють обидва описані механізми.

Якщо електричне поле змінюється з часом за гармонічним законом, поляризація, очевидно, зазнає коливань за таким самим законом (принаймні в полі невеликої амплітуди).

Зазначимо, що поляризація молекул з ненульовим дипольним моментом супроводжується їхнім поворотом у просторі (в ньому беруть участь і ядра атомів). Поляризація ж молекул без власного дипольного моменту реально зводиться лише до зсуву електронної хмари. Тому на високих частотах роль поляризації, пов'язаної з поворотом молекул, зменшується в силу більшої інерційності цього процесу.

Для знаходження діелектричної проникності нейтрального газу скористаємося феноменологічним підходом. Приймемо для спрощення, що газ складається атомів без дипольного моменту. Вважатимемо, що мале відхилення електронної хмари від положення рівноваги породжує повертальну силу $\alpha\vec{r}$, пропорційну цьому відхиленню (аналог сили пружності). Вважатимемо також, що в системі присутнє дисипація (аналог в'язкого тертя), яку ми подамо у формі $\nu m d\vec{r}/dt$. Тоді рівняння руху для електронної хмари (ядро можна вважати нерухомим через його велику масу) можна подати у формі:

$$Zm\frac{d^2\vec{r}}{dt^2} = -Ze\vec{E}(t) - \alpha\vec{r} - \nu Zm\frac{d\vec{r}}{dt}, \qquad \vec{E}(t) = \vec{E}_m \exp(i\omega t), \tag{8.1.27}$$

Z – кількість електронів в атомі, ν – параметр дисипації.

Комплексна амплітуда вимушених коливань, описуваних рівнянням (8.1.27), має вигляд:

$$\vec{r}_m = \frac{Ze\vec{E}_m}{\omega^2 Zm - \alpha - i\omega\nu Zm} = \frac{e}{m}\frac{\vec{E}_m}{\omega^2 - \omega_0^2 - i\omega\nu}, \qquad \omega_0^2 = \frac{\alpha}{Zm}. \tag{8.1.28}$$

Ці коливання породжують густину струму з комплексною амплітудою

$$\vec{j}_m = -eZn_a\frac{d\vec{r}_m}{dt} = -i\frac{e^2 Zn_a}{m}\frac{\omega\vec{E}_m}{\omega^2 - \omega_0^2 - i\omega\nu}, \tag{8.1.29}$$

n_a – концентрація нейтральних атомів газу.

Як і в плазмі, діелектричну проникність знаходимо з умови

$$\frac{4\pi}{c}\vec{j} + \frac{i\omega}{c}\vec{E} = \frac{i\omega}{c}\varepsilon\vec{E}.$$

Отримаємо:

$$\varepsilon = 1 - \frac{4\pi e^2 Zn_a}{m\left(\omega^2 - \omega_0^2 - i\omega\nu\right)} = 1 - \frac{\omega_1^2}{\omega^2 - \omega_0^2 - i\omega\nu}, \qquad \omega_1^2 = \frac{4\pi e^2 Zn_a}{m}. \tag{8.1.30}$$

Частота ω_0 характеризує коливання електронної хмари окремого атома навколо положення рівноваги. Частоту ω_1 можна розглядати як деякий аналог плазмової частоти, вона пропорційна кореневі з концентрації газу n_a. Видно, що зі зростанням концентрації газу відхилення діелектричної проникності від одиниці зростає.

Конкретизуємо формулу (8.1.30) для земної атмосфери, вважаючи виконаною умову $\omega \ll \omega_0$, і врахуємо, що $p = nk_BT$. Тоді вказана формула набуває вигляду:

$$\varepsilon \approx 1 + \sum_i \frac{4\pi e^2 Zn_{a\,i}}{m\omega_{0i}^2} = 1 + \sum_i \frac{4\pi e^2 Zp_i}{m\omega_{0i}^2 k_B T}, \tag{8.1.31}$$

де сума береться по сортах атомів (молекул).

Очевидно, що поляризація (принаймні, на низьких частотах) більшою мірою проявляється для молекул з ненульовим дипольним моментом. В атмосфері це, в першу чергу, водяна пара (рис. 8.1.7).

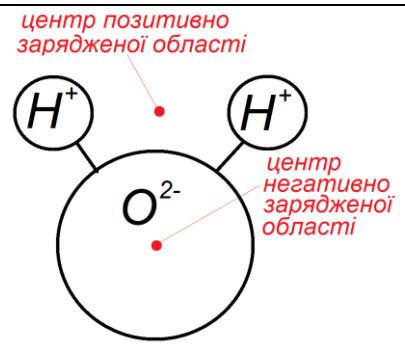


Рис. 8.1.7. Схематичне зображення молекули води

Строго кажучи, записані раніше формули (8.1.27)-(8.1.30) незастосовні до такого газу. Тим не менше, на низьких частотах (формула (8.1.31)) вони якісно правильно описують залежність діелектричної проникності від параметрів моделі. Слід лише врахувати, що тепловий рух перешкоджає процесу поляризації, тому можна вважати, що для таких молекул параметр α обернено пропорційний температурі. Тоді, вважаючи, що парціальний тиск водяної пари значно менший від атмосферного, формулу (8.1.31) можна подати у вигляді

$$\varepsilon = 1 + \frac{\alpha_1 p}{T} + \frac{\alpha_2 p_w}{T^2}, \qquad \alpha_1 = \sum_i \frac{4\pi e^2 Z}{m\omega_{0i}^2 k_B T}\frac{p_i}{p} \tag{8.1.32}$$

де перший після одиниці доданок відповідає внеску молекул без дипольного моменту, а другий – внеску водяної пари (p_w – парціальний тиск водяної пари). Коефіцієнти α_1 та α_2 знаходяться з експериментів.

Підставивши до (8.1.32) числові значення, отримуємо:

$$\varepsilon = 1 + \frac{1.552\cdot 10^{-4}}{T}\left(p + \frac{4810\, p_w}{T}\right), \tag{8.1.33}$$

де p та p_w беруться в мілібарах (1 мбар=102.6 Па=0.75 мм рт. ст.), T – в Кельвінах.

Для сухого повітря за атмосферного тиску (1000 мбар) і кімнатної температури (300 К) маємо: $\varepsilon \approx 1.0005$. тобто діелектрична проникність мало відрізняється від одиниці. Тиск насиченої пари за кімнатної температури складає близько 30 мм рт. ст., або 40 мбар. Тоді для атмосферного повітря з 100% вологістю $\varepsilon \approx 1.0008$. Отже, діелектрична проникність атмосфери за відсутності іонізації дуже мало відрізняється від одиниці. Таким чином, якщо ми маємо газ із не надто малим ступенем іонізації, то його діелектрична проникність визначається іонізованою компонентою.

Тим не менше, зміна густини нейтрального газу в просторі (наприклад, атмосферні флуктуації) породжує неоднорідності діелектричної проникності (і, відповідно, показника заломлення), на яких можливе розсіювання електромагнітних хвиль відповідної довжини. Таке розсіювання хвиль, випромінюваних передавачем, породжує додатковий шум на вході приймача.

Контрольні питання до підрозділу 8.1

1. Як враховуються струми провідності, що протікають у середовищі, при розрахунку його діелектричної провідності?
2. До якої категорії належать струми, які протікають у діелектрику, вміщеному в змінне

електричне поле: струмів провідності чи струмів зміщення?

3. У чому відмінність між струмами, що протікають у плазмі й у нейтральному газі під дією змінного електричного поля?
4. Чому діелектрична проникність плазми без зіткнень завжди менша за одиницю, а на низьких частотах стає від'ємною?
5. В яких випадках можна вважати, що діелектрична проникність газу близька до одиниці?
6. Чому водяна пара, пори її відносно малий парціальний тиск, дає суттєвий внесок до діелектричної проникності атмосфери?

Задачі до підрозділу 8.1

8.1.1. Плоска електромагнітна хвиля поширюється в діелектрику з проникністю $\varepsilon > 1$ й відбивається від його плоскої межі з вакуумом. Коли хвиля відбиватиметься повністю?

8.1.2. Концентрація електронів у плазмі – 10^{12} см$^{-3}$, частота зіткнень електронів з важкими частинками – 10^5 с$^{-1}$. Частота коливань удвічі більша за електронну ленгмюрівську частоту. Чи можна для таких коливань вважати діелектричну проникність плазми дійсною?

8.1.3. При якій вологості повітря внесок водяної пари до діелектричної проникності тропосфери буде таким самим, як тропосфери без водяної пари? Тиск нормальний, температура кімнатна.

8.2. Електричні властивості земної атмосфери

Радіохвилі, які забезпечують радіозв'язок, поширюються в атмосфері, властивості якої мають вирішальний вплив на їхнє поширення. Тому, перш ніж розглядати поширення радіохвиль, слід обговорити властивості атмосфери.

8.2.1. Структура та властивості атмосфери

Атмосфера – це шар газу, який оточує Землю. Умовно вважають, що в спокійних умовах вона простягається до висот у 2-3 радіуси Землі (радіус Землі складає приблизно 6,4 тис. км). Під час магнітних бурь висота атмосфери збільшується і може досягати 20 радіусів Землі. Концентрація атмосфери з висотою зменшується (рис. 8.2.1 а).

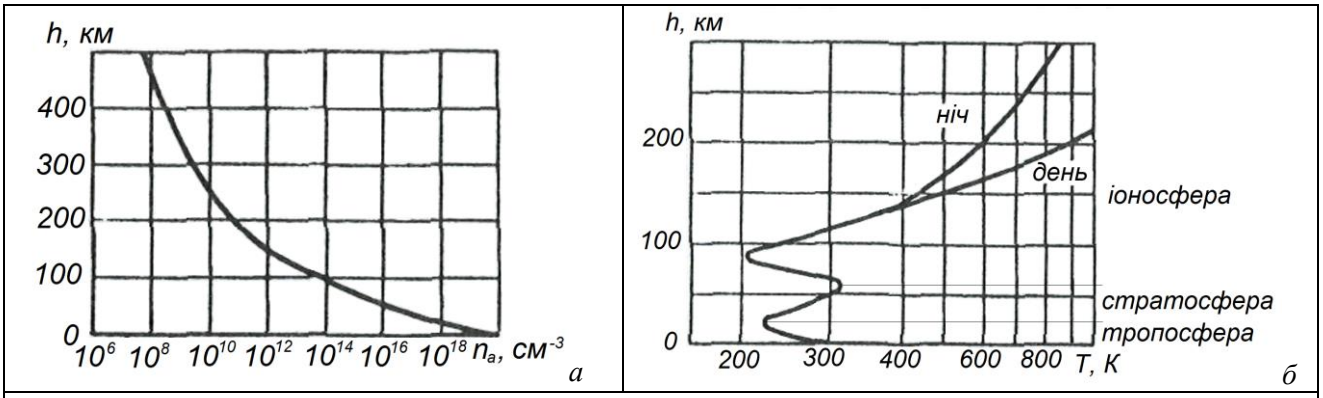


Рис. 8.2.1. Графіки висотних залежностей середньої концентрації нейтральних частинок (а) і температури (б) в атмосфері

Хімічний склад атмосфери залишається однорідним до висот приблизно 90 км через постійне перемішування повітряними потоками. У цьому проміжку висот атмосфера складається переважно з азоту й кисню в молекулярному стані. На більших висотах ці молекули дисоціюють (розпадаються на атоми) під дією ультрафіолетового випромінювання Сонця.

На висотах у кілька сотень кілометрів і вище відбувається дифузне розділення газів за їхньою молекулярною масою. В результаті на висотах понад 1000 км атмосфера складається переважно з водню.

Атмосферу традиційно поділяють на три шари – тропосферу, стратосферу та іоносферу.

Тропосфера простягається від поверхні Землі до висот 7-10 км у полярних областях, 10-12 км у середніх широтах та 16-18 км в екваторіальних областях. Вона складається з нейтрального газу У тропосфері температура з висотою зменшується (рис. 8.2.1 б), оскільки ця область нагрівається переважно від поверхні Землі, яку, в свою чергу, нагріває сонячне випромінювання.

В тропосфері присутня водяна пара, яка помітно впливає на поширення радіохвиль. Її концентрація швидко спадає з висотою й на верхній межі тропосфери зменшується більш ніж на два порядки.

Стратосфера простягається від верхнього краю тропосфери до висоти приблизно 60 км. Як і тропосфера, вона складається з нейтрального газу, але температури стратосфери з висотою зростає (рис. 8.2.1 б). Це зумовлено інтенсивним поглинанням сонячного випромінювання (зокрема, ультрафіолетового) на висотах близько 60 км, де в атмосфері наявна помітна частка озону.

Іоносфера розташована вище стратосфери. Її основна особливість полягає в тому, що газ у ній виявляється помітно іонізованим (зокрема, за рахунок поглинання сонячного ультрафіолету). Хоча ступінь іонізації іоносфери (відношення концентрації вільних електронів до сумарної концентрації нейтральних атомів та іонів) невеликий і складає 10^{-3} і менше, такий

слабкоіонізований газ уже поводить себе як плазма (з високою частотою зіткнень електронів з важкими частинками через велику концентрацію нейтральних атомів) – див. вище п. 8.1.2. Висотний хід температури в іоносфері є немонотонним (рис. 8.2.1 б). Це пов'язано з тим, що з одного боку, як уже відзначалося, на висоті близько 60 км атмосфера локально нагрівається за рахунок поглинання ультрафіолету озоновим шаром, а з іншого – з висотою зростає інтенсивність сонячного опромінення, яке безпосередньо нагріває атмосферу.

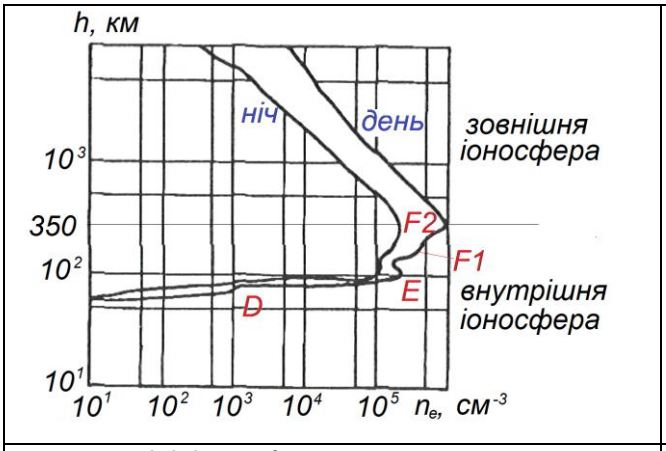


Рис. 8.2.2. Графік висотної залежності концентрації електронів в іоносфері

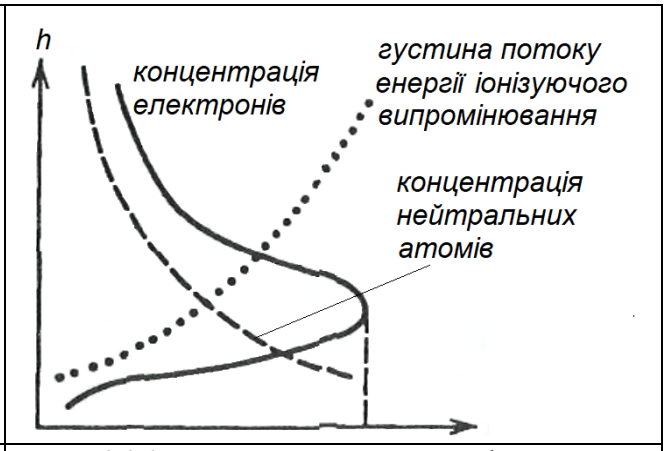


Рис. 8.2.3. До пояснення механізму формування головного максимуму концентрації електронів

Висотна залежність концентрації електронів в іоносфері (рис. 8.2.2) теж немонотонна. До того ж вона змінюється протягом доби.

Основним джерелом іонізації виступає фотоіонізація сонячним ультрафіолетом (енергії квантів випромінювання з більшими довжинами хвиль, $\lambda > 134$ *нм* бракує для іонізації). Механізм формування основного максимуму густини електронів зрозумілий з графіків, показаних на рис. 8.2.3: з висотою концентрація нейтральних атомів спадає, а густина потоку енергії іонізуючого випромінювання, навпаки, зростає. Кількість актів іонізації за одиницю часу в одиниці об'єму пропорційна добутку цих величин, який на деякій висоті досягає максимуму. Нарешті, в плазмі існує динамічна рівновага між процесами іонізації та рекомбінації, тому її рівноважна концентрація пропорційна інтенсивності іонізації.

Тепер обговоримо висотну залежність концентрації електронів детальніше.

В структурі іоносфери традиційно виділяють окремі висотні проміжки – області (шари) D, E та F (вони показані на рис. 8.2.2).

Шар D знаходиться на висотах 60-90 км, концентрація електронів там порівняно низька й складає 10^2 - 10^3 см$^{-3}$. Основний внесок до іонізації забезпечує рентгенівське випромінювання Сонця. Додаткові джерела – метеорні тіла, що згорають у щільних шарах атмосфери, космічні промені, а під час магнітних бур – ще й частинки високих енергій, що заносяться з магнітосфери Землі, розташованої вище від її атмосфери. В темну пору доби концентрація електронів у шарі D різко зменшується.

Шар E знаходиться на висотах 90-120 км, концентрація електронів там складає до 10^5 см$^{-3}$ і також змінюється протягом доби. Різкому зменшенню її вночі протидіє дифузія плазми з шару F, який розташований вище і де концентрація електронів суттєво більша. Додатковими джерелами нічної іонізації виступають метеорні тіла, космічні промені тощо.

На висотах 100-120 км майже завжди спостерігається так званий спорадичний шар E_S, щільний ($n_e \geq 10^5$ см$^{-3}$), але дуже тонкий (0,5-1 км)

Шаром F зараз називають усю іоносферу на висотах 130-140 км. Максимальна іонізація відбувається на висотах 150-200 км. Однак через відносно тривалий час життя плазма дифундує вгору, і максимальна концентрація електронів досягається на висотах 250-400 км.

У денний час внаслідок іонізації сонячним ультрафіолетом на висотах 150-200 км спостерігається щось на кшталт локального максимуму електронної концентрації. Ця область відома в літературі як шар F1. Частину шару F вище від F1 називають шаром F2. Саме тут досягається максимальна електронна концентрація – 10^5 - 10^6 $см^{-3}$.

Області нижче й вище шару F2 називають відповідно внутрішньою та зовнішньою іоносферою. У зовнішній іоносфері концентрація електронів повільно зменшується з висотою (рис. 8.2.2).

Концентрація електронів зазнає регулярних квазіперіодичних змін, зумовлених зміною густини потоку енергії іонізуючого випромінювання. Виділяють добові зміни, сезонні зміни, а також зміни, спричинені 11-річними циклами сонячної активності.

Як уже зазначалося (див. вище п. 8.1.3), нагрівання плазми зумовлене зіткненнями електронів з важкими частинками – іонами та нейтральними атомами. Графік висотної залежності ефективної частоти ν електронних зіткнень показаний на рис. 8.2.4.

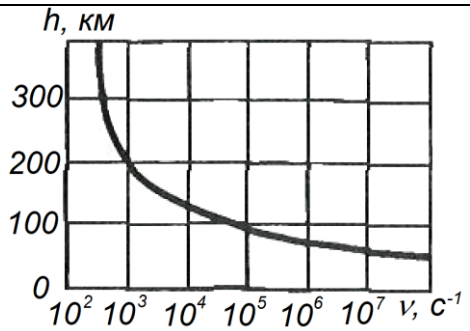


Рис. 8.2.4.Графік висотної залежності ефективної частоти ν електронних зіткнень

8.2.2. Електричні властивості тропосфери та стратосфери

Оскільки ступінь іонізації газу в тропосфері й стратосфері дуже низький, діелектрична проникність цих областей добре описується формулами (8.1.32)-(8.1.33) для нейтрального газу, які можна подати у вигляді

$$\varepsilon = 1 + \frac{\alpha_1 p}{T} + \frac{\alpha_2 p_w}{T^2} = 1 + \frac{1.552 \cdot 10^{-4}}{T[\text{К}]}\left(p[\text{мбар}] + \frac{4810 p_w[\text{мбар}]}{T[\text{К}]} \right). \qquad (8.2.1)$$

Графік висотної залежності середньої концентрації нейтральних атомів у тропосфері й стратосфері показана на рис. 8.2.1 а. Але, крім цієї регулярної неоднорідності, існують також різноманітні флуктуації концентрації і відповідні флуктуації діелектричної проникності (рис. 8.2.5). Величина і просторовий розподіл цих флуктуацій постійно змінюються з часом.

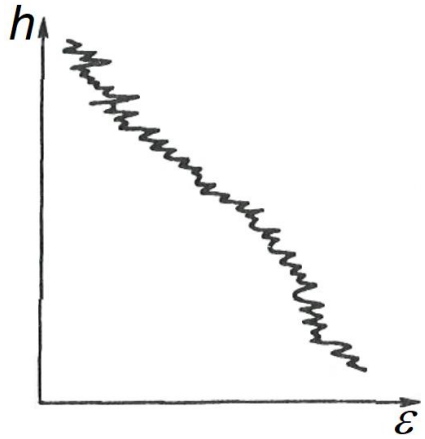


Рис. 8.2.5. Графік залежності миттєвого значення діелектричної проникності тропосфери від висоти вздовж фіксованої вертикалі

Флуктуації густини тропосфери й стратосфери поділяють на турбулентні й шаруваті. Турбулентні флуктуації зумовлені турбулентним (вихровим) рухом атмосферного повітря

(через малу в'язкість повітря його течія практично завжди є турбулентною). Їхні розміри приблизно однакові в усіх напрямках і варіюються від кількох міліметрів до кількох десятків метрів, а збурення діелектричної проникності складає величину порядку 10^{-6} на висотах 1-3 км.

Шаруватим флуктуаціям притаманні великі поперечні розміри й значно менша товщина. Причиною їхнього виникнення може бути температурна інверсія (коли шар холодного повітря знаходиться над шаром теплого), а також хмарність. Товщина шаруватих неоднорідностей лежить в інтервалі від кількох дециметрів до кількох сотень метрів, відповідно поперечні розміри варіюються від десятків метрів до десятків кілометрів і більше. Відповідне збурення діелектричної проникності змінюється в межах від 10^{-6} до 10^{-4}. Кількість та інтенсивність шаруватих флуктуацій зменшується з висотою.

8.2.3. Електричні властивості іоносфери

Діелектрична проникність іоносфери визначається її іонізованою складовою. В загальному випадку вона описується формулами (8.1.23)-(8.1.24):

$$\varepsilon = 1 - i\frac{4\pi e^2 n_e}{m\omega(i\omega+\nu)} = 1 - \frac{\omega_p^2}{\omega(\omega - i\nu)} = 1 - \frac{\omega_p^2}{\omega^2+\nu^2} - i\frac{\nu\omega_p^2}{\omega\left(\omega^2+\nu^2\right)}, \qquad \omega_p = \sqrt{\frac{4\pi e^2 n}{m}}$$

Частота ν зіткнень електронів з іонами та нейтральними атомами досягає максимуму в шарі D і складає там приблизно 10^7 c^{-1}. На частотах $\omega \gg \nu$ (тобто вище 10 МГц) можна користуватися спрощеною формулою (8.1.26):

$$\varepsilon \approx 1 - \frac{\omega_p^2}{\omega^2} = 1 - 80.8\frac{n_e\left[\text{м}^{-3}\right]}{\left(f\left[\text{Гц}\right]\right)^2}. \qquad (8.2.2)$$

На рис. 8.2.6 показано графіки модельної залежності концентрації плазми від висоти (з характерним максимумом) і відповідних залежностей діелектричної проникності на трьох різних частотах. Видно, що на найнижчій частоті в околі максимуму концентрації з'являється область від'ємних значень ε. Як уже зазначалося, в такій області поширення електромагнітних хвиль стає неможливим (див. вище формулу (8.1.9) та її обговорення). Це означає, що електромагнітні хвилі достатньо низької частоти, які випромінюються з поверхні Землі, відбиваються від щільних шарів іоносфери.

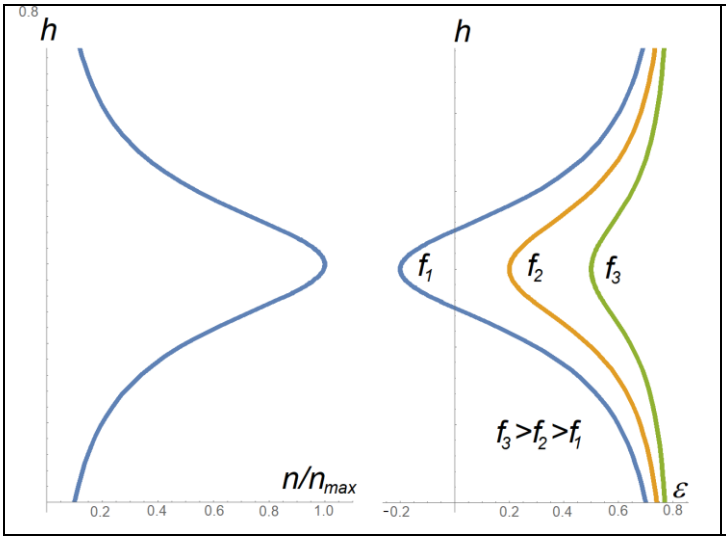


Рис. 8.2.6. Ліворуч – графік висотного ходу електронної концентрації (показаний якісно), праворуч – відповідні йому профілі діелектричної проникності на різних частотах. На найнижчій частоті поблизу максимуму концентрації утворюється область від'ємних значень діелектричної проникності

На рис. 8.2.7 подано графіки висотних залежностей електронної концентрації, ефективної частоти зіткнень електронів з важкими частинками та добутку цих величин. Останній визначає уявну частину діелектричної проникності (див. формулу (8.1.26)), та, отже, втрати

за рахунок дисипації (нагрівання плазми). З рисунку видно, що найбільш інтенсивне розсіювання енергії хвиль відбувається в D-шарі на висотах близько 90 км. Як уже згадувалося, цей шар існує лише в денний час. Тому вдень електромагнітні хвилі поглинаються дужче, ніж уночі.

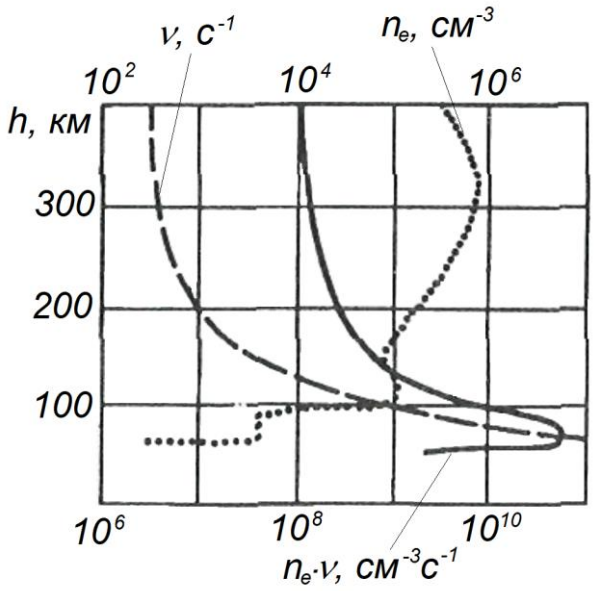


Рис. 8.2.7. Графіки висотних залежностей денної концентрації електронів n_e, ефективної частоти електронних зіткнень ν та добутку цих величин $n_e\nu$, який визначає активну провідність плазми

В іоносфері виникають також нерегулярні неоднорідності концентрації плазми, подібно до виникнення флуктуацій густини нейтрального газу в тропосфері та стратосфері. Дрібномасштабні неоднорідності з горизонтальними розмірами в кілька сотень метрів і відносним збуренням густини плазми порядку 10^{-2} виникають на висотах 80-400 км внаслідок турбулентного руху плазми та процесів дифузії.

Великомасштабні неоднорідності з характерними горизонтальними розмірами десятки – сотні кілометрів мають форму еліпсоїдів, зорієнтованих уздовж ліній геомагнітного поля, і виникають через коливні процеси в іоносфері.

Окремо виділяють іоносферні збурення (бурі) тривалістю не менше години. Такі збурення можуть породжуватись імпульсними потоками заряджених частинок, які надходять від Сонця. Вони можуть призводити до порушення радіозв'язку. Іоносферні бурі протікають порізному в залежності від широти точки спостереження.

У низьких та середніх широтах (тобто в широких областях, які з обох боків прилягають до екватора) бурі зазвичай впливають лише на шар F, нижні шари збурюються лише під час дуже сильних бур. Для цих областей характерні так звані негативні збурення, коли електронна концентрація зменшується на 30-40%. Крім того, в околі шару F2 помітно зростає рівень флуктуацій електронної концентрації.

У високих широтах (приполярних областях) електронна концентрація зростає всюди в іоносфері, аж до шару D включно. Зростання електронної концентрації супроводжується збільшенням активної провідності іоносфери й відповідним збільшенням поглинання електромагнітних хвиль.

Спостерігаються й бурі хвильового походження, породжені потужним рентгенівським випромінюванням за рахунок так званих хромосферних[1] спалахів на Сонці. Бурі такого типу супроводжуються різким зростанням поглинання і зазвичай тривають від кількох хвилин до

[1] Хромосфера – один із зовнішніх шарів Сонця. Характеризується аномально високою температурою – від $4\cdot10^4$ К до $1.5\cdot10^5$ К.

1-2 годин, охоплюючи всю освітлену поверхню Землі.

8.2.4. Геомагнітне поле

Магнітне поле Землі (геомагнітне поле) на висотах до приблизно двох радіусів Землі нагадує поле магнітного диполя (рис. 8.2.7). Насправді розподіл магнітного поля по земній поверхні носить складніший характер і характеризується відсутністю симетрії щодо осі, яка сполучає північний та південний полюси (рис. 8.2.8). До того ж цей розподіл змінюється з часом.

З рис. 8.2.8 видно, що магнітне поле над поверхнею Землі змінюється в межах 20-60 мкТл (0.2-0.6 Гс).

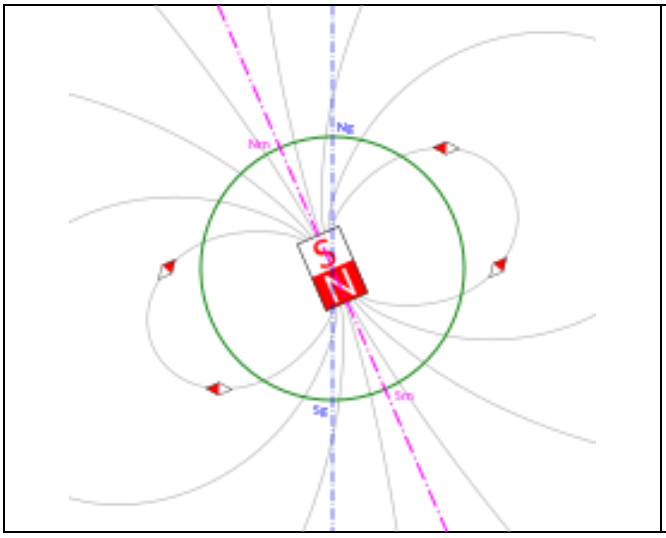


Рис. 8.2.7. Схематичне зображення силових ліній геомагнітного поля Землі. Видно, що положення магнітних і географічних полюсів відрізняються

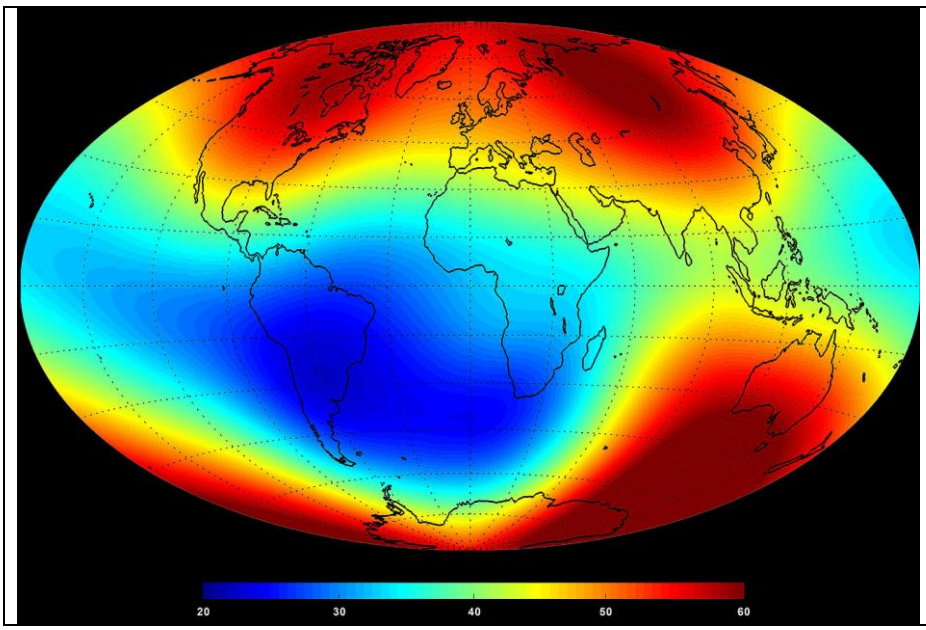


Рис. 8.2.8. Діаграма розподілу індукції магнітного поля по поверхні Землі. Шкала – в мкТл (10^{-2}Гс)

Наявність геомагнітного поля суттєво ускладнює рух електронів і зв'язане з ним поширення радіохвиль. Діелектрична проникність стає тензором. У відповідному частотному діапазоні спостерігається, зокрема, подвійне заломлення променів, коли на тій самій частоті поширюються дві хвилі з відмінними за величиною хвильовими векторами (звичайна й незвичайна хвилі).

Впливом магнітного поля на властивості плазми на частоті ω зазвичай можна знехтувати за умов, що

$$\omega, \omega_p >> \omega_c, \tag{8.2.3}$$

де

$$\omega_c = \frac{eB}{mc}, \tag{8.2.4}$$

або

$$f_c[\text{Гц}] = 2.8 \cdot 10^6 B[\text{Гс}] = 2.8 \cdot 10^{10} B[\text{Тл}] \tag{8.2.5}$$

– електронна циклотронна частота (частота обертання електронів навколо силових ліній магнітного поля).

Контрольні питання до підрозділу 8.2

1. Чим відрізняються тропосфера й стратосфера? Де проходить їхня межа?
2. Якісно поясніть висотний хід концентрації іоносферної плазми.
3. З чим пов'язані добові та річні коливання концентрації іоносферної плазми?
4. На яких частотах в іоносфері з'являються області з від'ємною діелектричною проникністю? Де вони розташовані?
5. Яка природа випадкових неоднорідностей діелектричної проникності в тропосфері, стратосфері та іоносфері?
6. В яких областях на поверхні Землі геомагнітне поле найбільше, а в яких – найменше?

Задачі до підрозділу 8.2

8.2.1. Для яких частот іоносфера може бути прозорою для електромагнітних хвиль? Відповіді дати для денних та нічних умов.

8.2.2. Для яких частот вплив геомагнітного поля на властивості іоносферної плазми в максимумі її концентрації в денний час виявляється суттєвим? Відповіді дати для полярних та екваторіальних областей. Ослабленням магнітного поля з висотою знехтувати.

8.2.3. Чи існують у внутрішній іоносфері області, де на частотах $\omega \approx \omega_p$ влив геомагнітного поля на властивості плазми буде суттєвим? Якщо так, то де вони розташовані? Ослабленням магнітного поля з висотою знехтувати.

8.3. Поширення радіохвиль у вільному просторі

Спочатку отримаємо формули, які дозволяють розрахувати електромагнітне поле в точці приймання сигналу за відомими характеристиками передавальної антени та проміжку простору між передавачем і приймачем.

Щоб правильно розрахувати електромагнітне поле, яке передавальна антена створює в точці приймання сигналу, потрібно знати властивості середовища (газ, плазма та інше), через яке поширюються електромагнітні хвилі, та об'єктів, від яких вони можуть відбиватися (наприклад, поверхня Землі). Слід підкреслити, що розрахунок такого роду, виконаний у наближенні геометричної оптики, буде некоректним, оскільки не враховує явищ інтерференції та дифракції (приклад розрахунку з урахуванням інтерференції для найпростішої моделі див. вище у п. 7.3.3).

8.3.1. Електричне поле, створене передавальною антеною в точці приймання сигналу

Нагадаємо (див. вище п. 7.2), що середня густина потоку енергії (радіальна компонента вектора Пойнтінга) в далекій зоні випромінювання у вакуумі для хвилі, електричне поле якої лежить у вертикальній площині, в системі СІ (пор. з формулою (7.2.27), поданою в системі Гаусса) записується у формі

$$\Pi_r = \frac{1}{2} E_\theta H_\varphi^* \approx \frac{1}{2} E_m H_m^* \qquad (8.3.1)$$

(в далекій зоні випромінювання фронт хвилі можна вважати плоским з комплексними амплітудами електричного та магнітного полів E_m, H_m). Амплітуди E_m, H_m пов'язані співвідношенням

$$\frac{E_m}{H_m} \equiv Z_{0v} = \sqrt{\frac{\mu_0}{\varepsilon_0}} \approx 120\pi\,\text{Ом} \approx 377\,\text{Ом}, \qquad (8.3.2)$$

де Z_{0v} – опір вільного простору (хвильовий опір вакууму; див. формулу (7.2.26)). Виключивши магнітне поле з (8.3.1) за допомогою (8.3.2), отримуємо:

$$\Pi_r = \frac{\left|E_m^2\right|}{2Z_{0v}}, \qquad (8.3.3)$$

звідки

$$\left|E_m\right| = \sqrt{2\Pi_r Z_{0v}}\,. \qquad (8.3.4)$$

В далекій зоні антену можна розглядати як точковий випромінювач, і її потужність у середовищі без втрат має дорівнювати потоку енергії через будь-яку поверхню, яка її оточує. Приймемо за таку поверхню сферу радіуса r, центр якого збігається з антеною. В цьому випадку густина потоку енергії складає Π_r, а сам потік енергії – $4\pi r^2 \Pi_r$. Отже, густину потоку енергії в далекій зоні для ізотропного випромінювача можна записати через його випромінювану потужність P_r та віддаль до точки приймання r таким чином:

$$\Pi_r = \frac{P_r}{4\pi r^2} \qquad (8.3.5)$$

Підставивши (8.3.5) до (8.3.4) з урахуванням явного вигляду Z_{0v}, отримуємо:

$$E_m = \frac{\sqrt{60 P_r}}{r}\,. \qquad (8.3.6)$$

Якщо передавальна антена характеризується коефіцієнтом спрямованої дії K_{dir} (див. вище п. 7.4.3) порівняно з ізотропною антеною, формула (8.3.6) набуває вигляду

$$E_m = \frac{\sqrt{60 P_r K_{dir}}}{r} = \frac{\sqrt{60PG}}{r}, \qquad (8.3.7)$$

де P – потужність, що підводиться до антени, G – коефіцієнт підсилення антени[2] (пор. з формулою (7.4.11)).

В реальності електромагнітна хвиля поширюється не у вакуумі, а в середовищі (атмосфері), де вона може додатково ослаблюватися за рахунок поглинання і розсіювання на флуктуаціях густини. Ці явища враховують за допомогою так званого множника ослаблення V (в загальному випадку комплексного), тому остаточно формула (8.3.7) записується у формі:

$$E_m = \frac{\sqrt{60PG}}{r} V. \qquad (8.3.8)$$

Цією формулою ми будемо надалі користуватися для розрахунку поля передавальної антени в точці приймання сигналу.

8.3.2. Поняття про суттєву область

Інтуїтивно зрозуміло, що вплив на електромагнітне поле, яке формується випромінювальною антеною в точці приймання сигналу, має деяка обмежена область простору, яка оточує пряму, що сполучає передавач та приймач (очевидно, властивості середовища на далеких віддалях від цієї лінії будуть несуттєвими). В теорії антен її інколи називають суттєвою областю.

Розміри суттєвої області можна визначити експериментально. Розмістимо в деякій точці прямої, що сполучає передавач *A* і приймач *B*, непрозорий плоский екран, перпендикулярний до цієї прямої. Зробимо в ній отвір у формі кола, центр якого лежить на згаданій прямій (рис. 8.3.1 а)

При дуже великому діаметрі d отвору екран не впливає на електромагнітне поле в точці В. При зменшенні d до деякого порогового значення такий вплив з'являється (очевидно, при $d = 0$ поле в точці В взагалі зникає). Це порогове значення й можна вважати діаметром суттєвої області в площині розташування екрану. Пересуваючи екран між точками *A* і *B*, можна таким способом визначити форму суттєвої області. Якщо середовище однорідне, то ця область, вочевидь, буде аксіально симетричною щодо прямої *AB*.

Спробуємо знайти форму і розміри суттєвої області теоретично.

8.3.3. Зони Френеля

В основу розрахунку покладемо відомий з оптики принцип Гюйгенса – Френеля. Відповідно до цього принципу будь-яка точка простору, якої досягла первинна хвиля, стає джерелом вторинних хвиль, що поширюються в усіх напрямках. Ці вторинні хвилі додаються, підсилюючи одна одну в одних напрямках і гасячи в інших, залежно від їхньої різниці фаз. Принцип Гюйгенса – Френеля пояснює, зокрема, явища дифракції, інтерференції, відбиття та розсіювання хвиль довільної природи (зокрема, й електромагнітних).

Відповідно до принципу Гюйгенса – Френеля поверхню, яка є джерелом вторинних хвиль, можна обирати довільно – аби тільки вона була замкнутою й охоплювала випромінювач. В даному разі її зручно обрати у вигляді половини сфери S_∞ дуже великого (формально нескінченного) радіусу та площини S_0, яка лежить між точками *A* та *B* і перпендикулярна до відрізку *AB* – точніше, частини цієї площини у формі кругу, що потрапляє всередину згаданої сфери (рис. 8.3.1 б). Для знаходження сумарного електромагнітного поля в точці *B* необхідно підсумувати в ній вторинні хвилі, випромінювані поверхнями S_∞ та S_0 (з урахуванням їхніх

[2] Нагадаємо, що коефіцієнт підсилення антени – це добуток к.с.д. антени на її к.к.д.

фаз).

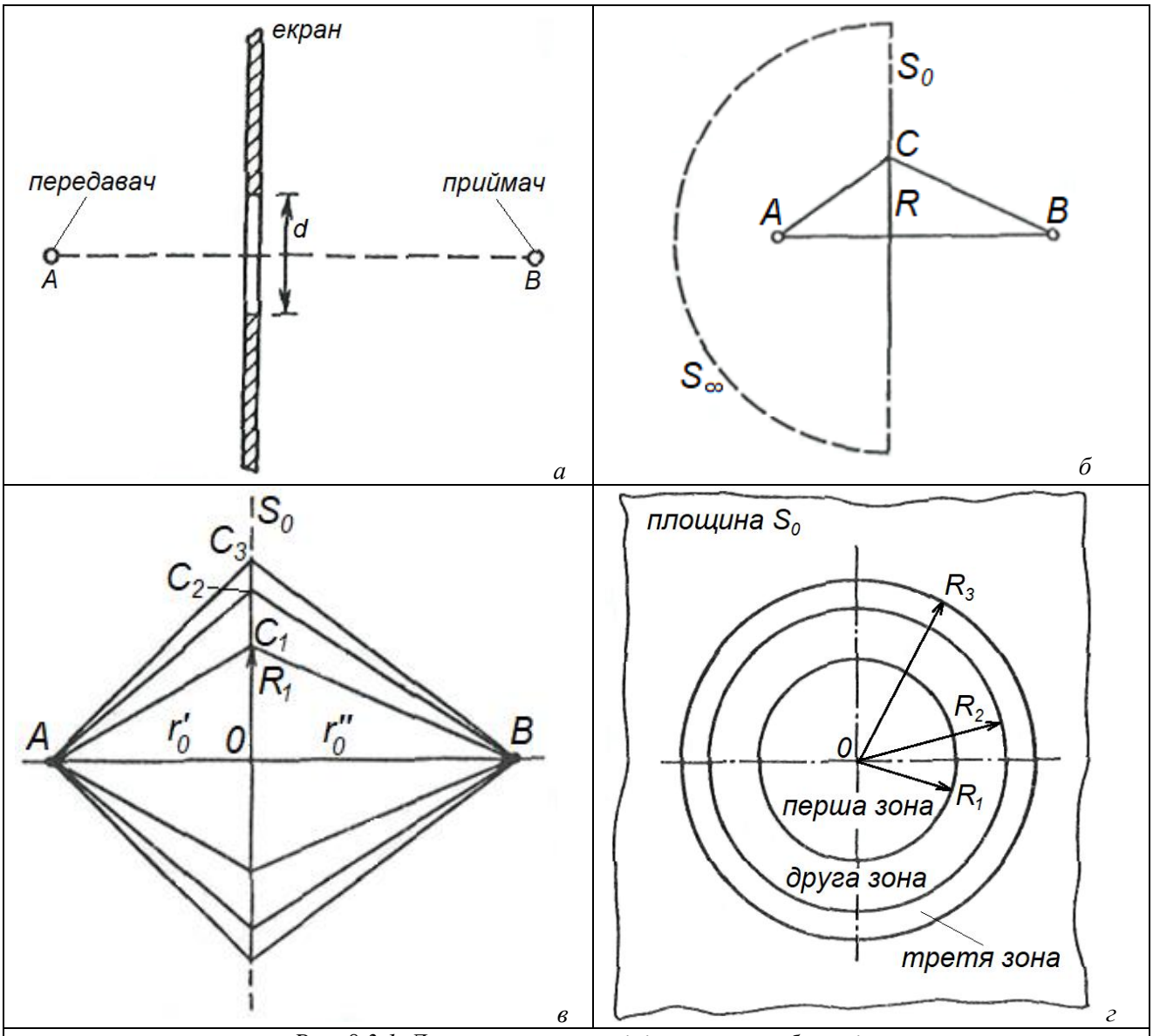


Рис. 8.3.1. До визначення розмірів суттєвої області

Півсфера S_∞ нескінченно віддалена від точки *B*, тому вторинні хвилі, джерелом яких вона є, не впливатимуть на поле в цій точці. Те саме можна сказати про точки площини S_0, значно віддалені від осі *AB*. Отже, поле в точці *B* визначається сумою вторинних хвиль, які приходять туди з тієї частини площини S_0, яка лежить недалеко від осі *AB*.

Початкова фаза вторинних хвиль, випромінюваних деякою точкою *C*, яка лежить на площині S_0, визначається віддаллю *AC* від цієї точки до джерела первинних хвиль (рис. 8.3.1 б). При зміні віддалі R від точки *C* до осі *AB*, ця фаза, очевидно, буде змінюватися неперервно. Однак для спрощення розрахунку зручно розбити площину S_0 на області, всі точки яких випромінюють приблизно в однаковій фазі – так звані зони Френеля. Зони Френеля визначають з умови, що фази полів вторинних хвиль, які приходять у точку *B* з різних точок тієї самої зони, відрізняються не більше ніж на π. Іншими словами, для точок однієї й тієї самої зони Френеля віддалі AC_1B (див. рис. 8.3.1 в) не повинні відрізнятися більше ніж на половину довжини хвилі.

Якщо точка C_1 є межею першої зони Френеля (яка вміщує точку перетину площини

S_0 з відрізком AB), то для неї має виконуватися умова:

$$AC_1B - AB = \lambda/2, \tag{8.3.9}$$

а радіус першої зони R_1 відповідно може бути знайдений з умови

$$\sqrt{(r_0')^2 + R_1^2} + \sqrt{(r_0'')^2 + R_1^2} - r_0' - r_0'' = \lambda/2 \tag{8.3.10}$$

(див. рис. 8.3.1 в).

Аналогічно для зовнішнього радіусу R_n зони Френеля з номером п може бути знайдений з умови

$$\sqrt{(r_0')^2 + R_n^2} + \sqrt{(r_0'')^2 + R_n^2} - r_0' - r_0'' = n\lambda/2. \tag{8.3.11}$$

Вважатимемо, що віддаль *AB* набагато перевищує довжину хвилі. Як буде показано далі, для розрахунку радіусу суттєвої області достатньо врахувати зони Френеля з номерами менше або порядку 10. Тоді можна прийняти, що для не надто малих значень r_0' та r_0'' виконуються умови

$$R_n \ll r_0', \tag{8.3.12}$$

$$R_n \ll r_0''. \tag{8.3.13}$$

Тоді можна наближено записати, що

$$\sqrt{(r_0')^2 + R_n^2} - r_0' = \frac{(r_0')^2 + R_n^2 - (r_0')^2}{\sqrt{(r_0')^2 + R_n^2} + r_0'} \approx \frac{R_n^2}{2r_0'}, \qquad \sqrt{(r_0'')^2 + R_n^2} - r_0'' \approx \frac{R_n^2}{2r_0''}, \tag{8.3.14}$$

звідки

$$\frac{R_n^2}{2r_0'} + \frac{R_n^2}{2r_0''} = \frac{n\lambda}{2}, \tag{8.3.15}$$

або

$$R_n = \sqrt{\frac{n\lambda r_0' r_0''}{r_0' + r_0''}}. \tag{8.3.16}$$

Максимальний зовнішній радіус *n*-ї зони Френеля досягається на середині відрізку *AB*, де $r_0' = r_0'' \equiv L$:

$$R_{n\,\max} = \sqrt{n\lambda L/2}. \tag{8.3.17}$$

Як уже говорилося, коли середовище, в якому поширюються хвилі, однорідне, то система характеризується аксіальною симетрією щодо осі *AB*. Тому перша зона Френеля – це круг радіусу R_1, зони з номерами $n > 1$ – кільця, внутрішній і зовнішній радіуси яких дорівнюють відповідно R_{n-1} і R_n (рис. 8.3.1 г).

З формули (8.3.11) випливає, що електромагнітні поля, створені в точці спостереження будь-якою парою сусідніх зон Френеля, будуть протифазними.

8.3.4. Розрахунок напруженості електричного поля в точці приймання сигналу

Повне електричне поле $\vec{E}_0$ у точці приймання сигналу може бути записане як векторна сума внесків від кожної із зон Френеля:

$$\vec{E}_0 = \vec{E}_1 + \vec{E}_2 + \vec{E}_3 + \vec{E}_4 + \ldots \tag{8.3.18}$$

Акуратні розрахунки показують, що електричні поля, створені в точці спостереження різними зонами Френеля, будуть практично колінеарними (це дозволяє перейти від векторів до їхніх проєкцій на напрямок сумарного поля), а внесок від окремої зони зменшується зі зростанням її номеру (рис. 8.3.2), що дозволяє обмежитися в нескінченній сумі (8.3.18) скінченною кількістю доданків.

Як уже відзначалося, внесок від сусідніх зон Френеля протифазний. Вважатимемо

внески до сумарного поля від зон з непарними номерами додатними, відповідно внески від зон з парними номерами – від'ємними. Тоді ряд (8.3.18) можна подати у формі:

$$E_0 = \frac{E_1}{2} + \left(\frac{E_1 + E_3}{2} - E_2\right) + \left(\frac{E_3 + E_5}{2} - E_4\right) + \ldots \qquad (8.3.19)$$

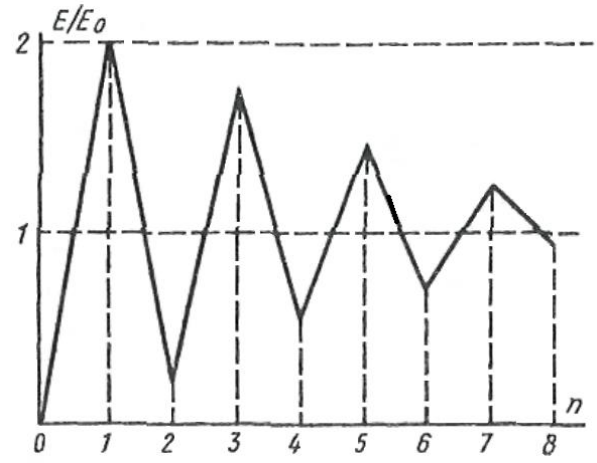


Рис. 8.3.2. Внесок перших n зон Френеля до електричного поля в точці спостереження (E_0 – сумарне поле в точці спостереження)

В круглих дужках стоять різниці (зі знаком мінус) між внеском від зон з парними номерами і середнім арифметичним від внесків сусідніх зон з непарними номерами. Очевидно, ці різниці значно менші від першого доданку в ряді (8.3.19). Отже, в нульовому наближенні можна вважати, що $E_0 \approx E_1/2$. Для точнішого розрахунку значення поля E_0 в ряді (8.3.19) треба враховувати наступні доданки.

На практиці звичайно обмежуються внеском перших восьми зон Френеля, що забезпечує точність розрахунку 16%. Тоді межею суттєвої області можна вважати межу області простору, обмежену зовнішніми межами восьмих зон Френеля, побудованих на різних віддалях від передавача.

8.3.5. Форма та розміри суттєвої області

Як уже вказувалося, аналізована задача є аксіально-симетричною щодо осі *AB* (рис. 8.3.1), тому суттєва область також є аксіально-симетричною. Розглянемо переріз цієї області, що вміщує вісь симетрії, і з'ясуємо, яку форму має межа цього перерізу.

Розглянемо загальний випадок зони Френеля з номером *n*.

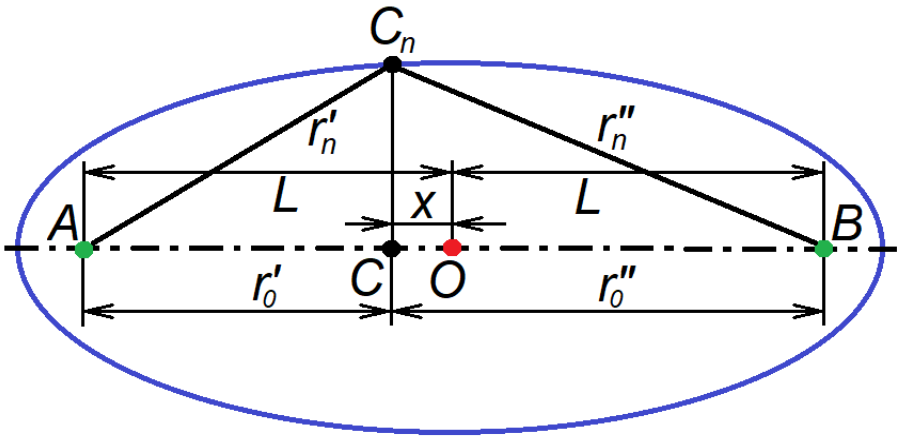


Рис. 8.3.3. До розрахунку форми суттєвої області

З формули (8.3.11)

$$\sqrt{(r_0')^2 + R_n^2} + \sqrt{(r_0'')^2 + R_n^2} - r_0' - r_0'' = n\lambda/2$$

випливає, що для всіх можливих точок C_n сума довжин відрізків AC_n та C_nB зберігається:

$$\sqrt{(r_0')^2 + R_n^2} + \sqrt{(r_0'')^2 + R_n^2} = r_0' + r_0'' + n\lambda/2 = 2L + n\lambda/2$$

Як відомо, геометричне місце точок, сума віддалей яких до деяких двох точок стала – це еліпс, а згадані дві точки є його фокусами. Таким чином, переріз суттєвої області, що вміщує вісь AB – це еліпс, у фокусах якого розміщені передавач і приймач.

Цей самий результат можна отримати з рівняння (8.3.11) більш формально. Нехай довжина відрізку AB дорівнює $2L$, а точка O є серединою цього відрізку (рис. 8.3.3). Нехай x – віддаль між точками O і C. Тоді

$$r_0' + r_0'' = 2L, \quad r_0' = L - x, \quad r_0'' = L + x, \tag{8.3.20}$$

і рівняння (8.3.11) набуває вигляду

$$\sqrt{(L-x)^2 + R_n^2} + \sqrt{(L+x)^2 + R_n^2} = 2L + n\lambda/2. \tag{8.3.21}$$

Тепер x виступає в ролі абсциси, R_n – у ролі ординати, а початок координат збігається з точкою O.

Піднісши до квадрату обидві частини рівняння (8.3.21) і спростивши, отримаємо:

$$\left(\sqrt{L^2 + x^2 + R_n^2 - 2Lx} + \sqrt{L^2 + x^2 + R_n^2 + 2Lx}\right)^2 = (2L + n\lambda/2)^2;$$

$$\begin{gathered} L^2 + x^2 + R_n^2 - 2Lx + L^2 + x^2 + R_n^2 + 2Lx + \\ +2\sqrt{\left(L^2 + x^2 + R_n^2\right)^2 - 4L^2x^2} = 4(L + n\lambda/4)^2; \\ \sqrt{\left(L^2 + x^2 + R_n^2\right)^2 - 4L^2x^2} = 2(L + n\lambda/4)^2 - \left(L^2 + x^2 + R_n^2\right). \end{gathered} \tag{8.3.22}$$

Ще раз піднісши до квадрату обидві частини й спростивши, отримаємо:

$$\left(L^2 + x^2 + R_n^2\right)^2 - 4L^2x^2 = 4(L + n\lambda/4)^4 + \left(L^2 + x^2 + R_n^2\right)^2 - 4(L + n\lambda/4)^2\left(L^2 + x^2 + R_n^2\right);$$

$$(L + n\lambda/4)^2\left(x^2 + R_n^2\right) = (L + n\lambda/4)^4 + L^2x^2 - L^2(L + n\lambda/4)^2;$$

$$\left[L^2 + n\lambda L/2 + (n\lambda/4)^2\right]\left(x^2 + R_n^2\right) = (L + n\lambda/4)^4 + L^2x^2 - L^2(L + n\lambda/4)^2;$$

$$x^2\left[n\lambda L/2 + (n\lambda/4)^2\right] + R_n^2(L + n\lambda/4)^2 = (L + n\lambda/4)^4 - L^2(L + n\lambda/4)^2;$$

$$x^2(n\lambda/4)(2L + n\lambda/4) + R_n^2(L + n\lambda/4)^2 = (L + n\lambda/4)^2(n\lambda/4)(2L + n\lambda/4) \tag{8.3.23}$$

Нарешті, поділивши рівняння (8.3.23) на його праву частину, отримує відоме з аналітичної геометрії рівняння еліпса:

$$\frac{x^2}{(L + n\lambda/4)^2} + \frac{R_n^2}{(n\lambda/4)(2L + n\lambda/4)} = 1, \tag{8.3.24}$$

або, після заміни R_n на y,

$$\left(\frac{x}{a_n}\right)^2 + \left(\frac{y}{b_n}\right)^2 = 1 \tag{8.3.24 а}$$

з півосями

$$a_n = L + n\lambda/4, \quad b_n = \sqrt{(n\lambda/4)(2L + n\lambda/4)}. \tag{8.3.25}$$

Зі зробленого раніше припущення $L \gg \lambda$ випливає нерівність $a_n \gg b_n$.

Якщо у формулі для b_n відкинути доданки вищих порядків за λ, приходимо до раніше отриманої формули (8.3.17).

Таким чином, суттєва область є еліпсоїдом обертання з півосями a_n, b_n (8.3.25). У цих формулах, як уже згадувалося, прийнято обирати n=8. Видно, що зі збільшенням λ довжина малої півосі b_n також збільшується.

Візьмемо $L = 50$ км, $\lambda = 1$ м[3], n=8. Тоді $a_8 \approx 50$ км, $b_8 \approx 2\sqrt{\lambda L} \approx 0.5$ км. Це означає, що при помірних висотах розміщення передавальної та приймальної антен (десятки метрів) і довжинах хвиль від метрів і більше треба обов'язково враховувати вплив на поширення радіохвиль земної поверхні, оскільки вона потрапляє до суттєвої області.

Отримані результати справедливі, строго кажучи, лише тоді, коли діаграми спрямованості передавальної та приймальної антен є ізотропними[4]. В протилежному випадку форма суттєвої області ускладняється й залежить від названих діаграм спрямованості.

Контрольні питання до підрозділу 8.3

1. Що таке зона Френеля?
2. Що таке суттєва зона формування поля? Яку форму вона має?
3. Чим визначається поперечний розмір суттєвої зони формування поля?
4. Як можна експериментально визначити поперечний розмір суттєвої зони формування поля?

Задачі до підрозділу 8.3

8.3.1. Віддаль між випромінювальною та приймальною антенами, що перебувають у вакуумі, складає 20 км. Частота хвилі – 1 ГГц. Знайдіть внутрішній та зовнішній радіус четвертої зони Френеля, розташованої на віддалі 5 км від передавальної антени.

8.3.2. Віддаль між випромінювальною та приймальною антенами, що перебувають у вакуумі, складає 100 км. Частота хвилі – 1.5 ГГц. Знайти поздовжній та поперечний розміри суттєвої області, вважаючи, що остання обмежується 10-ю зоною Френеля.

---

[3] Нагадаємо, що ця довжина хвилі відповідає межі метрового та дециметрового діапазону (див. Вступ).

[4] Якщо, наприклад, приймальна антена має дуже вузьку головну пелюстку діаграми спрямованості, орієнтовану на передавач, вона, очевидно, сприймає лише випромінювання з першої зони Френеля. Аналогічна діаграма спрямованості передавальної антени, орієнтована на приймач, приведе то того, що сторинні хвилі збуджуватимуться тільки в першій зоні Френеля.

8.4. Збудження земної хвилі високо піднятою антеною

З оцінки, зробленої наприкінці п. 8.3.4, випливає, що для не дуже коротких хвиль (з довжинами понад 1м) і розумних висот розташування передавальної антени (десятки метрів) до суттєвої області обов'язково потрапляє частина земної поверхні. Хвилю, яка в процесі поширення зазнає значного впливу земної поверхні, називають земною хвилею.

Розрахунок збудження земної хвилі в загальному випадку є складною й громіздкою задачею, яку, однак, вдається спростити для деяких крайніх випадків. Зокрема, такими випадками є високо підняті випромінювачі, висота яких h задовольняє умові $h \gg \lambda$, і низько розміщені випромінювачі, для яких виконано протилежну умову $h \ll \lambda$. Оскільки на практиці висота h зазвичай складає десятки метрів, то перша з цих умов може вважатися виконаною при $\lambda \le 10м$, друга – при $\lambda \ge 100м$.

У цьому підрозділі ми розглянемо збудження земної хвилі високо піднятими антенами.

8.4.1. Способи розрахунку поля земної хвилі

Відповідно до принципу Гюйгенса – Френеля поле земної хвилі можна розглядати як суперпозицію полів вторинних джерел у повітрі та на поверхні Землі, які збуджуються первинним джерелом. Треба брати до уваги скінченну провідність Землі, яка призводить до поглинання електромагнітних хвиль. Різні частини поверхні Землі – гори, рослинний покрив, водні поверхні – відбивають і поглинають електромагнітні хвилі по-різному.

Сферичність земної поверхні приводить до того, що в точки цієї поверхні на достатній віддалі від випромінювача хвилі можуть потрапити лише внаслідок дифракції (огинання перешкоди). Звідси зрозуміло, що в загальному випадку поширення земної хвилі неможливо описати в наближенні геометричної (променевої оптики).

В результаті задача про поширення земної хвилі стає надзвичайно складною, і розв'язати її аналітично вдається лише за рахунок тих чи інших спрощень моделі. Наприклад, було отримано аналітичний розв'язок задачі про дифракцію для моделі сферичної Землі з однорідними електричними параметрами та однорідною атмосферою. Але навіть для такої значно спрощеної моделі отриманий результат має вигляд нескінченної суми деяких спеціальних функцій (так званих функцій Ейрі). Скільки-небудь реалістичну модель вдається проаналізувати лише числовими методами.

Тому на практиці зазвичай користуються розв'язками, справедливими лише в деяких обмежених областях і заснованих на максимально спрощених моделях. В деяких випадках критерієм застосування тих чи інших моделей виступає висота h підняття випромінювальної антени над земною поверхнею і віддаль між точками випромінювання й приймання радіохвилі.

За виконання умови $h \gg \lambda$ говорять про високо підняті антени. Ця модель добре працює для ультракоротких хвиль. Навпаки, модель низько розміщених антен ($h \ll \lambda$) застосовна для середніх та довгих хвиль.

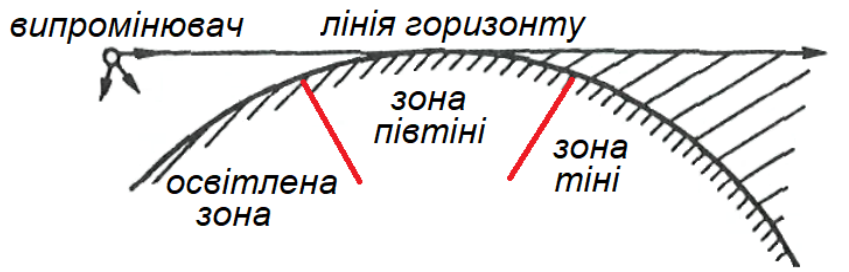


Рис. 8.4.1. Зони з різною освітленістю від високо піднятої антени

Для великих віддалей між випромінювачем і приймачем і високо піднятих антен на

поверхні Землі розрізняють освітлену зону, зону півтіні та зону тіні (рис. 8.4.1). Розрахунок електромагнітного поля в освітленій зоні зводиться до врахування інтерференції між прямою хвилею та хвилею, відбитою від поверхні Землі. В зону тіні хвилі можуть потрапити лише в результаті дифракції, тому в цій області розрахунок помітно ускладняється. Те саме стосується зони півтіні.

Для низько розміщених антен освітлена зона відсутня. На малих віддалях від такої антени кривина земної поверхні несуттєва, і застосовна модель плоскої Землі. В цій області розрахунок є також є порівняно простим (див. вище п. 7.3.3, де описано модель та результати такого розрахунку для півхвильового вібратора).

Оцінимо віддаль r_s між випромінювачем сигналу та точкою його приймання вздовж поверхні Землі (рис. 8.4.2; індекс – від англ. surface, поверхня).

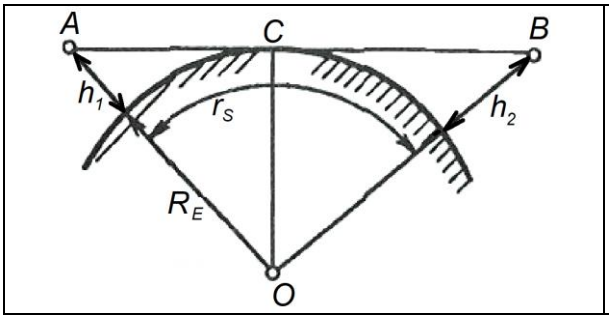


Рис. 8.4.2. До розрахунку віддалі між випромінювачем сигналу та приймачем уздовж поверхні Землі

Висоти h_1, h_2 підняття передавальної та приймальної антен з великим запасом задовольняють умові $h_1, h_2 \ll R_E$, де $R_E \approx 6400$ км – радіус Землі (індекс – від англ. Earth, Земля). Тому можна вважати, що $r_s \approx AB = AC + CB$. Застосувавши теорему Піфагора до трикутників ACO та CBO, маємо:

$$r_s = \sqrt{(R_E + h_1)^2 - R_E^2} + \sqrt{(R_E + h_2)^2 - R_E^2} = \\ = \sqrt{2h_1R_E + h_1^2} + \sqrt{2h_2R_E + h_2^2} \approx \sqrt{2R_E}\left(\sqrt{h_1} + \sqrt{h_2}\right), \tag{8.4.1}$$

або

$$r_s[км] = 3.57\left(\sqrt{h_1[м]} + \sqrt{h_2[м]}\right). \tag{8.4.2}$$

Взявши для прикладу $h_1 = h_2 = 25м$, отримаємо $r_s = 35.7км$.

Очевидно, віддаль вздовж земної поверхні між точками A і C (рис. 8.4.2) можна визначити як $\sqrt{2h_1R_E}$. Нехай тепер r це віддаль уздовж поверхні Землі між передавачем і приймачем. Тоді можна оцінити, в якій із характерних зон освітленості (див. рис. 8.4.1) знаходиться приймач: $r < \sqrt{2h_1R_E}$ – освітлена зона, $r \approx \sqrt{2h_1R_E}$ – зона півтіні, $r > \sqrt{2h_1R_E}$ – зона тіні.

8.4.2. Розрахунок впливу Землі шляхом урахування відбитих від неї хвиль

Електричне поле, створене випромінювальною антеною в точці спостереження у вільному просторі, задається формулою (8.3.8):

$$E_m = \frac{\sqrt{60PG}}{r}V.$$

Вплив поверхні Землі на це поле формально зводиться до знаходження відповідного множника ослаблення V.

Один із способів розрахунку впливу Землі (див. вище п. 7.3.3) полягає в урахуванні фіктивного джерела випромінювання, розташованого дзеркально до реального джерела щодо поверхні Землі (рис. 8.4.3). Зміну амплітуди й фази сигналу в результаті відбиття від поверхні

зі скінченною провідністю враховують за рахунок введення комплексного коефіцієнту відбиття $R = |R|\exp(i\theta)$ – так званого коефіцієнту Френеля. Він залежить від характеру земної поверхні, кута падіння Δ (рис. 8.4.3) та поляризації хвилі. Якісні залежності $|R|$ та θ від Δ для *s*-поляризованих (електричне поле перпендикулярне до площини падіння) та *p*-поляризованих (електричне поле лежить у площині падіння) хвиль представлені на рис. 8.4.4. Видно, що для будь-яких кутів падіння $R_{\perp} > R_{//}$.

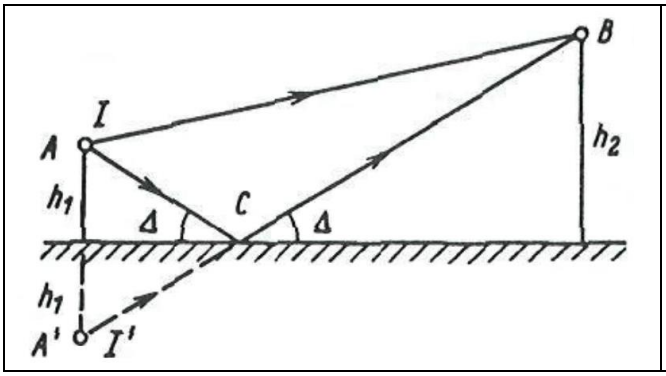


Рис. 8.4.3. До розрахунку впливу поверхні Землі через урахування відбитих від неї хвиль

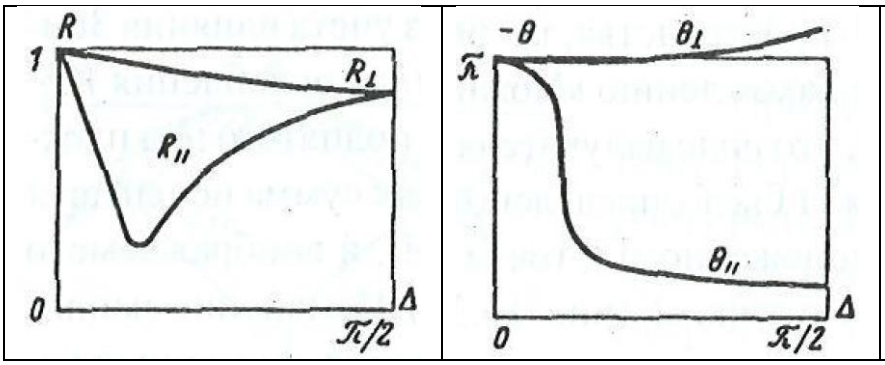


Рис. 8.4.4. Залежності модуля та фази коефіцієнту відбиття від частково провідної поверхні залежно від кута падіння для s-поляризованих ($R_{\perp}$, $\theta_{\perp}$) та p-поляризованих ($R_{//}$, $\theta_{//}$) хвиль

Якщо (відносна) діелектрична проникність ε речовини, від якої відбивається електромагнітна хвиля, значно більша за одиницю ($\varepsilon \gg 1$), то ця речовина поводить себе як ідеальний провідник. При цьому $|R| \to 1$.

Електричні параметри деяких видів земної поверхні в радіодіапазоні наведені в табл. 8.4.1. Видно, що найбільшу діелектричну проникність має морська та річкова вода, найбільшу провідність – морська вода. Найменшу діелектричну проникність та провідність має сніг.

Таблиця 8.4.1

Вид земної поверхні	*Діелектрична проникність* ε	*Питома провідність* σ *[См/м]*
Морська вода	80	1-6
Вода річок та озер	80	10^{-3}-10^{-2}
Волога земля	10-30	$3\cdot10^{-3}$-$3\cdot10^{-2}$
Суха земля	3-6	10^{-5}-$5\cdot10^{-3}$
Мерзла земля	3-6	10^{-3}-10^{-2}
Сніг при -10ºС	1	10^{-6}
Крига при -10ºС	4-5	10^{-2}-10^{-1}

8.4.3. Суттєва область відбиття

В моделі променевої оптики відбиття окремого променю відбивається від певної точки поверхні. Врахування процесів інтерференції хвиль, як ми бачили, породжує поняття суттєвої області. Переріз суттєвої області (між уявним джерелом A' і точкою спостереження B) поверхнею Землі показує, яка область поверхні впливає на відбиття хвилі з певним напрямком

хвильового вектора (рис. 8.4.5). Цю область називають суттєвою областю відбиття.

Розрахунок розмірів суттєвої області відбиття знано спрощується за умови, коли $h_1 = h_2 = h$, тобто точка C_n – центр області земної поверхні, від якої відбивається хвиля (рис. 8.4.5) – у цьому випадку збігається з центром еліпса. Тоді розмір b_n збігається з малою піввіссю суттєвої області (див. формулу (8.3.25)):

$$b_n \approx \sqrt{n\lambda L/2}\,. \qquad (8.4.3)$$

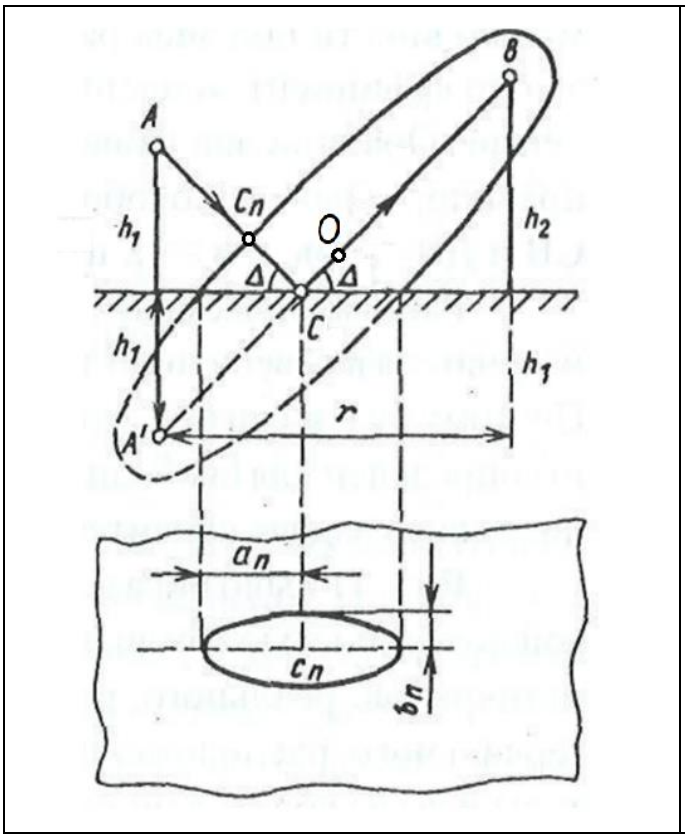


Рис. 8.4.5. Область поверхні, що впливає на відбиття хвилі певного напрямку – суттєва область відбиття (пор. з рис. 8.3.3)

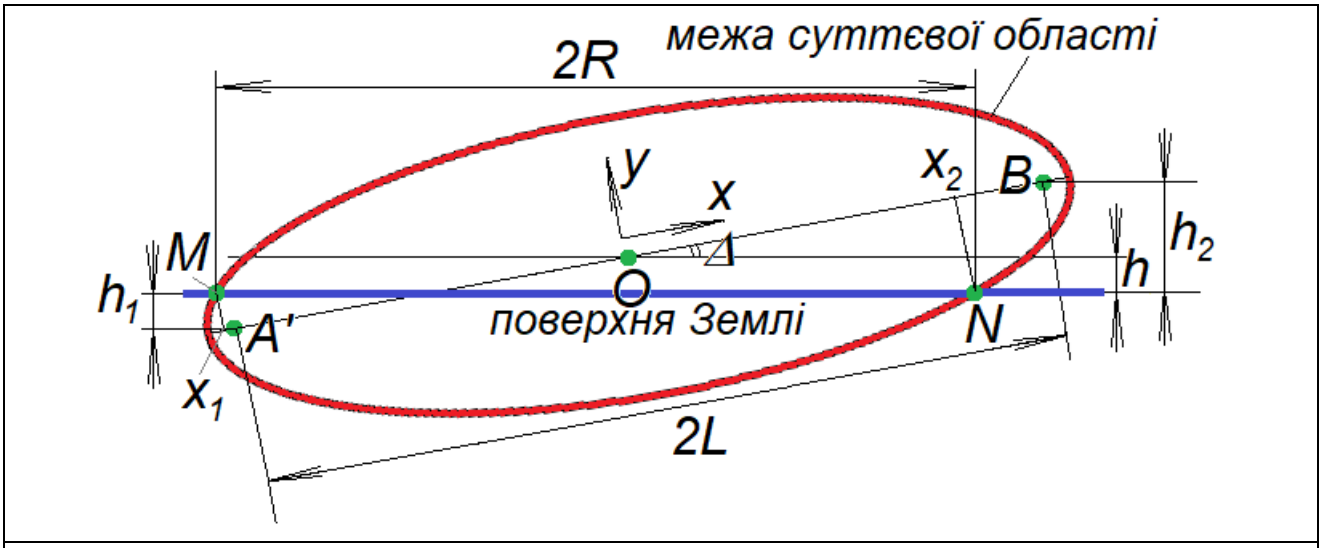


Рис. 8.4.6. Віддаль $2L$ між уявним джерелом відбитої хвилі та приймачем і довжина суттєвої області відбиття $2R$

Межі суттєвої області відбиття M і N належать водночас до межі вертикального перерізу суттєвої області через точку O (рис. 2.4.6), яка в координатах x, y має вигляд

$$\left(\frac{x}{a_n}\right)^2+\left(\frac{y}{b_n}\right)^2=1$$

(див. рівняння (8.3.24 а)), і того самого перерізу земної поверхні, яка в тих самих координатах відповідає рівнянню $y=-x\,\mathrm{tg}\,\Delta-h/\cos\Delta=-\left(x\sin\Delta+h\right)/\cos\Delta$. $h=\left(h_2-h_1\right)/2$ (позначення зрозумілі з рис. 2.4.6). Тоді x-координати точок M і N задовольняють рівняння

$$\left(\frac{x}{a_n}\right)^2+\left(\frac{x\sin\Delta+h}{b_n\cos\Delta}\right)^2=1,$$

звідки

$$\left(\frac{x}{a_n}\right)^2+\left(\frac{x\sin\Delta+h}{b_n\cos\Delta}\right)^2=1,$$

або

$$\left(\frac{1}{a_n^2}+\frac{\mathrm{tg}^2\Delta}{b_n^2}\right)x^2+\frac{2h\sin\Delta}{b_n^2\cos^2\Delta}x+\left(\frac{h^2}{b_n^2\cos^2\Delta}-1\right)=0,$$

або

$$\left(b_n^2\cos^2\Delta+a_n^2\sin^2\Delta\right)x^2+2ha_n^2x\sin\Delta+a_n^2\left(h^2-b_n^2\cos^2\Delta\right)=0.$$

Корені цього рівняння:

$$x_{1,2}=\frac{-ha_n^2\sin\Delta\pm\sqrt{h^2a_n^4\sin^2\Delta-a_n^2\left(b_n^2\cos^2\Delta+a_n^2\sin^2\Delta\right)\left(h^2-b_n^2\cos^2\Delta\right)}}{b_n^2\cos^2\Delta+a_n^2\sin^2\Delta}=$$

$$=\frac{a_n}{b_n^2\cos^2\Delta+a_n^2\sin^2\Delta}\left[-ha_n\sin\Delta\pm\sqrt{h^2a_n^2\sin^2\Delta-\left(b_n^2\cos^2\Delta+a_n^2\sin^2\Delta\right)\left(h^2-b_n^2\cos^2\Delta\right)}\right]=$$

$$=\frac{a_n}{b_n^2\cos^2\Delta+a_n^2\sin^2\Delta}\left(-ha_n\sin\Delta\pm b_n\cos\Delta\sqrt{b_n^2\cos^2\Delta-h^2+a_n^2\sin^2\Delta}\right). \qquad (8.4.4)$$

З рис. 8.4.6 видно, що

$$R=\frac{x_2-x_1}{2\cos\Delta}=\frac{a_nb_n\sqrt{b_n^2\cos^2\Delta-h^2+a_n^2\sin^2\Delta}}{b_n^2\cos^2\Delta+a_n^2\sin^2\Delta}. \qquad (8.4.5)$$

Тоді з урахуванням (8.3.25) та умови $\lambda\ll L$, а також того, що $H\ll R$, де $H=\left(h_1+h_2\right)/2$, можна записати:

$$\Delta\approx H/R\ll 1,$$

$$R\approx\frac{a_nb_n\sqrt{b_n^2-h^2+a_n^2H^2/R^2}}{b_n^2+a_n^2H^2/R^2}\approx\frac{\left(L+n\lambda/4\right)\sqrt{n\lambda L/2}\sqrt{n\lambda L/2-h^2+H^2}}{n\lambda L/2+H^2}\approx$$

$$\approx\left(L+n\lambda/4\right)\left(1+\frac{H^2-h^2}{n\lambda L/4}\right)\left(1-\frac{H^2}{n\lambda L/2}\right)\approx\left(L+n\lambda/4\right)\left(1-\frac{H^2+h^2}{n\lambda L/4}\right)\approx$$

$$\approx L+\frac{n\lambda}{4}-\frac{2\left(h_1^2+h_2^2\right)}{n\lambda} \qquad (8.4.6)$$

При отриманні формул (8.4.3) та (8.4.6) ми, по суті, користувалися моделлю плоскої Землі. Спробуємо отримати кількісну умову її застосовності.

На рис. 8.4.7 точка A, як і раніше, відповідає положенню передавача ($AC=h$), чорним кольором показана земна поверхня, синім – умовна поверхня плоскої землі. Точка B на поверхні Землі відповідає розміщенню приймача. За віддаль між передавачем і приймачем приймемо довжину відрізка CD ($CD\approx r$). Відмінність між плоскою та криволінійною поверхнями в точці розташування приймача можна характеризувати довжиною δ відрізку BD, яка може

бути знайдена за теоремою Піфагора з трикутника OCD:

$$R_E^2 + r^2 = \left(R_E + \delta\right)^2,$$

звідки

$$\delta = \sqrt{R_E^2 + r^2} - R_E = \frac{R_E^2 + r^2 - R_E^2}{\sqrt{R_E^2 + r^2} + R_E} \approx \frac{r^2}{2R_E} \qquad (8.4.7)$$

(враховано, що $r \ll R_E$, де, як і раніше, R_E – радіус Землі).

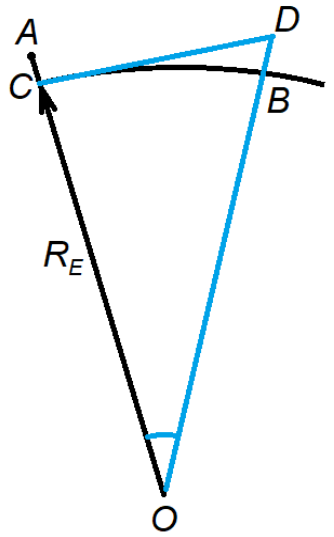


Рис. 8.4.7. До знаходження умови застосовності моделі плоскої Землі

За критерій застосовності моделі плоскої Землі приймемо умову $\delta \ll h$ (h – віддаль між токами A й C, див. рис. 8.4.7). Він буде задовольнятися, якщо з запасом хоча б на півпорядку справджується нерівність

$$r < \sqrt{2hR_E}\,. \qquad (8.4.8)$$

Візьмемо, як і раніше, $h = 25$ м. Тоді з (8.4.8) отримуємо: $r < 18$ км. Отже, застосовність формул (8.4.3) та (8.4.6) досить обмежена.

Тепер можна оцінити розміри суттєвої області відбиття за формулами (8.4.3) та (8.4.6). Для оцінки візьмемо $r \approx 2L = 6$ км, $\lambda = 1$ м, $n = 8$. Отримаємо:

$$b_8 = 2\sqrt{1 \cdot 6000} \approx 150 \text{ м};$$

$$R = 3\sqrt{1 - \frac{2 \cdot 625}{8 \cdot 1 \cdot 3000}} \approx 2.9 \text{ км}.$$

Бачимо, що поперечний (щодо траси сигналу) розмір суттєвої області відбиття вийшов на порядок меншим за поздовжній.

Фактично для обраних параметрів суттєва область відбиття майже збігається з перерізом еліпсоїда обертання, що визначає межі суттєвої зони, через велику вісь. Тому в цьому випадку $R \approx L$ (рис. 8.4.6).

Описаний підхід коректний лише в тому випадку, коли в суттєвій області відбиття коефіцієнт відбиття хвилі можна вважати практично незмінним.

8.4.4. Електромагнітне поле в освітленій зоні без урахування кривини Землі

Перейдемо до безпосереднього розрахунку електромагнітного поля, створеного високо піднятою антеною, в точці спостереження, яка знаходиться в освітленій зоні. Кривину Землі поки що не враховуватимемо.

Миттєве значення довільної компоненти електричного поля E_E (індекс – від англ.

Earth, Земля) земної хвилі в точці спостереження *B* можна подати як суму полів прямої E_{dir} (індекс – від англ. direct, прямий) та відбитої E_{ref} (індекс – від англ. reflected, відбитий) хвиль (рис. 8.4.8):

$$E_E = E_{dir} + E_{ref} \,. \tag{8.4.9}$$

Амплітуда електричного поля, створеного прямою хвилею в точці спостереження, задається формулою (8.3.8), яка в припущенні про відсутність поглинання й розсіювання хвиль (тобто вважається, що множник ослаблення $V = 1$) має вигляд :

$$E_m = \frac{\sqrt{60 P_r G}}{r_1} \,. \tag{8.4.10}$$

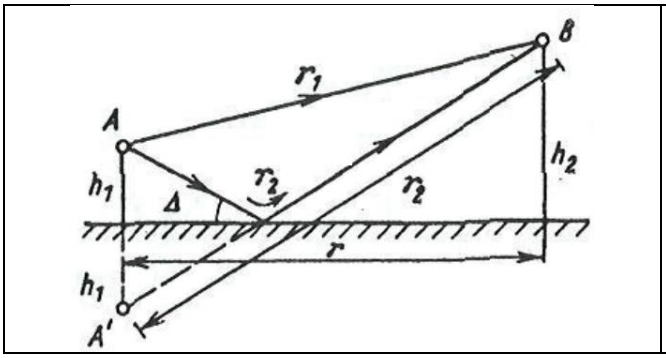


Рис. 8.4.8. До розрахунку поля в освітленій зоні без урахування кривини Землі

Врахувавши набіг фази на шляху $AB = r_1$, комплексну амплітуду в точці спостереження можна подати у формі

$$E_{dir} = \frac{\sqrt{60 P_r K_{dir}}}{r_1} \exp\left(-i\frac{2\pi r_1}{\lambda}\right) \tag{8.4.11}$$

При записі поля відбитої хвилі, крім очевидної заміни r_1 на r_2, треба також врахувати комплексний коефіцієнт відбиття хвилі $\mathrm{R}\exp(i\theta)$ від поверхні Землі:

$$E_{ref} = \frac{\sqrt{60 P_r K_{dir}}}{r_2} R \exp\left(i\theta - i\frac{2\pi r_2}{\lambda}\right) \tag{8.4.12}$$

(тут враховано, що в силу умови $r \gg h_1, h_2$ напрямки випромінювання прямої та відбитої хвиль випромінювальною антеною майже однакові, тому коефіцієнт спрямованої дії антени K_{dir} для них також можна вважати однаковим). З тієї ж причини при знаходженні амплітуд хвиль вважатимемо, що $r_2 \approx r_1 = r$. В результаті після підстановки (8.4.11)- (8.4.12) до (8.4.9) отримаємо:

$$E_E = \frac{\sqrt{60 P_r K_{dir}}}{r} \exp\left(-i2\pi\frac{r_1}{\lambda}\right)\left[1 + R\exp\left(i\theta - i2\pi\frac{r_2 - r_1}{\lambda}\right)\right]. \tag{8.4.13}$$

Множник у квадратних дужках у правій частині (8.4.13) – це і є множник ослаблення сигналу за рахунок впливу земної поверхні, або інтерференційний множник:

$$V = 1 + R\exp\left(i\theta - i2\pi\frac{r_2 - r_1}{\lambda}\right). \tag{8.4.14}$$

При зміні різниці ходу $r_2 - r_1 \equiv \Delta r$ прямої та відбитої хвиль його абсолютна величина

$$|V| = \sqrt{\left[1 + R\cos\left(\theta - 2\pi\frac{\Delta r}{\lambda}\right)\right]^2 + R\sin^2\left(\theta - 2\pi\frac{\Delta r}{\lambda}\right)} =$$
$$= \sqrt{1 + R^2 + 2R\cos\left(\theta - 2\pi\frac{\Delta r}{\lambda}\right)} \qquad (8.4.15)$$

змінюється в межах від $1-R$ до $1+R$.

Різницю ходу Δr можна подати через параметри моделі (рис. 8.4.9):

$$\Delta r = r_2 - r_1 = \sqrt{r^2 + (h_1 + h_2)^2} - \sqrt{r^2 + (h_1 - h_2)^2} =$$
$$= \frac{r^2 + (h_1 + h_2)^2 - \left[r^2 + (h_1 - h_2)^2\right]}{\sqrt{r^2 + (h_1 + h_2)^2} + \sqrt{r^2 + (h_1 - h_2)^2}} \approx \frac{4h_1h_2}{2r} = \frac{2h_1h_2}{r}. \qquad (8.4.16)$$

Видно, що різниця ходу Δr прямує до нуля при $r \to \infty$.

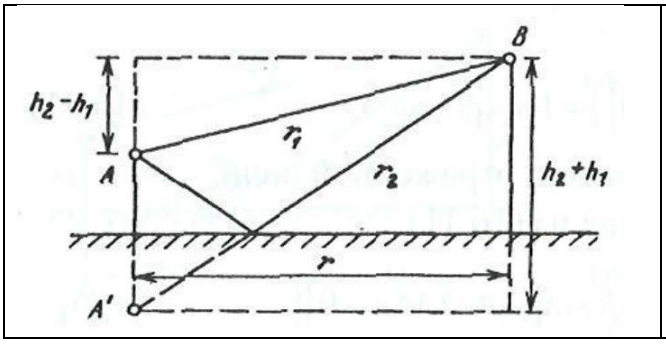


Рис. 8.4.9. До розрахунку різниці ходу прямої та відбитої хвиль

Об'єднуючи формули (8.4.12) та (8.4.15)-(8.4.16), отримуємо:

$$|E_E| = \frac{\sqrt{60P_rK_{dir}}}{r}\sqrt{1 + R^2 + 2R\cos\left(\theta - 4\pi\frac{h_1h_2}{\lambda r}\right)} \qquad (8.4.17)$$

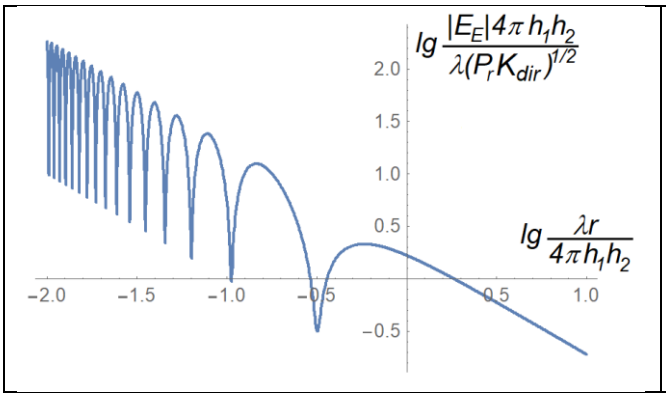


Рис. 8.4.10. Графік залежності амплітуди електромагнітного поля в точці спостереження в освітленій зоні від віддалі r між передавачем і приймачем (у двічі логарифмічному масштабі) для $R = 0.9$, $\theta = 0$

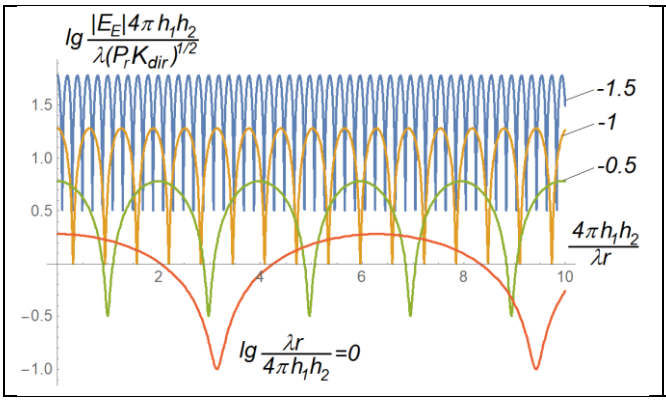


Рис. 8.4.11. Графіки залежності амплітуди електромагнітного поля в точці спостереження в освітленій зоні (у логарифмічному масштабі) від висоти підняття приймача h_2 для $R = 0.9$, $\theta = 0$

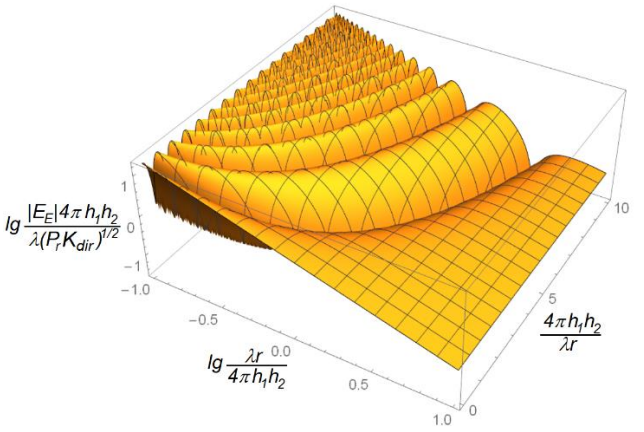


Рис. 8.4.12. Діаграма залежності амплітуди електромагнітного поля в точці спостереження в освітленій зоні (у логарифмічному масштабі) від нормованої відстані між передавачем і приймачем r та від нормованої висоти підняття приймача h_2

Залежності нормованого значення $|E_E|$ від нормованої відстані між передавачем і приймачем r та від нормованої висоти підняття приймача h_2 подані на рис. 8.4.10-8.4.12. Обом залежностям притаманні осциляції. В залежності від h_2 вони періодичні, їхні амплітуда та період залежать від r (рис. 8.4.11). В залежності від r вони зникають при великих r, оскільки при цьому, як уже зазначалося, $\Delta r \to 0$. Межу зникнення осциляцій r_{max} знаходимо з умови

$$\left(\theta - 4\pi\frac{h_1h_2}{\lambda r}\right)\Bigg|_{r\to\infty} - \left(\theta - 4\pi\frac{h_1h_2}{\lambda r}\right)\Bigg|_{r_{max}} = \pi,$$

звідки

$$r_{max} = \frac{4h_1h_2}{\lambda}. \qquad (8.4.18)$$

Взявши для оцінки, як і раніше, $h_1 = h_2 = 25$ м, $\lambda = 1$ м, отримуємо: $r_{max} \approx 2.5$ км.

8.4.5. Вплив кривини земної поверхні на електромагнітне поле в освітленій зоні

Раніше було показано, що модель плоскої Землі виявляється застосовною за виконання умови (8.4.8), тобто на відстані щонайбільше кількох десятків кілометрів. Радіус же освітленої зони (див. п. 8.4.2) принаймні вдвічі більший. Тому в багатьох випадках кривину земної поверхні необхідно враховувати.

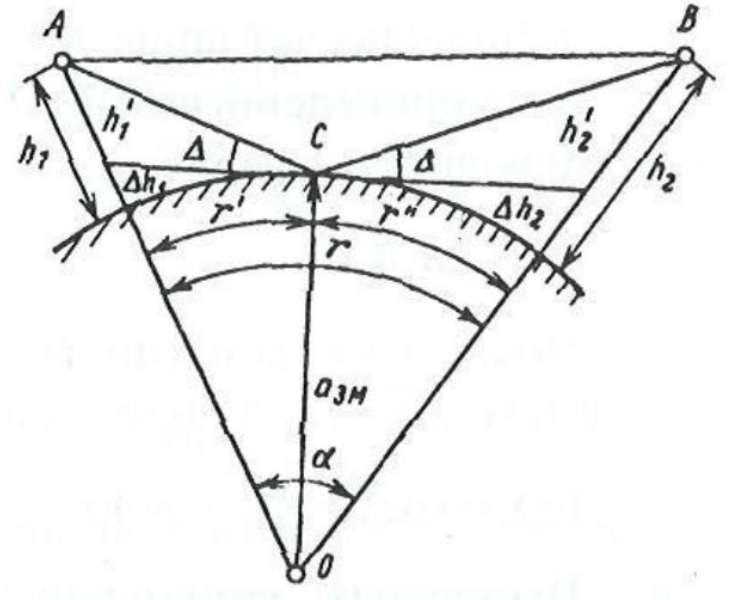


Рис. 8.4.13. До оцінки впливу кривини земної поверхні на поле в освітленій зоні

На жаль, формули, що враховують сферичність поверхні Землі, досить громіздкі, тому ми обмежимося обговоренням цієї проблеми на якісному рівні.

Порівняємо між собою рис. 8.4.13, де показано траєкторії прямої (АВ) та відбитої (СВ) хвиль з урахуванням сферичності поверхні Землі, з рис. 8.4.9, який відповідає моделі плоскої Землі. Видно, що врахування кривини земної поверхні приводить до зміни довжин траєкторій прямої та відбитої хвиль. В результаті змінюється й різниця ходу Δr, яка впливає на інтерференційний множник.

Точкове джерело породжує розбіжні (конічні) пучки променів. При відбитті від плоскої поверхні кут розходження такого променю не змінюється, при відбитті від сферичної поверхні він зростає (рис. 8.4.14; як відомо, відбиття від опуклого сферичного дзеркала приводить до розходження первісно паралельних променів). Ця обставина приводить до додаткового ослаблення відбитого променю в точці спостереження.

В результаті врахування названих факторів поле в точці спостереження може помітно відрізнятися від того, що передбачає модель плоскої Землі.

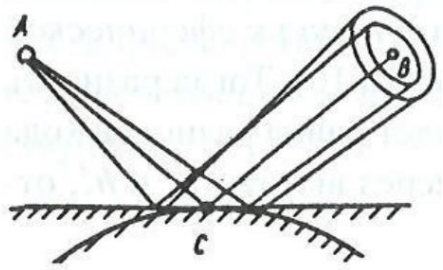


Рис. 8.4.14. Вплив кривини поверхні на відбиття від неї розбіжного пучка променів

На закінчення слід указати, що всі розрахунки, наведені в цьому підрозділі, виконані на основі врахування інтерференції прямої та відбитої хвиль. Ці хвилі можна розділити за умови, що кут Δ між горизонтом та відбитою хвилею (див. рис. 8.4.8) не дуже малий, а точніше, задовольняє умову

$$\sin\Delta \approx \frac{h_1 + h_2}{r} > \left(\frac{\lambda}{\pi R_E}\right)^{1/3}$$

де, як і раніше, $R_E = 6400$ км – радіус Землі. Взявши $\lambda = 1$ м, отримаємо: $\sin\Delta > 0.0037$.

8.4.6. Електромагнітне поле в зонах півтіні й тіні

Як уже вказувалося, електромагнітне поле випромінювання високо піднятої антени в зонах півтіні й тіні значною мірою визначається дифракційними ефектами і тому в принципі не може бути розраховане як сума прямої хвилі й хвилі, відбитої від Землі. Розрахунок в обох зонах виконується на основі формул строгої дифракційної теорії, які мають вигляд суми нескінченного ряду. Відмінність зони півтіні й тіні лише в тому, що в зоні тіні можна враховувати меншу кількість доданків цього ряду.

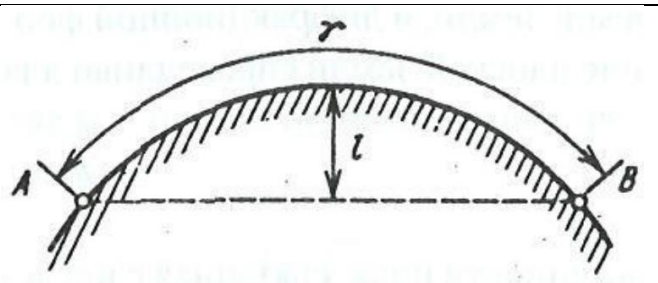


Рис. 8.4.15. До розрахунку поля передавальної антени в зонах півтіні й тіні

Оскільки розрахункові формули дуже громіздкі, знову обмежимося обговоренням ефекту на якісному рівні.

Опуклу земну поверхню між передавачем *A* і приймачем *B* (рис. 8.4.15) можна розглядати як перешкоду, яку огинає випромінювана передавачем хвиля. Характер такого огинання (дифракції) залежить від параметра l/λ, де l – висота кульового сегменту (рис. 8.4.15). Справді, зі збільшенням l усе більшу частину суттєвої зони між точками А та В заповнює опукла поверхня Землі. Скінченна провідність земного покрову приводить до суттєвого згасання земної хвилі на шляху її поширення.

При збільшенні віддалі r між передавачем і приймачем напруженість поля зменшується, перш за все, за рахунок природного розходження променів у пучку. Водночас збільшується висота кульового сегмента l, що призводить до зростання втрат на дифракційне огинання перешкоди. Нарешті, зі збільшенням r зростають омічні втрати в приповерхневому шарі земного покрову. В результаті дії цих трьох факторів залежність напруженості поля в точці спостереження від віддалі r має спадний характер, а швидкість цього спадання залежить від довжини хвилі λ.

Справді, при зростанні λ властивості земної поверхні все краще описуються моделлю ідеального провідника, тобто згасання земної хвилі зменшується. Зменшується також параметр l/λ, тобто зменшуються втрати на дифракційне огинання Землі. Тому можна чекати, що в зоні тіні при зростанні λ поле у фіксованій точці спостереження має збільшуватися.

Висловлені міркування ілюструє рис. 8.4.16, на якому представлені графіки залежності коефіцієнту ослаблення сигналу V від віддалі між передавачем та приймачем r для різних довжин хвилі λ. Представлені залежності охоплюють освітлену зону, зону півтіні та зону тіні.

Область осциляцій залежностей $V(r)$ відповідає освітленій зоні, а самі осциляції є наслідком інтерференції між прямою та відбитою від поверхні Землі хвилею. Видно, що розміри цієї області зростають зі зменшенням довжини хвилі λ, що добре узгоджується з отриманою раніше формулою (8.4.18).

Праворуч від освітленої зони знаходиться область, де залежності $V(r)$ монотонно спадають. За початок зони тіні на рис. 8.4.16 можна умовно прийняти точку $r = 80$ км, праворуч від якої величина $V(r)$ зростає зі збільшенням довжини хвилі. Видно, що в загоризонтній області швидкість спадання функції $V(r)$ збільшується зі зменшенням довжини хвилі. Іншими словами, чим більша довжина хвилі, тим на більшу віддаль за горизонт випромінювача вона поширюється.

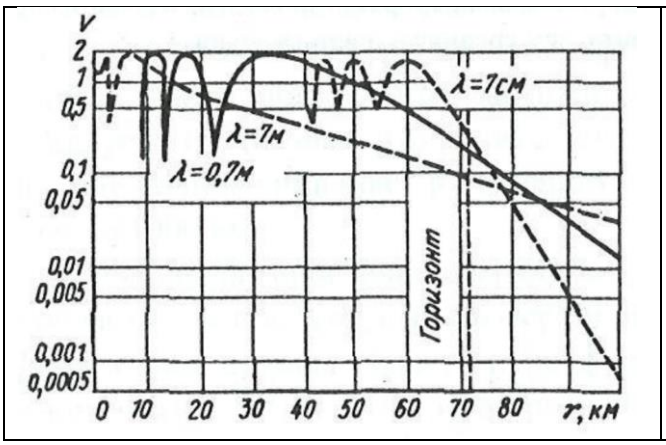


Рис. 8.4.16. Графіки залежності коефіцієнту ослаблення сигналу від віддалі між передавачем та приймачем (у напівлогарифмічному масштабі) для різних довжин хвилі й висоти підняття антени 400 м: розрахунок на основі теорії дифракції для освітленої зони, зони півтіні й зони тіні

Контрольні питання до підрозділу 8.4

1. Що розуміють під земною хвилею?
2. Коли поверхня Землі впливає на поширення радіохвиль?
3. Як поверхня Землі впливає на поширення радіохвиль?
4. Які характерні зони на поверхні Землі виділяють при різних висотах розміщення випромінювальної антени?
5. Що таке суттєва область відбиття електромагнітних хвиль?
6. В яких наближеннях можна розрахувати електромагнітне поле випромінювальної антени в освітленій зоні та в зоні тіні?
7. Опишіть модель, що використовується для розрахунку електромагнітного поля випромінювальної антени в освітленій зоні для моделі плоскої Землі.
8. Як впливає кривина Землі на електромагнітне поле випромінювальної антени в освітленій зоні?

Задачі до підрозділу 8.4

8.4.1. Висота розміщення передавальної антени – $h = 100$ м. Оцінити віддаль від неї до зони півтіні.

8.4.2. Віддаль від передавальної до приймальної антени – $r = 100$ км, частота хвилі – $f = 60$ МГц, висота розміщення передавальної антени – $h = 50$ м, $n = 10$. Знайти розміри суттєвої області відбиття в наближенні плоскої Землі. Чи буде це наближення справедливим для даної задачі?

8.4.3. Передавальна антена піднята на висоту $h = 400$ м. Оцінити коефіцієнт ослаблення сигналу з частотою $f = 43$ МГц на віддалях 50 км та 90 км від неї.

8.5. Електромагнітне поле низько розміщеного випромінювача

Як уже вказувалося, модель низько розміщеного випромінювача ($h \ll \lambda$) застосовна для середніх та довгих хвиль ($\lambda \geq 100$ м). Розрахунки для цієї моделі виявляються значно складнішими, ніж для моделі високо піднятого випромінювача, оскільки в даному разі освітлена зона відсутня, а існують лише зона півтіні й зона тіні, де суттєву роль відіграють дифракційні ефекти, а розрахунки на основі врахування відбитої від Землі хвилі (див. вище п. 8.4) не справджуються.

Тому ми будемо вести мову лише про електромагнітне поле на невеликих віддалях від випромінювача, де кривиною Землі можна знехтувати.

8.5.1. Опис моделі без урахування кривини Землі та поняття числової віддалі

Як уже згадувалося, низько розміщені антени збуджують хвилю, яка поширюється безпосередньо вздовж земної поверхні. Вважатимемо, що її джерелом є вертикально орієнтований вібратор Герца, розташований безпосередньо на поверхні Землі (рис. 8.5.1). На частотах, які відповідають середнім і довгим хвилям, земну поверхню в багатьох випадках можна розглядати як ідеальний провідник.

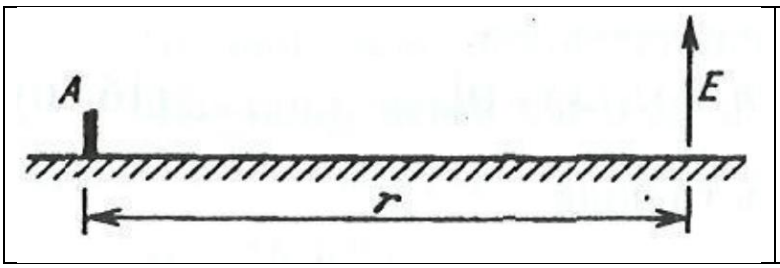


Рис. 8.5.1. Схематичне зображення вертикального вібратора Герца, розміщеного в точці А на провідній поверхні

Отримання точного розв'язку цієї задачі вимагає розв'язання рівнянь Максвелла з урахуванням граничних умов, зокрема й на поверхні провідника. Такий розв'язок був отриманий ще в 1909 р. А. Зомерфельдом.

На практиці користуються, як правило, наближеними розв'язками або результатами числових розрахунків.

Якщо поверхня Землі має достатню провідність, то вона стає непрозорою для електромагнітних хвиль. В результаті потік енергії випромінювання існує лише у верхній половині простору, а його середня густина (для ізотропного випромінювача) виявляється вдвічі більшою, ніж у вільному просторі, а поле в тоці спостереження зростає в $h \ll \lambda$ разів. У результаті формула (8.3.8) набуває вигляду

$$E_m = \frac{\sqrt{120 P_r G}}{r} V, \qquad (8.5.1)$$

а розв'язання задачі формально зводиться до знаходження множника ослаблення V.

Множник ослаблення V зручно записати як функцію деякого безрозмірного параметра $\tilde{r}$ – так званої числової віддалі: $V = V(\tilde{r})$.

Нагадаємо (див. вище п. 8.1.1), що діелектрична проникність провідного середовища для змінних (точніше, гармонічних) електричних полів є комплексною. Згідно формули (8.1.6) вона може бути подана у вигляді

$$\varepsilon = \varepsilon' - i\varepsilon'' = \varepsilon' - i\frac{4\pi\sigma}{\omega},$$

де ε' – діелектрична проникність без врахування провідності, σ – питома високочастотна електропровідність, ω – частота прикладеного поля. Формула (8.1.6) записана в системі оди-

ниць Гаусса. Врахувавши зв'язок між частотою і довжиною хвилі $\lambda = 2\pi c/\omega$ і перейшовши до системи СІ, можна переписати цю формулу у вигляді

$$\varepsilon = \varepsilon' - i60\lambda\sigma. \qquad (8.5.2)$$

Тепер можна ввести поняття *числової віддалі* $\tilde{r}$:

$$\tilde{r} = \frac{\pi r}{\lambda}\left|\frac{\varepsilon - 1}{\varepsilon^2}\right|, \qquad (8.5.3)$$

де ε – комплексна діелектрична проникність (8.5.2) для хвиль довжини λ, r – як і раніше, віддаль між передавачем і приймачем (рис. 8.5.1). Видно, що при поширенні хвилі у вакуумі, де $\varepsilon = 1$, $\tilde{r} = 0$. Так само поводить себе $\tilde{r}$ при $|\varepsilon| \to \infty$. Навпаки, коли $|\varepsilon| \to 0$ (таке може трапитися в плазмі, див. п. 8.1.2), то $\tilde{r} \to \infty$. Отже, числова віддаль кількісно враховує вплив середовища на поширення радіохвилі від передавача до приймача.

Підставивши (8.5.2) до (8.5.3), отримуємо:

$$\tilde{r} = \frac{\pi r}{\lambda}\left|\frac{\varepsilon' - i60\lambda\sigma - 1}{(\varepsilon' - i60\lambda\sigma)^2}\right| = \frac{\pi r}{\lambda}\frac{\sqrt{(\varepsilon'-1)^2 + (60\lambda\sigma)^2}}{(\varepsilon')^2 + (60\lambda\sigma)^2}. \qquad (8.5.4)$$

У деяких крайніх випадках формулу (8.5.4) вдається спростити.

Зокрема, за умови $1 \ll \varepsilon' \sim 60\lambda\sigma$ (середовище з великим ε' і значною провідністю) маємо:

$$\rho \approx \frac{\pi r}{\lambda\sqrt{(\varepsilon')^2 + (60\lambda\sigma)^2}}. \qquad (8.5.5)$$

За умов $\varepsilon' \gg 60\lambda\sigma$ та $\varepsilon' \gg 1$ (добрий діелектрик)

$$\rho \approx \frac{\pi r}{\lambda\varepsilon'}. \qquad (8.5.6)$$

За умови $60\lambda\sigma \gg \varepsilon'$ (провідник)

$$\rho \approx \frac{\pi r}{60\lambda^2\sigma} \qquad (8.5.7)$$

(записана умова пояснює, чому для довгих хвиль більшість різновидів земної поверхні можна розглядати як провідники).

При $\varepsilon' = 1 \gg 60\lambda\sigma$ (діелектрик з $\varepsilon' = 1$)

$$\rho = 60\pi\sigma r. \qquad (8.5.8)$$

Щоб зрозуміти, яким моделям відповідають різні види земного покрову, доповнимо таблицю 8.3.1 (див. табл. 8.5.1). Для оцінок візьмемо $\lambda = 1000 м$, що добре відповідає моделі низько розміщених антен.

Таблиця 8.5.1

Вид земної поверхні	*Діелектрична проникність* ε'	*Питома провідність* σ *[См/м]*	$60\lambda\sigma$	*Модель для опису*
Морська вода	80	1-6	(0.6-3.6) 10^5	провідник
Вода річок та озер	80	10^{-3}-10^{-2}	60-600	середовище з великим ε' і значною провідністю
Волога земля	10-30	$3\cdot10^{-3}$-$3\cdot10^{-2}$	180-1800	провідник
Суха земля	3-6	10^{-5}-$5\cdot10^{-3}$	0.6-300	добрий діелектрик, середовище з великим ε' і значною провідністю або провідник
Мерзла земля	3-6	10^{-3}-10^{-2}	60-600	провідник
Сніг при -10ºС	1	10^{-6}	0.06	діелектрик з $\varepsilon' = 1$
Крига при -10ºС	4-5	10^{-2}-10^{-1}	600-6000	провідник

Як бачимо з таблиці, для обраного значення λ земну поверхню справді в більшості

випадків можна розглядати як провідник. При збільшенні λ застосовність цієї моделі зростає. З іншого боку, добрим діелектриком (щоправда, з $\varepsilon' = 1$) виступає лише сухий сніг, в окремих випадках – суха земля. Прикладом середовища з великим ε' і значною провідністю є вода річок та озер, інколи – суха земля.

8.5.2. Результати розрахунку множника ослаблення сигналу без урахування кривини Землі

Як показують розрахунки, для вертикально розташованого вібратора коефіцієнт ослаблення має вигляд

$$V(\tilde{r}) = \left| 1 - \sqrt{\tilde{r}} \exp(-\tilde{r}) \left[2 \int_0^{\sqrt{\tilde{r}}} \exp\left(x^2\right) dx + i\sqrt{\pi} \right] \right|. \qquad (8.5.9)$$

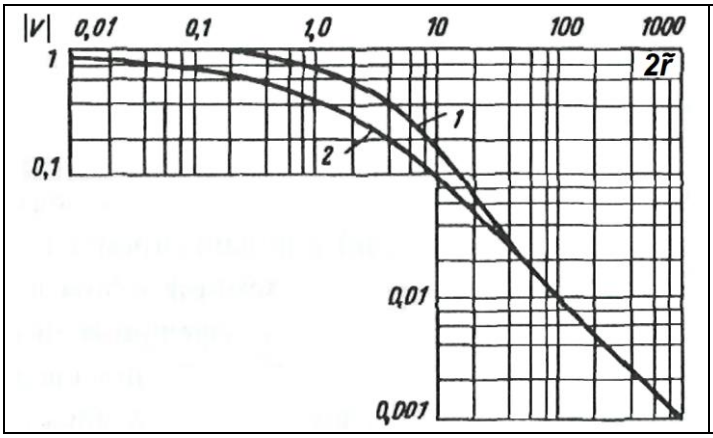


Рис. 8.5.2. Графік залежності коефіцієнту ослаблення від числової віддалі: 1 – провідники ($60\lambda\sigma \gg \varepsilon'$), 2 – діелектрики ($60\lambda\sigma \ll \varepsilon'$)

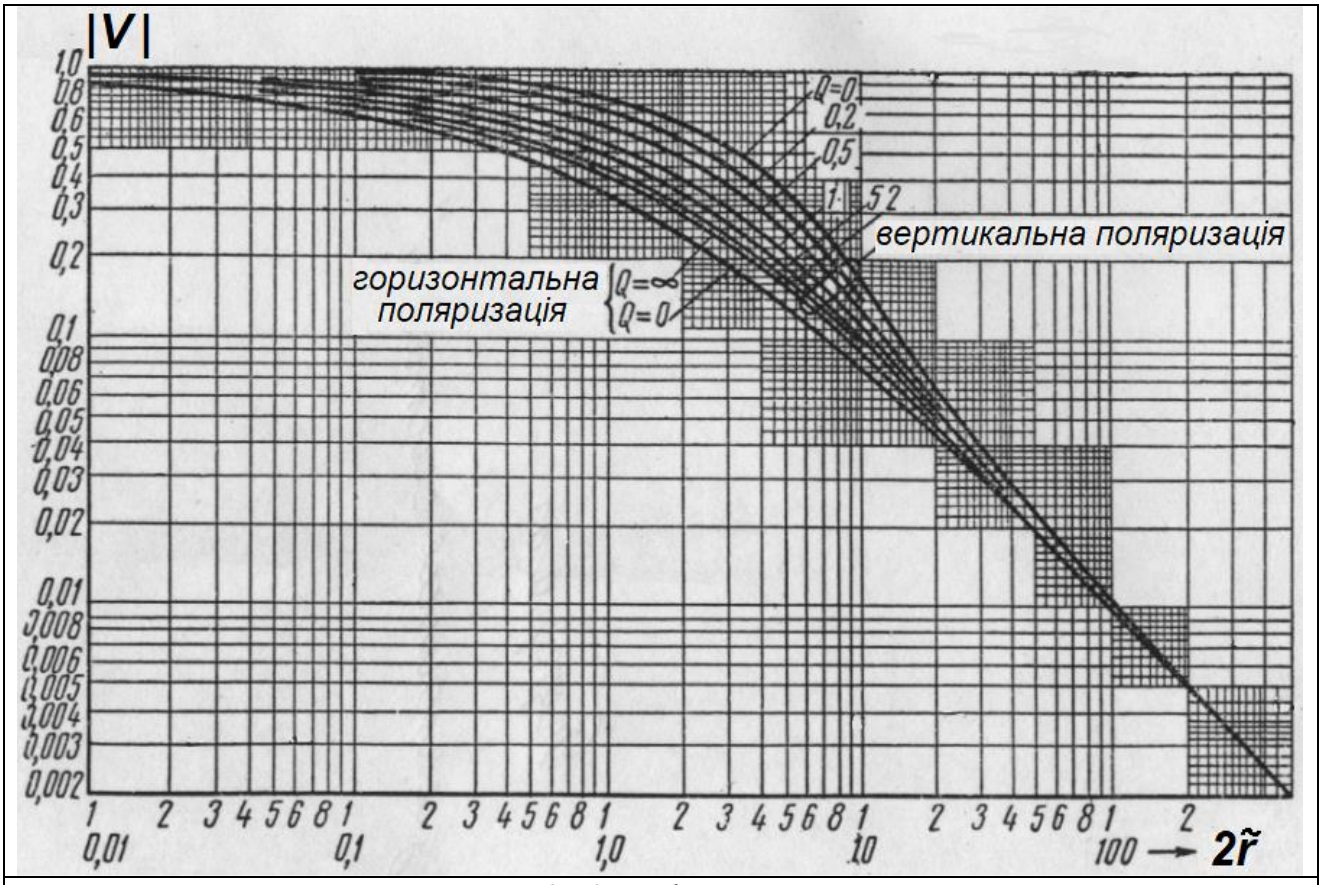


Рис. 8.5.3. Графіки Берроуза

Графіки залежностей $V(\rho)$ для випадків, коли земну поверхню можна вважати добрим провідником (крива 1) та добрим діелектриком (крива 2), наведені на рис. 8.5.2.

Повнішу інформацію про залежність $V(\rho)$ дають так звані графіки Берроуза (рис. 8.5.3), які включають дані рис. 8.3.2 як частинний випадок. По перше, на графіки Берроуза подані результати не тільки для вертикальної поляризації поля, джерелом якого є вертикально орієнтований електричний диполь, але й для горизонтальної поляризації поля, збудженого вертикально орієнтованим магнітним диполем. По друге, тут подані залежності для різних значень параметра

$$Q = \frac{\varepsilon'}{\varepsilon''} = \frac{\varepsilon'}{60\lambda\sigma}, \qquad (8.5.10)$$

який визначає відношення струму зміщення до струму провідності в тому самому змінному електричному полі. Очевидно, для діелектриків $Q \to \infty$, для провідників $Q \to 0$.

Отже, графіки Берроуза дозволяють знайти поле випромінювачів з різною поляризацією, розміщених на земній поверхні з різними типами провідності в точці спостереження на різних відстанях від випромінювача.

Порівняння отриманих результатів з точними розв'язками, які враховують кривину земної поверхні, показує, що модель плоскої Землі чинна за виконання умови

$$r[м] < 7 \cdot 10^3 \sqrt[3]{\lambda[км]},$$

тобто для $\lambda = 1$ км маємо $r < 7$ км. При цьому похибка розрахунків не перевищує 10%.

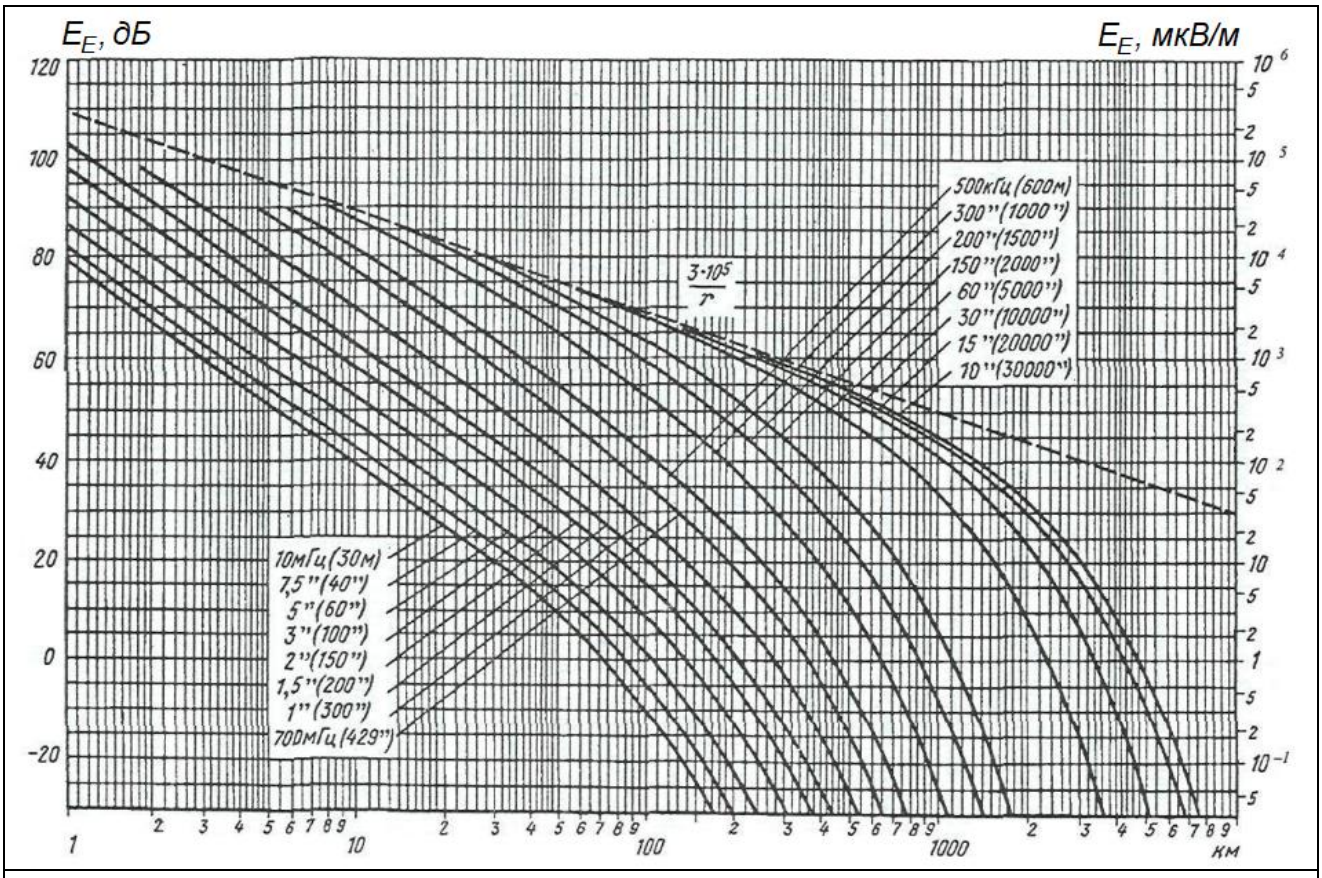


Рис. 8.5.4. Графіки МККР для $\varepsilon' = 4$, $\sigma = 10^{-3}$ *См / м*

8.5.3. Результати розрахунків поля антени з урахуванням кривини Землі

Як уже вказувалося, врахування кривини Землі принципово неможливе без урахування дифракційних ефектів, тому відповідні аналітичні розв'язки різко ускладняються. На практи-

ці користуються результатами числових розрахунків, виконаних на основі цих аналітичних розв'язків. Результат розрахунків залежить від електричних властивостей земної поверхні, вздовж якої поширюється хвиля, а також від характеру її рельєфу, вплив якого враховується через так звану удавану діелектричну проникність. Наприклад, для міських умов вона складає $\varepsilon' = (3-5)$, $\sigma = \left(10^{-4} - 10^{-3}\right)$ Сим/м.

Для практичного розрахунку полів, збуджених антеною, використовуються, зокрема, так звані графіки МККР[5].Приклад такого графіка для параметрів $\varepsilon' = 4$, $\sigma = 10^{-3}$ См/м наведений на рис. 8.5.4. Вказані на ньому значення напруженості електричного поля відповідають ізотропному випромінювачу з потужністю 1 кВт, тому їх треба помножити на величину $\sqrt{P_r\,[\text{кВт}]\,K_{dir}}$.

Рис. 8.5.4 підтверджує зроблений раніше висновок, що збільшення довжини хвилі приводить до зростання віддалі, якої досягає поле антени. Нахил штрихової лінії у верхній частині цього рисунку відповідає ослабленню поля лише за рахунок природного розходження променів. Видно, що графіки мають більший нахил, особливо на великих віддалях від антени. Це пов'язано з додатковими механізмами ослаблення поля (поглинання хвиль земною поверхнею, втрати на дифракційне огинання поверхні та інші).

Користуватися графіками МКРР простіше, ніж графіками Берроуза, оскільки з них відразу можна отримати (з урахуванням зробленого зауваження) напруженість поля на заданій відстані від випромінювача. Однак для практичної роботи (різні значення ε' та σ) потрібно мати цілий масив таких графіків, тоді як графік Берроуза більш універсальний.

Хоча на графіку МКРР присутні криві для довжин хвиль від 30 м до 30 км включно, насправді ними можна користуватися лише для діапазону 100 м – 2 км. При менших довжинах хвиль порушується наближення низько розміщеної антени, при більших включається інший механізм поширення, пов'язаний з існування хвилеводу Земля – іоносфера.

Контрольні питання до підрозділу 8.5

1. Як можна точно розрахувати поле випромінювання низько розміщеної антени? Які припущення використовують для спрощення цього розрахунку?
2. Що таке числова віддаль між точкою випромінювання й точкою спостереження? Чи може вона обертатися на нуль при ненульовій віддалі між цими точками?
3. Які моделі земної поверхні застосовують у розрахунках поля випромінювання низько розміщеної антени? Коли ці моделі можна застосовувати?
4. Чому для довгих хвиль поверхню Землі можна вважати ідеальним провідником?
5. За яких умов поверхню Землі можна розглядати як діелектрик?

Приклади розв'язання задач до підрозділу 8.5

П 8.5.1. Частота хвилі, випромінюваної вертикально розташованим на поверхні Землі диполем Герца – 150 кГц, поверхню Землі вважаємо добрим діелектриком з $\varepsilon' = 3$. На якій віддалі від випромінювача коефіцієнт ослаблення сигналу дорівнюватиме 0.01? Скористатися моделлю плоскої Землі.

[5] Міжнародний консультативний комітет з радіозв'язку (МККР), після 1992 р. – Сектор радіозв'язку Міжнародного союзу електрозв'язку – міжнародна організація, що займається вивченням технічних та експлуатаційних проблем радіозв'язку і виробленням відповідних рекомендацій.

Розв'язок

Для розв'язання задачі скористаємося кривою 1 на рис. 8.5.2. З нього видно, що значення $V(\tilde{r}) = 0.01$ досягається, коли $2\tilde{r} = 100$, або $\tilde{r} = 50$. Тоді з формули (8.5.8) випливає, що

$$r = \lambda \frac{\tilde{r}\varepsilon'}{\pi} = 2\frac{50 \times 3}{3.14} = 96 \text{ км}.$$

Тут враховано, що

$$\lambda = \frac{c}{f} = \frac{3 \cdot 10^8}{1.5 \cdot 10^5} = 2 \cdot 10^3 \text{ м} = 2 \text{ км}.$$

Задачі до підрозділу 8.5

8.5.1. Частота хвилі, випромінюваної вертикально розташованим на поверхні Землі диполем Герца – 300 кГц, поверхню Землі вважаємо добрим діелектриком з $\varepsilon' = 5$. На якій віддалі від випромінювача коефіцієнт ослаблення сигналу дорівнюватиме 0.1? Скористатися моделлю плоскої Землі.

8.5.2. Коефіцієнт спрямованої дії низько розміщеного випромінювача – 10^4, його потужність – 30 кВт, частота – 10 МГц. Параметри земного покрову – $\varepsilon' = 4$, $\sigma = 10^{-3}$ Сим/м. Знайти напруженість електричного поля випромінювання на віддалях 10 км та 100 км від випромінювача з урахуванням кривини Землі.

8.6. Відбиття електромагнітних хвиль від слабконеоднорідної плазми

В багатьох випадках на поширення хвиль у приземному просторі вливає іоносферна плазма. Як уже зазначалося, на низьких частотах вона стає непрозорою для радіовипромінювання.

Вище згадувалося, що іоносферна плазма є неоднорідною (в першу чергу, за висотою). Ця неоднорідність впливає на поширення електромагнітних хвиль в іоносфері. Особливо це стосується довгих та наддовгих хвиль, для яких деякі частини іоносфери (в першу чергу, шар F2, якому відповідає максимальна концентрація іоносферної плазми) можуть стати непрозорими.

У цьому підрозділі ми розглянемо задачу про поширення електромагнітних хвиль у неоднорідній плазмі в загальному вигляді, не прив'язуючись до іоносфери, а в наступному – предметно проаналізуємо вплив іоносферної плазми на поширення радіохвиль у приземному просторі.

8.6.1 Хвильові рівняння для електромагнітних хвиль у плоскошаруватій плазмі

Будемо розглядати холодну ізотропну плазму без зіткнень, концентрація якої залежить від однієї координати z (так звана плоскошарувата модель). В реальних ситуаціях ця модель застосовна, коли поперечні розміри променю радіохвиль, що падає на плазмове утворення, значно менші за характерний розмір неоднорідності в поперечному напрямку (рис. 8.6.1).

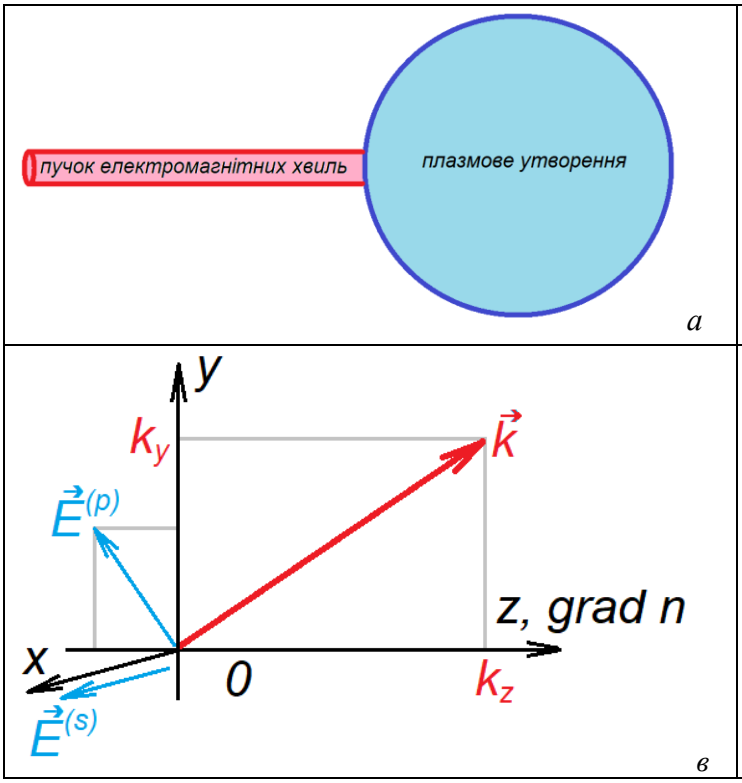


Рис. 8.6.1. Плоскошарувате плазмове утворення: а – умови застосовності моделі; б – профіль концентрації плазми; в – орієнтація хвильового вектора та електричних полів s- і p-поляризованої хвилі щодо градієнту концентрації плазми

Будемо описувати неоднорідну плазму її діелектричною проникністю ε на частоті ω. Це можливо, якщо характерний розмір неоднорідності

$$L = \left| \frac{1}{n_e} \frac{dn_e}{dz} \right|^{-1} \tag{8.6.1}$$

буде значно більшим за дебаївський радіус (див. вище п. 8.1.2), $L >> r_D \equiv v_{Te}/\omega_p$. Якщо ця умова не задовольняється, величину ε у неоднорідній плазмі ввести не можна.

Нехай у такій плазмі поширюється гармонічна електромагнітна хвиля з частотою ω,

$\vec{E}(\vec{r},t)=\vec{E}(\vec{r})\exp(i\omega t)$.

Хвильове рівняння для такої моделі можна отримати з відповідних рівнянь Максвелла. Воно має вигляд

$$\left[\nabla\times\left[\nabla\times\vec{E}\right]\right]-k_0^2\varepsilon(z)\vec{E}=0, \qquad k_0=\frac{\omega}{c}, \tag{8.6.2}$$

або

$$\Delta\vec{E}-\nabla\left(\nabla\cdot\vec{E}\right)+k_0^2\varepsilon(z)\vec{E}=0. \tag{8.6.3}$$

Приймемо, що компоненти електричного поля не залежать від x, тобто хвиля поширюється перпендикулярно осі x, і, відповідно, її хвильовий вектор лежить у площині yz (рис. 8.6.1 в). Тоді проекції хвильового рівняння (8.6.3) на осі координат мають вигляд:

$$\frac{\partial^2 E_x}{\partial y^2}+\frac{\partial^2 E_x}{\partial z^2}+k_0^2\varepsilon(z)E_x=0; \tag{8.6.4}$$

$$\begin{cases}\dfrac{\partial^2 E_y}{\partial y^2}+\dfrac{\partial^2 E_y}{\partial z^2}-\dfrac{\partial}{\partial y}\left(\dfrac{\partial E_y}{\partial y}+\dfrac{\partial E_z}{\partial z}\right)+k_0^2\varepsilon(z)E_y=0;\\ \dfrac{\partial^2 E_z}{\partial y^2}+\dfrac{\partial^2 E_z}{\partial z^2}-\dfrac{\partial}{\partial z}\left(\dfrac{\partial E_y}{\partial y}+\dfrac{\partial E_z}{\partial z}\right)+k_0^2\varepsilon(z)E_z=0.\end{cases} \tag{8.6.5}$$

Система отриманих таким чином рівнянь розпадається на рівняння (8.6.4) для компоненти E_x (це рівняння описує $s-$поляризовані хвилі) і підсистему (8.6.5) для компонент E_y та E_z (вона відповідає $p-$поляризованій хвилі)[6]. Напрямок електричних полів s- і p-поляризованої хвилі щодо хвильового вектора та градієнту концентрації плазми показаний на рис. 8.6.1 в.

Розглянемо спочатку $s-$поляризовані хвилі. Будемо шукати розв'язок хвильового рівняння (8.6.4) у формі $E_x(\vec{r})=E_x(z)\exp(-ik_y y)$. Після його підстановки отримаємо:

$$\frac{d^2E_x}{dz^2}+k_z^2(z)E_x=0, \qquad k_z^2(z)=\varepsilon(z)k_0^2-k_y^2. \tag{8.6.6}$$

8.6.2. Наближення геометричної оптики

В однорідній плазмі, де $k_z(z)=const$, розв'язок рівняння (8.6.6) має вигляд

$$E_x(z)=A\exp(\mp ik_z z). \tag{8.6.7}$$

Нехай неоднорідність плазми є слабкою, тобто виконано умову

$$k_z L>>1. \tag{8.6.8}$$

Іншими словами, характерний розмір неоднорідності L значно перевищує просторовий період електромагнітної хвилі в напрямку градієнту концентрації плазми. В цьому випадку в розв'язку (8.6.7) можна припустити, що амплітуда A та хвильове число k_z є повільними функціями координати z. Іншими словами, шукатимемо розв'язок рівняння (8.6.6) у формі:

$$E_x(z)=A(z)\exp\left[i\psi(z)\right], \tag{8.6.9}$$

---

[6] Терміни "p-поляризація" та "s-поляризація" походять від англ. parallel (паралельний) та нім. senkrecht (перпендикулярний) і вказують на орієнтацію вектора напруженості електричного поля щодо площини падіння хвилі.

що відповідає наближенню геометричної оптики[7]. Введена таким чином величина ψ називається ейконалом. Амплітуда та ейконал мають задовольняти умовам

$$\frac{dA}{dz} \sim \frac{A}{L}, \qquad \frac{d^2\psi}{dz^2} \sim \frac{1}{L}\frac{d\psi}{dz} \tag{8.6.10}$$

(перша похідна від ейконалу відповідає $z-$компоненті хвильового вектора, тому при $L \to \infty$ залежності A та $d\psi/dz$ від z повинні зникнути). З умов (8.6.8) та (8.6.10) випливають нерівності

$$\left|\frac{dA}{dz}\right| << k_z |A|, \qquad \left|\frac{d^2\psi}{dz^2}\right| << k_z \left|\frac{d\psi}{dz}\right| \tag{8.6.11}$$

(сам ейконал не є повільною функцією координати).

Випишемо в явному вигляді першу та другу похідні від E_x:

$$\frac{dE_x}{dz} = \overset{I}{\frac{dA}{dz}} \exp(i\psi) + iA \overset{0}{\frac{d\psi}{dz}} \exp(i\psi); \tag{8.6.12}$$

$$\frac{d^2E_x}{dz^2} = \overset{II}{\frac{d^2A}{dz^2}} \exp(i\psi) + 2i \overset{I}{\frac{dA}{dz}\frac{d\psi}{dz}} \exp(i\psi) + iA \overset{I}{\frac{d^2\psi}{dz^2}} \exp(i\psi) - A\overset{0}{\left(\frac{d\psi}{dz}\right)^2} \exp(i\psi)$$

(над доданками у правих частинах виразів указаний їхній порядок мализни за параметром $1/k_z L$).

Тепер підставимо розв'язок (8.6.9) з урахуванням (8.6.12) до рівняння (8.6.6), залишаючи лише доданки нульового та першого порядків мализни. В цьому виразі прирівняємо до нуля окремо доданки нульового і першого порядків. Отримане таким чином рівняння для доданків нульового порядку мализни має вигляд:

$$-A\left(\frac{d\psi}{dz}\right)^2 + k_z^2(z)A = 0, \tag{8.6.13}$$

звідки можна знайти явний вираз для ейконалу:

$$\psi = \pm\int_{z_0}^{z} k_z(z')dz'. \tag{8.6.14}$$

Рівняння для доданків першого порядку має вигляд

$$2\frac{dA}{dz}\frac{d\psi}{dz} + A\frac{d^2\psi}{dz^2} = 0, \tag{8.6.15}$$

або, з урахуванням явного вигляду ейконалу,

$$\pm 2\frac{dA}{dz}k_z \pm A\frac{dk_z}{dz} = 0. \tag{8.6.16}$$

Інтегруючи його, можна знайти явний вигляд просторової залежності амплітуди A:

$$A(z) = \frac{A_0}{\sqrt{k_z(z)}}. \tag{8.6.17}$$

Отже, в загальному випадку розв'язок рівняння (8.6.6), отриманий методом геометричної оптики, можна подати у формі:

$$E_x(z) = \frac{1}{\sqrt{k_z(z)}}\left\{A_1 \exp\left[i\int_{z_0}^{z} k_z(z')dz'\right] + A_2 \exp\left[-i\int_{z_0}^{z} k_z(z')dz'\right]\right\}\exp(i\omega t - ik_y y). \tag{8.6.18}$$

Цей розв'язок описує дві хвилі – пряму та зворотну.

---

[7] В літературі метод геометричної оптики називають також методом Вентцеля – Крамерса – Бріллюена (ВКБ).

Відзначимо, що в розв'язку (8.6.18) сталі інтегрування A_1 та A_2 – комплексні амплітуди зворотної та прямої хвиль – залежать від параметра z_0 (початку відліку) таким чином, що розв'язок у цілому не залежить від цього параметра.

У щільній плазмі, де $k_z^2 < 0$, також можна записати розв'язок хвильового рівняння у наближенні геометричної оптики:

$$E_x(z) = \frac{1}{\sqrt{\chi_z(z)}}\left\{B_1 \exp\left[-\int_{z_0}^{z} \chi_z(z')dz'\right] + B_2 \exp\left[\int_{z_0}^{z} \chi_z(z')dz'\right]\right\}\exp\left(i\omega t - ik_y y\right), \qquad (8.6.19)$$

де введене позначення $\chi_z^2(z) = -k_z^2(z)$.

8.6.3. Інтерпретація геометрикооптичного розв'язку

Нехай у неоднорідному середовищі присутня тільки пряма хвиля, тобто $A_1 = 0$, $A_2 \neq 0$. Знайдемо $z-$компоненту вектора Пойнтінга для такої хвилі. Як відомо, вектор Пойнтінга пропорційний добутку групової швидкості хвилі на густину її енергії, так що $\Pi_z \sim v_{zg} E_m^2$, де $E_m = A_0 \big/ \sqrt{k_z(z)}$. Розрахунок $z-$компоненти групової швидкості з урахуванням явної залежності $\varepsilon(\omega)$ дає:

$$v_{zg} = \frac{\partial\omega}{\partial k_z} = \left(\frac{\partial k_z}{\partial\omega}\right)^{-1} = \left(\frac{\partial}{\partial\omega}\sqrt{k_0^2\varepsilon(z) - k_y^2}\right)^{-1} = \left(\frac{\partial}{\partial\omega}\sqrt{\frac{\omega^2 - \omega_p^2}{c^2} - k_y^2}\right)^{-1} = \left(\frac{1}{2k_z}\frac{2\omega}{c^2}\right)^{-1} = c\frac{k_z}{k_0}. \qquad (8.6.20)$$

Отже,

$$\Pi_z(z) \sim \frac{A_0{}^2}{k_z(z)} c \frac{k_z(z)}{k_0} = const. \qquad (8.6.21)$$

Як бачимо, просторова зміна амплітуди, що описується формулою (8.6.17), обумовлена зміною групової швидкості хвилі при збереженні густини потоку її енергії. Іншими словами, амплітуда хвилі зростає там, де зменшується її групова швидкість, так що густина потоку енергії в усіх точках залишається однаковою.

8.6.4. Розподілене відбиття хвилі в слабконеоднорідному середовищі

Знайдемо тепер, у яких областях наближення геометричної оптики (8.6.8) порушується. Очевидно, це відбувається при $k_z \to 0$, тобто в околі точки, де $\varepsilon(z) = \left(k_y / k_0\right)^2$. Оскільки у вакуумі $k_y = k_0 \sin\theta$ (θ – кут падіння, що відраховується від осі z), то в цій точці $\varepsilon(z) = \sin^2\theta$. Ця точка називається точкою повороту хвилі, оскільки вона розділяє області прозорості та непрозорості для хвилі з даними значеннями ω та k_y.

Порушення справедливості наближення геометричної оптики в околі точки повороту означає, що в цій точці можливе взаємне перетворення прямої та зворотної хвиль.

Для того, щоб з'ясувати, в чому саме полягає порушення наближення геометричної оптики в околі точки повороту, розглянемо знову $s-$поляризовану хвилю в плоскошаруватій плазмі. Неоднорідність тепер уже не вважатимемо слабкою.

Підставимо розв'язок рівняння (8.6.6) у формі (8.6.18), вважаючи, однак, що «геометрикооптичні амплітуди» $A_{1,2}$ залежать від координат. Розрахуємо спочатку першу похідну dE_x / dz:

$$\frac{dE_x}{dz}=\left\{\frac{1}{\sqrt{k_z}}\frac{dA_1}{dz}-\frac{A_1}{2k_z^{3/2}}\frac{dk_z}{dz}+i\sqrt{k_z}A_1\right\}\exp(i\psi)+\left\{\frac{1}{\sqrt{k_z}}\frac{dA_2}{dz}-\frac{A_2}{2k_z^{3/2}}\frac{dk_z}{dz}-i\sqrt{k_z}A_2\right\}\exp(-i\psi). \tag{8.6.22}$$

Оскільки ми замість однієї невідомої функції E_x ввели дві нові – A_1 та A_2, то ми можемо накласти на них деяку додаткову умову. Подамо її у формі

$$\left\{\frac{1}{\sqrt{k_z}}\frac{dA_1}{dz}-\frac{A_1}{2k_z^{3/2}}\frac{dk_z}{dz}\right\}\exp(i\psi)+\left\{\frac{1}{\sqrt{k_z}}\frac{dA_2}{dz}-\frac{A_2}{2k_z^{3/2}}\frac{dk_z}{dz}\right\}\exp(-i\psi)=0. \tag{8.6.23}$$

З урахуванням (8.6.23) можна записати, що

$$\frac{d^2E_x}{dz^2}=\left\{i\sqrt{k_z}\frac{dA_1}{dz}+\frac{iA_1}{2\sqrt{k_z}}\frac{dk_z}{dz}-k_z^{3/2}A_1\right\}\exp(i\psi)+\left\{-i\sqrt{k_z}\frac{dA_2}{dz}-\frac{iA_2}{2\sqrt{k_z}}\frac{dk_z}{dz}-k_z^{3/2}A_2\right\}\exp(-i\psi). \tag{8.6.24}$$

Підставивши (8.6.24) до (8.6.6), отримаємо:

$$\left\{i\sqrt{k_z}\frac{dA_1}{dz}+\frac{iA_1}{2\sqrt{k_z}}\frac{dk_z}{dz}\right\}\exp(i\psi)+\left\{-i\sqrt{k_z}\frac{dA_2}{dz}-\frac{iA_2}{2\sqrt{k_z}}\frac{dk_z}{dz}\right\}\exp(-i\psi)=0. \tag{8.6.25}$$

Співвідношення (8.6.23) та (8.6.25) можна розглядати як систему лінійних неоднорідних алгебраїчних рівнянь щодо похідних $dA_{1,2}/dz$. Розв'язавши її, отримаємо:

$$\frac{dA_{1,2}}{dz}=\frac{A_{2,1}}{2k_z}\frac{dk_z}{dz}\exp(\mp i2\psi). \tag{8.6.26}$$

Як видно з (8.6.26), зміна геометрикооптичної амплітуди A_1 (A_2) можлива лише за рахунок іншої геометрикооптичної амплітуди A_2 (A_1). Для того, щоб така зміна відбувалася, необхідний градієнт поздовжнього хвильового числа (тобто, по суті, градієнт ε). Цей ефект – розподілене відбиття хвилі – має місце, зокрема, й у слабконеоднорідній плазмі. Він не враховується в першому наближенні методу геометричної оптики, яким ми користувалися вище.

Найбільш помітна зміна геометрикооптичних амплітуд, як видно з (8.6.26), відбувається в околі точки повороту, тобто там, де в слабконеоднорідній плазмі порушується наближення геометричної оптики.

8.6.5. Відбиття s-поляризованої хвилі від слабконеоднорідної плазми

Розглянемо тепер задачу про відбиття $s-$поляризованої електромагнітної хвилі від плазми з розмитою межею. Концентрація плазми, як і раніше, залежить лише від координати z. Нехай область $z<0$ являє собою вакуум, а в області $z>0$ концентрація плазми монотонно зростає (рис. 8.6.2 а).

Нехай на плазму під кутом θ падає $s-$поляризована хвиля. Хвильовий вектор хвилі лежить у площині yz. Електричне поле $s-$поляризованої хвилі перпендикулярне до площини падіння, тобто воно спрямоване вздовж осі x. Таким чином, поведінка електричного поля хвилі описується рівнянням (8.6.6).

Якщо неоднорідність плазми вважати слабкою, то поза околом точки повороту, в якій $\varepsilon=\sin^2\theta$, справедливі геометрикооптичні розв'язки (8.6.18) (ліворуч від точки повороту) та (8.6.19) (праворуч від точки повороту). В останньому розв'язку слід покласти $B_2=0$, оскільки поле хвилі повинне залишатися обмеженим при $z\to\infty$.

Ми вважаємо хвилю, що падає на плазму, заданою (тобто задаємо амплітуду A_2). Тоді, зв'язавши між собою розв'язки хвильового рівняння по обидва боки від точки повороту, можна знайти амплітуди A_1 та B_1.

Оскільки в нашій моделі відсутнє згасання ($\operatorname{Im}\varepsilon=0$), то амплітуда відбитої хвилі має

дорівнювати амплітуді падаючої хвилі ($|A_1| = |A_2|$). Іншими словами, в тій області, де електромагнітна хвиля може поширюватися, вона буде стоячою.

Для того, щоб зв'язати геометрикооптичні розв'язки по обидва боки від точки повороту, можна покласти, що в деякому невеликому околі названої точки профіль концентрації плазми є лінійним. У цьому випадку рівняння (8.6.6) вдається розв'язати точно через деякі спеціальні функції – так звані функції Ейрі. В свою чергу, асимптотика цих функцій може бути ототожнена з геометрикооптичними розв'язками.

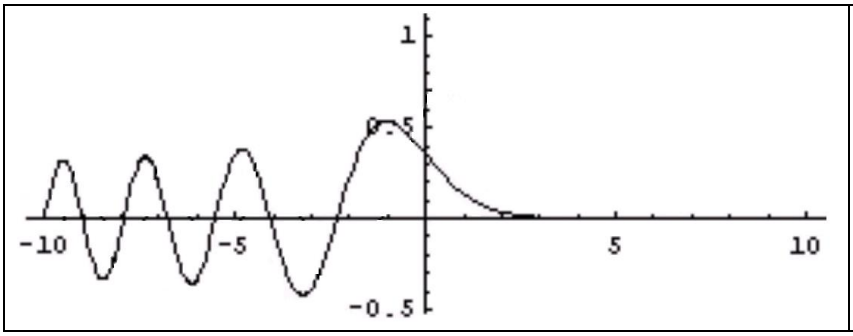


Рис. 8.6.2. Графік функції Ейрі Ai(x)

Залежність інтенсивності електричного поля хвилі від координати z, побудована описаним вище способом, наведена на рис. 8.6.2 б.

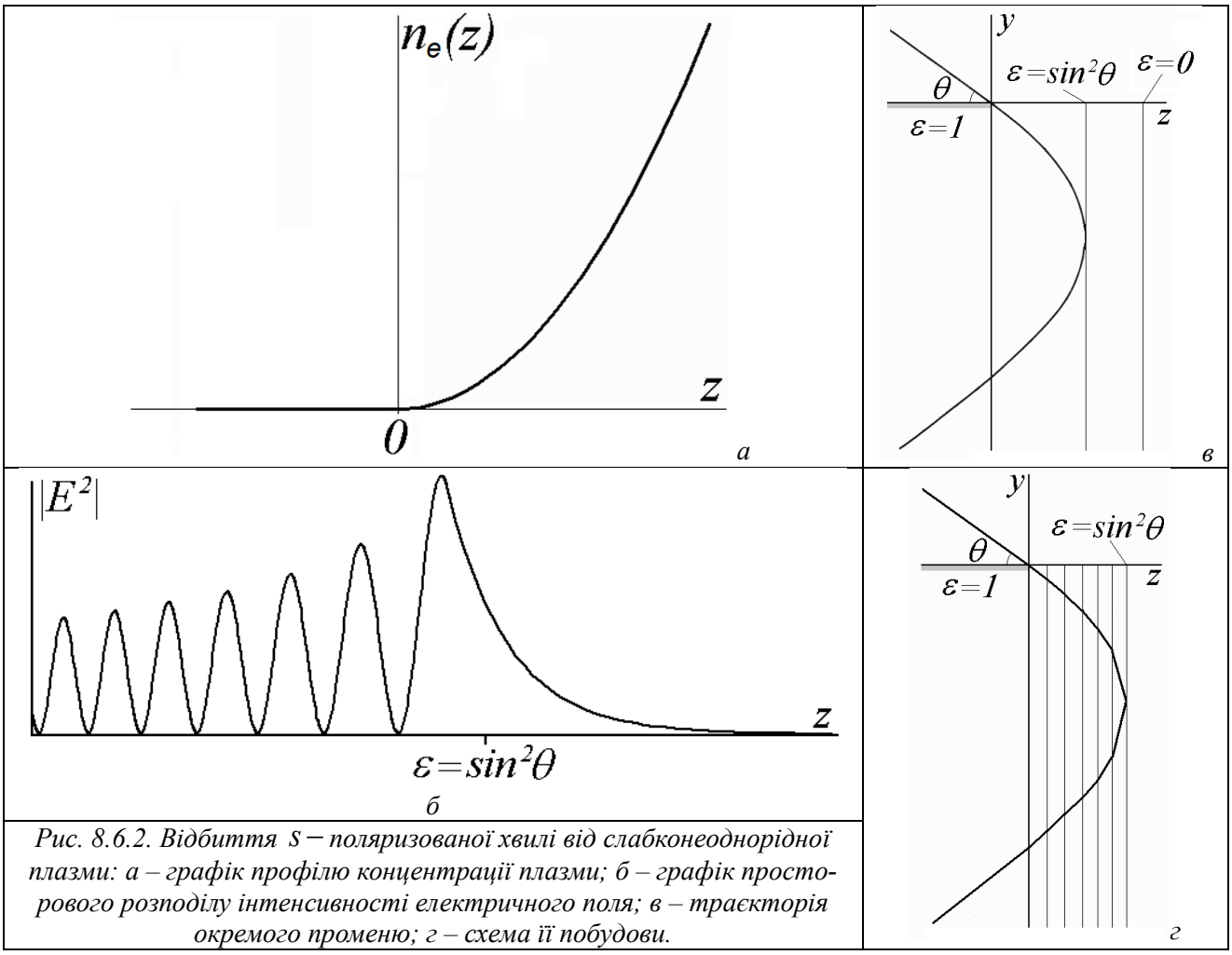


Рис. 8.6.2. Відбиття s – поляризованої хвилі від слабконеоднорідної плазми: а – графік профілю концентрації плазми; б – графік просторового розподілу інтенсивності електричного поля; в – траєкторія окремого променю; г – схема її побудови.

Щоб зрозуміти, як виглядає траєкторія окремого променю (вона показана на рис. 8.6.2 в), можна подумки розбити плазму на плоскі однорідні шари (рис. 8.6.2 г). При зро-

станні концентрації плазми її діелектрична проникність ε зменшується, і показник заломлення $n=\sqrt{\varepsilon}$ також зменшується. В результаті при заломленні на кожній межі шарів промінь все дужче відхиляється від нормалі до меж цих шарів. У точці повороту відбувається повне внутрішнє відбиття, і промінь змінює напрямок руху вздовж осі z на протилежний.

Бачимо, що в неоднорідній плазмі (і взагалі в середовищі, де показник заломлення плавно змінюється в просторі) промінь перестає бути прямою або ламаною лінією, а натомість являє собою деяку неперервну криву. Це явище відоме як рефракція.

8.6.6. Відбиття $p-$поляризованої хвилі від розмитої межі плазми. Точка локального плазмового резонансу

Розглянемо тепер падіння на розмиту межу плазми, описану вище (п. 8.6.4), $p-$поляризованої хвилі, у якої відмінні від нуля компоненти електричного поля E_x та E_z. Для неї зручніше розв'язувати хвильове рівняння для магнітного поля, яке має єдину ненульову компоненту B_x. Це хвильове рівняння неважко отримати з рівнянь Максвелла таким самим способом, як ми раніше отримували хвильові рівняння для електричного поля (8.6.4)-(8.6.6).

Запишемо рівняння, що визначає магнітне поле струму, з урахуванням струму зміщення (оскільки струм провідності плазми враховано в її діелектричній проникності, то вважаємо, що струм провідності відсутній, $\vec{j}=0$):

$$\left[\nabla\times\vec{B}\right]=ik_0\varepsilon(z)\vec{E} \tag{8.6.27}$$

(прийнято, що часова залежність поля є гармонічною з частотою ω, і використано позначення $k_0=\omega/c$). Узявши ротор від обох частин цього рівняння, отримаємо:

$$\left[\nabla\times\left[\nabla\times\vec{B}\right]\right]=ik_0\left[\nabla\times\left(\varepsilon\vec{E}\right)\right] \tag{8.6.28}$$

Тепер перетворимо праву та ліву частини (8.6.28) за відомими формулами векторного аналізу:

$$\left[\nabla\times\left(\varepsilon\vec{E}\right)\right]=\varepsilon\left[\nabla\times\vec{E}\right]+\left[\left(\nabla\varepsilon\right)\times\vec{E}\right];$$

$$\left[\nabla\times\left[\nabla\times\vec{B}\right]\right]=\nabla\left(\nabla\cdot\vec{B}\right)-\Delta\vec{B}=-\Delta\vec{B} \tag{8.6.29}$$

(в останньому співвідношенні враховано теорему Гаусса для магнітного поля).

Підставивши (8.6.29) до (8.6.28) і виключивши електричне поле за допомогою (8.6.27) та закону електромагнітної індукції

$$\left[\nabla\times\vec{E}\right]=-ik_0\vec{B}, \tag{8.6.30}$$

отримаємо:

$$-\Delta\vec{B}=k_0^2\varepsilon\vec{B}+\left[\left(\nabla\varepsilon\right)\times\left[\frac{1}{\varepsilon}\nabla\times\vec{B}\right]\right], \tag{8.6.31}$$

або остаточно

$$\Delta\vec{B}+\frac{1}{\varepsilon}\left[\left(\nabla\varepsilon\right)\times\left[\nabla\times\vec{B}\right]\right]+k_0{}^2\varepsilon\vec{B}=0. \tag{8.6.32}$$

Врахуємо, що $\varepsilon=\varepsilon(z)$, а залежність від x відсутня. Тоді векторне рівняння (8.6.27) можна переписати у формі двох скалярних:

$$E_z=-\frac{1}{ik_0\varepsilon(z)}\frac{\partial B_x}{\partial y}, \qquad E_y=\frac{1}{ik_0\varepsilon(z)}\frac{\partial B_x}{\partial z}. \tag{8.6.33}$$

Шукатимемо розв'язок хвильового рівняння (8.6.32) у формі гармонічної залежності

від y і t, яка нав'язується падаючою хвилею, $\vec{B}(\vec{r},t)=\vec{e}_x B(z)\exp(i\omega t-ik_y y)$. Врахувавши, що

$$rot\,\vec{B}=\begin{pmatrix}0\\ dB/dz\\ ik_y B\end{pmatrix},\qquad \nabla\varepsilon=\vec{e}_z\frac{d\varepsilon}{dz},\qquad \left[\nabla\varepsilon\times rot\,\vec{B}\right]=\begin{vmatrix}\vec{e}_x & \vec{e}_y & \vec{e}_z\\ 0 & 0 & d\varepsilon/dz\\ 0 & dB/dz & ik_y B\end{vmatrix}=-\vec{e}_x\frac{d\varepsilon}{dz}\frac{dB}{dz},$$

можна переписати (8.6.32) у вигляді:

$$\frac{d^2B}{dz^2}-\frac{1}{\varepsilon}\frac{d\varepsilon}{dz}\frac{dB}{dz}+\left(\varepsilon k_0{}^2-k_y^2\right)B=0. \tag{8.6.34}$$

Можна показати, що в околі точки, в якій $\varepsilon=0$, розв'язок рівняння (8.6.34) не має особливості, тобто там наближено виконується умова $B(z)\approx const$. Тому в цій області, відповідно до першого зі співвідношень (8.6.33),

$$E_z=-\frac{k_y B}{k_0\varepsilon(z)}, \tag{8.6.35}$$

тобто електричне поле у точці $\varepsilon=0$ має полюс першого порядку (рис. 8.6.3 а, пор. з рис. 8.6.2 б). Ця особливість електричного поля відповідає збудженню коливань у точці локального плазмового резонансу z_{res}, де частота хвилі збігається з електронною ленгмюрівською частотою, $\omega=\omega_p(z_{res})$.

Справді, електричне поле $p-$поляризованої хвилі має $z-$компоненту, паралельну до градієнту концентрації. Ця компонента збуджує резонансні коливання електронів у точці локального резонансу, які супроводжуються порушенням електронейтральності плазми, неоднорідної в напрямку z.

Врахування зіткнень електронів із важкими частинками або ненульової температури електронів обмежує амплітуду коливань електричного поля в резонансній точці. Цей результат цілком аналогічний до обмеження особливості резонансної кривої ідеального коливного контуру при врахуванні дисипації. При врахуванні зіткнень енергія електричного поля витрачається на розігрів плазми. При врахуванні температури плазми в околі точки локального плазмового резонансу буде збуджуватись ленгмюрівська хвиля[8], що відходить у бік розрідженої плазми. На збудження цієї хвилі також витрачається енергія.

Можна показати, що електричне поле в околі точки локального плазмового резонансу при врахуванні лише зіткнень електронів з важкими частинками можна записати у формі

$$\left|E_z(z_0)\right|=\frac{\omega E_m\Phi(\tau)}{\nu\sqrt{2\pi k_0 L}},\qquad \tau=(k_0L)^{1/3}\sin\theta, \tag{8.6.36}$$

де E_m – амплітуда падаючої хвилі, ν – частота зіткнень електронів з важкими частинками, L – характерний розмір неоднорідності плазми в області локального плазмового резонансу. Графік функції $\Phi(\tau)$ поданий на рис. 8.6.3 б.

При малих кутах θ падіння $p-$поляризованої хвилі на плазму амплітуда поздовжніх коливань у точці локального плазмового резонансу прямує до нуля, оскільки при $\theta=0$ електричне поле падаючої хвилі стає суто поперечним (щодо градієнту концентрації плазми) і не може розгойдувати поздовжні коливання в резонансній точці. Навпаки, при великих θ зростає бар'єр непрозорості між точкою повороту, де $\varepsilon=\sin^2\theta$, і точкою резонансу, де $\varepsilon=0$, в якому поле падаючої хвилі спадає за законом експоненціального типу (рис. 8.6.3 а).

Коливання в точці локального плазмового резонансу можна збудити й іншими спосо-

[8] Ленгмюрівська хвиля – це аналог акустичної хвилі в електронному газі (при цьому іони вважаються нерухомими). В такій хвилі збурення електронної концентрації приводить до порушення електронейтральності і появи потенціального електричного поля.

бами – наприклад, помістивши неоднорідну плазму в змінне електричне поле, паралельне до градієнту її концентрації.

Таким чином, специфіка відбиття $p-$поляризованої хвилі від слабконеоднорідної плазми полягає в можливому збудженні нею коливань в області локального плазмового резонансу. В усьому іншому картина аналогічна відбиттю $s-$поляризованої хвилі.

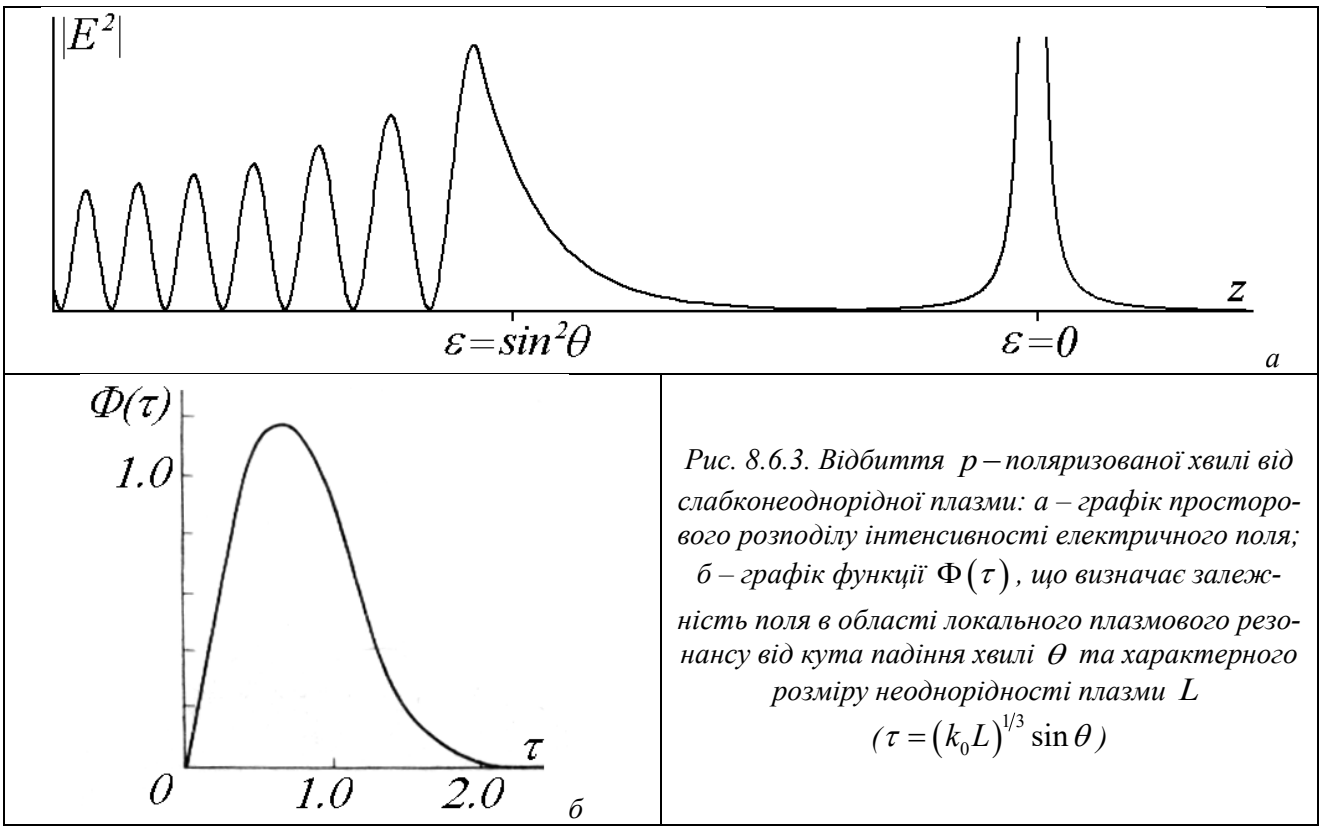


Рис. 8.6.3. Відбиття $p-$поляризованої хвилі від слабконеоднорідної плазми: а – графік просторового розподілу інтенсивності електричного поля; б – графік функції $\Phi(\tau)$, що визначає залежність поля в області локального плазмового резонансу від кута падіння хвилі θ та характерного розміру неоднорідності плазми L ($\tau = (k_0 L)^{1/3} \sin\theta$)

Контрольні питання до підрозділу 8.6

1. Коли в задачі про відбиття електромагнітних хвиль від слабконеоднорідної плазми останню можна вважати плоскошаруватою?
2. Що розуміють під *s*- і *p*-поляризованими хвилями в задачі про відбиття електромагнітних хвиль від слабконеоднорідної плазми?
3. В чому полягає наближення геометричної оптики при аналізі поширення хвиль у слабконеоднорідному середовищі?
4. Чому при наближенні до точки повороту амплітуда електромагнітної хвилі в плазмі зростає?
5. Що таке розподілене відбиття хвилі від слабконеоднорідної плазми? В яких областях воно максимальне?
6. Що таке область локального плазмового резонансу? Чому коливання в ній збуджуються лише *p*-поляризованими хвилями?

Задачі до підрозділу 8.6

8.6.1. Поширюючись уздовж осі z, нормальної до межі плазми, хвиля з частотою 0.3 ГГц переходить з вакууму в плазму з концентрацією $3\cdot10^7\,\text{см}^{-3}$. Межа плазми розмита, так що відбиттям від неї можна знехтувати. Як зміниться в плазмі амплітуда та групова швидкість хвилі порівняно з вакуумом?

8.6.2. Профіль концентрації плазми має вигляд

$$n_e(z) = \frac{n_0}{2}\left[1 + th(z/L)\right],$$

$$n_0 = 10^8\,\text{см}^{-3},\ L = 2\,\text{м}.$$

Електромагнітна хвиля з частотою 1.5 ГГц поширюється вздовж осі z. Оцініть коефіцієнт її відбиття від неоднорідності плазми.

8.6.3. Концентрація плазми змінюється за законом

$$n_e(z) = \begin{cases} 0, & z \le 0; \\ n_0 (z/L)^2, & z > 0, \end{cases}$$

$$n_0 = 10^{10}\,\text{см}^{-3},\ L = 10\,\text{м}.$$

Хвильовий вектор падаючої на плазму *p*-поляризованої хвилі з частотою 1 ГГц при $z \le 0$ спрямований під кутом 30º до осі z. Знайти координати точки повороту та точки локального плазмового резонансу.

8.7. Поширення радіохвиль в атмосфері Землі

Тепер застосуємо уявлення, викладені в п. 8.6, до поширення радіохвиль в іоносфері, нехтуючи цього разу впливом Землі. Такий підхід застосовний, якщо головна пелюстка діаграми спрямованості передавальної антени спрямована у верхню півкулю під не дуже малим кутом до горизонту, коли поверхня Землі майже не потрапляє до суттєвої області (рис. 8.7.1). Іншими словами, це крайній випадок, протилежний збудженню земної хвилі.
Очевидно, що в загальному випадку на поширення радіохвиль впливатиме як іоносфера, так і поверхня Землі.

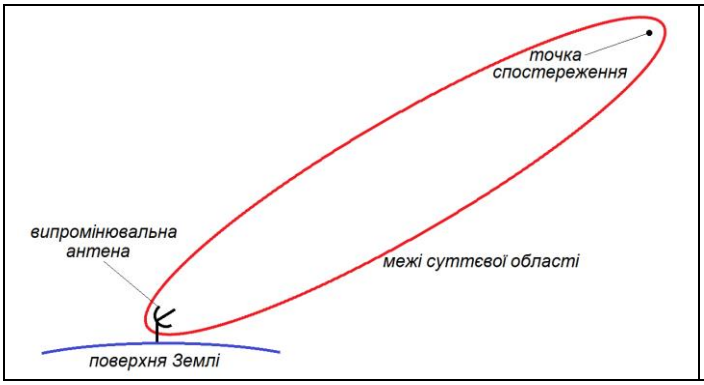


Рис. 8.7.1. До пояснення умови відсутності впливу земної поверхні на поширення радіохвиль

8.7.1. Деякі числові оцінки параметрів іоносфери

У попередньому підрозділі задача про поширення електромагнітних хвиль розглядалася для моделі ізотропної плазми без зіткнень, а її діелектрична проникність записувалася у формі (8.1.23):

$$\varepsilon = 1 - \frac{\omega_p^2}{\omega^2}, \qquad \omega_p = \sqrt{\frac{4\pi e^2 n_e}{m}},$$

ω_p – електронна ленгмюрівська частота.

З'ясуємо умови застосовності цієї моделі до іоносферної плазми.

Умовно вважатимемо, що іоносферна плазма суттєво впливає на поширення радіохвиль тоді, коли в області, де її концентрація досягає максимуму, діелектрична проникність стає від'ємною або, принаймні, обертається на нуль. Іншими словами, вплив іоносферної плазми суттєвий на частотах $\omega \le \omega_p(n_{max})$.

Як уже зазначалося, максимум електронної концентрації іоносфери досягається в шарі F2 (на висоті приблизно 350 км). Його величина вдень складає 10^6 см$^{-3}$, а вночі – $3\cdot10^5$ см$^{-3}$ (рис. 8.2.2). Цим концентраціям відповідають плазмові частоти $5.7\cdot10^7$ с$^{-3}$ ($\lambda = 30$ м) та $3.1\cdot10^7$ с$^{-3}$ ($\lambda = 60$ м).

Таким чином, у діапазоні $\lambda \ge 30$ м вплив іоносфери можна вважати суттєвим.

Тепер з'ясуємо, коли плазму можна вважати ізотропною, тобто знехтувати впливом геомагнітного поля. Для цього повинна виконуватися умова

$$\omega_{ce} \ll \omega_p, \qquad \omega_{ce} = \frac{eB}{mc} \tag{8.7.1}$$

ω_{ce} – електронна циклотронна частота (частота обертання електронів навколо силових ліній магнітного поля)[9], B – індукція геомагнітного поля. Як уже вказувалося, над поверхнею Землі ця величина змінюється в межах (0.2-0.6) Гс; це поле максимальне над магнітними полюсами і мінімальне в екваторіальній області. Йому відповідають циклотронні частоти (1.8-

[9] Для числових розрахунків зручно користуватися формулою $\omega_{ce}/2\pi\,[\text{МГц}] = 2.80\,B\,[\text{Гс}]$.

5.3)$\cdot 10^6$ с$^{-3}$. Порівнюючи з отриманими раніше значеннями ω_p, бачимо, що умова (8.7.1) добре виконується в шарі F2, особливо вдень та в екваторіальній області.

Нарешті, зіткненнями електронів з іонами та нейтральними атомами можна нехтувати за умови, що їхня частота ν значно менша від частоти сигналу ω. В шарі F2 частота ν близька до 10^2Гц, тобто умова $\nu \ll \omega$ на відповідних висотах добре виконується для частот понад 1 кГц ($\lambda \leq 300$ км).

Таким чином, модель ізотропної плазми без зіткнень правильно описує поведінку іоносфери в шарі F2 для радіохвиль з довжинами 100 м-300 км.

8.7.2. Форма променю в приземній атмосфері

Обговоримо спершу в загальному вигляді влив зміни показника заломлення приземної атмосфери з висотою на форму променів, які через неї проходять. В тропосфері зміна показника заломлення зумовлена зміною властивостей нейтральної атмосфери (концентрації азоту, кисню та водяної пари) з висотою. В іоносфері така зміна значно помітніша й зумовлена висотною зміною концентрації іоносферної плазми (див. рис. 8.2.2).

Для визначення форми та кривини променю[10] в приземній атмосфері використаємо найпростішу модель сферично шаруватого середовища, тобто вважатимемо, що останнє складається з однорідних сферичних шарів (сферичний шар – область, обмежена двома концентричними сферами з дещо відмінними радіусами) – див. рис. 8.7.2-8.7.3. Вважаємо ці шари тонкими. Тоді відмінність їхніх показників заломлення мала, і амплітуда відбитої хвилі теж настільки мала, що її можна не враховувати (всюди, крім околу точки повороту, див. вище п. 8.6). Всередині кожного шару промінь є прямим, а на межах сусідніх шарів він заломлюється.

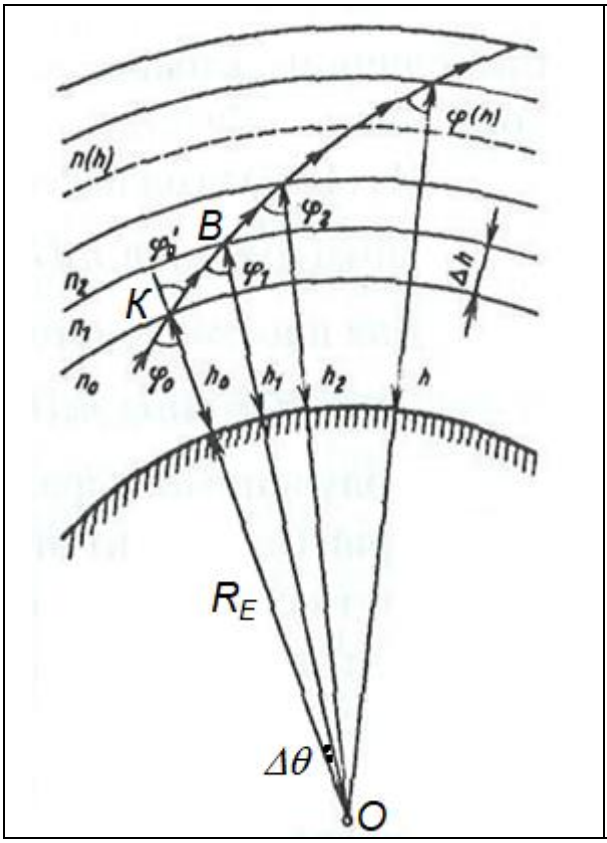


Рис. 8.7.2. До визначення ходу променю в іоносфері

[10] Тут і нижче ми користуємося наближенням геометричної оптики, нехтуючи можливими ефектами дифракції та інтерференції, оскільки аналіз цих ефектів в неоднорідному середовищі різко ускладнює розрахунок.

Форму променю описуватимемо залежністю $\varphi(h)$, де $\varphi(h)$ – кут між променем та радіус-вектором, проведеним з центру Землі в точку променю на висоті h. В іоносфері вона залежить від початкового значення $\varphi_0(h_0)$, де висота h_0 відповідає нижньому краю іоносфери. Показник заломлення нижче від іоносфери (тобто в стратосфері) – n_0.

Нехай найнижчому сферичному шару іоносфери відповідає кут заломлення φ_0' і показник заломлення n_1. Тоді для нижньої межі цього шару закон заломлення набуває вигляду

$$n_0 \sin\varphi_0 = n_1 \sin\varphi_0'. \tag{8.7.2}$$

З трикутника *OKB* (рис. 8.7.2) за теоремою синусів

$$\frac{\sin\varphi_1}{R_E + h_0} = \frac{\sin(\pi - \varphi_0')}{R_E + h_1} = \frac{\sin\varphi_0'}{R_E + h_1}. \tag{8.7.3}$$

(R_E, як і раніше – радіус Землі). Підставляючи $\sin\varphi_0'$ з (8.7.2), отримуємо:

$$\sin\varphi_1 = \sin\varphi_0 \frac{n_0(h_0)}{n_1(h_1)} \frac{R_E + h_0}{R_E + h_1} \tag{8.7.4}$$

Послідовно використовуючи формули вигляду (8.7.4) до сусідніх шарів з більшими номерами, можна отримати для довільної висоти h

$$\sin\varphi(h) = \sin\varphi_0 \frac{n_0(h_0)}{n(h)} \frac{R_E + h_0}{R_E + h}. \tag{8.7.5}$$

Формула (8.7.5) описує форму $\varphi(h)$ променю в сферично-симетричній іоносфері (або в будь-якому іншому сферично-симетричному середовищі, де на висоті h_0 напрямок променю характеризується кутом $\varphi_0(h_0)$).

Спрямовуючи R_E до нескінченності (і тим самим нехтуючи кривиною поверхні Землі, тобто повертаючись до моделі плоскошаруватої плазми), приходимо до співвідношення

$$\sin\varphi(h) = \sin\varphi_0 \frac{n_0(h_0)}{n(h)},$$

або, врахувавши, що $n_0(h_0) = 1$

$$\sin\varphi(h) = \frac{1}{n(h)} \sin\varphi_0, \tag{8.7.6}$$

яке має вигляд звичайного закону заломлення і відповідає плоско-шаруватому середовищу.

Спробуємо тепер записати траєкторію променю в полярній системі координат $r = R_E + h, \theta$, центр якої суміщений з центром Землі (площина цієї полярної системи координат збігається з площиною поширення променю). Кут θ характеризує напрямок на точку спостереження, тому ним зручніше користуватися, ніж кутом φ.

У першому шарі приросту радіуса $\Delta r = \Delta h$ відповідає приріст полярного кута

$$r\Delta\theta \approx \Delta h \, tg\varphi_1, \tag{8.7.7}$$

звідки

$$\Delta\theta = \frac{\Delta h}{r} \frac{\sin\varphi_1}{\sqrt{1 - \sin^2\varphi_1}}. \tag{8.7.8}$$

Врахувавши (8.7.6), отримуємо

$$\Delta\theta = \frac{\Delta h}{r}\frac{\sin\varphi_0 \dfrac{n_0(h_0)}{n(h)}}{\sqrt{1-\left(\sin\varphi_0 \dfrac{n_0(h_0)}{n(h)}\right)^2}}, \tag{8.7.9}$$

або, після переходу від скінченних приростів до диференціалів,

$$d\theta = \frac{dh}{r}\frac{n_0(h_0)\sin\varphi_0}{\sqrt{n^2(h)-\left(n_0(h_0)\sin\varphi_0\right)^2}} = \frac{dh}{R_E+h}\frac{n_0(h_0)\sin\varphi_0}{\sqrt{n^2(h)-\left(n_0(h_0)\sin\varphi_0\right)^2}}. \tag{8.7.10}$$

Знаючи явний вигляд залежності $n(h)$, можна проінтегрувати рівняння (8.7.10) і отримати явний вигляд траєкторії променю.

Знаменник у правій частині формули (8.7.10) обертається на нуль у точці, де

$$n^2(h)-\left(n_0(h_0)\sin\varphi_0\right)^2 = 0,$$

або

$$\varepsilon(h) = \varepsilon(h_0)\sin^2\varphi_0. \tag{8.7.11}$$

На нижній межі іоносфери можна вважати, що $\varepsilon(h_0)=1$, тобто для іоносфери умова (8.7.11) набуває вигляду

$$\varepsilon(h) = \sin^2\varphi_0. \tag{8.7.12}$$

Точка, в якій виконується ця умова – це вже знайома нам точка повороту електромагнітної хвилі[11] (див. вище п.8.6). З формули (8.7.10) видно, що в ній $dh/d\theta = 0$, тобто промінь іде паралельно до земної поверхні.

Досі йшлося про довільне сферично-симетричне середовище. Тепер конкретизуємо отримані формули для плазми.

Підставивши до (8.7.12) формулу (8.1.26)

$$\varepsilon = 1-\frac{\omega_p^2}{\omega^2},$$

отримаємо:

$$n_e = \frac{m\omega^2\cos^2\varphi_0}{4\pi e^2}, \tag{8.7.13}$$

або

$$\omega = \frac{\omega_p}{\cos\varphi_0} = \frac{e}{\cos\varphi_0}\sqrt{\frac{4\pi n_e}{m}}. \tag{8.7.14}$$

Формула (8.7.13) визначає концентрацію електронів у шарі слабконеоднорідної плоскошаруватої плазми, від якого відбивається хвиля частоти ω, що падає на цю плазму з вакууму під кутом φ_0. Формула (8.7.14), навпаки, визначає частоту хвилі, яка відіб'ється від шару з концентрацією n_e при її падінні на плазму з вакууму під кутом φ_0.

8.7.3. Радіус кривини променю в іоносферній плазмі

Крім залежності $\varphi(h)$ (або $\theta(r)$), форму променю можна характеризувати також радіусом кривини в різних його точках. Як ми побачимо далі, радіус кривини визначає поведінку променю в сферично-симетричному середовищі (див. нижче п. 8.7.5).

---

[11] За певних умов поява точки повороту може бути пов'язана з висотною зміною не тільки концентрації іоносферної плазми, а й властивостей тропосфери, див. нижче п. 8.7.4.

Для того, щоб визначити радіус кривини, розглянемо тонкий сферичний шар завтовшки Δh, що знаходиться на висоті h (рис. 8.7.3 а). На нижній межі цього шару промінь заломлюється, і кут падіння змінюється на величину $\Delta\varphi$. Радіус кривини променю визначається як

$$R = \lim_{\Delta h \to 0} \frac{KB}{\Delta\varphi} = \lim_{\Delta h \to 0} \frac{\Delta h}{\Delta\varphi \cos\varphi(h)} = \frac{dh/d\varphi}{\cos\varphi(h)}, \tag{8.7.15}$$

де KB – довжина відрізку між точками K і B на рис. 8.7.3 а.

Геометричний зміст радіуса кривини зрозумілий з рис. 8.7.3 б: це радіус кола, яке найкращим чином апроксимує плоску криву в околі деякої її точки.

Формулу (8.7.6) перепишемо у вигляді

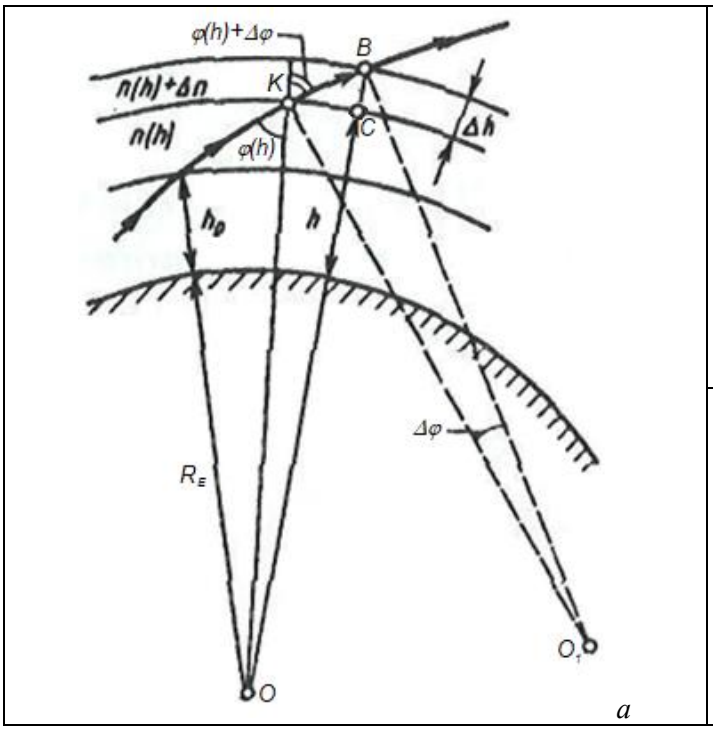
$$n(h)\sin\varphi(h) = \sin\varphi_0 \tag{8.7.16}$$

і візьмемо похідну за h від її обох частин. Отримаємо:

$$\frac{dn}{dh}\sin\varphi(h) - n(h)\cos\varphi(h)\frac{d\varphi}{dh} = 0 \tag{8.7.17}$$

звідки

$$\frac{d\varphi}{dh} = \frac{dn}{dh}\frac{tg\varphi(h)}{n(h)}. \tag{8.7.18}$$



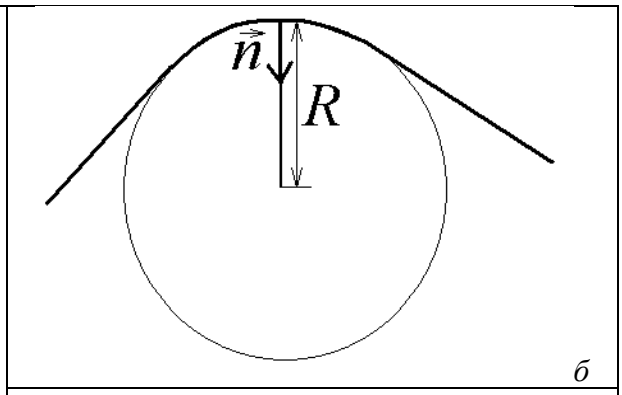


Рис. 8.7.3. До визначення радіуса кривини променю в іоносфері (а) та до пояснення геометричного змісту радіуса кривини плоскої кривої (б)

Підставивши (8.7.18) до (8.7.15), отримаємо остаточну формулу для радіуса кривини:

$$R = \left(\frac{d\varphi}{dh}\right)^{-1}\frac{1}{\cos\varphi(h)} = \left(\frac{dn}{dh}\frac{tg\varphi(h)}{n(h)}\right)^{-1}\frac{1}{\cos\varphi(h)} = \left(\frac{dn}{dh}\right)^{-1}\frac{n(h)}{\cos\varphi(h)tg\varphi(h)} = \left|\frac{dn}{dh}\right|^{-1}\frac{n(h)}{\sin\varphi(h)}. \tag{8.7.19}$$

Модуль у формулі (8.7.19) записаний для того, щоб радіус кривини завжди був додатним.

Величину, обернену до радіусу кривини лінії, називають її кривиною a. В нашому випадку

$$a(h) = \frac{1}{R(h)} = \left|\frac{dn}{dh}\right| \frac{\sin\varphi(h)}{n(h)}. \qquad (8.7.20)$$

В однорідному середовищі, де $dn/dh = 0$, отримуємо, що $\rho \to \infty$ – промінь має форму прямої, якій відповідає нескінченний радіус кривини й, відповідно, нульова кривина.

8.7.4. Траєкторії хвиль у тропосфері

Нагадаємо, що тропосфера простягається від поверхні Землі до висот 7-10 км (у полярних областях) до висот 16-18 км (в екваторіальних областях). В тропосфері температура зменшується з висотою. На таких висотах атмосферу можна вважати нейтральною і користуватися для діелектричної проникності формулою (8.1.32)

$$\varepsilon = 1 + \frac{\alpha_1 p}{T} + \frac{\alpha_2 p_w}{T^2},$$

де другий доданок характеризує внесок кисню та азоту (p – атмосферний тиск) а третій – внесок водяної пари (p_w – її парціальний тиск).

Атмосферний тиск спадає з висотою, тому для сухого повітря ($p_w \to 0$) діелектрична проникність і, відповідно, показник заломлення з висотою зменшуються.

Для вологого повітря, коли основну роль відіграє третій доданок у (8.1.32), висотна залежність діелектричної проникності й показника заломлення визначається функцією $p_w(h)$, яка може з висотою як спадати, так і зростати.

Стандартний градієнт діелектричної проникності в тропосфері складає величину $d\varepsilon/dh \equiv g = -7.85 \cdot 10^{-8}\, м^{-1}$ (знак «мінус» вказує на спадання ε з висотою). Відповідно характерний розмір неоднорідності $L = \varepsilon \left| d\varepsilon/dh \right|^{-1}$ можна оцінити як 13000 км, тобто в межах тропосфери зміна показника заломлення майже непомітна.

Як уже зазначалося, висотний розподіл водяної пари може помітно вплинути на профіль $\varepsilon(h)$ і, відповідно, на поширення радіохвиль у тропосфері (конкретніше – на характер їхньої рефракції). Розглянемо на якісному рівні, до яких результатів це може привести (рис. 8.7.4).

Крива 1 відповідає від'ємній рефракції, або субрефракції. Для цього показник заломлення повинен зростати з висотою. Така залежність можлива при зростанні з висотою тиску водяної пари.

Крива 2 (прямолінійне поширення хвиль) спостерігається за в однорідному середовищі.

Криві 3-5 відповідають зменшенню показника заломлення з висотою (додатна рефракція).

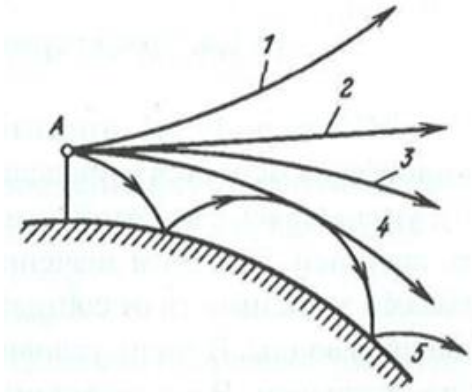


Рис. 8.7.4. Різні типи рефракції радіохвиль у тропосфері (рисунок носить якісний характер)

Тут розрізняють кілька окремих випадків у залежності від величини g:

– для стандартного градієнту діелектричної проникності, як уже зазначалося, форма променів майже не відрізняється від прямолінійної;

– крива 3 відповідає підвищеній рефракції ($g \le -8 \cdot 10^{-8}$ м$^{-1}$);

– крива 4 відповідає так званій критичній рефракції, коли промінь стає паралельним земній поверхні. Такий режим реалізується при $g = -31.4 \cdot 10^{-8}$ м$^{-1}$;

– крива 5 відповідає надрефракції, або хвилеводному режиму ($g < -31.4 \cdot 10^{-8}$ м$^{-1}$). При цьому спостерігається розподілене відбиття хвиль від верхніх шарів тропосфери. За відповідних умов та сама хвиля відбивається й від Землі, після чого ефект повторюється. Забігаючи наперед, можна вказати, що хвилеводний режим більш характерний для іоносфери.

На завершення зазначимо, що в стратосфері (висоти від 7-18 км до приблизно 60 км) атмосфера залишається нейтральною, а концентрація нейтральних атомів на 1-2 порядки менша, ніж біля земної поверхні (див. вище рис.8.2.1 а). Тому відмінність діелектричної проникності від одиниці в середньому на порядок менша, ніж у тропосфері, і практично не впливає на поширення радіохвиль.

8.7.5. Траєкторії радіохвиль в іоносфері

Як уже зазначалося, в іоносфері діелектрична проникність, яка залежить від концентрації електронів і частоти хвилі, завжди менша за одиницю і може стати від'ємною (див. вище рис. 8.2.5). Тому в іоносфері явище рефракції включно з розподіленим відбиттям проявляється набагато сильніше, ніж у тропосфері.

У внутрішній іоносфері (на висотах до 350 км – див. рис. 8.2.2) промені відхиляються вниз (додатна рефракція), у зовнішній – угору (від'ємна рефракція). Додатна рефракція у внутрішній іоносфері може привести до розподіленого відбиття променів (рис. 8.7.5, пор. з рис. 8.6.1 в).

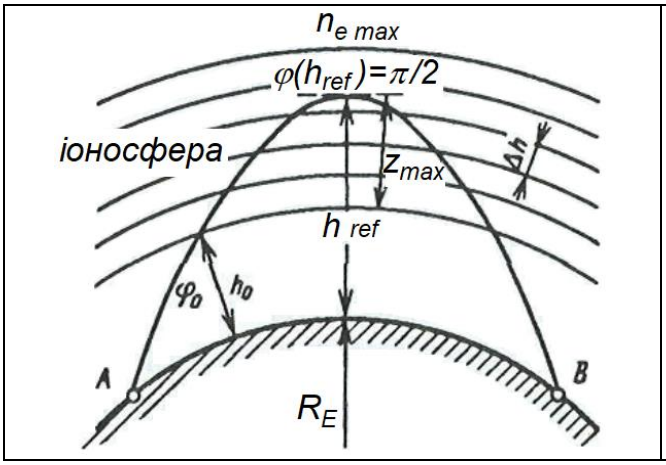


Рис. 8.7.5. Розподілене відбиття променів від нижньої іоносфери

Якщо хвиля рухається зі стратосфери в іоносферу, а кут падіння складає φ_0, то висоту розташування точки повороту h_{ref} (без урахування кривини Землі) можна знайти з умови (8.7.13):

$$n_e\left(h_{ref}\right) = \frac{m\omega^2 \cos^2 \varphi_0}{4\pi e^2}.$$

Врахування кривини Землі призводить до того, що не всі відбиті від іоносфери хвилі можуть повернутися на її поверхню. Таке повернення можливе за виконання наближеної умови

$$R\left(h_{ref}\right) < R_E + h_{ref}\,, \tag{8.7.21}$$

де $R\left(h_{ref}\right)$ – радіус кривини траєкторії хвилі в точці повороту. Зміст цієї умови зрозумілий з рис. 8.7.6. Наближення полягає в тому, що траєкторія ототожнюється з колом, радіус якого дорівнює радіусу кривини траєкторії в точці повороту.

Радіус кривини в точці повороту променю можна знайти з формули (8.7.19)

$$R = \left|\frac{dn}{dh}\right|^{-1} \frac{n(h)}{\sin\varphi(h)}\,,$$

врахувавши, що $\varphi\left(h_{ref}\right) = \pi/2$:

$$R = \frac{n(h)}{dn/dh}\,. \tag{8.7.22}$$

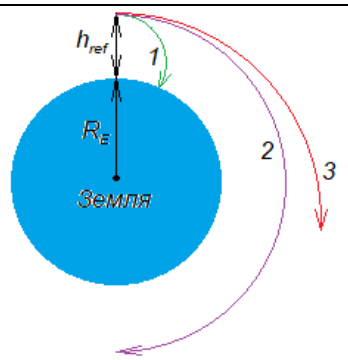


Рис. 8.7.6. До інтерпретації умови повернення на поверхню Землі відбитого від іоносфери променю: крива 1 – $R\left(h_{ref}\right) < R_E + h_{ref}$; *крива 2 –* $R\left(h_{ref}\right) = R_E + h_{ref}$; *крива 3 –* $R\left(h_{ref}\right) > R_E + h_{ref}$

Нагадаємо, що тут $n(h)$ – висотний профіль показника заломлення:

$$n(h) = \sqrt{\varepsilon(h)} = \sqrt{1 - \frac{4\pi n_e(h) e^2}{m\omega^2}}\,,$$

де $n_e(h)$ – висотний профіль електронної концентрації. Тоді

$$\frac{dn}{dh} = \frac{1}{2\sqrt{\varepsilon(h)}}\left(-\frac{4\pi e^2}{m\omega^2}\frac{dn_e}{dh}\right) = -\frac{2\pi e^2}{m\omega^2\sqrt{\varepsilon(h)}}\frac{dn_e}{dh} \tag{8.7.23}$$

Грубо апроксимуючи висотний профіль електронної концентрації іоносфери (див. рис. 8.2.2) прямою вигляду

$$n_e(h) = n_{e\,max} h / h_{max}\,, \tag{8.7.24}$$

тобто не враховуючи існування тропосфери та стратосфери, можна отримати оцінку для радіусу кривини променю в точці повороту:

$$R\left(h_{ref}\right) = \frac{n\left(h_{ref}\right)}{dn/dh} = n\left(h_{ref}\right)\left[-\frac{2\pi e^2}{m\omega^2 n\left(h_{ref}\right)}\frac{dn_e}{dh}\right]^{-1} = n\left(h_{ref}\right)\left[\frac{2\pi e^2}{m\omega^2 n\left(h_{ref}\right)}\frac{n_{e\,\max}}{h_{\max}}\right]^{-1} =$$

$$= \frac{m\omega^2 n^2\left(h_{ref}\right) h_{\max}}{2\pi e^2 n_{e\,\max}} = \frac{m\omega^2 h_{\max}}{2\pi e^2 n_{e\,\max}}\varepsilon\left(h_{ref}\right) = \frac{m\omega^2 h_{\max}}{2\pi e^2 n_{e\,\max}}\sin^2\varphi_0 = \frac{2\omega^2}{\omega_{p\,\max}^2} h_{\max}\sin^2\varphi_0\,. \tag{8.7.25}$$

Як бачимо, радіус кривини зменшується зі зменшенням частоти хвилі й кута падіння (відрахованого від нормалі до поверхні Землі). Зокрема, при $\varphi_0 = 0$ промінь іде вертикально вгору і після відбиття від іоносфери повертається в те саме місце, тобто його радіус кривини в точці повороту справді дорівнює нулеві. Зростання ρ при зростанні частоти пов'язане з тим, що при цьому точка повороту піднімається вгору, і відповідно збільшується довжина траєкторії хвилі в іоносфері, де спостерігається рефракція.

Приймемо, що $\omega/\omega_{p\,\max} = 1/\sqrt{2}$, $\varphi_0 = \pi/4$, $\sin\varphi_0 = 1/\sqrt{2}$. Тоді з (8.7.13), (8.7.24) отримаємо:

$$n_e\left(h_{ref}\right) = n_{e\,\max}\frac{h_{ref}}{h_{\max}} = \frac{m\omega^2\cos^2\varphi_0}{4\pi e^2},$$

або

$$h_{ref} = h_{\max}\cos^2\varphi_0\frac{m\omega^2}{4\pi n_{e\,\max}e^2} = h_{\max}\cos^2\varphi_0\left(\frac{\omega}{\omega_{p\,\max}}\right)^2, \qquad (8.7.26)$$

звідки $h_{ref} \approx 90$ км. Відповідно до (8.7.25) $\rho\left(h_{ref}\right) \approx 175$ км.

Формула (8.7.14) при для антени, що випромінює вертикально вгору, набуває вигляду

$$\omega_{cr} = \sqrt{\frac{4\pi n_e e^2}{m}}$$

і визначає так звану критичну частоту плазми з концентрацією n_e (вона збігається за величиною з електронною ленгмюрівською частотою для такої концентрації). Ця частота відповідає хвилі, що випромінюється вертикально вгору й відбивається від точки з концентрацією n_e. Цю величину визначають за допомогою іоносферних станцій – радіолокаторів зі змінною частотою, у яких головна пелюстка діаграми спрямованості напрямлена вгору. На рис. 8.7.7 подані графіки, що ілюструють зміну критичних частот різних шарів іоносфери під дією різних факторів: добового обертання Землі, обертання її навколо Сонця, густини потоку сонячної радіації W та широти точки спостереження.

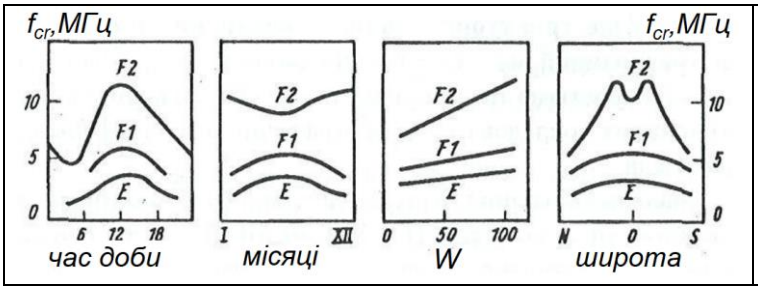


Рис. 8.7.7. Графіки залежності критичних частот різних шарів іоносфери від часу доби, пори року, густини потоку енергії Сонця та широти

Обговоримо тепер хід траєкторій радіохвиль, частота яких дещо більша за критичну частоту, що відповідає шару F2 (або максимальній концентрації іоносферної плазми). При вертикальному випромінюванні (або при кутах випромінювання до горизонту, близьких до 90º) такі хвилі не відбиваються, а проходять крізь іоносферу. При випромінюванні під невеликим кутом до горизонту (але так, щоб поверхня Землі не потрапляла до суттєвої зони) хвиля може відбитися донизу (рис. 8.7.8).

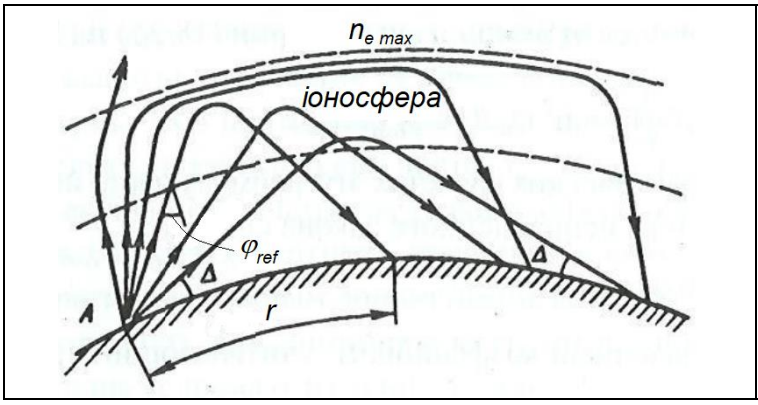


Рис. 8.7.8. Різні варіанти траєкторій радіохвиль в іоносфері (схематичне зображення)

8.7.6. Поглинання та розсіювання радіохвиль в атмосфері

Ослаблення електромагнітних хвиль при їхньому поширенні у вакуумі зумовлене розходженням променів від точкового джерела. При цьому амплітуда їхнього поля залежить від віддалі як $1/r$. У середовищі (або через вплив земної поверхні) може спостерігатися додаткове ослаблення електромагнітних хвиль. Один з можливих механізмів такого ослаблення – це розсіювання на випадкових неоднорідностях атмосферного газу або іоносферної плазми (див. вище п. 8.2). Іншим можливим механізмом виступає поглинання в середовищі. В останньому випадку інтенсивність променю (розходженням променю нехтуємо) спадає вздовж його траєкторії за експоненціальним законом:

$I(r) = I_0 \exp(-\gamma r)$

(γ – коефіцієнт поглинання). Ослаблення променю подається в децибелах,

$$10\lg(I/I_0) = 10\lg\left[\exp(-\gamma r)\right] = -10\gamma r \lg e = -43.4\gamma r. \qquad (8.7.27)$$

В залежності від механізму розрізняють нерезонансне та резонансне поглинання. Нерезонансне поглинання супроводжується нагріванням середовища, резонансне – збудженням молекул на вищі енергетичні рівні.

Причинами нерезонансного поглинання можуть бути скінченна провідність поверхні Землі (при поширенні земної хвилі), втрати на тертя молекул газу з ненульовими електричними та/або магнітними моментами при їхньому русі в полі хвилі (див. вище п. 8.1.3), джоулеві втрати при протіканні струму в середовищі з ненульовою провідністю (зокрема, в іоносферній плазмі) в електричному полі хвилі тощо.

<u>Поглинання радіохвиль у газах.</u> Резонансне поглинання в радіодіапазоні спостерігається для хвиль з частотами понад 10 ГГц ($\lambda < 3$ см). Таке поглинання відбувається, в першу чергу, в тропосфері, де атмосферний тиск максимальний. До вказаного діапазону потрапляють частоти, що відповідають коливним рівням деяких молекул. З числа газів, що входять до складу атмосфери, це молекулярний кисень та водяна пара (рис. 8.7.9). Видно, що молекулярний кисень має лінії поглинання на частотах 22 ГГц, 60 ГГц та 120 ГГц, а водяна пара – на частотах 183 ГГц та 320 ГГц

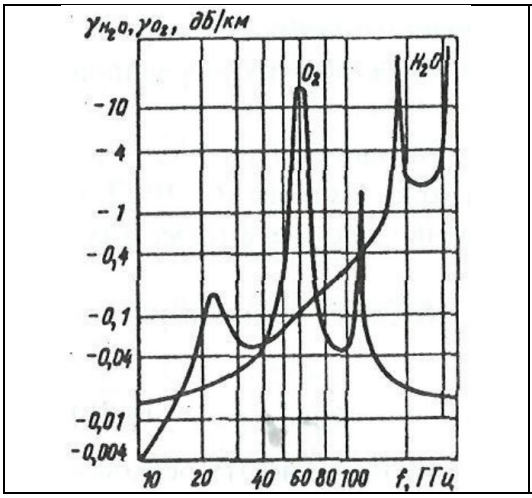


Рис. 8.7.9. Графік частотної залежності коефіцієнту поглинання для молекулярного кисню й водяної пари поблизу поверхні Землі

Ослаблення амплітуди хвилі внаслідок поглинання описують модулем множника ослаблення V (див. вище пп. 8.3.1, 8.4.4). Його величина задається співвідношенням

$$V[\text{дБ}] = \gamma_{H_2O}[\text{дБ/км}]\, r_{H_2O}[\text{км}] + \gamma_{O_2}[\text{дБ/км}]\, r_{O_2}[\text{км}], \qquad (8.7.28)$$

(пор. з формулою (8.7.27)), де γ_{H_2O} і γ_{O_2} – коефіцієнти ослаблення відповідно для водяної пари й молекулярного кисню, r_{H_2O} і r_{O_2} – ефективні довжини траєкторій хвиль для тих самих

газів, які враховують нерівномірне поглинання вздовж траєкторії. На наземних трасах $r_{H_2O} \approx r_{O_2} \approx r$, де r – довжина траси.

На рис. 8.7.10 наведено графіки залежностей множника ослаблення від частоти при різних кутах між променем і горизонтом, розрахованих за умови, що промінь повністю перетинає тропосферу, яка перебуває в спокійному стані.

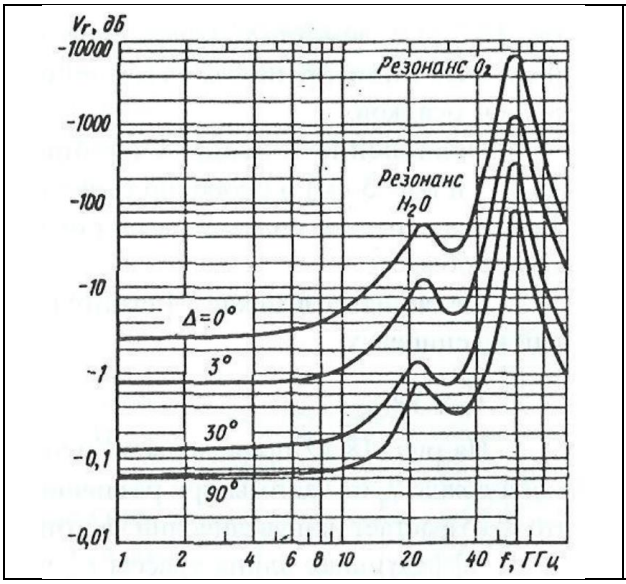


Рис. 8.7.10. Графіки залежностей множника ослаблення сигналу в тропосфері від частоти при різних кутах між променем і горизонтом (розрахунок). Промінь повністю перетинає тропосферу

Ослаблення радіохвиль в опадах. Ослаблення радіохвиль можливе й тоді, коли вони проходять через область, зайняту водяними краплями (туман, хмари, дощ), снігом або градом. Причини такого ослаблення – нерезонансне поглинання в макрочастинках, пов'язане з протіканням у них струму під дією електричного поля хвилі та розсіювання електромагнітної хвилі на макрочастинках. Ці ефекти суттєві на частотах вище 10 ГГц ($\lambda < 3$ см).

При поширенні радіохвилі через область дощу множник ослаблення можна записати аналогічно до (8.7.28):

$$V[\text{дБ}] = \gamma_{rain}[\text{дБ/км}]r_{eq}[\text{км}]. \qquad (8.7.29)$$

Коефіцієнт ослаблення γ_{rain} залежить від частоти хвилі та інтенсивності дощу (рис. 8.7.11).

У свою чергу, еквівалентна довжина траси може бути подана у формі

$$r_{eq} = k_{rain}(r)r, \qquad (8.7.30)$$

де коефіцієнт k_{rain} залежить від інтенсивності дощу та відстані r (тобто залежність $V[\text{дБ}]$ від r є нелінійною), але, на відміну від γ_{rain}, не залежить від частоти хвилі (рис. 8.7.12).

Множник ослаблення в хмарах та тумані записується аналогічно до (8.7.29),

$$V[\text{дБ}] = \gamma_{cl}[\text{дБ/км}]r[\text{км}] \qquad (8.7.31)$$

а відповідний коефіцієнт ослаблення γ_{cl} пропорційний масі води в одиниці об'єму M_{cl}:

$$\gamma_{cl} = k_{cl}M_{cl}. \qquad (8.7.32)$$

Для туману $M_{cl} \approx 0.25$ г/м^3, для хмар $M_{cl} \approx (0.1-8)$ г/м^3.

Питомий коефіцієнт ослаблення k_{cl} залежить від температури (рис. 8.7.13).

Ослаблення радіохвиль в іоносферній плазмі. Цей ефект зумовлений активною провідністю плазми на високих частотах, яку, відповідно до формули (8.1.21), можна подати у формі

$$\sigma' = \frac{e^2 n\nu}{m\left(\omega^2 + \nu^2\right)}. \qquad (8.7.2533$$

Видно, що активна провідність визначається частотою ν зіткнень електронів з важкими частинками – іонами та нейтральними атомами (див. рис. 8.2.6).

Ослаблення радіохвиль в іоносфері може також бути зумовлене їхнім розсіюванням на випадкових неоднорідностях іоносферної плазми (див. вище п. 8.2).

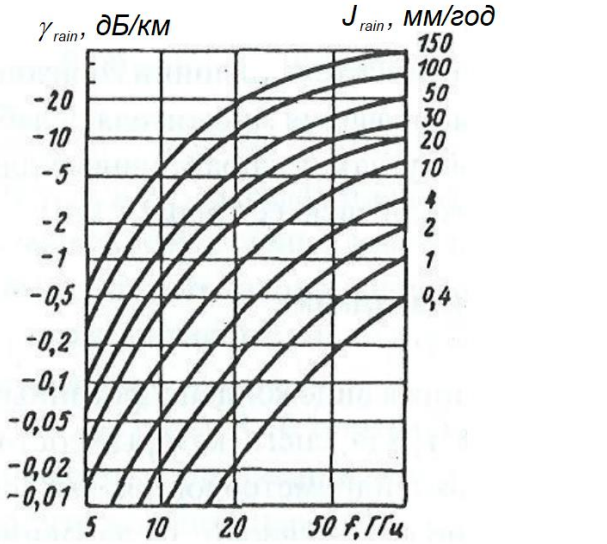


Рис. 8.7.11. Графіки залежності коефіцієнта ослаблення γ_{rain} під час дощу від частоти хвилі при різній інтенсивності дощу

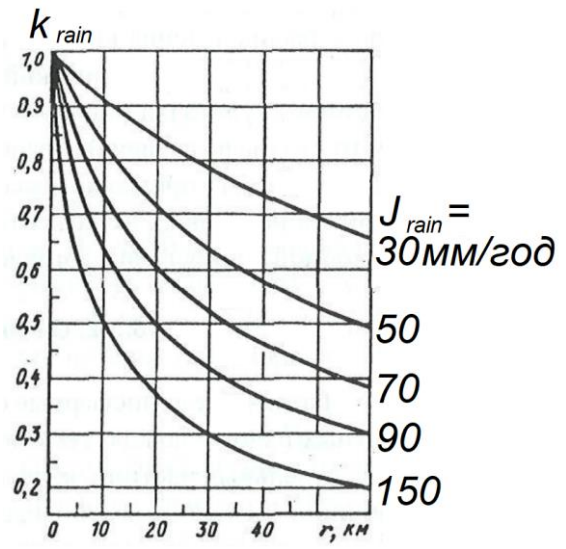


Рис. 8.7.12. Графіки залежності коефіцієнта k_{rain} під час дощу від частоти хвилі при різній інтенсивності дощу

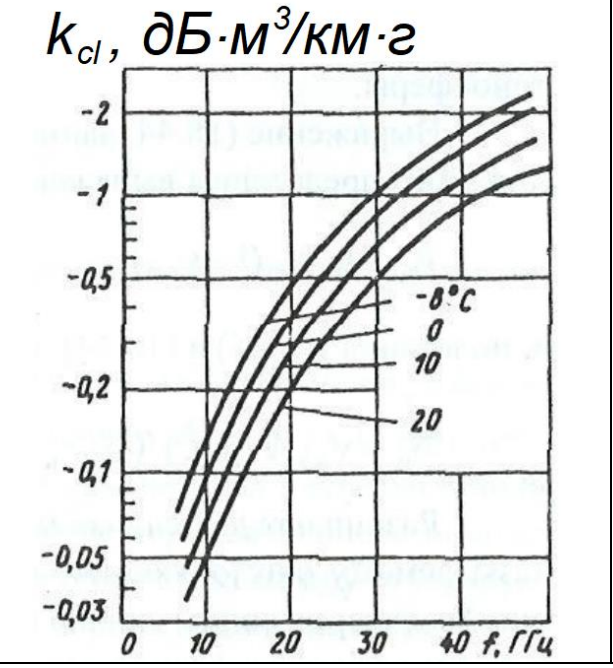


Рис. 8.7.13. Залежність питомого коефіцієнта ослаблення в хмарах та тумані від частоти хвилі при різних температурах

Контрольні питання до підрозділу 8.7

1. Чому в іоносфері промені мають криволінійну форму?
2. Чим визначається радіус кривини променю в іоносфері?
3. Охарактеризуйте різновиди рефракції радіохвиль в атмосфері Землі. Від чого зале-

жить їхня реалізація?

4. В чому полягає якісна відмінність між рефракцією хвиль в тропосфері та в іоносфері?
5. Які механізми поглинання радіохвиль в приземній атмосфері ви знаєте?

Задачі до підрозділу 8.7

8.7.1. Денний профіль концентрації плазми у внутрішній іоносфері апроксимується лінійною залежністю $n(h) = n_0 h/L$, $n_0 = 10^6$ см$^{-3}$, $L = 350$ км. Антена з голкоподібною діаграмою спрямованості випромінює під кутом 45º до горизонту. На якій висоті знаходиться точка повороту?

8.7.2. Для умов задачі 8.7.1 розрахувати радіус кривини променю електромагнітних хвиль в околі точки повороту.

8.7.3. Віддаль між передавачем і приймачем – 50 км. Знайти множник ослаблення сигналу за рахунок поглинання. Вважати, що основну роль у поглинанні сухої атмосфери відіграє кисень, а ефективні довжини траєкторії хвиль збігаються з геометричною довжиною траєкторії. Для розрахунку скористатися рис. 8.7.9.

8.7.4. Промінь електромагнітних хвиль з частотою 20 ГГц поширюється в тумані на віддаль 5 км. Температура повітря – +10ºС. Знайти множник ослаблення сигналу.

8.7.5. Випромінювач розташований на борту космічного апарата, що знаходиться на висоті 300 км над поверхнею Землі. Випромінювач має голкоподібну діаграму спрямованості й збуджує хвилю з частотою 20 МГц, яка поширюється паралельно до земної поверхні. На якій віддалі амплітуда хвилі зменшиться вдвічі, якщо вважати, що її згасання зумовлене зіткненнями електронів у плазмі? Розбіжністю променю знехтувати. Для розв’язання задачі скористатися рис. 8.2.6.

Частина IV. Елементи статистичної радіотехніки

При розгляді пристроїв сигналів у радіотехніці (див. вище частина I) ми вважали сигнали регулярними й відтворюваними. Ті самі припущення робилися при розгляді сигналів у пристроях аналогової радіоелектроніки (див. вище розділи 4-5). Насправді в силу різних причин радіотехнічні сигнали завжди містять деяку випадкову (флуктуаційну) складову. Вона може бути зумовлена зовнішніми завадами або власними шумами схеми.

Завади – це сторонні сигнали, що потрапляють до антени приймача. Вони можуть мати природне або штучне походження.

До завад природного походження належать, зокрема:

- космічне випромінювання;
- теплове випромінювання Землі;
- випромінювання блискавкових розрядів;
- випромінювання, спричинене магнітними бурями;
- флуктуації радіосигналу, зумовлені розсіюванням на випадкових неоднорідностях атмосфери (газу чи плазми);
- завмирання (флуктуації радіосигналу, зумовлені інтерференцією земної та іоносферної (тропосферної) хвиль);

Приклади завад штучного походження:

- сигнали від сторонніх радіостанцій;
- електромагнітне випромінювання промислового походження;
- електромагнітне випромінювання, спричинене транспортними засобами (в першу чергу, електротранспортом);
- електромагнітне випромінювання від побутової апаратури (комп'ютери, електроприлади тощо).

Шуми (внутрішні шуми) – це випадкові напруги або струми, які завжди виникають у радіотехнічних схемах. Їхньою першопричиною звичайно є тепловий рух заряджених частинок та дискретність електричних зарядів.

Наприклад, якщо вхід підсилювача замкнений накоротко, на його виході однаково спостерігається деяка змінна напруга, яка змінюється в часі випадковим чином і не відтворюється від експерименту до експерименту.

Розділ 9. Методи опису та аналізу випадкових процесів

У цьому розділі ми розглянемо питання, що стосуються методів опису та аналізу випадкових величин та випадкових процесів.

9.1. Випадкові процеси та їхні характеристики

Ми почнемо розгляд з методів математичного опису випадкових процесів, які включають елементи теорії ймовірностей. До характеристик випадкових процесів належать густина розподілу ймовірності, автокореляційна функція, спектральна інтенсивність та інші.

9.1.1. Випадкові величини, випадкові процеси та випадкові поля

Випадкову величину можна розглядати як результат вимірювання, який не можна передбачити і який не відтворюється від експерименту до експерименту. Наприклад, якщо ми маємо ящик яблук і по черзі зважуємо їх із великою точністю, то вага кожного наступного яблука може розглядатись як випадкова величина.

Випадковий процес – це така зміна деякої функції з часом, що значення цієї функції в кожний момент часу може розглядатися як випадкова величина (рис. 9.1.1 – 9.1.2). Прикладом може слугувати сила струму в мікрофоні біля вікна під час дощу або точне значення тиску газу, зумовлене ударами його атомів/молекул об стінку посудини.

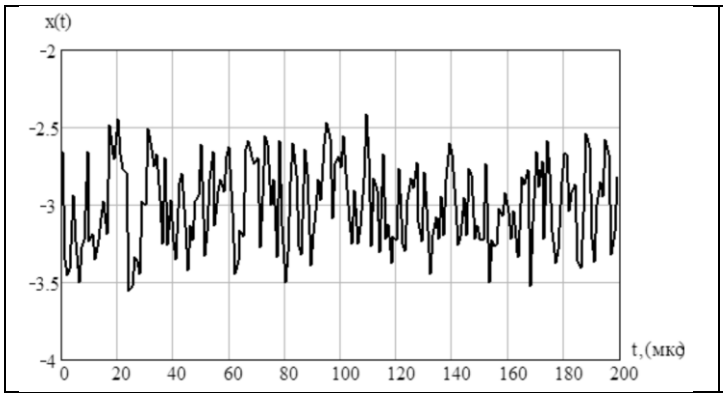


Рис. 9.1.1. Реалізація випадкового процесу

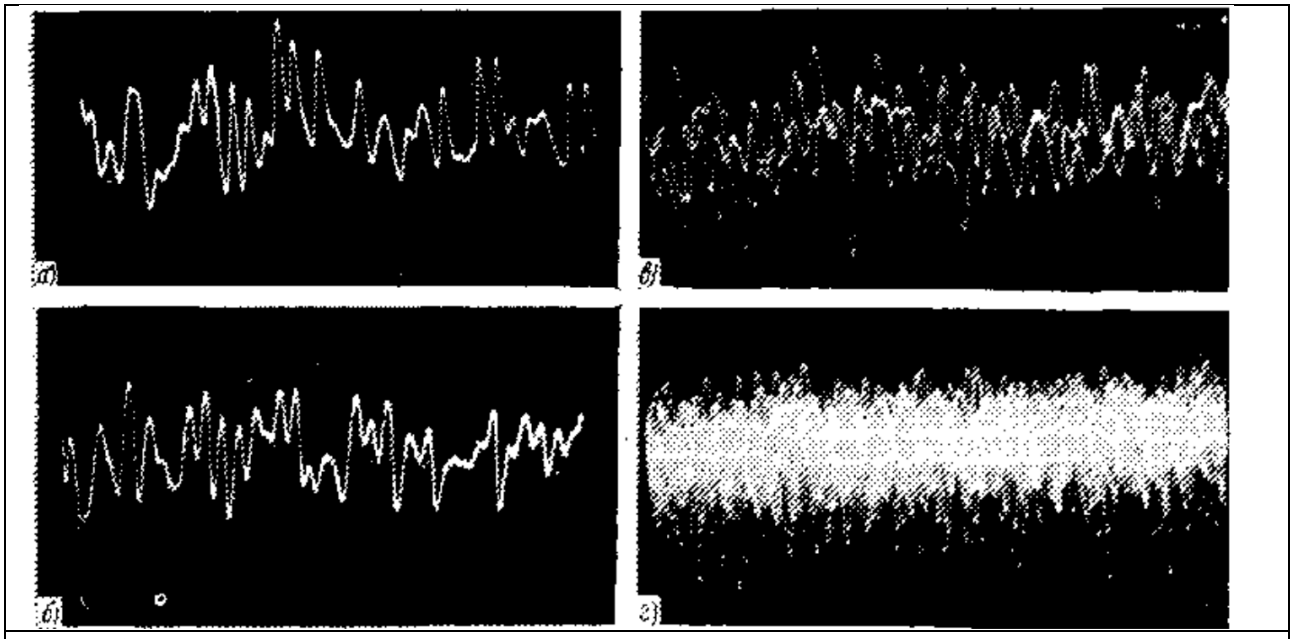


Рис. 9.1.2. Осцилограми напруги на кінцях резистора, зумовленої тепловим рухом електронів: а, б – окремі реалізації випадкового процесу; в – накладання трьох реалізацій; г – накладання ста реалізацій

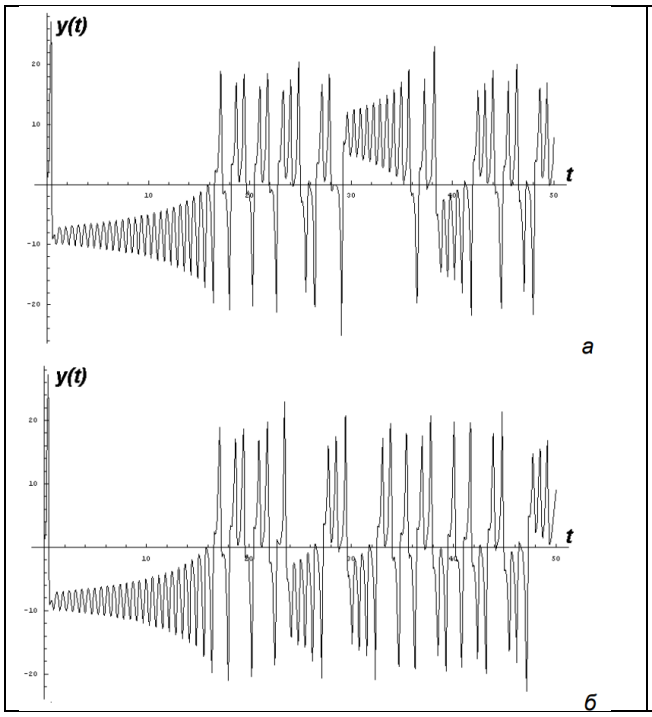


Рис.9.1.3. Розв'язки, що описують поведінку нестійкої системи (так званої системи Лоренца) для невеликої відмінності в початкових умовах:
а – y(0)=1.00; б – y(0)=1.01

Вимірювана величина може випадково змінюватися не тільки з часом, але одночасно і в просторі. В цьому випадку говорять про випадкові поля. Прикладом випадкового поля може служити випадкова зміна швидкості турбулентної течії рідини в просторі й у часі.

Детерміновані величини, процеси та поля відтворюються від одного експерименту до іншого, випадкові – ні. З математичної точки зору детермінована функція повністю задається диференціальним рівнянням (з невипадковими коефіцієнтами й зовнішньою силою) і набором початкових умов.

Відсутність відтворюваності найчастіше пов'язана з неможливістю точного розв'язання набору диференціальних рівнянь, що описують аналізовану систему (наприклад, для газу, що складається з дуже великої кількості молекул; див. також рис. 9.1.2) або нестійкістю її руху, тобто збільшенням відмінності між двома розв'язками того само диференціального рівняння, які трохи відрізняються початковими умовами (рис. 9.1.3).

В цій частині курсу ми будемо вести мову переважно про випадкові процеси. Їх називають також стохастичними процесами.

9.1.2. Густина розподілу ймовірності

Нехай ми виконали N ($N >> 1$) вимірювань деякої дійсної випадкової величини x і отримали набір значень $x_1, x_2, x_3, ..., x_N$, що лежать в інтервалі $\left[x_{\min}, x_{\max}\right]$. Оберемо всередині цього інтервалу деякий менший інтервал $\left[x_0, x_0 + \Delta x\right]$. Нехай до обраного інтервалу потрапило N_1 ($N_1 < N$) результатів вимірювань. Тоді ймовірність того, що випадкова величина x потрапляє до інтервалу $\left[x_0, x_0 + \Delta x\right]$, можна оцінити як

$$P\left(x_0 \le x \le x_0 + \Delta x\right) \approx \frac{N_1}{N}.$$

Статистичний опис можливий, якщо відношення N_1/N прямує до деякої межі при $N \to \infty$. В цьому випадку замість оцінки отримуємо точне значення ймовірності:

$$P\left(x_0 \le x \le x_0 + \Delta x\right) = \lim_{N\to\infty} \frac{N_1}{N} \qquad (9.1.1)$$

Тепер спрямуємо Δx до нуля. Тоді функція

$$w\left(x_0\right) = \lim_{\Delta x\to 0} \frac{P\left(x_0 \le x \le x_0 + \Delta x\right)}{\Delta x} = \lim_{\Delta x\to 0} \lim_{N\to\infty} \frac{1}{\Delta x} \frac{N_1}{N} \qquad (9.1.2)$$

називається густиною розподілу ймовірності випадкової величини x (рис. 9.1.4). Нижче ми інколи називатимемо густину розподілу ймовірності просто розподілом.

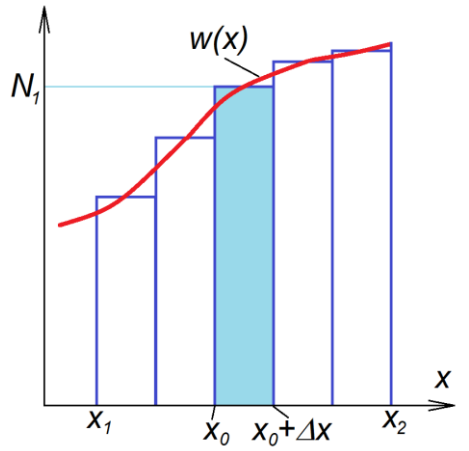


Рис. 9.1.4. До визначення густини розподілу ймовірності

Ймовірність $P\left(x_1 \le x \le x_2\right)$ того, що випадкова величина потрапить до заданого

інтервалу, й густина розподілу ймовірності $w(x)$ пов'язані співвідношенням

$$P(x_1 \le x \le x_2) = \int_{x_1}^{x_2} w(x)\,dx \tag{9.1.3}$$

(див. рис. 9.1.4).

Якщо всі значення випадкової величини x потрапляють до інтервалу $[x_{\min}, x_{\max}]$, то можна записати, що

$$\int_{x_{\min}}^{x_{\max}} w(x)\,dx = 1 \tag{9.1.4}$$

(умова нормування густини розподілу ймовірності). Якщо ж $-\infty < x < \infty$, то умова нормування набуває вигляду

$$\int_{-\infty}^{\infty} w(x)\,dx = 1. \tag{9.1.5}$$

Нехай випадкові величини x, y пов'язані співвідношеннями

$$y = f(x),\ x = f^{-1}(y), \tag{9.1.6}$$

де $f^{-1}(x)$ – функція, обернена щодо $f(x)$, тобто $f^{-1}(f(x)) = x$. Нехай розподіл $w(x)$ відомий. Щоб знайти розподіл $w(y)$, запишемо ймовірність того, що величина x потрапляє до інтервалу $[x_1, x_2]$, скориставшись формулою (9.1.3):

$$P(x_1 \le x \le x_2) = \int_{x_1}^{x_2} w(x)\,dx.$$

Перейдемо в правій частині до змінної $y = f(x)$, узявши до уваги, що $y_{1,2} = f(x_{1,2})$ (вважаємо функцію $f(x)$ на інтервалі $[x_1, x_2]$ монотонною). Тоді, оскільки $x = f^{-1}(y)$, можна записати:

$$\begin{aligned} P(x_1 \le x \le x_2) &= P(y_1 \le y \le y_2) = \int_{y_1}^{y_2} w\left(f^{-1}(y)\right) d\left(f^{-1}(y)\right) = \\ &= \int_{y_1}^{y_2} w\left(f^{-1}(y)\right) \frac{d\left(f^{-1}(y)\right)}{dy}\,dy = \int_{y_1}^{y_2} w(y)\,dy. \end{aligned} \tag{9.1.7}$$

Оскільки остання рівність справедлива для довільних значень $y_{1,2}$, то можна записати, що

$$w(y) = w\left(f^{-1}(y)\right) \frac{d\left(f^{-1}(y)\right)}{dy}. \tag{9.1.8}$$

Очевидно, всі наведені міркування справедливі й для випадкового процесу, значення якого беруться у фіксований момент часу t. Взагалі кажучи, для різних моментів часу густини розподілу ймовірності можуть відрізнятися, тому для випадкових процесів $w = w(x,t)$.

Випадковий процес називають стаціонарним, якщо його ймовірнісні закономірності не змінюються з часом. Зокрема, для стаціонарних процесів $w(x,t) = w(x)$. Процеси, які не є стаціонарними, називаються нестаціонарними. Далі ми дамо акуратніше визначення стаціонарного випадкового процесу.

Комплексний випадковий процес $z(t) = x(t) + iy(t)$ визначається двома дійсними випадковими функціями $x(t)$ та $y(t)$.

9.1.3. Статистичні середні

Знаючи $w(x,t)$, можна розрахувати середні по ансамблю реалізацій (статистичні середні) різних величин. Наприклад,

$$\langle x(t)\rangle = \lim_{N\to\infty}\frac{\sum_{i=1}^{N} x_i(t)}{N} = \lim_{N\to\infty}\frac{\sum_{k=1}^{K} x_k(t)N_k}{N} = \sum_{k=1}^{K} x_k(t)\Delta x \lim_{N\to\infty}\frac{N_k}{\Delta x N} = \int_{-\infty}^{\infty} x w(x,t)\,dx\,. \qquad (9.1.9)$$

Величина $\langle x(t)\rangle$ – це регулярна (передбачувана) складова процесу $x(t)$. Відповідно

$$\tilde{x}(t) \equiv x(t) - \langle x(t)\rangle \qquad (9.1.10)$$

– це флуктуаційна складова деякої реалізації. Очевидно,

$$\langle \tilde{x}(t)\rangle = 0\,. \qquad (9.1.11)$$

На рис. 9.1.5 показано п’ять реалізацій випадкового процесу і середнє по ансамблю реалізацій. Окремі реалізацій флуктуаційної складової показані на рис. 9.1.6.

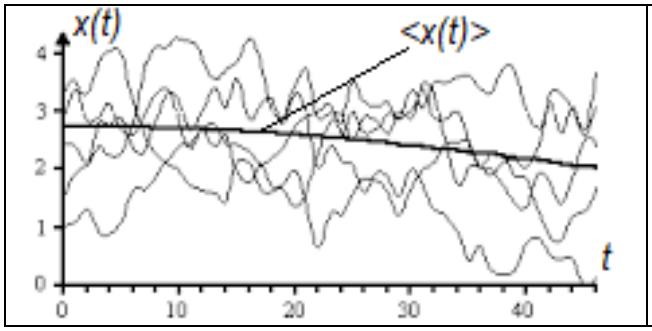


Рис. 9.1.5. Регулярна складові випадкового процесу

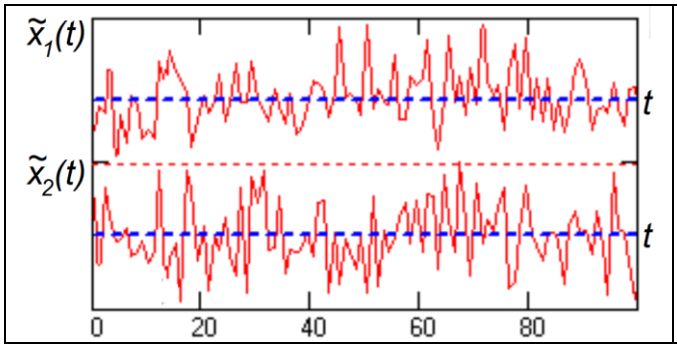


Рис. 9.1.6. Дві реалізації флуктуаційної складової випадкового процесу

Аналогічно до (9.1.9) можна записати статистичне середнє будь-якої функції випадкової змінної $F(x(t))$:

$$\langle F(x(t))\rangle = \int_{-\infty}^{\infty} F(x)w(x,t)\,dx\,. \qquad (9.1.12)$$

Величина

$$m_n(t) \equiv \langle x^n(t)\rangle = \int_{-\infty}^{\infty} x^n w(x,t)\,dx \qquad (9.1.13)$$

називається моментом *n*-го порядку розподілу $w(x,t)$, а величина

$$\mu_n(t) \equiv \langle \tilde{x}^n(t)\rangle = \int_{-\infty}^{\infty} \tilde{x}^n w(x,t)\,dx \qquad (9.1.14)$$

– центральним моментом.

Центральний момент другого порядку, який називають дисперсією, визначає інтенсивність флуктуацій:

$$\sigma^2(t)=\mu_2(t)\equiv\langle\tilde{x}^2(t)\rangle=\left\langle\left[x(t)-\langle x(t)\rangle\right]^2\right\rangle=\left\langle x^2(t)-2x(t)\langle x(t)\rangle+\langle x(t)\rangle^2\right\rangle=$$
$$=\langle x^2(t)\rangle-2\langle x(t)\rangle^2+\langle x(t)\rangle^2=\langle x^2(t)\rangle-\langle x(t)\rangle^2. \qquad (9.1.15)$$

Можна показати, що повний набір моментів містить ту саму інформацію про випадковий процес, що й густина розподілу ймовірностей.

Величина

$$\theta(u)=\langle\exp(iux)\rangle=\int_{-\infty}^{\infty}\exp(iux)w(x,t)dx \qquad (9.1.16)$$

називається характеристичною функцією. Її можна інтерпретувати як Фур'є-спектр густини розподілу $w(x,t)$.

Контрольні питання до підрозділу 9.1

1. Яка природа випадкових процесів у радіоелектроніці?
2. Що таке ансамбль реалізацій випадкового процесу?
3. Що таке густина розподілу ймовірностей?
4. Як за ансамблем реалізацій можна визначити густину розподілу ймовірностей як функцію часу?
5. Як за ансамблем реалізацій можна визначити середню за ансамблем складову випадкового процесу?
6. Як можна кількісно охарактеризувати інтенсивність флуктуацій?

Задачі до підрозділу 9.1

9.1.1. Випадкова величина характеризується розподілом $w(x)=\begin{cases}1/2a, & |x|\le a;\\ 0, & |x|>a;\end{cases}$. Знайти розподіл випадкової величини $y=x^2$ та намалювати його графік.

9.1.2. Випадкова величина характеризується

а) рівномірним розподілом $w(x)=\begin{cases}C, & |x-x_0|\le a;\\ 0, & |x-x_0|>a;\end{cases}$;

б) гауссовим розподілом $w(x)=C\exp\left[-(x-x_0)^2/a^2\right]$;

в) лоренцовим розподілом $w(x)=C\Big/\left[(x-x_0)^2+a^2\right]$.

Знайти сталі нормування *C*.

9.1.3. Для розподілів, наведених у задачі 9.1.1, знайти середнє значення $\langle x\rangle$ та дисперсію σ^2 випадкової величини x.

9.1.4. Для розподілів, наведених у задачі 9.1.1, знайти характеристичні функції.

9.2. Багатовимірні статистичні характеристики

Виявляється, що густина розподілу ймовірності $w(x,t)$ випадкового процесу $x(t)$ не визначає повністю його властивості. Додаткову інформацію про цей процес дають багатовимірні статистичні характеристики, зокрема, автокореляційна функція.

9.2.1. Багатовимірні розподіли ймовірності

Нехай набір випадкових величин x, y, z характеризується розподілами $w(x), w(y), w(z)$. Щоб з'ясувати, чи пов'язані ці випадкові величини між собою, необхідно знати також двовимірні розподіли $w(x,y), w(x,z), w(y,z)$ і тривимірний розподіл $w(x,y,z)$.

Двовимірний розподіл $w(x,y)$ дозволяє знайти ймовірність $w(x_0,y_0)dxdy$ того, що величини x та y одночасно потрапляють відповідно до інтервалів $[x_0, x_0+dx]$ та $[y_0, y_0+dy]$. Аналогічно $w(x_0,y_0,z_0)dxdydz$ – це ймовірність того, що величини x, y, z одночасно потрапляють відповідно до інтервалів $[x_0, x_0+dx]$, $[y_0, y_0+dy]$ та $[z_0, z_0+dz]$.

Очевидно, багатовимірні розподіли мають задовольняти умовам нормування типу

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, w(x,y) = 1, \qquad \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz\, w(x,y,z) = 1 \tag{9.2.1}$$

(пор. з формулою (9.1.5)) та умовам відповідності типу

$$\int_{-\infty}^{\infty} dy\, w(x,y) = w(x), \qquad \int_{-\infty}^{\infty} dz\, w(x,y,z) = w(x,y). \tag{9.2.2}$$

Якщо випадкові величини $x_{1,2}$ та $y_{1,2}$ пов'язані між собою співвідношеннями

$$y_1 = f_1(x_1,x_2),\ y_2 = f_2(x_1,x_2), \tag{9.2.3}$$

то, за аналогією з (9.1.8) можна записати зв'язок між двовимірними розподілами $w(x_1,x_2)$ і $w(y_1,y_2)$ у формі

$$w(y_1,y_2) = w\left(x_1 = f_1^{-1}(y_1,y_2), x_2 = f_2^{-1}(y_1,y_2)\right)\left|\frac{\partial\left(f_1^{-1}, f_2^{-1}\right)}{\partial(y_1,y_2)}\right|, \tag{9.2.4}$$

де

$$\left|\frac{\partial\left(f_1^{-1}, f_2^{-1}\right)}{\partial(y_1,y_2)}\right| = \begin{vmatrix} \partial f_1^{-1}/\partial y_1 & \partial f_1^{-1}/\partial y_2 \\ \partial f_2^{-1}/\partial y_1 & \partial f_2^{-1}/\partial y_2 \end{vmatrix} \tag{9.2.5}$$

– якобіан перетворення.

Якщо замість випадкових величин розглядати випадкові процеси $x(t), y(t), z(t)$, то дво- й тривимірні розподіли будуть залежати від часу як від параметра: $w(x,y;t)$, $w(x,z;t)$, $w(y,z;t)$ та $w(x,y,z;t)$.

Якщо взяти значення випадкових процесів у відмінні моменти часу, наприклад, $x(t_1)$, $y(t_2)$, $z(t_3)$, то відповідні розподіли також залежатимуть від цих моментів часу, наприклад, $w(x,y;t_1,t_2)$, $w(x,y,z;t_1,t_2,t_3)$.

Дуже часто необхідно знати зв'язок між значеннями випадкового процесу в різні моменти часу: $x_1 = x(t_1), x_2 = x(t_2), x_3 = x(t_3), \ldots$. Для цього потрібні багатовимірні розподіли вигляду $w(x_1,x_2;t_1,t_2)$, $w(x_1,x_2,x_3;t_1,t_2,t_3), \ldots$ Вони повинні задовольняти умовам відповідності типу

$$\lim_{t_2 \to t_1} w(x_1,x_2;t_1,t_2) = w(x_1,t_1)\delta(x_1 - x_2). \tag{9.2.6}$$

Справді, виконання умови (9.2.6) автоматично задовольняє співвідношення відповідності типу першої з умов (9.2.2).

Користуючись такими багатовимірними розподілами, можна повністю задати випадковий процес $x(t)$. Для цього для довільного набору моментів часу $t_1, t_2, t_3, ..., t_n$ ($n > 1$ – ціле число) треба задати багатовимірний розподіл $w(x_1, x_2, x_3, ..., x_n; t_1, t_2, t_3, ,...t_n)$. Тоді величина $w(x_1, x_2, x_3, ..., x_n; t_1, t_2, t_3, ,...t_n)dx_1dx_2dx_3...dx_n$ визначає ймовірність того, що в момент часу t_1 величина $x(t)$ потрапить до інтервалу $[x_1, x_1 + dx_1]$, в момент часу t_2 – до інтервалу $[x_2, x_2 + dx_2]$, тощо (рис. 9.2.1).

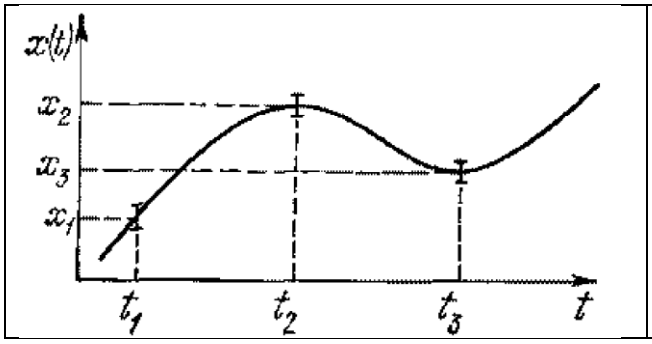


Рис. 9.2.1. До задання випадкового процесу за допомогою багатовимірного розподілу

Можна сказати, що задання багатовимірного розподілу $w(x_1, x_2, x_3, ..., x_n; t_1, t_2, t_3, ,...t_n)$ є статистичним аналогом визначення детермінованої функції часу через набір її значень для довільних моментів часу $t_1, t_2, t_3, ,...t_n$.

Ми вже говорили, що стаціонарний випадковий процес характеризується тим, що його статистичні характеристики не змінюються з часом. Акуратніше визначення можна дати за допомогою багатовимірного розподілу $w(x_1, x_2, x_3, ..., x_n; t_1, t_2, t_3, ,...t_n)$. Для стаціонарних випадкових процесів справджується рівність

$$w(x_1, x_2, x_3, ..., x_n; t_1, t_2, t_3, ,...t_n) = w(x_1, x_2, x_3, ..., x_n; t_1 + T, t_2 + T, t_3 + T, ,...t_n + T), \qquad (9.2.7)$$

де $T \neq 0$ – довільний проміжок часу. Зокрема, як уже відзначалося, одновимірний розподіл стаціонарного процесу взагалі не залежить від часу: $w(x;t) = w(x)$.

Прикладами стаціонарних випадкових процесів можуть бути флуктуації в системах, властивості яких не змінюються з часом – наприклад, флуктуації напруги на резисторі, вміщеному в термостат.

9.2.2. Умовні ймовірності

При проходженні медичного огляду зазвичай вимірюють зріст людини та її вагу. Розглядатимемо їх як випадкові величини x та y, що змінюються від людини до людини. Виберемо інтервали значень $[x_0, x_0 + \Delta x]$ та $[y_0, y_0 + \Delta y]$. Нехай всього виконано N вимірювань, причому значення x потрапило до інтервалу $[x_0, x_0 + \Delta x]$ всього N_x разів, значення y до інтервалу $[y_0, y_0 + \Delta y]$ – N_y разів, а значення x та y одночасно потрапили до відповідних інтервалів N_{xy} разів (зрозуміло, що $N > N_{x,y} > N_{xy}$; див. рис. 9.2.2).

Тоді, спрямовуючи N до нескінченності, можна записати ймовірність того, що величина x потрапить до інтервалу $[x_0, x_0 + \Delta x]$ як

$$P(x_0) \equiv P(x \in [x_0, x_0 + \Delta x]) = \frac{N_x}{N}, \qquad (9.2.8)$$

і аналогічно для y

$$P(y_0) \equiv P\left(y \in [y_0, y_0 + \Delta y]\right) = \frac{N_y}{N}. \qquad (9.2.9)$$

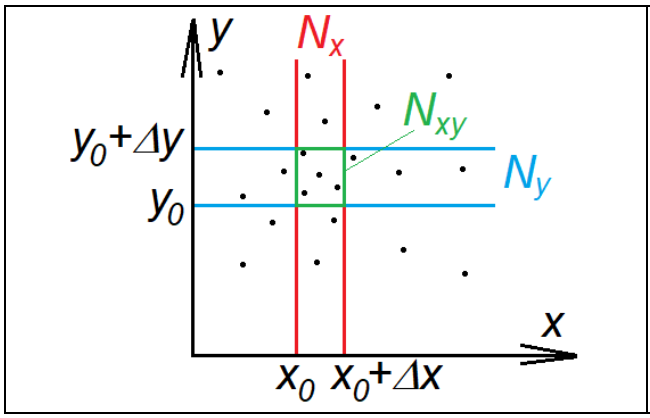


Рис. 9.2.2. До визначення умовної ймовірності

Величина

$$P(x, y) \equiv P\left(x \in [x_0, x_0 + \Delta x] \cap y \in [y_0, y_0 + \Delta y]\right) = \frac{N_{xy}}{N} \qquad (9.2.10)$$

визначає ймовірність того, що величини x та y одночасно потраплять відповідно до інтервалу $[x_0, x_0 + \Delta x]$ та $[y_0, y_0 + \Delta y]$ (тут $\cap$ – символ кон'юнкції; умови, об'єднані цим символом, мають виконуватись одночасно).

Величини

$$P(x \mid y) = \frac{N_{xy}}{N_y} \qquad (9.2.11)$$

та

$$P(y \mid x) = \frac{N_{xy}}{N_x} \qquad (9.2.12)$$

називають умовними ймовірностями. $P(x \mid y)$ – це ймовірність того, що величина x потрапить до інтервалу $[x_0, x_0 + \Delta x]$ за умови, що величина y уже потрапила до інтервалу $[y_0, y_0 + \Delta y]$. Аналогічний зміст має умовна ймовірність $P(y \mid x)$.

Оскільки справедливі співвідношення

$$\frac{N_{xy}}{N} = \frac{N_{xy}}{N_y}\frac{N_y}{N} = \frac{N_{xy}}{N_x}\frac{N_x}{N}, \qquad (9.2.13)$$

то з урахуванням (9.2.8)-(9.2.12) можна записати:

$$P(x, y) = P(x \mid y)P(y) = P(y \mid x)P(x). \qquad (9.2.14)$$

Спрямовуючи ширину інтервалів Δx та Δy до нуля, отримаємо аналогічні співвідношення для густин розподілу ймовірності:

$$w(x, y) = w(x \mid y)w(y) = w(y \mid x)w(x). \qquad (9.2.15)$$

В умовному розподілі другий аргумент відіграє роль параметра, тому умови нормування мають вигляд

$$\int_{-\infty}^{\infty} w(x \mid y)dx = 1, \qquad \int_{-\infty}^{\infty} w(y \mid x)dy = 1. \qquad (9.2.16)$$

Випадкові величини x та y називають статистично незалежними за виконання умов

$$w(x \mid y) = w(x), \qquad w(y \mid x) = w(y). \qquad (9.2.17)$$

У цьому випадку випадкові величини x та y не впливають одна на одну.

Для статистично незалежних випадкових величин з (9.2.15) випливає, що

$$w(x,y)=w(x)w(y). \tag{9.2.18}$$

В протилежному випадку говорять, що між ними існує статистичний зв'язок. Наприклад, медицина встановила існування статистичного зв'язку між групою крові та схильністю до деяких хвороб (наприклад, до виразки шлунку).

9.2.3. Кореляційні та автокореляційні функції

Кількісною мірою статистичного зв'язку двох випадкових процесів виступає взаємна кореляційна функція.

Розглянемо два випадкових процеси $x(t)$ та $y(t)$. Статистичний зв'язок між цими величинами визначає (взаємна) кореляційна функція

$$\begin{aligned} B_{xy} &= \langle \tilde{x}(t)\tilde{y}(t)\rangle = \left\langle \left[x(t)-\langle x(t)\rangle\right]\left[y(t)-\langle y(t)\rangle\right]\right\rangle = \\ &= \left\langle x(t)y(t)-x(t)\langle y(t)\rangle - y(t)\langle x(t)\rangle + \langle x(t)\rangle\langle y(t)\rangle\right\rangle = \\ &= \langle x(t)y(t)\rangle - \langle x(t)\rangle\langle y(t)\rangle. \end{aligned} \tag{9.2.19}$$

Відповідно можна записати, що

$$\langle x(t)y(t)\rangle = B_{xy} + \langle x(t)\rangle\langle y(t)\rangle. \tag{9.2.20}$$

Якщо випадкові процеси $x(t)$ та $y(t)$ незалежні, то $\langle x(t)y(t)\rangle = \langle x(t)\rangle\langle y(t)\rangle$ і $B_{xy}=0$.

Коефіцієнт кореляції:

$$R_{xy} \equiv \frac{B_{xy}}{\sigma_x\sigma_y} = \frac{\langle x(t)y(t)\rangle - \langle x(t)\rangle\langle y(t)\rangle}{\sqrt{\left[\langle x^2(t)\rangle - \langle x(t)\rangle^2\right]\left[\langle y^2(t)\rangle - \langle y(t)\rangle^2\right]}}, \quad -1 < R_{xy} < 1. \tag{9.2.21}$$

Часто розглядають випадок, коли $y(t)=x(t+\tau)\equiv x_\tau$. Тоді величина

$$B(t,\tau) = \langle x x_\tau\rangle - \langle x\rangle\langle x_\tau\rangle \tag{9.2.22}$$

називається автокореляційною функцією випадкового процесу $x(t)$, а величина

$$R(t,\tau) = \frac{B(t,\tau)}{\sigma\sigma_\tau} \tag{9.2.23}$$

– коефіцієнтом автокореляції. Видно, що $B(t,0)=\langle x^2\rangle - \langle x\rangle^2 = \sigma^2$. Для стаціонарних випадкових процесів $B(t,\tau)=B(\tau)$.

Автокореляційна функція, як випливає з її визначення, встановлює статистичний зв'язок між значеннями випадкового процесу в різні моменти часу. Звичайно $B(t,\tau\to\infty)=0$ (рис. 15.4).

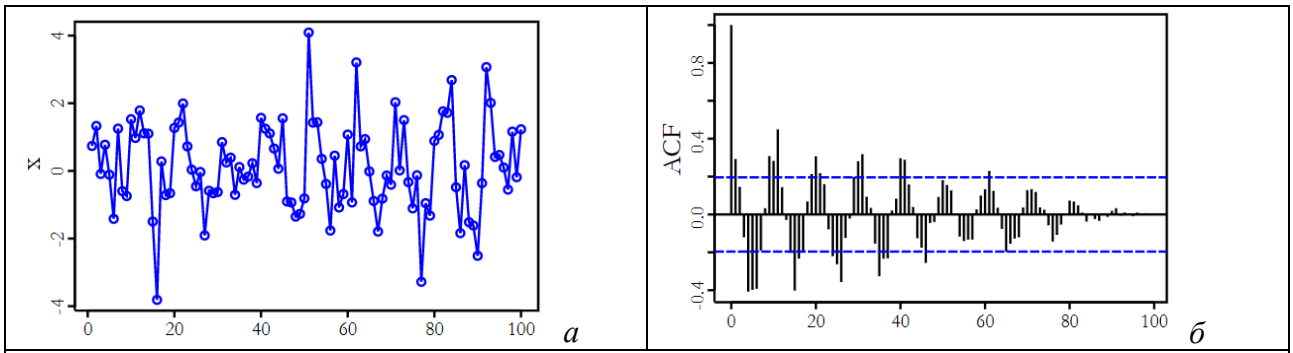


Рис. 9.2.3. Приклад випадкового процесу (а) та його автокореляційної функції (б)

Час τ_c, за який автокореляційна функція за модулем помітно зменшується (обертається на нуль) – це час автокореляції, тобто тривалість «пам'яті» системи. Строге означення часу автокореляції обирається залежно від конкретної задачі. З часом (точніше, при $t \gg \tau_c$ випадковий процес «забуває» своє минуле.

Стаціонарний випадковий процес, для якого $\tau_c = 0$, називають δ-корельованим. Для нього

$B(\tau) = C\delta(\tau)$.

Крім парних автокореляційних функцій, інколи розглядають автокореляційні функції вищих порядків: $\langle \tilde{x}(t)\tilde{x}(t+\tau_1)\tilde{x}(t+\tau_2)\rangle$, $\langle \tilde{x}(t)\tilde{x}(t+\tau_1)\tilde{x}(t+\tau_2)\tilde{x}(t+\tau_3)\rangle$ тощо.

9.2.4. Теорема Вінера - Хінчина

Ми бачили, що детерміновані сигнали можна подавати як функцію часу або як функцію частоти (тобто в спектральній формі). Такий підхід застосовується й для випадкових процесів, зокрема, в задачах про фільтрацію корисних сигналів за наявності шумів або завад.

Позначимо надалі через $\xi(t)$ флуктуаційну складову стаціонарного випадкового процесу $x(t)$:

$$\xi(t) \equiv x(t) - \langle x(t)\rangle. \tag{9.2.24}$$

Її можна подати у формі інтегралу Фур'є:

$$\xi(t) = \int_{-\infty}^{\infty} \xi(\omega)\exp(i\omega t)\,d\omega, \tag{9.2.25}$$

де

$$\xi(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \xi(t)\exp(-i\omega t)\,dt \tag{9.2.26}$$

– спектр випадкового процесу $\xi(t)$.

Очевидно, що спектр $\xi(\omega)$ – це випадкова функція частоти, причому $\langle \xi(\omega)\rangle = 0$.

Введемо спектральну густину $G(\omega)$ випадкового процесу (спектр потужності) так, що

$$\langle \xi^2(t)\rangle \equiv \sigma^2 = \int_{-\infty}^{\infty} G(\omega)\,d\omega. \tag{9.2.27}$$

Ця формула не є строгим визначенням $G(\omega)$, таке визначення буде дане далі.

Запишемо автокореляційну функцію через спектр випадкового процесу:

$$\begin{aligned} B(\tau) &= \langle \xi\xi_\tau\rangle = \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' \langle \xi_\omega \xi_{\omega'}\rangle \exp\left[i\omega t + i\omega'(t+\tau)\right] = \\ &= \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' \langle \xi_\omega \xi_{\omega'}\rangle \exp\left[i(\omega+\omega')t + i\omega'\tau\right] \end{aligned}. \tag{9.2.28}$$

Для стаціонарного процесу автокореляційна функція $B(\tau)$ не повинна залежати від часу t. Для цього під інтегралом повинна з'явитися δ-функія аргументу $\omega + \omega'$, тобто $\delta(\omega+\omega')$. Це можливо лише тоді, коли спектри ξ_ω є δ-корельованими, тобто задовольняється співвідношення

$$\langle \xi_\omega \xi_{\omega'}\rangle = G(\omega)\delta(\omega+\omega'). \tag{9.2.29}$$

Тоді

$$B(\tau)=\int_{-\infty}^{\infty}d\omega\int_{-\infty}^{\infty}d\omega' G(\omega)\delta(\omega+\omega')\exp\left[i\omega t+i\omega'(t+\tau)\right]=\int_{-\infty}^{\infty}d\omega G(\omega)\exp(-i\omega\tau). \qquad (9.2.30)$$

Відповідно

$$G(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}B(\tau)\exp(i\omega\tau)d\tau. \qquad (9.2.31)$$

Співвідношення (9.2.30)-(9.2.31) відомі в літературі як теорема Вінера – Хінчина. Вона вказує, що автокореляційна функція $B(\tau)$ та спектральна інтенсивність $G(\omega)$ пов'язані між собою перетворенням Фур'є.

Оскільки $\left\langle \xi^2(t)\right\rangle = B(\tau=0)$, то зі співвідношення (9.2.30) справді випливає записане раніше співвідношення (9.2.27).

В силу теореми Вінера – Хінчина, час автокореляції (проміжок часу, на якому $B(\tau)\neq 0$) і ширина $\Delta\omega$ спектру $G(\omega)$ (за частотою) обернено пропорційні одне одному. Зокрема, для δ-корельованого випадкового процесу, коли $B(\tau)=C\delta(\tau)$, спектральна густина рівномірно розподілена по всіх частотах (так званий білий шум).

Контрольні питання до підрозділу 9.2

1. Що розуміють під умовною ймовірністю?
2. Що таке взаємна кореляційна функція?
3. Що таке автокореляційна функція випадкового процесу?
4. Що таке час автокореляції?
5. В чому полягає зміст теореми Вінера – Хінчина?
6. Запропонуйте спосіб практичного визначення спектру потужності (спектральної інтенсивності) випадкового процесу.

Задачі до підрозділу 9.2

9.2.1. Двовимірний розподіл для випадкових величин x, y має вигляд $w(x,y)=C\exp\left[-\left(x^2+y^2\right)/a^2\right]$. Чи існує статистичний зв'язок між цими величинами? Знайдіть сталу нормування C та відповідні одновимірні розподіли $w(x)$ та $w(y)$.

9.2.2. Двовимірний розподіл для випадкових величин x, y має вигляд $w(x,y)=C\Big/\left(x^2+y^2+a^2\right)^2$. Чи існує статистичний зв'язок між цими величинами? Знайдіть сталу нормування C, відповідні одновимірні розподіли $w(x)$ та $w(y)$ і відповідні умовні розподіли $w(x\,|\,y)$ та $w(y\,|\,x)$.

9.2.3. Спектральна інтенсивність випадкового процесу $x(t)$ має вигляд:

а) $G(\omega)=\begin{cases} G_0, & |\omega|\le h; \\ 0, & |\omega|>h; \end{cases}$

б) $G(\omega)=\dfrac{G_0h^2}{h^2+\omega^2}$;

в) $G(\omega)=G_0\exp\left(-\omega^2/h^2\right)$;

г) $G(\omega)=\begin{cases} G_0\left(1-|\omega|/h\right), & |\omega|\le h; \\ 0, & |\omega|>h; \end{cases}$

д) $G(\omega)=\begin{cases} G_0, & \omega_0-h\le|\omega|\le\omega_0+h; \\ 0, & |\omega|<\omega_0-h\cup|\omega|>\omega_0+h; \end{cases}$

е) $G(\omega)=G_0\left\{\exp\left[-(\omega-\omega_0)^2/4h^2\right]+\exp\left[-(\omega+\omega_0)^2/4h^2\right]\right\}$.

Знайти дисперсію, автокореляційну функцію та коефіцієнт автокореляції випадкового процесу $x(t)$.

9.2.4. Автокореляційна функція випадкового процесу $x(t)$ має вигляд:

а) $B(\tau)=B_0\dfrac{\sin h\tau}{h\tau}$;

б) $B(\tau)=B_0\dfrac{\sin h\tau}{h\tau}\cos\omega_0\tau$;

в) $B(\tau)=B_0\exp(-h|\tau|)$

г) $B(\tau)=B_0\exp\left[-(h\tau)^2\right]$.

Знайти спектральну інтенсивність випадкового процесу $x(t)$.

9.3. Моделі випадкових процесів

У цьому підрозділі ми обговоримо моделі, які використовуються для опису випадкових процесів у радіотехніці.

Обговоримо, як можна класифікувати моделі випадкових процесів. Оскільки такі процеси характеризуються багатьма ознаками, відповідно до них існують різні варіанти класифікації.

Якщо результати усереднення випадкового процесу за ансамблем реалізацій (статистичне усереднення) та за достатнім проміжком часу (середнє за часом) збігаються, такий процес називають ергодичним. У протилежному випадку говорять про неергодичний процес.

За виглядом функції розподілу ймовірності найпоширенішими є нормальні (гауссівські) процеси, які відповідно описуються нормальним розподілом (розподілом Гаусса).

За виглядом спектральної інтенсивності найчастіше виділяють білий шум, або δ-корельований випадковий процес із нескінченно широким рівномірним спектром. В протилежному випадку, коли ширина спектру значно менша за його середню частоту, говорять про вузькосмуговий випадковий процес.

Названі моделі випадкових процесів трапляються найчастіше. Існують і більш специфічні моделі випадкових процесів.

Модель випадкової імпульсної послідовності (набір однакових імпульсів, моменти появи та повна кількість яких на заданому проміжку часу є випадковими величинами) використовується для опису так званого дробового шуму (див. нижче п. 10.1.2).

Модель дифузійного (вінерівського) випадкового процесу, у якого дисперсія з часом зростає, служить прикладом нестаціонарного випадкового процесу (про стаціонарні й нестаціонарні процеси див. вище пп. 9.1.2, 9.2.1).

Модель коливань, модульованих шумом (за амплітудою, частотою або фазою) можна розглядати як частинний випадок вузькосмугового випадкового процесу.

Дві останні моделі використовуються при аналізі флуктуацій сигналу автогенератора (див. нижче пп. 10.2.1-10.2.8).

9.3.1. Ергодичні процеси

Як уже вказувалося, ансамбль реалізацій випадкового процесу $x(t)$ дозволяє визначити його густину розподілу ймовірностей $w(x,t)$ (див. п. 9.1) та статистичні характеристики цього процесу (середнє значення, дисперсію тощо). Але для деяких стаціонарних випадкових процесів статистичні характеристики можна визначити за однією єдиною реалізацією.

Розглянемо деяку функцію часу

$$f(t)=\int_{-\infty}^{\infty} f(\omega)\exp(i\omega t)\,d\omega. \tag{9.3.1}$$

Середнє значення цієї функції на проміжку часу T можна записати через її спектр $f(\omega)$:

$$\begin{aligned}
\bar{f}_T(t)&=\frac{1}{T}\int_t^{t+T} f(t')\,dt'=\frac{1}{T}\int_t^{t+T} dt'\int_{-\infty}^{\infty} f(\omega)\exp(i\omega t')\,d\omega=\int_{-\infty}^{\infty} d\omega f(\omega)\frac{1}{T}\int_t^{t+T} dt'\exp(i\omega t')=\\
&=\int_{-\infty}^{\infty} d\omega f(\omega)\frac{1}{i\omega T}\left\{\exp\left[i\omega(t+T)\right]-\exp(i\omega t)\right\}=\\
&=\int_{-\infty}^{\infty} d\omega f(\omega)\frac{\exp\left[i\omega(t+T/2)\right]}{i\omega T}\left\{\exp(i\omega T/2)-\exp(i\omega T/2)\right\}=\\
&=\int_{-\infty}^{\infty} d\omega f(\omega)\frac{\exp\left[i\omega(t+T/2)\right]}{i\omega T}2i\sin(\omega T/2)=\int_{-\infty}^{\infty} f(\omega)\frac{\sin(\omega T/2)}{\omega T/2}\exp\left[i\omega(t+T/2)\right]d\omega.
\end{aligned} \tag{9.3.2}$$

Видно, що при усередненні за часом придушується високочастотна частина спектру (з частотами $\omega \geq 2\pi/T$) – див. рис. 9.3.1.

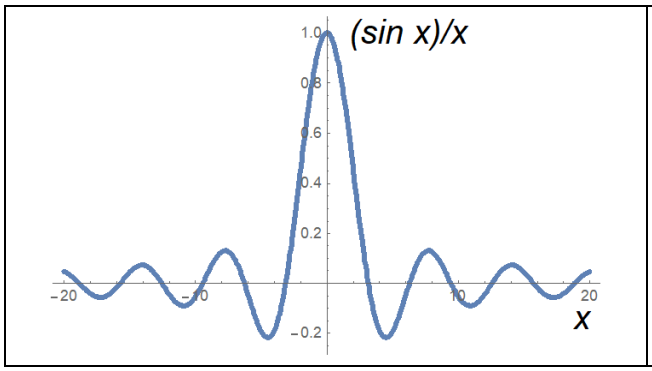


Рис. 9.3.1. Графік функції $\operatorname{sinc} x \equiv \dfrac{\sin x}{x}$

Тепер розглянемо усереднення за часом стаціонарного випадкового процесу $x(t)$, який можна подати у формі

$$x(t) = \langle x \rangle + \xi_0 + \xi(t), \tag{9.3.3}$$

де $\langle x \rangle$ – середнє значення (константа), $\xi(t)$ – флуктуаційна складова, а ξ_0 – стала, яка випадково змінюється від реалізації до реалізації (наприклад, випадкова похибка виставлення нуля вимірювального приладу).

Усереднення за часом T впливає лише на флуктуаційну складову, так що

$$\sigma_T^2 \equiv \left\langle \left(\bar{\xi}_T \right)^2 \right\rangle \xrightarrow[T\to\infty]{} 0. \tag{9.3.4}$$

Таким чином,

$$\bar{x} \equiv \lim_{T\to\infty} \bar{x}_T(t) = \langle x \rangle + \xi_0. \tag{9.3.5}$$

Якщо ж $\xi_0 = 0$, то

$$\bar{x} = \langle x \rangle. \tag{9.3.6}$$

Процеси, для яких виконується така рівність, називаються ергодичними.

Зрозуміло, що для ергодичних процесів знаходження статистичних характеристик значно спрощується порівняно з процесами, які не є ергодичними – для знаходження цих характеристик не потрібен ансамбль реалізацій, а достатньо однієї єдиної реалізації випадкового процесу (див. нижче).

Знайдемо закон спадання величини σ_T^2 при збільшенні часу усереднення T. Маємо:

$$\begin{aligned}
&\sigma_T^2 \equiv \left\langle \left(\bar{\xi}_T \right)^2 \right\rangle = \\
&= \left\langle \int_{-\infty}^{\infty} d\omega \xi(\omega) \frac{\sin(\omega T/2)}{\omega T/2} \exp\left[i\omega(t+T/2) \right] \int_{-\infty}^{\infty} \xi(\omega') \frac{\sin(\omega' T/2)}{\omega' T/2} \exp\left[i\omega'(t+T/2) \right] d\omega' \right\rangle = \\
&= \left\langle \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' \xi(\omega)\xi(\omega') \frac{\sin(\omega T/2)}{\omega T/2} \frac{\sin(\omega' T/2)}{\omega' T/2} \exp\left[i\omega(t+T/2) \right] \exp\left[i\omega'(t+T/2) \right] \right\rangle = \\
&= \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' \left\langle \xi(\omega)\xi(\omega') \right\rangle \frac{\sin(\omega T/2)}{\omega T/2} \frac{\sin(\omega' T/2)}{\omega' T/2} \exp\left[i\omega(t+T/2) \right] \exp\left[i\omega'(t+T/2) \right] = \\
&= \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' G(\omega) \delta(\omega+\omega') \frac{\sin(\omega T/2)}{\omega T/2} \frac{\sin(\omega' T/2)}{\omega' T/2} \exp\left[i\omega(t+T/2) \right] \exp\left[i\omega'(t+T/2) \right] = \\
&= \int_{-\infty}^{\infty} G(\omega) \left[\frac{\sin(\omega T/2)}{\omega T/2} \right]^2 d\omega.
\end{aligned} \tag{9.3.7}$$

Нехай спектр потужності на нульовій частоті є скінченним, $G(0)<\infty$, а в смузі частот шириною Ω в околі нуля його можна вважати сталим, $G(|\omega|\le\Omega)\approx G(0)$. Тоді за виконання умови $\Omega T\gg 1$ (рис. 9.3.2) можна записати:

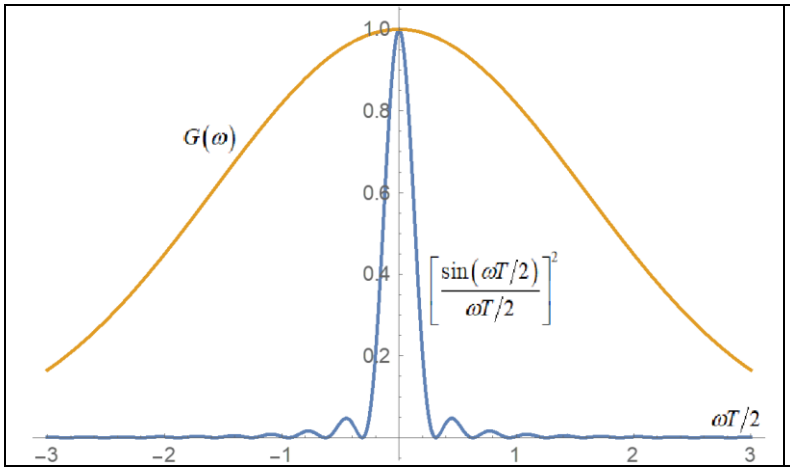


Рис. 9.3.2. До оцінки швидкості спадання усередненої за часом дисперсії

$$\sigma_T^2=\int_{-\infty}^{\infty}G(\omega)\left[\frac{\sin(\omega T/2)}{\omega T/2}\right]^2 d\omega\approx G(0)\int_{-\infty}^{\infty}\left[\frac{\sin(\omega T/2)}{\omega T/2}\right]^2 d\omega=\frac{2\pi G(0)}{T}, \qquad (9.3.8)$$

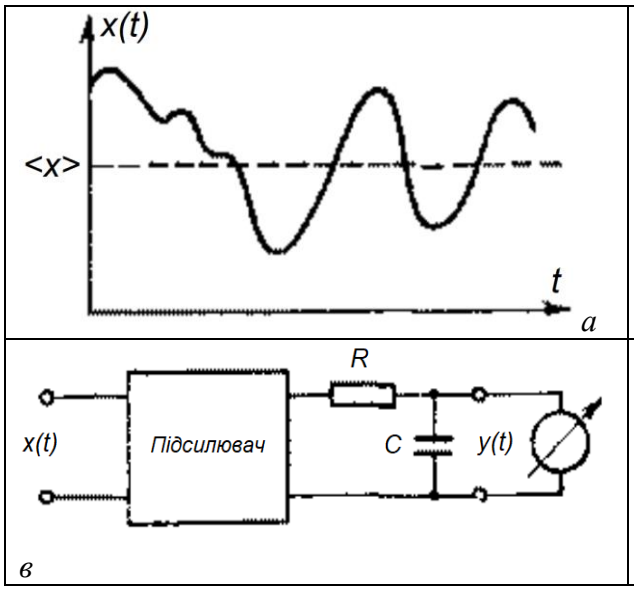
тобто σ_T^2 спадає як 1/T.

Як уже вказувалося, статистичні характеристики ергодичного процесу можна визначити за однією реалізацією. Проілюструємо це на прикладі визначення середнього значення. Схема відповідного вимірювального пристрою подана на рис. 9.3.3 в.

Підсилений сигнал $x(t)$ (рис. 9.3.3 а) надходить на інтегрувальну RC-ланку. Напруга на її виході має вигляд

$$y(t)=\frac{K}{\tau}\int_0^{\infty}x(t-\theta)\exp\left(-\frac{\theta}{\tau}\right)d\theta, \qquad (9.3.9)$$

де К – коефіцієнт підсилення підсилювача, $\tau=RC$, $\exp(-\theta/\tau)$ - імпульсна функція (функція Гріна) RC-ланки. Якщо при цьому виконано умову $\Omega\tau\gg 1$, де Ω – ширина спектру кореляцій, то $y(t)\approx\langle x\rangle$ (рис. 9.3.3 б).



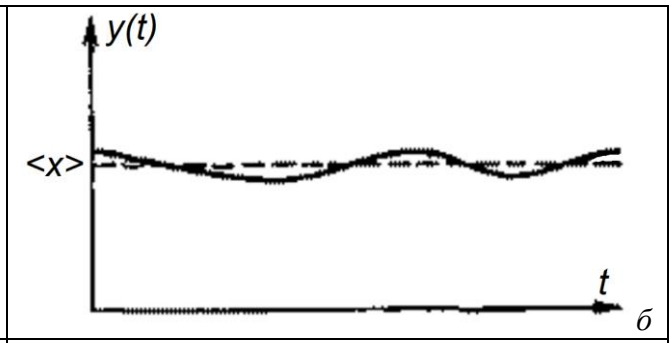


Рис. 9.3.3. Вимірювання середнього за часом значення випадкового процесу: а – вхідний сигнал, б – вихідний сигнал, в – схема вимірювального пристрою

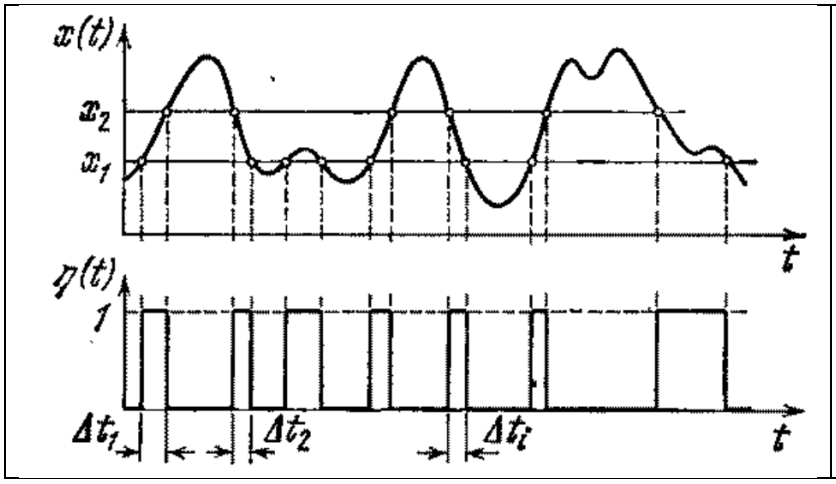


Рис. 9.3.4. До оцінки густини розподілу ймовірностей за однією реалізацією випадкового процесу x(t). Площа під кривою η(t) дорівнює часу перебування величини в інтервалі [x_1,x_2]

За однією реалізацією ергодичного процесу можна також оцінити густину розподілу ймовірностей. Для цього область значень функції $x(t)$ слід розбити на невеликі інтервали. Ймовірність $P(x)$ потрапляння випадкової функції $x(t)$ до окремого інтервалу визначається відносним часом перебування виміряного значення виміряної функції $x(t)$ в цьому інтервалі (рис. 9.3.4):

$$P(x)=\overline{\eta}(x)=\frac{1}{T}\int_0^T \eta(t)\,dt\,. \qquad (9.3.10)$$

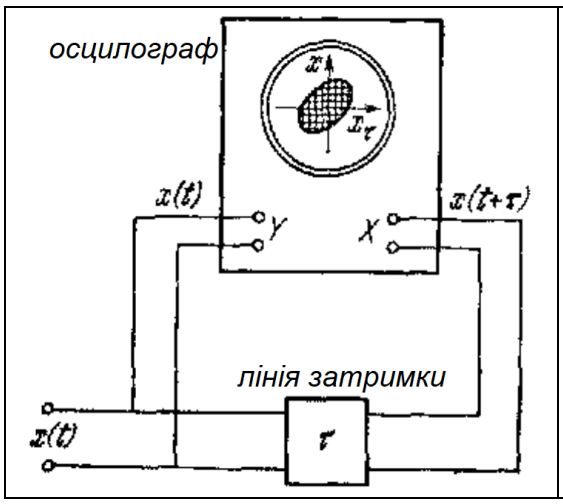


Рис. 9.3.5. Вимірювання двовимірного розподілу $w(x,x_\tau;\tau)$ стаціонарного випадкового процесу $x(t)$ (аналогова схема)

Для стаціонарного випадкового процесу $x(t)$ можна також експериментально виміряти двовимірний розподіл $w(x,x_\tau;\tau)$ (рис. 9.3.5), якому відповідає розподіл яскравості на екрані осцилографа (для заданого значення τ). Зараз такі завдання розв'язуються шляхом оцифрування реалізації випадкового процесу та її подальшої комп'ютерної обробки.

9.3.2. Нормальний (гауссівський) випадковий процес

У багатьох задачах, в тому числі в радіотехніці, трапляються випадкові процеси $x(t)$, для яких у кожний момент часу густина розподілу ймовірностей має вигляд

$$w(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \qquad (9.3.11)$$

– так званий нормальний розподіл, або розподіл Гаусса (рис. 9.3.6). Тут $\mu=\langle x\rangle$ – середнє

значення випадкової величини x, $\sigma^2 = \left\langle \left(x - \langle x \rangle \right)^2 \right\rangle$ – її дисперсія. Оскільки функція розподілу $w(x)$ не залежить від часу, то відповідний процес є стаціонарним.

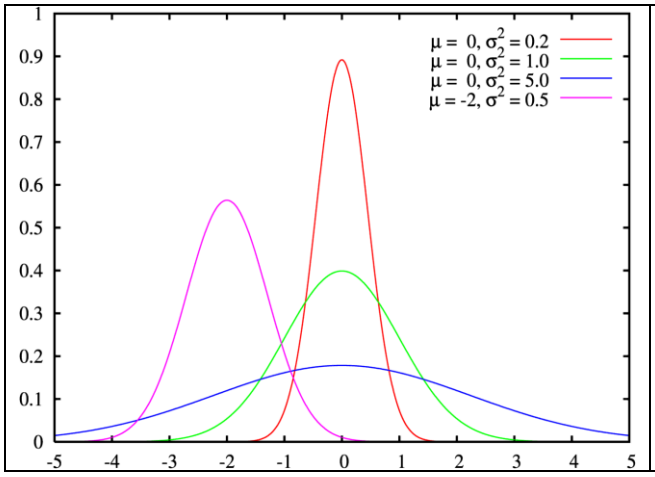


Рис. 9.3.6. Графіки нормального розподілу для різних значень параметрів

Для гауссівського розподілу центральні моменти непарних порядків дорівнюють нулеві, а центральні моменти парних порядків виражаються через дисперсію:

$$\left\langle \left(x - \langle x \rangle \right)^n \right\rangle = \begin{cases} 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (n-1)\sigma^2, & n - \textit{парне}; \\ 0, & n - \textit{непарне}. \end{cases} \tag{9.3.12}$$

Аналогічно кореляційні функції довільного порядку $B_{p\ldots s} = \left\langle \xi_p \ldots \xi_s \right\rangle$ для флуктуацій $\xi_p = x_p - \left\langle x_p \right\rangle$ статистично незалежних гауссових процесів $x_p, \ldots, x_s$ виражаються через кореляційні функції другого порядку:

$$B_{p\ldots s} = \begin{cases} \sum B_{pq} \ldots B_{rs}, & \textit{якщо порядок кореляційної функції парний}; \\ 0, & \textit{якщо порядок кореляційної функції непарний}. \end{cases} \tag{9.3.13}$$

Сума у формулі (9.3.13) охоплює всі можливі парні сполучення індексів. Наприклад,

$$B_{pqrs} = B_{pq} B_{rs} + B_{pr} B_{qs} + B_{ps} B_{qr}. \tag{9.3.14}$$

Зокрема, для автокореляційних функцій

$$B_{1122} = \left\langle \xi^2 \xi_\tau^2 \right\rangle = \sigma^4 + 2B^2(\tau); \tag{9.3.15}$$

(враховано, що для стаціонарного випадкового процесу $B_{11} = B_{22} = \sigma^2$, $B_{12} = B(\tau)$).

В теорії ймовірностей доводиться так звана центральна гранична теорема. Вона стверджує, що розподіл суми N статистично незалежних випадкових величин при $N \to \infty$ прямує до нормального. Цим і пояснюється поширення такого розподілу.

9.3.3. Білий шум

Випадковий процес із рівномірним спектром потужності $G(\omega) = const$ у літературі називають білим шумом. Як випливає з теореми Вінера – Хінчина, його автокореляційна функція має вигляд $B(\tau) = B_0 \delta(\tau)$, тобто час автокореляції білого шуму дорівнює нулеві. Такий процес, як ми вже вказували (п. 9.2.4), називають δ-корельованим. Видно, що $B(0) = \infty$, тобто білий шум несе нескінченно велику потужність. Отже, реалізувати такий процес неможливо – потужність будь-якого реального процесу завжди скінченна, а його спектр обмежений з боку верхніх частот. Тим не менше, моделлю білого шуму як вхідного сигналу можна користуватися, якщо ширина спектру такого сигналу значно більша за ширину смуги пропускання пристрою, на який він подається (рис. 9.3.7).

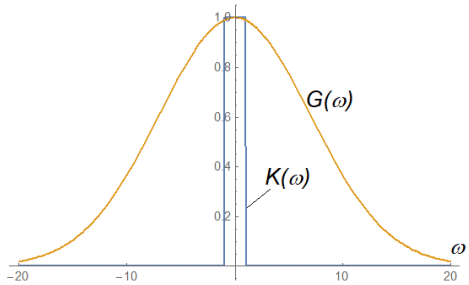


Рис. 9.3.7. Спектр потужності шуму $G(\omega)$, який для пристрою з передавальною функцією $K(\omega)$ може вважатися білим

9.3.4. Вузькосмуговий випадковий процес

В багатьох задачах радіотехніки трапляються так звані вузькосмугові випадкові процеси, для яких виконане співвідношення $\Delta\omega/\omega_0 \ll 1$, де $\Delta\omega$ і ω_0 - відповідно ширина і середня частота спектру такого процесу. Очевидно, чим менше $\Delta\omega$, тим ближче такий процес до гармонічного коливання. Його можна представити у вигляді

$$x(t) = \rho(t)\cos\left[\omega_0 t + \varphi(t)\right], \tag{9.3.16}$$

де $\rho(t)$ і $\varphi(t)$ - повільні в масштабі частоти ω_0 випадкові амплітуда (обвідна) та фаза (вони мало змінюються на проміжку часу $2\pi/\omega_0$).

Перетворимо вираз для $x(t)$:

$$\begin{aligned} x(t) &= \rho(t)\cos\left[\omega_0 t + \varphi(t)\right] = \rho(t)\cos\left[\varphi(t)\right]\cos\omega_0 t - \rho(t)\sin\left[\varphi(t)\right]\sin\omega_0 t = \\ &= a(t)\cos\omega_0 t - b(t)\sin\omega_0 t. \end{aligned} \tag{9.3.17}$$

Введені таким чином функції $a(t)$ та $b(t)$ називають квадратурними компонентами.

Обвідна, фаза та квадратурні компоненти пов'язані між собою співвідношеннями

$$a = \rho\cos\varphi,\ b = \rho\sin\varphi,\ \rho = \sqrt{a^2 + b^2},\ \varphi = arctg\frac{b}{a} \tag{9.3.18}$$

(рис. 9.3.8).

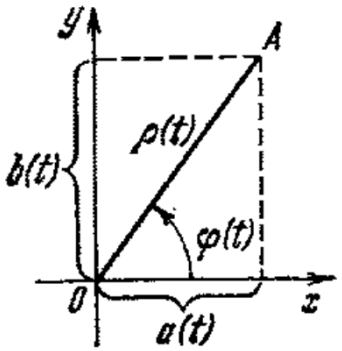


Рис. 9.3.8. Геометрична інтерпретація обвідної, фази та квадратурних компонент

Ввівши випадкові обвідну та фазу або квадратурні компоненти, ми тим самим замість однієї випадкової функції $\xi(t)$ ввели дві. Щоб однозначно визначити обвідну, вважатимемо, що вона знаходиться зі співвідношення

$$\rho(t) = \sqrt{\xi^2(t) + \left[\dot{\xi}(t)/\omega_0\right]^2} \tag{9.3.19}$$

(мається на увазі, що $\dot{\xi}(t) = -\omega_0\rho(t)\sin\left[\omega_0 t + \varphi(t)\right]$, тобто при записі цієї похідної ми нехтуємо похідними від повільних функцій). За відомою обвідною можна однозначно визначити фазу.

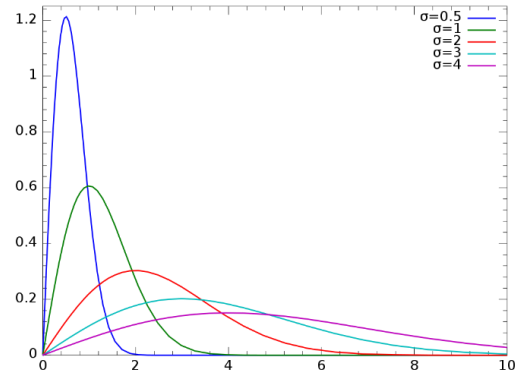


Рис. 9.3.9. Графіки розподілу Рейлі для різних значень параметрів

Якщо квадратурні компоненти є стаціонарними випадковими функціями з гауссівським розподілом, то випадкова фаза розподілена на проміжку $[0, 2\pi]$ рівномірно, $w(\varphi) = 1/2\pi$, а обвідна має розподіл

$$w(\rho) = \frac{\rho}{\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \tag{9.3.20}$$

відомий у літературі як розподіл Рейлі, див. рис. 9.3.9. Він уперше був запропонований Джоном Вільямом Стреттом (лордом Рейлі) в зв'язку з задачею про додавання гармонічних коливань із випадковими фазами. В радіотехніці цей розподіл використовують для опису флуктуацій амплітуди радіосигналу.

9.3.5. Випадкова імпульсна послідовність

Випадковою імпульсною послідовністю називають випадковий процес, коли на заданому проміжку часу T з'являється послідовність однакових імпульсів $F(t - t_p)$, кожен з яких виникає незалежно від інших, причому момент початку кожного імпульсу t_p та їхня загальна кількість n є статистично незалежними випадковими величинами:

$$x(t) = \sum_{i=1}^{n} F(t - t_i) \tag{9.3.21}$$

(рис. 9.3.10).

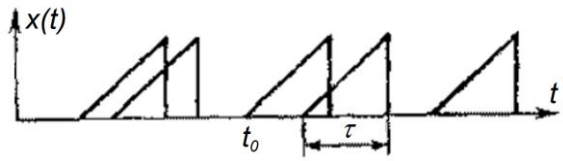


Рис. 9.3.10. Випадкова імпульсна послідовність

Перш ніж розглядати випадкову імпульсну послідовність, розберемо простішу модель – одиночний імпульс заданої форми $F(t)$, що з'являється у випадковий момент часу t_0:

$$x(t) = F(t - t_0). \tag{9.3.22}$$

Нехай $w(t_0)$ – густина розподілу ймовірності появи імпульсу в момент часу t_0. Тоді статистичні середні визначаються формулами

$$\langle x \rangle = \int_{-\infty}^{\infty} F(t - t_0) w(t_0)\, dt_0, \tag{9.3.23}$$

$$\langle x x_\tau \rangle = \int_{-\infty}^{\infty} F(t - t_0) F(t - t_0 + \tau) w(t_0)\, dt_0. \tag{9.3.24}$$

Нехай тривалість імпульсу $F(t)$ складає τ_p, а $w(t_0) \neq 0$ на проміжку тривалістю T

(нагадаємо, що в силу умови нормування $w(t_0) \to 0$ при $t_0 \to \pm\infty$). Нехай функція $F(t)$ досягає максимуму в точці $t = t'$, а добуток $F(t)F(t+\tau)$ – в точці $t = t''$.

Припустимо, що виконана умова $\tau_p \ll T$. Тоді на проміжку τ_p функція $w(t_0)$ майже не змінюється, і її можна винести з-під інтегралів (9.3.23)-(9.3.24) зі значеннями в точках, де величини відповідно $F(t-t_0)$ і $F(t-t_0)F(t-t_0+\tau)$ максимальні. Отримаємо:

$$\langle x \rangle \approx w(t-t') \int_{-\infty}^{\infty} F(\theta) d\theta, \tag{9.3.25}$$

$$\langle x x_\tau \rangle \approx w(t-t'') \int_{-\infty}^{\infty} F(\theta) F(\theta+\tau) d\theta. \tag{9.3.26}$$

Тепер конкретизуємо вигляд функції $w(t_0)$. Нехай ймовірність появи імпульсу однакова на всьому проміжку часу тривалістю T. Тоді з умови нормування випливає, що

$$w(t_0) = \begin{cases} 1/T, t_0 \in [0,T]; \\ 0, \quad t_0 \notin [0,T], \end{cases} \tag{9.3.27}$$

звідки

$$\langle x \rangle \approx \frac{1}{T} \int_{-\infty}^{\infty} F(\theta) d\theta; \tag{9.3.28}$$

$$\langle x x_\tau \rangle \approx \frac{1}{T} \int_{-\infty}^{\infty} F(\theta) F(\theta+\tau) d\theta. \tag{9.3.29}$$

Тепер подамо імпульс $F(t)$ через його спектр $F(\omega)$:

$$F(t) = \int_{-\infty}^{\infty} F(\omega) \exp(i\omega t) d\omega, \tag{9.3.30}$$

де

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(t) \exp(-i\omega t) dt, \qquad F(-\omega) = F^*(\omega). \tag{9.3.31}$$

Підставивши (9.3.30) до (9.3.31), отримуємо:

$$\langle x x_\tau \rangle \approx \frac{1}{T} \int_{-\infty}^{\infty} d\theta \int_{-\infty}^{\infty} F(\omega) \exp(i\omega\theta) d\omega \int_{-\infty}^{\infty} F(\omega') \exp[i\omega'(\theta+\tau)] d\omega' =$$

$$= \frac{1}{T} \int_{-\infty}^{\infty} F(\omega) d\omega \int_{-\infty}^{\infty} F(\omega') \exp(i\omega'\tau) d\omega' \int_{-\infty}^{\infty} \exp[i\theta(\omega+\omega')] d\theta =$$

$$= \frac{1}{T} \int_{-\infty}^{\infty} F(\omega) d\omega \int_{-\infty}^{\infty} F(\omega') \exp(i\omega'\tau) d\omega' 2\pi\delta(\omega+\omega') =$$

$$= \frac{2\pi}{T} \int_{-\infty}^{\infty} F(\omega) F(-\omega) \exp(-i\omega\tau) d\omega = \frac{2\pi}{T} \int_{-\infty}^{\infty} |F(\omega)|^2 \exp(-i\omega\tau) d\omega. \tag{9.3.32}$$

Порівнюючи (9.3.32) з теоремою Вінера – Хінчина (9.2.30), можна записати спектр потужності одиночного імпульсу, що починається у випадковий момент часу:

$$G(\omega) = \frac{2\pi}{T} |F(\omega)|^2. \tag{9.3.33}$$

Тепер розглянемо випадкову імпульсну послідовність (9.3.21):

$$x(t) = \sum_{i=1}^{n} F(t-t_i).$$

Нагадаємо, що тут випадковими величинами є момент t_i появи кожного імпульсу та

їхня загальна кількість n. Всі ці випадкові величини вважатимемо статистично незалежними, так що

$$w(n;t_1,t_2,...t_n)=w(n)w(t_1)w(t_2)...w(t_n). \tag{9.3.34}$$

Всі розподіли $w(t_i)$ вважаємо однаковими, тому статистичні середні

$$\langle F(t-t_i)\rangle \equiv \langle F\rangle, \qquad \langle F(t-t_i+\tau)\rangle \equiv \langle F_\tau\rangle, \qquad \langle F(t-t_i)F(t-t_i+\tau)\rangle \equiv \langle FF_\tau\rangle \tag{9.3.35}$$

не залежать від індексу i.

Тоді

$$\langle x(t)\rangle = \left\langle \sum_{n=0}^{\infty} F(t-t_i)\right\rangle = \sum_{n=0}^{\infty} w(n)\int_{-\infty}^{\infty} w(t_1)dt_1\int_{-\infty}^{\infty} w(t_2)dt_2...\int_{-\infty}^{\infty} w(t_n)dt_n\sum_{i=1}^{n} F(t-t_i)=$$

$$=\sum_{n=0}^{\infty} w(n)\int_{-\infty}^{\infty} w(t_1)dt_1\int_{-\infty}^{\infty} w(t_2)dt_2...\int_{-\infty}^{\infty} w(t_{n-1})dt_{n-1}\left[\sum_{i=1}^{n}\int_{-\infty}^{\infty} w(t_n)dt_nF(t-t_i)\right]=$$

$$=\sum_{n=0}^{\infty} w(n)\int_{-\infty}^{\infty} w(t_1)dt_1\int_{-\infty}^{\infty} w(t_2)dt_1...\int_{-\infty}^{\infty} w(t_{n-1})dt_{n-1}\left[\int_{-\infty}^{\infty} w(t_n)dt_nF(t-t_n)+\sum_{i\neq n}^{n} F(t-t_i)\int_{-\infty}^{\infty} w(t_n)dt_n\right]=$$

$$=\sum_{n=0}^{\infty} w(n)\int_{-\infty}^{\infty} w(t_1)dt_1\int_{-\infty}^{\infty} w(t_2)dt_2...\int_{-\infty}^{\infty} w(t_{n-1})dt_{n-1}\left[\int_{-\infty}^{\infty} w(t_n)dt_nF(t-t_n)+\sum_{i=1}^{n-1} F(t-t_i)\right]=$$

$$=\sum_{n=0}^{\infty} w(n)\int_{-\infty}^{\infty} w(t_1)dt_1\int_{-\infty}^{\infty} w(t_2)dt_2...\int_{-\infty}^{\infty} w(t_{n-1})dt_{n-1}\left[\langle F\rangle+\sum_{i}^{n-1} F(t-t_i)\right]=$$

$$=\sum_{n=0}^{\infty} w(n)\int_{-\infty}^{\infty} w(t_1)dt_1\int_{-\infty}^{\infty} w(t_2)dt_2...\int_{-\infty}^{\infty} w(t_{n-2})dt_{n-2}\left[2\langle F\rangle+\sum_{i}^{n-2} F(t-t_i)\right]...=\sum_{n=0}^{\infty} w(n)n\langle F\rangle=$$

$$=\langle F\rangle\sum_{n=0}^{\infty} w(n)n=\langle F\rangle\langle n\rangle, \tag{9.3.36}$$

$$\langle x(t+\tau)\rangle=\langle F_\tau\rangle\langle n\rangle, \tag{9.3.37}$$

$$\langle x(t)x(t+\tau)\rangle=\left\langle\sum_{i=1}^{n}F(t-t_i)\sum_{i'=1}^{n}F(t-t_{i'}+\tau)\right\rangle=\left\langle\sum_{i=1}^{n}\sum_{i'=1}^{n}F(t-t_i)F(t-t_{i'}+\tau)\right\rangle=$$

$$=\left\langle\sum_{i=i'}F(t-t_i)F(t-t_{i'}+\tau)+\sum_{i\neq i'}F(t-t_i)F(t-t_{i'}+\tau)\right\rangle= \tag{9.3.38}$$

$$=\left\langle n\langle FF_\tau\rangle+\left(n^2-n\right)\langle F\rangle\langle F_\tau\rangle\right\rangle=\langle n\rangle\langle FF_\tau\rangle+\left(\langle n^2\rangle-\langle n\rangle\right)\langle F\rangle\langle F_\tau\rangle.$$

Автокореляційна функція процесу $x(t)$:

$$\begin{aligned}B(\tau)&=\langle xx_\tau\rangle-\langle x\rangle\langle x_\tau\rangle=\langle n\rangle\langle FF_\tau\rangle+\left(\langle n^2\rangle-\langle n\rangle\right)\langle F\rangle\langle F_\tau\rangle-\langle n\rangle^2\langle F\rangle\langle F_\tau\rangle=\\&=\langle n\rangle\langle FF_\tau\rangle+\left(\langle n^2\rangle-\langle n\rangle^2-\langle n\rangle\right)\langle F\rangle\langle F_\tau\rangle.\end{aligned} \tag{9.3.39}$$

Як і для одиночного імпульсу, вважатимемо виконаними умову $T\gg\tau$, де τ – тривалість окремого імпульсу, T – інтервал спостереження та рівність (9.3.27)

$$w(t_i)=\begin{cases}1/T, t_i\in[0,T];\\0, \quad t_i\notin[0,T].\end{cases}$$

Тоді можна вважати, що $\langle F\rangle=\langle F_\tau\rangle$, і величини $\langle F\rangle$ та $\langle FF_\tau\rangle$ визначаються формулами (9.3.28)-(9.3.29).

Тепер знайдемо розподіл випадкової величини n, що відповідає цій моделі. Для цього розіб'ємо інтервал спостереження T на проміжки $\Delta t=T/N$, настільки малі, що в кожному з них починається тільки один випадковий імпульс (із ймовірністю p) або не починається

жодного (із ймовірністю $q = 1 - p$). Тоді ймовірність того, що на N інтервалах тривалістю Δt з'явиться n імпульсів, складає величину

$$P(n) = C_N^n p^n q^{N-n}, \qquad C_N^n = \frac{N!}{n!(N-n)!}, \qquad k! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (k-1) \cdot k. \tag{9.3.40}$$

Тут C_N^n – кількість способів, якими можна розташувати n точок (початку імпульсів) в N дискретних інтервалах (за домовленістю вважають, що $0! = 1$).

Якщо $N \gg n$, можна записати, що $C_N^n \approx N^n / n!$ і

$$P(n) \approx N^n p^n q^{N-n} / n! = C\alpha^n / n!, \qquad \alpha = Np. \tag{9.3.41}$$

Константу C можна знайти з умови нормування, яка в даному разі має вигляд

$$1 = \sum_{n=0}^{\infty} P(n) = \sum_{n=0}^{\infty} C\alpha^n / n! = C \sum_{n=0}^{\infty} \alpha^n / n! = C \exp(\alpha), \tag{9.3.42}$$

звідки

$$C = \exp(-\alpha) \tag{9.3.43}$$

Таким чином, ймовірність того, що на інтервалі T буде спостерігатися n імпульсів, дається формулою

$$P(n) = \frac{\alpha^n}{n!} \exp(-\alpha) \tag{9.3.44}$$

– так званим розподілом Пуассона (рис. 9.3.10).

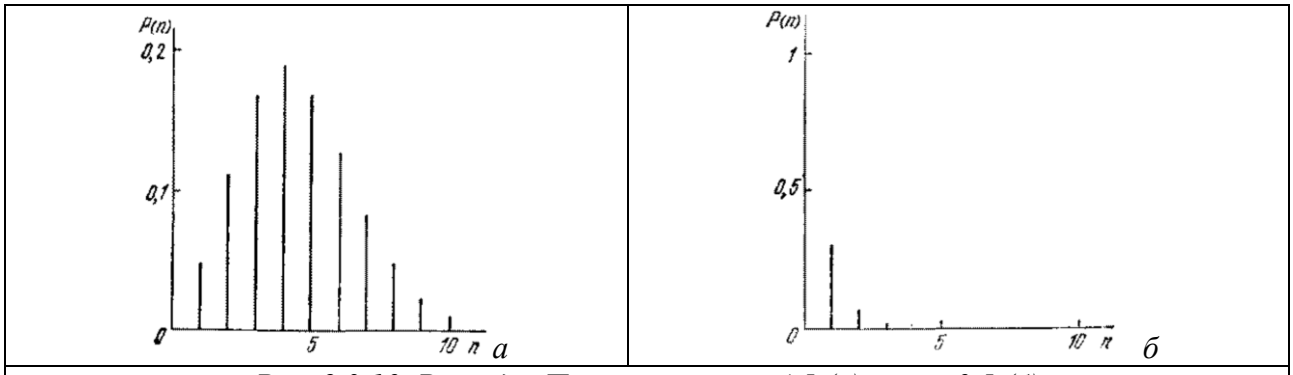


Рис. 9.3.10. Розподіл Пуассона при α=4.5 (а) та α=0.5 (б)

Користуючись розподілом Пуассона, знайдемо $\langle n \rangle$:

$$\langle n \rangle = \sum_{n=0}^{\infty} nP(n) = \sum_{n=0}^{\infty} n \frac{\alpha^n}{n!} \exp(-\alpha) = \alpha \exp(-\alpha) \sum_{n=1}^{\infty} \frac{\alpha^{n-1}}{(n-1)!} = \alpha. \tag{9.3.45}$$

Відповідно

$$\langle n^2 \rangle = \sum_{n=0}^{\infty} n^2 \frac{\alpha^n}{n!} \exp(-\alpha) = \exp(-\alpha)\left[\left(\alpha + \alpha^2\right)\exp(\alpha)\right] = \alpha + \alpha^2. \tag{9.3.46}$$

Підставимо (9.3.44)-(9.3.45) до (9.3.36) та (9.3.39) з урахуванням (9.3.28)-(9.3.29). Відповідно отримаємо:

$$\langle x \rangle = \langle F \rangle \langle n \rangle = \alpha \langle F \rangle = \frac{\alpha}{T} \int_{-\infty}^{\infty} F(\theta) d\theta = \Omega \int_{-\infty}^{\infty} F(\theta) d\theta, \tag{9.3.47}$$

$$\begin{aligned} B(\tau) &= \langle n \rangle \langle FF_\tau \rangle + \left(\langle n^2 \rangle - \langle n \rangle^2 - \langle n \rangle\right) \langle F \rangle \langle F_\tau \rangle = \\ &= \alpha \langle FF_\tau \rangle + \left(\alpha^2 + \alpha - \alpha^2 - \alpha\right) \langle F \rangle \langle F_\tau \rangle = \end{aligned} \tag{9.3.48}$$

$$=\alpha\langle FF_\tau\rangle=\Omega\int_{-\infty}^{\infty}F(\theta)F(\theta+\tau)d\theta,$$

де $\Omega=\alpha/T$ – середня частота надходження імпульсів на інтервалі спостереження.

Скориставшись теоремою Вінера – Хінчина, за відомою автокореляційною функцією можна знайти спектральну інтенсивність $G(\omega)$ випадкового процесу $x(t)$:

$$G(\omega)=2\pi\Omega\left|F^2(\omega)\right|,\qquad F(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}F(t)\exp(-i\omega t)dt. \tag{9.3.49}$$

(пор. з формулою (9.3.33)).

9.3.6. Дифузійний випадковий процес

Дифузійний випадковий процес $\xi(t)$ описується диференціальним рівнянням вигляду

$$\frac{d\xi}{dt}=\eta(t), \tag{9.3.50}$$

де $\eta(t)$ – випадкова функція часу. Очевидно, розв'язок рівняння (9.3.50) можна подати у вигляді

$$\xi(t)=\int_0^t\eta(t')dt'. \tag{9.3.51}$$

Як буде показано далі, до вигляду (9.3.50) зводиться, наприклад, рівняння для флуктуацій фази LC-автогенератора.

Нехай

$$\langle\eta(t)\rangle=0,\quad \langle\eta^2(t)\rangle=\sigma_0^2,\qquad \langle\eta\eta_\tau\rangle=B_0(\tau)=\int_{-\infty}^{\infty}G_0(\omega)\exp(i\omega\tau)d\omega, \tag{9.3.52}$$

то, відповідно до (9.3.51),

$$\langle\xi(t)\rangle=0, \tag{9.3.53}$$

$$\begin{aligned}\frac{d}{dt}\sigma^2(t)&=\frac{d}{dt}\langle\xi^2(t)\rangle=2\langle\xi(t)\eta(t)\rangle=2\left\langle\eta(t)\int_0^t\eta(t')dt'\right\rangle=\\&=2\int_0^t\langle\eta(t)\eta(t')\rangle dt'=2\int_0^tB_0(t-t')dt'=\begin{vmatrix}t-t'=\tau\\dt'=-d\tau\end{vmatrix}=-2\int_t^0B_0(\tau)d\tau=2\int_0^tB_0(\tau)d\tau.\end{aligned} \tag{9.3.54}$$

Тоді

$$\sigma^2(t)=2\int_0^tdt'\int_0^{t'}B_0(\tau)d\tau. \tag{9.3.55}$$

Врахуємо, що

$$\begin{aligned}\int_0^t\tau B_0(\tau)d\tau&=\begin{vmatrix}B_0(\tau)d\tau=dv & \tau=u\\ v=\int B_0(\tau)d\tau & du=d\tau\end{vmatrix}=\left[\tau\int B_0(\tau)d\tau-\int d\tau\int B_0(\tau)d\tau\right]_0^t=\\&=t\int_0^tB_0(\tau)d\tau-\int_0^tdt'\int_0^{t'}B_0(\tau)d\tau.\end{aligned} \tag{9.3.56}$$

Тоді

$$\int_0^tdt'\int_0^{t'}B_0(\tau)d\tau=t\int_0^tB_0(\tau)d\tau-\int_0^t\tau B_0(\tau)d\tau=\int_0^t(t-\tau)B_0(\tau)d\tau,$$

і остаточно

$$\sigma^2(t) = 2\int_0^t (t-\tau) B_0(\tau) d\tau . \qquad (9.3.57)$$

Видно, що дисперсія дифузійного процесу залежить від часу, тобто він нестаціонарний.

Для малих проміжків часу t ($t \ll \tau_C$, де τ_C, як і раніше – час спадання функції $B_0(\tau)$ або $t \ll T$, де T – період її осциляцій), можна вважати, що

$$2\int_0^t (t-\tau) B_0(\tau) d\tau \approx 2B_0(0)\int_0^t (t-\tau) d\tau = 2B_0(0)\left(t\tau - \tau^2/2\right)\Big|_0^t = B_0(0)t^2 = \sigma_0^2 t^2 ,$$

отримуємо:

$$\sigma^2(t) = \sigma_0^2 t^2 . \qquad (9.3.58)$$

Навпаки, при великих t ($t \gg \tau_C$) можна вважати, що $B_0(\tau = t) = 0$, і в інтегралі (9.3.57) верхню межу можна покласти нескінченною. Тоді

$$\sigma^2(t) = 2\int_0^\infty (t-\tau) B_0(\tau) d\tau = 2\left[t\int_0^\infty B_0(\tau) d\tau - \int_0^\infty \tau B_0(\tau) d\tau \right] =$$
$$= 2(t-\tau_0)\int_0^\infty B_0(\tau) d\tau, \qquad (9.3.59)$$

де

$$\tau_0 = \frac{\int_0^\infty \tau B_0(\tau) d\tau}{\int_0^\infty B_0(\tau) d\tau} . \qquad (9.3.60)$$

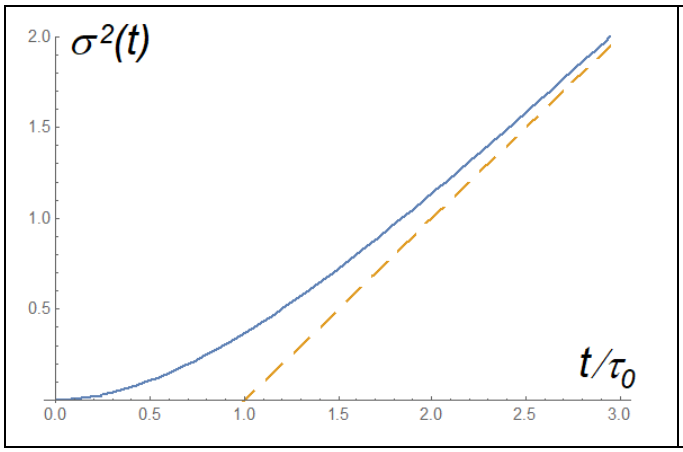


Рис. 9.3.11. Часова залежність дисперсії дифузійного випадкового процесу

Щоб зрозуміти зміст параметра τ_0, оберемо функцію $B_0(\tau)$ у формі

$$B_0(\tau) = B_0\Theta(\tau_c - |t|),$$

де $\Theta(x)$ – функція Хевісайда, τ_c – час автокореляції. Тоді

$$\tau_0 = \frac{\int_0^{\tau_c} \tau B_0 d\tau}{\int_0^{\tau_c} B_0 d\tau} = \frac{B_0\tau_c^2/2}{B_0\tau_c} = \frac{\tau_c}{2} .$$

Отже, величина τ_c має той самий порядок, що й час автокореляції.

Таким чином, у початкові моменти часу $\sigma^2(t)$ зростає за квадратичним законом, а в

пізні – за лінійним.

Нехай випадковий процес $\eta(t)$ – білий шум. У цьому випадку процес $\xi(t)$ називають вінерівським. Оскільки білий шум характеризується рівномірним спектром, $G_0(\omega) = G_0 = const$, і, відповідно до теореми Вінера – Хінчина,

$$B_0(\tau) = 2\pi G_0 \delta(\tau) \tag{9.3.61}$$

(як уже зазначалося, білий шум є δ-корельованим). Тоді

$$\sigma^2(t) = 2\int_0^t (t-\tau) 2\pi G_0 \delta(\tau) d\tau = 4\pi G_0 \int_0^t (t-\tau)\delta(\tau) d\tau = 2\pi G_0 t\,. \tag{9.3.62}$$

Отже, для вінерівського процесу дисперсія лінійно зростає з часом (рис. 9.3.12 а).

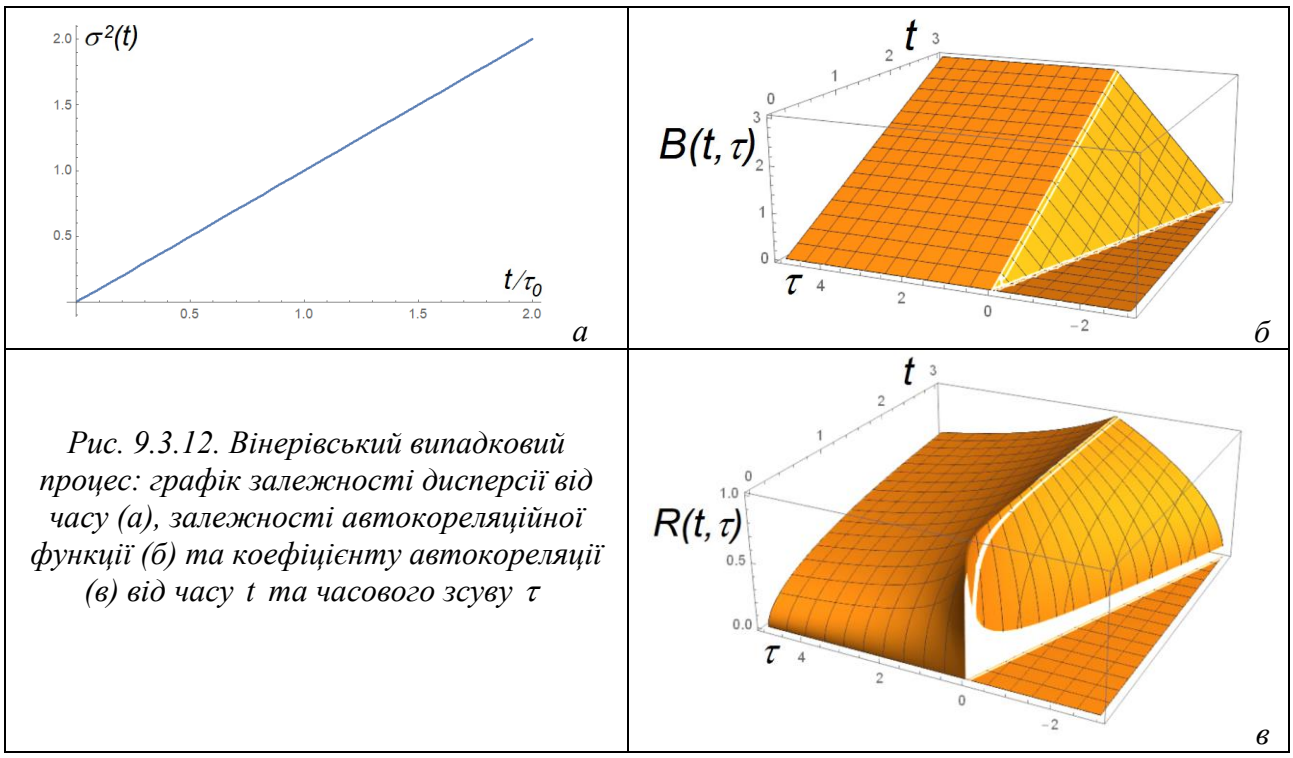


Рис. 9.3.12. Вінерівський випадковий процес: графік залежності дисперсії від часу (а), залежності автокореляційної функції (б) та коефіцієнту автокореляції (в) від часу t та часового зсуву τ

Автокореляційна функція вінерівського процесу має вигляд:

$$B(t,\tau) = \langle \xi(t)\xi(t+\tau)\rangle = \left\langle \int_0^t \eta(t')dt' \int_0^{t+\tau} \eta(t'')dt'' \right\rangle = \int_0^t dt' \int_0^{t+\tau} dt'' \langle \eta(t')\eta(t'')\rangle =$$
$$= \int_0^t dt' \int_0^{t+\tau} dt'' 2\pi G_0 \delta(t'-t'') = 2\pi G_0 \int_0^t dt' \int_0^{t+\tau} dt'' \delta(t'-t''). \tag{9.3.63}$$

Функція $\xi(t)$ визначена лише при $t \ge 0$, тому вважатимемо цю нерівність виконаною.

Якщо $\tau \ge 0$, то

$$\int_0^t dt' \int_0^{t+\tau} dt'' \delta(t'-t'') = \int_0^t dt' = t\,.$$

Якщо $t+\tau < 0$, або $\tau < -t$, функція $\xi(t+\tau) = 0$, і

$B(t,\tau) = 0$.

Нарешті, при $-t < \tau < 0$

$$\int_0^t dt' \int_0^{t+\tau} dt'' \delta(t'-t'') = \int_0^t dt' \int_0^{t-|\tau|} dt'' \delta(t'-t'') = \int_0^{t-|\tau|} dt'' \int_0^t dt' \delta(t'-t'') = \int_0^{t-|\tau|} dt'' = t - |\tau|\,.$$

Таким чином,

$$B(t,\tau)=\begin{cases} 2\pi G_0 t, & \tau\ge 0; \\ 2\pi G_0\left(t-|\tau|\right), & -t<\tau<0; \\ 0, & \tau<-t \end{cases}$$

(рис. 9.3.12 б).

Коефіцієнт кореляції:

$$R(t,\tau)=\frac{B(t,\tau)}{\sigma(t)\sigma(t+\tau)}=\begin{cases} 1\big/\sqrt{1+\tau/t}, & \tau\ge 0; \\ \sqrt{1+\tau/t}, & -t<\tau<0; \\ 0, & \tau<-t \end{cases} \qquad (9.3.64)$$

(рис. 9.3.12 в). Видно, що для вінерівського процесу коефіцієнт кореляції зростає з часом і завжди прямує до 1 при $t\to\infty$ (тобто при $\tau/t\to 0$).

9.3.7. Коливання, модульовані шумом

Коливання, модульовані шумом – це випадковий процес, коли гармонічне коливання зазнає випадкової модуляції за амплітудою, фазою або частотою. Коливання, модульовані шумом, можуть розглядатись як частинний випадок вузькосмугового шуму (9.3.16)

$$x(t)=\rho(t)\cos\left[\omega_0 t+\varphi(t)\right]$$

(див. вище п. 9.3.4).

Модель коливань, модульованих шумом, використовується, наприклад, в задачі про автогенератор, на який діє випадкова сила (в ролі цієї сили виступають зазвичай внутрішні шуми). Розв'язання цієї задачі дозволяє знайти форму й ширину спектру автоколивань.
Як і при модуляції детермінованим сигналом, розрізняють амплітудну, фазову та частотну модуляцію шумом.

Одна з цілей наведених далі розрахунків – з'ясувати для кожного типу модуляції зв'язок між спектрами керувального (модулюючого) сигналу та модульованого сигналу. В усіх випадках застосовується приблизно однаковий підхід:

- задаємо властивості керувального сигналу (випадкового процесу);
- знаходимо (з більшими чи меншими зусиллями) автокореляційну функцію модульованого сигналу;
- користуючись теоремою Вінера – Хінчина, за відомою автокореляційною функцією знаходимо спектр модульованого сигналу.

<u>Амплітудна модуляція шумом</u>. Відповідний сигнал має вигляд

$$x(t)=a_0\left[1+\eta(t)\right]\cos\left[\omega_0 t+\varphi_0\right]. \qquad (9.3.65)$$

Вважатимемо, що $\eta(t)$ – стаціонарний випадковий процес, причому $1+\eta(t)\ge 0$, а початкова фаза φ_0 – випадкова величина, рівномірно розподілена на інтервалі $\left[-\pi,\pi\right]$, тобто

$$w(\varphi_0)=\frac{1}{2\pi},\quad -\pi\le\varphi_0\le\pi. \qquad (9.3.66)$$

Тоді випадковий процес $x(t)$ також буде стаціонарним.

Нехай

$$\langle\eta(t)\rangle=0, \qquad \langle\eta\eta_\tau\rangle=B_0(\tau)=\int_{-\infty}^{\infty}G_0(\omega)\exp(i\omega\tau)\,d\omega. \qquad (9.3.67)$$

Тоді

$$\overline{x}=0,$$

$$B(\tau)\equiv\langle xx_\tau\rangle=\left\langle a_0\left[1+\eta(t)\right]\cos\left[\omega_0 t+\varphi_0\right]a_0\left[1+\eta(t+\tau)\right]\cos\left[\omega_0(t+\tau)+\varphi_0\right]\right\rangle=$$

$$=\frac{a_0^2}{2}\left\langle\left[1+\eta(t)+\eta(t+\tau)+\eta(t)\eta(t+\tau)\right]\left\{\cos\left[\omega_0(2t+\tau)+2\varphi_0\right]+\cos\omega_0\tau\right\}\right\rangle.$$

Оскільки φ_0 і $\eta(t)$ – статистично незалежні, то усереднення можна виконувати спершу за φ_0, а потім за t. Виконаємо спершу усереднення за φ_0. Оскільки

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\cos\left[\omega_0(2t+\tau)+2\varphi_0\right]d\varphi_0=0,$$

маємо:

$$B(\tau)=\frac{a_0^2}{2}\cos\omega_0\tau\left\langle\left[1+\eta(t)+\eta(t+\tau)+\eta(t)\eta(t+\tau)\right]\right\rangle=$$

$$=\frac{a_0^2}{2}\cos\omega_0\tau\left[1+\langle\eta(t)\rangle+\langle\eta(t+\tau)\rangle+\langle\eta(t)\eta(t+\tau)\rangle\right]=\frac{1}{2}a_0^2\left[1+B_0(\tau)\right]\cos\omega_0\tau. \tag{9.3.68}$$

Відповідна спектральна інтенсивність, згідно теореми Вінера – Хінчина, має вигляд

$$G(\omega)=\frac{1}{4}a_0^2\left[\delta(\omega-\omega_0)+\delta(\omega+\omega_0)+G_0(\omega-\omega_0)+G_0(\omega+\omega_0)\right]. \tag{9.3.69}$$

Враховано, що, згідно теореми Вінера – Хінчина,

$$G(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}B(\tau)\exp(i\omega\tau)\,d\tau,$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}B_0(\tau)\cos\omega\tau\exp(i\omega\tau)\,d\tau=\frac{1}{4\pi}\int_{-\infty}^{\infty}B_0(\tau)\cos\omega\tau\exp\left[i(\omega+\omega_0)\tau\right]d\tau+$$

$$+\frac{1}{4\pi}\int_{-\infty}^{\infty}B_0(\tau)\cos\omega\tau\exp\left[i(\omega-\omega_0)\tau\right]d\tau=\frac{1}{2}\left[G(\omega+\omega_0)+G(\omega-\omega_0)\right].$$

Перші два доданки в правій частині рівняння (9.3.69) відповідають детермінованій складовій процесу $\xi(t)$, два інші – шумовій модуляції. Загалом співвідношення між спектрами модуляції та модульованого сигналу таке саме, як при амплітудній модуляції детермінованим сигналом.

Експериментально можна визначити величину

$$G^+(\omega)=\begin{cases}2G(\omega), & \omega>0;\\ 0, & \omega<0.\end{cases} \tag{9.3.70}$$

В даному разі

$$G^+(\omega)=\frac{1}{2}a_0^2\left[\delta(\omega-\omega_0)+G_0(\omega-\omega_0)\right] \tag{9.3.71}$$

(рис. 9.3.13). Видно, що спектр $G^+(\omega)$ симетричний щодо частоти ω_0.

Розподіл для процесу $x(t)$ визначається співвідношенням

$$w(x)=\frac{1}{\pi}\int_0^{\infty}d\eta\,\frac{w(\rho)}{\rho}\bigg|_{\rho=\sqrt{x^2+\eta^2}}=\frac{1}{\pi}\int_{|x|}^{\infty}\frac{w(\rho)\,d\rho}{\sqrt{\rho^2-x^2}}, \tag{9.3.72}$$

де $w(\rho)$ – розподіл обвідної, який, у свою чергу, визначається статистикою амплітудної модуляції $\eta(t)$.

<u>Фазова модуляція шумом</u>. Відповідний випадковий процес має вигляд

$$x(t)=a_0\cos\left[\omega_0 t+\varphi(t)\right]=\frac{1}{2}a_0\exp\left[i\omega_0 t+i\varphi(t)\right]+к.с., \tag{9.3.73}$$

де амплітуда a_0 стала, а фаза $\varphi(t)$ зазнає випадкових флуктуацій. Відповідно

$$\langle x(t)\rangle = \frac{1}{2}a_0 \exp(i\omega_0 t)\langle \exp[i\varphi(t)]\rangle + к.с. \qquad (9.3.74)$$

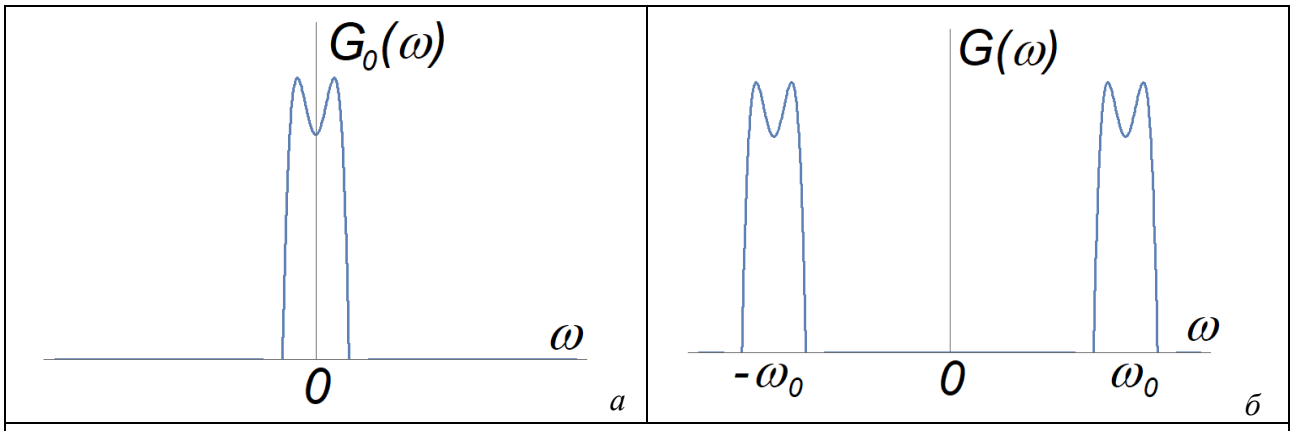


Рис. 9.3.13. Спектри сигналу модуляції $G_0(\omega)$ та відповідного амплітудно-модульованого сигналу $G(\omega)$

Нехай $\varphi(t)$ – стаціонарний гауссівський процес, так що

$$\langle\varphi\rangle = 0, \qquad \langle\varphi\varphi_\tau\rangle = B_0(\tau) = \sigma_0^2 R_0(\tau) = \int_{-\infty}^{\infty} G_0(\omega)\exp(i\omega t)\,d\omega. \qquad (9.3.75)$$

Тоді з урахуванням явного вигляду гауссівського розподілу

$$\langle\exp(i\varphi)\rangle = \exp(-\sigma_0^2/2), \quad \langle\exp[i(\varphi\pm\varphi_\tau)]\rangle = \exp\{-\sigma_0^2[1\pm R(\tau)]\}. \qquad (9.3.76)$$

Звідси випливає, що

$$\langle x(t)\rangle = a_0\exp(-\sigma_0^2/2)\cos\omega_0 t, \qquad (9.3.77)$$

$$\begin{aligned}
\sigma^2 = \langle x^2\rangle &= \frac{1}{4}a_0^2\left\langle\left\{\exp(i\omega_0 t)\exp[i\varphi(t)] + \exp(-i\omega_0 t)\exp[-i\varphi(t)]\right\}^2\right\rangle = \\
&= \frac{1}{4}a_0^2\left\langle\exp(i2\omega_0 t)\exp[i2\varphi(t)] + 2 + \exp(-i2\omega_0 t)\exp[-i2\varphi(t)]\right\rangle = \\
&= \frac{1}{4}a_0^2\left\{2 + \exp(i2\omega_0 t)\langle\exp[i2\varphi(t)]\rangle + \exp(-i2\omega_0 t)\langle\exp[-i2\varphi(t)]\rangle\right\} = \\
&= \frac{1}{2}a_0^2\left[1 + \exp(-2\sigma_0^2)\cos 2\omega_0 t\right].
\end{aligned} \qquad (9.3.78)$$

Нехай

$$\xi(t) = x(t) - \langle x(t)\rangle = \frac{1}{2}a_0\left\{\exp[i\omega_0 t + i\varphi(t)] - \exp(i\omega_0 t)\exp(-\sigma_0^2/2)\right\} + c.c. =$$

$$= \frac{1}{2}a_0\exp(i\omega_0 t)\left[\exp[i\varphi(t)] - \exp(-\sigma_0^2/2)\right] + c.c.$$

– флуктуаційна складова процесу $x(t)$. Тоді

$$B(t,\tau) \equiv \langle\xi(t)\xi(t+\tau)\rangle = \frac{1}{4}a_0^2 \times$$

$$\times\left\langle\left\{\exp(i\omega_0 t)\left[\exp[i\varphi(t)] - \exp(-\sigma_0^2/2)\right] + \exp(-i\omega_0 t)\left[\exp[-i\varphi(t)] - \exp(-\sigma_0^2/2)\right]\right\}\times\right.$$

$$\times\left\{\exp[i\omega_0(t+\tau)]\left[\exp[i\varphi(t+\tau)] - \exp(-\sigma_0^2/2)\right] + \right.$$

$$\left.\left. + \exp[-i\omega_0(t+\tau)]\left[\exp[-i\varphi(t+\tau)] - \exp(-\sigma_0^2/2)\right]\right\}\right\rangle =$$

$$=\frac{1}{4}a_0^2\left\langle\exp\left[i\omega_0\left(2t+\tau\right)\right]\left[\exp\left[i\varphi\left(t\right)\right]-\exp\left(-\sigma_0^2/2\right)\right]\left[\exp\left[i\varphi\left(t+\tau\right)\right]-\exp\left(-\sigma_0^2/2\right)\right]+\right.$$

$$+\exp\left(-i\omega_0\tau\right)\left[\exp\left[i\varphi\left(t\right)\right]-\exp\left(-\sigma_0^2/2\right)\right]\left[\exp\left[-i\varphi\left(t+\tau\right)\right]-\exp\left(-\sigma_0^2/2\right)\right]+$$

$$+\exp\left(i\omega_0\tau\right)\left[\exp\left[-i\varphi\left(t\right)\right]-\exp\left(-\sigma_0^2/2\right)\right]\left[\exp\left[i\varphi\left(t+\tau\right)\right]-\exp\left(-\sigma_0^2/2\right)\right]+$$

$$\left.+\exp\left[-i\omega_0\left(2t+\tau\right)\right]\left[\exp\left[-i\varphi\left(t\right)\right]-\exp\left(-\sigma_0^2/2\right)\right]\left[\exp\left[-i\varphi\left(t+\tau\right)\right]-\exp\left(-\sigma_0^2/2\right)\right]\right\}\Big\rangle= \quad (9.3.79)$$

$$=\frac{1}{2}a_0^2\exp\left(-\sigma_0^2\right)\left\{\exp\left[B_0\left(\tau\right)\right]-1\right\}\left[\cos\omega_0\tau-\cos\left(2\omega_0 t+\omega_0\tau\right)\right].$$

Отже, процес $\xi(t)$ є нестаціонарним.

На практиці звичайно мають справу із середніми за часом величинами. Тоді

$$\bar{B}\left(t,\tau\right)=\frac{1}{2}a_0^2\exp\left(-\sigma_0^2\right)\cos\omega_0\tau\left\{\exp\left[B_0\left(\tau\right)\right]-1\right\}. \quad (9.3.80)$$

Як випливає з формул (9.3.77) та (9.3.80), спектр випадкового процесу $x(t)$ складається з дискретної лінії на частоті-носії ω_0 з інтенсивністю

$$I_{discr}=\frac{1}{2}a_0^2\exp\left(-\sigma_0^2\right) \quad (9.3.81)$$

і неперервного спектру флуктуацій (за умови, що спектр $G_0(\omega)$ сигналу модуляції неперервний)

$$\bar{G}\left(\omega\right)=\frac{1}{8\pi}a_0^2\exp\left(-\sigma_0^2\right)\int_{-\infty}^{\infty}d\tau\left\{\exp\left[B_0\left(\tau\right)\right]-1\right\}\left\{\exp\left[i\left(\omega_0-\omega\right)\tau\right]+\exp\left[-i\left(\omega_0+\omega\right)\tau\right]\right\}. \quad (9.3.82)$$

Видно, що $\bar{G}(\omega)=0$, якщо $B_0(\tau)=0$.

Неперервній частині спектру відповідає інтегральна інтенсивність

$$I_{cont}=\frac{1}{2}a_0^2\left[1-\exp\left(-\sigma_0^2\right)\right], \quad (9.3.83)$$

так що

$$I_{discr}+I_{cont}=\frac{1}{2}a_0^2. \quad (9.3.84)$$

Іншими словами, фазова модуляція впливає на вигляд спектру сигналу, але інтегральна інтенсивність при цьому зберігається.

В загальному випадку форма спектру фазово-модульованого сигналу відрізняється від форми спектру сигналу модуляції (рис. 9.3.14).

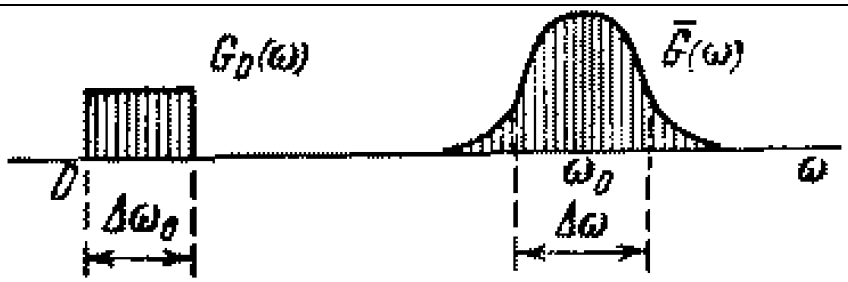


Рис. 9.3.14. Спектр сигналу модуляції $G_0(\omega)$ та неперервна частина спектру відповідного фазово-модульованого сигналу $\bar{G}(\omega)$

Однак, якщо модуляція дуже слабка ($\left|B_0(\tau)\right|\le\sigma_0^2\ll 1$), то, розклавши $\exp\left[B_0(\tau)\right]$ в ряд

Тейлора, можна отримати, що

$$\bar{G}(\omega)=\frac{1}{4}a_0^2\left[G_0(\omega-\omega_0)+G_0(\omega+\omega_0)\right]. \tag{9.3.85}$$

Як бачимо, в цьому разі неперервна частина спектру фазово-модульованого сигналу повторює спектр сигналу модуляції.

В протилежному випадку глибокої фазової модуляції ($\sigma_0^2 \gg 1$) можна вважати, що функція $B_0(\tau)$ швидко спадає зі зростанням $|\tau|$. Оберемо модельну залежність

$$R_0(\tau)\approx 1-\frac{1}{2}(\alpha\tau)^2 \tag{9.3.86}$$

і, відповідно,

$$\exp\left[B_0(\tau)\right]-1\approx\exp\left\{\sigma_0^2\left[1-\frac{1}{2}(\alpha\tau)^2\right]\right\}. \tag{9.3.87}$$

Тоді

$$\bar{G}(\omega)=\frac{a_0^2}{4\alpha\sqrt{2\pi}}\left\{\exp\left[-\frac{(\omega_0-\omega)^2}{2\alpha^2\sigma_0^2}\right]+\exp\left[-\frac{(\omega_0+\omega)^2}{2\alpha^2\sigma_0^2}\right]\right\} \tag{9.3.88}$$

Ширину $\Delta\omega_0$ спектру $G_0(\omega)$ можна оцінити як $\Delta\omega_0\sim\alpha$, ширину $\Delta\omega$ спектру $\bar{G}(\omega)$ – як $\Delta\omega\sim\sigma_0\alpha$. Тому з умови $\sigma_0^2\gg 1$ випливає, що

$$\Delta\omega\gg\Delta\omega_0. \tag{9.3.89}$$

Іншими словами, при глибокій фазовій модуляції спектр модульованого сигналу виходить значно ширшим, ніж спектр сигналу модуляції.

<u>Частотна модуляція шумом</u>. Розглянемо тепер квазігармонічний процес вигляду

$$x(t)=a_0\cos\left[\omega_0 t+\int_0^t dt'\eta(t')\right]. \tag{9.3.90}$$

Як бачимо, процес (9.3.90) відрізняється від фазово-модульованого сигналу (9.3.73)

$$x(t)=a_0\cos\left[\omega_0 t+\varphi(t)\right]$$

тим, що тепер випадковий процес $\eta(t)$ стоїть під знаком інтеграла.

Вважатимемо, що

$$\langle\eta(t)\rangle=0,\quad \langle\eta\eta_\tau\rangle=B_0(\tau)=\sigma_0^2R_0(\tau)=\int_{-\infty}^{\infty}G_0(\omega)\exp(i\omega t)\,d\omega. \tag{9.3.91}$$

Миттєва частота процесу $x(t)$:

$$\omega(t)=\frac{d}{dt}\left[\omega_0 t+\int_0^t dt'\eta(t')\right]=\omega_0+\eta(t). \tag{9.3.92}$$

Зсув фази цього сигналу

$$\varphi(t)=\int_0^t dt'\eta(t') \tag{9.3.93}$$

є, очевидно, дифузійним випадковим процесом (див. вище п. 9.3.6). Його дисперсія, відповідно до формули (9.3.57), може бути подана у формі

$$\langle\varphi^2(t)\rangle=\sigma^2(t)=2\int_0^t(t-\theta)B_0(\theta)\,d\theta. \tag{9.3.94}$$

В цьому випадку в усталеному режимі (при $t\to\infty$) дисперсія, як уже зазначалося, зростає з часом лінійно:

$$\sigma^2(t) = 2\pi G_0(0)t\,. \tag{9.3.95}$$

Звідси випливає, що незалежно від статистики частотної модуляції можна вважати, що розподіл $w(\varphi)$, приведений до інтервалу $[-\pi,\pi]$, з часом стає рівномірним:

$$\lim_{t\to\infty} w(\varphi) = \frac{1}{2\pi}\,. \tag{9.3.96}$$

Таким чином, в усталеному режимі частотно-модульоване коливання $x(t)$ є стаціонарним випадковим процесом з розподілом

$$w(x) = \begin{cases} \dfrac{1}{\pi\sqrt{a_0^2 - x^2}}, & |x| \le a_0; \\ 0, & |x| > a_0 \end{cases} \tag{9.3.97}$$

(рис. 9.3.15).

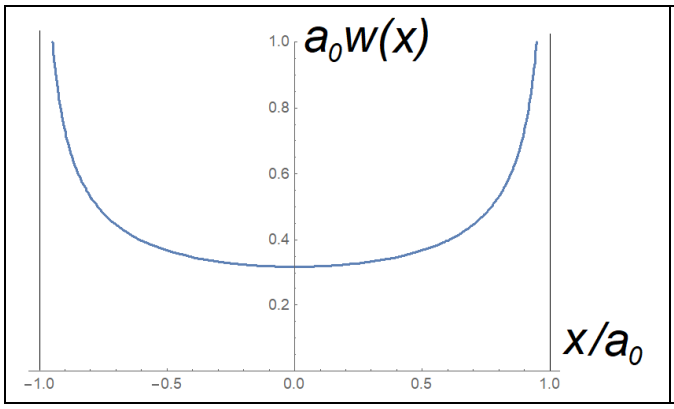


Рис. 9.3.15. Нормований розподіл $w(x)$ частотно-модульованого коливання $x(t)$ в усталеному режимі

Зазначимо, що в спектрі частотно-модульованого сигналу $x(t)$ немає дискретної лінії на частоті ω_0.

Нехай випадковий процес $\eta(t)$ є гауссівським. У такому разі фазові флуктуації $\varphi(t)$ також будуть гауссівськими, і

$$\langle \exp(i\varphi)\rangle = \langle \exp(i\varphi + i\varphi_\tau)\rangle = 0\,, \tag{9.3.98}$$

$$\langle \exp(i\varphi - i\varphi_\tau)\rangle = \left\langle \exp\left[i\int_0^\tau \eta(\theta + t)d\theta \right]\right\rangle = \exp\left[\sigma^2(\tau)/2\right], \tag{9.3.99}$$

де $\sigma^2(\tau) = \sigma^2(t=\tau)$ задається формулою (9.3.94).

Введемо позначення

$$\psi(\tau) = \frac{\sigma^2(\tau)}{2\sigma_0^2} = |\tau|\int_0^\tau R_0(\theta)d\theta - \int_0^{|\tau|} \theta R_0(\theta)d\theta\,. \tag{9.3.100}$$

Тоді можна записати, що

$$B(\tau) = \langle x x_\tau\rangle = \frac{1}{2}a_0^2 \exp\left[-\sigma_0^2\psi(\tau)\right]\cos\omega_0\tau\,. \tag{9.3.101}$$

Функція $\psi(\tau)$ при малих τ зростає за квадратичним законом, при великих – за лінійним:

$$\psi(\tau) = \begin{cases} \dfrac{1}{2}\tau^2, & \tau \ll \tau_0 \\ \dfrac{\pi G_0(0)}{\sigma_0^2}(|\tau| - \tau_0), & \tau \gg \tau_0 \end{cases}\,. \tag{9.3.102}$$

Вважатимемо модуляцію глибокою, так що $\sigma_0^2\tau_0^2 \gg 1$. Тоді функція $B(\tau)$ встигає зменшитися майже до нуля на ділянці $\tau \ll \tau_0$. У цьому випадку можна прийняти, що

$$B(\tau) = \langle xx_\tau \rangle = \frac{1}{2}a_0^2 \exp\left(-\frac{1}{2}\sigma_0^2\tau^2\right)\cos\omega_0\tau . \qquad (9.3.103)$$

Відповідна спектральна інтенсивність має вигляд

$$G(\omega) = \frac{a_0^2}{4\sqrt{2\pi}\sigma_0}\left\{\exp\left[-\frac{(\omega_0-\omega)^2}{2\sigma_0^2}\right] + \exp\left[-\frac{(\omega_0+\omega)^2}{2\sigma_0^2}\right]\right\}. \qquad (9.3.104)$$

Ширину цього спектру можна оцінити як $\Delta\omega \sim \sigma_0$. Можна показати, що $\Delta\omega \gg \Delta\omega_0$, де $\Delta\omega_0$ – ширина спектру сигналу модуляції.

Умову $\sigma_0^2\tau_0^2 \gg 1$ можна переписати у формі

$$G_0(0) \gg \Delta\omega_0 , \qquad (9.3.105)$$

тобто спектр $G_0(\omega)$ вищий і вужчий від спектру $G(\omega)$ (рис. 9.3.16).

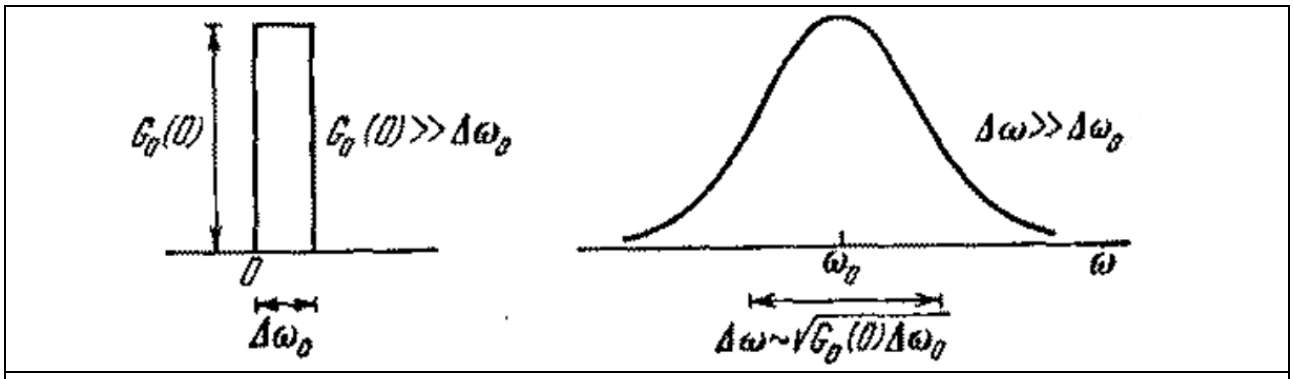


Рис. 9.3.16. Спектри сигналу модуляції $G_0(\omega)$ та частотно-модульованого сигналу $G(\omega)$ у випадку глибокої модуляції

Контрольні питання до підрозділу 9.3

1. Що таке ергодичний випадковий процес?
2. Як за однією реалізацією стаціонарного ергодичного процесу можна визначити густину розподілу ймовірності?
3. Чому висота графіка гауссівської функції розподілу зменшується зі зростанням його ширини?
4. Що таке вузькосмуговий випадковий процес? Як його можна реалізувати?
5. Коли амплітуда (обвідна) вузькосмугового випадкового процесу описується розподілом Рейлі?
6. Що таке білий шум? Який вигляд має його спектр?
7. Коли реальний випадковий процес можна описувати моделлю білого шуму?
8. Що таке випадкова імпульсна послідовність?
9. Яку випадкову величину, що характеризує випадкову імпульсну послідовність, описує розподіл Пуассона?

Задачі до підрозділу 9.3

9.3.1. Для стаціонарного гауссівського випадкового процесу знайти автокореляційну функцію шостого порядку вигляду $B_{111222} = \langle x^3 x_\tau^3 \rangle$.

9.3.2. Якому значенню амплітуди (обвідної) відповідає максимум розподілу Рейлі?

9.3.3. Випадкова імпульсна послідовність $x(t)$ складається з прямокутних імпульсів з амплітудою X_m та тривалістю τ_p. Кількість таких імпульсів на проміжку спостереження T визначається розподілом Пуассона. Знайти статистичне середнє $\langle x \rangle$, автокореляційну функцію $\langle x x_\tau \rangle$ та спектральну інтенсивність $G(\omega)$.

9.4. Стохастичні диференціальні рівняння

Як відомо, для опису динамічних процесів, які протікають у часі, використовують диференціальні рівняння в повних похідних, доповнені початковими умовами. Аналогічно для опису випадкових процесів також використовують диференціальні рівняння в повних похідних з початковими умовами, але при цьому деякі величини, через які виражається розв'язок, є випадковими. В залежності від того, яка саме величина є випадковою, розрізняють три типи стохастичних диференціальних рівнянь:

- рівняння з випадковими початковими умовами;
- рівняння з випадковими зовнішніми силами;
- рівняння з випадковими коефіцієнтами.

9.4.1. Розв'язання стохастичних диференціальних рівнянь шляхом усереднення точних розв'язків

Відомі два підходи до розв'язання стохастичних диференціальних рівнянь:

- метод усереднення точного розв'язку;
- стохастичні методи.

Проілюструємо метод усереднення точного розв'язку на прикладі лінійного диференціального рівняння з коефіцієнтом, залежним від часу, та зовнішньою силою,

$$\frac{dx}{dt}+a(t)x=f(t) \tag{9.4.1}$$

і початковою умовою

$$x(t=0)=X_0. \tag{9.4.2}$$

Загальний розв'язок рівняння (9.4.1) можна знайти методом варіації довільної сталої. Відповідне цьому рівнянню однорідне рівняння може бути подане у вигляді

$$\frac{dx}{x}=-a(t)dt \tag{9.4.3}$$

звідки

$$\ln x=-\int_0^t a(t')dt'+C_1; \qquad x(t)=C\exp\left[-\int_0^t a(t')dt'\right] \tag{9.4.4}$$

Тепер скористаємося методом варіації довільної сталої. Підставимо до (9.4.1) розв'язок

$$\begin{gathered}
x(t)=C(t)\exp\left[-\int_0^t a(t')dt'\right]\\
\frac{dx}{dt}=\frac{dC}{dt}\exp\left[-\int_0^t a(t')dt'\right]-C(t)a(t)\exp\left[-\int_0^t a(t')dt'\right]\\
\frac{dC}{dt}\exp\left[-\int_0^t a(t')dt'\right]-C(t)a(t)\exp\left[-\int_0^t a(t')dt'\right]+a(t)C(t)\exp\left[-\int_0^t a(t')dt'\right]=f(t)\\
\frac{dC}{dt}=f(t)\exp\left[\int_0^t a(t')dt'\right]\\
C(t)=C_0+\int_0^t dt'f(t')\exp\left[\int_0^{t'} a(t'')dt''\right]\\
x(t)=\left\{C_0+\int_0^t dt'f(t')\exp\left[\int_0^{t'} a(t'')dt''\right]\right\}\exp\left[-\int_0^t a(t')dt'\right]
\end{gathered} \tag{9.4.5}$$

Врахувавши початкову умову (9.4.2), маємо:

$$x(t)=\left\{X_0+\int_0^t dt' f(t')\exp\left[\int_0^{t'} a(t'')dt''\right]\right\}\exp\left[-\int_0^t a(t')dt'\right]. \qquad (9.4.6)$$

Розглянемо спершу рівняння з випадковими початковими умовами. Тоді середнє для процесу $x(t)$ має вигляд:

$$\langle x(t)\rangle=\left\{\langle X_0\rangle+\int_0^t dt' f(t')\exp\left[\int_0^{t'} a(t'')dt''\right]\right\}\exp\left[-\int_0^t a(t')dt'\right]. \qquad (9.4.7)$$

Так само просто можна записати середнє для процесу з випадковою зовнішньою силою:

$$\langle x(t)\rangle=\left\{X_0+\int_0^t dt' \langle f(t')\rangle\exp\left[\int_0^{t'} a(t'')dt''\right]\right\}\exp\left[-\int_0^t a(t')dt'\right]. \qquad (9.4.8)$$

Слід зазначити, що в обох випадках випадковий процес $x(t)$ буде нестаціонарним. Відносна простота процедури усереднення в даному разі зумовлена лінійністю доданків розв'язку (9.4.6) як щодо X_0, так і щодо $f(t)$.

9.4.2. Приклад: струм в RL-ланцюжку при надходженні на його вхід випадкового стрибка напруги

Простим прикладом кола, описуваного рівнянням типу (9.4.1), може бути *RL*-ланцюжок (рис. 9.4.1 а), на вхід якого в момент часу t_0 надходить стрибок напруги у вигляді функції Хевісайда з випадковою амплітудою U_m (рис. 9.4.1 б):

$$L\frac{di}{dt}+Ri=U_m\theta(t-t_0). \qquad (9.4.10)$$

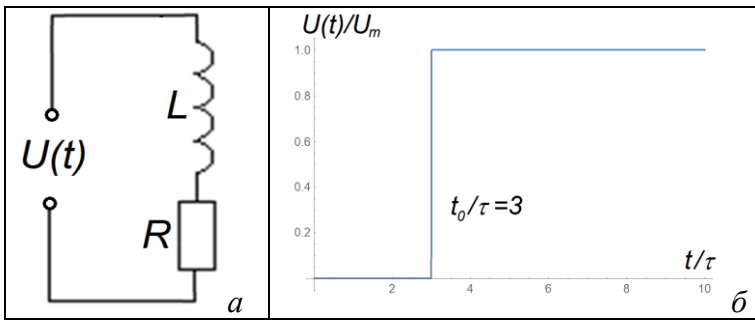


Рис. 9.4.1. Схема RL-ланцюжка (а) та часова залежність напруги, що подається на його вхід (б)

Початкові умови вважаємо нульовими. Порівнюючи з формулами (9.4.1)-(9.4.2), маємо:

$$a=\frac{R}{L}=\frac{1}{\tau},\qquad f(t)=\frac{U_m}{R}\theta(t-t_0),\qquad X_0=0,$$

$$i(t)=\frac{U_m}{L}\exp(-t/\tau)\int_0^t \theta(t'-t_0)\exp(t'/\tau)dt'=$$

$$=\frac{U_m}{L}\begin{cases}0, & t_0\ge t;\\ \exp(-t/\tau)\tau\left[\exp(t/\tau)-\exp(t_0/\tau)\right], & t_0<t\end{cases}=$$

$$=\frac{U_m}{R}\begin{cases}0, & t_0\ge t;\\ 1-\exp\left[(t_0-t)/\tau\right], & t_0<t\end{cases} \qquad (9.4.11)$$

(рис. 9.4.2)

Відповідно

$$\langle i(t)\rangle=\frac{\langle U_m\rangle}{R}\begin{cases}0, & t_0\ge t;\\ 1-\exp\left[(t-t_0)/\tau\right], & t_0<t.\end{cases} \qquad (9.4.12)$$

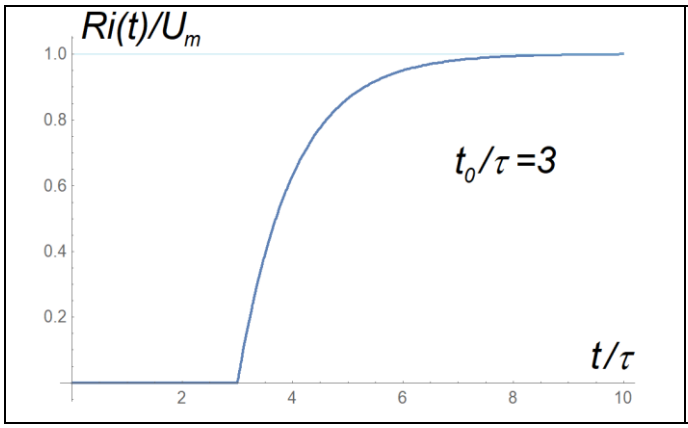


Рис. 9.4.2. Часова залежність струму в RL-ланцюжку при поданні на вхід стрибка напруги

Тепер припустимо, що випадковою величиною є не амплітуда U_m, а момент часу t_0. Нехай він характеризується розподілом

$$w(t_0)=\begin{cases}1/T, t_0\in[0,T];\\ 0, \quad t_0\notin[0,T].\end{cases} \tag{9.4.13}$$

Тоді

$$\langle i(t)\rangle=\int_{-\infty}^{\infty} i(t)w(t_0)dt_0=\frac{U_m}{R}\int_0^T\frac{1}{T}dt_0\begin{cases}0, & t_0\ge t;\\ 1-\exp\left[(t_0-t)/\tau\right], & t_0<t\end{cases}=$$

$$=\frac{U_m}{RT}\begin{cases}\int_0^t dt_0\left\{1-\exp\left[(t_0-t)/\tau\right]\right\}, t\le T;\\ \int_0^T dt_0\left\{1-\exp\left[(t_0-t)/\tau\right]\right\}, t>T\end{cases}=\frac{U_m}{RT}\begin{cases}t-\tau\left[\exp(t/\tau)-1\right]\exp(-t/\tau), t\le T;\\ T-\tau\left[\exp(T/\tau)-1\right]\exp(-t/\tau), t\le T\end{cases}= \tag{9.4.14}$$

$$=\frac{U_m}{RT}\begin{cases}t-\tau\left[1-\exp(-t/\tau)\right], t\le T;\\ T-\tau\left[\exp(T/\tau)-1\right]\exp(-t/\tau), t>T\end{cases}$$

(рис. 9.4.3). На інтервалі $[t_0,t_0+T]$ середньостатистичний струм із часом зростає, оскільки зростає ймовірність потрапляння стрибка напруги в інтервал $[0,t]$. На проміжку (T,∞) величина $\langle i(t)\rangle$ поступово досягає насичення на рівні $\langle i(t)\rangle R/U_m=1$ (пор. з рис. 9.4.2). Випадковий процес $\langle i(t)\rangle$, очевидно, є нестаціонарним.

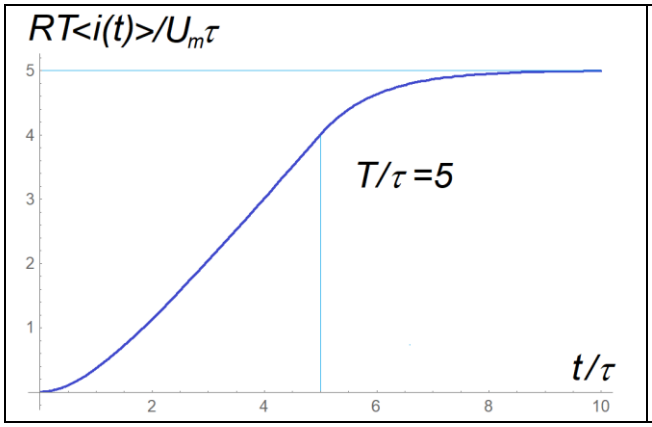


Рис. 9.4.3. Залежність $\langle i(t)\rangle$ в припущенні, що ймовірність стрибка напруги рівномірно розподілена на інтервалі $[0,T]$

9.4.3. Стохастичні методи: отримання рівняння для середніх

В багатьох задачах (особливо коли йдеться про нелінійні диференціальні рівняння)

точні розв'язки, що описують випадковий процес, невідомі. Однак існує ряд методів, які й у цьому випадку дозволяють знаходити статистичні характеристики такого процесу. Їх називають стохастичними методами.

Використовуючи стохастичні методи, вдається із стохастичного диференціального рівняння, не розв'язуючи його, знайти рівняння, які описують статистичні характеристики випадкового процесу, наприклад, моменти або функцію розподілу.

Проілюструємо, як можна отримати рівняння для середнього, на прикладі лінійного диференціального рівняння першого порядку з δ-корельованим випадковим коефіцієнтом

$$\dot{x}+\left[h+\xi(t)\right]x=\varphi(t). \tag{9.4.15}$$

Вважатимемо, що

$$\langle\xi\rangle=0, \qquad \langle\xi\xi_\tau\rangle=2D\delta(\tau). \tag{9.4.16}$$

Усереднивши рівняння (9.4.15) за ансамблем, маємо:

$$\langle\dot{x}\rangle+h\langle x\rangle+\langle\xi(t)x\rangle=\varphi(t). \tag{9.4.17}$$

Тепер потрібно виразити доданок $\langle\xi(t)x\rangle$ через $\langle x\rangle$.

Подамо x у вигляді ряду за ступенями ξ:

$$x=x_0+x_1+x_2+\ldots, \qquad x_n\sim\xi^n. \tag{9.4.18}$$

Підставивши (9.4.18) до (9.4.15), виберемо доданки однакового порядку мализни й отримаємо набір рекурентних рівнянь:

$$\begin{aligned}
&\dot{x}_0+hx_0=\varphi(t);\\
&\dot{x}_1+hx_1=-\xi(t)x_0;\\
&\dot{x}_2+hx_2=-\xi(t)x_1;\\
&\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned} \tag{9.4.19}$$

Скориставшись методом імпульсної функції, можна записати:

$$\begin{aligned}
&x_0(t)=\int_0^\infty\exp(-h\theta)\varphi(t-\theta)\,d\theta;\\
&x_1(t)=-\int_0^\infty\exp(-h\theta)x_0(t-\theta)\xi(t-\theta)\,d\theta=\\
&=-\int_0^\infty\exp(-h\theta)\xi(t-\theta)\,d\theta\int_0^\infty\exp(-h\theta')\varphi(t-\theta-\theta')\,d\theta';\\
&x_2(t)=-\int_0^\infty\exp(-h\theta)x_1(t-\theta)\xi(t-\theta)\,d\theta=\\
&=\int_0^\infty\exp(-h\theta)\xi(t-\theta)\,d\theta\int_0^\infty\exp(-h\theta')\xi(t-\theta-\theta')\,d\theta'\times\\
&\times\int_0^\infty\exp(-h\theta'')\varphi(t-\theta-\theta'-\theta'')\,d\theta'';\\
&\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned} \tag{9.4.20}$$

Очевидно, для доданків непарних порядків $\langle x_{2n+1}\rangle=0$, тобто вони мають лише флуктуаційні складові,

$$x_{2n+1}=\tilde{x}_{2n+1}.$$

Доданки ж парних порядків мають як флуктуаційні, так і регулярні компоненти:

$x_{2n} = \langle x_{2n} \rangle + \tilde{x}_{2n}$.

Зокрема,

$$\tilde{x}_2(t) = \int_0^\infty \exp(-h\theta) d\theta \int_0^\infty \exp(-h\theta') d\theta' \left[\xi(t-\theta)\xi(t-\theta-\theta') - \langle \xi(t-\theta)\xi(t-\theta-\theta') \rangle \right] \times$$
$$\times \int_0^\infty \exp(-h\theta'')\varphi(t-\theta-\theta'-\theta'') d\theta''. \tag{9.4.21}$$

Підрахуємо взаємну кореляційну функцію

$$\left\langle \xi(t) \int_0^t \tilde{x}_1(\theta) d\theta \right\rangle = -\lim_{\varepsilon\to 0} \int_0^{t+\varepsilon} d\theta \int_0^\infty d\theta' \exp(-h\theta') x_0(\theta-\theta') \langle \xi(t)\xi(\theta-\theta') \rangle =$$
$$= -2D \lim_{\varepsilon\to 0} \int_0^{t+\varepsilon} d\theta \int_0^\infty d\theta' \exp(-h\theta') x_0(\theta-\theta') \delta(t-\theta+\theta') \tag{9.4.22}$$

Внутрішній інтеграл відмінний від нуля лише за наявності точки, в якій $t-\theta+\theta'=0$. Оскільки $\theta' \geq 0$, для цього необхідно задовольнити умову $t-\theta \leq 0$, що можливо лише при $t \leq \theta \leq t+\varepsilon$. Отже,

$$\left\langle \xi(t) \int_0^t \tilde{x}_1(\theta) d\theta \right\rangle = -2D \lim_{\varepsilon\to 0} \int_t^{t+\varepsilon} d\theta \int_0^\infty d\theta' \exp(-h\theta') x_0(\theta-\theta') \delta(t-\theta+\theta') =$$
$$= -2D \lim_{\varepsilon\to 0} \int_t^{t+\varepsilon} d\theta \exp\left[-h(\theta-t)\right] x_0(t) = -2Dx_0(t) \lim_{\varepsilon\to 0} \int_t^{t+\varepsilon} d\theta \exp\left[-h(\theta-t)\right] = 0. \tag{9.4.23}$$

Аналогічно можна показати, що для довільного n

$$\left\langle \xi(t) \int_0^t \tilde{x}_n(\theta) d\theta \right\rangle = 0, \tag{9.4.24}$$

звідки випливає, що й

$$\left\langle \xi(t) \int_0^t \tilde{x}(\theta) d\theta \right\rangle = 0. \tag{9.4.25}$$

Аналогічно можна показати, що

$$\left\langle \xi(t) \int_0^t \xi(\theta) \tilde{x}(\theta) d\theta \right\rangle = 0. \tag{9.4.26}$$

Тоді

$$\left\langle \xi(t) \int_0^t \xi(\theta) x(\theta) d\theta \right\rangle = \left\langle \xi(t) \int_0^t \xi(\theta) \left[\langle x(\theta) \rangle + \tilde{x}(\theta) \right] d\theta \right\rangle =$$
$$= \int_0^t \langle x(\theta) \rangle \langle \xi(t)\xi(\theta) \rangle d\theta = \int_0^t 2D \langle x(\theta) \rangle \delta(t-\theta) d\theta = -D \langle x(t) \rangle. \tag{9.4.27}$$

З рівняння (9.4.15) випливає, що

$$x = \int_0^t d\theta \left\{ \varphi(\theta) - \left[h + \xi(\theta) \right] x(\theta) \right\}. \tag{9.4.28}$$

Тоді з урахуванням (9.4.26)-(9.4.27)

$$\langle \xi(t) x(t) \rangle = \left\langle \xi(t) \int_0^t d\theta \left\{ \varphi(\theta) - \left[h + \xi(\theta) \right] x(\theta) \right\} \right\rangle = \langle \xi(t) \rangle \int_0^t d\theta \varphi(\theta) -$$
$$- \int_0^t d\theta h \langle \xi(t) \langle x(\theta) \rangle \rangle - \left\langle \xi(t) \int_0^t d\theta \xi(\theta) x(\theta) \right\rangle = \left\langle \xi(t) \int_0^t d\theta \xi(\theta) x(\theta) \right\rangle = D \langle x(t) \rangle. \tag{9.4.29}$$

З урахуванням (9.4.29) рівняння (9.4.17) набуває вигляду

$$\frac{d}{dt}\langle x\rangle+(h+D)\langle x\rangle=\varphi(t). \tag{9.4.30}$$

Розв'язок цього рівняння легко записати по аналогії з $x_0(t)$ (9.4.20):

$$\langle x(t)\rangle=\int_0^\infty \exp\left[-(h+D)\theta\right]\varphi(t-\theta)\,d\theta. \tag{9.4.31}$$

Таким чином, ми знайшли середнє за ансамблем від розв'язку стохастичного диференціального рівняння (9.4.17), не отримуючи його точного розв'язку.

9.4.3. Стохастичні методи: рівняння Фоккера – Планка

Якщо в звичайному стохастичному диференціальному рівнянні (лінійному чи нелінійному) для випадкового процесу $x(t)$ випадковий коефіцієнт та/або випадкова зовнішня сила є δ-корельованими, то можна записати диференціальне рівняння для густини розподілу ймовірностей $w(x;t)$. Це рівняння відоме як рівняння Фоккера – Планка.

Для ілюстрації візьмемо нелінійне диференціальне рівняння першого порядку з випадковим коефіцієнтом

$$\dot{x}=a(x)+b(x)\xi(t),\qquad \langle\xi\rangle=0,\qquad \langle\xi\xi_\tau\rangle=2D\delta(\tau). \tag{9.4.32}$$

Нехай $\Delta x=x_\tau-x$. Звичайні моменти

$$m_n=\left\langle(\Delta x)^n\right\rangle=\left\langle(x_\tau-x)^n\right\rangle. \tag{9.4.33}$$

враховують випадковість обох величин – x та x_τ. Але, крім них, можна також розглянути умовні моменти

$$m_n^{conv}(x)=\left\langle(\Delta x)^n\right\rangle^{conv}, \tag{9.4.34}$$

коли випадковою вважається лише величина x_τ, а величина x є фіксованою й виступає як параметр.

Перепишемо рівняння (9.4.32) у формі

$$\Delta x(\tau)=\int_0^\tau d\theta\left\{a\left[x(t+\theta)\right]+b\left[x(t+\theta)\right]\xi(t+\theta)\right\}. \tag{9.4.35}$$

Шукатимемо розв'язок (9.4.35) методом послідовних наближень, вважаючи, що

$$x(t+\theta)=x(t)+\Delta x(\theta),\qquad 0<\theta<\tau. \tag{9.4.36}$$

Тоді

$$\Delta x_1(\tau)=\int_0^\tau d\theta\left\{a+b\xi(t+\theta)\right\};$$

$$\Delta x_2(\tau)=\int_0^\tau d\theta\left\{a+a'\Delta x_1(\theta)+\left[b+b'\Delta x_1(\theta)\right]\xi(t+\theta)\right\}; \tag{9.4.37}$$

$$\Delta x_3(\tau)=\int_0^\tau d\theta\left\{a+a'\Delta x_2(\theta)+\frac{1}{2}a''\Delta x_2^2(\theta)+\left[b+b'\Delta x_2(\theta)+\frac{1}{2}b''\Delta x_2^2(\theta)\right]\xi(t+\theta)\right\};$$

……

Знайдемо умовні моменти:

$$\langle\Delta x_1(\tau)\rangle^{conv}=\left\langle\int_0^\tau d\theta\left\{a\left[x(t)\right]+b\left[x(t)\right]\xi(t+\theta)\right\}\right\rangle^{conv}=a\int_0^\tau d\theta=a\tau;$$

$$\left\langle \Delta x_1^2(\tau)\right\rangle^{conv} = \left\langle \int_0^\tau d\theta \left\{ a\left[x(t)\right] + b\left[x(t)\right]\xi(t+\theta)\right\} \int_0^\tau d\theta' \left\{ a\left[x(t)\right] + b\left[x(t)\right]\xi(t+\theta')\right\} \right\rangle^{conv} =$$

$$= \int_0^\tau d\theta \int_0^\tau d\theta' \left\{ a^2\left[x(t)\right] + b^2\left[x(t)\right]\left\langle \xi(t+\theta)\xi(t+\theta')\right\rangle \right\} = a^2\tau^2 + 2Db^2 \int_0^\tau d\theta \int_0^\tau d\theta' \delta(\theta-\theta') =$$

$$= a^2\tau^2 + 2Db^2\tau \approx 2Db^2\tau; \tag{9.4.38}$$

……

(враховано, що $\langle \xi \rangle = 0$, $\langle \xi\xi_\tau \rangle = 2D\delta(\tau)$).

У наступному порядку маємо:

$$\left\langle \Delta x_2(\tau)\right\rangle^{conv} = \left\langle \int_0^\tau d\theta \left\{ a + a'\Delta x_1(\theta) + \left[b + b'\Delta x_1(\theta)\right]\xi(t+\theta)\right\} \right\rangle^{conv} =$$

$$= \left\langle \int_0^\tau d\theta \left\{ a + b\xi(t+\theta) + \left[a' + b'\xi(t+\theta)\right] \int_0^\theta d\theta' \left[a + b\xi(t+\theta')\right] \right\} \right\rangle^{conv} =$$

$$= a\tau + \int_0^\tau d\theta \int_0^\theta d\theta' \left[aa' + bb' \left\langle \xi(t+\theta)\xi(t+\theta')\right\rangle \right] = \tag{9.4.39}$$

$$= a\tau + \frac{1}{2}aa'\tau^2 + 2Dbb' \int_0^\tau d\theta \int_0^\theta d\theta' \delta(\theta-\theta') \approx (a + Dbb')\tau;$$

$$\left\langle \Delta x_2^2(\tau)\right\rangle^{conv} =$$

$$= \left\langle \int_0^\tau d\theta \left\{ a + a'\Delta x_1(\theta) + \left[b + b'\Delta x_1(\theta)\right]\xi(t+\theta)\right\} \int_0^\tau d\theta' \left\{ a + a'\Delta x_1(\theta') + \left[b + b'\Delta x_1(\theta')\right]\xi(t+\theta')\right\} \right\rangle^{conv} =$$

$$= \left\langle \int_0^\tau d\theta \int_0^\tau d\theta' \left\{ a + a'\Delta x_1(\theta) + \left[b + b'\Delta x_1(\theta)\right]\xi(t+\theta)\right\} \left\{ a + a'\Delta x_1(\theta') + \left[b + b'\Delta x_1(\theta')\right]\xi(t+\theta')\right\} \right\rangle =$$

$$= \int_0^\tau d\theta \int_0^\tau d\theta' \left\{ a^2 + a^2 a'(\theta+\theta') + b^2 \left\langle \xi(t+\theta)\xi(t+\theta')\right\rangle + \ldots \right\} \approx 2Db^2\tau. \tag{9.4.40}$$

Враховано, що доданки типу

$$\left\langle \xi(t+\theta)\Delta x_1(\theta)\right\rangle^{conv} = \left\langle \int_0^\tau d\theta' \xi(t+\theta)\left\{ a + b\xi(t+\theta')\right\} \right\rangle^{conv} =$$

$$= b\int_0^\tau d\theta' \left\langle \xi(t+\theta)\xi(t+\theta')\right\rangle = b\int_0^\tau d\theta' \delta(\theta-\theta') = \begin{cases} b, & 0 < \theta < \tau; \\ 0, & \theta < 0 \cup \theta > \tau \end{cases} \tag{9.4.41}$$

у виразі для $\left\langle \Delta x_2^2(\tau)\right\rangle^{conv}$ дають внесок лише другого та вищих порядків за τ.

Таким чином, з точністю до першого порядку за τ маємо:

$$m_1^{conv}(x) = \left[a(x) + b(x)b'(x)D\right]\tau;$$

$$m_2^{conv}(x) = 2Db^2(x)\tau; \tag{9.4.42}$$

$$m_n^{conv}(x) = 0, \quad n \ge 3.$$

Як уже відзначалося (див. вище пп. 9.3.1-9.3.2), двовимірні, одновимірні та умовні ймовірності пов’язані співвідношеннями

$$w(x,t;x_\tau,t+\tau) = w(x_\tau,t+\tau \mid x,t)\,w(x,t). \tag{9.4.43}$$

Тоді

$$w(x_\tau, t+\tau) = \int_{-\infty}^{\infty} w(x,t;x_\tau,t+\tau)\,dx = \int_{-\infty}^{\infty} w(x_\tau, t+\tau \mid x,t)\,w(x,t)\,dx\,. \qquad (9.4.44)$$

Умовний розподіл $w(x_\tau \mid x)$ щодо аргументу x_τ має такі самі властивості, як звичайний розподіл. Зокрема, можна ввести умовну характеристичну функцію

$$\theta^{conv}(u,x) = \int_{-\infty}^{\infty} dx_\tau w(x_\tau, t+\tau \mid x,t)\exp\left[iu(x_\tau - x)\right]. \qquad (9.4.45)$$

Зворотне перетворення Фур'є має вигляд

$$w(x_\tau, t+\tau \mid x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} du\theta^{conv}(u,x)\exp\left[-iu(x_\tau - x)\right] \qquad (9.4.46)$$

Характеристичну функцію можна розкласти в ряд по моментах, розклавши для цього в ряд експоненту під інтегралом у правій частині (9.4.45). Маємо:

$$\theta^{conv}(u,x) = \sum_{k=0}^{\infty}\frac{(iu)^k}{k!}m_k^{conv}(x)\,. \qquad (9.4.47)$$

Підставивши ряд (9.4.47) до правої частини (9.4.46), отримаємо:

$$w(x_\tau, t+\tau \mid x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} du\exp\left[-iu(x_\tau - x)\right]\sum_{k=0}^{\infty}\frac{(iu)^k}{k!}m_k^{conv}(x) =$$

$$= \sum_{k=0}^{\infty}\frac{m_k^{conv}(x)}{k!}\frac{1}{2\pi}\int_{-\infty}^{\infty} du(iu)^k\exp\left[-iu(x_\tau - x)\right] = \qquad (9.4.48)$$

$$= \sum_{k=0}^{\infty}\frac{m_k^{conv}(x)}{k!}\left(-\frac{\partial}{\partial x_\tau}\right)^k\frac{1}{2\pi}\int_{-\infty}^{\infty} du\exp\left[-iu(x_\tau - x)\right] = \sum_{k=0}^{\infty}\frac{m_k^{conv}(x)}{k!}\left(-\frac{\partial}{\partial x_\tau}\right)^k\delta(x_\tau - x).$$

Підставивши (9.4.48) до (9.4.44), отримаємо:

$$w(x_\tau, t+\tau) = \int_{-\infty}^{\infty} w(x_\tau, t+\tau \mid x,t)\,w(x,t)\,dx =$$

$$= \int_{-\infty}^{\infty} w(x,t)\,dx\sum_{k=0}^{\infty}\frac{m_k^{conv}(x)}{k!}\left(-\frac{\partial}{\partial x_\tau}\right)^k\delta(x_\tau - x) =$$

$$= w(x_\tau,t) + \sum_{k=1}^{\infty}\frac{1}{k!}\left(-\frac{\partial}{\partial x_\tau}\right)^k\int_{-\infty}^{\infty} w(x,t)\,dx\,m_k^{conv}(x)\delta(x_\tau - x) = \qquad (9.4.49)$$

$$= w(x_\tau,t) + \sum_{k=1}^{\infty}\frac{1}{k!}\left(-\frac{\partial}{\partial x_\tau}\right)^k\left[w(x_\tau,t)m_k^{conv}(x_\tau)\right].$$

Перепишемо співвідношення (9.4.49) у вигляді

$$\frac{w(x_\tau,t+\tau) - w(x_\tau,t)}{\tau} = \sum_{k=1}^{\infty}\frac{1}{k!}\left(-\frac{\partial}{\partial x_\tau}\right)^k\left[w(x_\tau,t)\frac{m_k^{conv}(x_\tau)}{\tau}\right]. \qquad (9.4.50)$$

Тепер спрямуємо τ до нуля. В лівій частині (9.4.50) маємо:

$$\lim_{\tau\to 0}\frac{w(x_\tau,t+\tau) - w(x_\tau,t)}{\tau} = \frac{\partial w(x_\tau,t)}{\partial t}\,. \qquad (9.4.51).$$

З урахуванням (9.4.42) можна записати, що

$$\lim_{\tau\to 0}\frac{m_1^{conv}(x_\tau)}{\tau} = a(x_\tau) + Db(x_\tau)b'(x_\tau) = K_1(x_\tau);$$

$$\lim_{\tau\to 0}\frac{m_2^{conv}(x_\tau)}{\tau} = 2Db^2(x_\tau) = K_2(x_\tau); \qquad (9.4.52)$$

$$\lim_{\tau\to 0}\frac{m_k^{conv}\left(x_\tau\right)}{\tau}=0,\qquad k\ge 3.$$

Підставивши (9.4.51)-(9.4.52) до (9.4.50) і замінивши x_τ на x, отримуємо:

$$\frac{\partial w}{\partial t}=-\frac{\partial}{\partial x}\left(wK_1\right)+\frac{1}{2}\frac{\partial^2}{\partial x^2}\left(wK_2\right). \tag{9.4.53}$$

Це і є рівняння Фоккера – Планка. Його треба доповнити початковою умовою

$$w\left(x,t=0\right)=w_0\left(x\right). \tag{9.4.54}$$

Рівняння (9.4.53) залишається справедливим і тоді, коли функції $a\left(x\right)$ та $b\left(x\right)$ у рівнянні (9.4.32) залежать також і від часу. Якщо ж така залежність відсутня, рівняння Фоккера – Планка вдається розв'язати аналітично для усталеного режиму (тобто при $t\to\infty$), коли $w\left(x,t\right)=w\left(x\right)$.

У стаціонарному випадку рівняння Фоккера – Планка спрощується й перетворюється на рівняння в повних похідних:

$$\frac{d}{dx}\left(wK_1\right)=\frac{1}{2}\frac{d^2}{dx^2}\left(wK_2\right). \tag{9.4.55}$$

Проінтегруємо рівняння (9.4.55) один раз і подамо його у вигляді:

$$\frac{1}{wK_2}\frac{d}{dx}\left(wK_2\right)=2\frac{K_1}{K_2}.$$

Проінтегрувавши ще раз, отримаємо:

$$w\left(x\right)K_2\left(x\right)=C'\exp\left(2\int_{-\infty}^{x}\frac{K_1\left(x'\right)}{K_2\left(x'\right)}dx'\right),$$

або

$$w\left(x\right)=\frac{C'}{K_2\left(x\right)}\exp\left(2\int\frac{K_1\left(x\right)}{K_2\left(x\right)}dx\right). \tag{9.4.56}$$

Підставивши (9.4.52) до (9.4.56), можна записати $w\left(x\right)$ через коефіцієнти рівняння (9.4.32):

$$\begin{aligned}
w\left(x\right)&=\frac{C}{b^2\left(x\right)}\exp\left[\frac{1}{D}\int\frac{a\left(x\right)+Db\left(x\right)b'\left(x\right)}{b^2\left(x\right)}dx\right]=\\
&=\frac{C}{b^2\left(x\right)}\exp\left[\frac{1}{D}\int\frac{a\left(x\right)}{b^2\left(x\right)}dx+\int\frac{b'\left(x\right)}{b\left(x\right)}dx\right]=\\
&=\frac{C}{b^2\left(x\right)}\exp\left\{\frac{1}{D}\int\frac{a\left(x\right)}{b^2\left(x\right)}dx+\ln\left[b\left(x\right)\right]\right\}=\frac{C}{b\left(x\right)}\exp\left[\frac{1}{D}\int\frac{a\left(x\right)}{b^2\left(x\right)}dx\right].
\end{aligned} \tag{9.4.57}$$

Сталу інтегрування C знаходять з умови нормування.

Питання до підрозділу 9.4

1. Які типи стохастичних диференціальних рівнянь вам відомі?
2. Які стохастичні диференціальні рівняння зручно розв'язувати шляхом усереднення точного розв'язку?
3. Що розуміють під стохастичними методами?
4. В яких випадках зі стохастичного диференціального рівняння можна отримати рівняння для середнього?
5. Що таке рівняння Фоккера – Планка?

6. Коли розв’язок рівняння Фоккера – Планка вдається записати аналітично?

Задачі до підрозділу 9.4

9.4.1. На вхід послідовного коливного контуру в заданий момент часу надходить стрибок напруги з випадковою амплітудою. Знайти середнє за ансамблем значення струму в контурі.

9.4.2. Розв’язати попередню задачу, вважаючи випадковим не амплітуду стрибка напруги, а момент її надходження, що характеризується розподілом

$$w(t_0)=\begin{cases}1/T, t_0\in[0,T];\\ 0, \quad t_0\notin[0,T].\end{cases}$$

9.4.3. Для стохастичного диференціального рівняння вигляду

$\dot{x}=a(x)+b(x)\xi(t)$, $a(x)=a_0+a_1x$, $b(x)=b_0+b_1x^2$, $\langle\xi\rangle=0$, $\langle\xi\xi_\tau\rangle=2D\delta(\tau)$.

записати рівняння Фоккера – Планка та його стаціонарний розв’язок.

Розділ 10. Шуми та завади в радіотехнічних колах

У попередньому розділі ми ознайомилися з математичним апаратом, який використовується для опису та аналізу випадкових процесів. Наступний розділ присвячений аналізу шумів, що існують у радіотехнічних колах.

Спочатку ми розглянемо шуми, які виникають в окремих елементах кіл. Це теплові шуми, пов'язані з активними опорами (зокрема, опорами резисторів), а також інші типи шумів, притаманні активним елементам (біполярним і польовим транзисторам).

Потім ми проаналізуємо вплив цих шумів на роботу автогенераторів: розглянемо питання про форму і ширину спектру автоколивань у стаціонарному режимі, а також задачу про встановлення коливань автогенератора від рівня його власних шумів.

Окремий підрозділ присвячено стохастичним коливанням, які можуть виникати в простих системах (системах з невеликою кількістю ступенів вільності). Для прикладу розглянуто стохастичний режим генератора шуму Кияшка – Піковського – Рабиновича.

Останній підрозділ присвячений питанням електромагнітної сумісності радіоелектронних пристроїв, тобто їхньому взаємному впливу та методам його мінімізації.

10.1. Шуми пасивних та активних елементів радіотехнічних кіл

Як уже зазначалося, шуми (внутрішні шуми) – це випадкові напруги або струми, які завжди виникають у радіотехнічних схемах. Їхньою першопричиною звичайно є тепловий рух заряджених частинок та дискретність електричних зарядів. Як правило, основними джерелами шумів у радіотехнічних колах є резистори та активні елементи (в першу чергу, біполярні транзистори).

У цьому підрозділі ми розглянемо основні типи шумів пасивних та активних елементів: тепловий, дробовий, флікер-шум та інші.

10.1.1. Тепловий шум

Теплові шуми (їх інколи називають шумами Джонсона – Найквіста) виникають унаслідок теплового руху електронів у провідниках. У результаті в колі, що містить резистор (рис. 10.1.1, провідники вважаємо ідеальними), протікає тепловий струм, який носить флуктуаційний характер. На опорі цей струм буде створювати флуктуаційну напругу. Її можна подати як деяке еквівалентне джерело шумової е.р.с. $E_T(t)$, що відповідає тепловому шуму.

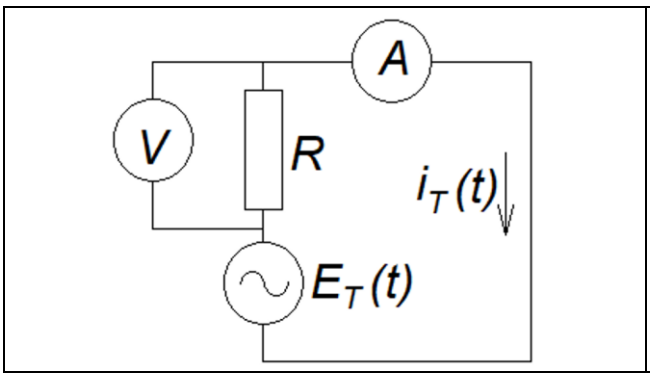


Рис. 10.1.1. До пояснення джерела теплового шуму

Розрахувати спектр потужності теплового шуму можна з термодинамічних міркувань. Для цього розглянемо замкнене коло, складене з двох комплексних опорів $Z_{1,2} = R_{1,2} + iX_{1,2}$ і двох еквівалентних джерел теплової е.р.с. $E_T^{(1,2)}(t)$ і вміщене в термостат, де підтримується стала температура T (рис. 10.1.2 а). Нехай P_{12} і P_{21} – потужності, що виділяються відповідно

джерелом $E_T^{(1)}(t)$ на резисторі R_2 та джерелом $E_T^{(2)}(t)$ на резисторі R_1. Їх можна записати через відповідні спектральні інтенсивності $G_T^{(1,2)}(\omega)$:

$$P_{12} = \int_{-\infty}^{\infty} \frac{G_T^{(1)}(\omega)}{|Z_1 + Z_2|^2} R_2 d\omega, \qquad P_{21} = \int_{-\infty}^{\infty} \frac{G_T^{(2)}(\omega)}{|Z_1 + Z_2|^2} R_1 d\omega. \tag{10.1.1}$$

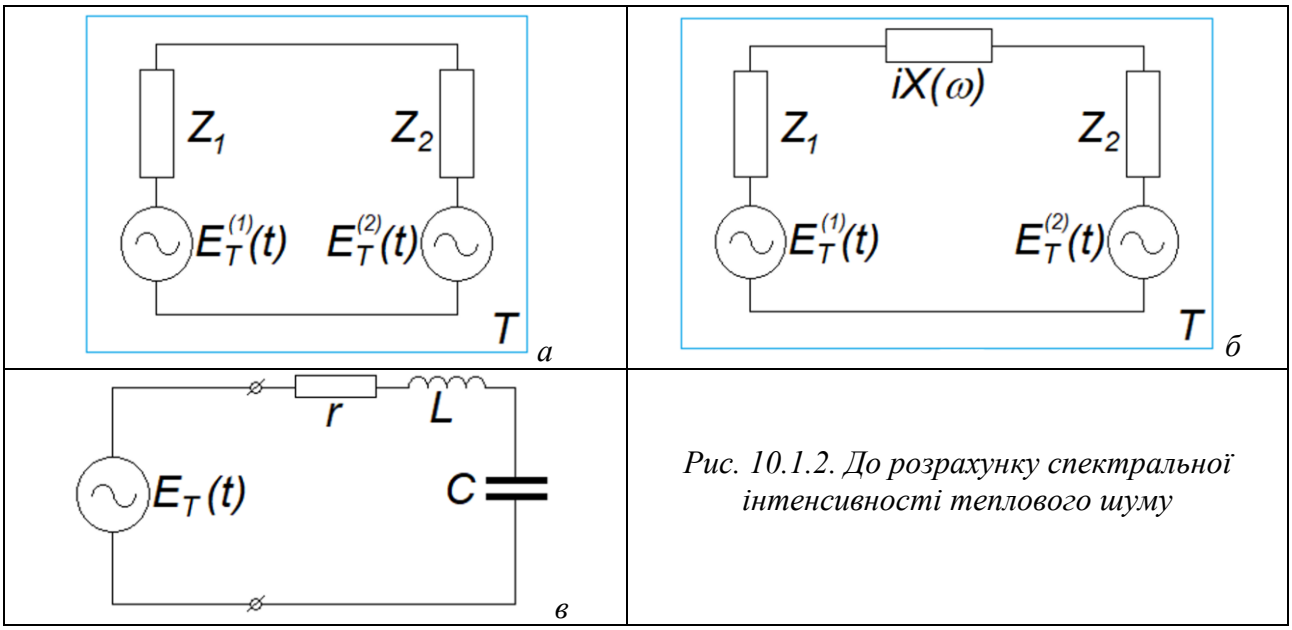


Рис. 10.1.2. До розрахунку спектральної інтенсивності теплового шуму

В силу другого початку термодинаміки ці потужності мають бути однаковими:

$$P_{12} = P_{21}. \tag{10.1.2}$$

Тепер дещо ускладнимо коло, додавши в нього суто реактивний опір $iX(\omega)$, який довільним чином залежить від частоти (рис. 10.1.2 б). Тоді умову (10.1.2) можна переписати у вигляді

$$\int_{-\infty}^{\infty} \frac{G_T^{(1)}(\omega)}{|Z_1 + Z_2 + iX(\omega)|^2} R_2 d\omega = \int_{-\infty}^{\infty} \frac{G_T^{(2)}(\omega)}{|Z_1 + Z_2 + iX(\omega)|^2} R_1 d\omega. \tag{10.1.3}$$

Припустимо, що

$$X(\omega) = \begin{cases} 0, & \omega_0 - \Delta\omega \le \omega \le \omega_0 + \Delta\omega; \\ \infty, & \omega < \omega_0 - \Delta\omega \cup \omega > \omega_0 + \Delta\omega, \end{cases} \tag{10.1.4}$$

причому $\Delta\omega$ настільки мала, що в смузі частот $\omega_0 - \Delta\omega \le \omega \le \omega_0 + \Delta\omega$ спектральні інтенсивності $G_T^{(1)}(\omega)$ можна вважати сталими (рис. 10.1.3). Тоді рівність (10.1.3) набуває вигляду

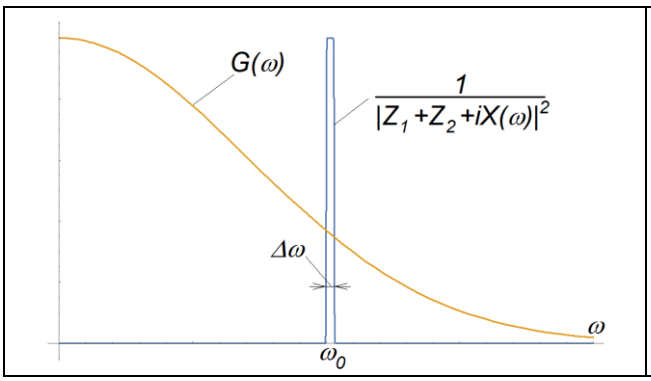


Рис. 10.1.3. Співмножники підінтегрального виразу (10.1.3)для потужності при відповідному підборі залежності $X(\omega)$ (10.1.4)

$$G_T^{(1)}(\omega_0)R_2\int\limits_{\omega_0-\Delta\omega}^{\omega_0+\Delta\omega}\frac{d\omega}{|Z_1+Z_2|^2}=G_T^{(2)}(\omega_0)R_1\int\limits_{\omega_0-\Delta\omega}^{\omega_0+\Delta\omega}\frac{d\omega}{|Z_1+Z_2|^2}, \tag{10.1.5}$$

або

$$G_T^{(1)}(\omega_0)R_2=G_T^{(2)}(\omega_0)R_1. \tag{10.1.6}$$

Оскільки частота ω_0 обиралася довільним чином, то остання нерівність має виконуватися для довільних ω, тобто

$$\frac{G_T^{(1)}(\omega)}{R_1}=\frac{G_T^{(2)}(\omega)}{R_2}=u(\omega,T), \tag{10.1.7}$$

де $u(\omega,T)$ – деяка універсальна функція.

Для знаходження цієї функції розглянемо схему, що включає джерело шумової напруги та послідовний коливний контур (рис. 10.1.2 в). Якщо $q_T(t)$ – заряд на конденсаторі С, то він задовольняє диференціальне рівняння

$$L\frac{d^2q_T}{dt^2}+r\frac{dq_T}{dt}+\frac{1}{C}q_T=E_T(t) \tag{10.1.8}$$

Передавальна функція, що пов'язує спектри струму $i_T(t)$ з е.р.с. $E_T(t)$, має вигляд

$$K_i(\omega)=\frac{1}{i\omega L\left(\omega_0^2-\omega^2+i2\delta\omega\right)},\qquad \omega_0^2=\frac{1}{LC},\qquad \delta=\frac{R}{2L}. \tag{10.1.9}$$

Оскільки $i_T(t)=dq_T/dt$ і, відповідно, $i_T(\omega)=i\omega q_T(\omega)$, то передавальна функція, що пов'язує спектри $q_T(\omega)$ і $E_T(\omega)$, має вигляд

$$K_q(\omega)=\frac{1}{L\left(\omega_0^2-\omega^2+i2\delta\omega\right)}. \tag{10.1.10}$$

Таким чином,

$$q_T(t)=\int\limits_{-\infty}^{\infty}K_q(\omega)E_T(\omega)d\omega, \tag{10.1.11}$$

$$\left\langle q_T^2\right\rangle=\left\langle\int\limits_{-\infty}^{\infty}K_q(\omega)E_T(\omega)d\omega\int\limits_{-\infty}^{\infty}K_q(\omega)E_T(\omega)d\omega\right\rangle=\int\limits_{-\infty}^{\infty}d\omega K_q(\omega)\int\limits_{-\infty}^{\infty}d\omega' K_q(\omega')\left\langle E_T(\omega)E_T(\omega')\right\rangle=$$
$$=\int\limits_{-\infty}^{\infty}d\omega K_q(\omega)\int\limits_{-\infty}^{\infty}d\omega' K_q(\omega')G_T(\omega)\delta(\omega+\omega')=\int\limits_{-\infty}^{\infty}d\omega K_q(\omega)K_q(-\omega)G_T(\omega)=\int\limits_{-\infty}^{\infty}|K_q(\omega)|^2G_T(\omega)d\omega. \tag{10.1.12}$$

Вважатимемо контур високодобротним, так що в межах його смуги пропускання завширшки δ функцію $G_T(\omega)$ можна вважати сталою. Тоді її можна винести з-під знаку інтеграла в (10.1.12) і записати, що

$$\left\langle q_T^2\right\rangle=G_T(\omega_0)\int\limits_{-\infty}^{\infty}|K_q(\omega)|^2d\omega=\frac{\pi G_T(\omega_0)}{2\delta\omega_0^2L^2}. \tag{10.1.13}$$

Аналогічно

$$\left\langle i_T^2\right\rangle=G_T(\omega_0)\int\limits_{-\infty}^{\infty}|K_i(\omega)|^2d\omega=\frac{\pi G_T(\omega_0)}{2\delta L^2}. \tag{10.1.14}$$

В результаті повну енергію в контурі можна подати як

$$\langle W\rangle=\frac{\left\langle q_T^2\right\rangle}{2C}+L\frac{\left\langle i_T^2\right\rangle}{2}=\frac{\pi G_T(\omega_0)}{2R}. \tag{10.1.15}$$

З іншого боку, в класичній області частот, де $k_B T \gg \hbar\omega$, діє принцип рівномірного розподілу енергії по ступенях вільності, тобто

$$\langle W \rangle = N k_B T, \tag{10.1.16}$$

де N – кількість ступенів вільності, k_B – стала Больцмана. Для коливного контуру $N = 1$, звідки

$$\frac{\pi G_T(\omega_0)}{2R} = k_B T, \tag{10.1.17}$$

і, таким чином,

$$G_T(\omega) = \frac{2}{\pi} k_B T R. \tag{10.1.18}$$

Спектральна інтенсивність $G_T(\omega)$ визначена в смузі частот $(-\infty, \infty)$. В експерименті (тобто в смузі частот $(0, \infty)$) вимірюється величина $G_T^+(\omega) = 2G_T(\omega)$. Спектральна інтенсивність $G_T(\omega)$ відповідає інтенсивності випадкових коливань напруги й пропорційна відповідному активному опору R. Тоді виміряна потужність шумів у смузі частот Δf має вигляд

$$P(\Delta f) = \frac{G_T}{R} 2\pi\Delta f = 4 k_B T \Delta f. \tag{10.1.19}$$

Це співвідношення відоме як формула Найквіста.

У біполярних транзисторах основну роль у генерації теплового шуму відіграє база, як область із найнижчим ступенем легування і, відповідно, з найбільшим питомим опором.

10.1.2. Дробовий шум

Дробовий шум (або шум Шотткі) виникає внаслідок дискретності носіїв заряду. В результаті струм являє собою послідовність окремих імпульсів.

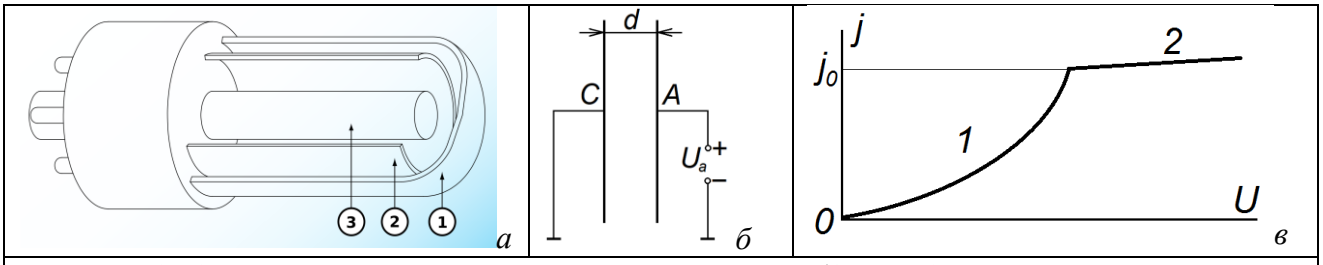


Рис. 10.1.4. Електровакуумний діод: а – конструкція (1 – колба, 2 – анод, 3 – катод; б – модель плоского діода; в – графік вольт-амперної характеристики (1 – область дії закону трьох других; 2 – режим насичення)

Найпростіше розрахувати характеристики дробового шуму для електровакуумного діода в режимі насичення (рис. 10.1.4), коли об'ємний заряд у міжелектродному проміжку відсутній, і рух кожного окремого електрона можна вважати незалежним від інших. Тим не менше отримані результати якісно правильно описують дробовий шум і в напівпровідникових приладах, зокрема, в біполярних транзисторах.

Для опису дробового шуму скористаємося моделлю випадкової імпульсної послідовності (див. вище п. 9.3.5). Для такої моделі спектральна інтенсивність визначається формулою

$$G(\omega) = 2\pi\Omega \left| F(\omega) \right|^2, \tag{10.1.20}$$

де $\Omega = \overline{n}/T$ – середня частота проходження імпульсів (середня кількість імпульсів, що спостерігаються за одиницю часу), $F(t)$ – форма окремого імпульсу, а

$$F(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}F(t)\exp(-i\omega t)\,dt \qquad (10.1.21)$$

– його спектр.

Таким чином, знаходження спектральної інтенсивності дробового шуму вимагає знання форми імпульсів струму, які спостерігаються в результаті руху окремого електрона від катода до анода.

Нехай між катодом і анодом прикладена напруга V, яка породжує електричне поле $E=V/d$ (d – віддаль між електродами). Електрони рухаються вдовж осі z, $z(t)$ – координата електрона в момент часу t, $i(t)$ – струм. Тоді енергія, яка виділяється струмом за малий проміжок часу, має дорівнювати роботі електричного поля на відповідному шляху:

$$i(t)Vdt=eEdz, \qquad (10.1.22)$$

звідки можна знайти миттєвий струм

$$i(t)=\frac{eE}{V}\frac{dz}{dt}=\frac{eEv(t)}{V}=\frac{e}{d}v(t), \qquad (10.1.23)$$

де $v(t)$ – миттєва швидкість електрона. Рух електрона в сталому однорідному електричному полі відбувається з прискоренням $a=eE/m$ протягом пролітного часу тривалістю $\tau=\sqrt{2d/a}=\sqrt{2md/eE}$. Отже,

$$i(t)=\frac{e}{d}v(t)=\begin{cases}e^2Et/md, & 0\le t\le\tau,\\ 0, & t<0\cup t>\tau\end{cases}=\begin{cases}2et/\tau^2, & 0\le t\le\tau,\\ 0, & t<0\cup t>\tau\end{cases} \qquad (10.1.24)$$

(рис. 10.1.5).

Спектр окремого імпульсу:

$$i(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}i(t)\exp(-i\omega t)\,dt=\frac{1}{2\pi}\int_{0}^{\tau}\frac{2et}{\tau^2}\exp(-i\omega t)\,dt=$$

$$=\frac{e}{\pi\tau^2}\int_{0}^{\tau}t\exp(-i\omega t)\,dt=\frac{e}{\pi\theta^2}\int_{0}^{\theta}x\exp(-ix)\,dx, \qquad (10.1.25)$$

$$\theta=\omega\tau, \qquad x=\omega t.$$

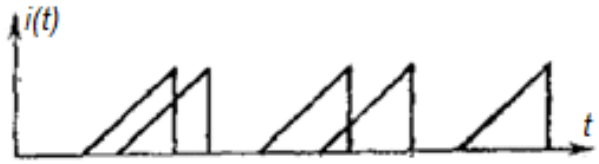


Рис. 10.1.5. Часова залежність струму вакуумного діода в режимі насичення як суперпозиція випадкових трикутних імпульсів, створюваних окремими електронами

Інтегруючи частинами, отримуємо:

$$i(\omega)=\frac{e}{\pi\theta^2}\int_{0}^{\theta}x\exp(-ix)\,dx=\frac{e}{\pi\theta^2}\left[ix\exp(-ix)\Big|_0^{\theta}-i\int_{0}^{\theta}\exp(-ix)\,dx\right]=$$

$$=\frac{e}{\pi\theta^2}\left[i\theta\exp(-i\theta)+\exp(-ix)\Big|_0^{\theta}\right]=\frac{e}{\pi\theta^2}\left[i\theta\exp(-i\theta)+\exp(-i\theta)-1\right]. \qquad (10.1.26)$$

$$|i(\omega)|^2 = \left(\frac{e}{\pi\theta^2}\right)^2 \left[i\theta\exp(-i\theta)+\exp(-i\theta)-1\right]\left[-i\theta\exp(i\theta)+\exp(i\theta)-1\right] =$$

$$= \left(\frac{e}{\pi\theta^2}\right)^2 \left[\theta^2 + i\theta - i\theta\exp(-i\theta) - i\theta + 1 - \exp(-i\theta) + i\theta\exp(i\theta) - \exp(i\theta) + 1\right] =$$

$$\left(\frac{e}{\pi\theta^2}\right)^2 \left\{\theta^2 + 2 + i\theta\left[\exp(i\theta)-\exp(-i\theta)\right] - \left[\exp(i\theta)+\exp(-i\theta)\right]\right\} = \tag{10.1.27}$$

$$= \left(\frac{e}{\pi\theta^2}\right)^2 \left(\theta^2 + 2 - 2\theta\sin\theta - 2\cos\theta\right).$$

При $\theta << 1$

$$\sin\theta \approx \theta - \frac{1}{6}\theta^3, \qquad \cos\theta \approx 1 - \frac{1}{2}\theta^2 + \frac{1}{24}\theta^4. \tag{10.1.28}$$

$$|i(\omega)|^2 \approx \left(\frac{e}{\pi\theta^2}\right)^2 \left[\theta^2 + 2 - 2\theta\left(\theta - \frac{1}{6}\theta^3\right) - 2\left(1 - \frac{1}{2}\theta^2 + \frac{1}{24}\theta^4\right)\right] =$$

$$= \left(\frac{e}{\pi\theta^2}\right)^2 \theta^4 \left(\frac{1}{3} - \frac{1}{12}\right) = \frac{\theta^4}{4}\left(\frac{e}{\pi\theta^2}\right)^2 = \left(\frac{e}{2\pi}\right)^2. \tag{10.1.29}$$

Маємо:

$$G(\omega) = 2\pi\Omega|i(\omega)|^2 = 2\pi\frac{I_0}{e}\left(\frac{e}{2\pi}\right)^2 = \frac{1}{2\pi}eI_0, \tag{10.1.30}$$

де I_0 – середній струм через діод (враховано, що $\Omega = I_0/e$). Ця формула відома в літературі як формула Шотткі. Видно, що в цьому наближенні дробовий шум має рівномірний спектр. Спадання функції $G(\omega)$ спостерігається при $\theta \approx 1$, або при $\omega \approx 1/\tau$ (рис. 10.1.6).

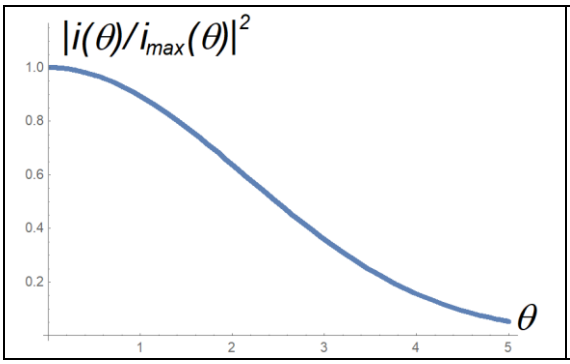


Рис. 10.1.6. Графік залежності нормованої спектральної інтенсивності дробового шуму від частоти ($\theta = \omega\tau$)

Взаємний вплив руху електронів (наприклад, у біполярному транзисторі) приводить до деякого зменшення спектральної інтенсивності дробового шуму. Однак формулою Шотткі користуються і в цьому випадку, ввівши до неї поправочний коефіцієнт Γ – так званий коефіцієнт депресії:

$$G(\omega) = \frac{\Gamma}{2\pi}eI_0. \tag{10.1.31}$$

Розрахунок коефіцієнту депресії – це окрема складна задача.

10.1.3. Флікер-шум

Флікер-шум (шум *1/f*) притаманний широкому класу процесів. В електронних приладах він зумовлений випадковими повільними змінами емісії з катоду (в тріоді) або випадковими повільними змінами провідності напівпровідника (в біполярному транзисторі). Природа

цих процесів не до кінця зрозуміла. Характерною особливістю флікер-шуму є спектр із особливістю в нулі:

$$G(\omega)=\begin{cases} AI_0^n\left(\dfrac{\omega_0}{\omega}\right)^{\mu}, & \omega\le\omega_0; \\ 0, & \omega>\omega_0, \end{cases} \qquad 0.6<\mu<2, \qquad n\approx 2, \tag{10.1.32}$$

n – показник струмової залежності (рис. 10.1.6).

Ми говорили, що для ергодичних процесів усереднення за часом еквівалентне дії фільтру нижніх частот (див. вище п. 9.3.1). Відповідно до формули (9.3.7)

$$\sigma_T^2\equiv\left\langle\left(\bar{\xi}_T\right)^2\right\rangle=\int_{-\infty}^{\infty}G(\omega)\left[\frac{\sin(\omega T/2)}{\omega T/2}\right]^2 d\omega .$$

Але флікер-шум має максимум спектральної інтенсивності на нульовій частоті, тому такі шуми погано відфільтровуються при усередненні за часом – рис. 10.1.7.

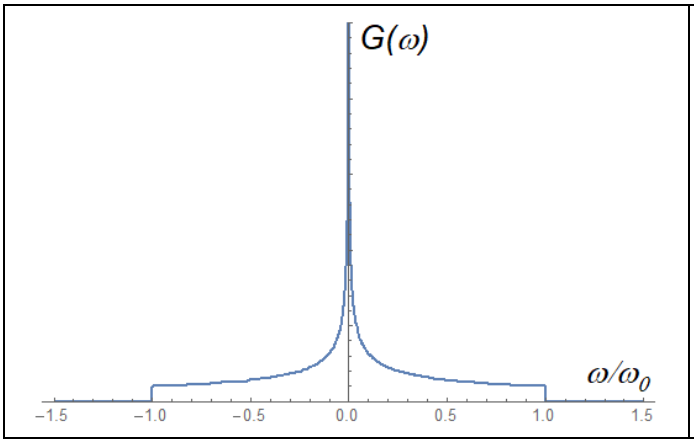


Рис. 10.1.6. Графік модельного розподілу спектральної інтенсивності флікер-шуму

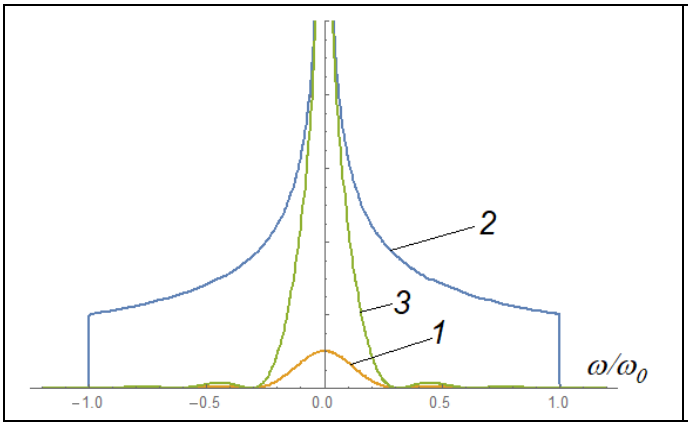


Рис. 10.1.7. До усереднення флікер-шуму за часом: крива 1 – передавальна функція, що відповідає усередненню за часом; 2, 3 – спектральна інтенсивність флікер-шуму відповідно до й після усереднення

Властивості флікер-шуму мало залежать від температури.

Однією з різновидів флікер-шуму в напівпровідникових приладах є так званий вибуховий шум. Джерело вибухового шуму поки що не до кінця зрозуміле, але вважається, що він пов'язаний з наявністю тонких, сильно легованих емітерних переходів і захопленням там носіїв заряду дислокаціями. Дислокація – це дефект кристалічної ґратки, який супроводжується згущенням чи розрідженням у розташуванні атомів (рис. 10.1.8). Дислокація в напівпровіднику може служити пасткою для носіїв заряду.

Поява і зникнення імпульсів вибухового шуму пов'язується з однією пасткою в області просторового заряду. Найбільш правдоподібною теорією вибухового шуму слід вважати дислокаційну теорію, яка перебуває у добрій згоді з експериментом.

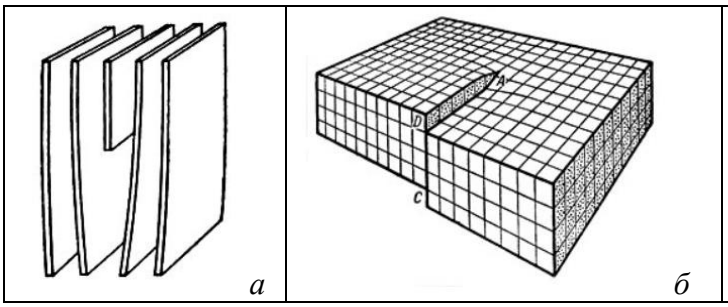


Рис. 10.1.8. Крайова (а) та гвинтова (б) дислокації кристалічної ґратки

10.1.4. Шуми розподілу струмів

У приладах з кількома електродами (тріод та багатосіткові лампи, біполярний транзистор) розподіл носіїв струмів між електродами (катодного струму між сіткою і анодом, емітерного струму між базою й колектором) носить випадковий характер. Через це виникає додатковий вид шумів – шуми розподілу струмів. Їхня спектральна інтенсивність для кожної з двох компонент, на які розгалужується вхідний струм I_0, має вигляд

$$G_{id} = 2\lambda(1-\lambda)eI_0\left|i(\omega)\right|^2, \qquad (10.1.33)$$

де λI_0 та $(1-\lambda)I_0$ – струми вихідних електродів, $i(\omega)$ – спектр струму, створеного одним носієм заряду. Ці шуми будуть максимальними, якщо вхідний струм ділиться навпіл між вихідними електродами, тобто $\lambda = 0.5$ (рис. 10.1.9), чого немає ні в тріоді, ні в біполярному транзисторі.

Спектр частот шумів розподілу струмів простягається до частот (0.1-1) ГГц.

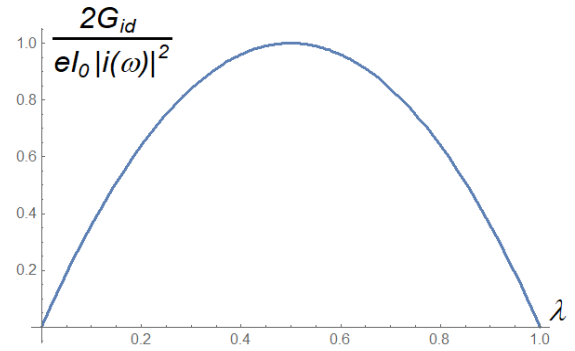


Рис. 10.1.9. Графік залежності спектральної інтенсивності шумів розподілу струмів від частки вхідного струму, що відгалужується на перший вихід

10.1.5. Шуми генерації-рекомбінації

Як відомо, в напівпровіднику носії заряду виникають в результаті генерація пар електрон – дірка. Можливий і зворотний процес – при зустрічі електрона й дірки вони можуть рекомбінувати (зникнути). Ці процеси носять випадковий характер. Зокрема, швидкості генерації та рекомбінації флуктуюють. В результаті виникає такий вид шуму, притаманний напівпровідниковим приладам, як шуми генерації – рекомбінації. В біполярному транзисторі рекомбінація відбувається переважно в базі, куди неосновні (для цієї області) носії інжектуються з емітера.

Спектральна інтенсивність шумів генерації-рекомбінації може бути подана у вигляді:

$$G_{gr}(\omega) = \frac{4I_0^2\left\langle \Delta N^2 \right\rangle}{N_0^2}\frac{\tau}{1+\omega^2\tau^2}, \qquad (10.1.34)$$

де I_0 – струм через зразок, N_0 – середня рівноважна кількість носіїв у ньому, $\Delta N = N - N_0$, τ – середній час життя носіїв у напівпровіднику (рис. 10.1.10).

З рис. 10.1.10 видно, що спектральна інтенсивність шумів генерації-рекомбінації відмінна від нуля до частот порядку кількох кГц.

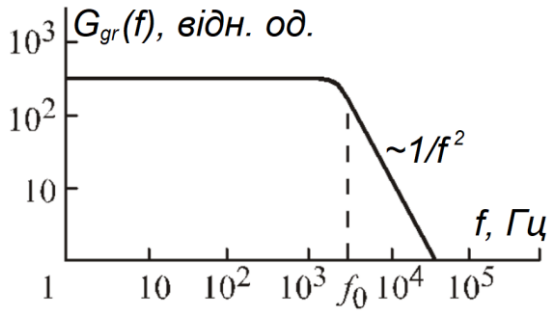


Рис. 10.1.10. Графік спектральної інтенсивності шумів генерації-рекомбінації (спектр Лоренца – Дебая)

10.1.6. Дифузійний шум

Електронний газ у напівпровіднику може стати нерівноважним. Для цього, наприклад, можна накласти сильне електричне поле. Нерівноважність виникає також при інжекції великої кількості неосновних носіїв з емітера в базу. При цьому розподіл за енергіями відрізняється від максвеллівського. В цьому випадку формула Найквіста виявляється несправедливою, і виникає деякий додатковий шум, який називають дифузійним. Його виникнення пов'язане з випадковим характером обміну енергією між гарячими електронами та кристалічною ґраткою та з флуктуаціями потужності від зовнішнього джерела, яка поглинається напівпровідником.

У випадку, коли носії мають максвеллівський розподіл, дифузійний шум зводиться до теплового.

10.1.7. Шуми напівпровідникових приладів у різних частотних діапазонах

На завершення цього розділу обговоримо, чим визначаються шуми напівпровідникових приладів (зокрема й біполярних транзисторів) у різних частотних діапазонах (рис. 10.1.11).

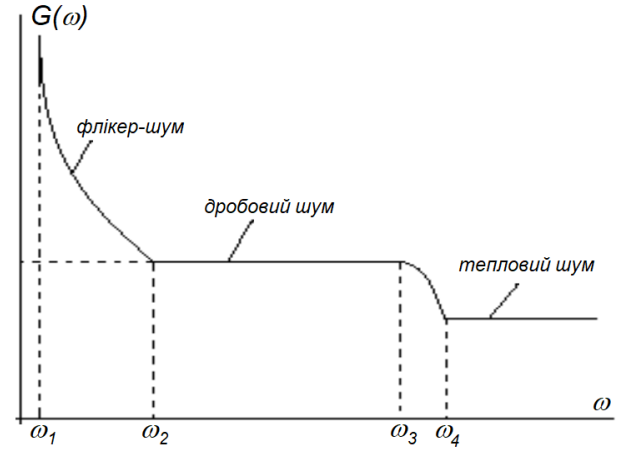


Рис. 10.1.11. Частотна залежність спектральної інтенсивності шумів напівпровідникових приладів у різних частотних діапазонах

На низьких частотах переважає флікер-шум. Для напівпровідникових приладів $f_1 = \omega_1/2\pi \approx 10^{-7}\,Гц$, $f_2 = \omega_2/2\pi \approx 5кГц$.

На середніх частотах основну роль відіграє дробовий шум. Верхню граничну частоту для нього в біполярному транзисторі можна оцінити як $f_3 = \omega_3/2\pi \sim 1/\tau$, де τ – час руху неосновних носіїв через базу.

На ще вищих частотах основну роль відіграє тепловий шум. Як уже відзначалося, це переважно тепловий шум бази.

У польових транзисторах з керувальним *p-n* переходом основну роль відіграє тепловий шум, зумовлений активним опором каналу, а також дробовий шум затвору, що залежить

від його сталого зворотного струму. На низьких частотах суттєвим стають шуми генерації-рекомбінації.

У польових транзисторах з ізольованим затвором на середніх частотах основну роль відіграють теплові шуми різного типу (зокрема, тепловий шум каналу). На низьких частотах основну роль відіграє флікер-шум.

Загалом польові транзистори мають нижчий рівень власних шумів порівняно з біполярними.

Контрольні питання до підрозділу 10.1

1. Опишіть природу теплового шуму.
2. Який вигляд має спектральна інтенсивність теплового шуму? До яких частот вона простягається?
3. Опишіть механізм виникнення дробового шуму. Де він спостерігається?
4. В чому полягають специфічні особливості флікер-шуму?
5. Опишіть природу шуму розподілу струмів. Де він спостерігається?
6. Опишіть механізм виникнення та вигляд спектру шуму генерації-рекомбінації.

Задачі до підрозділу 10.1

10.1.1. Знайдіть середню амплітуду (корінь з дисперсії) напруги теплового шуму на резисторі 1 кОм у смузі пропускання 10 МГц при кімнатній температурі.

10.1.2. Колекторний струм біполярного транзистора – 20 мА. Знайдіть потужність дробового шуму в смузі частот від 0 до 10 МГц. Вважати, що коефіцієнт депресії дорівнює одиниці.

10.2. Флуктуації в автогенераторах

Ми розглянули шуми окремих елементів радіоелектронних схем – резисторів та транзисторів. Тепер подивимося, що являють собою шуми схем у цілому.

З багатьох точок зору найбільший інтерес у цьому плані являють автогенератори. Зокрема, автогенератор є необхідним елементом схеми будь-якого радіопередавача та супергетеродинного приймача (див. вступ). Шуми сигналу, що відповідає частоті-носію, зберігаються при його подальшій обробці (зокрема, при модуляції такого сигналу). Флуктуації частоти-носія обмежують знизу можливі частоти модуляції.

В цьому розділі ми розглянемо дві задачі, що стосуються автогенераторів. Перша з них – це визначення форми та ширини спектру генерованого сигналу. Друга – встановлення коливань генератора, що зростає від рівня флуктуацій.

10.2.1. Технічна та природна ширина лінії автогенератора

В реальних автогенераторах радіодіапазону завжди присутні низькочастотні флуктуації на частотах до (0.1-1) кГц, зумовлені, зокрема, флікер-шумом та іншими випадковими процесами. Вплив такого шуму на автогенератор призводить до того, що сигнал останнього набуває вигляду вузькосмугового випадкового процесу вигляду

$$x(t) = \rho(t)\cos\left[\omega_0 t + \xi(t)\right]. \tag{10.2.1}$$

Амплітуда флуктуює навколо деякого середнього значення $\langle\rho\rangle$. Флуктуації амплітуди і фази в загальному випадку пов'язані між собою (не є статистично незалежними).

За відомими $\rho(t)$ та $\xi(t)$ можна знайти форму спектральної лінії. В багатьох випадках адекватною є модель коливань із фазою, модульованою шумом (див. вище п. 9.3.7), причому ця модуляція є глибокою. Такому сигналу відповідає спектральна інтенсивність

$$G^{+}(\omega) = \frac{\langle\rho\rangle^2}{\sqrt{2\pi}\sigma_\omega}\exp\left[-\frac{(\omega-\omega_0)^2}{2\sigma_\omega^2}\right] \tag{10.2.2}$$

(див. вище п. 9.3.7, формула (9.3.88)). Ширину $\Delta\omega_{techn} = \sigma_\omega$ називають технічною шириною спектральної лінії автогенератора. Вона залежить від конструкції автогенератора, стабілізації його параметрів тощо, тобто визначається технічними факторами (наприклад, наявність кварцової стабілізації частоти).

Однак навіть в ідеальному автогенераторі зі стабільними параметрами ширина спектральної лінії завжди відмінна від нуля. Причиною є тепловий (резистори та транзистори різних типів) та дробовий шум (біполярні транзистори, польові транзистори з керувальним *p-n* переходом) елементів схеми, який принципово неможливо усунути. Цей шум з широким спектром можна розглядати як випадкову зовнішню силу, яка діє на автогенератор.

Тоді вихідний сигнал також може бути поданий у вигляді вузькосмугового випадкового процесу (10.2.1). Але форма й ширина спектральної лінії відрізняються від (10.2.2). У цьому випадку добре працює наближення слабких та швидких флуктуацій частоти, і форма лінії виявляється лоренцівською,

$$G^{+}(\omega) \sim \frac{C}{(\omega-\omega_0)^2 + \Delta\Omega^2}. \tag{10.2.3}$$

В цьому разі говорять про природну ширину спектральної лінії $\Delta\omega_{nat} = \Delta\Omega$.

Можна сказати, що природна ширина лінії автогенератора – це її значення, менше якого вона принципово не може бути. Зазвичай $\Delta\omega_{techn} \gg \Delta\omega_{nat}$.

Задачу про природну ширину лінії автогенератора й буде розглянуто далі.

10.2.2. Рівняння для автогенератора Ван-дер-Поля з урахуванням власних шумів

Розглянемо природні флуктуації автогенератора на прикладі схеми на польовому транзисторі з ізольованим затвором. Джерелами шуму в схемі автогенератора Ван-дер-Поля виступають, по-перше, сам транзистор (на схемі, поданій на рис. 10.2.1, цим шумам відповідає еквівалентне джерело шумового струму i_n), по-друге – активний опір контуру (цим шумам відповідає еквівалентне джерело напруги u_n).

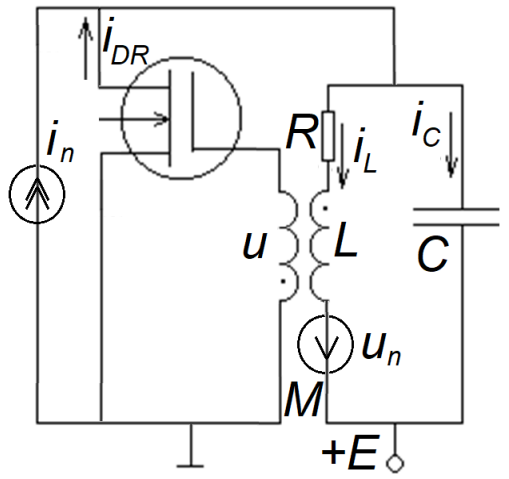


Рис. 10.2.1. Автогенератор Ван-дер-Поля з джерелами природних флуктуацій

Як уже вказувалося (див. вище п. 10.1.7), основними джерелами шуму польового транзистора з ізольованим затвором є тепловий шум, зумовлений активною провідністю каналу. Для інших типів активних елементів (лампи, біполярні транзистори, польові транзистори з керувальним *p-n* переходом) помітним, а в деяких випадках і визначальним є внесок дробового шуму.

Отримаємо рівняння для автогенератора Ван-дер-Поля з урахуванням шумових джерел. Як і раніше (див. п. 5.2.1, формула (5.2.1)) апроксимуємо прохідну характеристику польового транзистора кубічною параболою:

$$I_{DR}(U) = I_0 + SU - \frac{SU^3}{3K^2}. \tag{10.2.4}$$

Запишемо рівняння за другим законом Кірхгофа для коливного контуру:

$$L\frac{di_L}{dt^2} + Ri_L - \frac{1}{C}\int i_C dt = u_n(t). \tag{10.2.5}$$

За першим законом Кірхгофа

$$i_{DR} + i_n = i_L + i_C. \tag{10.2.6}$$

Тоді

$$L\frac{di_L}{dt^2} + Ri_L + \frac{1}{C}\int i_L dt - \frac{1}{C}\int (i_{DR} + i_n)dt = u_n(t). \tag{10.2.7}$$

Врахуємо, що напруга на затворі транзистора визначається формулою

$$u = M\frac{di_L}{dt}. \tag{10.2.8}$$

Тоді змінну складову детермінованого стокового струму (10.2.4) можна подати у формі

$$i_{DR}(u) = Su - \frac{Su^3}{3K^2} = MS\frac{di_L}{dt}\left[1 - \frac{1}{3}\left(\frac{M}{K}\right)^2\left(\frac{di_L}{dt}\right)^2\right]. \tag{10.2.9}$$

Підставивши (10.2.9) до (10.2.7) і диференціюючи за часом, отримуємо:

$$L\frac{d^2 i_L}{dt^2} + R\frac{di_L}{dt} + \frac{i_L}{C} - \frac{MS}{C}\frac{di_L}{dt}\left[1 - \frac{1}{3}\left(\frac{M}{K}\right)^2\left(\frac{di_L}{dt}\right)^2\right] = \frac{du_n}{dt} + \frac{i_n}{C}. \tag{10.2.10}$$

Нарешті, поділивши на L і скориставшись позначеннями

$$\frac{1}{LC}=\omega_0^2; \qquad \frac{MS}{LC}-\frac{R}{L}=\alpha; \qquad \frac{M^3S}{3K^2LC}=\gamma, \tag{10.2.11}$$

остаточно маємо:

$$\frac{d^2i_L}{dt^2}-\left[\alpha-\gamma\left(\frac{di_L}{dt}\right)^2\right]\frac{di_L}{dt}+\omega_0^2 i_L=\xi(t), \tag{10.2.12}$$

де

$$\xi(t)=\frac{1}{L}\frac{du_n}{dt}+\omega_0^2 i_n. \tag{10.2.13}$$

Рівняння (10.2.12) без правої частини – це знайоме нам рівняння Рейлі (див. п. 5.2.1, рівняння (5.2.7); див. рис. 10.2.2). Тепер у правій частині цього рівняння стоїть випадкова зовнішня сила $\xi(t)$, яка враховує власні шуми польового транзистора та активного опору в контурі. Таким чином, рівняння (10.2.12) – це звичайне нелінійне стохастичне диференціальне рівняння з випадковою зовнішньою силою.

Вважатимемо надалі, що $\alpha>0$. Як було показано в п. 5.2.2, ця нерівність забезпечує виконання амплітудної умови самозбудження автогенератора.

10.2.3. Вкорочені рівняння для автоколивань з урахуванням шумів

Будемо вважати контур автогенератора високодобротним. Тоді він реагує лише на ті спектральні складові зовнішньої сили, частоти яких близькі до його резонансної частоти. Це дає змогу розв'язувати рівняння (10.2.12) методом повільних амплітуд, вважаючи, що

$$i_L(t)=A(t)\exp(i\omega_0 t)+c.c., \tag{10.2.14}$$

$$\omega_0^2|A|\gg\omega_0|\dot{A}|\gg|\ddot{A}|.$$

Формально ми вважаємо параметри α та γ малими того самого порядку, що $\dot{A}$.

Фактично розв'язок (10.2.14) відповідає вузькосмуговому випадковому процесу, а точніше – коливанням з амплітудою, модульованою шумом (пор. з формулою (9.3.65)).

З точністю до доданків відповідно нульового та першого порядків мализни маємо:

$$\frac{di_L(t)}{dt}=i\omega_0 A\exp(i\omega_0 t)+c.c. \tag{10.2.15}$$

$$\frac{d^2i_L(t)}{dt^2}=\left(-\omega_0^2A+2i\omega_0\dot{A}\right)\exp(i\omega_0 t)+c.c \tag{10.2.16}$$

Випадковий процес $\xi(t)$ (10.2.13), що стоїть у правій частині рівняння (10.2.12), також подамо у формі коливань з амплітудою, модульованою шумом:

$$\xi(t)=i\eta(t)\exp(i\omega_0 t)+c.c. \tag{10.2.17}$$

Підставивши (10.2.14)-(10.2.17) до (10.2.12) і відкинувши третю гармоніку частоти ω_0, отримуємо:

$$\left(-\omega_0^2A+2i\omega_0\dot{A}-\alpha i\omega_0A+\omega_0^2A-i\eta\right)\exp(i\omega_0 t)+c.c=3\gamma(i\omega_0)^3A^2A^*\exp(i\omega_0 t)+c.c. \tag{10.2.18}$$

(тут права частина відповідає кубічному доданку, а доданок, що відповідає зовнішньому шуму, перенесений ліворуч).

Прирівнюючи амплітуди при $\exp(i\omega_0 t)$ у лівій та правій частинах рівняння (10.2.18), отримуємо вкорочене рівняння вигляду

$$\dot{A}-\frac{\alpha}{2}A+\frac{3}{2}\gamma\omega_0^2A^2A^*=\frac{1}{2\omega_0}\eta(t). \tag{10.2.19}$$

Ліва частина (10.2.19) збігається з укороченим рівнянням Рейлі (див. п. 5.2.5, рівняння

(5.2.20)).

Комплексну амплітуду $A(t)$ запишемо через модуль і фазу,

$$A(t)=\rho(t)\exp\left[i\varphi(t)\right], \tag{10.2.20}$$

а повільну зовнішню силу $\eta(t)$ – через квадратурні компоненти,

$$\eta(t)=a(t)+ib(t). \tag{10.2.21}$$

Підставимо (10.2.20)-(10.2.21) до (10.2.19) і поділимо на $\exp\left[i\varphi(t)\right]$. Маємо:

$$\dot{\rho}+i\dot{\varphi}\rho-\frac{\alpha}{2}\rho+\frac{3}{2}\gamma\omega_0^2\rho^3=\frac{1}{2\omega_0}\left[a\exp(-i\varphi)+ib\exp(-i\varphi)\right]. \tag{10.2.22}$$

Прирівнявши дійсні та уявні частини в лівій та правій частинах (10.2.22), отримуємо систему:

$$\begin{cases}\dot{\rho}-\dfrac{\alpha}{2}\rho+\dfrac{3}{2}\gamma\omega_0^2\rho^3=\varsigma_\rho(t);\\ \dot{\varphi}=\dfrac{1}{\rho}\varsigma_\varphi(t),\end{cases} \tag{10.2.23}$$

де

$$\varsigma_\rho(t)=\frac{1}{2\omega_0}(a\cos\varphi+b\sin\varphi),\qquad \varsigma_\varphi(t)=\frac{1}{2\omega_0}(-a\sin\varphi+b\cos\varphi). \tag{10.2.24}$$

За відсутності флуктуацій розв’язок системи (10.2.23) досліджувався в п. 5.2.5 (формули (5.2.24)-(5.2.25)). В усталеному режимі (при $t\to\infty$) незалежно від початкових умов установлюється амплітуда коливань

$$\rho_\infty=\frac{1}{\omega_0}\sqrt{\frac{\alpha}{3\gamma}} \tag{10.2.25}$$

(рис. 10.2.2). Початкова фаза коливань за відсутності шумів зберігається: $\varphi=\varphi_0$. Природно вважати фазу φ_0 випадковою величиною з розподілом

$$w(\varphi_0)=\frac{1}{2\pi},\qquad -\pi\le\varphi_0\le\pi. \tag{10.2.26}$$

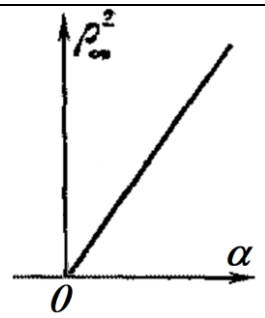


Рис. 10.2.2. Залежність квадрату усталеної амплітуди коливань автогенератора від параметра α за відсутності власних шумів

10.2.4. Спрощення вкорочених рівнянь

Вважатимемо шум $\xi(t)$ в правій частині рівняння Рейлі (див. формулу (10.2.13)) гауссівським та δ-корельованим з нульовим середнім, так що

$$\langle\xi(t)\rangle=0,\qquad B(\tau)=\langle\xi\xi_\tau\rangle=2\pi G_\xi\delta(\tau). \tag{10.2.27}$$

Процеси $\eta(t)$, $a(t)$ та $b(t)$ у вкорочених рівняннях (10.2.19), (10.2.23)-(10.2.24) є повільними в масштабі частоти ω_0. Однак у масштабі характерних часів зміни величин $A(t)$,

$\rho(t)$ й $\varphi(t)$ (або, що те саме, в масштабі частоти $\Delta\omega << \omega_0$ – ширини смуги пропускання контуру автогенератора – їх також можна вважати білим шумом.

Як було показано раніше, при амплітудній модуляції шумом форма спектру керувального та модульованого сигналів однакова (див. вище п. 9.3.7). Тоді

$$\langle a(t)\rangle = \langle b(t)\rangle = 0, \quad \langle ab_\tau\rangle = \langle a_\tau b\rangle = 0, \qquad \langle aa_\tau\rangle = \langle bb_\tau\rangle = \frac{\pi}{2}G_\xi\delta(\tau). \tag{10.2.28}$$

Відповідно, з урахуванням формули (10.2.21), маємо:

$$\langle \eta(t)\rangle = 0, \quad \langle \eta\eta_\tau\rangle = 0, \qquad \langle \eta\eta_\tau^*\rangle = \pi G_\xi\delta(\tau), \tag{10.2.29}$$

Стохастичні рівняння (10.2.23) утворюють систему. Однак їх можна перетворити таким чином, що $\rho(t)$ й $\varphi(t)$ перестануть залежати одне від одного. Для цього треба виділити швидкі й повільні компоненти процесів ς_ρ та ς_φ.

Нехай помітна зміна величини $\rho(t)$ відбувається за характерний час $\tau_\rho \sim 1/\Delta\omega$, а величин $a(t)$ та $b(t)$ – за характерний час $\tau_\eta \ll \tau_\rho$ (для аналізованої моделі $\tau_\eta \to 0$).

Тепер розглянемо випадковий процес $\varsigma_\rho(t)$. Подамо його як суму повільної компоненти $\xi_0(t)$ з характерним часом зміни порядку τ_ρ та швидкої компоненти $\xi_\rho(t)$ з характерним часом зміни порядку τ_η:

$$\varsigma_\rho(t) = \xi_0(t) + \xi_\rho(t). \tag{10.2.30}$$

Щоб виділити з $\varsigma_\rho(t)$ повільну компоненту $\xi_0(t)$, необхідно виконати усереднення на проміжку часу T такому, що $\tau_\rho \gg T \gg \tau_\eta$:

$$\xi_0(t) = \overline{\varsigma_\rho(t)}_T = \left(\overline{a\cos\varphi}\right)_T + \left(\overline{b\sin\varphi}\right)_T. \tag{10.2.31}$$

Тоді, очевидно,

$$\xi_\rho(t) = \varsigma_\rho(t) - \xi_0(t). \tag{10.2.32}$$

В усталеному режимі внутрішні шуми автогенератора можна вважати стаціонарним випадковим процесом. В результаті, як випливає з другого рівняння системи (10.2.23),

$$\dot{\varphi} = \frac{1}{\rho}\varsigma_\varphi(t),$$

фаза $\varphi(t)$ являє собою дифузійний випадковий процес (точніше, вінерівський, оскільки ми вважаємо процес $\varsigma_\varphi(t)$ δ-корельованим) – пор. з рівнянням (9.3.50). Його дисперсія на проміжку $(-\infty,\infty)$ необмежено зростає з часом (див. вище п. 9.3.6). Однак, якщо фазу привести до інтервалу завширшки 2π, вона буде там розподілена рівномірно незалежно від статистики самого процесу $x(t)$:

$$w(\varphi) = \frac{1}{2\pi}, \qquad -\pi \le \varphi \le \pi. \tag{10.2.33}$$

Випадкові процеси $\cos\varphi$ та $\sin\varphi$ стаціонарні. Вважаючи їх ергодичними, замінимо в (10.2.31) усереднення за часом на статистичне усереднення:

$$\xi_0(t) = \frac{1}{2\omega_0}\left(\langle a\cos\varphi\rangle + \langle b\sin\varphi\rangle\right). \tag{10.2.34}$$

Розіб'ємо $\cos\varphi$ та $\sin\varphi$ на частини, які корелюють з процесами $a(t)$ й $b(t)$, та частини, які з ними не корелюють:

$$\cos\varphi = (\cos\varphi)^{corr} + (\cos\varphi)^{uncorr}, \qquad \sin\varphi = (\sin\varphi)^{corr} + (\sin\varphi)^{uncorr}. \tag{10.2.35}$$

Тоді

$$\xi_0(t) = \left\langle a(\cos\varphi)^{corr} \right\rangle + \left\langle b(\sin\varphi)^{corr} \right\rangle, \tag{10.2.36}$$

$$\xi_\rho(t) = a\left\langle (\cos\varphi)^{uncorr} \right\rangle + b\left\langle (\sin\varphi)^{uncorr} \right\rangle \tag{10.2.37}$$

Відповідно в силу формул (10.2.29)

$$\left\langle \xi_\rho(t) \right\rangle = \langle a \rangle \left\langle (\cos\varphi)^{uncorr} \right\rangle + \langle b \rangle \left\langle (\sin\varphi)^{uncorr} \right\rangle = 0. \tag{10.2.38}$$

Аналогічно можна записати, що

$$\varsigma_\varphi(t) = \xi_1(t) + \xi_\varphi(t), \tag{10.2.39}$$

$$\xi_1(t) = \left\langle b(\cos\varphi)^{corr} \right\rangle - \left\langle a(\sin\varphi)^{corr} \right\rangle, \tag{10.2.40}$$

$$\xi_\varphi(t) = b\left\langle (\cos\varphi)^{uncorr} \right\rangle - a\left\langle (\sin\varphi)^{uncorr} \right\rangle, \tag{10.2.41}$$

$$\left\langle \xi_\varphi(t) \right\rangle = 0. \tag{10.2.42}$$

Для знаходження $(\cos\varphi)^{corr}$ домножимо друге з рівнянь (10.2.23) на $\sin\varphi$. Отримаємо:

$$\frac{d}{dt}(\cos\varphi) = -\frac{1}{\rho}\varsigma_\varphi(t)\sin\varphi. \tag{10.2.43}$$

Інтегрування дає:

$$\cos\varphi = -\int_{t_0}^{t} \frac{\varsigma_\varphi(t')\sin\varphi dt'}{\rho} = -\frac{1}{\rho}\int_{t_0}^{t}\left[\xi_1(t') + \xi_\varphi(t')\right]\sin\varphi dt'. \tag{10.2.44}$$

Внесок до $(\cos\varphi)^{corr}$ дає лише частина, пропорційна $a(t)$ або $b(t)$, тобто ξ_φ (див. вище формули (10.2.40)-(10.2.41)), тому

$$\begin{aligned}
(\cos\varphi)^{corr} &= -\frac{1}{\rho}\int_{t_0}^{t}\xi_\varphi(t')(\sin\varphi)^{uncorr} = -\frac{1}{\rho}\int_{t_0}^{t}\left[b\left\langle (\cos\varphi)^{uncorr}\right\rangle - a\left\langle (\sin\varphi)^{uncorr}\right\rangle\right](\sin\varphi)^{uncorr}\,dt' = \\
&= -\frac{1}{\rho}\int_{t_0}^{t}\left\{\frac{1}{2}b\left\langle (\sin 2\varphi)^{uncorr}\right\rangle - a\left\langle \left[(\sin\varphi)^{uncorr}\right]^2\right\rangle\right\}dt' = \\
&= -\frac{1}{\rho}\left\{\frac{1}{2}B\left\langle (\sin 2\varphi)^{uncorr}\right\rangle - A\left\langle \left[(\sin\varphi)^{uncorr}\right]^2\right\rangle\right\},
\end{aligned} \tag{10.2.45}$$

де

$$A(t) = \int_{t_0}^{t} a(t')\,dt', \qquad B(t) = \int_{t_0}^{t} b(t')\,dt'. \tag{10.2.47}$$

Тоді

$$\begin{aligned}
\left\langle a(t)(\cos\varphi)^{corr}\right\rangle &= \frac{1}{\rho}\left\{\left\langle \left[(\sin\varphi)^{uncorr}\right]^2\right\rangle \int_{t_0}^{t}\left\langle a(t')a(t)\right\rangle dt' - \frac{1}{2}\left\langle (\sin 2\varphi)^{uncorr}\right\rangle \int_{t_0}^{t}\left\langle b(t')a(t)\right\rangle dt'\right\} = \\
&= \frac{1}{\rho}\left\langle \left[(\sin\varphi)^{uncorr}\right]^2\right\rangle \frac{\pi}{2}G_\xi = \frac{\pi}{4\rho}G_\xi.
\end{aligned} \tag{10.2.48}$$

Аналогічно можна показати, що

$$\left\langle b(t)(\sin\varphi)^{corr}\right\rangle = \frac{\pi}{4\rho}G_\xi. \tag{10.2.49}$$

Тоді, відповідно до (10.2.36),

$$\xi_0(t)=\frac{\pi}{2\rho(t)}G_\xi\,. \qquad (10.2.50)$$

Автокореляційна функція процесу $\xi_\rho(t)$:

$$B_{\xi_\rho}(\tau)=\left\langle\xi_\rho(t)\xi_\rho(t+\tau)\right\rangle=\left\langle\left[a\left\langle(\cos\varphi)^{uncorr}\right\rangle+b\left\langle(\sin\varphi)^{uncorr}\right\rangle\right]\left[a_\tau\left\langle(\cos\varphi)^{uncorr}\right\rangle_\tau+b_\tau\left\langle(\sin\varphi)^{uncorr}\right\rangle_\tau\right]\right\rangle=$$

$$=\left\langle aa_\tau\right\rangle\left\langle(\cos\varphi)^{uncorr}\right\rangle\left\langle(\cos\varphi)^{uncorr}\right\rangle_\tau+\left\langle bb_\tau\right\rangle\left\langle(\sin\varphi)^{uncorr}\right\rangle\left\langle(\sin\varphi)^{uncorr}\right\rangle_\tau=\frac{\pi}{2}G_\xi\delta(\tau)\,. \qquad (10.2.51)$$

Аналогічно можна показати, що

$$\xi_1(t)=0,\qquad \left\langle\xi_\varphi(t)\xi_\varphi(t+\tau)\right\rangle=\frac{\pi}{2}G_\xi\delta(\tau),\qquad \left\langle\xi_\rho\xi_\varphi\right\rangle=0\,. \qquad (10.2.52)$$

В результаті систему (10.2.23)

$$\begin{cases}\dot\rho-\dfrac{\alpha}{2}\rho+\dfrac{3}{2}\gamma\omega_0^2\rho^3=\varsigma_\rho(t);\\ \dot\varphi=\dfrac{1}{\rho}\varsigma_\varphi(t),\end{cases}$$

можна переписати у формі:

$$\begin{cases}\dot\rho-\dfrac{\alpha}{2}\rho+\dfrac{3}{2}\gamma\omega_0^2\rho^3=\dfrac{\pi}{4\rho}G_\xi+\dfrac{1}{2\omega_0}\xi_\rho(t);\\ \dot\varphi=\dfrac{1}{\rho}\xi_\varphi(t).\end{cases} \qquad (10.2.53)$$

Шуми $\xi_\rho(t)$ й $\xi_\varphi(t)$ є статистично незалежними й δ-корельованими.

Перше рівняння системи (10.2.53) можна розв’язувати незалежно від другого.

10.2.5. Функція розподілу й моменти амплітуди автоколивань

Перше рівняння системи (10.2.53) можна розглядати як частинний випадок рівняння (9.4.32)

$$\dot x=a(x)+b(x)\xi(t),\qquad \left\langle\xi\xi_\tau\right\rangle=2D\delta(\tau),$$

в якому

$$x\to\rho,\qquad a(\rho)=\frac{\alpha}{2}\rho-\frac{3}{2}\gamma\omega_0^2\rho^3+\frac{\pi}{4\rho}G_\xi,\qquad b=\frac{1}{2\omega_0}.\qquad 2D=\frac{\pi}{2}G_\xi \qquad (10.2.54)$$

Тоді для розподілу $w(\rho,t)$ випадкового процесу $\rho(t)$ можна відразу записати рівняння Фоккера – Планка вигляду (9.4.53)

$$\frac{\partial w}{\partial t}=-\frac{\partial}{\partial x}(wK_1)+\frac{1}{2}\frac{\partial^2}{\partial x^2}(wK_2),$$

де, відповідно до (9.4.52)

$$K_1(\rho)=a(\rho)+Db(\rho)b'(\rho)=\frac{\alpha}{2}\rho-\frac{3}{2}\gamma\omega_0^2\rho^3+\frac{\pi}{4\rho}G_\xi\,; \qquad (10.2.55)$$

$$K_2(\rho)=2Db^2(\rho)=\frac{\pi}{8\omega_0^2}G_\xi\,. \qquad (10.2.56)$$

Стаціонарний розподіл $w(\rho)$, який встановлюється при $t\to\infty$, відповідно до формули (9.4.56) має вигляд

$$w(\rho)=C_0\rho\exp\left(-\frac{\rho^4-2q\rho^2}{4N}\right),\qquad N=\frac{\pi}{6\gamma}G_\xi,\qquad q=\frac{2\alpha}{3\gamma\omega_0^2}\,. \qquad (10.2.57)$$

Константа C_0 визначається з умови нормування.

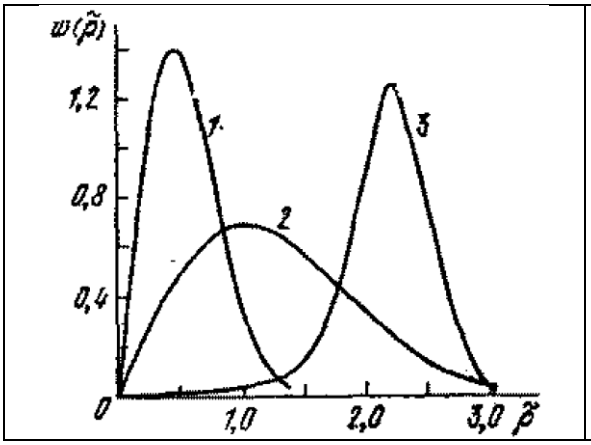


Рис. 10.2.3. Графіки функції розподілу приведеної амплітуди $\tilde{\rho} = \rho / \sqrt[4]{N}$ для різних значень параметра $\tilde{q} = q / \sqrt{N}$: $\tilde{q} = -5$ (крива 1), $\tilde{q} = 0$ (крива 2), $\tilde{q} = 5$ (крива 3)

Графіки функції $w(\rho)$ подані на рис. 10.2.3. Параметр $\tilde{q}$ прямо пропорційний інкременту малих коливань α. Таким чином, крива 1 відповідає схемі, в якій амплітудна умова самозбудження не виконана (при цьому амплітуда коливань відмінна від нуля за рахунок шумів), крива 2 – автогенератору на межі самозбудження (при цьому дисперсія максимальна), крива 3 – режиму розвиненої генерації.

Знання розподілу $w(\rho)$ дозволяє підрахувати його моменти $\langle \rho \rangle$, $\langle \rho^2 \rangle$ та дисперсію $\sigma_\rho^2 = \langle \rho^2 \rangle - \langle \rho \rangle^2$.

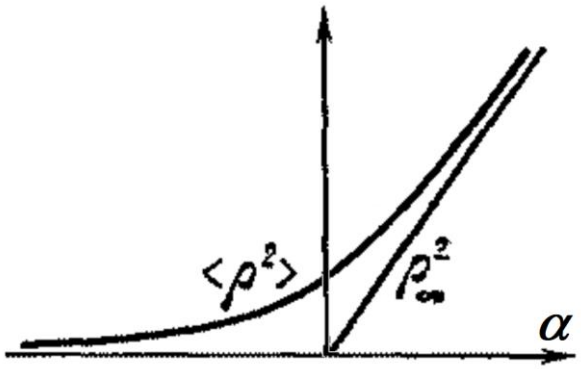


Рис. 10.2.4. Графік залежності квадрату усталеної амплітуди коливань автогенератора від параметра α за відсутності (ρ_∞^2) та за наявності ($\langle \rho^2 \rangle$) власних шумів

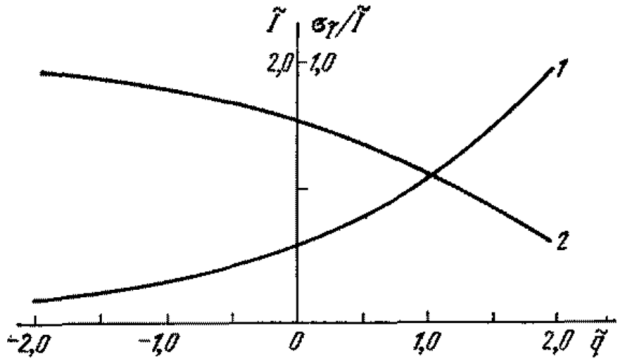


Рис. 10.2.5. Графік залежності приведеної середньої інтенсивності коливань автогенератора $\tilde{I} = \langle \rho^2 \rangle / 2\sqrt{N}$ (крива 1) та їхньої відносної дисперсії $\sigma_\rho^2 / \langle \rho^2 \rangle$ (крива 2) від нормованого інкременту малих коливань $\tilde{q}$

Значення $\langle \rho^2 \rangle$ як функція α, показане на рис. 10.2.4. Наявність шумів приводить до того, що $\langle \rho^2 \rangle \neq 0$ навіть при $\alpha \leq 0$. В режимі розвиненої генерації, тобто при $\tilde{q} \geq 1$, або

$$\alpha \geq \omega_0^2 \sqrt{(3/8)\pi\gamma G_\xi}\,, \tag{10.2.58}$$

шум перестає помітно впливати на значення $\langle \rho^2 \rangle$.

На рис. 10.2.5 показано залежність приведеної середньої інтенсивності коливань автогенератора та їхньої відносної дисперсії від нормованого інкремента малих коливань. Перша з цих залежностей, по суті, повторює хід відповідного графіка на рис. 10.2.4. Друга крива підтверджує висловлену раніше думку, що в режимі розвиненої генерації вплив шумів на сигнал автогенератора стає несуттєвим.

10.2.6. Спектр флуктуацій амплітуди автоколивань

Для знаходження спектру флуктуацій амплітуди автоколивань необхідно спершу отримати розв'язок відповідного диференціального рівняння. Потім за відомим розв'язком можна знайти автокореляційну функцію. Нарешті, застосувавши перетворення Фур'є, за автокореляційною функцією знаходять спектральну інтенсивність.

Рівняння для амплітуди автоколивань (перше рівняння системи (10.2.53)) нелінійне:

$$\dot{\rho} - \frac{\alpha}{2}\rho + \frac{3}{2}\gamma\omega_0^2\rho^3 = \frac{\pi}{4\rho}G_\xi + \omega_0\xi_\rho(t).$$

Однак, як уже відзначалося, в режимі розвиненої генерації відносна дисперсія флуктуацій амплітуди мала. Тому в цьому випадку рівняння для флуктуацій $\tilde{\rho} = \rho - \rho_\infty$ можна лінеаризувати. Отримаємо:

$$\frac{d\tilde{\rho}}{dt} - \frac{\alpha}{2}(\rho_\infty + \tilde{\rho}) + \frac{3}{2}\gamma\omega_0^2(\rho_\infty + \tilde{\rho})^3 = \frac{\pi}{4(\rho_\infty + \tilde{\rho})}G_\xi + \omega_0\xi_\rho(t)$$

$$\frac{d\tilde{\rho}}{dt} - \frac{\alpha}{2}(\rho_\infty + \tilde{\rho}) + \frac{3}{2}\gamma\omega_0^2(\rho_\infty^3 + 3\rho_\infty^2\tilde{\rho}) = \frac{\pi}{4\rho_\infty}\left(1 - \frac{\tilde{\rho}}{\rho_\infty}\right)G_\xi + \omega_0\xi_\rho(t)$$

$$\frac{d\tilde{\rho}}{dt} - \frac{\alpha}{2}\tilde{\rho} + \frac{9}{2}\gamma\omega_0^2\rho_\infty^2\tilde{\rho} + \frac{\pi G_\xi}{4\rho_\infty^2}\tilde{\rho} = \omega_0\xi_\rho(t)$$

$$-\frac{\alpha}{2}\rho_\infty + \frac{3}{2}\gamma\omega_0^2\rho_\infty^3 = \frac{\pi G_\xi}{4\rho_\infty}$$

$$\rho_\infty^2 \approx \frac{\alpha}{3\gamma\omega_0^2}$$

$$\frac{d\tilde{\rho}}{dt} - \left(\frac{\alpha}{2} - \frac{3\alpha}{2} - \frac{3\pi\gamma\omega_0^2 G_\xi}{4\alpha}\right)\tilde{\rho} = \omega_0\xi_\rho(t)$$

$$\frac{d\tilde{\rho}}{dt} + \beta\tilde{\rho} = \omega_0\xi_\rho(t), \qquad \beta = \alpha + \frac{3\pi\gamma\omega_0^2}{4\alpha}G_\xi \tag{10.2.59}$$

Розв'язок рівняння (10.2.59) можна знайти методом варіації довільної сталої:

$$\tilde{\rho}(t) = C(t)exp(-\beta t);$$

$$\frac{d\tilde{\rho}}{dt} = \left(\frac{dC}{dt} - \beta C\right)exp(-\beta t);$$

$$\left(\frac{dC}{dt} - \beta C\right)exp(-\beta t) + \beta C exp(-\beta t) = \omega_0\xi_\rho(t);$$

$$C(t) = \omega_0 \int_0^t \xi_\rho(t')exp(\beta t')dt' \tag{10.2.60}$$

(вважаємо початкові умови нульовими). Тоді остаточно маємо:

$$\tilde{\rho}(t) = \omega_0\, exp(-\beta t)\int_0^t \xi_\rho(t')exp(\beta t')dt'. \tag{10.2.61}$$

Автокореляційна функція амплітудних флуктуацій:

$$B_{\tilde{\rho}}(\tau) = \langle \tilde{\rho}\tilde{\rho}_\tau \rangle =$$

$$= \langle \omega_0\, exp(-\beta t) \int_0^t \xi_\rho(t') exp(\beta t') dt' \omega_0\, exp[-\beta(t+\tau)] \int_0^{t+\tau} \xi_\rho(t'') exp[\beta(t''+\tau)] dt'' \rangle =$$

$$= \omega_0^2\, exp[-\beta(2t+\tau)] \int_0^t dt' \int_0^{t+\tau} dt'' \langle \xi_\rho(t') \xi_\rho(t'') \rangle exp[\beta(t'+t'')] = \qquad (10.2.62)$$

$$= \omega_0^2\, exp[-\beta(2t+\tau)] \int_0^t dt' \int_0^{t+\tau} dt'' \frac{\pi}{2} G_\xi \delta(t'-t'') exp[\beta(t'+t'')] =$$

$$= \frac{\pi}{2} \omega_0^2 G_\xi exp[-\beta(2t+\tau)] \int_0^t dt' exp(2\beta t') = \frac{\pi}{4\beta} \omega_0^2 G_\xi exp[-\beta(2t+\tau)][exp(2\beta t) - 1] =$$

$$= \frac{\pi}{4\beta} \omega_0^2 G_\xi exp(-\beta\tau)[1 - exp(-2\beta t)]$$

(враховано формулу (10.2.51)).

В усталеному режимі (при $\beta t >> 1$) вираз спрощується:

$$B_{\tilde{\rho}}(\tau) = \frac{\pi}{4\beta} \omega_0^2 G_\xi exp(-\beta\tau). \qquad (10.2.63)$$

Відповідно до теореми Вінера – Хінчина (формула (9.2.31)) маємо:

$$G_\rho(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, B_{\tilde{\rho}}(\tau) exp(i\omega\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \frac{\pi}{4\beta} \omega_0^2 G_\xi exp(-\beta|\tau|) exp(i\omega\tau) =$$

$$= \frac{\pi}{4\beta} \omega_0^2 G_\xi \left\{ \int_{-\infty}^{0} d\omega\, exp[(\beta + i\omega)\tau] + \int_0^{\infty} d\omega\, exp[(-\beta + i\omega)\tau] \right\} =$$

$$= \frac{\omega_0^2 G_\xi}{4\beta} \left(\frac{1}{\beta + i\omega} - \frac{1}{-\beta + i\omega} \right) = \frac{\omega_0^2 G_\xi}{2(\omega^2 + \beta^2)} \qquad (10.2.64)$$

(рис. 10.2.6). Видно, що максимум спектральної інтенсивності флуктуацій амплітуди автогенератора припадає на нульову частоту. Півширина цього спектру – β (при малих іньенсивностях шуму $\beta \approx \alpha$), тобто вона визначається інкрементом малих коливань автогенератора. Зі зростанням інтенсивності шуму названа півширина росте.

Правильність формули (10.2.64) неодноразово перевірялася в експериментах. Для вимірювання спектру $G_\rho(\omega)$ сигнал автогенератора слід пропустити через квадратичний детектор.

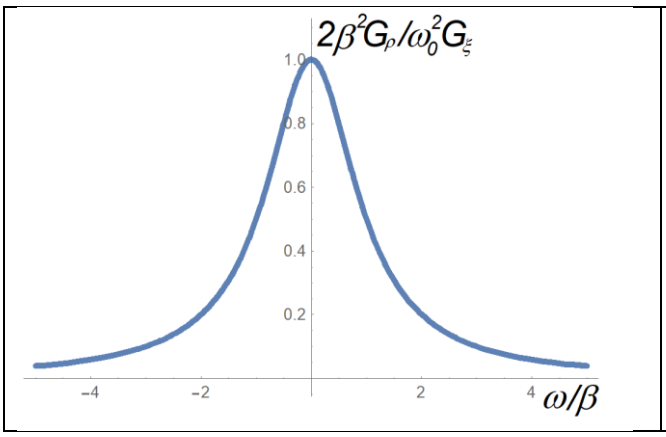


Рис. 10.2.6. Графік спектру флуктуацій амплітуди автогенератора

10.2.7. Флуктуації фази автогенератора та їхній спектр

Флуктуації фази автоколивань описуються другим з рівнянь системи (10.2.53):

$$\dot{\varphi} = \frac{1}{\rho(t)} \xi_\varphi(t).$$

В загальному випадку розв'язати це рівняння складно, оскільки амплітуда коливань $\rho(t)$ флуктуює. Але в режимі розвиненої генерації ці флуктуації малі, й випадковий процес $\rho(t)$ можна наближено замінити його середнім значенням:

$$\dot{\varphi} = \frac{1}{\langle \rho \rangle} \xi_\varphi(t). \qquad (10.2.65)$$

Процес $\xi_\varphi(t)$ в правій частині (10.2.65), як уже вказувалося, можна вважати δ-корельованим (в масштабі зміни $\rho(t)$).

В цьому наближенні випадковий процес $\varphi(t)$ є вінерівським. Можна записати, що приріст фази $\Delta\varphi(\tau)$ на проміжку часу τ складає

$$\Delta\varphi(\tau) = \frac{1}{\langle \rho \rangle} \int_t^{t+\tau} \xi_\varphi(t')\,dt'. \qquad (10.2.66)$$

Дисперсія процесу $\Delta\varphi(\tau)$ лінійно зростає зі зростанням τ (див. вище п. 9.3.6):

$$\sigma_\varphi^2(\tau) = \left\langle \left[\Delta\varphi(\tau)\right]^2 \right\rangle = D\tau, \qquad D = \frac{\pi\omega_0^2}{2\langle \rho \rangle^2} G_\xi. \qquad (10.2.67)$$

Тут D – коефіцієнт дифузії фази.

Для стохастичного диференціального рівняння (10.2.65) можна записати рівняння Фоккера – Планка. Його розв'язок має вигляд гауссівського розподілу:

$$w(\Delta\varphi) = \frac{1}{\sqrt{2\pi}\sigma_\varphi(\tau)} \exp\left[-\frac{(\Delta\varphi)^2}{2\sigma_\varphi^2(\tau)}\right]. \qquad (10.2.68)$$

Очевидно, повна фаза коливань складає величину

$$\varphi(t) = \varphi_0 + \frac{1}{\langle \rho \rangle} \int_0^t \xi_\varphi(t')\,dt'. \qquad (10.2.69)$$

Її моменти:

$$\langle \varphi(t) \rangle = \varphi_0, \qquad \sigma_\varphi^2(\tau) = \left\langle \left[\varphi(t) - \varphi_0\right]^2 \right\rangle = Dt. \qquad (10.2.70)$$

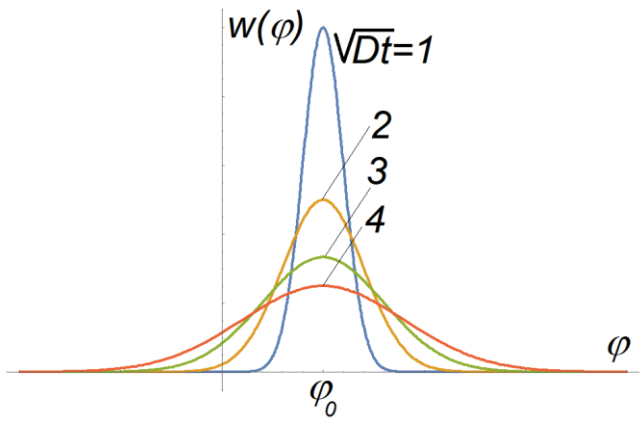


Рис. 10.2.7. Зміна густини розподілу фази автоколивань з часом

Функція розподілу фази:

$$w(\varphi) = \frac{1}{\sqrt{2\pi}\sigma_\varphi(t)} \exp\left[-\frac{(\varphi - \varphi_0)^2}{2\sigma_\varphi^2(t)}\right] \qquad (10.2.71)$$

(рис. 10.2.7).

Перейдемо до розрахунку спектральної густини флуктуацій фази. Миттєва частота автоколивань може бути записана у формі

$$\omega(t) = \frac{d}{dt}\left[\omega_0 t + \varphi(t)\right] = \omega_0 + \frac{1}{\langle \rho \rangle} \xi_\varphi(t). \qquad (10.2.72)$$

Відповідно флуктуації частоти мають вигляд

$$\delta\omega(t)=\omega(t)-\omega_0=\frac{1}{\langle\rho\rangle}\xi_\varphi(t). \qquad (10.2.73)$$

Оскільки випадковий процес $\xi_\varphi(t)$ вважається δ-корельованим, таким самим буде й процес $\delta\omega(t)$. Відповідно в цьому наближенні його спектральна інтенсивність стала:

$$G_{\delta\omega}(\omega)=\left(\frac{1}{2\langle\rho\rangle}\right)^2 G_\xi. \qquad (10.2.74)$$

Нагадаємо, що процес $\xi_\varphi(t)$ можна вважати δ-корельованим лише в масштабі характерного часу зміни амплітуди автоколивань τ_ρ. Це означає, що на частотах ω таких, що $\omega\tau_\rho \gg 1$, спектр $G_{\delta\omega}(\omega)$ перестає бути рівномірним.

10.2.8. Форма та природна ширина спектральної лінії автогенератора

Для знаходження спектральної інтенсивності автоколивань вигляду (10.2.1)

$$x(t)=\rho(t)\cos\left[\omega_0 t+\varphi(t)\right]$$

необхідно спочатку знайти їхню автокореляційну функцію, як ми вже робили це раніше. Ця задача полегшується завдяки тому, що, як було показано вище, флуктуації амплітуди $\rho(t)$ й фази $\varphi(t)$ можна вважати статистично незалежними. Таким чином, можна записати, що

$$B_x(\tau)=\langle x x_\tau\rangle=\langle\rho(t)\rho(t+\tau)cos[\omega_0 t+\varphi(t)]cos[\omega_0(t+\tau)+\varphi(t+\tau)]\rangle=$$
$$=\langle(\langle\rho\rangle+\tilde{\rho})(\langle\rho\rangle+\tilde{\rho}_\tau)\rangle\frac{1}{2}\langle cos[\omega_0(2t+\tau)+\varphi(t)+\varphi(t+\tau)]+cos[\omega_0\tau+\varphi(t+\tau)-\varphi(t)]\rangle$$
$$=\frac{1}{2}(\langle\rho\rangle^2+\langle\tilde{\rho}\tilde{\rho}_\tau\rangle)\langle\cos[\omega_0\tau+\Delta\varphi(\tau)]\rangle. \qquad (10.2.75)$$

Тут враховано, що в стаціонарному режимі розподіл фази можна вважати рівномірним на проміжку $[-\pi,\pi]$, тому

$$\left\langle\cos\left[\omega_0(2t+\tau)+\varphi(t)+\varphi(t+\tau)\right]\right\rangle=0. \qquad (10.2.76)$$

Величина $B_{\tilde{\rho}}(\tau)=\langle\tilde{\rho}\tilde{\rho}_\tau\rangle$ була знайдена раніше (див. формулу (10.2.63)):

$$B_{\tilde{\rho}}(\tau)=\frac{\pi}{4\beta}\omega_0^2 G_\xi exp(-\beta\tau).$$

Скориставшись розподілом (10.2.68),

$$w(\Delta\varphi)=\frac{1}{\sqrt{2\pi}\sigma_\varphi(\tau)}\exp\left[-\frac{(\Delta\varphi)^2}{2\sigma_\varphi^2(\tau)}\right],$$

можна розрахувати величину

$$\langle\cos\Delta\varphi\rangle=\int_{-\infty}^{\infty}d(\Delta\varphi)\cos\Delta\varphi w(\Delta\varphi)=\frac{1}{\sqrt{2\pi}\sigma_\varphi(\tau)}\int_{-\infty}^{\infty}d(\Delta\varphi)\cos\Delta\varphi\exp\left[-\frac{(\Delta\varphi)^2}{2\sigma_\varphi^2(\tau)}\right]=$$
$$=\frac{1}{\sqrt{2\pi}\sigma_\varphi(\tau)}\sqrt{2\pi}\sigma_\varphi(\tau)\exp\left[-\frac{1}{2}\sigma_\varphi^2(\tau)\right]=\exp\left[-\frac{1}{2}\sigma_\varphi^2(\tau)\right]=\exp\left(-\frac{1}{2}D\tau\right) \qquad (10.2.77)$$

(враховано формулу (10.2.67)). Аналогічно можна показати, що

$$\langle\sin\Delta\varphi\rangle=0. \qquad (10.2.78)$$

Тоді

$$\left\langle\cos\left[\omega_0\tau+\Delta\varphi(\tau)\right]\right\rangle=\left\langle\cos\omega_0\tau\cos\left[\Delta\varphi(\tau)\right]-\sin\omega_0\tau\sin\left[\Delta\varphi(\tau)\right]\right\rangle=$$
$$=\cos\omega_0\tau\left\langle\cos\left[\Delta\varphi(\tau)\right]\right\rangle-\sin\omega_0\tau\left\langle\sin\left[\Delta\varphi(\tau)\right]\right\rangle=\cos\omega_0\tau\exp\left(-\frac{\pi\omega_0^2}{4\langle\rho\rangle^2}G_\xi\tau\right). \qquad (10.2.79)$$

Підставляючи (10.2.63) та (10.2.79) до (10.2.75), знаходимо, що

$$B_x(\tau) = \frac{1}{2}(\langle\rho\rangle^2 + \langle\tilde{\rho}\tilde{\rho}_\tau\rangle)\langle\cos[\omega_0\tau + \Delta\varphi(\tau)]\rangle =$$

$$= \frac{1}{2}\left[\langle\rho\rangle^2 + \frac{\pi}{4\beta}\omega_0^2 G_\xi exp(-\beta\tau)\right] exp\left(\frac{\pi\omega_0^2 G_\xi}{4\langle\rho\rangle^2}\tau\right) = B_{x1}(\tau) + B_{x2}(\tau). \qquad (10.2.80)$$

Тоді, відповідно до теореми Вінера – Хінчина, можна записати спектральну інтенсивність автоколивань у формі

$$G_x(\tau) = G_{x1}(\tau) + G_{x2}(\tau) \qquad (10.2.81)$$

$$G_{x1}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} B_{x1}(\tau)\exp(i\omega\tau)d\omega = \frac{\langle\rho\rangle^2 D^2}{8\pi\left[(\omega-\omega_0)^2 + D^2/4\right]}; \qquad (10.2.82)$$

$$G_{x2}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} B_{x2}(\tau)exp(i\omega\tau)d\omega = \frac{(\beta+D)\omega_0^2 G_\xi}{8\pi(\omega^2+4\beta^2)\{(\omega-\omega_0)^2+[(\beta+D)/2]^2\}}. \qquad (10.2.83)$$

Обидві складові спектральної інтенсивності – це лоренцівські лінії з максимумами на частоті ω_0. Вони мають амплітуди

$$G_{x1}(\omega_0) = \frac{\langle\rho\rangle^2}{2\pi}, \qquad G_{x2}(\omega_0) = \frac{\omega_0^2 G_\xi}{2\pi(\omega_0^2+4\beta^2)(\beta+D)} \approx \frac{G_\xi}{2\pi\beta}, \qquad G_{x1}(\omega_0) \gg G_{x2}(\omega_0) \qquad (10.2.84)$$

(враховано, що в режимі розвиненої генерації квазігармонічних коливань $\omega_0^2 \gg \beta \gg D$ та $\langle\rho\rangle^2 \gg G_\xi/\beta$).

Відповідно півширина цих складових

$$\Delta\omega_1 = D/2 = \frac{\pi\omega_0^2}{4\langle\rho\rangle^2}G_\xi, \qquad \Delta\omega_2 = (\beta + D)/2 \approx \beta/2, \qquad \Delta\omega_1 \ll \Delta\omega_2. \qquad (10.2.85)$$

Видно, що за порядком величини $\Delta\omega_2$ збігається з півшириною спектру амплітудних флуктуацій (див. вище формулу (10.2.64) та рис. 10.2.6).

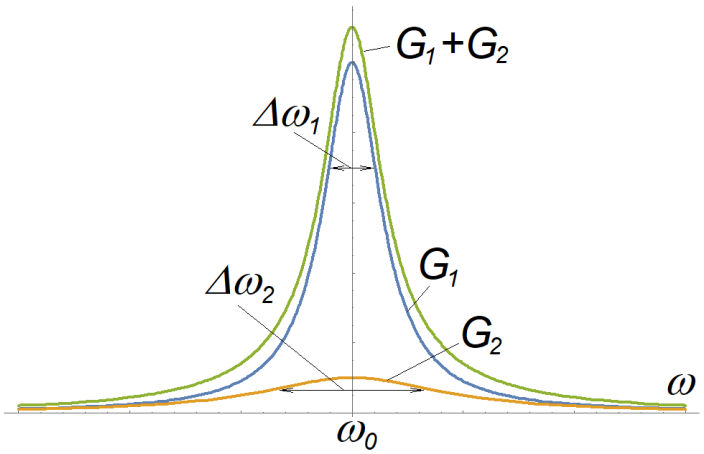


Рис. 10.2.8. Спектральна лінія автогенератора та її складові (схематичне зображення)

Як випливає з формул (10.2.82)-(10.2.83), складова $G_{x1}(\omega)$ зумовлена флуктуаціями фази, а складова $G_{x2}(\omega)$ – як флуктуаціями фази, так і флуктуаціями амплітуди (рис. 10.2.8). Оскільки складова $G_{x2}(\omega)$ помітна лише при великих розстроюваннях, то природна ширина лінії визначається величиною $\Delta\omega_1$, тобто інтенсивністю власних шумів схеми.

10.2.9. Встановлення коливань автогенератора від рівня шумів: розподіл для миттєвої амплітуди

Процес встановлення автоколивань від рівня власних шумів автогенератора можна

описати за допомогою першого з рівнянь системи (10.2.53):

$$\dot{\rho} - \frac{\alpha}{2}\rho + \frac{3}{2}\gamma\omega_0^2\rho^3 = \frac{\pi}{4\rho}G_\xi + \frac{1}{2\omega_0}\xi_\rho(t).$$

Природно вважати шуми в контурі автогенератора вузькосмуговим випадковим процесом. Тоді початкова амплітуда коливань $\rho_0 = \rho(t=0)$ описується розподілом Рейлі (див. вище п. 9.3.4, формула (9.3.20)):

$$w(\rho_0) = \frac{\rho_0}{\sigma_0^2}\exp\left(-\frac{\rho_0^2}{2\sigma_0^2}\right).$$

В детермінованому режимі (за відсутності шумів) зростання амплітуди коливань описується співвідношенням (див. п. 5.2.5, формула (5.2.24)):

$$\rho(t) = \frac{\rho_\infty}{\sqrt{1+\left(\rho_\infty^2/\rho_0^2 - 1\right)\exp(-\alpha t)}}, \qquad \rho_\infty^2 \approx \frac{\alpha}{3\gamma\omega_0^2}.$$

Знаючи зв'язок величин ρ_0 і $\rho(t)$ та розподіл $w(\rho_0)$, можна відразу записати розподіл $w\left[\rho(t)\right]$, скориставшись співвідношенням (9.1.8) (див. п. 9.1.2).

$$w(y) = w\left(f^{-1}(y)\right)\frac{d\left(f^{-1}(y)\right)}{dy}.$$

В нашому випадку $x \to \rho_0$, $y \to \rho(t)$,

$$\rho_0 = f^{-1}\left[\rho(t)\right] = \frac{\rho_\infty}{\sqrt{1+\left[\frac{\rho_\infty^2}{\rho^2(t)} - 1\right]\exp(\alpha t)}}. \tag{10.2.86}$$

Маємо:

$$w\left[\rho(t)\right] = \frac{2K_0\rho_\infty^2\rho(t)}{\left\{\rho^2(t) + \left[\rho_\infty^2 - \rho^2(t)\right]\exp(\alpha t)\right\}^2}\exp\left\{\alpha t - \frac{K_0\rho^2(t)}{\rho^2(t) + \left[\rho_\infty^2 - \rho^2(t)\right]\exp(\alpha t)}\right\}, \qquad K_0 = \frac{\rho_\infty^2}{2\sigma_0^2}. \tag{10.2.87}$$

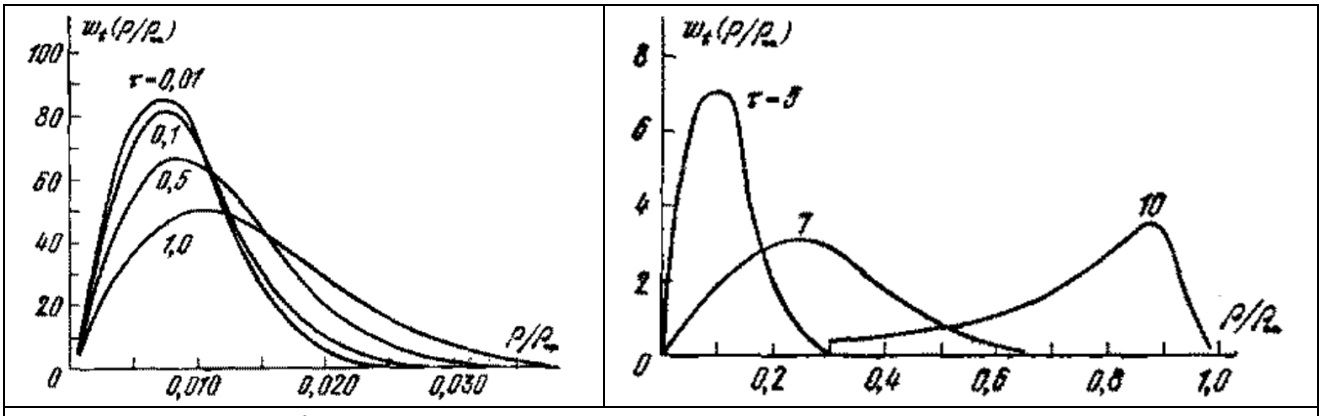


Рис. 10.2.9. Графіки розподілу для амплітуди автоколивань у різні моменти часу при $K_0 = 2\cdot 10^4$

При $\alpha t = 0$ розподіл (10.2.87) переходить у розподіл Рейлі, як і має бути. При $\alpha t \gg 1$ з (10.2.87) отримуємо:

$$w\left[\rho(\alpha t \gg 1)\right] = \frac{\rho_\infty^4\rho\exp(-\alpha t)}{\sigma_0^2\left[\rho_\infty^2 - \rho^2\right]^2}\exp\left\{-\frac{\rho_\infty^2\rho^2\exp(-\alpha t)}{2\sigma_0^2\left[\rho_\infty^2 - \rho^2\right]}\right\}. \tag{10.2.88}$$

При $\alpha t \to \infty$ розподіл (10.2.88) набуває вигляду δ-функції, $w(\rho) = C\delta(\rho - \rho_\infty)$.

Графіки розподілів (10.2.87) для різних значень параметра $\alpha t = \tau$ подані на рис. 10.2.9.

На рис. 10.2.10 подані експериментально виміряні розподіли $w\left[\rho(t)\right]$ для відбивального клістрона. Видно, що вони подібні до розрахованих (рис. 10.2.9).

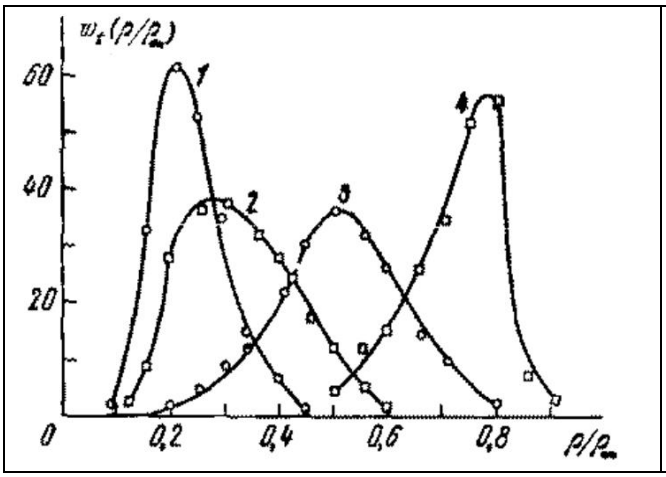


Рис. 10.2.10. Графіки експериментальних розподілів $w\left[\rho(t)\right]$ для відбивального клістрона в моменти часу 520 нс (крива 1), 550 нс (крива 2), 590 нс (крива 3), 660 нс (крива 4)

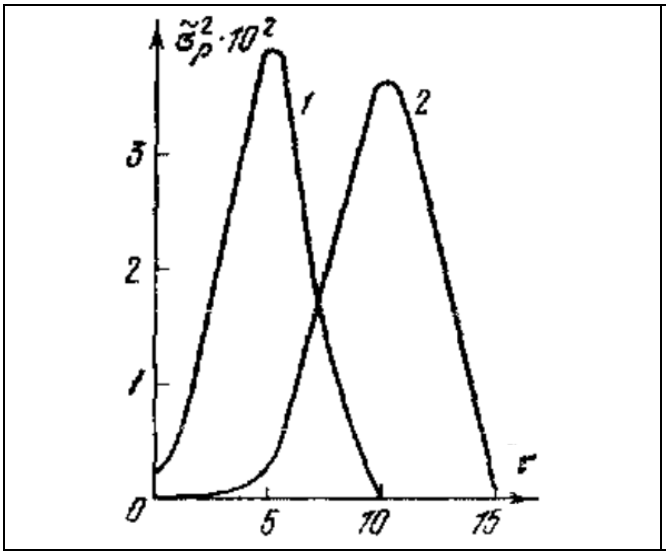


Рис. 10.2.11. Графіки часової залежності дисперсії амплітуди автоколивань, що зростають від шумів:
1 – $K_0 = 200$, 2 – $K_0 = 2 \cdot 10^4$

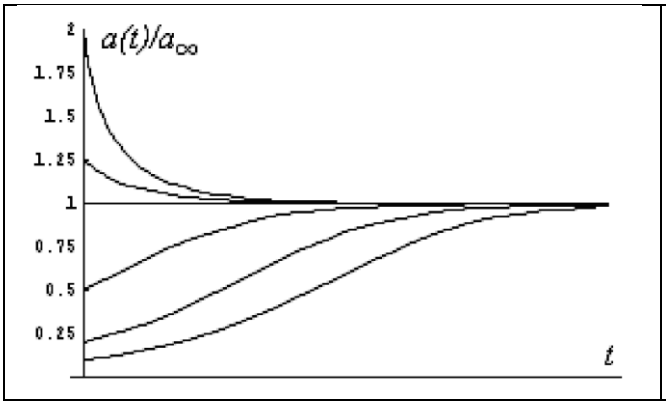


Рис. 5.2.4. Графіки часової залежності амплітуди коливань автогенератора для різних початкових амплітуд

За відомим розподілом (10.2.87) можна підрахувати його моменти $\langle \rho(t) \rangle$ та $\langle \rho^2(t) \rangle$, а потім і дисперсію $\sigma_\rho^2(t) = \langle \rho^2(t) \rangle - \langle \rho(t) \rangle^2$. Відповідні формули дуже громіздкі, тому ми їх не наводимо. Часові залежності $\sigma_\rho^2(t)$ подані на рис. 10.2.11. Видно, що в процесі встанов-

лення коливань дисперсія спершу (при $\alpha t \leq 1$) зростає, а потім (при $\rho \to \rho_\infty$) спадає до нуля (пор. з рис. 5.2.4). Час встановлення коливань зростає зі збільшенням параметра K_0, тобто зі зменшенням рівня власних шумів.

10.2.10. Розподіл для часу встановлення сигналу автогенератора

Ще однією випадковою величиною, яка характеризує процес встановлення коливань автогенератора, є час τ_K досягнення інтенсивності коливань $K\rho_\infty^2$ ($0 < K < 1$). Формально його можна знайти зі співвідношення (5.2.24)

$$\rho(t) = \frac{\rho_\infty}{\sqrt{1 + \left(\rho_\infty^2/\rho_0^2 - 1\right)\exp(-\alpha t)}},$$

замінивши t на τ_K і $\rho^2(t)$ на $K\rho_\infty^2$. Отримаємо:

$$\tau_K = \frac{1}{\alpha}\ln\frac{K\left(\rho_\infty^2 - \rho_0^2\right)}{\rho_0^2(1-K)}. \qquad (10.2.89)$$

Задавши значення параметра K і вважаючи, як і раніше, що початкова амплітуда розподілена за Рейлі (див. формулу (9.3.20)), з рівнянь (9.3.20) та (10.2.89) можна знайти розподіл $w(\tau_K)$:

$$w(\alpha\tau_K) = \frac{K_0(1-K)}{\left[K + (1-K)\exp(\alpha\tau_K)\right]^2}\exp\left\{\alpha\tau_K - \frac{K_0 K}{K + (1-K)\exp(\alpha\tau_K)}\right\}. \qquad (10.2.90)$$

Сім'я залежностей $w(\alpha\tau_K)$ для різних значень K подана на рис. 10.2.12. Як і слід було чекати, час τ_K збільшується зі зростанням K.

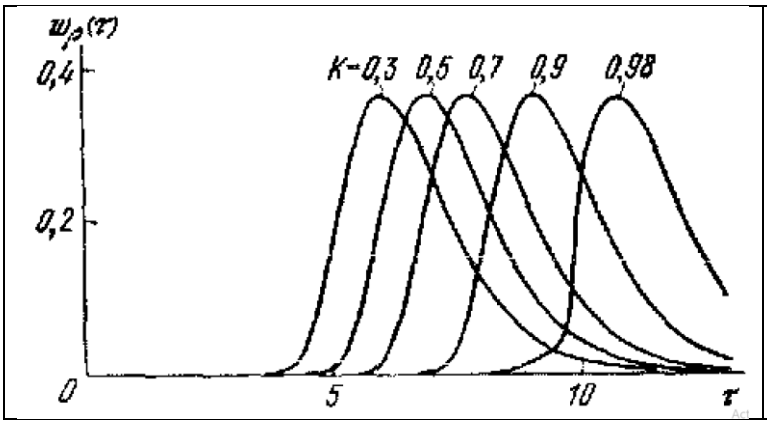


Рис. 10.2.12. Графіки розподілів часу встановлення інтенсивності коливань на рівні K від її усталеного значення

Контрольні питання до підрозділу 10.2

1. Які фактори визначають технічну ширину лінії автогенератора?
2. Що розуміють під природною шириною лінії автогенератора? Які ефекти її визначають?
3. Поясніть відмінність між інтенсивностями усталених коливань автогенератора без урахування та з урахуванням його власних шумів. Коли така відмінність максимальна?
4. В чому полягає метод знаходження форми та природної ширини лінії автогенератора?
5. Чому процес встановлення коливань автогенератора має стохастичний характер?
6. Як змінюється дисперсія коливань автогенератора в процесі їхнього встановлення?

Задачі до підрозділу 10.2

10.2.1. Знайдіть сталу нормування стаціонарного розподілу $w(\rho)$, який встановлюється в автогенераторі Ван-дер-Поля (див. формулу (10.2.57)).

10.2.2. Побудуйте сім‘ю залежностей нормованого часу $\alpha\tau_K$ досягнення інтенсивності коливань $K\rho_\infty^2$ від параметра ρ_0/ρ_∞ ($(\rho_0/\rho_\infty)^2$<K) для різних значень K ($0 < K < 1$).

10.3. Генератор шуму Кияшка – Піковського – Рабиновича

За певних умов поведінка досить простої нелінійної системи може стати непередбачуваною. Для того, щоб стохастична поведінка реалізувалася в дисипативній системі за відсутності дії зовнішньої періодичної сили, необхідно, щоб ця система була відкритою (тобто могла отримувати енергію ззовні). Іншими словами, це має бути деякий аналог автогенератора, причому він повинен мати принаймні півтора ступеня вільності[1].

Напевне, найпростішим прикладом автономних систем, здатних демонструвати непередбачувану поведінку, є так званий генератор шуму Кияшка – Піковського – Рабіновича (генератор шуму КПР). Його роботу ми й розглянемо.

Генератори шуму різноманітних конструкцій знаходять широке застосування в численних областях сучасної техніки, починаючи, наприклад, з імітації теплового випромінювання і закінчуючи розігрівом плазми в установках для реалізації керованого термоядерного синтезу.

10.3.1. Непередбачувана поведінка простих систем

Про стохастичну динаміку у фізиці зазвичай кажуть тоді, коли розглядаються системи, складені з великої кількості елементів. Наприклад, при аналізі руху газу або рідини немає можливості записати та розв'язати набір рівнянь руху для кожної молекули або атома через їх велику кількість. Тому вводять деякі середні величини (тиск, температуру, швидкість спрямованого руху та ін.), якими характеризують систему як ціле. Непередбачувані відхилення величин (напр. тиску) від їхніх середніх значень – флуктуації – пояснюються тим, що при такому описі мікроскопічні ефекти не враховуються.

Коли ж ідеться про порівняно прості системи з невеликою кількістю ступенів вільності, то донедавна прийнято було вважати, що поведінка таких систем (принаймні, коли не йдеться про квантову механіку, де всі процеси мають принципово ймовірнісний характер) є цілком передбачуваною.

Насправді останнє міркування є справедливим лише для стійких систем, у яких сусідні зображувальні точки у фазовому просторі (див. вище п. 5.2.6) з часом не розбігаються. Коли ж система є нестійкою, ситуація змінюється. Початкові умови завжди відомі нам з певною скінченною точністю. Поведінку системи можна вважати передбачуваною лише тоді, коли початкова невизначеність із часом не зростає. Проте для нестійких систем (напр., лінійний осцилятор з від'ємним квадратом частоти або з від'ємною дисипацією, рис. 10.3.1) ця умова порушується. У результаті невизначеність із часом зростає й може стати як завгодно великою.

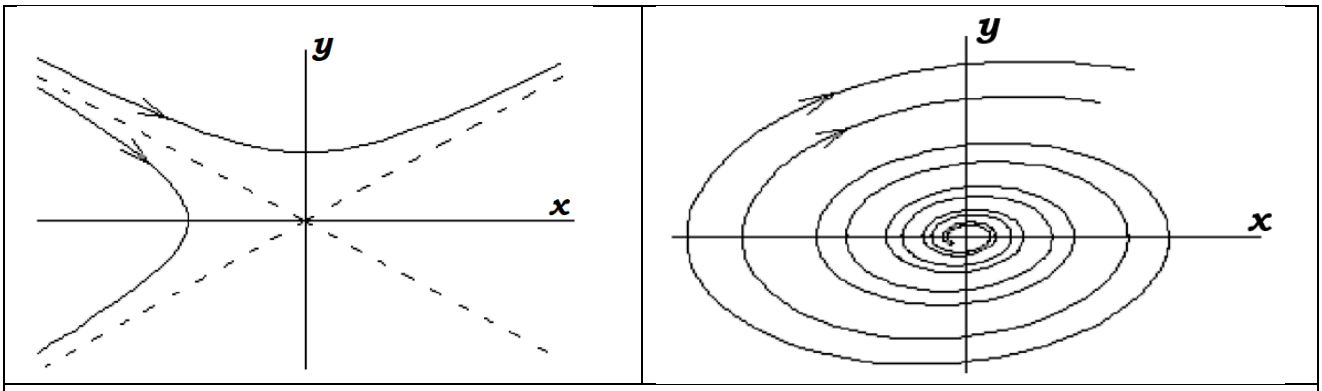


Рис. 10.3.1. Розбігання сусідніх зображувальних точок: а) лінійний осцилятор з від'ємним квадратом частоти, первісно точки розташовані по різні боки від сепаратриси; б) лінійний осцилятор з від'ємною дисипацією

[1] Тобто описуватися системою звичайних диференціальних рівнянь вигляду $dx_i/dt = f(x_1, x_2, \ldots x_n)$, де i змінюється від *1* до *n*, а $n \geq 3$. Детальніше про кількість ступенів вільності див., наприклад, [16].

Зазначимо, що причиною розбігання зображувальних точок для маятника з від'ємним квадратом частоти (рис. 10.3.1, *а*) є розташування цих точок у початковий момент часу по різні боки від сепаратриси (про сепаратриси див. вище п. 5.2.6), а для маятника з від'ємною дисипацією (рис. 10.3.1, *б*) – обумовлене від'ємною дисипацією зростання амплітуди коливань із часом. Проте в обох щойно розглянутих моделях рух системи є інфінітним (необмеженим), що, очевидно, не відповідає реальним ситуаціям.

Для того, щоб можна було казати про стохастичну поведінку системи й описувати її в термінах теорії ймовірності, необхідно, щоб вона характеризувалася деякими середніми величинами, а для цього її рух має бути фінітним. Поєднання нестійкості з фінітністю руху (або, що те саме, обмеженістю області фазового простору, у якій відбувається еволюція системи) і приводить до непередбачуваної поведінки.

Оскільки фазові траєкторії не можуть перетинатися, на фазовій площині (тобто для системи з одним ступенем вільності) поєднання нестійкості та обмеженості руху виявляється неможливим. Для цього необхідний принаймні тривимірний фазовий простір, тобто система щонайменше з півтора ступенями вільності. Один із найпростіших прикладів таких систем – це деякі особливі режими коливань консервативного нелінійного осцилятора[2] під дією зовнішньої періодичної сили. Перейдемо безпосередньо до розгляду цієї задачі.

10.3.2. Генератор шуму КПР: схема та рівняння руху

Схему генератора шуму КПР наведено на рис. 10.3.2, *а*. Вона нагадує схему звичайного генератора Ван-дер-Поля з контуром у колі затвору, але в цей контур додатково ввімкнений тунельний діод з N-подібною вольт-амперною характеристикою $I_D(V)$ (рис. 10.3.2, *б*). Остання має дві зростаючі ділянки – *A* та *B*, розділені спадною ділянкою (стани, що їй відповідають, є нестійкими і не реалізуються). Омічний опір діода на ділянці *A* (в області від'ємних і малих додатних напруг) можна вважати невеликим, на початку ж ділянки *B* він значно зростає. При дуже великих додатних напругах на ділянці *B* омічний опір знову зменшується.

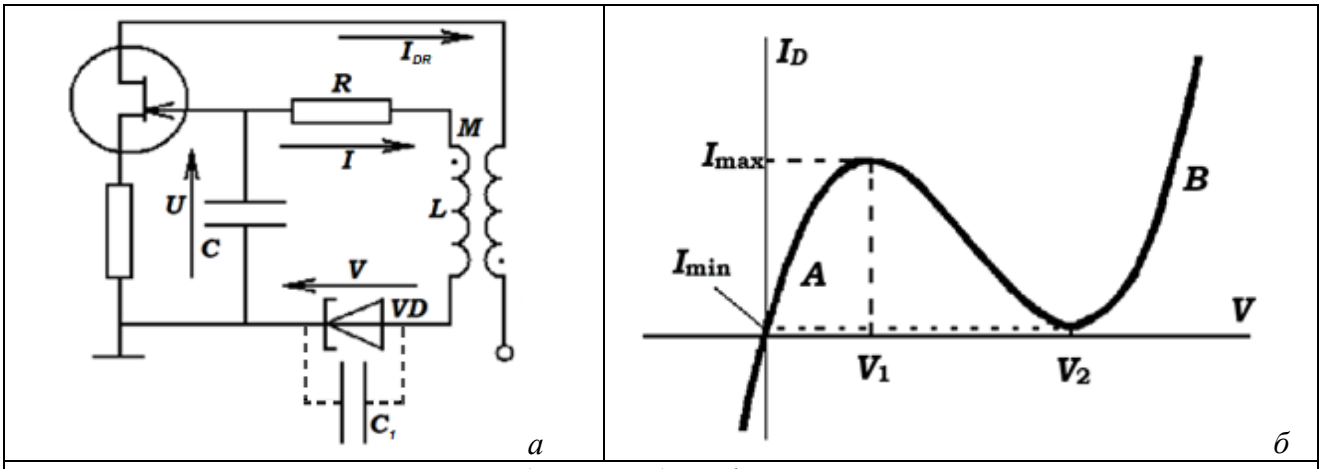


Рис. 10.3.2. Генератора КПР: а) схема; б) графік вольт-амперної характеристики тунельного діода

Окрім тунельного діода, схема має ще один нелінійний елемент – польовий транзистор[3] з прохідною характеристикою, яку ми, як і раніше, апроксимуватимемо кубічним поліномом:

$$I_{DR} = SU - \frac{SU^3}{3K^2}.$$

---

[2] Формально рух такої системи буде відбуватися в обмеженій області фазового простору, якщо вісь t замкнути в кільце, довжина якого дорівнює, напр., періоду зовнішньої сили.

[3] Замість польового транзистора в перших експериментах фігурував тріод.

Складемо систему диференціальних рівнянь, що описує роботу генератора. Позначивши через I струм у коливному контурі, а через U і V – падіння напруги відповідно на конденсаторі та на тунельному діоді (рис. 10.3.2, *а*), можна записати рівняння за другим правилом Кірхгофа для контуру (з урахуванням взаємної індукції *M* та апроксимації) у формі

$$L\frac{dI}{dt}+RI-M\frac{d}{dt}\left(SU-\frac{SU^3}{3K^2}\right)+U+V=0 \qquad (10.3.1)$$

(пор. з виведенням рівняння Рейлі, п. 5.2.1).

Диференціюючи рівняння, що пов'язує між собою заряд конденсатора та напругу на ньому, дістанемо:

$$C\frac{dU}{dt}=I. \qquad (10.3.2)$$

Врахуємо, що $p-n$-перехід тунельного діода має деяку ємність C_1, яку ми вважатимемо малою ($\varepsilon\equiv C_1/C<<1$), тому повний струм через цей діод складатиметься зі струму провідності $I_D(V)$ і струму зміщення $C_1\,dV/dt$:

$$I=I_D(V)+C_1\frac{dV}{dt}. \qquad (10.3.3)$$

Виключимо з (10.3.1) похідну dU/dt за допомогою (10.3.2) і перейдемо до безрозмірних змінних:

$$\tau=\omega_0 t, \qquad u=\frac{U}{V_1}, \qquad \upsilon=\frac{V}{V_1}, \qquad i=\frac{\rho I}{V_1}, \qquad (10.3.4)$$

де $\omega_0^2=1/LC$, $\rho^2=L/C$, величина V_1 показана на рис. 10.3.2, *б*. Тоді рівняння (10.3.1) – (10.3.3) можна переписати у вигляді

$$\frac{di}{d\tau}=\alpha i\left[1-2\left(\frac{u}{u_\infty}\right)^2\right]-\upsilon-u; \qquad \frac{du}{d\tau}=i; \qquad \varepsilon\frac{d\upsilon}{d\tau}=i-i_D(\upsilon), \qquad (10.3.5)$$

де введено позначення

$$\alpha=\frac{R}{\rho}\left(\frac{MS}{RC}-1\right), \qquad u_\infty^2=\frac{2\alpha K^2}{\omega_0 MSV_1^2}, \qquad i_D(\upsilon)=\frac{\rho I_D(V_1\upsilon)}{V_1}. \qquad (10.3.6)$$

10.3.3. Якісний аналіз рівнянь руху в режимах регулярних коливань

Нехай тунельний діод відсутній (або його опором можна знехтувати). Тоді, поклавши $\upsilon=0$ у першому з рівнянь (10.3.5) і відкинувши останнє, легко отримати систему, що зводиться до рівняння

$$\frac{d^2u}{d\tau^2}-\alpha\frac{du}{d\tau}\left[1-2\left(\frac{u}{u_\infty}\right)^2\right]+u=0, \qquad (10.3.7)$$

тобто до рівняння Ван-дер-Поля. При $\alpha<0$ (коли амплітудну умову самозбудження не виконано) коливання згасатимуть.

За виконання умови $1>>\alpha>0$ амплітуда майже гармонічних коливань, що встановляться в системі, дорівнюватиме

$$u_m=u_\infty. \qquad (10.3.8)$$

Отриманий результат якісно відповідатиме реальній поведінці генератора КПР, якщо ця амплітуда лежатиме на ділянці *A* вольт-амперної характеристики тунельного діода (рис. 10.3.2, *б*), тобто при

$$u_m < \frac{\rho I_{\max}}{V_1}. \qquad (10.3.9)$$

При достатньо великих α, коли буде виконано нерівність

$$u_m >> \frac{\rho I_{\max}}{V_1}, \qquad (10.3.10)$$

немонотонність характеристики тунельного діода стане малопомітною в масштабі u_m, а його омічний опір знову зменшиться, тому у схемі встановляться періодичні автоколивання релаксаційного типу. Цей результат підтверджується числовим моделюванням.

За виконання умови

$$u_m \geq \frac{\rho I_{\max}}{V_1} \qquad (10.3.11)$$

немонотонність вольт-амперної характеристики діода стає принциповою, і для аналізу поведінки генератора КПР слід користуватися повною системою рівнянь (10.3.5).

10.3.4. Фазовий портрет стохастичних коливань

Система (10.3.5) містить три диференціальних рівняння першого порядку з незалежними від часу правими частинами і, отже, описує систему з півтора ступенями вільності. Вона містить малий параметр ε при похідній в одному з рівнянь, тому її розв'язок можна будувати шляхом виділення ділянок швидкого та повільного руху.

Поклавши в останньому з рівнянь (10.3.5) $\varepsilon = 0$, отримаємо, що траєкторії повільного руху лежать на поверхні $i = i_D(v)$ (рис. 10.3.3, *а*). Область *A* цієї поверхні відповідає першій зростаючій ділянці (*A*) характеристики тунельного діода (рис. 10.3.2, *б*), а область *B* – іншій зростаючій ділянці характеристики (*B*).

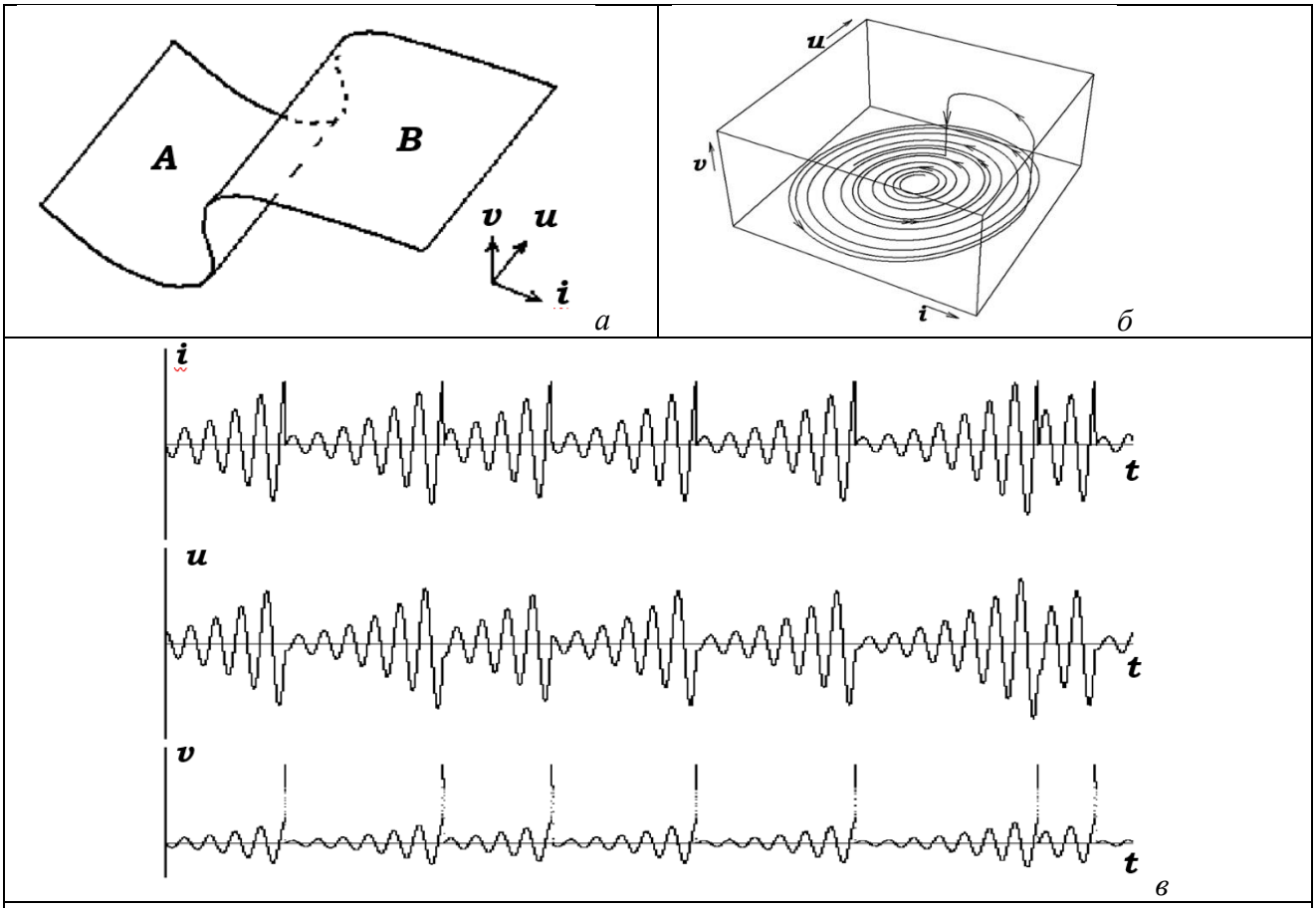


Рис. 10.3.3. Стохастичний режим генератора КПР: а) поверхня повільного руху; б) фазовий портрет; в) епюри коливань

Траєкторії швидкого руху – це прямі, паралельні до осі υ. Справді, розділивши третє рівняння (10.3.5) на перше та друге, отримаємо відповідно, що за межами поверхні повільного руху

$$\frac{d\upsilon}{di} \sim \frac{d\upsilon}{du} \sim \frac{1}{\varepsilon},$$

тобто ці похідні прямують до нескінченності при $\varepsilon \to 0$.

За виконання умови (10.3.11), коли робоча точка розташована на ділянці *A* вольт-амперної характеристики діода (рис. 10.3.2, *б*), а зображувальна точка – відповідно на поверхні *A* повільного руху (рис. 10.3.3, *а*), в автогенераторі виконано амплітудну умову самозбудження, і фазова траєкторія на поверхні *A* являє собою спіраль, що розкручується. Згідно з умовою (10.3.11) нелінійність транзистора на межі поверхні *A* ще не обмежує зростання коливань. Тому, коли зображувальна точка потрапляє на край поверхні *A*, вона опиняється на траєкторії швидкого руху й перестрибує на ділянку *B* поверхні повільного руху (яка відповідає ділянці *B* характеристики діода). При цьому омічний опір тунельного діода різко зростає, амплітудна умова генерації порушується, і в контурі починається згасання коливань в аперіодичному режимі. У результаті зображувальна точка рухається по поверхні *B* у напрямку до початку координат. Коли вона доходить до краю поверхні *B*, вона знову опиняється на траєкторії швидкого руху й перестрибує на поверхню *A*. Слід краю поверхні *B* на поверхні *A*, куди потрапить зображувальна точка, являє собою деякий відрізок, що містить континуум точок. Тому ймовірність того, що зображувальна точка може потрапити на свою попередню фазову траєкторію, дорівнює нулеві. У результаті фазова траєкторія виявляється незамкненою[4]. Це означає, що коливання в контурі починають зростати з деякою іншою початковою амплітудою та фазою, і цикл коливань до наступного стрибка не буде повторювати попередній.

Фазовий портрет, що відповідає описаному режиму, подано на рис. 10.3.3, *б*, епюри коливань – на рис. 10.3.3, *в*. Видно, що величини i та u змінюються з часом неперервно, а величина υ зазнає стрибків, що узгоджується із фазовим портретом.

Зрозуміло, що фазова траєкторія на ділянці зростання коливань (на поверхні *A*) є нестійкою, чим і обумовлена чутливість поведінки системи до малої зміни початкових умов. Тому динаміка системи на проміжках часу порядку тривалості кількох пачок коливань буде непередбачуваною.

Описаний режим стохастичних коливань генератора КПР спостерігався в лабораторному експерименті.

10.3.5. Біфуркації генератора КПР при зміні керувального параметра

На рис. 10.3.4, *а–г* показано зміни фазового портрета системи (10.3.5) при зміні параметра α. В експерименті цей параметр можна змінювати, наприклад, шляхом зміни взаємної індуктивності M котушок, що забезпечують позитивний зворотний зв'язок.

Як уже вказувалося, при малих α коливання в генераторі згасають (рис. 10.3.4, *а*). При цьому початок відліку (на поверхні повільного руху *A*) являтиме собою стійкий фокус. При поступовому зростанні α на місці стійкого фокуса виникає нестійкий, оточений стійким граничним циклом (рис. 10.3.4, *б*).

За виконання умови (10.3.11) граничний цикл, фігурально висловлюючись, вже не вміщується на поверхні *A* повільного руху, і зображувальна точка перестрибує на поверхню *B*. В результаті коливання, як було зазначено, стають хаотичними (рис. 10.3.4, *в*).

За дуже великих значень α, коли уявні розміри граничного циклу на поверхні А стають значно більшими за розміри ділянки немонотонності вольт-амперної характеристики діода, стохастичний режим зникає, і система стрибкоподібно переходить у режим регулярних релаксаційних коливань. Фазовий портрет системи для цього випадку подано на рис. 10.3.4, *г*, епюри коливань – на рис. 10.3.4, *д*.

---

[4] Наведене пояснення, звичайно, не є строгим. Незамкненість фазових траєкторій генератора КПР у певних режимах доводиться за допомогою спеціального математичного методу – т. зв. методу точкових відображень.

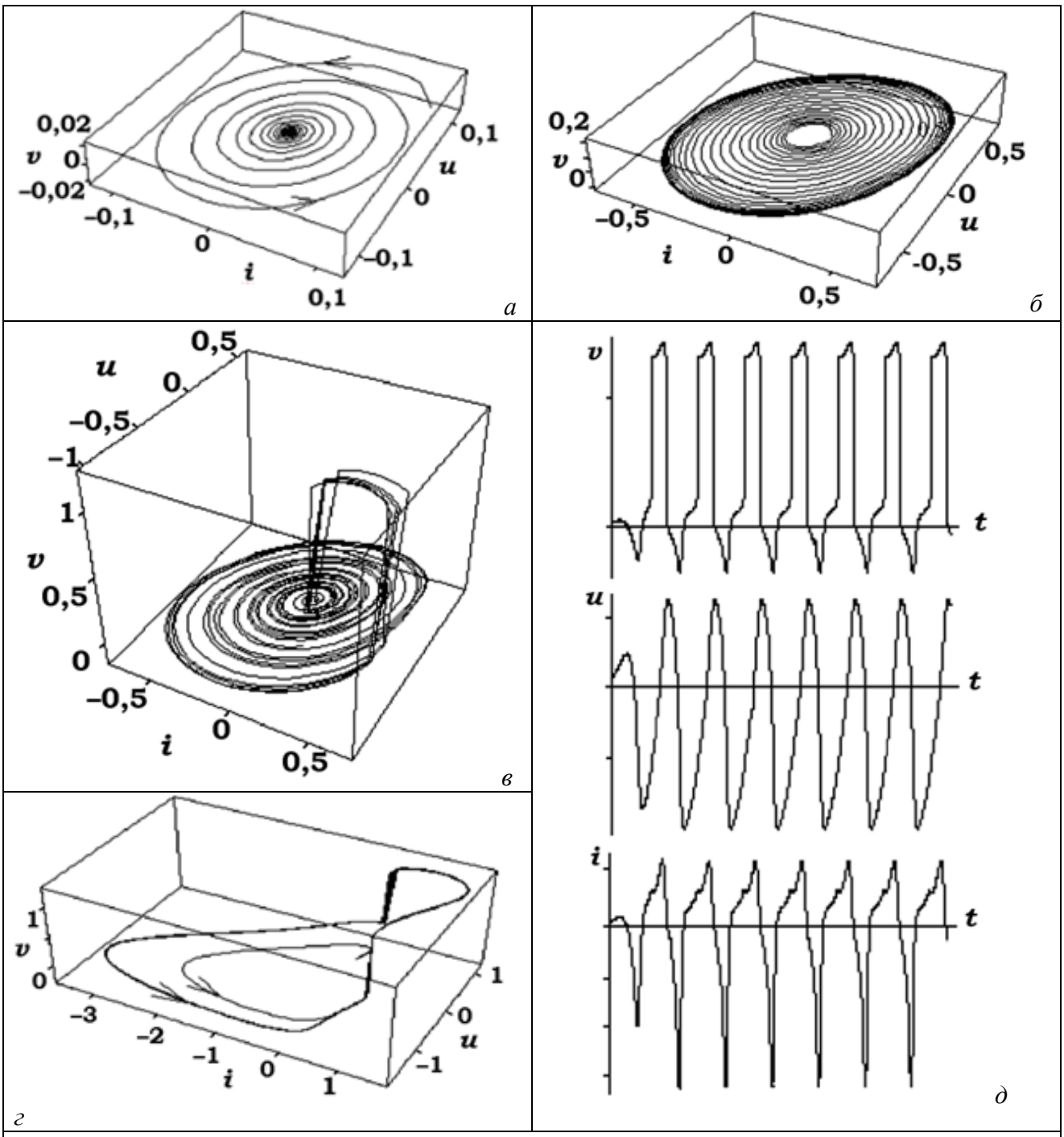


Рис. 10.3.4. Фазові портрети генератора КПР у режимах згасаючих (а), квазігармонічних (б), стохастичних (в) і релаксаційних (г) коливань і епюри релаксаційних коливань (д)

Таким чином, режим хаотичних коливань існує в обмеженій області параметрів генератора КПР. Це типово для стохастичної поведінки динамічних систем.

На закінчення слід вказати, що генератори шуму знаходять застосування в різних галузях науки й техніки. Розробка генераторів шуму із заданими властивостями – одна з цікавих областей сучасної радіоелектроніки.

Контрольні питання до підрозділу 10.3

1. Чому в простих системах може виникнути непередбачувана поведінка?
2. Чим генератор шуму КПР відрізняється від звичайного автогенератора Ван-дер-Поля?

3. Опишіть стохастичні коливання генератора шуму КПР.
4. Чим обмежується зростання коливань у стохастичному режимі генератора шуму КПР?

Задачі до підрозділу 10.3

10.3.1. Яким повинен бути омічний
опір тунельного діода на ділянках А і В його вольт-амперної характеристики, щоб у схемі генератора КПР реалізувався режим стохастичних коливань?

10.3.2. Вольт-амперна характеристика тунельного діода має вигляд $i = av - bv^2 + cv^3$, $a, b, c > 0$. Яка максимальна напруга досягається на тунельному діоді в режимі стохастичних коливань?

10.4. Поняття про електромагнітну сумісність

Досі ми обговорювали внутрішні шуми радіоелектронних схем. Але не менш важливу роль відіграють завади, про які йтиметься в цьому підрозділі.

Як уже зазначалося, в реальних умовах у місці розташування радіоелектронних пристроїв на них діє велика кількість різних (за походженням та властивостями) видів зовнішніх електромагнітних сигналів – так звані завади, англ. hindrance (рис. 10.4.1). В загальному випадку опис завад можливий за допомогою методів теорії випадкових процесів. Радіоелектронні пристрої, по-перше, повинні мати змогу нормально працювати за наявності завад, і, по-друге, їхній вплив не повинен заважати нормально працювати іншим пристроям.

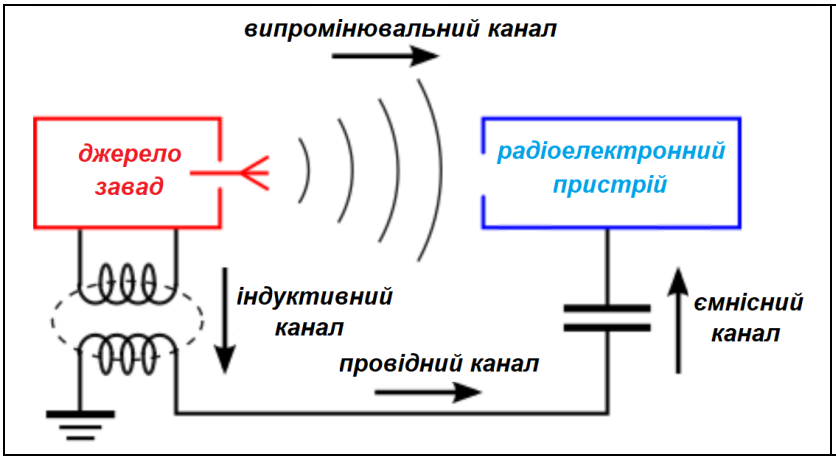


Рис. 10.4.1. Канали впливу зовнішніх завад на радіоелектронні пристрої (умовне зображення)

Під електромагнітною сумісністю якраз і розуміють здатність радіоелектронних пристроїв нормально функціонувати за наявності зовнішніх завад і водночас не створювати неприпустимих завад іншим радіоелектронним пристроям.

Кількісно рівень завад характеризує електромагнітна обстановка (electromagnetic environment) – сукупність електромагнітних сигналів різного походження, наявних у даному місці.

Можливі три способи забезпечення електромагнітної сумісності: придушення завад безпосередньо в їхньому джерелі, створення пристроїв, нечутливих до завад, і мінімізація передачі завад через можливі канали зв’язку між їхнім джерелом і пристроєм.

10.4.1. Нормативні документи, що визначають вимоги до електромагнітної сумісності

В країнах Європейського Союзу чинні вимоги до електрообладнання та електроприладів, які стосуються електромагнітної сумісності, визначаються Директивою 2014/30/EU, виданою 26 лютого 2014 року на заміну попередньої директиви 2004/108/EU. Згідно Директиви, електроприлади і електрообладнання повинні бути спроектовані, вироблені та встановлені таким чином, щоб їхня робота не створювала електромагнітних завад, здатних порушити роботу іншого електрообладнання; водночас має бути забезпечена можливість нормально працювати в існуючій електромагнітній обстановці. Ця директива не встановлює будь-яких конкретних значень допустимого рівня електромагнітних завад; вони встановлюються відповідними гармонізованими стандартами.

Виробникам, які надають свою продукцію на європейський ринок, необхідно забезпечити її відповідність названій Директиві. Повна відповідність вимогам Директиви 2014/30/EU дає право на маркування СЄ на відповідних виробах, їхній випуск і вільний обіг на європейському ринку електрообладнання. Маркування CE (скорочення французьких слів Conformité Européenne — «європейська відповідність») – це спеціальний знак (рис. 10.4.2), який наноситься на виріб і засвідчує, що цей виріб відповідає вимогам директив ЄС і гармо-

нізованим стандартам ЄС та пройшов процедуру оцінки відповідності директивам.

CE	*Рис. 10.4.2. Маркування CE*

10.4.2. Електромагнітні завади природного походження

Як уже вказувалося (див. початок розділу 9), електромагнітні завади за походженням можна поділити на природні й штучні (техногенні).

Природні завади можуть бути земні (зокрема, атмосферні) та позаземні – випромінювання Сонця, зорь та інших космічних об'єктів.

До атмосферних механізмів радіовипромінювання належить, зокрема, електромагнітне випромінювання блискавок. В атмосфері Землі в середньому за 1с б'ють 100 блискавок. Радіовипромінювання блискавок може поширюватися на віддалі в десятки тисяч кілометрів (декілька радіусів Землі).

Ще один механізм радіовипромінювання Землі пов'язаний з існуванням радіаційних поясів – областей магнітосфери Землі, де утримуються заряджені частинки (рис. 10.4.3). Магнітні бурі (збурення магнітосфери, зумовлені сплесками сонячного вітру) можуть порушити умови утримання, і тоді в області магнітних полюсів з'являються потоки електронів, які висипаються з радіаційних поясів уздовж магнітних силових ліній. Ці потоки служать одним з джерел радіовипромінювання (і джерелом полярних сяйв). Вважається, що саме з цим механізмом пов'язане авроральне[5] кілометрове випромінювання.

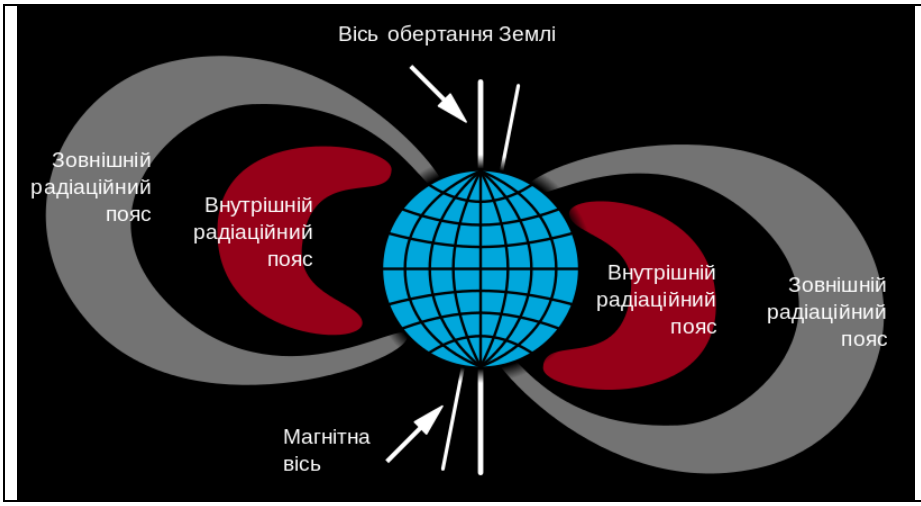


Рис. 10.4.3. Радіаційні пояси Землі: внутрішній, заповнений здебільшого протонами, та зовнішній, заповнений в основному високоенергетичними електронами

Внесок до радіовипромінювання Землі дає також явище радіовипромінювання широких атмосферних злив космічних променів. Космічні промені – це заряджені частинки (переважно протони) космічного походження, енергія яких досягає 10^{20}еВ (16 Дж). В результаті зіткнень з атомами чи молекулами в атмосфері вони породжують каскад перетворень, який супроводжується виникненням нових високоенергетичних частинок (атмосферна злива). Атмосферні зливи породжують електромагнітне випромінення, яке займає діапазон від γ-квантів до радіохвиль.

У діапазоні високих частот слід брати до уваги теплове випромінювання земної атмосфери, а також прилеглої до точки спостереження частини земної поверхні.

[5] Авроральні області – приполярні області атмосфери, де спостерігається полярне сяйво (лат. aurora polaris).

Сонце – найбільш яскраве джерело радіовипромінювання на небі. Випромінювання спокійного Сонця пов’язане з випромінюванням електронів, які зазнають прискорення в електричному полі іонів (так зване гальмівне випромінювання). Лежить переважно в діапазоні довжин хвиль від міліметрів до метрів.

Повільно змінюване радіовипромінювання Сонця пов’язане з сонячними плямами. За своєю природою це випромінювання є гальмівним або магнітогальмівним (зумовленим обертанням електронів навколо силових ліній магнітного поля). Лежить переважно в діапазоні 5-20 см, спостерігається в періоди сонячної активності.

Сплески сонячного радіовипромінювання можуть перевищувати випромінювання спокійного Сонця в мільйони разів. Переважають у метровому діапазоні, хоча інколи спостерігаються навіть у міліметровому діапазоні. Джерелом цього випромінювання є релятивістські частинки, що виникають під час сонячних спалахів. Один з можливих механізмів – магнітогальмівне випромінювання.

Аналогічну природу має радіовипромінювання інших зорь.

Прикладами випромінювання інших космічних об’єктів може бути випромінювання пульсарів (нейтронних зорь[6], які обертаються та мають потужне магнітне поле, і тому є космічними джерелами пульсуючого електромагнітного випромінювання з магнітних полюсів), чорних дір (астрофізичних об'єктів, які створюють настільки велику силу тяжіння, що жодні, навіть найшвидші частинки, не можуть покинути їхню поверхню), міжзоряного середовища тощо.

Таким чином, електромагнітна обстановка, зумовлена завадами природного походження, тісно пов’язана з сонячною активністю.

З точки зору ЕМС вимоги до електроапаратури зводяться до того, що вона повинна бути малочутливою до завад природного походження.

10.4.3. Техногенні завади

Зараз електромагнітна обстановка визначається переважно техногенними джерелами завад (чим, власне, й зумовлена проблема ЕМС).

Серед джерел завад штучного походження можна, в першу чергу, назвати лінії передач та випромінювачі радіохвиль. Це, зокрема, радіорелейні лінії зв’язку (ланцюжки приймально-передавальних станцій), радіомовлення та телебачення, радіонавігаційні та радіолокаційні системи, системи мобільного зв’язку.

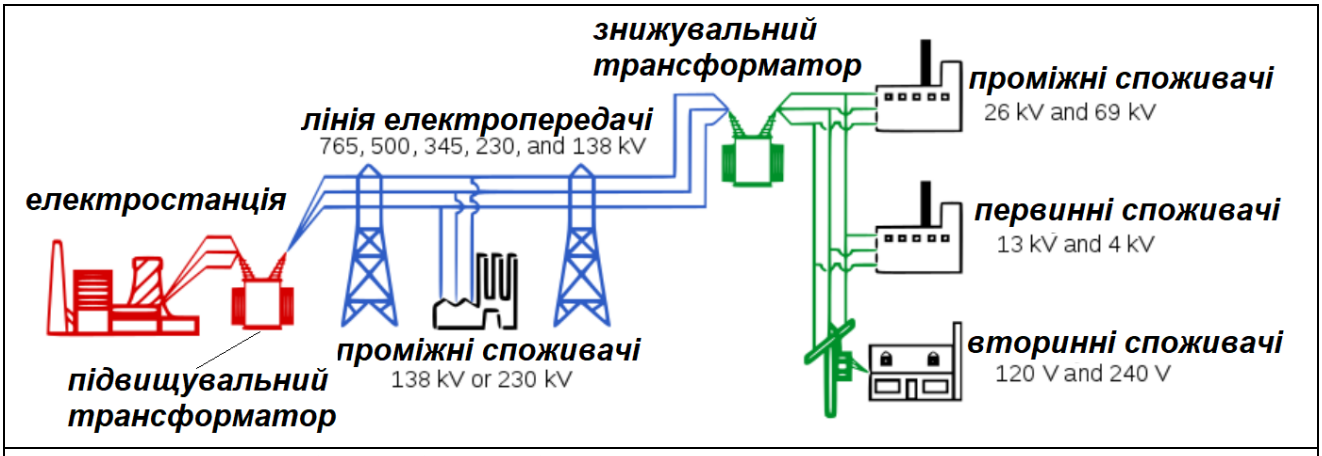


Рис. 10.4.4. Схема забезпечення споживачів електроенергією

[6] Нейтронна зоря - зоря на завершальному етапі своєї еволюції, що не має внутрішніх джерел енергії та складається переважно з нейтронів. Густина такого об'єкта сумірна з густиною атомного ядра.

Джерелами завад є також системи електроживлення (рис. 10.4.4): генератори електростанцій (причиною появи радіовипромінювання є перемикання обмоток, в яких вихрове електричне поле, зумовлене обертанням магніту, наводить е.р.с. – див. рис. 10.4.5), перетворювачі (випрямлячі, стабілізатори, підвищувальні трансформатори; зокрема, в останніх змінне магнітне поле, спричинене змінним струмом, породжує вихрове електричне поле), лінії електропередачі (де протікає змінний струм, в результаті чого вони через свою велику довжину відіграють роль антен), засоби розподілення електроенергії (знижувальні трансформаторні підстанції).

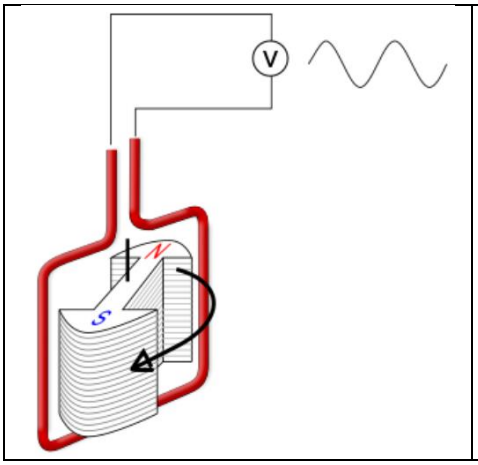


Рис. 10.4.5. Найпростіша схема електричного генератора з магнітом, що обертається. Це обертання створює змінний магнітний потік через нерухому обмотку Нерухомих обмоток може бути багато; тоді вони підключаються до виходу по черзі. Струм через обмотку протікає тільки тоді, коли вона підключена до виходу, тобто він складається з окремих імпульсів

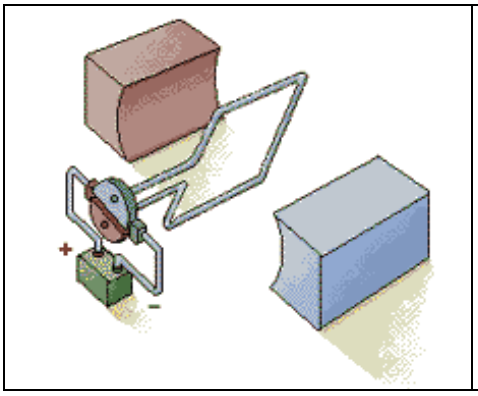


Рис. 10.4.6. Найпростіша схема електродвигуна. Полярність струму в рамці, який створюється джерелом змінної напруги, періодично змінюється, завдяки чому сила Ампера також змінює свій напрямок, і рамка обертається в постійному магнітному полі

Ще одним джерелом завад є промислове обладнання: будь-які установки великої потужності, верстати, конвеєри, зварювальні апарати тощо.
В окрему групу можна виділити двигуни, зокрема, двигуни транспортних засобів. Електродвигуни мають, по суті, таку саму конструкцію, як уже згадані генератори електроенергії (рис. 10.4.6). Двигуни внутрішнього згоряння для запалювання суміші повітря й пального використовують іскровий розряд, який, як і блискавки, є джерелом електромагнітного випромінювання.

Нарешті, слід згадати електроапаратуру загального та побутового призначення: комп'ютери (де протікають імпульсні струми) та відповідні периферійні пристрої, медичну апаратуру (зокрема, ультразвукові пристрої), побутову техніку, засоби освітлення.

Далі ми обговоримо роботу деяких джерел техногенних завад більш детально.

10.4.4. Електромагнітна обстановка

Ми вже користувалися поняттям електромагнітної обстановки. Тепер обговоримо його детальніше.

Отже, електромагнітна обстановка – це сукупність електромагнітних полів у деякій точці простору, яка впливає на якість роботи розташованої там радіоелектронної апаратури

(надалі умовно називатимемо її приймачем).

Розрізняють зовнішню та внутрішню електромагнітну обстановку. Зовнішня обстановка створюється джерелами, розташованими в далекій зоні, які впливають на приймач через електромагнітне випромінювання. Відповідно внутрішня обстановка створюється джерелами, розташованими в ближній зоні, які діють на приймач через індуктивний, ємнісний або провідний канали впливу (див. рис. 10.4.1).

Однією з характеристик зовнішньої електромагнітної обстановки є потужність завад на вході приймача. Вона, очевидно, визначається потужністю джерела завад, втратами в середовищі поширення та в приймачі. Всі ці фактори можна вважати випадковими. Наприклад, втрати при поширенні радіохвиль у середовищі (на поглинання та розсіювання) залежать від чинників, що мають випадковий характер (температура, вологість, атмосферний тиск тощо), а також від характеру та рельєфу місцевості тощо. Потужність джерела завад і втрати в приймачі також можуть зазнавати випадкових змін, спричинених, наприклад, змінами напруги живлення або температури, старінням елементів схем тощо. Нарешті, слід узяти до уваги, що приймач зазвичай піддається одночасній дії кількох джерел завад. Тому для опису завад користуються теорією випадкових процесів.

Сукупність корисних сигналів та завад називають повним груповим сигналом. В реальних умовах груповий сигнал містить багато компонентів. Ті компоненти, що не стосуються корисного сигналу, є завадами. На рівні моделей виділяють завади, зосереджені за спектром (вузькосмуговий випадковий процес) і завади, зосереджені за часом (послідовність імпульсів), які, відповідно, мають широкий спектр. У багатьох випадках спектр завад визначається тривалістю фронтів імпульсів напруги та струму, які генерує джерело завад.

Існує кілька методів теоретичного (розрахункового) визначення електромагнітної обстановки. До найбільш поширених належать електродинамічний метод, енергетичний метод та ймовірнісний метод. Загалом кожен з них має свою область застосування.

- Електродинамічний метод базується на розв'язанні системі рівнянь Максвелла з урахуванням усіх наявних джерел завад. В результаті визначається електромагнітне поле в точці спостереження. Цим методом можна користуватися, коли джерел завад небагато, а відповідні сигнали є детермінованими (невипадковими).

- Енергетичний метод базується на моделях завад, зосереджених за спектром або за часом. Як уже вказувалося, йдеться відповідно про вузькосмугові сигнали та сигнали невеликої тривалості. Зокрема, за наявності кількох джерел вузькосмугових завад, спектри яких не перекриваються, інтерференція сигналів від різних джерел несуттєва, і сумарна спектральні інтенсивність знаходиться як сума спектральних інтенсивностей окремих джерел.

- Нарешті, ймовірнісний метод базується на визначенні ймовірностей відхилень від обумовлених рівнів реального сигналу, спричинених завадами. При такому підході основну роль відіграють випадкові джерела завад.

10.4.5. Радіопередавальні пристрої як джерела завад

Радіопередавальні пристрої (передавачі), крім корисних сигналів, випромінюють ряд інших сигналів, які можна розглядати як завади.

Корисний сигнал – це сигнал, промодульований за амплітудою або за фазою (нагадаємо, що саме модуляція переносить корисну інформацію). Він характеризується деякою частотою-носієм ω_0 та певною смугою частот $\Delta\omega_0$ навколо неї.

Але через нелінійність елементів схеми передавача спектр його вихідного сигналу містить цілий ряд інших частот, так що реальна ширина спектру $\Delta\omega$, випромінюваного в ефір, виявляється значно ширшою за $\Delta\omega_0$ (рис. 10.4.7). У спектрі випромінювання зазвичай присутні кратні гармоніки $n\omega_0$ та субгармоніки ω_0/n ($n = 1,2,3,\ldots$).

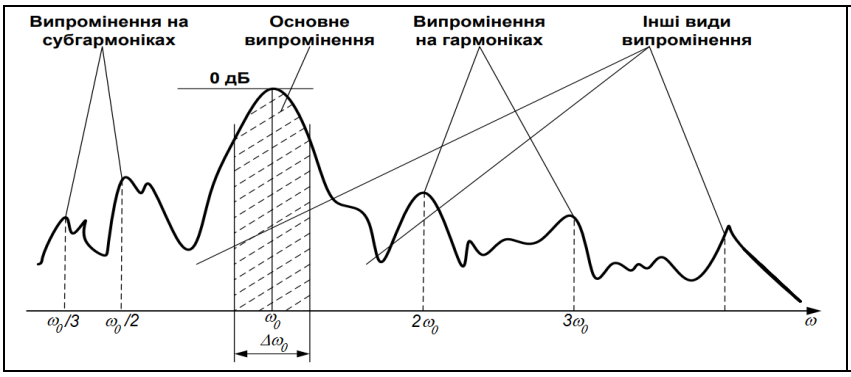


Рис. 10.4.7. Умовний вигляд спектру випромінювання радіопередавача

Додатково до кратних гармонік та субгармонік може з'явитися випромінювання на інтермодуляційних (комбінаційних) частотах $m\omega_1 + n\omega_2$, де $m,n = \pm1, \pm2, \pm3\ldots$), які породжуються в результаті нелінійної взаємодії гармонік ω_1, ω_2 в спектрі сигналу (див. вище п. 5.7.4). Інтермодуляційні частоти можуть виникати про взаємодії не тільки двох, а й більшої кількості гармонік.

Нерідко присутнє також так зване паразитне випромінювання, зумовлене самозбудженням підсилювальних каскадів за рахунок паразитного зворотного зв'язку, та вищі гармоніки відповідної частоти.

Передавач також може бути джерелом шумового випромінювання, зумовленого його внутрішніми шумами. Таке випромінювання зазвичай має невелику амплітуду, але широкий спектр.

Позбутися цих видів завад можна, поставивши на вході приймача смуговий фільтр, який повинен пропускати лише корисний сигнал з середньою частотою ω_0 і шириною спектру $\Delta\omega_0$.

10.4.6. Небажані канали приймання супергетеродинних радіоприймачів

Блок-схема супергетеродинного радіоприймача (див. Вступ, п. В.3) подана на рис. 10.4.8.

Як уже вказувалося, основне підсилення в ньому відбувається на проміжній частоті (intermediate frequency) $\omega_{\text{int}} = \omega_{get} - \omega_0$, де ω_{get} – частота гетеродина, ω_0 – частота-носій сигналу на вході приймача. При перелаштуванні приймача на іншу частоту ω_0 змінюється ω_{get}, так що ω_{int} залишається незмінною. Це дає змогу використовувати для підсилення проміжної частоти складні й досконалі схеми.

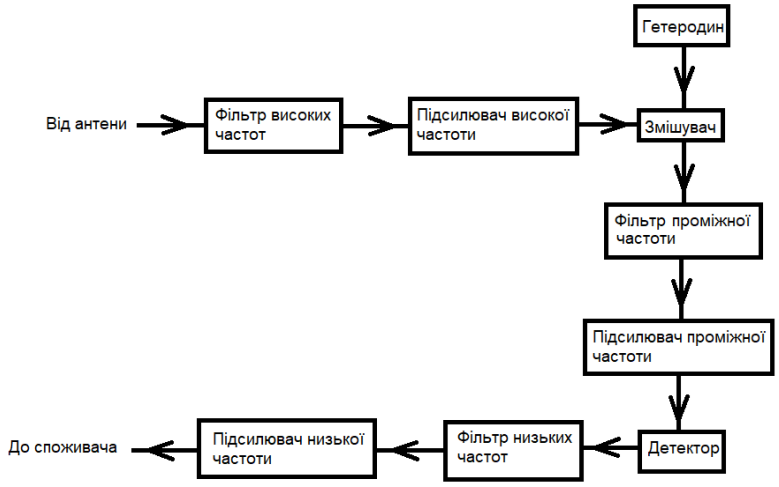


Рис. 10.4.8. Блок-схема супергетеродинного приймача

Недоліком супергетеродинних приймачів є існування так званого дзеркального каналу з середньою частотою ω_0', яка задовольняє умові $\omega_{\text{int}} = \omega_0' - \omega_{get}$. Якщо завада з частотою ω_0' потрапить на вхід змішувача, вона призведе до появи додаткового сигналу з частотою ω_{int} – дзеркальної завади. Смуговий фільтр на вході супергетеродинного приймача якраз і призначений для придушення дзеркальної завади.

Супергетеродинний приймач має й ряд інших небажаних каналів. Перш за все, це комбінаційні частоти, що породжуються вхідним сигналом із середньою частотою ω_0 і сигналом гетеродина з частотою ω_{get}, які задовольняють умові $\omega_{\text{int}} = m\omega_0 + n\omega_{get}$, $m,n = \pm1, \pm2, \pm3, \ldots$ Дзеркальній заваді відповідають значення $m = 1, n = -1$, основному каналу – $m = -1, n = 1$. Амплітуда сигналів з комбінаційними частотами звичайно швидко зменшується зі зростанням m та n.

Інколи нелінійна взаємодія гармонік ω_1 та ω_2 може відбуватися безпосередньо на вході підсилювача. Якщо задовольняється умова $m\omega_1 + n\omega_2 = \omega_0$, $m,n = \pm1, \pm2, \pm3, \ldots$, відповідні вхідні завади породжують гармоніку, яка проходить через підсилювач високої частоти на змішувач і далі обробляється подібно до сигналу по основному каналу.
Нарешті, вхідний сигнал з частотою ω_{int} також може потрапити зі входу в підсилювач проміжної частоти, формуючи ще один небажаний канал.

В супергетеродинних радіоприймачах спостерігається також ефект блокування завадою – зміна режиму роботи підсилювачів під впливом завади великої амплітуди. Це може бути, наприклад, зменшення середньої крутості відповідної прохідної характеристики підсилювача, яке призводить до відповідного зменшення коефіцієнту підсилення (як при вимушеній синхронізації автогенератора зовнішнім сигналом, див. вище п. 5.2.9).

10.4.7. Сполучні провідники як передавачі та приймачі завад

Як уже вказувалося, між пристроями, розташованими в ближній зоні один одного, можливі три типи зв'язку (рис. 10.4.1): провідний, ємнісний та індуктивний.

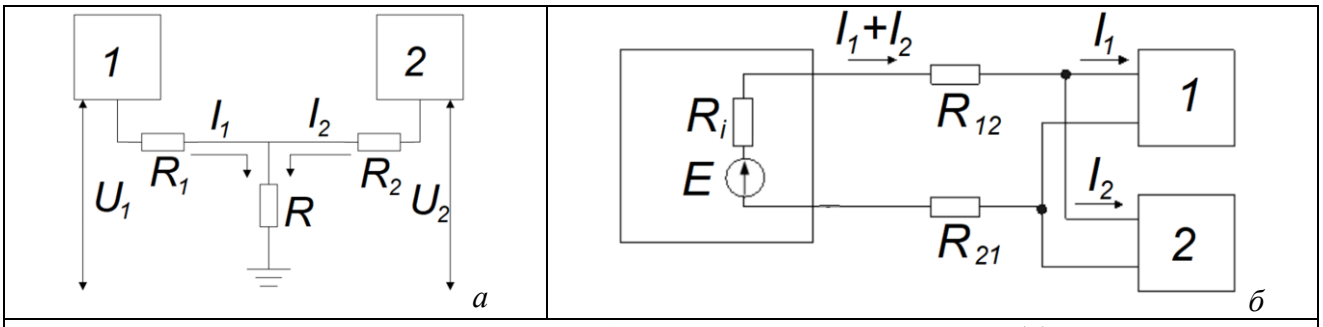


Рис. 10.4.9. Провідний зв'язок пристроїв 1 і 2 через спільне заземлення (а) та через спільне джерело живлення (б)

Провідний (кондуктивний) зв'язок, або зв'язок через спільний опір, виникає, якщо струми кількох пристроїв протікають через той самий активний опір заземлення або внутрішній опір джерела живлення (рис. 10.4.9).

Ємнісний, або електричний зв'язок виникає, коли електричне поле одного пристрою через (паразитну) ємність зв'язку наводить таке саме поле в інших пристроях (рис. 10.4.10 а). Для зменшення ємнісного зв'язку можна екранувати один з провідників, а екран заземлити (рис. 10.4.10 б). В цьому випадку ємності C_{1s} та C_{2s} виявляються значно меншими від ємності C_{12}, а ємність C_{sg} буде, навпаки, дуже великою. Тому рівень завади в другому провіднику,

спричинений ємнісним зв'язком, може бути зменшений принаймні на порядок.

Індуктивний, або магнітний зв'язок виникає, коли змінне магнітне поле, спричинене протіканням відповідного струму в одному пристрої, породжує е.р.с. індукції в інших пристроях (рис. 10.4.10). В даному разі провідний екран з немагнітного матеріалу, заземлений в одній точці, не дозволяє зменшити рівень завади в провіднику 2, оскільки змінний струм в екрані породжує змінне магнітне поле, яке впливає на провідник 2. Тому зменшити магнітний зв'язок можна, лише збільшивши віддаль між пристроями.

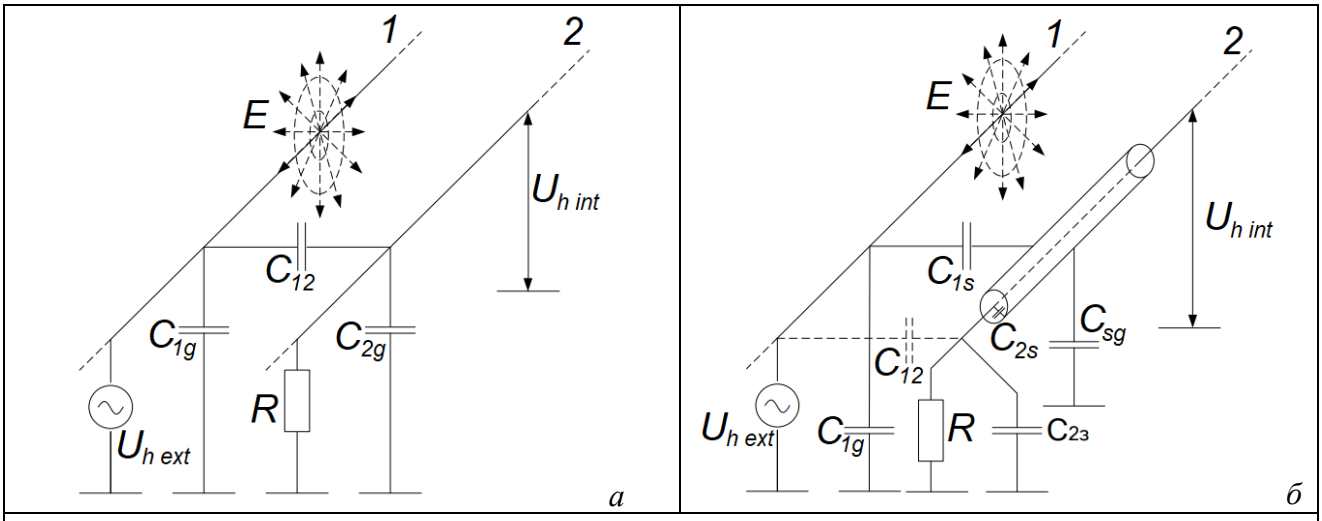


Рис. 10.4.10. Еквівалентна схема ємнісного зв'язку між провідниками 1 і 2 за відсутності (а) та за наявності (б) заземленого екрану: C_{12} – паразитна ємність зв'язку; C_{1g} та C_{2g} – ємності між провідниками 1 і 2 та землею; C_{1s} та C_{2s} – паразитні ємності між провідниками 1 і 2 та екраном; C_{sg} – ємність між екраном та землею; R – опір навантаження провідника 2; $U_{h\ ext}$ та $U_{h\ int}$ – напруги завади в лініях 1 і 2

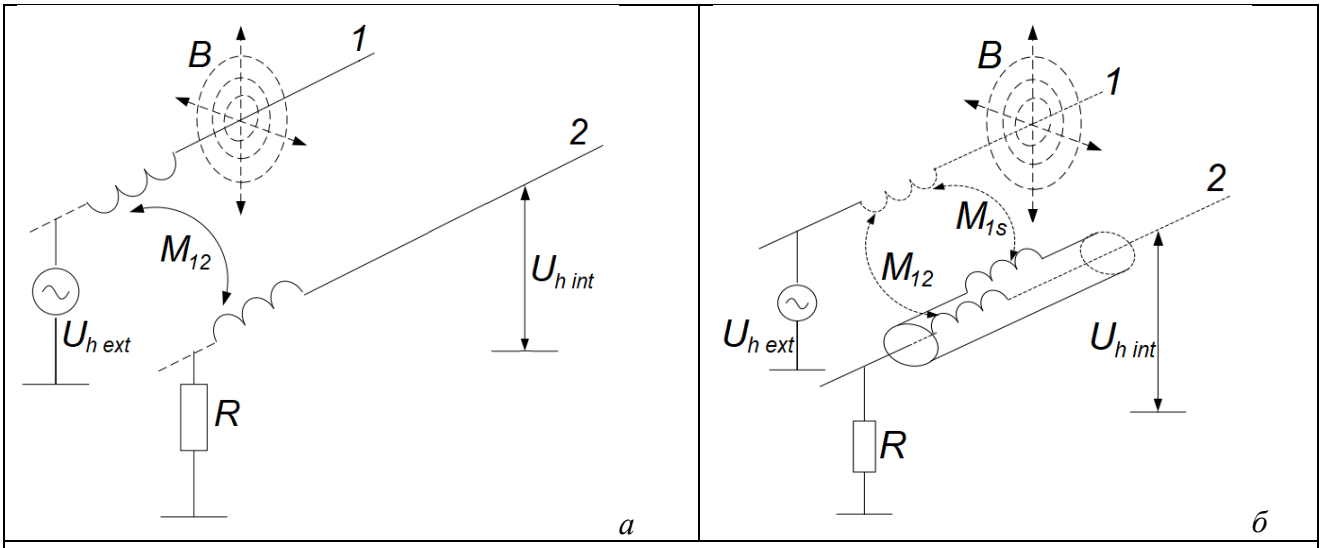


Рис. 10.4.11. Еквівалентна схема індуктивного зв'язку між провідниками 1 і 2 за відсутності (а) та за наявності (б) заземленого екрану

Відзначимо, що екранування та інші методи зменшення рівня завад для інших пристроїв, про які йтиметься далі, в багатьох випадках водночас виступають методами запобігання несанкціонованого витоку інформації.

10.4.8. Особливості приймальних антен, спрямовані на забезпечення електромагнітної сумісності

Як правило, завади сприймаються антенами через задню та бічні пелюстки діаграми спрямованості. Тому основна вимога до антен, виконання якої зменшує вплив завад – це ослаблення задньої та бічних пелюсток. Методи такого ослаблення найбільш розроблені для антен апертурного типу.

Як відомо, діаграми спрямованості антени на передавання та приймання сигналу однакові (див. вище п. 7.4.5). Тому надалі ми говоритимемо про випромінювачі.

Нагадаємо (див. вище п. 7.5.5), що апертурні випромінювачі (наприклад, параболічні антени або рупорні антени) характеризуються протіканням струмів по деякій провідній поверхні (апертурі). Саме ці струми безпосередньо породжують випромінювання. Зазвичай розміри апертурних антен більші (інколи значно більші) за довжину випромінюваних хвиль, тому такі антени використовуються для коротких хвиль. Саме такі антени найчастіше використовуються й у системах супутникового зв'язку.

Низький рівень бічних пелюсток можна забезпечити спаданням амплітудного розподілу опромінення антени до країв апертури (нагадаємо, що струми в апертурі в багатьох випадках збуджуються деяким точковим випромінювачем). Однак це знижує коефіцієнт використання площі антени[7]. Існує можливість суттєво знизити рівень бічних пелюсток і водночас зберегти високий коефіцієнт використання площі. Для цього застосовують спеціальні розподіли поля в апертурі антени:

− напруженість поля має монотонно спадати при віддаленні від центру антени;

− для забезпечення низького рівня бічних пелюсток напруженість поля на межі апертури має бути невеликою;

− кривина амплітудного розподілу опромінення антени в центрі розкриву апертури та на її межі повинна мати різні знаки (рис. 10.4.12).

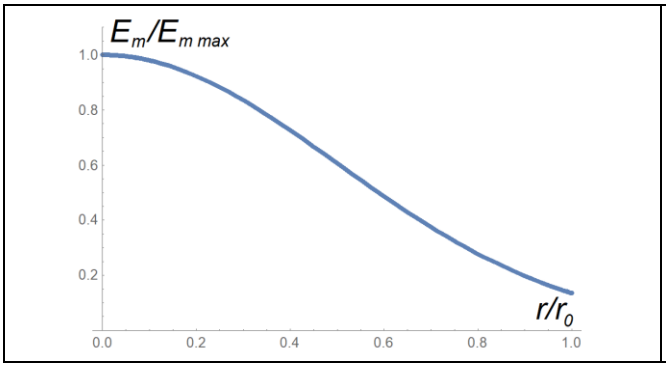


Рис. 10.4.12. Приклад оптимального радіального розподілу електричного поля для аксіально-симетричної апертурної антени

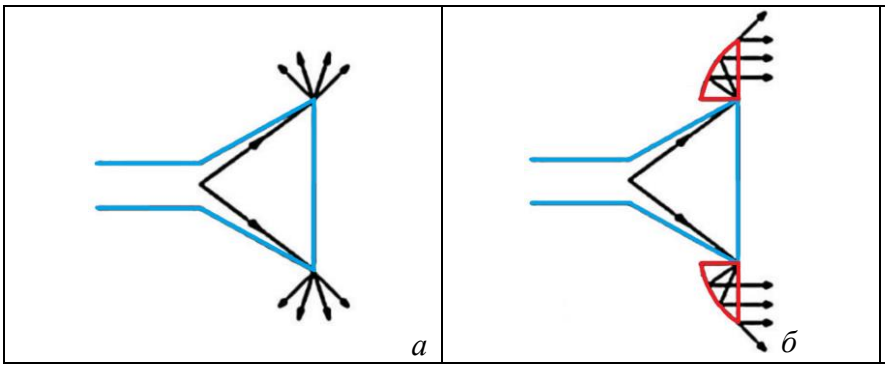


Рис. 10.4.13. Дифракційне випромінювання рупорної антени без екрану (а) та зі спеціальним екраном (б)

[7] Коефіцієнт використання площі антени – це відношення її ефективної площі S_0 до геометричної площі розкриву S. Ефективна площа антени – це відношення потужності P сигналу, який приймає антена, до густини потоку енергії Π хвилі, яка на неї падає: $S_0 = P/\Pi$.

Тепер обговоримо методи придушення бічних пелюсток діаграми спрямованості, зумовлені дифракцією електромагнітних хвиль на межі екрану (рис. 10.4.13 а).

Для рупорної антени це може бути використання спеціальних екранів, які відбивають поле бічних пелюсток (рис. 10.4.13 б). Якщо екрани мають параболічний переріз, відбиті промені будуть спрямовані вздовж осі антени. Однак виготовити екрани з параболічним перерізом складно, тому вони не набули широкого розповсюдження.

Випромінювання задньої пелюстки в параболічних дзеркальних антенах можна зменшити за допомогою екранів конічної форми (рис. 10.4.14). Один екран зменшує її рівень на 8-14 дБ, два екрани – на 20-25 дБ.

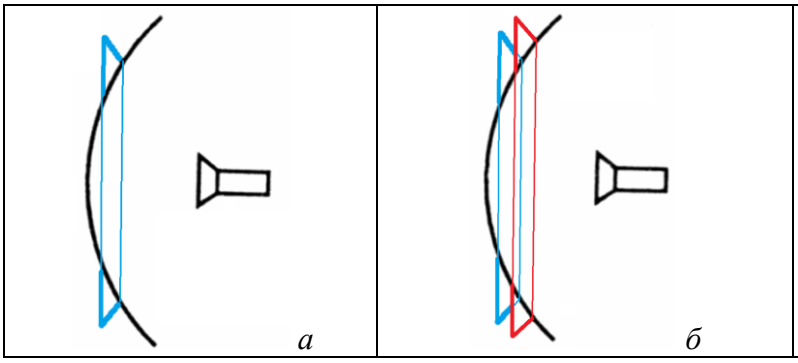


Рис. 10.4.14. Зменшення задньої пелюстки діаграми спрямованості параболічної антени (показано її переріз) за допомогою одного (а) та двох (б) екранів (показані блакитним)

Ще один метод зниження інтенсивності задньої пелюстки полягає в деформації контуру апертури антени. Якщо антена аксіально-симетрична, вся її межа збуджується синфазно. Це обумовлює високий рівень задньої пелюстки. Для зменшення цього рівня край антени деформують (рис. 10.4.15). Це дозволяє зменшити задню пелюстку на (5...6) дБ.

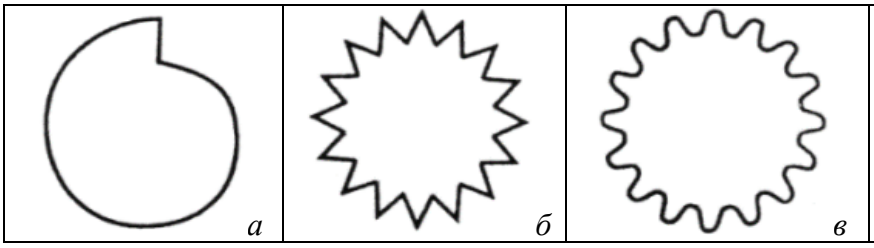


Рис. 10.4.15. Межі антен, що забезпечують ослаблення задньої пелюстки: а – спіральна, б – трикутна, в – пелюсткова

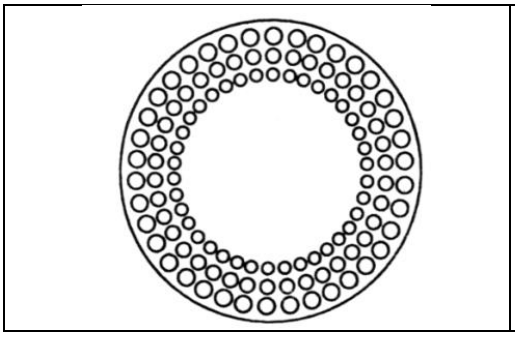


Рис. 10.4.16. Перфорація периферії металевого дзеркала антени для зменшення дифракційного випромінювання

Поля на периферії дзеркала можна зменшити, вкривши цю область радіопоглинальним матеріалом. Однак відсутність ефективних і недорогих поглинальних матеріалів не дозволяє широко застосовувати такі методи.

Легше знизити бічне випромінювання антени, виготовивши її периферійну частину з частково радіопрозорого матеріалу. Такий самий результат дає металеве дзеркало, поверхня якого перфорована. Підібравши розміри отворів та їхню щільність, можна досягти помітного зменшення інтенсивності бічних пелюсток (рис. 10.4.16).

Для зменшення бічного випромінювання застосовують захисні екрани. На рис. 10.4.17 а показано циліндричний екран (бленду), встановлений вздовж контуру антени. Він дозволяє на 5-10 дБ знизити рівень бічних та задньої пелюсток. Довжину екрана підбирають так, щоб рівень збудження його кромки наближався до нуля. Зменшення задньої пелюстки можна реалізувати, якщо фаза струму на межі екрану буде неоднакова. Для цього застосовують скошені бленди (рис. 10.4.17 б).

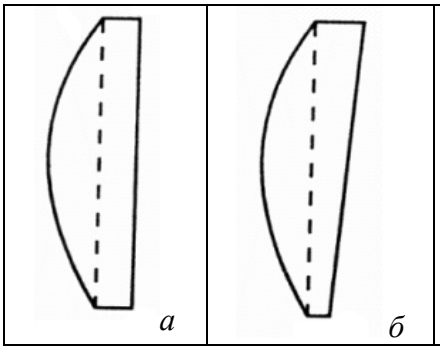


Рис. 10.4.17. Пряма (а) та скошена (б) бленди – циліндричні екрани для ослаблення задньої та бічних пелюсток

Якщо корисний сигнал плоскополяризований, а завада має іншу поляризацію, то рівень останньої можна знизити за допомогою системи паралельних дротин, розміщених із зовнішнього боку антени. Дротини мають бути перпендикулярні електричному полю сигнальної хвилі.

Застосування кільцевих екранів з внутрішнім і зовнішнім радіусами кільця, що дорівнюють відповідно радіусам першої та другої зон Френеля (рис. 10.4.18 а), істотно підсилює головну пелюстку. Інші розміри екрана можуть дати протилежний ефект – придушення випромінювання.

Секторні екрани (рис. 10.4.18 б) забезпечують нульове поле на осі системи. Радіус секторного екрана повинен дорівнювати радіусу першої зони Френеля. Затінити потрібно половину першої зони Френеля. Цей результат має прозоре фізичне пояснення. Якщо сумарний внесок другої та наступних зон Френеля в поле на осі складає $-E_0$, то перша зона дасть $+2E_0$. Якщо затінити половину першої зони, то можна одержати нульове поле на осі.

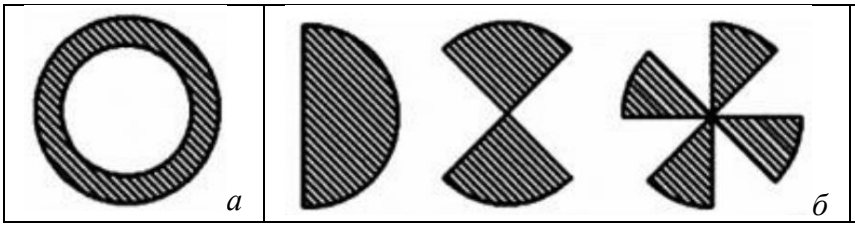


Рис. 10.4.18. Кільцевий (а) та секторні (б) екрани для управління діаграмою спрямованості

10.4.10. Екранування та заземлення

Екран – це провідна конструкція (суцільний металевий лист або сітка), яка відокремлює дві сусідні області. Екран утруднює проникнення електромагнітного поля з однієї області в іншу.

Екран навколо джерела завад поліпшує навколишню електромагнітну обстановку. Екран може також захищати чутливі елементи від зовнішніх полів (рис. 10.4.19).

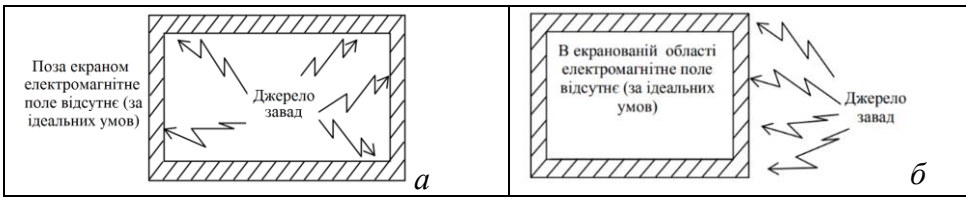


Рис. 10.4.19. Різновиди захисних екранів

Екранувати можна електронні прилади, електричні кола, функціональні вузли, кабелі і системи, які є джерелами або рецепторами завад, або і тим, та іншим водночас.

Для електромагнітної хвилі, яка падає на провідну поверхню, існує два види втрат. Хвиля частково відбивається від поверхні, а заломлена хвиля, яка поширюється в провіднику, згасає внаслідок поглинання (рис. 10.4.20).

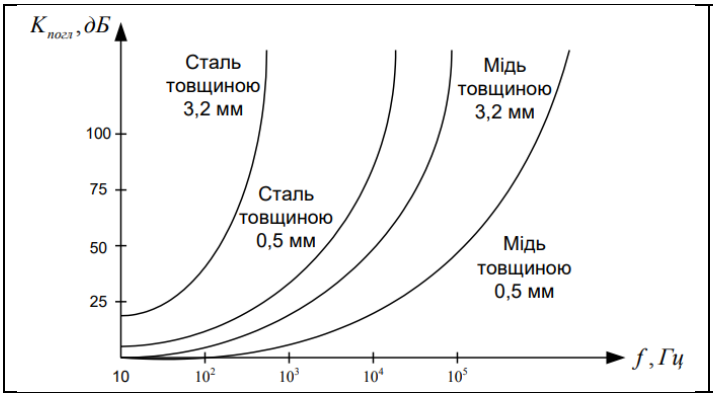


Рис. 10.4.20. Графіки залежності коефіцієнту екранування внаслідок поглинання для сталі та міді від частоти

Ми говорили, що провідні екрани не здатні завадити проникненню магнітного поля. Це частково вдається зробити, якщо екран виготовити з феромагнетику з великою магнітною проникністю. Але треба брати до уваги, що на високих частотах магнітна проникність швидко спадає (рис. 10.4.21).

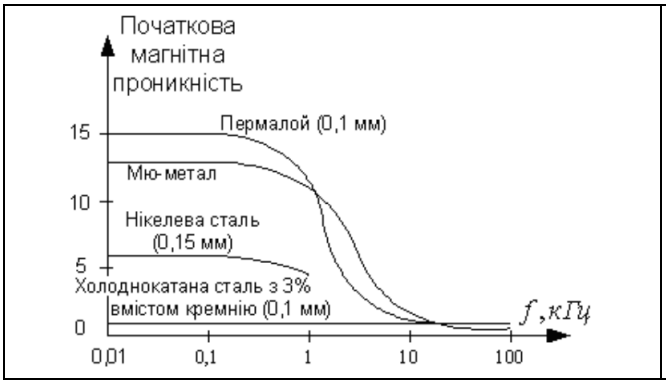


Рис. 10.4.21. Графіки залежності магнітної проникності різних магнетиків від частоти

Вище (п. 10.4.7) згадувалося, що кондуктивний канал впливу завад може бути пов’язаний зі спільним заземленням. У зв’язку з цим обговоримо деякі питання, що стосуються заземлення, детальніше.

Шина заземлення (нульова шина) – це точка, еквіпотенціальна лінія або поверхня, потенціал якої для деякого пристрою чи системи служить рівнем відліку напруги. Приймають, що лінія або поверхня еквіпотенціальна, якщо різниця потенціалів між різними її точками не перевищує:

– для аналогових систем – 100 мВ;

– для цифрових – 200 мВ.

Існує три основні причини заземлення електроапаратури:

- забезпечення електробезпеки;
- забезпечення опорного еквіпотенціального рівня для сигнальних напруг;
- забезпечення електромагнітної сумісності.

Шина захисного заземлення завжди повинна мати потенціал землі; для шини сигнального заземлення ця умова необов'язкова.

Вже згадувалося, що заземлення кількох пристроїв через спільний провідник поро-

джує провідний канал проникнення завад (рис. 10.4.22 а). Для уникнення такого зв’язку використовують системи з окремими провідниками заземлення (рис. 10.4.22 б) або система заземлення в кількох точках (рис. 10.4.22 в). В останньому випадку пристрої підключають до ближньої заземлювальної поверхні (так зване шасі) з малим імпедансом.

З'єднання між кожним з пристроїв та заземлювальною поверхнею треба робити якнайкоротшими, щоб зменшити їхній імпеданс.

На низьких частотах заземлення в кількох точках треба уникати, оскільки зворотні струми усіх кіл протікають крізь загальний імпеданс заземлювальної лінії чи поверхні. На частотах до 1МГц застосовують систему заземлення в одній точці; вище 10 МГц найкращі результати забезпечує заземлення в кількох точках. У смузі 1-10 МГц можна застосовувати заземлення в одній точці, якщо найдовший провідник заземлення коротший від 1/20 довжини хвилі (0.05 λ).

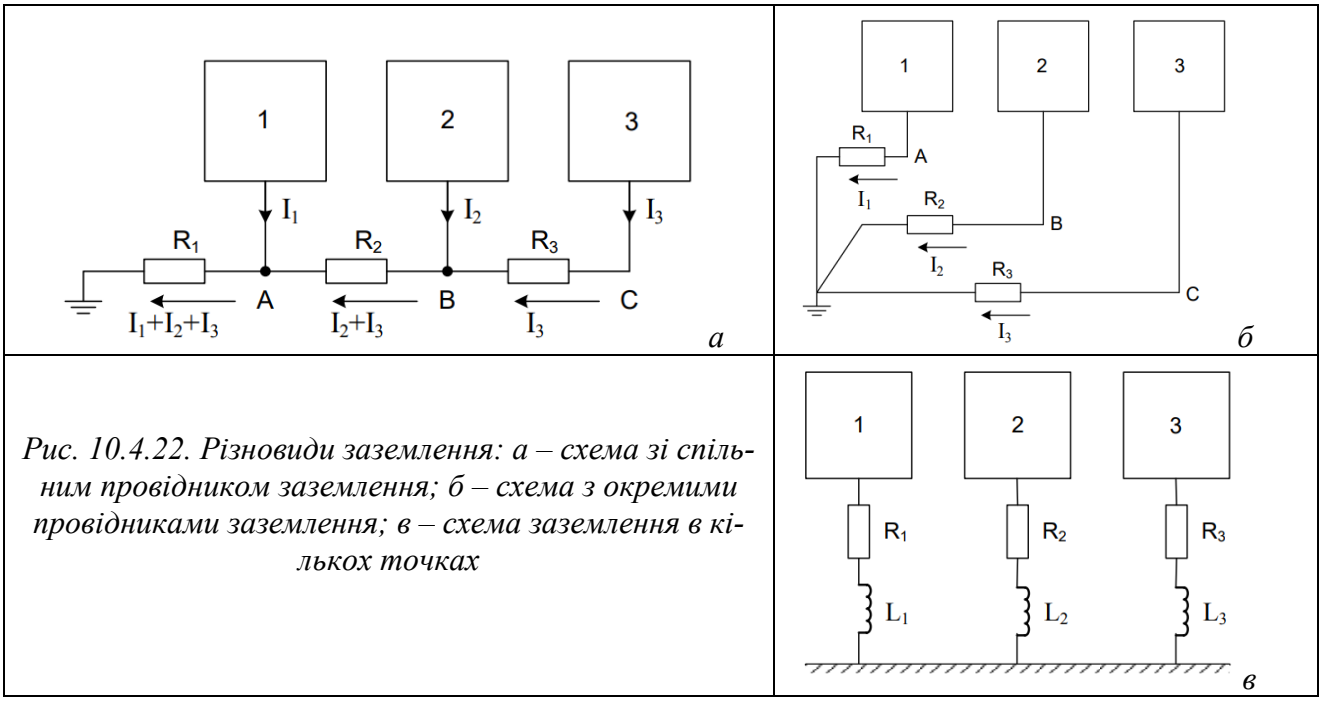


Рис. 10.4.22. Різновиди заземлення: а – схема зі спільним провідником заземлення; б – схема з окремими провідниками заземлення; в – схема заземлення в кількох точках

Ще одна проблема, пов’язана із заземленням – це так званий контур заземлення, який може відігравати роль рамкової антени, що приймає завади. Такий контур виникає, якщо заземлення виконується на кінцях лінії, до якої в кількох точках приєднані зв’язані між собою пристрої (рис. 10.4.23 а). Якщо необхідний додатковий захист від завад, треба розірвати контур заземлення, щоб зменшити площу контуру, в якому є силові лінії поля.

Контур заземлення можна розірвати за допомогою розділового трансформатора (рис. 10.4.23 б). Однак, у деяких ситуаціях необхідно, щоб між двома колами існував зв'язок за постійним струмом або на дуже низькій частоті, і застосовувати розділовий трансформатор не можна. У таких випадках трансформатор застосовують як дросель (рис. 10.4.23 в).

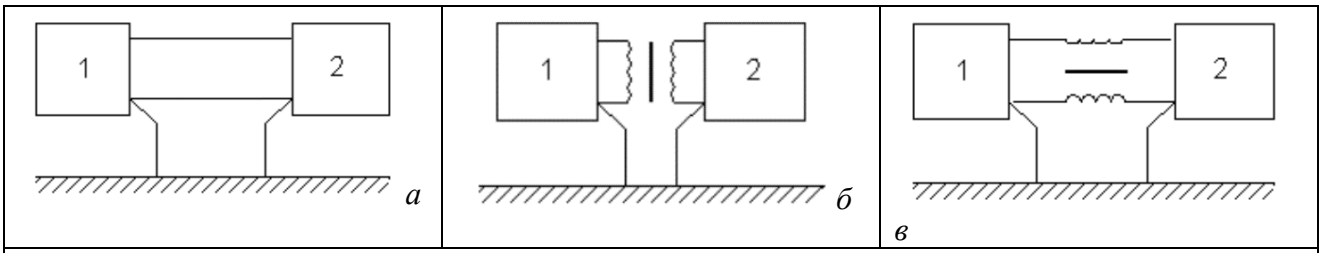


Рис. 10.4.23. Контур заземлення: схема утворення (а) та схеми усунення за допомогою трансформатора (б) та із застосуванням трансформатора як дроселя (в)

10.4.11. Протизавадні фільтри

Протизавадні фільтри застосовують з метою забезпечення електромагнітної сумісності в сигнальних (інформаційних) та силових колах (колах живлення).

Функція фільтрів у сигнальних колах – це обмеження робочої смуги частот тракту, аби туди не потрапляли завади, спектри яких лежать в інших діапазонах (див., наприклад, рис. 10.4.7).

Фільтри в силових колах виконують три основні функції:
– ослаблення провідних каналів завад у мережі електроживлення;
– запобігання несанкціонованого доступу до конфіденційної інформації;
– запобігання проходженню електромагнітних завад від електромережі до кола навантаження рис. 10.4.24).

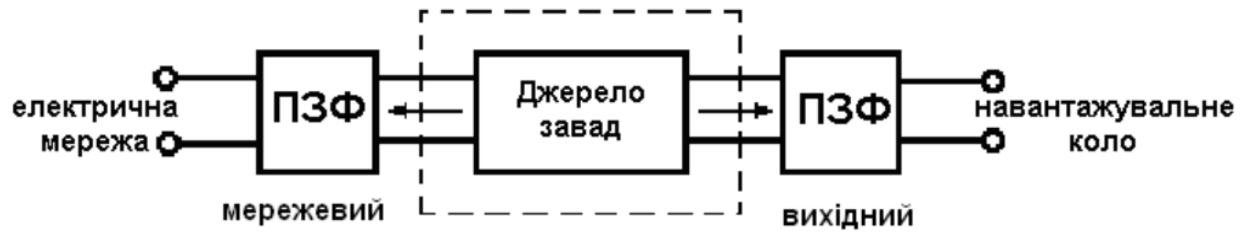


Рис. 10.4.24. Одна з можливих схем увімкнення протизавадних фільтрів

В основі роботи протизавадних фільтрів лежить два фізичні явища: відбиття електромагнітного сигналу та його поглинання.

Відбиття використовується у фільтрах із смугою загасання до 300 МГц - 1 ГГц, поглинання – у фільтрах, що забезпечують загасання в смузі частот до десятків гігагерц.

Контрольні питання до підрозділу 10.4

1. Що розуміють під електромагнітною сумісністю радіоелектронної апаратури? Які шляхи її досягнення?
2. Назвіть відомі вам типи завад природного походження.
3. Назвіть відомі вам типи завад техногенного походження.
4. Назвіть відомі вам канали впливу зовнішніх завад на радіоелектронні пристрої та методи боротьби з ними.

Розділ 11. Відокремлення корисних сигналів від завад і шумів

В реальному житті на вході приймача поруч із корисним сигналом майже завжди присутні шуми (внутрішні або зовнішні). Якщо рівень шумів сумірний із рівнем сигналу, задача отримання корисної інформації набуває статистичного характеру. В цьому розділі ми розглянемо різні методи відокремлення корисного сигналу від завад.

11.1. Найпростіші моделі фільтрації сигналу за наявності шумів та завад

В багатьох випадках відділити корисний сигнал від завад можна за допомогою лінійних фільтрів. Зокрема, це вдається зробити тоді, коли спектри завади й сигналу перекриваються слабко або тоді, коли спектр корисного сигналу значно вужчий за спектр завад чи шумів.

11.1.1. Лінійні фільтри

Будемо вважати приймач деяким лінійним фільтром. Можна відразу сказати, що такий фільтр дозволяє повністю відділити сигнал від шуму, якщо їхні спектри не перекриваються. В цьому випадку передавальна функція фільтру підбирається такою, щоб вона пропускала сигнал і придушувала шум (рис. 11.1.1).

Така ситуація можлива, коли в ролі шумів виступає зовнішня завада. Але для внутрішніх шумів, які мають широкий спектр, так зробити не вдається.

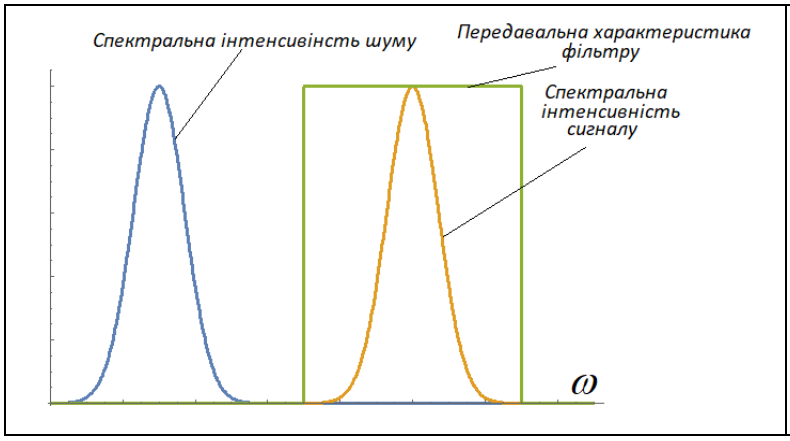


Рис. 11.1.1. Повне придушення шумів у випадку, коли спектри сигналу і шуму не перекриваються

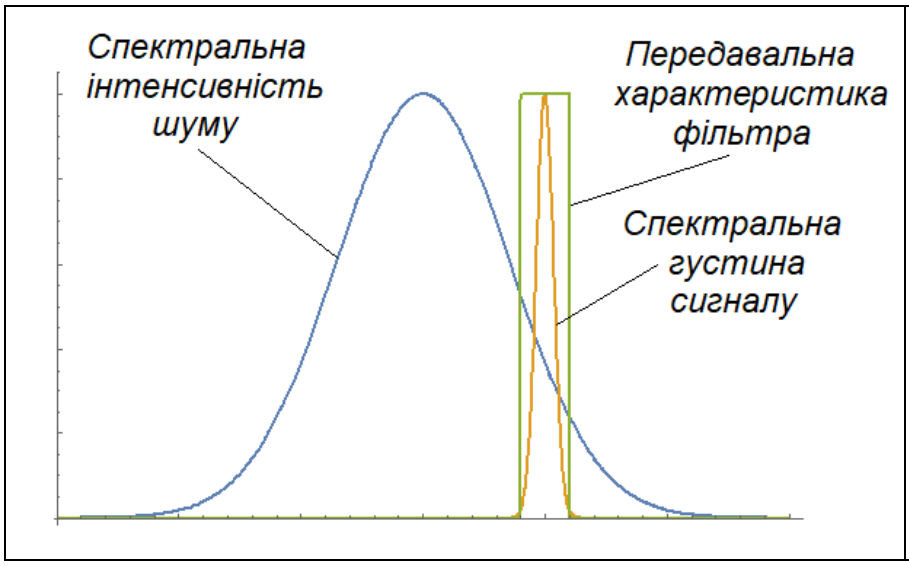


Рис. 11.1.2. Виділення вузькосмугового сигналу на фоні широкосмугового шуму

11.1.2. Фільтрація монохроматичних сигналів

Обговоримо тепер інший випадок, коли спектри шуму й сигналу перекриваються, але

спектр сигналу значно вужчий за спектр шуму (рис. 1.11.2). В цьому випадку за допомогою фільтру, смуга пропускання якого відповідає спектру сигналу, можна добитися того, що на вихід сигнал пройде повністю, а більша частина шумів буде відфільтрована. Якщо сигнал монохроматичний, то смугу пропускання фільтру можна зробити як завгодно вузькою, і на його вихід шум взагалі не потрапить.

Підтвердимо ці якісні міркування розрахунком.

Нехай на вхід приймача з передавальною функцією $K(\omega)$ надходить сума регулярного сигналу $S_{in}(t)$ і шуму $\xi(t)$:

$$x(t) = S_{in}(t) + \xi(t). \tag{11.1.1}$$

Нехай регулярний сигнал на вході є гармонічним:

$$S_i(t) = a\cos(\omega_1 t + \varphi_1) = \frac{a}{2}\left\{\exp\left[i(\omega_1 t + \varphi_1)\right] + \exp\left[-i(\omega_1 t + \varphi_1)\right]\right\}. \tag{11.1.2}$$

Інтенсивність цього сигналу:

$$I_{in} = \overline{\left|S_{in}(t)\right|^2} = a^2\overline{\cos^2(\omega_1 t + \varphi_1)} = \frac{a^2}{2}. \tag{11.1.3}$$

На виході маємо:

$$S_{out}(t) = \frac{a}{2}K(\omega_1)\exp\left[i(\omega_1 t + \varphi_1)\right] + c.c., \qquad I_{out} = \overline{\left|S_{out}(t)\right|^2} = \frac{a^2}{2}\left|K(\omega_1)\right|^2. \tag{11.1.4}$$

Якщо $G(\omega)$ – спектральна інтенсивність шуму $\xi(t)$, то його повна інтенсивність на вході й на виході складає відповідно

$$\sigma_{in}^2 = \int_{-\infty}^{\infty} G(\omega)\,d\omega = 2\int_0^{\infty} G(\omega)\,d\omega, \qquad \sigma_{out}^2 = \int_{-\infty}^{\infty} G(\omega)\left|K(\omega)\right|^2 d\omega. \tag{11.1.5}$$

Відношення сигнал-шум s/n – це відношення відповідних інтенсивностей. Отже,

$$\left(\frac{s}{n}\right)_{in} = \frac{I_{in}}{\sigma_{in}^2} = \frac{a^2}{4\int_0^{\infty} G(\omega)\,d\omega}, \qquad \left(\frac{s}{n}\right)_{out} = \frac{I_{out}}{\sigma_{out}^2} = \frac{a^2\left|K(\omega_1)\right|^2}{4\int_0^{\infty} G(\omega)\left|K(\omega)\right|^2 d\omega}. \tag{11.1.6}$$

Тоді

$$\frac{(s/n)_{out}}{(s/n)_{in}} = \frac{\left|K(\omega_1)\right|^2\int_0^{\infty} G(\omega)\,d\omega}{\int_0^{\infty} G(\omega)\left|K(\omega)\right|^2 d\omega}. \tag{11.1.7}$$

Для того, щоб максимально збільшити відношення сигнал-шум на виході, потрібно:

- покласти $\omega_0 = \omega_1$, де ω_0 – частота, на якій досягається максимум $\left|K(\omega)\right|$;
- обрати смугу пропускання $\Delta\omega_0$ фільтра $\left|K(\omega)\right|$ значно меншою за ширину спектру шуму на вході $\Delta\omega_n$, $\Delta\omega_0 \ll \Delta\omega_n$.

Тоді в смузі $\Delta\omega_0$ функцію $G(\omega)$ можна вважати сталою і винести її з-під знаку інтеграла. Маємо:

$$\frac{(s/n)_{out}}{(s/n)_{in}} = \frac{|K(\omega_0)|^2 \int_0^\infty G(\omega)\,d\omega}{G(\omega_0)\int_0^\infty |K(\omega)|^2\,d\omega}. \qquad (11.1.8)$$

Визначимо шумову смугу пропускання фільтра як

$$\Delta\omega_0 = \frac{\int_0^\infty |K(\omega)|^2\,d\omega}{|K(\omega_0)|^2} \qquad (11.1.9)$$

(рис. 11.1.3).

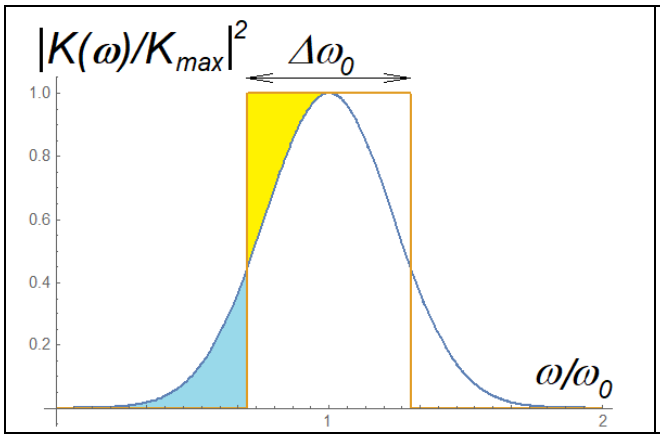


Рис. 11.1.3. До визначення шумової смуги пропускання фільтру (площі блакитної та жовтої областей однакові

Аналогічно визначимо ефективну ширину спектра шуму:

$$\Delta\omega_n = \frac{\int_0^\infty G(\omega)\,d\omega}{G(\omega_0)}. \qquad (11.1.10)$$

Тоді можна записати, що поліпшення співвідношення сигнал-шум у результаті фільтрації буде

$$\frac{(s/n)_{out}}{(s/n)_{in}} = \frac{\Delta\omega_n}{\Delta\omega_0}. \qquad (11.1.11)$$

Оскільки в даній моделі регулярний сигнал є монохроматичним, то шумову смугу пропускання фільтру $\Delta\omega_0$ можна зробити як завгодно вузькою і тим самим досягти як завгодно великого поліпшення співвідношення сигнал-шум у результаті фільтрації.

В реальних ситуаціях подібна фільтрація, як уже вказувалося, буде ефективною, коли ширина спектру сигналу значно менша від ширини спектру шуму (рис. 11.1.2).

Насправді реальні сигнали завжди мають суттєво ненульову ширину спектру, і для них задача про фільтрацію виглядає складнішою.

Контрольні питання

1. Що таке лінійні фільтри?
2. У яких випадках лінійні фільтри дозволяють повністю позбутися шумів?
3. Чому гармонічний сигнал завжди можна за допомогою лінійного фільтра виділити на фоні шумів?

Задачі до підрозділу 11.1

11.1.1. На вхід лінійного фільтра надходить амплітудно-модульований сигнал (частота-носій – ω_0, частота модуляції – Ω $(\Omega \ll \omega_0)$, глибина модуляції – m) та білий шум зі спектральною інтенсивністю G_0. Знайти співвідношення сигнал-шум на його виході. Вважати, що фільтр має прямокутну АЧХ, повністю пропускає сигнал і максимально придушує шум.

11.1.2. Розв'язати задачу 11.1.1, вважаючи, що вхідний шум має лоренцівський спектр,

$$G(\omega) = \frac{G_0}{(\Delta\omega)^2 + (\omega - \omega_1)^2}, \quad \Delta\omega \gg \Omega, \ \omega_0 \ll \omega_1 .$$

11.2. Фільтрація сигналів за наявності шуму: задача виявлення

Можна сформулювати дві можливі постановки задачі про фільтрацію сигналу за наявності шумів.

Перша – це задача виявлення сигналу, коли ми повинні дати відповідь на питання, чи надходив сигнал (заздалегідь відомої форми) на вхід приймача протягом деякого проміжку часу. Відтворення форми корисного сигналу при цьому не вимагається.

Друга – це задача виділення сигналу, коли на виході приймача потрібно отримати сигнал, форма якого (за деяким критерієм) максимально близька до форми корисного сигналу на вході.

Обговоримо спершу задачу виявлення.

11.2.1. Постановка задачі про оптимальний фільтр

Вважатимемо, що регулярний сигнал $S_{in}(t)$ на вході

$$S_{in}(t)=\int_{-\infty}^{\infty} S_{in}(\omega)\exp(i\omega t)\,d\omega\,. \tag{11.2.1}$$

має обмежену тривалість:

$$\lim_{t\to\pm\infty} S_{in}(t)=0\,. \tag{11.2.2}$$

Його миттєва інтенсивність – $I_{in}(t)=S_{in}^2(t)$, повна енергія –

$$E_{in}=\int_{-\infty}^{\infty} S_{in}^2(t)\,dt=2\pi\int_{-\infty}^{\infty}\left|S_{in}^2(\omega)\right|d\omega \tag{11.2.3}$$

(див. вище п. 1.3.4). Нехай на вхід фільтра разом із сигналом надходить стаціонарний шум $\xi(t)$ ($\langle\xi(t)\rangle=0$) зі спектральною інтенсивністю $G(\omega)$. Знайдемо передавальну функцію $K_{opt}(\omega)$ так званого оптимального фільтра, для якого відношення сигнал-шум на виході буде максимальним.

На виході фільтра для сигналу маємо:

$$S_{out}(t)=\int_{-\infty}^{\infty} S_{in}(\omega)K_{opt}(\omega)\exp(i\omega t)\,d\omega, \tag{11.2.4}$$

$$I_{out}(t)=\left|S_{out}^2(t)\right|=\left|\int_{-\infty}^{\infty} S_{in}(\omega)K_{opt}(\omega)\exp(i\omega t)\,d\omega\right|^2, \tag{11.2.5}$$

$$E_{out}=2\pi\int_{-\infty}^{\infty}\left|S_{in}^2(\omega)\right|\left|K_{opt}^2(\omega)\right|d\omega\,. \tag{11.2.6}$$

Тоді на виході

$$\left(\frac{s}{n}\right)_{out}=\frac{I_{out}(t)}{\int_{-\infty}^{\infty} G(\omega)\left|K_{opt}(\omega)\right|^2 d\omega}\,. \tag{11.2.7}$$

11.2.2. Виявлення сигналу на фоні білого шуму

Розглянемо спершу простіший випадок, коли $G(\omega)=const=G_0$, тобто вхідний шум – білий. Відповідну передавальну функцію позначимо через $K_1(\omega)$. Тоді

$$\int_{-\infty}^{\infty} G(\omega)\left|K_1(\omega)\right|^2 d\omega=G_0\int_{-\infty}^{\infty}\left|K_1(\omega)\right|^2 d\omega\,. \tag{11.2.8}$$

У математиці доводиться співвідношення

$$\left|\int_{-\infty}^{\infty} f(x)g^*(x)dx\right|^2 \le \int_{-\infty}^{\infty}|f(x)|^2 dx \times \int_{-\infty}^{\infty}|g(x)|^2 dx, \qquad (11.2.9)$$

яке називають нерівністю Коші – Буняковського. Тут $f(x)$ і $g(x)$ – довільні квадратично інтегровні функції. Нерівність переходить у рівність, якщо $g(x) = C_0 f(x)$, C_0 – довільна константа.

За допомогою цієї нерівності оцінимо інтенсивність сигналу на виході, прийнявши, що $f(x) \to S_{in}(\omega)$, $g(x) \to K_1(\omega)\exp(-i\omega t_0)$:

$$I_{out}(t) = \left|\int_{-\infty}^{\infty} S_{in}(\omega)K_1(\omega)\exp(i\omega t)d\omega\right|^2 \le \int_{-\infty}^{\infty}|S_{in}(\omega)|^2 d\omega \times \int_{-\infty}^{\infty}|K_1(\omega)|^2 d\omega. \qquad (11.2.10)$$

Тоді

$$\left(\frac{s}{n}\right)_{out} \le \frac{\int_{-\infty}^{\infty}|S_{in}(\omega)|^2 d\omega \times \int_{-\infty}^{\infty}|K_1(\omega)|^2 d\omega}{G_0\int_{-\infty}^{\infty}|K_1(\omega)|^2 d\omega} = \frac{\int_{-\infty}^{\infty}|S_{in}(\omega)|^2 d\omega}{G_0}. \qquad (11.2.11)$$

Нерівність переходить у рівність для моменту часу $t = t_0$, коли $K_1^*(\omega)\exp(-i\omega t_0) = C_0 S_{in}(\omega)$, або

$$K_1(\omega) = C_0 S_i^*(\omega)\exp(-i\omega t_0). \qquad (11.2.12)$$

Останнє співвідношення й визначає передавальну функцію оптимального фільтра для білого шуму. Такий фільтр інколи називають узгодженим.

11.2.3. Зміна форми сигналу на виході оптимального фільтра

Сигнал на виході оптимального фільтра для білого шуму:

$$I_{out}(t) = \left|\int_{-\infty}^{\infty} S_{in}(\omega)K_1(\omega)\exp(i\omega t)d\omega\right|^2 = C_0^2\left|\int_{-\infty}^{\infty}|S_{in}(\omega)|^2\exp\left[i\omega(t-t_0)\right]d\omega\right|^2 =$$
$$= C_0^2\left|\int_{-\infty}^{\infty}|S_{in}(\omega)|^2\cos\omega(t-t_0)d\omega\right| \qquad (11.2.13)$$

(враховано, що $S_{in}(-\omega) = S_{in}^*(\omega)$, тобто $|S_{in}(\omega)|^2$ – парна функція частоти. Таким чином, функція $I_{out}(t)$ є симетричною щодо моменту часу $t = t_0$, коли вона досягає максимального значення:

$$I_{out}(t_0) = C_0^2\left|\int_{-\infty}^{\infty}|S_{in}(\omega)|^2 d\omega\right|^2 = \frac{C_0^2 E_{in}}{2\pi}. \qquad (11.2.14)$$

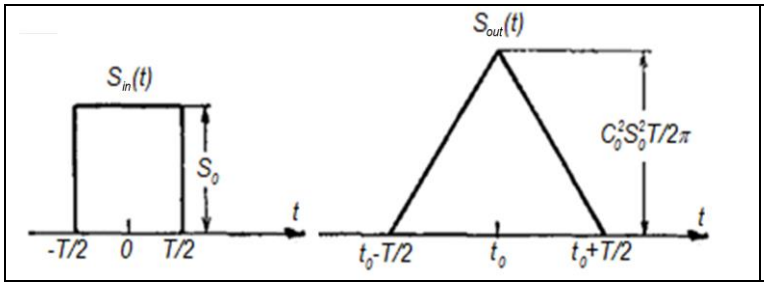


Рис. 11.2.1. Перетворення прямокутного вхідного сигналу в узгодженому лінійному фільтрі (шум вважається білим)

Можна показати, що для регулярного сигналу прямокутної форми залежність $I_{out}(t)$ матиме форму трикутника (рис. 11.2.1).

Фазово-частотна характеристика оптимального фільтру підібрана так, що в момент часу $t = t_0$ всі гармоніки вхідного сигналу додаються в однаковій фазі, чим і забезпечується максимальне значення вихідного сигналу. Саме в цей момент він стає помітним на фоні шумів.

11.2.4. Оптимальний фільтр для шумів з довільним спектром

Тепер перейдемо до випадку, коли $G(\omega)$ – довільна функція частоти. Зведемо її до попереднього випадку. Для цього потрібно перетворити шум із довільним розподілом спектральної інтенсивності на білий. Очевидно, для цього його треба пропустити через фільтр з передавальною функцією

$$K_2(\omega) = C_1 exp\left[i\varphi(\omega)\right] \Big/ \sqrt{G(\omega)}\,, \tag{11.2.15}$$

де C_1 – довільна константа, $\varphi(\omega)$ – довільна дійсна функція. Справді, в цьому випадку

$$G_1(\omega) = G(\omega)\left|K_2(\omega)\right|^2 = C_1^2\,, \tag{11.2.16}$$

і задача зводиться до попередньої. Отже,

$$K_{opt}(\omega) = K_1(\omega)K_2(\omega) = \frac{CS_i^*(\omega)exp\left[i\varphi(\omega) - i\omega t_0\right]}{\sqrt{G(\omega)}}\,, \tag{11.2.17}$$

де $C = C_1 C_2$. Сигнал на його виході:

$$\begin{aligned} S_{out}(t) &= \int_{-\infty}^{\infty} S_{in}(\omega)K_{opt}(\omega)\exp(i\omega t)\,d\omega = \\ &= \int_{-\infty}^{\infty} S_{in}(\omega)\frac{CS_{in}^*(\omega)\exp\left[i\varphi(\omega) - i\omega t_0\right]}{\sqrt{G(\omega)}}\exp(i\omega t)\,d\omega = \\ &= C\int_{-\infty}^{\infty}\frac{\left|S_{in}(\omega)\right|^2}{\sqrt{G(\omega)}}\exp\left[i\varphi(\omega) + i\omega(t - t_0)\right]d\omega \end{aligned} \tag{11.2.18}$$

Цей сигнал досягає максимуму при $t = t_0$ за умови $\varphi(\omega) = const$. Новим (у порівнянні з випадком білого шуму є те, що передавальна функція зменшується на тих частотах, де $G(\omega)$ максимальне (рис. 11.2.2).

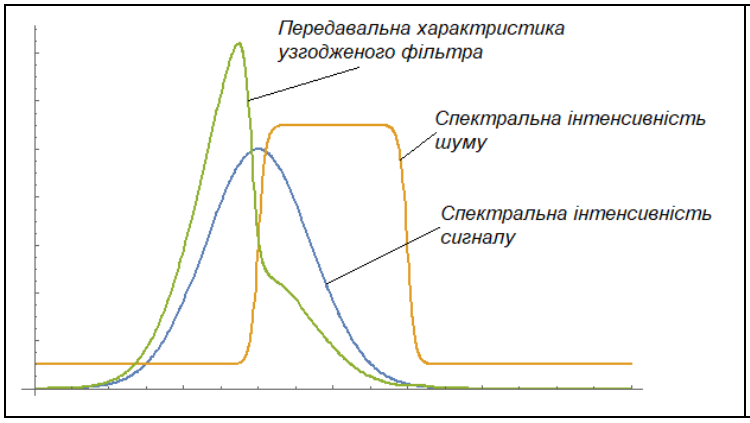


Рис. 11.2.2. Вплив спектральної інтенсивності шуму на передавальну характеристику узгодженого фільтра

Контрольні питання до підрозділу 11.2

1. Що таке задача виявлення і задача виділення сигналу на фоні шумів?
2. Опишіть принцип роботи узгодженого фільтра.
3. Чи зберігається форма сигналу при його проходженні через узгоджений фільтр? Відповідь обґрунтувати.
4. Поясніть вигляд фазово-частотної та амплітудно-частотної характеристик узгодженого фільтра.
5. Як спектр шуму впливає на амплітудно-частотну характеристику узгодженого фільтра?

Задачі до підрозділу 11.2

11.2.1. Корисний сигнал має форму прямокутного імпульсу тривалістю τ . Побудувати амплітудно-частотну та фазово-частотну характеристику узгодженого фільтра для його виявлення на фоні білого шуму.

11.2.2. Корисний сигнал на вході узгодженого фільтра має форму гауссівського імпульсу тривалістю τ . Як зміниться його форма на виході цього фільтра? Шум вважати білим.

11.3. Фільтрація сигналів за наявності шуму: задача виділення

Як уже вказувалося, задача виділення сигналу полягає в тому, що на виході приймача, на вхід якого надходить сума корисного сигналу й шуму, потрібно отримати сигнал, форма якого (за деяким критерієм) максимально близька до форми корисного сигналу на вході.

В цьому підрозділі з'ясовується характеристика фільтра, який дозволяє розв'язати задачу виділення, та розглядаються характерні режими фільтрації.

11.3.1. Постановка задачі

В багатьох випадках (мова людини, музика, тощо) сигнал, що несе корисну інформацію, може розглядатися як стаціонарний випадковий процес $\xi_1(t)$. Відповідно шум на вході приймача подамо як стаціонарний випадковий процес $\xi_2(t)$. Вважатимемо, що

$$\xi_{1,2}(t)=\int_{-\infty}^{\infty}\xi_{1,2}(\omega)\exp(i\omega t)\,d\omega,\qquad \langle\xi_{1,2}(t)\rangle=0,\qquad \langle\xi_{1,2}(\omega)\rangle=0, \tag{11.3.1}$$

$$\langle\xi_{1,2}(\omega)\xi_{1,2}(\omega')\rangle=G_{1,2}(\omega)\delta(\omega+\omega').$$

За спектральними інтенсивностями $G_{1,2}(\omega)$, користуючись теоремою Вінера – Хінчина, можна знайти відповідні автокореляційні функції $B_{1,2}(\tau)$.

Вважатимемо процеси $\xi_1(t)$ і $\xi_2(t)$ статистично незалежними:

$$\langle\xi_1(\omega)\xi_2(\omega')\rangle=0. \tag{11.3.2}$$

Отже, на вході фільтра

$$x_{in}(t)=\xi_1(t)+\xi_2(t). \tag{11.3.3}$$

Наше завдання – знайти передавальну функцію $K(\omega)$ такого фільтра, сигнал на виході якого

$$x_{out}(t)=\int_{-\infty}^{\infty}K(\omega)\left[\xi_1(\omega)+\xi_2(\omega)\right]\exp(i\omega t)\,d\omega \tag{11.3.4}$$

з деяким запізненням t_0 найкраще відтворює на виході корисний сигнал $\xi_1(t)$.

Для оптимізації потрібен кількісний критерій. Вважатимемо оптимальною передавальну функцію, що забезпечує мінімальну середньоквадратичну похибку $\langle\mu^2\rangle$, де

$$\mu(t)=x_{out}(t+t_0)-\xi_1(t). \tag{11.3.5}$$

11.3.2. Розрахунок передавальної функції

Очевидно, $\langle\mu\rangle=0$, тому середньоквадратична похибка відтворення на виході складає

$$\langle\mu^2\rangle=\left\langle\left[x_{out}(t+t_0)-\xi_1(t)\right]^2\right\rangle=\left\langle x_{out}^2(t+t_0)+\xi_1^2(t)-2x_{out}(t+t_0)\xi_1(t)\right\rangle=$$

$$=\left\langle\int_{-\infty}^{\infty}K(\omega)\left[\xi_1(\omega)+\xi_2(\omega)\right]\exp\left[i\omega(t+t_0)\right]d\omega\int_{-\infty}^{\infty}K(\omega')\left[\xi_1(\omega')+\xi_2(\omega')\right]\exp\left[i\omega'(t+t_0)\right]d\omega'\right\rangle+$$

$$+\left\langle\int_{-\infty}^{\infty}\xi_1(\omega)\exp(i\omega t)\,d\omega\int_{-\infty}^{\infty}\xi_1(\omega')\exp(i\omega' t)\,d\omega'\right\rangle-$$

$$-2\left\langle\int_{-\infty}^{\infty}K(\omega)\left[\xi_1(\omega)+\xi_2(\omega)\right]\exp\left[i\omega(t+t_0)\right]d\omega\int_{-\infty}^{\infty}\xi_1(\omega')\exp(i\omega' t)\,d\omega'\right\rangle=$$

$$= \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' K(\omega) K(\omega') \exp\left[i(\omega+\omega')(t+t_0) \right] \times$$
$$\times \left\langle \xi_1(\omega)\xi_1(\omega') + \xi_2(\omega)\xi_2(\omega') + \xi_1(\omega)\xi_2(\omega') + \xi_2(\omega)\xi_1(\omega') \right\rangle +$$
$$+ \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' \exp\left[i(\omega+\omega')t \right] \left\langle \xi_1(\omega)\xi_1(\omega') \right\rangle - \tag{11.3.6}$$
$$-2\int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' K(\omega) \exp\left[i(\omega+\omega')t + i\omega t_0 \right] \left\langle \xi_1(\omega)\xi_1(\omega') \right\rangle.$$

Врахувавши записані раніше співвідношення (11.3.1), отримаємо:

$$\left\langle \mu^2 \right\rangle = \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' K(\omega) K(\omega') \exp\left[i(\omega+\omega')(t+t_0) \right] \left[G_1(\omega) + G_2(\omega) \right] \delta(\omega+\omega') +$$
$$+ \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' \exp\left[i(\omega+\omega')t \right] G_1(\omega) \delta(\omega+\omega') -$$
$$-2\int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' K(\omega) \exp\left[i(\omega+\omega')t + i\omega t_0 \right] G_1(\omega) \delta(\omega+\omega') =$$
$$= \int_{-\infty}^{\infty} d\omega K(\omega) K(-\omega) \left[G_1(\omega) + G_2(\omega) \right] + \int_{-\infty}^{\infty} d\omega G_1(\omega) - 2\int_{-\infty}^{\infty} d\omega K(\omega) \exp(i\omega t_0) G_1(\omega) =$$
$$\int_{-\infty}^{\infty} d\omega \left\{ G_1(\omega) \left[K(\omega) K(-\omega) + 1 - 2K(\omega) \exp(i\omega t_0) \right] + G_2(\omega) K(\omega) K(-\omega) \right\} =$$
$$= \left\| \begin{matrix} K(-\omega) = K^*(\omega); \; K(\omega)\exp(i\omega t_0) \equiv K'(\omega) \\ K(\omega)K(-\omega) = K'(\omega)K'(-\omega) = \left[K'(\omega) \right]^2 \end{matrix} \right\| =$$
$$= \int_{-\infty}^{\infty} d\omega \left\{ G_1(\omega) \left[K'(\omega) - 1 \right]^2 + G_2(\omega) \left[K'(\omega) \right]^2 \right\}. \tag{11.3.7}$$

Тут t_0 – введений раніше час затримки сигналу на виході фільтру. Враховано, що за змістом задачі $K'(\omega)$ – дійсна функція.

Варіація $\left\langle \mu^2 \right\rangle$ по $K'(\omega)$:

$$\delta \left\langle \mu^2 \right\rangle = 2\int_{-\infty}^{\infty} d\omega \left\{ G_1(\omega) \left[K'(\omega) - 1 \right] + G_2(\omega) K'(\omega) \right\} \delta K'(\omega). \tag{11.3.8}$$

В мінімумі $\left\langle \mu^2 \right\rangle$ умова $\delta \left\langle \mu^2 \right\rangle = 0$ повинна виконуватися для довільного $\delta K'(\omega)$, тому

$$G_1(\omega) \left[K'(\omega) - 1 \right] + G_2(\omega) K'(\omega) = 0. \tag{11.3.9}$$

Звідси

$$K'(\omega) = \frac{G_1(\omega)}{G_1(\omega) + G_2(\omega)}, \tag{11.3.10}$$

або

$$K(\omega) = \frac{G_1(\omega) \exp(-i\omega t_0)}{G_1(\omega) + G_2(\omega)}. \tag{11.3.11}$$

Видно, що $\left| K(\omega) \right| \le 1$.

Середньоквадратична похибка відтворення сигналу на виході такого фільтра:

$$\left\langle \mu^2 \right\rangle = \int_{-\infty}^{\infty} d\omega \left\{ G_1(\omega)\left[K'(\omega)-1\right]^2 + G_2(\omega)\left[K'(\omega)\right]^2 \right\} =$$

$$= \int_{-\infty}^{\infty} d\omega \left\{ G_1(\omega)\left[\frac{G_1(\omega)}{G_1(\omega)+G_2(\omega)}-1\right]^2 + G_2(\omega)\left[\frac{G_1(\omega)}{G_1(\omega)+G_2(\omega)}\right]^2 \right\} =$$

$$= \int_{-\infty}^{\infty} d\omega \left\{ G_1(\omega)\frac{G_2^2(\omega)}{\left[G_1(\omega)+G_2(\omega)\right]^2} + G_2(\omega)\frac{G_1^2(\omega)}{\left[G_1(\omega)+G_2(\omega)\right]^2} \right\} = \tag{11.3.12}$$

$$= \int_{-\infty}^{\infty} d\omega \frac{G_1(\omega)G_2(\omega)}{G_1(\omega)+G_2(\omega)} \le \int_{-\infty}^{\infty} d\omega G_{1,2}(\omega) = \sigma_{1,2}^2$$

Видно, що середньоквадратична похибка $\left\langle \mu^2 \right\rangle$ не залежить від часу затримки t_0.

11.3.3. Характерні режими фільтрації

1. Сигнал значно більший за шум, їхні спектри перекриваються. Тоді на частотах, де $G_1(\omega) \ne 0$, $K(\omega) \approx 1$. Тоді $(s/n)_i \approx (s/n)_o >> 1$, тобто фільтрація не потрібна (рис. 11.3.1 а).

2. Спектри сигналу й шуму не перекриваються. Як і в попередньому випадку, на частотах, де $G_1(\omega) \ne 0$, $K(\omega) \approx 1$. При цьому весь шум відфільтровується, тому $(s/n)_o >> 1$ (рис. 11.3.1 б). В цьому випадку фільтрація найбільш ефективна.

3. Сигнал значно менший за шум, їхні спектри перекриваються. Тоді на частотах, де $G_1(\omega) \ne 0$, $K(\omega) \ll 1$. Тоді $(s/n)_i \approx (s/n)_o \ll 1$, і фільтрація не дає виграшу (рис. 11.3.1 в).

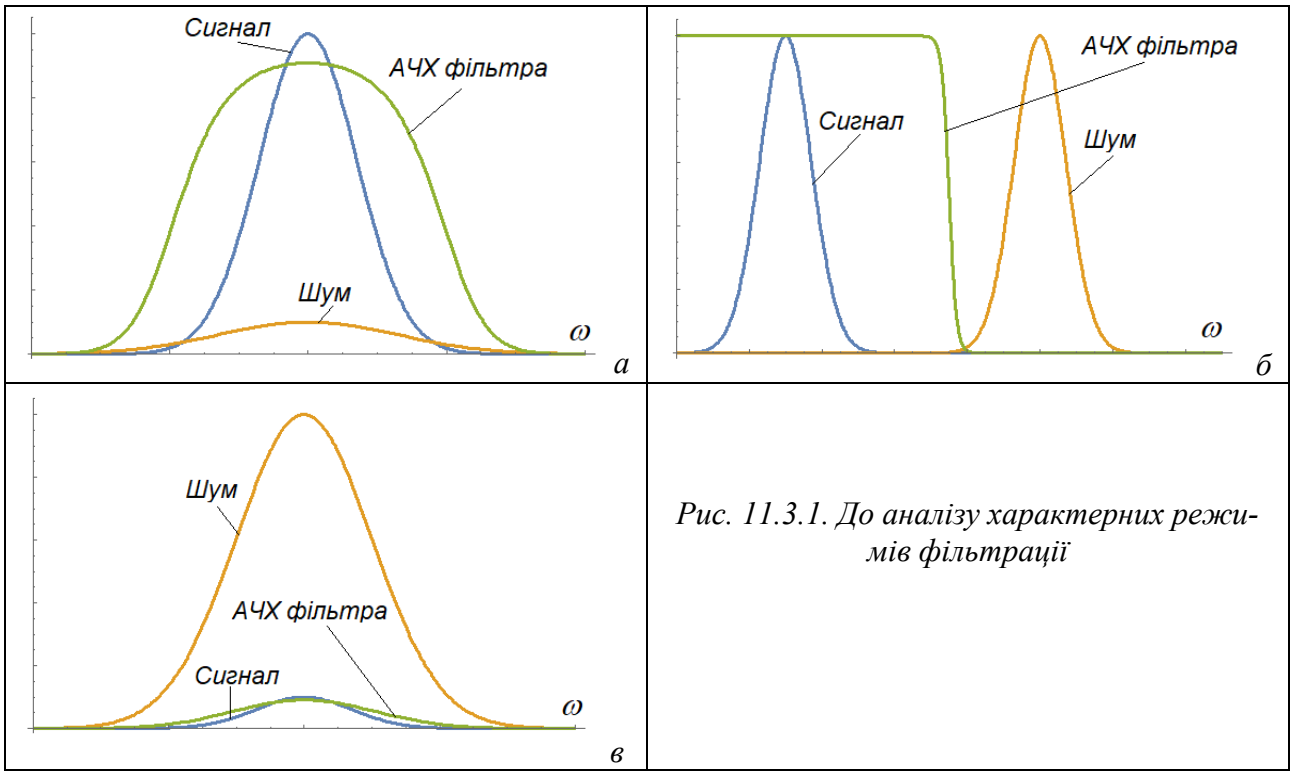


Рис. 11.3.1. До аналізу характерних режимів фільтрації

11.3.4. Рівняння Вінера - Хопфа

Ми розв'язували задачу виділення сигналу в спектральному поданні й знаходили передавальну функцію оптимального фільтра. Але цю задачу можна розв'язувати і в часовому поданні, шукаючи оптимальну функцію Гріна (імпульсну функцію). Для цього запишемо сигнал

на виході у формі

$$x_{out}(t)=\int_{-\infty}^{\infty}H(\theta)\left[\xi_1(t-\theta)+\xi_2(t-\theta)\right]d\theta. \tag{11.3.13}$$

Середньоквадратична похибка набуває вигляду

$$\begin{aligned}\left\langle\mu^2\right\rangle&=\left\langle\int_{-\infty}^{\infty}H(\theta)\left[\xi_1(t-\theta)+\xi_2(t-\theta)\right]d\theta\int_{-\infty}^{\infty}H(\theta')\left[\xi_1(t-\theta')+\xi_2(t-\theta')\right]d\theta'\right\rangle+\\&+\left\langle\xi_1^2(t+t_0)\right\rangle-2\left\langle\xi_1(t+t_0)\int_{-\infty}^{\infty}H(\theta)\left[\xi_1(t-\theta)+\xi_2(t-\theta)\right]d\theta\right\rangle=\\&=\int_{-\infty}^{\infty}d\theta\int_{-\infty}^{\infty}d\theta'H(\theta)H(\theta')\times\\&\times\left\langle\xi_1(t-\theta)\xi_1(t-\theta')+\xi_1(t-\theta)\xi_2(t-\theta')+\xi_2(t-\theta)\xi_1(t-\theta')+\xi_2(t-\theta)\xi_2(t-\theta')\right\rangle+\\&+\left\langle\xi_1^2(t+t_0)\right\rangle-2\int_{-\infty}^{\infty}H(\theta)\left[\left\langle\xi_1(t-\theta)\xi_1(t+t_0)\right\rangle+\left\langle\xi_2(t-\theta)\xi_1(t+t_0)\right\rangle\right]d\theta=\\&=\int_{-\infty}^{\infty}d\theta\int_{-\infty}^{\infty}d\theta'H(\theta)H(\theta')\left[B_1(\theta-\theta')+B_2(\theta-\theta')\right]+B_1(0)-2\int_{-\infty}^{\infty}H(\theta)B_1(\theta-t_0)d\theta.\end{aligned} \tag{11.3.14}$$

Варіація по $H(\theta)$:

$$\delta\left\langle\mu^2\right\rangle=2\int_{-\infty}^{\infty}d\theta\delta H(\theta)\left\{\int_{-\infty}^{\infty}d\theta'H(\theta')\left[B_1(\theta-\theta')+B_2(\theta-\theta')\right]-B_1(\theta-t_0)\right\}. \tag{11.3.15}$$

Мінімум варіації забезпечується за умови

$$\int_{-\infty}^{\infty}d\theta'H(\theta')\left[B_1(\theta-\theta')+B_2(\theta-\theta')\right]=B_1(\theta-t_0), \tag{11.3.17}$$

відомої як рівняння Вінера – Хопфа. Функція Гріна і передавальна функція пов'язані між собою перетворенням Фур'є. Тому зі знайденої раніше передавальної функції оптимального фільтра можна отримати функцію Гріна, яка задовольняє рівняння Вінера – Хопфа:

$$\begin{aligned}H(\theta)&=\frac{1}{2\pi}\int_{-\infty}^{\infty}K(\omega)\exp(i\omega\theta)d\omega=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{G_1(\omega)\exp(-i\omega t_0)}{G_1(\omega)+G_2(\omega)}\exp(i\omega\theta)d\omega=\\&=\|\theta-t_0\equiv\tau\|=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{G_1(\omega)\exp(i\omega\tau)}{G_1(\omega)+G_2(\omega)}d\omega=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{G_1(\omega)\cos\omega\tau}{G_1(\omega)+G_2(\omega)}d\omega\end{aligned} \tag{11.3.18}$$

(враховано, що спектральна інтенсивність – парна функція частоти). З останньої формули випливає, що функція Гріна, яка задовольняє рівняння Вінера – Хопфа, є парною функцією τ, тобто вона симетрична щодо моменту часу t_0. З іншого боку, виходячи з принципу причинності[1], для реальних систем має виконуватися умова $H(\theta<0)=0$.

Для розв'язання цієї суперечності слід узяти до уваги, що для передавальних функцій зі скінченною смугою пропускання $H(\theta)$ відмінна від нуля тільки для не дуже великих аргументів, $\theta<\theta_0$. Тому, взявши $t_0\geq\theta_0$, умову парності щодо точки t_0 і умову $H(\theta<0)=0$ вдається задовольнити одночасно (рис. 11.3.2).

[1] Йдеться про те, що в часі причина обов'язково має передувати наслідкам.

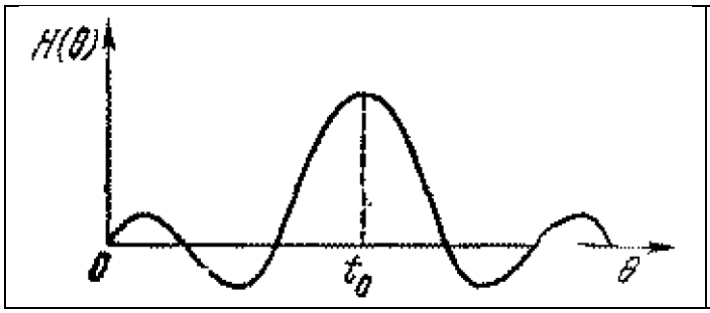


Рис. 11.3.2. Підбір часу затримки, що задовольняє одночасно умову парності щодо t_0 та принцип причинності

11.3.5. Оптимальні фільтри й корелятори

Лінійні фільтри є особливо ефективними, коли спектри сигналу й шуму не перекриваються. Але суттєві відмінності спектрів означають і відмінності кореляційних функцій. Тому для того, щоб відділити сигнал від шуму, можна використати автокореляційні та взаємно кореляційні функції.

Схема корелятора подана на рис. 11.3.3 а. Він дозволяє виміряти автокореляційну функцію сигналу (в припущення, що відповідний процес є ергодичним).

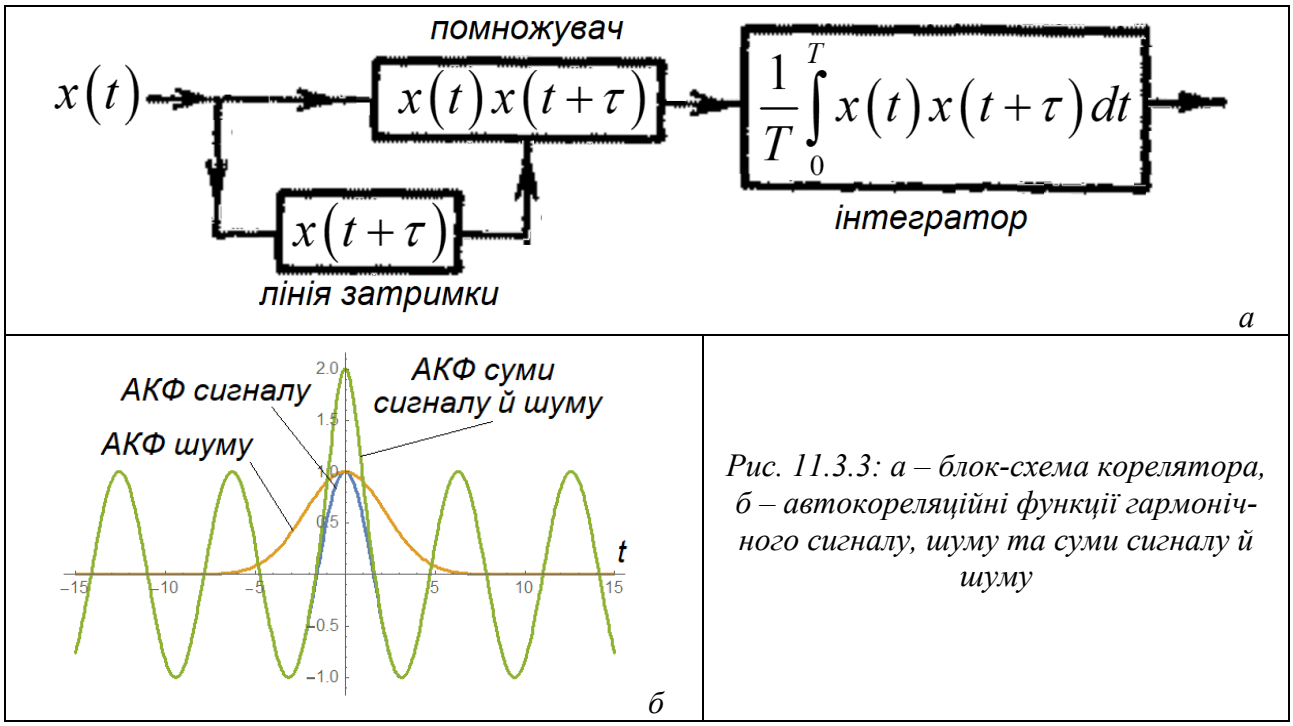


Рис. 11.3.3: а – блок-схема корелятора, б – автокореляційні функції гармонічного сигналу, шуму та суми сигналу й шуму

Нехай на вхід корелятора надходить сума гармонічного сигналу з випадковою початковою фазою й стаціонарного шуму:

$$x(t)=S(t)+\xi(t), \quad S(t)=a\cos(\omega t+\varphi), \qquad w(\varphi)=1/2\pi .$$

Вважаємо випадкову величину φ і шум ξ статистично незалежними. Тоді на виході корелятора маємо:

$$\langle xx_\tau\rangle=\langle[S(t)+\xi(t)][S(t+\tau)+\xi(t+\tau)]\rangle=$$

$$=\langle S(t)S(t+\tau)\rangle+\langle S(t)\rangle\langle\xi(t+\tau)\rangle+\langle S(t+\tau)\rangle\langle\xi(t)\rangle+\langle\xi(t)\xi(t+\tau)\rangle=$$

$$=\langle a^2\cos(\omega t+\varphi)\cos(\omega t+\omega\tau+\varphi)\rangle+B_\xi(\tau)=$$

$$=\left\langle\frac{1}{2}a^2\left[\cos(2\omega t+\omega\tau+2\varphi)+\cos\omega\tau\right]\right\rangle+B_\xi(\tau)=\frac{1}{2}a^2\cos\omega\tau+B_\xi(\tau).$$

Якщо τ перевищує час автокореляції шуму, то шуми перестануть давати внесок до

атокореляційної функції, і залишиться тільки корисний сигнал (рис. 11.3.3 б).

Контрольні питання до підрозділу 11.3

1. Дайте математичне формулювання задачі виділення сигналу на фоні шумів.
2. Дайте якісне пояснення вигляду амплітудно-частотної характеристики оптимального фільтра для задачі виділення. Коли такий фільтр буде ефективним?
3. Яку характеристику оптимального фільтра можна визначити з рівняння Вінера – Хопфа?
4. Опишіть принцип виділення сигналу на фону шуму за допомогою корелятора.

Задачі до підрозділу 11.3

11.3.1. Корисний сигнал має спектральну інтенсивність $G_1(\omega) = G_{10} \exp\left[-\frac{(\omega-\omega_{10})^2}{\Omega_{10}^2}\right]$, шум – спектральну інтенсивність $G_2(\omega) = G_{20} \exp\left[-\frac{(\omega-\omega_{20})^2}{\Omega_{20}^2}\right]$. Запишіть передавальну функцію оптимального фільтра виділення. За яких співвідношень між параметрами спектрів такий фільтр ефективно виділятиме сигнал на фоні шуму?

11.3.2. Розв’яжіть задачу 11.3.1, якщо $G_{1,2}(\omega) = \frac{G_{10,20}}{(\omega-\omega_{10,20})^2 + \Omega_{10,20}^2}$.

11.4. Радіометри

Ми бачили, що лінійні фільтри не дозволяють виділити слабкі сигнали на фоні сильних шумів, якщо їхні спектри значно перекриваються. Однак у деяких випадках цю задачу, тим не менше, вдається вирішити – зокрема, коли вхідний сигнал також можна розглядати як стаціонарний шум (як ми це вже робили при аналізі оптимального фільтру виділення).

Отже, нехай сигнал, який нас цікавить, надходить до приймача ззовні, а в ролі шумів виступають власні шуми приймача – в першу чергу, тепловий і дробовий. Для виділення корисного сигналу можна використати той факт, що його можна час від часу вимикати.

Припустимо, що на виході приймача стоїть детектор, і вихідна напруга пропорційна дисперсії. Тоді при вимкненому вхідному сигналі ця дисперсія відповідає лише шуму, а при увімкненому – сумі сигналу й шуму. Тому їхня різниця дозволяє визначити інтенсивність сигналу.

Прилади, побудовані на такому принципі, називають радіометрами. Відомі два типи радіометрів – компенсаційний та модуляційний.

11.4.1. Компенсаційний радіометр: перенесення енергії сигналу в постійну складову

Компенсаційний радіометр працює за принципом, описаним вище (рис. 11.4.1). При розімкненому ключі пристрій запам'ятовування запам'ятовує рівень сигналу, що відповідає власним шумам, а при замкненому пристрій віднімання віднімає його від рівня вихідного сигналу.

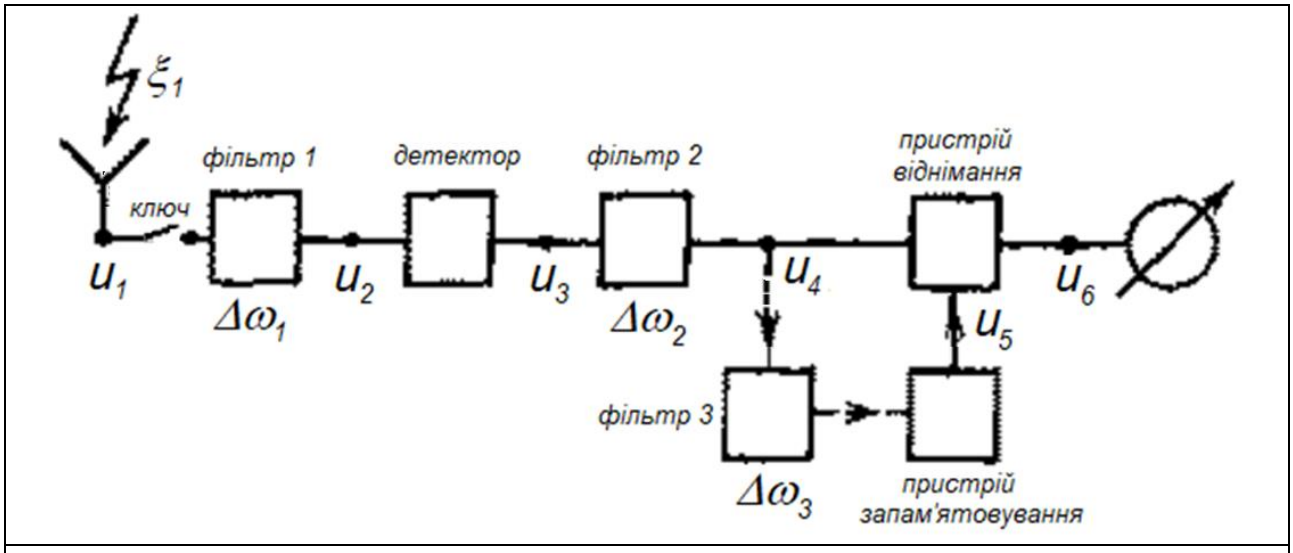


Рис. 11.4.1. Блок-схема компенсаційного радіометра: смуговий фільтр 1 пропускає весь спектр сигналу, ФНЧ 2 пропускає постійну складову сигналу, смуга пропускання ФНЧ 2 обирається якомога вужчою

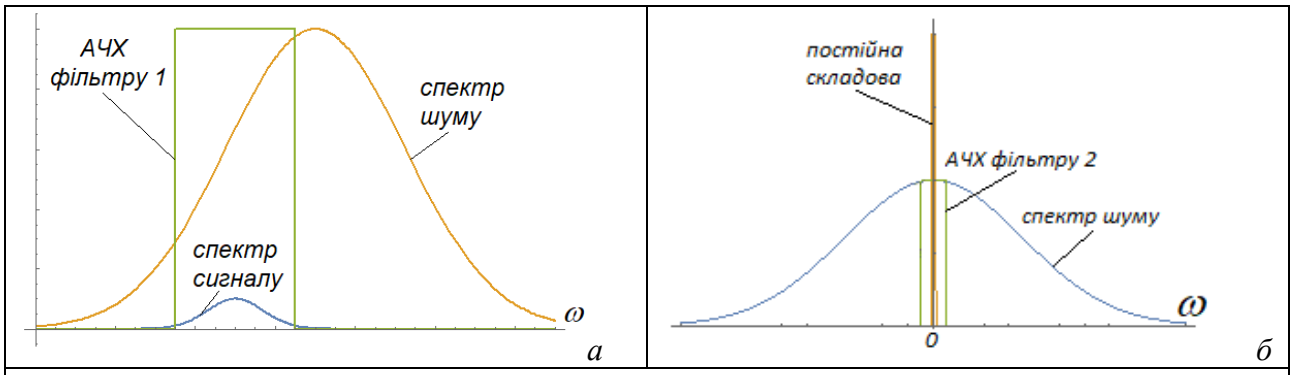


Рис. 11.4.2. До пояснення роботи компенсаційного радіометра: а – спектри сигналу та шуму на вході фільтру 1 і АЧХ цього фільтру; б – постійні складові, спричинені сигналом та шумом, і спектр шуму на вході фільтру 2 й АЧХ цього фільтру

Розглянемо роботу компенсаційного радіометра більш детально. Нехай ξ_1 – слабкий вхідний шум, який нам потрібно виміряти ($u_1 = \xi_1$), ξ_2 – власні шуми радіометра, приведені до його входу. Вважатимемо ці шуми гауссівськими й статистично незалежними.

Фільтр 1 – це, по суті, супергетеродинний радіоприймач. Його смуга пропускання $\Delta\omega_1$ підбирається так, щоб до неї потрапляв увесь спектр випадкового вхідного сигналу ξ_1. На його виході $u_2 = \xi_1 + \xi_2$ (рис. 11.4.2 а). На виході детектора, який вважатимемо квадратичним, маємо:

$$u_3 = \alpha u_2^2 = \alpha\left(\xi_1 + \xi_2\right)^2 = \alpha\left(\xi_1^2 + 2\xi_1\xi_2 + \xi_2^2\right). \tag{11.4.1}$$

$$u_3 = \langle u_3\rangle + \tilde{u}_3, \qquad \langle u_3\rangle = \alpha\left(\sigma_1^2 + \sigma_2^2\right). \tag{11.4.2}$$

Автокореляційна функція флуктуацій на виході детектора:

$$\begin{aligned}
&\langle\tilde{u}_3\tilde{u}_{3\tau}\rangle = \langle\alpha(\xi_1+\xi_2)^2 - \alpha(\sigma_1^2+\sigma_1^2)\rangle\langle\alpha(\xi_{1\tau}+\xi_{2\tau})^2 - \alpha(\sigma_1^2+\sigma_1^2)\rangle = \\
&= \alpha^2\langle[(\xi_1+\xi_2)^2 - (\sigma_1^2+\sigma_1^2)][(\xi_{1\tau}+\xi_{2\tau})^2 - (\sigma_1^2+\sigma_1^2)]\rangle = \\
&= \alpha^2\langle(\xi_1+\xi_2)^2(\xi_{1\tau}+\xi_{2\tau})^2 - (\sigma_1^2+\sigma_1^2)[(\xi_1+\xi_2)^2 + (\xi_{1\tau}+\xi_{2\tau})^2] + (\sigma_1^2+\sigma_1^2)^2\rangle = \\
&= \alpha^2\{\langle(\xi_1^2+2\xi_1\xi_2+\xi_2^2)(\xi_{1\tau}^2+2\xi_{1\tau}\xi_{2\tau}+\xi_{2\tau}^2)\rangle - (\sigma_1^2+\sigma_1^2)(2\sigma_1^2+2\sigma_1^2) + (\sigma_1^2+\sigma_1^2)^2\} =
\end{aligned} \tag{11.4.3}$$

$$\begin{aligned}
&= \alpha^2\left\{\left\langle\xi_1^2\xi_{1\tau}^2 + 2\xi_1^2\xi_{1\tau}\xi_{2\tau} + \xi_1^2\xi_{2\tau}^2 + 2\xi_1\xi_2\xi_{1\tau}^2 + 4\xi_1\xi_2\xi_{1\tau}\xi_{2\tau} + 2\xi_1\xi_2\xi_{2\tau}^2 + \xi_2^2\xi_{1\tau}^2 + 2\xi_2^2\xi_{1\tau}\xi_{2\tau} + \xi_2^2\xi_{2\tau}^2\right\rangle - \right. \\
&\left. -\left(\sigma_1^2+\sigma_2^2\right)^2\right\} = \alpha^2\left\{\left\langle\xi_1^2\xi_{1\tau}^2 + \xi_1^2\xi_{2\tau}^2 + 4\xi_1\xi_{1\tau}\xi_2\xi_{2\tau} + \xi_2^2\xi_{1\tau}^2 + \xi_2^2\xi_{2\tau}^2\right\rangle - \left(\sigma_1^2+\sigma_2^2\right)^2\right\} = \\
&= \alpha^2\left\{\left\langle\xi_1^2\xi_{1\tau}^2 + \xi_2^2\xi_{2\tau}^2\right\rangle + 2\sigma_1^2\sigma_2^2 + 4B_1(\tau)B_2(\tau) - \left(\sigma_1^2+\sigma_2^2\right)^2\right\}
\end{aligned}$$

Можна показати, що для гауссівських процесів

$$\left\langle\xi^2\xi_\tau^2\right\rangle = \sigma^4 + 2B^2(\tau) \tag{11.4.4}$$

(див. вище п. 9.3.2). Тоді

$$\begin{aligned}
&\langle\tilde{u}_3\tilde{u}_{3\tau}\rangle = \alpha^2\{\sigma_1^4 + \sigma_1^4 + 2B_1^2(\tau) + 2B_2^2(\tau) + 4B_1(\tau)B_2(\tau) + 2\sigma_1^2\sigma_1^2 - (\sigma_1^2+\sigma_1^2)^2\} = \\
&= \alpha^2\{2B_1^2(\tau) + 2B_2^2(\tau) + 4B_1(\tau)B_2(\tau)\} = 2\alpha^2[B_1(\tau)+B_2(\tau)]^2.
\end{aligned} \tag{11.4.5}$$

Врахуємо, що $B_1(\tau) \ll B_2(\tau)$. Отже, остаточно

$$\langle\tilde{u}_3\tilde{u}_{3\tau}\rangle = 2\alpha^2 B_2^2(\tau). \tag{11.4.6}$$

Спектральну інтенсивність цього шуму, відповідно до теореми Вінера – Хінчина, можна знайти, як перетворення Фур'є від його автокореляційної функції:

$$\begin{aligned}
&G_3(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\langle\tilde{u}_3\tilde{u}_{3\tau}\rangle exp(-i\omega\tau)d\tau = \frac{\alpha^2}{\pi}\int_{-\infty}^{\infty}B_2^2(\tau)exp(-i\omega\tau)d\tau = \\
&= \frac{\alpha^2}{\pi}\int_{-\infty}^{\infty}exp(-i\omega\tau)d\tau\int_{-\infty}^{\infty}G_2(\omega')exp(i\omega'\tau)d\omega'\int_{-\infty}^{\infty}G_2(\omega'')exp(i\omega''\tau)d\omega'' = \\
&= \frac{\alpha^2}{\pi}\int_{-\infty}^{\infty}G_2(\omega')d\omega'\int_{-\infty}^{\infty}G_2(\omega'')d\omega''\int_{-\infty}^{\infty}exp[-i(\omega-\omega'-\omega'')\tau]d\tau = \\
&= \frac{\alpha^2}{\pi}\int_{-\infty}^{\infty}G_2(\omega')d\omega'\int_{-\infty}^{\infty}G_2(\omega'')d\omega''\,2\pi\delta(\omega-\omega'-\omega'') = 2\alpha^2\int_{-\infty}^{\infty}G_2(\omega')G_2(\omega'-\omega)d\omega'.
\end{aligned} \tag{11.4.7}$$

На виході детектора енергія сигналу зосереджена в постійній складовій $\alpha\sigma_1^2$ (див. формулу (11.4.2)), енергія шуму – в постійній складовій $\alpha\sigma_2^2$ та спектрі $G_3(\omega)$ завширшки приблизно $4\Delta\omega_1$.

11.4.2. Компенсаційний радіометр: розрахунок відношення сигнал-шум

Щоб максимально позбутися шуму $G_3(\omega)$, після детектора стоїть фільтр низьких

частот зі смугою пропускання $2\Delta\omega_2 \ll \Delta\omega_1$ (рис. 11.4.2 б). Його передавальну функцію для простоти вважаємо прямокутною:

$$K_2(\omega) = \begin{cases} 1, |\omega| \le \Delta\omega_2; \\ 0, |\omega| > \Delta\omega_2. \end{cases} \tag{11.4.8}$$

У цій смузі $G_3(\omega) \approx G_3(0)$, тому на виході фільтру 2

$$\langle u_4 \rangle = \langle u_3 \rangle = \alpha\left(\sigma_1^2 + \sigma_2^2\right), \tag{11.4.9}$$

$$G_4(\omega) = G_3(0) K_2(\omega) = \begin{cases} 2\alpha^2 \int\limits_{-\infty}^{\infty} G_2^2(\omega')d\omega', |\omega| \le \Delta\omega_2 \\ 0, \quad |\omega| > \Delta\omega_2 \end{cases} \tag{11.4.10}$$

(враховано, що $G_2(-\omega') = G_2(\omega')$).

В моменти часу, коли ключ розімкнений, тобто $u_1 = 0$, на виході фільтру 2 постійна складова $\langle u_4 \rangle = \alpha\sigma_2^2$. Через фільтр низьких частот 3, смуга пропускання якого $2\Delta\omega_3$ обирається якомога вужчою, цей сигнал $u_5 = \langle u_4 \rangle = \alpha\sigma_2^2$ надходить на пристрій запам'ятовування, а після замикання ключа він віднімається від сигналу u_4. Таким чином, на виході радіометра

$$\langle u_6 \rangle = \langle u_4 \rangle - \alpha\sigma_2^2 = \alpha\sigma_1^2, \tag{11.4.11}$$

Тепер потужність шумів зосереджена лише в неперервному спектрі $G_6(\omega)$ і складає величину

$$P_n = G_6(\omega) 2\Delta\omega_2 = 4\alpha^2 \Delta\omega_2 \int\limits_{-\infty}^{\infty} G_2^2(\omega')d\omega'. \tag{11.4.12}$$

Таким чином, співвідношення сигнал-шум на виході радіометра складає

$$\left(\frac{s}{n}\right)_{out} = \frac{\langle u_6 \rangle^2}{P_n} = \frac{\alpha^2\sigma_1^4}{4\alpha^2\Delta\omega_2 \int\limits_{-\infty}^{\infty} G_2^2(\omega')d\omega'} = \frac{\sigma_1^4}{4\Delta\omega_2 \int\limits_{-\infty}^{\infty} G_2^2(\omega')d\omega'}. \tag{11.4.13}$$

Оцінимо σ_1^2 як $G_1 2\Delta\omega_1$ та

$$\int\limits_{-\infty}^{\infty} G_2^2(\omega')d\omega' = 2\Delta\omega_1 G_2^2,$$

де $G_{1,2}$ – середні значення $G_{1,2}(\omega)$ в смузі $\Delta\omega_1$. Тоді

$$\left(\frac{s}{n}\right)_{out} = \frac{\sigma_1^4}{4\Delta\omega_2 \int\limits_{-\infty}^{\infty} G_2^2(\omega')d\omega'} = \frac{\left(2G_1\Delta\omega_1\right)^2}{4\Delta\omega_2 G_2^2 2\Delta\omega_1} = \frac{\Delta\omega_1}{2\Delta\omega_2}\left(\frac{G_1}{G_2}\right)^2. \tag{11.4.14}$$

За умови, що $\Delta\omega_1/\Delta\omega_2 \gg 2(G_2/G_1)^2$, ця величина буде значно більшою за одиницю. Ми проаналізували роботу компенсаційного радіометра в припущенні, що корисний сигнал, який ми прагнемо виділити, є випадковим процесом. Але результати розрахунку виявляються справедливими й тоді, коли це буде регулярний процес, аби тільки виконувалася умова $\Delta\omega_1/\Delta\omega_2 \gg 1$. Справді, за виконання цієї умови на вихід радіометра потрапляє тільки постійна складова сигналу з виходу детектора, а змінна складова (шумова чи регулярна) на фоні шуму є несуттєвою.

11.4.3. Модуляційний радіометр

У розглянутій вище схемі компенсаційного радіометра на практиці величина $\alpha\sigma_2^2$ з часом може плавно змінюватися, що призведе до неправильного вимірювання інтенсивності сигналу. Цього недоліку позбавлена схема модуляційного радіометра (рис. 11.4.3). Модулятор на його вході модулює сигнал ξ_1. На його виході $u_1 = m(t)\xi_1 + \xi_2$, де $m(t)$ – періодична функція ($m(t+T) = m(t)$). Як і в компенсаційній схемі, смуговий фільтр Ф1 підбирається так, щоб повністю пропустити спектр сигналу ξ_1.

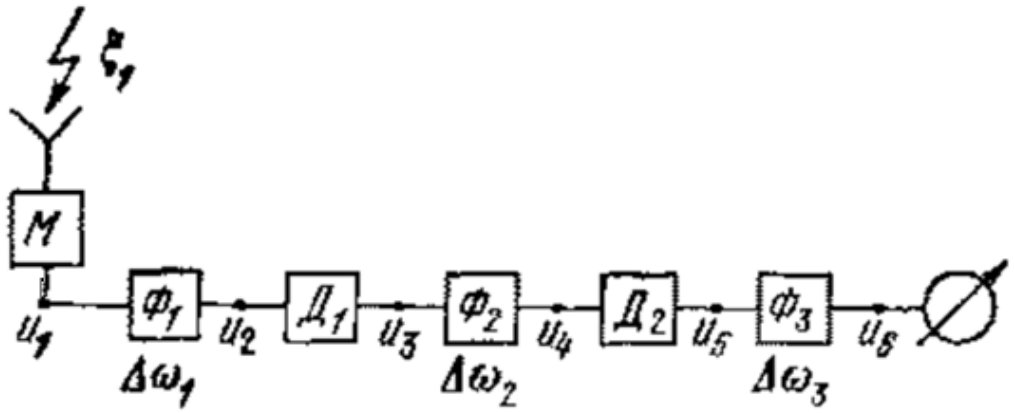


Рис. 11.4.3. Блок-схема модуляційного радіометра: смуговий фільтр 1 повністю пропускає спектр сигналу, вузькосмуговий фільтр 2 пропускає гармоніку частоти модуляції, ФНЧ 3 придушує другу гармоніку з виходу детектора 2

На виході квадратичного детектора Д1

$$u_3 = \alpha\left[m(t)\xi_1 + \xi_2\right]^2, \qquad \langle u_3\rangle = \alpha\left\langle m^2(t)\xi_1^2 + \xi_2^2\right\rangle = \alpha\left[m^2(t)\sigma_1^2 + \sigma_2^2\right]. \tag{11.4.15}$$

Розкладемо $m^2(t)$ в ряд Фур'є:

$$m^2(t) = \mu_0 + \sum_{n=1}^{\infty}\mu_n\cos(n\Omega t + \varphi_n), \qquad \Omega = 2\pi/T. \tag{11.4.16}$$

Тоді

$$\langle u_3\rangle = \alpha\left[m^2(t)\sigma_1^2 + \sigma_2^2\right] = \alpha\sigma_2^2 + \alpha\sigma_1^2\mu_0 + \alpha\sigma_1^2\sum_{n=1}^{\infty}\mu_n\cos(n\Omega t + \varphi_n). \tag{11.4.17}$$

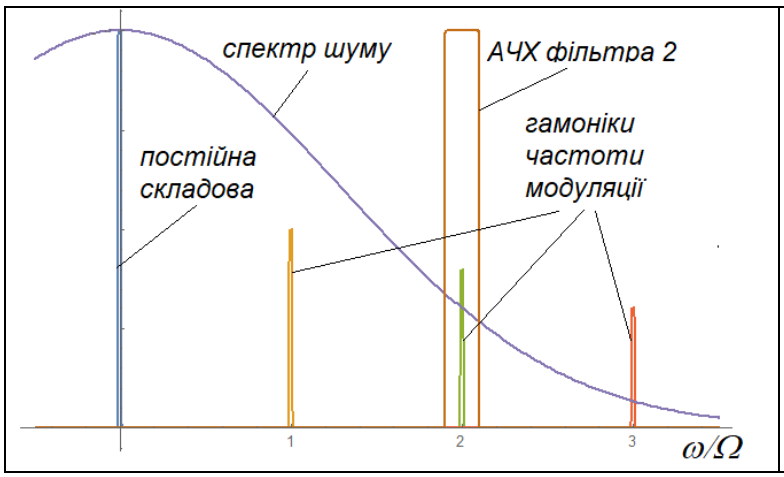


Рис. 11.4.4. До пояснення роботи модуляційного радіометра: вихід детектора 1

Фільтр Ф2 налаштований на одну з гармонік частоти Ω, а його смуга пропускання $\Delta\omega_2$

підбирається якомога вужчою, щоб пропустити якнайменше шумів (рис. 11.4.4). На його виході

$$\langle u_4 \rangle == \alpha\sigma_1^2 \mu_n \cos\left(n\Omega t + \varphi_n\right), \tag{11.4.18}$$

тобто регулярний сигнал, спричинений шумами, відсутній. Далі напруга u_4 надходить на квадратичний детектор Д2, фільтр низьких частот відфільтровує гармоніку $2n\Omega$, і на виході маємо

$$\langle u_6 \rangle == \frac{1}{2}\alpha^2\sigma_1^4\mu_n^2. \tag{11.4.19}$$

Для співвідношення сигнал-шум розрахунок дає:

$$\left(\frac{s}{n}\right)_{out} = \frac{\mu_n^2}{4}\frac{\Delta\omega_1}{2\Delta\omega_2}\left(\frac{G_1}{G_2}\right)^2. \tag{11.4.20}$$

Якщо прийняти, що закон модуляції вхідного сигналу гармонічний, тобто

$$m(t) = \cos\left(\Omega t + \varphi\right), \qquad m^2(t) = \frac{1}{2}\left[1 + \cos 2\left(\Omega t + \varphi\right)\right], \qquad \mu_0 = \mu_2 = \frac{1}{2}, \tag{11.4.21}$$

а фільтр 2 налаштований на другу гармоніку частоти модуляції, то

$$\left(\frac{s}{n}\right)_{out} = \frac{1}{32}\frac{\Delta\omega_1}{\Delta\omega_2}\left(\frac{G_1}{G_2}\right)^2. \tag{11.4.22}$$

Частота Ω підбирається так, щоб на її періоді Т величина $\alpha\sigma_2^2$ не встигала помітно змінитися.

Як і компенсаційний радіометр, модуляційна схема може виділяти й слабкі регулярні сигнали на фоні великих шумів.

11.4.4. Чутливість радіометра

Граничною чутливістю радіометра називають такий рівень сигналу на вході $G_{1\min}$, при якому відношення сигнал – шум на виході дорівнює одиниці. Для ідеального компенсаційного радіометра

$$G_{1\min} = \sqrt{2}G_2/q, \qquad q = \sqrt{\Delta\omega_1/\Delta\omega_2}, \tag{11.4.23}$$

відповідно для ідеального модуляційного радіометра

$$G_{1\min} = 4\sqrt{2}G_2/q, \tag{11.4.24}$$

а для звичайного радіоприймача

$$G_{1\min} = G_2. \tag{11.4.25}$$

Параметр q називають радіометричним виграшом. За порядком величини він показує, наскільки чутливість радіометра перевищує чутливість звичайного радіоприймача. На практиці радіометричний виграш складає $10^4 - 10^5$.

Зазначимо, що радіометр можна реалізувати не лише у вигляді аналогової схеми, але й у формі алгоритму обробки оцифрованого сигналу.

Контрольні питання до підрозділу 11.4

1. Для чого призначені радіометри?
2. Опишіть принцип роботи компенсаційного радіометра.
3. Порівняйте компенсаційний та модуляційний радіометри.
4. Що таке радіометричний виграш? Що він характеризує?

Задачі до підрозділу 11.4

11.4.1. Спектральна інтенсивність шуму на вході компенсаційного радіометра на порядок вища за спектральну інтенсивність корисного сигналу. Яким повинно бути відношення $\Delta\omega_1/\Delta\omega_2$, щоб забезпечити на виході співвідношення сигнал – шум, що дорівнює 10? Яким буде при цьому радіометричний виграш?

11.4.2. Розв'язати задачу 11.4.1 для модуляційного радіометра, налаштованого на другу гармоніку частоти модуляції.

Частина V. Елементи цифрової електроніки

Матеріал, поданий у цій частині, не претендує на скільки-небудь повне розкриття теми. Його призначення – перекинути місток між аналоговою та цифровою радіоелектронікою, детальне ознайомлення з якою передбачається в ряді інших курсів (або підручників).

Приблизно до 1980 р. радіоелектроніка мала справу майже виключно з аналоговими сигналами. Такі сигнали звичайно можна подати як неперервні гладенькі функції часу. Подібне подання відповідає звичайному протіканню процесів у природі (зміна положення тіла з часом, зміна температури й тиску повітря в атмосфері, зміна напруги на виході автогенератора тощо).

Однак пізніше аналогова електроніка майже повністю поступилася місцем цифровій. Цифрові електронні схеми обробляють цифрові сигнали або оцифровані аналогові сигнали. Діапазон сигналу в них представлений дискретними смугами (інтервалами) аналогового рівня, тобто не неперервною, а дискретною шкалою значень. Всі рівні в межах інтервалу являють собою той самий стан сигналу. Відносно невеликі зміни в аналоговому рівні сигналу в зв'язку з виробничими розбіжностями, ослабленням сигналу або паразитними шумами не виходять за дискретний інтервал, і, як наслідок ігноруються подальшою схемою, дозволяючи правильно сприйняти дискретний стан.

Для цифрових сигналів число дискретних станів зазвичай два, і вони представлені двома рівнями напруги: один біля опорного значення (зазвичай це «земля» або нуль вольт) і значення поблизу напруги живлення. Вони відповідають значенням «0» і «1» у двійковій системі.

Цифрові методи зручніші від аналогових: легше зробити електронний пристрій, який перемикається в один з відомих станів, ніж точно відтворити безперервний діапазон значень.

Цифрові електронні схеми, як правило, являють собою сукупність дуже великої кількості стандартних логічних елементів, причому кількість їхніх різновидів порівняно невелика.

Цифрові технології використовуються в обчислювальній електроніці, насамперед у комп'ютерах, у різних областях електротехніки, таких як телекомунікаційні системи, робототехніка, автоматизовані системи, вимірювальні прилади, цифровий звукозапис тощо.

Деякі поняття, які відносяться до застосування цифрової електроніки в задачах передавання даних, будуть обговорюватися в цій частині курсу.

Розділ 12. Поняття про цифрове передавання інформації

У цьому розділі ми постійно користуватимемося поняттями інформації та даних. Тому для початку зупинимося на їхньому змісті.

Інформація (від лат. informātiō «роз'яснення, уявлення, поняття про що-небудь») – це деякі відомості, незалежно від форми їхнього подання.

У своєму первісному значенні інформація – це відомості, що передаються людьми усним, письмовим або будь-яким іншим способом (за допомогою умовних сигналів, технічних засобів, тощо). Від середини XX століття термін «інформація» перетворився на загальнонаукове поняття, що включає обмін відомостями між людьми, людиною та автоматом, автоматом та автоматом; обмін сигналами між живими істотами; передачу ознак від клітини до клітини, від організму до організму (наприклад, генетична інформація).

Згідно з сучасними уявленнями, інформація вважається нематеріальною, а те, що міститься в структурі об'єктів, прийнято називати *даними* (англ. – representation form). На найфундаментальнішому рівні інформація стосується інтерпретації того, що можна відчути. Можна сказати, що будь-який природний процес, який не є повністю випадковим, і будь-який спостережуваний об'єкт чи явище у будь-якому середовищі передають певну кількість інформації.

Поняття даних (від лат. data – множина) у областях має різний науково формалізований

зміст та сутність. В окремих випадках воно може застосовуватись як синонім до подібних за змістом термінів «інформація», «відомості», «повідомлення», «сигнали», «коди» тощо):

Ми розумітимемо *дані* як форму представлення знань, інформації; тексти, таблиці, інструкції, відомості про факти, явища й таке інше, представлені у формі літер та цифр, чисел, текстів, звуків або графіків. Дані можуть зберігатися на різних носіях, в тому числі в комп'ютерах, пересилатися і піддаватися обробці.

У вужчому сенсі дані – це інформація (найчастіше цифрова), подана у формалізованому вигляді, прийнятному для оброблення автоматичними засобами за можливої участі людини.

В інформатиці дані символізують інформацію, що представлена у вигляді, необхідному для її опрацювання автоматичними засобами. Для цього інформацію кодують за допомогою знаків (алфавітів) відповідно до правил певного синтаксису. У сучасних комп'ютерах використовується двійковий метод запису даних за допомогою 0 та 1 (алфавіт з двох цифр). Для полегшення роботи для людини двійковий код перекодовується у зрозуміліші числа, букви тощо.

Хоча інформація має набути деякої форми подання (тобто перетворитися на дані), щоб її можна було зберігати або обмінюватися, інформація є, в першу чергу, інтерпретація (сенс) такого подання, тому, строго кажучи, інформація відрізняється від даних, хоча у неформальному контексті ці два терміни дуже часто використовують як синоніми.

12.1. Дискретизація та квантування сигналів

Прикладом, який демонструє можливості цифрових методів, є цифрове кодування аналогових сигналів для їхньої передачі без спотворень, пов'язаних зі впливом завад. Як ми бачили, до аналогових сигналів, що передаються по лініях зв'язку або через ефір, додаються завади, які потім важко відокремити від корисного сигналу. Якщо ж переданий сигнал перетворити на ряд чисел, що визначають його амплітуду в послідовні моменти часу, а потім ці числа передавати у вигляді цифрових сигналів, то аналоговий сигнал, відновлений на приймальній стороні (за допомогою так званого цифро-аналогового перетворювача), не міститиме помилок, якщо рівень шуму в каналі зв'язку не настільки високий, щоб перешкоджати розпізнаванню «одиниць» і «нулів». Цей метод, відомий під назвою імпульсно-кодової модуляції, особливо ефективний, коли сигнал повинен проходити через ряд ретрансляторів, наприклад, у радіорелейній лінії чи в системах міжконтинентального телефонного зв'язку, оскільки відновлення цифрового сигналу в кожному пункті ретрансляції гарантує завадостійкість передачі. Цифрові методи використовують і космічні зонди, які за допомогою імпульсно-кодової модуляції передають на Землю зображення та інші дані.

Цифрова електроніка оперує імпульсними сигналами. Зокрема, неперервні залежності величин від часу замінюються імпульсними послідовностями (рис. 12.1.1). При цьому відбувається дискретизація сигналу в часі (замість неперервного часу беруть періодичну послідовність моментів часу).

Наступний крок – це квантування амплітуди сигналів. Цей процес аналогічний округленню чисел до певних дискретних значень. Іншими словами, неперервні значення амплітуди набувають дискретного набору таких значень, і точні величини амплітуди імпульсів замінюють найближчими до них значеннями зі згаданого дискретного набору.

12.1.1. Поняття про дискретизацію, квантування та оцифрування сигналів

Дискретизований у часі сигнал може бути представлений у вигляді імпульсної послідовності, причому тривалість окремих імпульсів може бути значно меншою, ніж проміжок часу між сусідніми імпульсами (рис. 12.1.2).

Амплітуди імпульсів, які утворюють імпульсну послідовність, квантуються, тобто набувають дискретного набору значень (рис. 12.1.3). Потім вони кодуються, тобто подаються у

формі числа. Такий підхід пов'язаний із появою й удосконаленням комп'ютерів, здатних запам'ятовувати й обробляти масиви числових даних з високою швидкістю й точністю. Передавання даних також здійснюється в дискретизованому або цифровому форматі, який забезпечує кращий захист від завад і шумів.

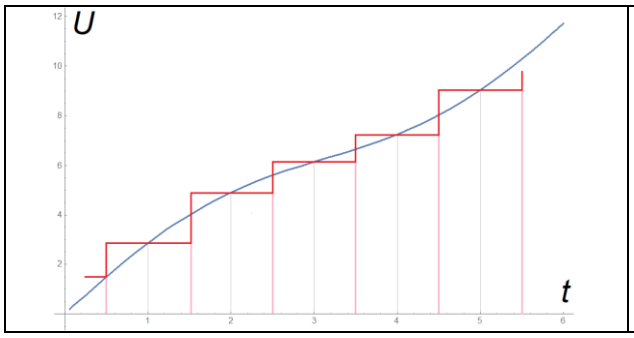


Рис. 12.1.1. Дискретизація сигналу в часі. Синя лінія – аналоговий сигнал, червона – сигнал, дискретизований за часом. Сірими лініями показані дискретні значення часу, рожевими – межі проміжків навколо цих значень.

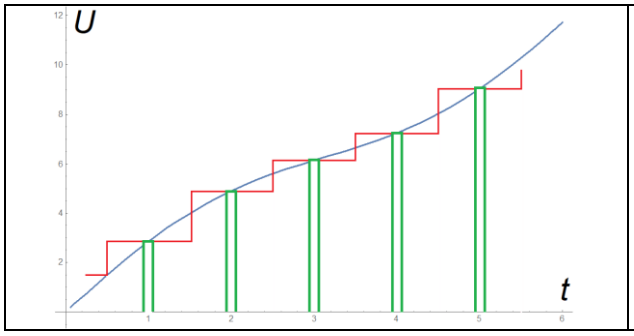


Рис. 12.1.2. Формування імпульсної послідовності (показана зеленим) на основі дискретизованого сигналу

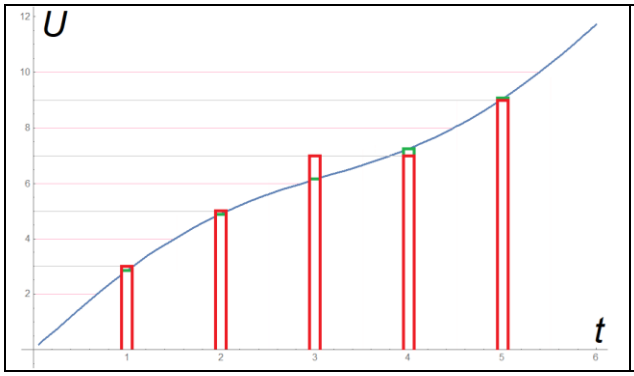


Рис. 12.1.3. Квантування рівня сигналу: зеленим показана імпульсна послідовність, що відповідає дискретизації за часом, червоним – та сама послідовність, квантована за рівнем сигналу. Синя лінія – первісний аналоговий сигнал, сірим показані дискретні значення рівня, рожевим – межі проміжків навколо цих значень.

12.1.2. Формування імпульсної послідовності, що відповідає дискретизації за часом

Для формування імпульсної послідовності, що відповідає дискретизації сигналу за часом, на вхід підсилювача подають суму цього сигналу та сигналу від генератора строб-імпульсів, тобто прямокутних імпульсів, тривалість яких значно менша за період повторення (рис. 12.1.4 а). Параметри схеми підбираються так, щоб за подання на вхід лише сигналу вхід підсилювача був запертий, а при поданні суми сигналу й прямокутних імпульсів робоча точка потрапляла на лінійну ділянку динамічної характеристики підсилювача (динамічна характеристика – це залежність вихідної напруги від вхідної). Роботу такої схеми ілюструють рис. 12.1.4 б-в.

Інший варіант схеми такого самого призначення на МДН-транзисторі поданий на рис. 12.1.5. Тут прямокутні імпульси від генератора виступають як напруга живлення, а сигнал подається на затвор. Тому на виході маємо послідовність прямокутних імпульсів, амплітуда яких пропорційна вхідному сигналу у відповідні моменти часу.

Обернена задача перетворення імпульсної послідовності на аналоговий сигнал розв'язується шляхом пропускання вказаної послідовності через фільтр низьких частот,

гранична частота якого дорівнює частоті повторення імпульсів.

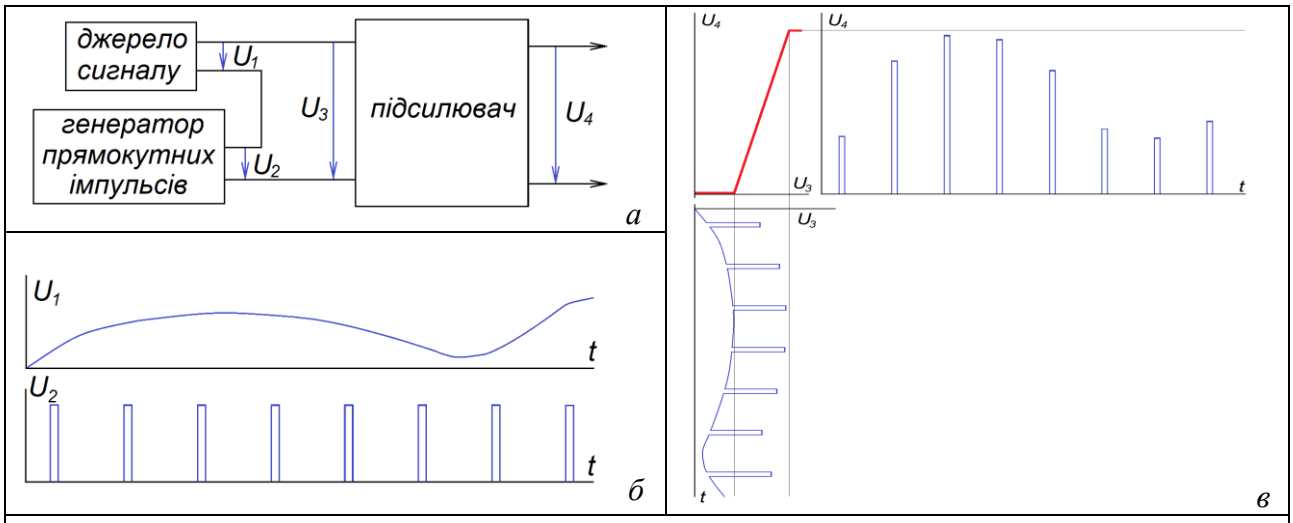


Рис. 12.1.4: а – блок-схема для отримання дискретизованої за часом імпульсної послідовності з використанням підсилювача; б – аналоговий сигнал та сигнал з генератора прямокутних імпульсів; в – динамічна характеристика підсилювача, сигнали на його вході (U_3) та виході (U_4)

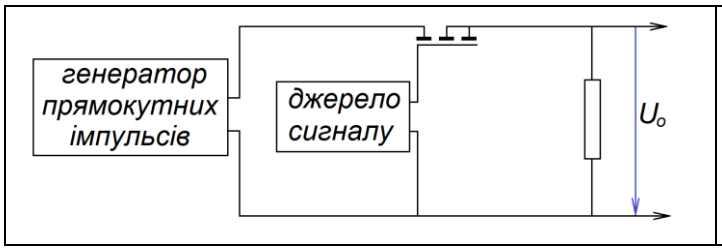


Рис. 12.1.5. Блок-схема для отримання дискретизованої за часом імпульсної послідовності на МДН-транзисторі

12.1.3. Теорема відліків Віттекера – Найквіста – Котельникова – Шеннона

При дискретизації сигналу за часом постає питання про те, якими треба обирати проміжки часу між відліками, або яким повинен бути період повторення вихідної імпульсної послідовності. Зрозуміло, що чим менше цей період, тим точніше буде відтворюватися аналоговий сигнал. З іншого боку, надто малий період призведе до різкого збільшення обсягу інформації про сигнал, яку треба запам'ятати.

Відповідь на питання про оптимальний крок дискретизації аналогового сигналу за часом дає теорема відліків Віттекера – Найквіста – Котельникова – Шеннона: функція часу з обмеженим частотним спектром повністю задається рядом дискретних значень (відліків), взятих через проміжки часу

$$\Delta t \le 1/2 f_{\max}, \qquad (12.1.1)$$

де $f_{\max}$ – максимальна частота в спектрі сигналу (її називають частотою Найквіста).

Іншими словами, на період спектральної складової сигналу з максимальною частотою має припадати хоча б два відліки.

12.1.4. Відлікові функції

Для доведення теореми відліків потрібно розкласти вхідний сигнал по так званих відлікових функціях вигляду

$$F_k(t,\omega_0) = \sqrt{\frac{\omega_0}{\pi}}\frac{\sin(\omega_0 t - k\pi)}{(\omega_0 t - k\pi)} = \sqrt{\frac{\omega_0}{\pi}}\,\mathrm{sinc}(\omega_0 t - k\pi), \qquad k = 0, \pm 1, \pm 2, \ldots \qquad (12.1.2)$$

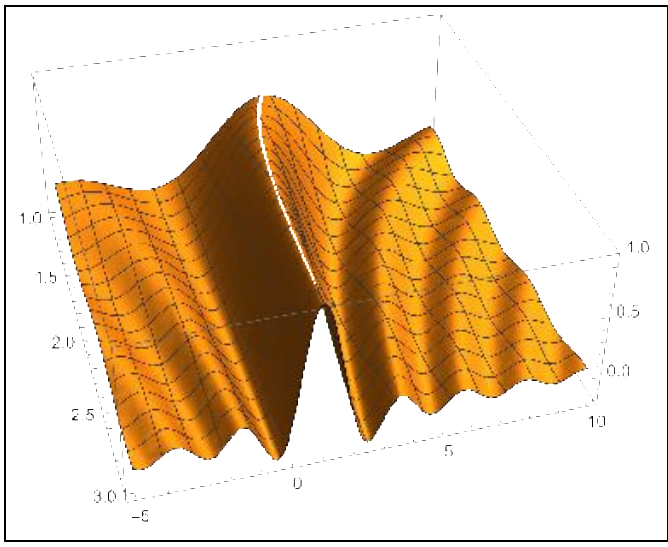


Рис. 12.1.4. Графік відлікової функції $F_k(t,\omega_0)$ при $k=1$ та зміні ω_0 від 1 до 3 і при зміні t від -5 до 10

Але спочатку треба показати, що набір таких функцій утворює ортонормований базис – так званий базис Котельникова. Для цього треба довести, що

$$\int_{-\infty}^{\infty} F_k(t,\omega_0)F_{k'}(t,\omega_0)\,dt = \int_{-\infty}^{\infty}\sqrt{\frac{\omega_0}{\pi}}\frac{\sin(\omega_0 t-k\pi)}{\omega_0 t-k\pi}\sqrt{\frac{\omega_0}{\pi}}\frac{\sin(\omega_0 t-k'\pi)}{\omega_0 t-k'\pi}\,dt = \begin{cases}1, & k'=k;\\ 0, & k'\neq k.\end{cases} \qquad (12.1.3)$$

Безпосередньо розрахувати інтеграл (12.1.3) не вдається, тому скористаємося для цього формулою Рейлі (див. вище п. 1.3.4):

$$\int_{-\infty}^{\infty} f_1(t)f_2(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} f_1(\omega)f_2^*(\omega)\,d\omega, \qquad (12.1.4)$$

де $f_{1,2}(t)$ – довільні функції часу, $f_{1,2}(\omega)$ – їхні спектри.

Раніше (див. задачу П. 1.3.1) було показано, що спектр прямокутного імпульсу тривалістю τ

$$f_0(t) = \begin{cases}1, & |t|\le \tau/2;\\ 0, & |t|>\tau/2\end{cases} \qquad (12.1.5)$$

має вигляд

$$f_0(\omega) = \int_{-\infty}^{\infty} f(t)exp(-i\omega t)dt = \int_{-\tau/2}^{\tau/2} exp(-i\omega t)dt = \tau\frac{sin(\omega\tau/2)}{\omega\tau/2}. \qquad (12.1.6)$$

Легко бачити, що зсув початку імпульсу на час t_0,

$$f(t,t_0) = \begin{cases}1, & |t-t_0|\le \tau;\\ 0, & |t-t_0|>\tau,\end{cases} \qquad (12.1.7)$$

приводить до деякої зміни спектру:

$$f(\omega,t_0) = \int_{t_0-\tau/2}^{t_0+\tau/2} exp(-i\omega t)dt = \left\|\begin{matrix}t-t_0=\theta\\ t=\theta+t_0\end{matrix}\right\| = \int_{\tau/2}^{\tau/2} exp[-i\omega(\theta+t_0)]d\theta ==$$

$$exp(-i\omega t_0)\int_{\tau/2}^{\tau/2} exp(-i\omega\theta)d\theta = \tau\frac{sin(\omega\tau/2)}{\omega\tau/2}exp(-i\omega t_0). \qquad (12.1.8)$$

Тепер розглянемо деяку функцію $f(t)$ і її спектр $s(\omega)$, які пов'язані співвідношеннями

$$s(\omega) = \int_{-\infty}^{\infty} f(t)\exp(-i\omega t)\,dt, \qquad f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} s(\omega)\exp(i\omega t)\,d\omega. \qquad (12.1.9)$$

Далі розглянемо деяку іншу функцію $f_1(t)$, форма якої повторює форму спектру $s(\omega)$:

$$f_1(t) = s(\omega)\big|_{\omega=t}. \qquad (12.1.10)$$

Очевидно, спектр функції $f_1(t)$ можна подати у вигляді

$$s_1(\omega)=\int_{-\infty}^{\infty} f_1(t)\exp(-i\omega t)\,dt\,. \qquad (12.1.11)$$

Тепер ліворуч і праворуч замінимо t на ω, а ω на t. Отримаємо з урахуванням формули (12.1.10):

$$s_1(t)=\int_{-\infty}^{\infty} f_1(\omega)\exp(-i\omega t)\,d\omega=\int_{-\infty}^{\infty} s(\omega)\exp(-i\omega t)\,d\omega=2\pi\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} s(\omega)\exp(-i\omega t)\,d\omega\right]. \qquad (12.1.12)$$

Якщо $s(\omega)$ – парна функція, то внесок до інтегралу дає лише парна частина $\exp(-i\omega t)$, і цю експоненту можна замінити комплексно спряженою величиною. Маємо в цьому випадку:

$$s_1(t)=2\pi\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} s(\omega)\exp(i\omega t)\,d\omega\right]=2\pi f(t)\,, \qquad (12.1.13)$$

тобто спектр функції $f_1(t)$ повторює форму функції $f(t)$ з точністю до константи 2π.

Тепер покладемо, що

$$f_1(t)=f_0(\omega)|_{\omega=t,\tau=\omega_0}=\tau\frac{sin(\omega\tau/2)}{\omega\tau/2}\Big|_{\omega=2t,\tau=\omega_0}=\omega_0\frac{sin(\omega_0 t)}{\omega_0 t}. \qquad (12.1.14)$$

Врахуємо, що спектр $f_0(\omega)$ – парна функція часу. Тоді відповідно до формул (12.1.13) та (12.1.5), можна записати, що

$$s_1(\omega)=2\pi f_0(t)|_{t=\omega,\tau=2\omega_0}=\begin{cases}2\pi, |\omega|\le\omega_0;\\ 0, |\omega|>\omega_0.\end{cases} \qquad (12.1.15)$$

Отже, для відлікової функції $F_k(t,\omega_0)$, яка відрізняється від $f_1(t)$ (12.1.14) множником $1/\sqrt{\pi\omega_0}$ і зсувом часу на $t_k=k\pi/\omega_0$, спектр має вигляд

$$F_k(\omega,\omega_0)=\sqrt{\frac{\pi}{\omega_0}}\exp(-i\omega t_k)\begin{cases}1, & |\omega|\le\omega_0;\\ 0, & |\omega|>\omega_0.\end{cases} \qquad (12.1.16)$$

Перейдемо нарешті до розрахунку інтеграла (12.1.3) з урахуванням того, що $t_k=k\pi/\omega_0$, і формули Рейлі (12.1.4):

$$\int_{-\infty}^{\infty} F_k(t,\omega_0)F_{k'}(t,\omega_0)\,dt=\frac{1}{2\pi}\int_{-\infty}^{\infty} F_k(\omega,\omega_0)F_{k'}^*(\omega,\omega_0)\,d\omega=$$

$$=\frac{1}{2\pi}\int_{-\omega_0}^{\omega_0}\sqrt{\frac{\pi}{\omega_0}}\exp(-i\omega t_k)\sqrt{\frac{\pi}{\omega_0}}\exp(i\omega t_{k'})\,d\omega=$$

$$=\frac{1}{2\omega_0}\int_{-\omega_0}^{\omega_0}\exp\left[i\omega(t_{k'}-t_k)\right]d\omega=\frac{1}{2\omega_0}\frac{\exp\left[i\omega(t_{k'}-t_k)\right]}{i(t_{k'}-t_k)}\Bigg|_{-\omega_0}^{\omega_0}= \qquad (12.1.17)$$

$$=\frac{1}{2\omega_0}\frac{\exp\left[i\omega_0(t_{k'}-t_k)\right]-\exp\left[-i\omega_0(t_{k'}-t_k)\right]}{i(t_{k'}-t_k)}=\frac{\sin\left[\omega_0(t_{k'}-t_k)\right]}{\omega_0(t_{k'}-t_k)}=$$

$$=\frac{\sin\left[\pi(k'-k)\right]}{\pi(k'-k)}=\begin{cases}1, k'=k;\\ 0, k'\ne k.\end{cases}$$

Отже, відлікові функції справді є ортонормованими і утворюють базис (як уже згадувалося, його називають базисом Котельникова).

12.1.5. Доведення теореми відліків

З ортонормованості відлікових функцій випливає, що довільну однозначну функцію часу $f(t)$ можна подати у формі

$$f(t)=\sum_{k=-\infty}^{\infty} C_k F_k(t,\omega_0). \qquad (12.1.18)$$

Для знаходження коефіцієнтів цього ряду домножимо обидві частини (12.1.18) на $F_{k'}(t,\omega_0)$ і проінтегруємо за часом у нескінченних межах. З урахуванням (12.1.17) отримаємо:

$$C_k=\int_{-\infty}^{\infty} f(t)F_k(t,\omega_0)\,dt\,. \qquad (12.1.19)$$

Знову скористаємося формулою Рейлі (12.1.4), формулою (2.1.16)

$$F_k(\omega,\omega_0)=\sqrt{\frac{\pi}{\omega_0}}\exp(-i\omega t_k)\begin{cases}1, & |\omega|\le\omega_0;\\ 0, & |\omega|>\omega_0\end{cases}$$

і тим, що спектр функції $f(t)$ обмежений:

$$\begin{aligned}C_k&=\int_{-\infty}^{\infty} f(\omega)F_k(\omega,\omega_0)\,d\omega=\int_{-\omega_0}^{\omega_0} f(\omega)F_k(\omega,\omega_0)\,d\omega=\\ &=\int_{-\omega_0}^{\omega_0} f(\omega)\exp(-i\omega t_k)\,d\omega=\int_{-\infty}^{\infty} f(\omega)\exp(-i\omega t_k)\,d\omega=\sqrt{\frac{\pi}{\omega_0}}f(t_k).\end{aligned} \qquad (12.1.20)$$

Остання рівність справедлива, якщо спектр $f(\omega)$ повністю вміщується на інтервалі $[-\omega_0,\omega_0]$, тобто максимальна частота в цьому спектрі не перевищує ω_0. Справді, в цьому випадку межі інтегрування в останньому інтегралі можна розширити до нескінченності, а сам інтеграл набуває форми перетворення Фур'є. При цьому крок за часом між сусідніми відліками складає величину $\Delta t=\pi/\omega_0=1/2f_0$ ($f_0=\omega_0/2\pi$), тобто частота відліків – $2f_0$.

Таким чином, набір відліків у точках $t_k=k\pi/\omega_0$ повністю задає функцію $f(t)$, якщо максимальна частота максимальної гармоніки в її спектрі не перевищує частоту Найквіста ω_0. Отже, теорема відліків доведена.

Як випливає з теореми відліків, дискретизований сигнал (або відповідна йому імпульсна послідовність) дає повну інформацію про початковий аналоговий сигнал лише тоді, коли спектр останнього обмежений згори. В багатьох випадках реальні сигнали мають спектри, які на високих частотах спадають, але не обертаються на нуль. Для дискретизації таких сигналів їх треба спершу пропустити через фільтр нижніх частот.

12.1.6. Шуми квантування

Як уже згадувалося, перед цим аналоговий сигнал потрібно дискретизувати за часом і квантувати за величиною. Почнемо з аналізу впливу заокруглення відліків при квантуванні їхньої величини на властивості сигналу.

Якщо спектр сигналу обмежений згори і набір відліків відповідає умовам теореми відліків, то дискретизована за часом імпульсна послідовність несе повну інформацію про відповідний сигнал. Однак квантування сигналу за величиною (рис. 12.1.3) призводить до неточності його зворотного відтворення, відповідно до розкладу сигналу по базису Котельникова (див. формулу (12.1.18)). Цю неточність можна інтерпретувати як появу деяких шумів – шумів квантування сигналу.

Очевидно, рівень шумів квантування зменшується зі зменшенням кроку квантування. Природною нижньою межею кроку квантування є рівень внутрішніх шумів або роздільна

здатність вимірювальних приладів.

Для оцінки рівня шумів квантування припустимо, що справжнє значення вимірюваної величини з однаковою ймовірністю може потрапити в будь-яку точку інтервалу квантування, довжину якого позначимо через a. При малих a таке припущення завжди відповідає дійсності. Тоді дисперсію процесу (фактично, його потужність) можна записати як

$$\sigma^2 = \langle \tilde{x}^2 \rangle = \int_{-a/2}^{a/2} \tilde{x}^2 w(\tilde{x}) d\tilde{x} = \int_{-a/2}^{a/2} \tilde{x}^2 \frac{1}{a} d\tilde{x} = \frac{1}{a} \frac{\tilde{x}^3}{3} \Big|_{-a/2}^{a/2} = \frac{a^2}{12}. \qquad (12.1.21)$$

Відповідно, середньоквадратична похибка, зумовлена квантуванням, складає

$$\sigma = \frac{a}{\sqrt{12}} \approx 0.29a\,. \qquad (12.1.23)$$

Припустимо, що значення сигналу лежать в інтервалі $[0, sa]$. Ймовірність розподілу, як і раніше, вважаємо рівномірною (насправді в даному разі це припущення є не дуже реалістичним). Потужність такого сигналу можна розрахувати аналогічно до (12.1.21):

$$P = \langle x^2 \rangle = \int_0^{sa} x^2 w(x) dx = \int_0^{sa} x^2 \frac{1}{sa} dx = \frac{1}{sa} \frac{x^3}{3} \Big|_0^{sa} = \frac{s^2 a^2}{3}\,. \qquad (12.1.24)$$

Тоді потужність квантованого сигналу буде

$$P_s = P - \sigma^2 = \frac{s^2 a^2}{3} - \frac{a^2}{12} = \frac{(4s^2 - 1) a^2}{12}\,, \qquad (12.1.26)$$

а відношення сигнал-шум –

$$\frac{s}{n} = \frac{P_s}{\sigma^2} = 4s^2 - 1\,. \qquad (12.1.27)$$

Отже, співвідношення сигнал-шум зростає при зменшенні кроку квантування (тобто при збільшенні кількості кроків s на заданому інтервалі значень вимірюваної величини).

Так, при передаванні телефонних розмов кількість рівнів складає кілька сотень, при високоякісному відтворенні музики – на порядок більше.

12.1.7. Основні види імпульсної модуляції

Досі, коли ми говорили про подання аналогового сигналу у вигляді імпульсної послідовності, йшлося про те, що інформація про сигнал закладена в амплітудах імпульсів. Це так звана амплітудна імпульсна модуляція (рис. 12.1.5 а). Однак існують і інші способи передачі інформації за допомогою імпульсної послідовності, які краще захищені від завад порівняно з амплітудною модуляцією. Краща захищеність пов'язана з незмінною амплітудою імпульсів, послідовність яких переносить інформацію.

Корисну інформацію може нести тривалість імпульсів, яка змінюється з часом пропорційно до аналогового сигналу (рис. 12.1.5 б). В цьому випадку говорять про широтну імпульсну модуляцію.

Інформацію про корисний сигнал може також нести зсув окремих імпульсів щодо немодульованої (строго періодичної) послідовності (рис. 12.1.5 в). Очевидно, такий зсув за часом не може перевищувати половину періоду повторення імпульсів немодульованої послідовності, а за фазою – значення π. В цьому випадку говорять про фазову імпульсну послідовність.

Нарешті, інформацію може нести частота повторення імпульсів, яка змінюється пропорційно корисному сигналу (рис. 12.1.5 г). Такий вид імпульсної модуляції називають частотним.

Інколи застосовують інші типи імпульсної модуляції.

При передачі інформації на віддаль за допомогою імпульсної послідовності звичайно використовують не відеоімпульси, а радіоімпульси з певною частотою-носієм.

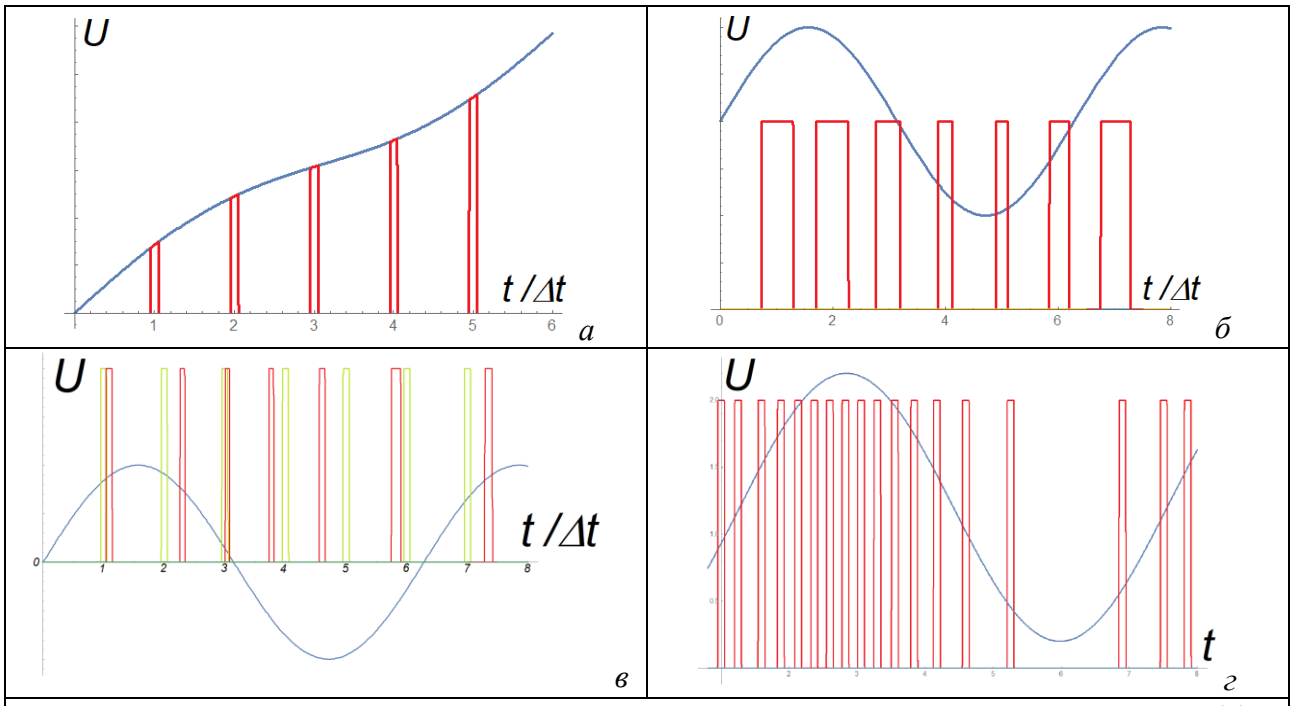


Рис. 12.1.5. Основні різновиди імпульсної модуляції: амплітудна імпульсна модуляція (а), широтна імпульсна модуляція (б), фазова імпульсна модуляція (в), частотна імпульсна модуляція (г). Синім показаний аналоговий сигнал, червоним – модульована імпульсна послідовність, зеленим – немодульована імпульсна послідовність

Контрольні питання до підрозділу 12.1

1. Чому в даний час цифрова електроніка значно поширеніша від аналогової?
2. Що розуміють під дискретизацією сигналу в часі?
3. Що розуміють під квантуванням сигналу за рівнем?
4. Чим визначається мінімальний крок дискретизації за часом?
5. Для яких сигналів застосовна теорема Котельникова?
6. Що таке відлікові функції?
7. Що таке базис Котельникова?

Задачі до підрозділу 12.1

12.1.1. Вхідний аналоговий сигнал $U_1(t)$ змінюється в межах від 0 до $U_{1\max}$, прямокутні імпульси мають амплітуду $U_{2\max}$ (див. рис. 12.1.4). В яких межах $U_{3\min}$, $U_{3\max}$ повинна лежати лінійна ділянка динамічної характеристики підсилювача, щоб забезпечити коректну дискретизацію вхідного аналогового сигналу?

12.1.2. Якою повинна бути мінімальна частота дискретизації амплітудно-модульованого сигналу з частотою-носієм ω_0 і частотою обвідної Ω?

12.1.3. Інтервал від 0 до 2 В квантується на 10 рівнів. Знайдіть середньоквадратичну похибку, зумовлену квантуванням, та співвідношення сигнал-шум для квантованого сигналу. Шум зумовлений квантуванням.

12.2. Обсяг цифрових даних

При цифровому представленні сигналів наступний крок після дискретизації сигналу за часом та його квантування за рівнем – це формування послідовності чисел, яка несе інформацію про послідовний набір квантованих значень рівня сигналу.

В цьому підрозділі ми поговоримо про те, як числа записуються за допомогою цифр (або інших умовних символів), зупинимося детальніше на двійковій системі числення і з'ясуємо, як визначається поняття кількості інформації.

12.2.1. Системи числення

Системою числення називається сукупність правил і знаків, за допомогою яких можна відобразити (кодувати) будь-яке невід'ємне число.

Протягом історії людства існували два основні підходи до запису чисел – непозиційний та позиційний.

Історично першими виникли непозиційні системи числення. Вони ґрунтуються на кількісному підході до визначення числа, який для кодування тих чи інших кількостей застосовував особливі знаки — числа. Кожному такому знаку відповідав кількісний еквівалент.

Одна з перших непозиційних систем числення існувала в Давньому Єгипті від середини III тис. до н.е. до кінця I тис. н.е. Спеціальними символами позначалися числа 1, 10, 100 тощо – до 1 млн.

У давньому Римі також використовували непозиційну систему числення. Числа різної величини позначали літерами (1 – I, 5 – V, 10 – X, 50 – L, 100 – C, 500 – D, 1000 – M), а записане число в найпростішому випадку, коли цифри розміщуються в порядку спадання, дорівнює сумі чисел, які його складали: наприклад, XXVII=10+10+5+1+1=27. Менша цифра може бути записана і ліворуч від більшої, тоді її слід відняти від більшої, що йде за нею. Наприклад, XIX=10+(10-1)=19.

У давній Русі-Україні використовувалася непозиційна система числення, де цифри позначалися кириличними літерами: 1 – А (аз), 2 – В (віди), 3 – Г (глаголь), 4 – Д (добро), тощо. Після одиниць ішли десятки й сотні. Для запису тисяч та більших чисел до літер домальовувалися спеціальні значки (рис. 12.2.1 а). Щоб відрізняти в тексті літери від цифр, над літерами з числовим значенням писали спеціальний значок – титло.

Так позначалися, наприклад, роки (до кінця 17 століття): символ, поданий на рис. 12.2.1 б, означає (7×1000+100+8+10; значок перед першою літерою означає тисячі, пор. з рис. 12.2.1 а) – це 7118 рік «від створення світу», або 7118–5508=1610 рік «від Різдва Христова» (Різдво Христове припадає на 5508 рік «від створення світу»).

У найпоширеніших сьогодні позиційних системах числення одна й та сама цифра (числовий знак) у записі числа набуває різних значень залежно від своєї позиції. Таким чином, позиція цифри має вагу у числі. Здебільшого вага кожної позиції кратна деякому натуральному числу.

Таким чином, у позиційних системах числення ціле число представляється у формі

$$N = A_0k^0 + A_1k^1 + A_2k^2 + A_3k^3 + A_4k^4 + \ldots \qquad (21.5)$$

(для будь-якого скінченного N цей ряд виявляється скінченним). Ціле число k ($k \geq 2$) називається основою системи числення, а цілі коефіцієнти A_i (цифри) набувають значень від 0 до $k-1$. Відповідне число записується як набір цифр, значення яких залежать від їхньої позиції. Найвідомішим прикладом позиційної системи є десяткова ($k=10$). Наприклад, в числі 777 перша з правого боку сімка має значення $7\times10^0=7$, друга – $7\times10^1=70$, третя – $7\times10^2=700$.

Десяткова система числення, як вважають, була запропонована індусами в I-IV століттях нової ери. Основою цієї системи є число пальців на руках. З Індії десяткова система поширилася в арабських країнах. До Західної Європи вона була занесена арабами наприкінці I тис.

н.е., тому наші цифри досі називають арабськими. В Російській імперії та на землях України, що входили до її складу, арабські цифри було введено з початку 18 століття.

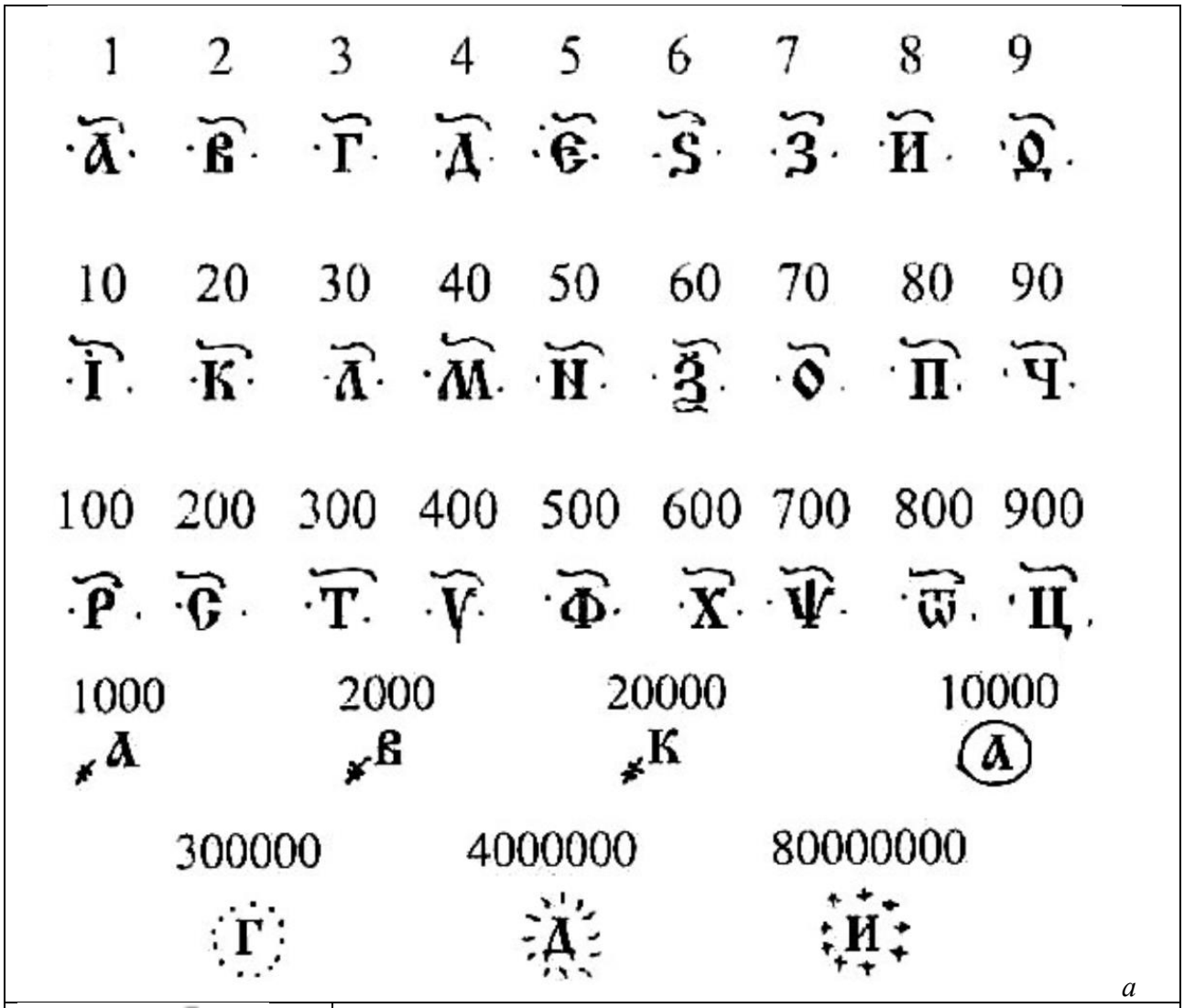


Рис. 12.2.1. Позначення цифр кириличними літерами та спеціальними значками (а) і позначення календарного року кириличними літерами (б)

Рис. 12.2.2. До походження дванадцяткової системи числення

Існували та продовжують існувати й інші позиційні системи.

За основу дванадцяткової системи числення взято кількість фаланг на чотирьох пальцях руки (рис. 12.2.2), тобто вона належить до числа найдавніших. Вважається, що вона виникла в давньому Шумері. В давніх Вавилоні, Ассирії та Єгипті вона використовувалася для поділу дня й ночі на 12 частин (прообраз теперішніх годин). Про цю систему досі нагадує слово «дюжина». На неї пропонувалося перейти під час Великої французької революції. Дванадцяткова система числення досі використовується деякими народностями Тибету й Нігерії.

У давньому Вавілоні використовувалася шістдесяткова система числення (її походження та історія використання залишаються нез'ясованими). Пам'ять по неї зберігають 60 секунд у хвилині та 60 хвилин у годині.

Мова програмування Асемблер використовує шістнадцяткову систему числення: після цифр від 0 до 9 йдуть латинські літери від A до F.

12.2.2. Двійкова система числення

Напевне, найпростішою є двійкова система. Її прообраз виник у давньому Китаї приблизно в XI ст. н.е. В нові часи (XVII ст.) цією системою займався Ф. Бекон, а в сучасній формі її запропонував у 1703 році Г. Лейбніц.

У двійковій системі числа записуються як ступені двійки, а цифр усього дві – 0 та 1. Наприклад, двійкове число 100101 – це $1\times2^0+0\times2^1+1\times2^2+0\times2^3+0\times2^4+1\times2^5= =1+4+32=37$.

Як уже зазначалося, двійкове число зручно подати у вигляді набору значень напруги (імпульсної послідовності): одиниці відповідає ненульова амплітуда імпульсу, нулю – нульова (рис. 12.2.3). Саме тому двійкова система використовується в комп'ютерах для запису чисел та операцій з ними.

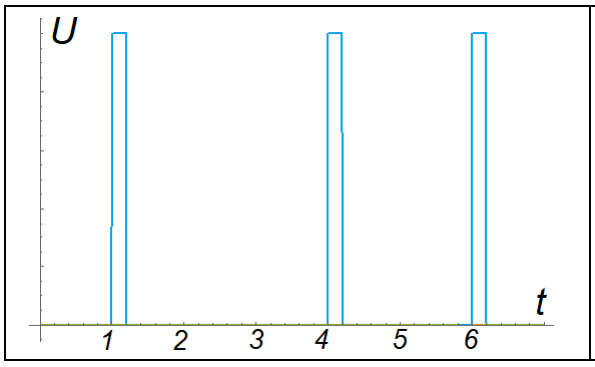


Рис. 12.2.3. Представлення числа 37 у двійковій системі числення як послідовності імпульсів напруги при послідовній передачі сигналу

Максимальне число, яке може бути записане n розрядами (тобто n цифрами) в двійковій системі складає 2^n (від 0 до 2^n-1), у десятковій – 10^n (від 0 до 10^n-1).
В цифровій електроніці існують спеціальні пристрої, що перетворюють аналоговий сигнал на двійковий код. Вони називаються аналого-цифровими перетворювачами (АЦП). Зворотну задачу – перетворення послідовності двійкових чисел на аналоговий сигнал – вирішують цифро-аналогові перетворювачі (ЦАП).

Передача двійкових чисел може бути послідовною (по одному провіднику послідовно передається послідовність імпульсів) або паралельною (число провідників дорівнює числу розрядів, по кожному з них передається двійкова цифра відповідного розряду).

12.2.3. Кількісна міра обсягу цифрових даних

Як уже вказувалося, найбільш універсальне визначення інформації – це відомості, які є об'єктом зберігання, передавання й перетворення. Дані – це форма представлення інформації.

Перехід до цифрового представлення даних (зокрема, числових) дозволяє кількісно

оцінити їхній обсяг. Зазвичай у цифровій електроніці числові дані подають у двійковій системі числення. Найменшою одиницею цифрової інформації є біт.

Біт (англ. bit, шматочок; альтернативне пояснення – скорочення від англ. binary digit, двійковий знак) – мінімальна одиниця кількості інформації, яка дорівнює одному двійковому розряду, який може дорівнювати одному з двох значень/станів (0 або 1) у двійковій системі числення. При записі двійкового числа, яке складається з більш ніж одного символу, число бітів даних – це кількість двійкових розрядів, необхідних для запису цього числа.

Масив даних зазвичай складається з багатьох чисел. Тоді кількість двійкових розрядів, необхідних для запису всіх цих чисел, називають обсягом цифрових даних (за Гартлі).

Щоб спростити запис великих обсягів даних в бітах, використовуються десяткові та двійкові префікси (табл. 12.2.1).

Припустимо, що в пам’яті комп’ютера кожен розряд двійкового числа зберігається одним тригером. Нехай загальна кількість тригерів – *N*. Тоді обсяг пам’яті складає *N* бітів.

Символьним позначенням біта є мала літера *b* латинського алфавіту.

Щоб полегшити людині роботу з цифровою інформацією, двійковий код перекодовується у зрозуміліші числа, букви тощо.

Як уже говорилося, 1 біт – це запис одного розряду двійкового числа. Для запису самого цього числа потрібно кілька розрядів, залежно від величини цього числа. Історично кількість бітів, яка використовувалася для кодування одного текстового символу в комп'ютері, назвали байтом. У багатьох випадках байт є найменшою адресною одиницею пам'яті комп'ютера. Сучасний стандарт байта розміром вісім бітів затверджений у стандарті ISO/IEC 2382-1:1993, він є зручним степенем двійки, який дає змогу одному в одному байті записати (в двійковій системі) числа від 0 до 255 ($2^8 = 256$ чисел, включно з нулем). Міжнародний стандарт IEC 80000-13 кодифікував це загальне визначення.

Таким чином, байт (англ. byte) – це одиниця вимірювання обсягу цифрової інформації, яка зазвичай містить вісім бітів і представлена двійковим числом. Похідні одиниці від байта утворюються так само, як похідні одиниці від біта (табл. 12.2.1).

Таблиця 12.2.1

Кількість біт

Десяткова система		Двійкові префікси	
Префікс	Степінь	Назва	Степінь
біт	10^0	біт	2^0=1
кілобіт	10^3	кібібіт	2^{10}=1 024
мегабіт	10^6	мебібіт	2^{20}=1 048 576
гігабіт	10^9	гібібіт	2^{30}=1 073 741 824
терабіт	10^{12}	тебібіт	2^{40}=1 099 511 627 776
петабіт	10^{15}	пебібіт	2^{50}=1 125 899 906 842 624
ексабіт	10^{18}	ексбібіт	2^{60}=1 152 921 504 606 846 976
зетабіт	10^{21}	зебібіт	2^{70}=1 180 591 620 717 411 303 424
йотабіт	10^{24}	йобібіт	2^{80}=1 208 925 819 614 629 174 706 176

Символьним позначенням байта є велика літера *B* латинського алфавіту. Одиниця октет (символьне позначення *o*) прямо вказує на послідовність з восьми бітів, усуваючи, таким чином, первісну неоднозначність стосовно розміру байта.

Машинне слово – це величина, що дорівнює розрядності регістрів процесора та/або розрядності шини даних (зазвичай це певний степінь двійки). Вимірюється в бітах або байтах.

У ранніх комп'ютерах розмір слова збігався з мінімальним розміром адресованої інформації (розрядністю даних, розташованих за однією адресою). У сучасних комп'ютерах

мінімальним блоком інформації, що адресуються, зазвичай є байт, а слово складається з кількох байтів.

Сучасні архітектури комп'ютерів зазвичай використовують 32-бітні чи 64-бітні слова, які складаються відповідно з чотирьох чи восьми байтів.

Розглянемо простий приклад визначення кількості інформації.

На сторінці тексту з розміром шрифту 12pt вміщується 50 рядків, кожен з яких має приблизно 85 символів (літер), сумарно – 4250 символів. Українська абетка містить 35 літер, або 70 символів разом з великими літерами. Додавши сюди латинську абетку (ще 52 символи), цифри (10), поширені символи та розділові знаки (17), отримаємо всього 149 символів. Для їхнього позначення знадобиться 8 двійкових розрядів, яким відповідає 256 чисел, або якраз 1 байт. Отже, для цифрового запису кожного символу, з яких складається текст, потрібен 1 байт. Таким чином, сторінка тексту з наведеними вище характеристиками містить 4250 байт, або 4.25 кБ, або 4,15 КіБ (враховано, що 1 кібібайт = 1.024 кілобайт, 4.25/1.024≈4.15). В бітах маємо 4250×8=34000 біт, або 34 кілобіт, або 33.2 кібібіт.

Кількість інформації, записана на сучасній флешці (USB-флешнакопи́чувачі, англ. USB Flash drive, скорочено UFD), складає сотні гігабайт – одиниці терабайт.

Контрольні питання до підрозділу 12.2

1. У чому недолік непозиційних систем числення?
2. Що таке позиційні системи числення?
3. У чому зручність двійкової системи числення?
4. Як кількісно визначається об'єм інформації?
5. Як співвідносяться 1 біт та 1 байт?
6. Скільки бітів міститься в одному кібібайті?
7. Скільки розрядів потрібно для запису числа в десятковій та двійковій системах числення? Відповіді пояснити.

Задачі до підрозділу 12.2

12.2.1. Запишіть поточний рік (нашої ери) за допомогою римських цифр, кириличних літер та в двійковій системі числення.

12.2.2. Запишіть рік свого народження в дванадцятковій та шістнадцятковій системах числення. Числа від 10 позначте літерами латинської абетки, починаючи з A.

12.2.3. Яка мінімальна кількість двійкових розрядів потрібна для запису україномовного тексту, що складається з малих літер, великих літер і розділових знаків? Скільки інформації (в бітах) при такому підході міститься в першій строфі «Заповіту» Т. Шевченка?

12.3. Канали зв'язку та їхня пропускна спроможність

На початку цього підручника йшлося про те, що корисний сигнал можна передати через ефір або лінію зв'язку, скориставшись амплітудною чи фазовою модуляцією частоти-носія.

В цифровій електроніці сигнал передається у формі деякої імпульсної послідовності. Ми почнемо цей розділ з аналізу цифрової системи передавання інформації, яка реалізує дану ідею.

Для підвищення швидкості передавання інформації використовують ущільнення каналів зв'язку, коли через той самий канал одночасно передається кілька корисних сигналів. Для одночасного передавання кількох аналогових сигналів використовують частотне ущільнення, для передавання кількох дискретних сигналів – часове ущільнення.

Наприкінці ми розглянемо питання про пропускну спроможність каналу зв'язку.

12.3.1. Системи передавання інформації

Більша частина цього курсу була присвячена розгляду каналу зв'язку, блок-схему якого подано на рис. 12.3.1. Через такий канал передаються неперервні (аналогові) сигнали. При цьому інформація передається за допомогою модуляції високочастотного сигналу корисними сигналами значно нижчої частоти.

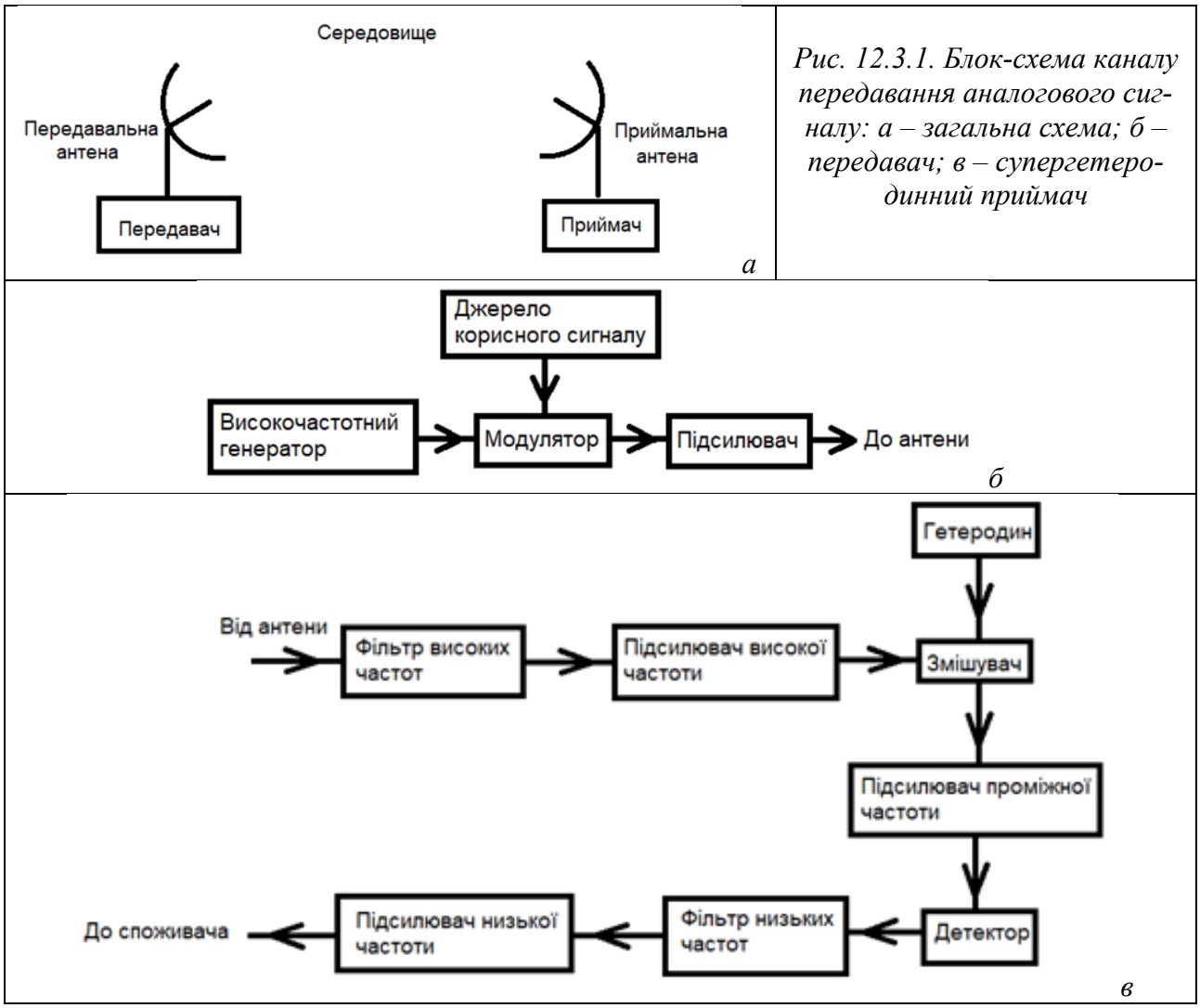


Рис. 12.3.1. Блок-схема каналу передавання аналогового сигналу: а – загальна схема; б – передавач; в – супергетеродинний приймач

В цифровій електроніці, як ми вже зазначали, використовуються дискретні сигнали. Блок-схема системи передавання інформації за допомогою цифрових сигналів подана на рис. 12.3.2.

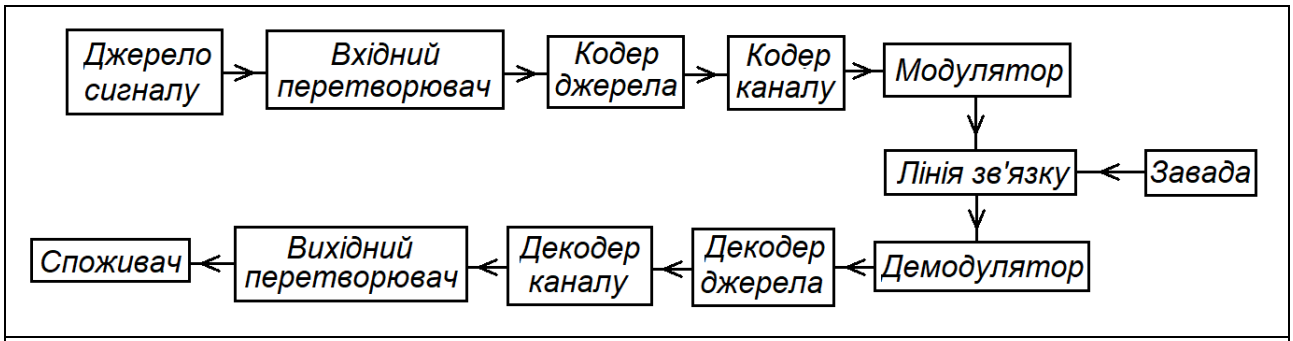


Рис. 12.3.2. Блок-схема цифрової системи передавання інформації

Вхідний сигнал, що надходить від джерела, є аналоговим. Вхідний перетворювач забезпечує його дискретизацію та квантування.

Кодер джерела перетворює дискретний квантований сигнал на двійковий код. Кодер каналу додатково перетворює двійковий сигнал, закладаючи в нього деяку надлишкову інформацію, яка дозволяє на виході системи позбутися впливу завад, які потрапляють до каналу зв'язку. Найпростішою формою такого надлишковою кодування початкової двійкової послідовності є просте повторення кожного двійкового символу декілька разів.

Модулятор перетворює вихідний сигнал кодера каналу на сигнал, який може поширюватися в лінії зв'язку. В найпростішому випадку він перетворює двійковий символ 0 на сигнал $s_0(t)$, а двійковий символ 1 – на сигнал $s_1(t)$ (так звана двійкова модуляція). При цьому кожен біт даних на виході кодера передається окремо. Складніші лінії зв'язку можуть передавати більшу кількість типів сигналів $s_i(t)$, $i = 0,1,2,...n-1$, що дозволяє одночасно передавати кілька бітів (n-позиційна модуляція).

На виході лінії зв'язку демодулятор знову перетворює прийнятий сигнал на двійковий код. Декодери каналу та джерела перетворюють сигнал демодулятора на дискретний квантований сигнал. З виходу вихідного перетворювача знімається аналоговий сигнал, який і надходить до споживача.

12.3.2. Ущільнення каналів зв'язку

Схеми аналогового та цифрового каналів зв'язку, показані на рис. 12.3.1-12.3.2, дозволяють одночасно передавати лише один корисний сигнал.

Якщо йдеться про низькочастотний аналоговий сигнал з максимальною частотою $\omega_{\max}$, то ширина його спектру (що лежить по обидва боки від частоти-носія) буде $\Delta\omega = 2\omega_{\max}$. Для телефонного зв'язку це складає 6.8 кГц, для високоякісного відтворення музики – приблизно 30-35 кГц.

При дискретизації того самого сигналу мінімальна частота повторення імпульсів, відповідно до теореми відліків, також складає $\Omega_{\min} = 2\omega_{\max}$. Відповідно тривалість окремих імпульсів не може перевищувати $\tau_{\max} = 2\pi/\Omega_{\min} = \pi/\omega_{\max}$, а ширина спектру таких імпульсів – $2\pi/\tau_{\max} = 2\omega_{\max}$.

Як бачимо, мінімальна ширина спектру, яку займає модульований сигнал або відповідна йому дискретна імпульсна послідовність, одна й та сама.

В багатьох випадках ширина $\Delta\Omega$ смуги пропускання лінії передачі може виявитися значно більшою за ширину спектру аналогового чи дискретизованого сигналу (особливо це стосується оптоволоконних ліній зв'язку). Тоді через неї можна передавати одночасно багато корисних сигналів. Такий прийом називають ущільненням каналів зв'язку.

Розрізняють частотне та часове ущільнення каналів. При часовому ущільненні частота-

носій модулюється сумою багатьох корисних сигналів, спектри яких не перекриваються. При часовому ущільненні через лінію зв'язку одночасно передається кілька послідовностей коротких імпульсів, які зсунуті одна щодо одної за фазою і через те не перекриваються.

12.3.3. Частотне ущільнення каналів

Блок-схема пристрою, який формує ущільнений за частотою сигнал, поданий на рис. 12.3.3, а спектри сигналів у різних точках цієї схеми – на рис. 12.3.4.

Вважатимемо, що всі низькочастотні (аналогові) сигнали, які ми хочемо передати, мають приблизно однакові спектри, вигляд яких поданий на рис. 12.3.4 а. Кожним таким сигналом модулюється своя проміжна частота $k\omega_1, k = 0,1,2,3,...$ В результаті спектр відповідного низькочастотного сигналу переноситься по осі частот на віддаль $\pm k\omega_1$ (рис. 12.3.4 б) і додатково з'являється лінія на частоті $\pm k\omega_1$, яка потім відфільтровується відповідним загороджувальним фільтром (рис. 12.3.4 в). Потім сигнали з різних каналів додаються (рис. 12.3.4 г), і цим сумарним сигналом модулюється частота-носій Ω. Нарешті, зумовлена нею складова так само відфільтровується загороджувальним фільтром. Спектр вихідного сигналу поданий на рис. 12.3.4 е.

Для максимального ущільнення сигналів слід, очевидно, обирати віддаль за частотою між сусідніми каналами, яка б збігалася з шириною спектру окремого сигналу, тобто $\omega_1 = \Delta\omega$ (рис. 12.3.4 а, г). У смузі пропускання каналу завширшки $\Delta\Omega$ може вміститися $N_{1\max} = \Delta\Omega/2\Delta\omega$ каналів (спектр кожного з них присутній ліворуч та праворуч від частоти-носія. З іншого боку, для низькочастотного сигналу спектр обмежений згори частотою $\omega_{\max} = \Delta\omega/2$ (рис. 12.3.4 а). Таким чином,

$$N_{1\max} = \Delta\Omega/4\omega_{\max}. \qquad (12.3.1)$$

Додатково збільшити число каналів можна, застосувавши так звану односмугову модуляцію. Йдеться про те, що спектр модульованого сигналу (в області додатних частот) складається з центральної лінії, яка відповідає частоті-носію, і симетричних щодо неї бічних частин (рис. 12.3.4 б, д). Очевидно, що вся інформація про корисний сигнал міститься тільки в одній бічній частині. Тому іншу бічну частину можна відфільтрувати, пропустивши її через фільтр низьких частот із граничною частотою Ω. В результаті частотна смуга, зайнята кожним із каналів, зменшиться вдвічі. Відповідно по лінії з фіксованою смугою пропускання можна передати вдвічі більше каналів, так що тепер

$$N_{1\max} = \Delta\Omega/2\omega_{\max}. \qquad (12.3.2)$$

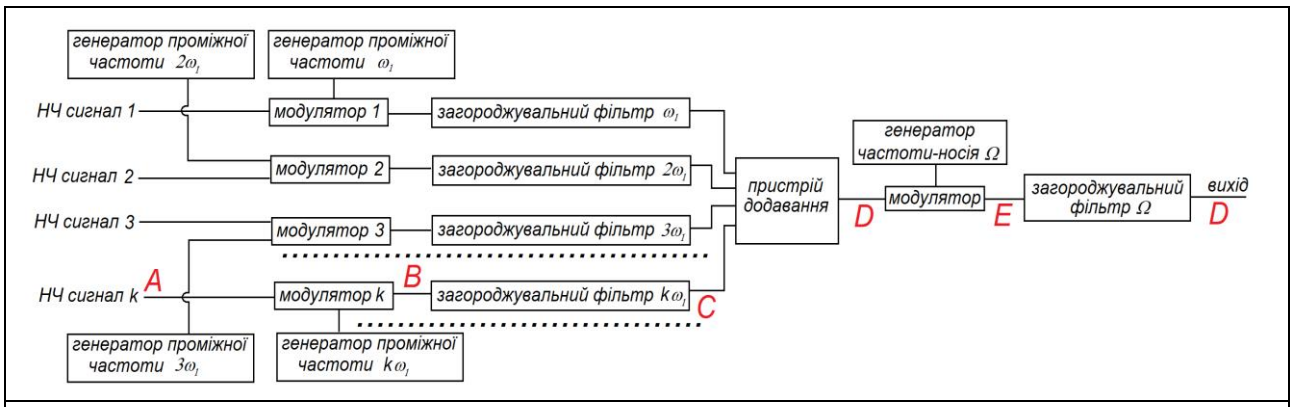


Рис. 12.3.3. Блок-схема формування ущільненого за частотою сигналу

У приймачі, на який надходить сигнал, ущільнений за частотою, спектри сигналів, які відповідають окремим каналам, виділяються за допомогою відповідного набору смугових

фільтрів. Потім сигнал кожного каналу обробляється окремо і передається споживачу.

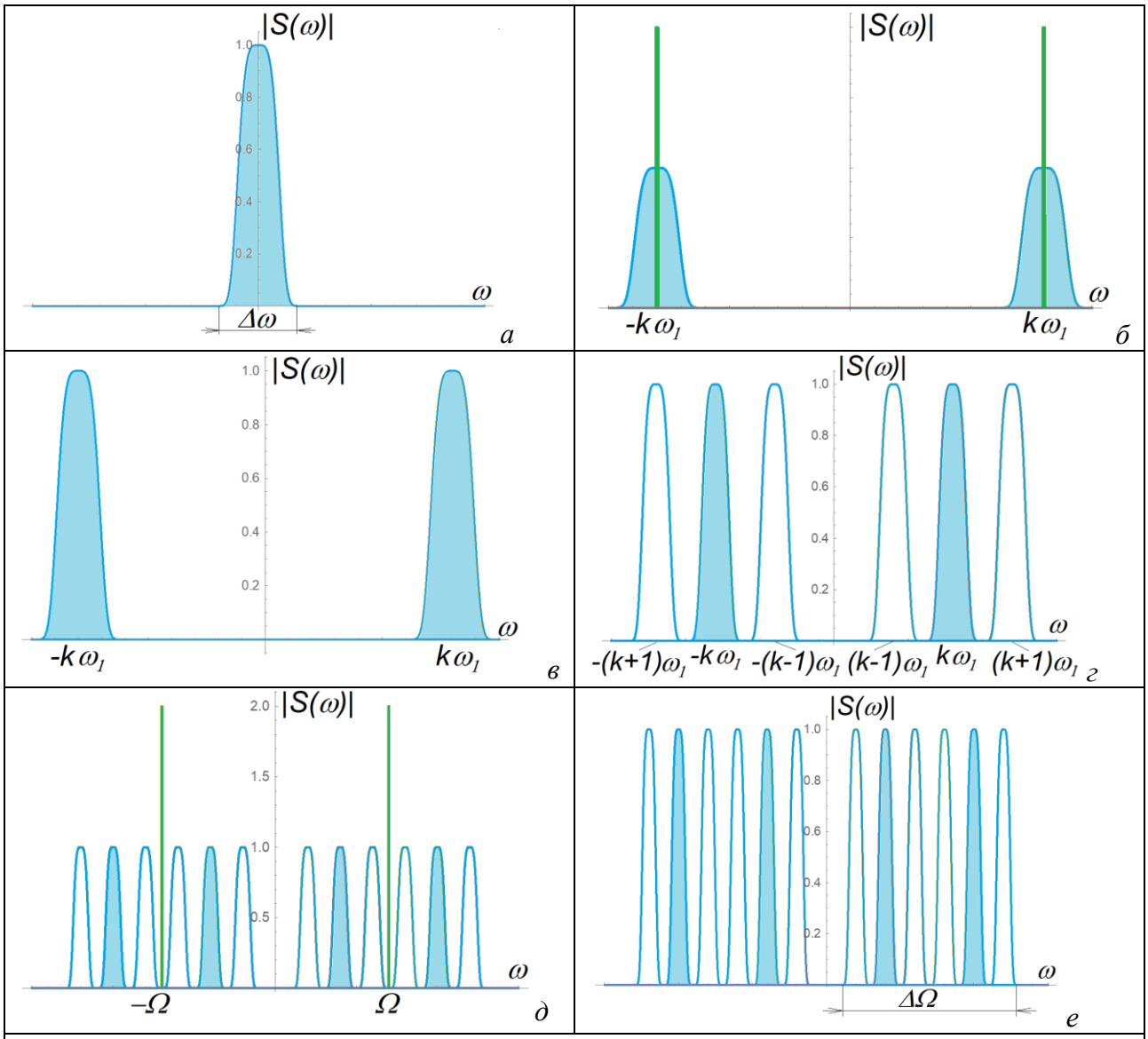


Рис. 12.3.4. Спектри сигналів у точках A (а), B (б), C (в), D (г), E (д), F (е) блок-схеми, поданої на рис. 12.3.3. Кольором виділені частини спектру, що відносяться до того самого каналу

12.3.4. Часове ущільнення каналів

Тепер обговоримо ущільнення каналів зв'язку при передаванні даних у формі імпульсної послідовності. В найпростішому випадку це може бути просто дискретизований сигнал. В загальному випадку цифрової системи передачі інформації (рис. 12.3.2) це буде сигнал з *n*-позиційною модуляцією з виходу модулятора.

Якщо смуга пропускання лінії зв'язку достатньо широка, через неї можна передавати одночасно декілька імпульсних послідовностей з амплітудною модуляцією. Для цього імпульси різних послідовностей потрібно зсунути в часі, як це показано на рис. 12.3.5.

Блок-схема для формування ущільненого в часі сигналу показана на рис. 12.3.6. Всі канали живляться від спільного генератора строб-імпульсів, тому відповідні дискретизовані за часом імпульсні послідовності виходять синфазними. Далі ці послідовності зсуваються в часі за допомогою ліній затримки (кожна на свій час) і додаються за допомогою відповідного пристрою. В результаті на виході формується сигнал, подібний до показаного на рис. 12.3.5. Максимальне число імпульсних послідовностей, які можна одночасно передавати за такою

схемою, буде, очевидно, $N_2 = T/\tau$, де τ – тривалість окремого імпульсу, T – період повторення.

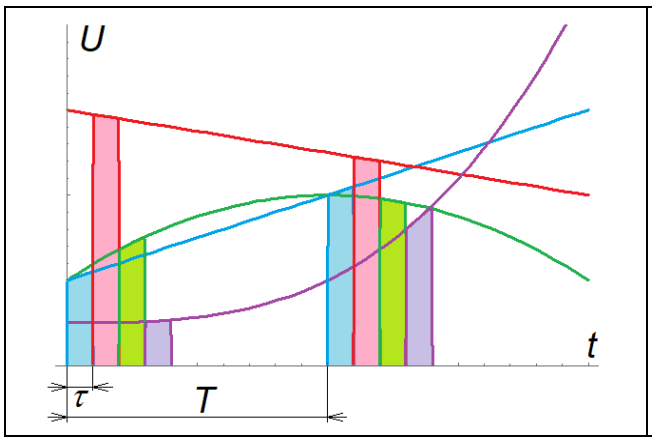


Рис. 12.3.5. Часове ущільнення каналу зв'язку (для дискретизованих сигналів)

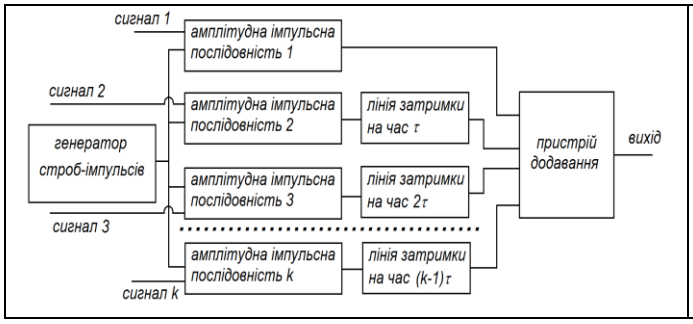


Рис. 12.3.6. Блок-схема формування ущільненого в часі сигналу

Здавалося б, величину N_1 можна необмежено збільшувати, зменшуючи τ. Насправді ж, як уже вказувалося, мінімальне значення τ визначається смугою пропускання лінії передачі Δf, $\tau_{min} \approx 1/\Delta f$. З іншого боку, з теореми відліків випливає, що $T_{\max} = 1/2 f_{\max}$. Таким чином,

$$N_2 = \Delta\omega / 2\omega_{\max} = \Delta f / 2 f_{\max} . \qquad (12.3.3)$$

Отже, чим більша смуга пропускання лінії передачі і чим вужчий спектр аналогових сигналів, що передаються, тим більше відповідних імпульсних послідовностей можна одночасно передати через неї.

Таким чином, при частотному ущільненні сигналу число каналів, які можна передати через задану лінію зв'язку, виходить таким самим, як і при часовому ущільненні.

12.3.5. Характеристики джерел інформації та каналів зв'язку

У схемах передавання даних джерелом інформації[1] виступає набір джерел аналогових сигналів. Продуктивність джерела інформації – це кількість інформації, що виробляється джерелом за одиницю часу.

Швидкість передачі даних – це кількість бітів , які передаються або обробляються за одиницю часу. Одиницею вимірювання цієї величини є бод: 1 бод = 1 біт/с.

Пропускна спроможність каналу зв'язку – це найвища швидкість передавання інформації, яка може бути досягнута за умови, що ймовірність похибки на виході буде як завгодно малою.

Всі названі вище величини – продуктивність джерела інформації, швидкість передачі даних і пропускна спроможність каналу – вимірюються в бітах на секунду, рідше – в байтах

[1] Строго кажучи, тут слід говорити про джерела даних, однак ми користуватимемося усталеною в літературі термінологією.

на секунду. У більшості обчислювальних і цифрових середовищ 1 байт/с відповідає 8 біт/с. Забезпечення необхідної пропускної спроможності каналу є однією з головних вимог при розробленні сучасних систем дротового та бездротового зв'язку. Розробка нових механізмів кодування з виправленням помилок привела до досягнення швидкості передачі інформації, близької до межі, що визначається пропускною спроможністю каналу.

За відсутності шумів пропускна спроможність каналу залежить тільки від ширини його смуги пропускання $\Delta\Omega$. Зокрема, як показав Г. Найквіст, число незалежних імпульсів, які можна передати за одиницю часу без спотворень, обмежене подвійною шириною смуги пропускання каналу зв'язку (див. вище, кінець п. 12.3.4). Це обмеження виникає через ефект міжсимвольної інтерференції (інтерференція при спотворенні, спричиненому запізненням).

12.3.6. Теорема Шеннона для каналу без завад

Засновник теорії інформації Клод Шеннон теоретично досліджував канал зв'язку, блок-схема якого показана на рис. 12.3.2, і сформулював для нього ряд важливих теорем, які стали математичною базою кодування інформації при передаванні повідомлень.

Обговоримо спершу визначення кількості інформації в рамках статистичної теорії, запропоноване Шенноном. В цьому разі інформація розглядається як реалізація випадкових подій, а кількість інформації залежить від імовірностей реалізації цих подій.

Нехай деяка фізична система характеризується N різними станами $x_1, x_2, x_3, ...x_N$ (це можуть бути, наприклад, рівні напруги квантованого сигналу). Нехай цим станам відповідають ймовірності їхньої реалізації $p_1, p_2, p_3, ...p_N$, причому

$$\sum_{i=1}^{N} p_i = 1, \tag{12.3.4}$$

тобто стани $x_1, x_2, x_3, ...x_N$ утворюють повну групу.

Кількість інформації, що відповідає реалізації даного стану, визначається мірою

$$I(x_i) = \log_a \frac{1}{p_i} = -\log_a p_i. \tag{12.3.5}$$

Тут $a \geq 2$ – ціле число, яке виступає основою системи числення (див. вище п. 12.2.1). Зокрема, при $a = 2$ отримаємо кількість інформації в бітах:

$$I(x_i) = -\log_2 p_i. \tag{12.3.5}$$

Середня кількість інформації по всіх станах системи може бути підрахована за звичайною формулою для середнього за ансамблем через задану густину розподілу ймовірності (див. вище п. 9.1.3):

$$I = -\sum_{i=1}^{N} p_i \log_2 p_i \equiv H(x). \tag{12.3.7}$$

Цю величину називають також ентропією системи $H(x)$.

Ентропія – це міра можливої невизначеності стану системи. Інформація усуває цю невизначеність.

Якщо всі стани системи реалізуються з однаковою ймовірністю, тобто $p_i = 1/N$, то

$$I = H(x) = -N\frac{1}{N}\log_2 \frac{1}{N} = \log_2 N, \tag{12.3.8}$$

що збігається з означенням, даним раніше (кількість інформації за Гартлі, див. вище п. 12.2.3).

Поняття кількості інформації використовується, зокрема, при кодуванні даних, тобто при їхньому перетворенні з одного виду в інший. При цьому виконується закон збереження інформації (в специфічному розумінні, про що йтиметься далі).

Щоб пояснити особливості закону збереження інформації, можна порівняти його з

законом збереження енергії. Останній встановлює точну рівність при перетворенні одного виду енергії в інший. При перетворенні ж даних справедливе співвідношення «не більше» (≤), тобто кількість інформації при перетворенні даних не може зрости. Знак рівності відповідає оптимальному кодуванню. Тому при формулюванні закону збереження інформації доводиться казати, що, можливо, існує таке кодування, за якого кількості інформації до й після кодування дорівнюють одна одній.

Теорема Шеннона для каналу без завад встановлює можливість передачі інформації по каналу зв'язку.

Нехай V – інформаційна продуктивність джерела, а R – пропускна спроможність каналу. Тоді, якщо виконано умову $R > V$, то існує спосіб кодування й декодування повідомлень джерела, який забезпечує як завгодно високу надійність передавання повідомлень. У протилежному випадку, коли $V > R$, такого способу немає.

12.3.7. Вплив завад на канал зв'язку

В реальності в каналі зв'язку завжди присутні завади (рис. 12.3.2), які додаються до сигналу, спотворюють його і приводять до помилок на виході декодера (рис. 12.3.7).

Однак вдале кодування дозволяє подолати цю проблему.

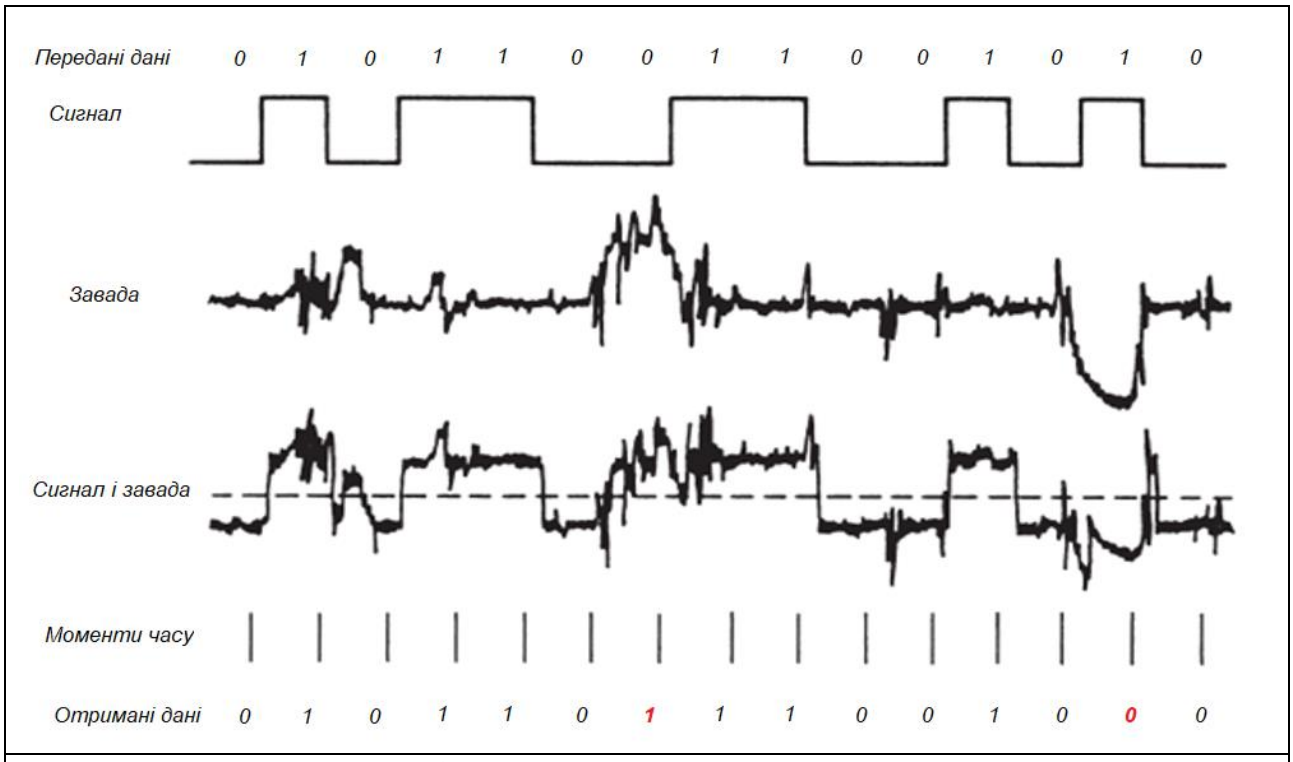


Рис. 12.3.7. Вплив шумів на канал зв'язку.
В останньому рядку червоним показані помилкові біти

Теорема Шеннона для каналу з завадами пов'язує пропускну спроможність каналу передавання даних та існування коду, який забезпечує передавання даних по каналу з помилкою, що прямує до нуля (при збільшенні довжини відповідного інформаційного блоку). Іншими словами, для каналу з завадами існує такий спосіб кодування, при якому забезпечується безпомилкове передавання всіх повідомлень джерела, якщо тільки пропускна спроможність каналу перевищує продуктивність джерела, тобто $R > V$.

12.3.8. Пропускна спроможність каналу зв'язку

Розглядатимемо неперервний часовий аналоговий канал зв'язку, спотворений

адитивним білим гауссівським шумом (шум вважається білим у межах смуги пропускання каналу). Для цієї моделі вдається знайти пропускну спроможність каналу – максимальний обсяг цифрових даних без помилок (тобто інформації), який може бути переданий за одиницю часу. Потужність сигналу вважається обмеженою, потужність гауссівського шуму (в заданій смузі частот) вважається відомою.

Якби існував позбавлений шумів аналоговий канал з нескінченно широкою смугою пропускання, то по ньому можна було б за одиницю часу передавати без помилок необмежений обсяг даних. Однак реальні канали мають обмежену смугу пропускання, і в них завжди присутній шум.

Теорема Шеннона – Гартлі, яка розглядає всі можливі багаторівневі та багатофазні методи кодування, визначає теоретичну верхню межу швидкості передавання даних через аналоговий канал зв'язку з адитивним білим гауссівським шумом, яку можна забезпечити з даною середньою потужністю сигналу. Пропускна спроможність R такого каналу (границя Шеннона) визначається формулою:

$$R = \Delta f \log_2\left(1+\frac{P_s}{P_n}\right), \qquad (12.3.9)$$

де Δf – ширина смуги пропускання каналу, P_s і P_n – відповідно середні потужності сигналу та шуму.

Теорема Шеннона – Гартлі показує, що для збільшення пропускної спроможності необхідно збільшувати ширину смуги пропускання лінії та потужність сигналу і зменшувати потужність шуму. Вона відіграє важливу роль у теорії зв'язку.

Формула Шеннона (12.3.9) дає теоретично можливий максимум пропускної спроможності каналу. На практиці, однак, досягаються значно менші швидкості передавання даних. Однією з причин є неврахування у формулі Шеннона імпульсних завад, амплітудних спотворень або спотворень, спричинених запізненням.

Спрямовуючи Δf до нескінченності, можна отримати максимальну пропускну спроможність каналу з адитивним білим гауссівським шумом. Для цього врахуємо, що $P_n = G\Delta f$, де G – спектральна інтенсивність шуму. Врахуємо також, що при $|x|<<1$

$$\log_2(1+x) = \log_2 e \ln(1+x) \approx x \log_2 e, \qquad e \approx 2.71828... \qquad (12.3.10)$$

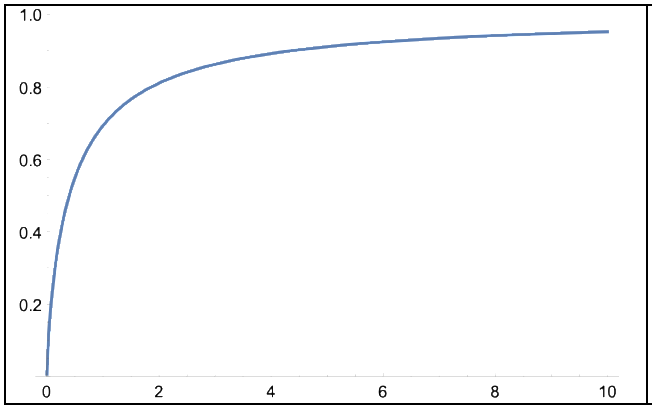


Рис. 12.3.8. Графік залежності нормованої пропускної спроможності каналу R/R_∞ від нормованої ширини його смуги пропускання $G\Delta f/P_s$

Тоді

$$R_\infty = \lim_{\Delta f\to\infty} \Delta f \log_2\left(1+\frac{P_s}{G\Delta f}\right) \to \log_2 e\, \Delta f \frac{P_s}{G\Delta f} = \frac{P_s}{G}\log_2 e, \qquad (12.3.11)$$

або

$$R_\infty\left[біт/c\right] \approx 1.443\frac{P_s}{G}. \qquad (12.3.12)$$

Пропускна спроможність каналу скінченної ширини завжди менша від цієї величини. З рис. 13.2.8 видно, що відхилення від цієї величини стає малим, якщо величина $G\Delta f/P_s$ буде більше або порядку одиниці.

Контрольні питання до підрозділу 12.3

1. Опишіть принцип роботи цифрової системи передавання інформації.
2. Як досягається ущільнення каналів зв'язку за часом?
3. Як досягається ущільнення каналів зв'язку за частотою?
4. Чим визначається максимальна кількість каналів, які можна передати по одній лінії зв'язку?
5. Дайте визначення кількості інформації за Шенноном (в рамках статистичної теорії) і поясніть його зміст.
6. В чому полягають переваги цифрових систем передавання інформації над аналоговими?
7. Дайте характеристику формули Шеннона для пропускної спроможності цифрового каналу зв'язку.

Задачі до підрозділу 12.3

12.3.1. Скільки акустичних каналів з частотами від 0 до 20 кГц можна одночасно передати через лінію зв'язку з шириною смуги пропускання 50 МГц?

12.3.2. Оцінити пропускну спроможність каналу з шириною смуги пропускання 1 ГГц, якщо співвідношення сигнал-шум складає 10^3. Як вона зміниться при необмеженому зростанні ширини смуги пропускання, якщо спектральну інтенсивність шуму вважати незмінною?