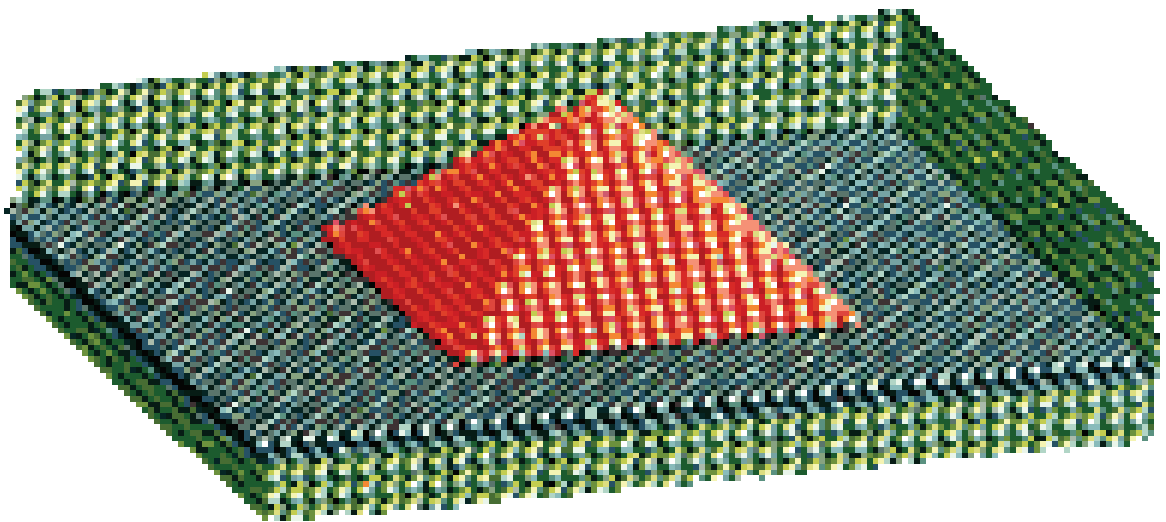


Ю.М. Поплавко, В.І. Ільченко,
С.А. Воронов, Ю.І. Якименко

Фізичне матеріалознавство.

Частина 4.

Напівпровідники



НТУУ «КПІ»

Київ - 2010

Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут»

**Ю.М. Поплавко, В.І. Ільченко,
С.А. Воронов, Ю.І. Якименко**

Фізичне матеріалознавство.

Частина 4.

Напівпровідники

Київ - 2010

УДК 621.38.0296 (075.8)

Поплавко Ю.М., В.І. Ільченко, Воронов С.А., Якименко Ю.І.

Фізичне матеріалознавство. Частина IV. Напівпровідники: Навчальний посібник.

Київ, видавництво «Політехніка» Національного Технічного університету України,

2010. — с.

Табл.Лл. Бібліогр.: назв.

ISBN

У посібнику розглянуті основні властивості напівпровідників. Наведені основи зонної теорії, розглянута статистика і кінетика електронів, викладено основи механізмів рекомбінації і поглинання світла у напівпровідниках, що необхідно для їх застосування у прикладній фізиці, зокрема в електроніці. Значна увага надається сучасним дослідженням нанофізики напівпровідників: аналізується вплив структури на фізичні властивості наноматеріалів і перспективи їхнього застосування в техніці. Обговорюються розмірні ефекти кристалах напівпровідників та приводяться дані по їх впливу на електричні і оптичні властивості матеріалів. Посібник складений на основі багаторічного досвіду авторів по читанню лекцій за курсами «Фізичне матеріалознавство», «Фізика твердого тіла» та «Фізика напівпровідників».

Для студентів вищих навчальних закладів технічних і фізичних спеціальностей у галузях прикладної фізики, електроніки та інформаційних технологій. Може бути використано аспірантами, інженерами і науковими співробітниками, які спеціалізуються у галузях мікроелектроніки та наноелектроніки.

Рецензенти:

УДК 621.38.0296 (075.8)

ББК.....

ISBN.....

Ю.М.Поплавко, Ільченко, Воронов С.А., Якименко Ю.І.

Зміст

Передмова	
Вступ	
1. Класифікація напівпровідникових матеріалів	
1.1. Моноелементні напівпровідники	
1.2. Напівпровідникові хімічні сполуки	
1.3. Сполуки типу $A^{III}B^V$	
1.4. Сполуки типу $A^{II}B^{VI}$, $A^{IV}B^{VI}$ та інші	
1.5. Аморфні напівпровідникові матеріали	
1.6. Органічні напівпровідники	
1.7. Методи виготовлення неорганічних напівпровідників	
1.8. Легування неорганічних напівпровідників	
2. Основи зонної теорії напівпровідників	
2.1. Кристалічні та аморфні структури твердих тіл	
2.2. Оператор повної енергії ідеальної кристалічної структури	
2.3. Адіабатичне наближення	
2.4. Одноелектронне наближення	
2.5. Періодичне поле кристала. Трансляційна симетрія	
2.6. Трансляційна властивість хвильової функції електрона	
2.7. Властивості хвильового вектора. Обернена ґратка	
2.8. Імпульс і квазіімпульс	
2.9. Енергетичний спектр електронів	
2.10. Наповнення дозволених зон	
2.11. Наближення ефективної маси	
2.12. Зонна структура і ефективна маса	
2.13. Ширина забороненої зони	
3. Статистика електронів і дірок у напівпровідниках.....	
3.1. Щільність енергетичних станів	
3.2. Розподіл носіїв заряду у дозволених зонах	
3.3. Динамічні властивості носіїв заряду	
3.4. Дифузійно – дрейфові процеси. Співвідношення Ейнштейна	
4. Кінетичне рівняння Больцмана	
4.1. Загальні положення	
4.2. Нерівноважна функція розподілу	
4.3. Імовірність пружного розсіювання	
4.4. Наближення часу релаксації	
4.5. Класифікація механізмів розсіювання	
4.6. Розсіювання на теплових коливаннях атомів (фононне розсіювання)	
4.7. Розсіювання на іонізованих домішках (іонне розсіювання)	
4.8. Залежність мобільності від температури	
4.9. Електропровідність напівпровідників	

4.10. Термоелектричні ефекти	
4.11. Гальваномагнітні ефекти	
4.12. Тензорезистивний ефект у твердих тілах	
4.13. Електропровідність у сильних полях	
5. Рекомбінація носіїв заряду у напівпровідниках.....	
5.1. Теплова генерація і рекомбінація	
5.2. Нетеплова генерація і рекомбінація	
5.3. Рекомбінація через захват	
5.4. Поверхнева рекомбінація і мобільність	
5.5. Міжзонна рекомбінація	
6. Рівняння неперервності	
6.1. Загальні положення	
6.2. Час релаксації Максвелла	
6.3. Дифузія і дрейф нерівноважних неосновних носіїв заряду	
6.4. Біполярні дифузія і дрейф нерівноважних носіїв заряду	
7. Поглинання світла у напівпровідниках	
7.1. Характеристики поглинання	
7.2. Відбиття світла	
7.3. Механізми поглинання	
7.4. Фоторезистивний ефект	
7.5. Фотовольтаїчні ефекти	
8. Квантово-розмірні ефекти	
8.1. Хвилові властивості та енергетичний спектр електронів.	
8.2. Кристал – тривимірна (3D) структура	
8.3. Квантова яма – двовимірна (2D) наноструктура	
8.4. Квантова нитка – одновимірна (1D) наноструктура	
8.5. Квантова точка – нульвимірна (0D) наноструктура	
9. Основні засади нанотехнології напівпровідників.....	
9.1. Нанолітографія	
9.2. Епітаксiальні методи одержання наноструктур	
9.3. Зондові методи одержання наноструктур	
9.4. Самоорганізація структур у нанотехнологіях	
9.5. Напівпровідникові гетероструктури	
10. Властивості напівпровідникових низькорозмірних структур.....	
10.1. Потенціальні ями, бар'єри і тунелювання	
10.2. Одноелектронне тунелювання і кулонівська блокада	
10.3. Резонансне тунелювання.	
10.4. Балістичний транспорт заряду	
10.5. Квантовий ефект Холла.	
Література	

Передмова

Прикладна фізика охоплює широке коло питань, але в даному навчальному посібнику автори описують, головним чином, електрофізичні властивості напівпровідників. Навчальний курс «Фізичне матеріалознавство» і пов'язаний з ним курси «Фізика твердого тіла» та «Фізика напівпровідників» є одними з основних навчальних курсів для бакалаврів і магістрів. Спеціальність ця відрізняється широким спектром сучасних науково-технічних проблем. У зв'язку з швидким розвитком сучасного матеріалознавства, тематика посібника «Фізичне матеріалознавство» відібрана на основі аналізу тематики бакалаврських та магістерських робіт студентів спеціальностей “Прикладна фізика”, а також «Мікроелектроніка і нанoeлектроніка», захищених у Національному технічному університеті «Київський політехнічний інститут» у 2004 – 2010 роках.

В галузі *фізичного матеріалознавства*, одного з основних навчальних курсів за фахом «Прикладна фізика», практично відсутня навчальна література, видана на Україні. Бракує сучасної літератури і для спеціалістів за фахом «Мікроелектроніка і нанoeлектроніка». Численні монографії і підручники у світовій літературі здебільшого видані англійською мовою і лише почасти – російською мовою. Посібники з фізики напівпровідників, що були видані російською мовою у 1970 – 1990 рр., у більшості своїй призначені для фізичних факультетів університетів. У той же час цей посібник розроблений спеціально для технічних університетів України.

Сучасна англomовна література дорога і практично недоступна; книги і наукові огляди, видані в Росії, тепер також стали рідкістю в Україні, що утрудняє самостійну роботу студентів. Крім того, у даний час не тільки навчальна, але і монографічна література не встигає за бурхливим розвитком технології і фізики напівпровідників. Особливо швидко розвиваються нанофізика і нанотехнології: зрозуміло, що про них немає і згадки у книжках з фізики, виданих 20 – 30 років тому.

В даному посібнику нанофізиці і технологіям приділена значна увага. Поглибленню і розширенню знань про фізичні властивості і перспективи застосування нових матеріалів сприяють сучасні високоточні методи дослідження і розвиток обчислювальних методів із застосуванням могутніх ЕОМ. Широко застосовуються ретельно контрольовані тонкі фізичні технології (*High technique*). У цей науково-технічний напрямок надаються величезні інвестиції, що також сприяє швидкому прогресові.

Даний навчальний посібник є останньою частиною повного курсу «Фізичне матеріалознавство», що поступово видається нами в чотирьох частинах:

Частина I. **«Перспективні напрямки матеріалознавства»**, видана в 2004 р. Розглянуто особливості нано-структурних матеріалів, фулерени, вуглецеві нанотрубки, аморфні метали і сплави, фрактальні структури, рідкі кристали, фізика деяких біологічних структур.

Частина II. **«Діелектрики»**, видана у 2007 р. Викладно основи симетрії кристалів і фазових переходів у них, фізичні основи електричної поляризації і електропереноса заряду, п'єзоелектрики і електрострикційні матеріали, піроелектрики, електрети, сегнетоелектрики, НВЧ діелектрики.

Частина III. **«Метали і магнетики»**, видана у 2010 р. Присвячена фізичним основам металофізики, нанорозмірним ефектам і перспективам їх застосування, металам і сплавам з «важкими» електронами, надпровідниковим матеріалам, магнітним явищам у твердих тілах, сучасним і перспективним магнітним матеріалам.

Частина IV. **Напівпровідники** (даний посібник, видання планується у 2011 р.) Присвячений викладанню фізики й основних властивості напівпровідників: зонній теорії, статистиці і кінетиці електронів, основам механізмів рекомбінації і поглинання світла у напівпровідниках, тощо. Значна увага надається сучасним дослідженням нанофізики напівпровідників:

Таким чином, сьогоденне видання являє собою четверту частину запланованого авторами навчального посібника з фізичного матеріалознавства. Цей посібник має традиційну побудову: питання розглядаються послідовно, «від простого до складного», з висновками, доказами і ін. Останні глави навчального посібника присвячені сучасним науково-технічним проблемам матеріалознавства. Такий підхід до викладу матеріалу обумовлений сучасною тенденцією до синтезу і взаємного проникнення знань, що відносяться, на перший погляд, до різних областей науки.

Вступ

Сучасна прикладна фізика, як правило, базується на електроніці, а електроніка, у свою чергу, побудована в основному на напівпровідниках.

Електроніка – галузь науки і техніки, у якій досліджуються електронні явища в речовинах (головним чином – твердотільних), і на основі результатів цих досліджень розробляються методи створення електронних приладів, електронних схем і систем. Можна також визначити електроніку як науку про взаємодію електронів з електромагнітними полями і про методи створення електронних матеріалів, приладів і пристроїв, у яких ця взаємодія використовується для перетворення електромагнітної енергії, в основному, для передачі, обробки і збереження інформації. Теоретичні задачі електроніки – це дослідження взаємодії електронів як з макроскопічними полями в робочому просторі електронного приладу, так і з мікроскопічними полями усередині атома, молекули або кристалічної ґратки. Практичні задачі електроніки зводяться до розробки електронних приладів і пристроїв, що виконують різні функції в системах перетворення і передачі інформації, у системах керування, в обчислювальній техніці, а також в енергетичних пристроях.

«Фізика напівпровідників» є одним з провідних навчальних курсів для бакалаврів електроніки. В даному навчальному посібнику автори описують, головним чином, електрофізичні властивості напівпровідників, що у першу чергу необхідно для студентів технічних університетів. Спеціалізації в галузі електроніки відрізняються різноманітним набором сучасних науково-технічних проблем, але всі вони завершуються втіленням у реальні прилади, які засновані на фізичних властивостях тих або інших твердих тіл, у тому числі прилади, реалізовані на основі нанотехнологій.

Таким чином, фізика напівпровідників, як найважливіша частина фізики твердого тіла і прикладної фізики, має велике значення для фахівців з електроніки.

Даний навчальний посібник в основному відповідає курсу «Фізики напівпровідників» і написаний згідно програми для фахівців з прикладної фізики і електроніки технічних університетів України. Посібник відрізняється тим, що математичні докази теорій спрощені й найперше виділяються основні поняття про фізичні явища у твердих тілах та їх порівняно прості пояснення.

Навчальний посібник значною мірою задовольняє поставлені завдання: майбутній фахівець з електроніки зможе ознайомитися з природою основних фізичних ефектів у напівпровідниках, а також з параметрами конкретних матеріалів, на яких базується сучасна електроніка та приладобудування. У даному навчальному посібнику використаний науковий внесок авторів у дану область, а також сучасна наукова література, представлена в основному закордонними (англомовними) джерелами.

У своєму розвитку електроніка пройшла три етапи: лампова електроніка, напівпровідникова електроніка і мікроелектроніка (інтегральна електроніка). Сучасний, четвертий етап електроніки це наноелектроніка.

В усіх трьох останніх етапах розвитку електроніки ключову роль відіграє фізика напівпровідників.

Напівпровідникові пристрої були основою вже другого етапу електроніки. Початок дослідженням напівпровідникових структур було покладено давно – німецьким фізиком Брауном, що відкрив у 1874 р. ефект однобічної провідності контакту метал-напівпровідник. Але електричні характеристики подібних структур були нестабільними, і тому у той час ці структури не могли використовуватися для створення електронних приладів. Однак тривалі дослідження, виконані багатьма вченими, дозволили до середини ХХ в. перебороти цей недолік.

Видатним кроком уперед – до сучасної напівпровідникової електроніки – виявився винахід Бардінім і Браттейном напівпровідникового **транзистора** в 1947 р. (Нобелівська премія по фізиці, 1956 р.). Як і тріод, транзистор має 3 електроди і може виконувати підсилювальні функції. Крім того, транзистор може служити електронним перемикачем, а каскад таких перемикачів здатний

виконувати логічні функції. Сучасну конструкцію і теорію транзистора розробив американський фізик У. Шоклі в 1951 р. У перших транзисторах використовувався кристалічний германій, лише пізніше знайшов застосування і дотепер домінує в технології більш термостабільний кремній. Довговічні, малогабаритні, які не потребують вакуумної скляної колби і живляться низькою електричною напругою, транзистори швидко замінили електронні лампи в більшості електронних схем.

Цікаво відзначити, що вперше подібний польовому транзисторові *«прилад, що контролює електричний струм»* був запатентований Лілієнфельдом ще в 1928 р. У ньому пропонувалося в якості «робочого середовища» використовувати напівпровідник Cu_2S , а в якості «підзатворного діелектрика» пропонувався діелектрик Al_2O_3 ; при цьому вперше у минулому застосовані сучасні терміни «витік», «стік» і «затвор». Однак через технологічні труднощі цей винахід не знайшов застосування. Тільки через тридцять років завдяки застосуванню сучасної кремнієвої технології з'явилася можливість реалізувати запропоновану ідею. Тепер такі прилади називають МДН-транзисторами.

Напівпровідникова структура являє собою границю між областями напівпровідника з особливими властивостями. Якщо ця границя розділяє шари напівпровідника з дірковим і електронним типом електропровідності, то вона називається *p-n*-переходом; границя шарів різних сполук називається гетеропереходом. Використовуються також структури, що включають шари металу, діелектрика і напівпровідника. Це так звані МДН або МОН-структури. У МОН-структурі діелектриком звичайно служить окисел напівпровідника, наприклад діоксид кремнію (SiO_2).

Таким чином, лампові діод і тріод були замінені кристалічними діодом і транзистором. Виникла електроніка, заснована на використанні дискретних напівпровідникових приладів. Надійність електронних схем і систем значно збільшилася. Істотно зменшилися їхні габарити, вага й енергоспоживання.

Однак «тиранія міжз'єднань» зберегалася. Собівартість виробництва також залишилася високою. Тому електроніка, заснована на дискретних напівпровідникових приладах, проіснувала недовго.

Основою третього етапу розвитку електроніки і подальшого прогресу електроніки стали **інтегральні схеми**. Хоча винахід транзисторів і став революцією в електроніці, але їхнє величезне число в обчислювальних пристроях (включаючи інші компоненти електроніки – резистори, конденсатори, індуктивності) приводило до великих складностей у реалізації безлічі міжз'єднань.

Рішення проблеми було знайдено Кілбі в 1958 р. на фірмі *Texas Instruments*, США (Нобелівську премію за свій винахід Кілбі одержав у 2000 р.). Транзистори, діоди, резистори, конденсатори й індуктивності – всі основні елементи схем разом з їх міжз'єднаннями були реалізовані в одному монолітному блоці напівпровідника (спочатку використовувався кремній). Процес виготовлення таких *інтегральних схем* допускав автоматизацію їхнього виготовлення і значне зниження габаритних розмірів. Першим застосуванням винаходу Кілбі стали мініатюрні калькулятори.

Таким чином, після 1960 р. почалася ера *інтегральної напівпровідникової електроніки* – **мікроелектроніки** (мікрон дорівнює 10^{-6} м). Почався третій етап розвитку електроніки. Основним елементом складних електронних систем стала інтегральна мікросхема (ІМС).

Інтегральна мікросхема – це мікромініатюрний функціональний вузол електронної апаратури, у якому активні, пасивні і сполучні елементи виготовляються в єдиному технологічному циклі на поверхні або в об'ємі матеріалу і мають загальну оболонку. Виготовлення всіх елементів мікросхеми в єдиному технологічному циклі й в одному матеріалі дозволяє використовувати замість послідовної (індивідуальної для кожного елемента) технології гуртову технологію, що є досить продуктивною. Собівартість

виробництва складних електронних систем у розрахунку на один схемний елемент різко зменшилася.

Можливість переходу на гуртову технологію зв'язана з тим, що напівпровідникова структура має практично всі необхідні схемні властивості. Так, окремий *p-n*-перехід може використовуватися як резистор, конденсатор, діод, перемикач сигналу, стабілізатор напруги, фоточутливий елемент, світлодіод, напівпровідниковий лазер, а в сполученні з іншими *p-n*-переходами – як транзистор, тиристор і т.д. Отже, формуючи на напівпровідниковому кристалі в одному технологічному циклі безліч *p-n*-переходів, можна створювати складну ІМС.

Розвиток твердотільної інтегральної електроніки (мікроелектроніки) йшов по шляху зменшення габаритів, збільшення швидкодії, обсягу пам'яті і надійності електронних систем. Саме на цьому шляху були розроблені методи мініатюризації елементів ІМС, що відкрили реальні можливості переходу до **наноелектроніки**.

Подальший прогрес в одержанні матеріалів з новими властивостями підготував реальні можливості до появи електроніки субмалих елементів (наноелектроніки). Цей прогрес зв'язаний з такими змінами в структурі речовини, що впливають на її фундаментальні властивості. Дотепер ці властивості були визначені, в основному, для макроскопічних об'єктів. Однак уже давно були помічені (і використовуються в техніці) *розмірні ефекти*, що спостерігаються в *тонких плівках*, у властивостях *поверхні* кристалів і у *надмалих* частках речовини.

Таким чином, з'являється можливість «керувати» властивостями речовини – шляхом зміни її фрагментації. У зв'язку з потребами техніки в даний час великий прогрес досягнуто в галузі **нанофізики** – нового наукового напрямку фізичного матеріалознавства, зв'язаного зі створенням і вивченням структури і властивостей матеріалів, що сконденсовані з дуже **малих кристалів, кластерів, фрагментів** (що мають обсяг усього $10^3 - 10^6$ атомів).

У науковій термінології «нано» означає 10^{-9} : один нанометр (нм) дорівнює одній тисячній частині мікрометра («мікрона»), або одній мільйонній частині міліметра, або одній мільярдній частині метра. Надтонка мікроструктура, що має середній розмір фаз або гранул (кластерів), шарів або ниток порядку (чи менше) за 100 нм вважається наноструктурованою. Саме завдяки малим розмірам блоків (часток, гранул, фаз), з яких вони побудовані, наноматеріали демонструють унікальні механічні, оптичні електричні і магнітні властивості.

Полікристалічні дрібнозернисті матеріали із середнім розміром зерна від 40 до 150 нм іноді називають *мікрокристалічними*; а якщо середній розмір зерна менш 40 нм – *нанокристалічними*. Відомо, що й у полікристалічній речовині зменшення розмірів «зерен»-кристалітів може приводити до істотної зміни його властивостей. Такі зміни стають особливо помітними, коли середній розмір кристалічних зерен знижується до 100 нм, але властивості дуже змінюються, коли розмір зерен стає менш 10 нм.

В останні роки розвивається також нова науково-технічна галузь – *магнітоелектроніка*, або, як тепер прийнято її називати, – «спінтроніка», зайнята вивченням і практичною реалізацією пристроїв, що використовують спіни електронів. «Спінтроніка» вивчає магнітні і магнітооптичні взаємодії в металевих і напівпровідникових структурах, а також квантові магнітні ефекти в структурах нанометрового розміру. Особливе значення для магнітоелектроніки мають сучасні нанотехнології, що дозволяють реалізувати на практиці нові досягнення в галузі нанофізики.

Таким чином, нанотехнологія – це сукупність методів виготовлення й обробки структур, що мають довжину 1 – 100 нм (хоча б в одному з вимірів). Загальновизнано, що нанотехнологія – найбільш перспективний напрямок сучасної технології. Вважається, що той, хто раніш опанує нанотехнологіями, займе провідне місце в техносфері XXI сторіччя. Саме тому економічно розвинуті світові держави виділяють мільярди доларів на розвиток нанотехнології. В даний час на світовому ринку вже реалізується понад 3000 товарів, вироблених з використанням нанотехнологій. У найближчі 10 років

світова потреба в дипломованих фахівцях в галузі нанотехнології буде обчислюватися мільйонами; вартість світового ринку нанотехнологічної продукції складе приблизно 1 трильйон доларів США. За останні роки створено понад 20 000 нанотехнологічних компаній, і їхнє кількість подвоюється кожні два роки.

Тому у даному учбовому посібнику приділена увага не тільки фізиці об'ємних кристалів напівпровідників, але й нанофізиці напівпровідникових структур.

1. Класифікація напівпровідникових матеріалів

Фізичні властивості напівпровідників вивчені найбільш повно у порівнянні з металами і діелектриками. Основним стимулом для вивчення напівпровідників є виробництво напівпровідникових приладів і інтегральних мікросхем. Крім того, для фізики твердого тіла важлива і величезна кількість ефектів у напівпровідниках, які не можна спостерігати в інших речовинах. Ці ефекти обумовлені особливостями електронної зонної структури різних напівпровідників, в том числі малою шириною забороненої зони.

Напівпровідники – широкий клас речовин, що по величині питомої провідності займають *проміжне місце* між провідниками і діелектриками: у металах провідність $\sigma > 10^5$ См/м, а для діелектриків $\sigma < 10^{-10}$ См/м (провідність зазначена при кімнатній температурі). Характерною рисою напівпровідників, що відрізняє їх від металів, є зростання провідності σ з ростом температури. Як правило, зростання $\sigma(T)$ експонентне і спостерігається у широкому інтервалі температур: $\sigma = \sigma_0 \exp(-E_g/2k_B T)$, де E_g – енергія активації електронів. З формули видно, що поблизу абсолютного нуля температури ($T \approx 0$) напівпровідники мають властивості ізоляторів: $\sigma \approx 0$.

Причиною температурної залежності σ є те, що з підвищенням температури тепловий рух розриває зв'язок електронів у кристалічній ґратці, так що частина електронів, пропорційна $\exp(-E_g/2k_B T)$, стає *вільними носіями заряду*. Зв'язок електронів у кристалі напівпровідника може бути розірваним не тільки тепловим рухом, але і різними зовнішніми впливами: *світлом, потоком швидких частинок, сильним електричним полем* і т.д. Крім різкої температурної залежності σ напівпровідники відрізняються від провідників також і сильною залежністю провідності від *концентрації домішок* та від *різних видів випромінювання*. Саме тому для напівпровідників характерна висока чутливість електропровідності до зовнішніх впливів, а також до концентрації домішок і дефектів в кристалах, оскільки в багатьох випадках енергія активації для

електронів, локалізованих поблизу домішок або дефектів, істотно менше, ніж в ідеальному кристалі даного напівпровідника. Можливість у широких межах керувати електропровідністю напівпровідників електричним і магнітним полем, зміною температури, введенням домішок і т.д. є основою їх численних і різноманітних застосувань.

Напівпровідникові матеріали (рис. 1.1) поділяються за хімічним складом на неорганічні і органічні, а за структурою – на кристалічні, аморфні і рідинні напівпровідники (останні тут не розглядатися). Найбільше практичне застосування знаходять неорганічні кристалічні напівпровідники, що у свою чергу розділяються на кілька основних груп.

1.1. Моноелементні напівпровідники.

У періодичній системі елементів до напівпровідників відноситься кілька простих речовин. Найбільш важлива частина моноелементних напівпровідників належать до **IV** групи елементів: вуглець С (графіт і алмаз), кремній Si, германій Ge, сіре олово α -Sn. Графіт по електричних властивостях близький до провідників ($E_g < 0,1$ еВ), а чисті природні алмази є діелектриками. Однак *штучні алмази* за рахунок домішок, що у них вводяться, здобувають властивості напівпровідників.

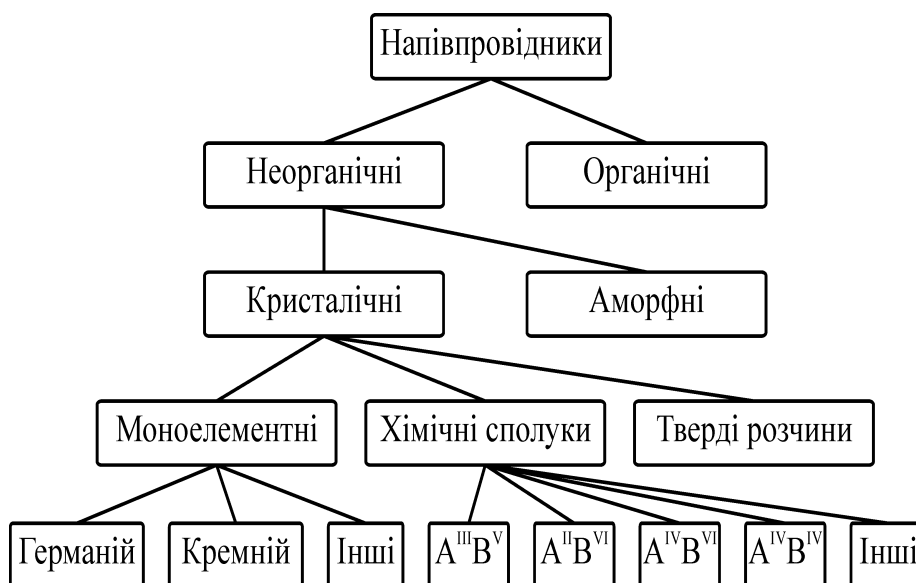


Рис. 1.1 Класифікація напівпровідникових матеріалів по складу і властивостям.

Найважливіші представники цієї групи – Ge і Si – мають кристалічну ґратку типу алмаза. Вони утворюють між собою *безперервний ряд твердих розчинів* Ge–Si, які також мають напівпровідникові властивості.

До моноелементних напівпровідників **III** групи таблиці Менделєєва можна віднести бор B; з **V** групи – фосфор P, арсен As і сурма Sb; з **VI** групи – сірка S, селен Se і телур Te, з **VII** групи – іод J.

1.2. Напівпровідникові хімічні сполуки.

Кристалічна структура багатьох таких сполук характеризується тетраедричною координацією атомів, як це має місце й у ґратці типу алмазу. Тому такі напівпровідникові сполуки одержали назва *алмазоподібних* напівпровідників. Серед них найбільший науковий і практичний інтерес представляють *бінарні* сполуки ($A^{III}B^V$, $A^{II}B^{VI}$), що у даний час є найважливішими матеріалами напівпровідникової оптоелектроніки. Сучасні технології напівпровідників використовують також *трійні* і більш складні хімічні сполуки

Треба також відзначити, що більшість алмазоподібних напівпровідників, що мають близькі властивості, утворюють між собою ізовалентні *тверді розчини*. У твердих розчинах шляхом зміни складу сполуки можна плавно й у досить широких межах керувати найважливішими властивостями напівпровідників, зокрема, шириною забороненої зони і рухливістю носіїв заряду. Це відкриває додаткові можливості для оптимізації параметрів напівпровідникових приладів, дозволяє домогтися кращого узгодження фізичних характеристик різних компонентів електронної апаратури.

Бінарні сполуки. Перш за все, це біля 30 кристалів сполук трьох та п'ятивалентних елементів $A^{III}B^V = A^{+3}B^{+5}$, з яких найбільш важливим є арсенід галію GaAs, а також Ga, InSb, AlAs, Ga та ін. Відомі також (і широко

застосовуються в електроніці) кристали групи $A^{II}B^{VI}$ – Zn, Zn, Cd, 3, Hg, HgTe і т.д., їх також біля 30.

До напівпровідників відносяться також і інші бінарні сполуки A^IB^{VII} (CuCl, AgBr, ...), A^IB^{VI} (Cu₂O, CuS, ...), A^IB^V (KSb, K₃Sb, ...), $A^{II}B^{IV}$ (Mg₂Sn, Ca₂Si, ...), $A^{III}B^{VI}$ (GaS, In₂Fe₃, ...), $A^{IV}B^{VI}$ (PbS, PbSe, ...), $A^{IV}B^{IV}$ (SiC), A^VB^{VI} (Sb₂Te₃, Bi₂S₃, ...)...

Трійні сполуки також можуть мати напівпровідникові властивості. З них найважливішими є куприти (CuAlS₂, CuInS₂, CuSbS₂, CuFeSe₂), сполуки цинкові (ZnSiAs₂, ZnGeAs₂), свинцю (PbBiSe₂), кадмію (CdHgTe) і т.д. Потрійні сполуки типу $A^{II}B^{IV}C^V$ кристалізуються звичайно в кристалічну ґратку халькопірита. Вони виявляють як магнітне, так і електричне упорядкування. Утворюють між собою тверді розчини. Сполуки $A^{II}B^{IV}C^V$ багато в чому є електронними аналогами сполук типу $A^{III}B^V$. Типові представники: CdSnAs₂, CdGeAs₂, ZnSnAs₂.

1.3. Сполуки типу $A^{III}B^V$.

Це найбільш важливі сполуки серед бінарних напівпровідників. Вони є найближчими електронними аналогами кремнію і германія. Кристали $A^{III}B^V$ утворюються в результаті взаємодії елементів III підгрупи періодичної таблиці (бора, алюмінію, галію, індію) з елементами V підгрупи (азотом, фосфором, миш'яком і сурмою). Слід зазначити при цьому, що вісмут і талій не утворюють сполук розглянутого ряду.

Напівпровідники групи $A^{III}B^V$ прийнято класифікувати по *металоїдному елементі*. Відповідно, розрізняють **нітриди, фосфіди, арсеніди й антимоніди**. За винятком нітридів усі сполуки $A^{III}B^V$ кристалізуються в ґратці цинкової обманки кубічного типу сфалериту. Для нітридів характерна структура гексагонального типу вюрциту. У ґратках того й іншого типів кожен атом елемента III групи знаходиться в *тетраедричному оточенні* чотирьох атомів елемента V групи і навпаки.

Структура сфалериту, на відміну від структури алмаза, не має центра симетрії. Ця особливість приводить до розходження у фізичних властивостях поверхонь (111) і ($\bar{1}\bar{1}\bar{1}$), що теоретично цілком складені з однотипних атомів або III, або V групи. Різне поводження граней виявляється при травленні, окислюванні і при вирощуванні кристалів. Напівпровідники зі структурою сфалериту мають *п'єзоелектричні властивості*.

Сполуки $A^{III}B^V$ характеризуються ковалентно-іонним характером хімічного зв'язку (частка іонної складової досягає 15 – 20%). Іноді хімічний зв'язок у цих сполуках трактується як особливий тип хімічного зв'язку, називаного *донорно-акцепторним* зв'язком. Така назва зумовлена тим, що з чотирьох ковалентних зв'язків, якими кожен атом вбудовується в ґратку, три утворюються усупільненням валентних електронів атомів A^{III} і B^V , але четвертий зв'язок здійснюється неподіленою парою валентних електронів атомів B^V . Утворення зв'язку такого типу відповідає енергетично вигідному переходу електронів від атома B^V в енергетичний стан, загальний для донора (атома B^V) і акцептора (атома A^{III}).

Багато напівпровідників типу $A^{III}B^V$ утворюють між собою безперервний ряд твердих розчинів – потрійних і більш складних ($Ga_xAl_{1-x}As$, $GaAs_xP_{1-x}$, $Ga_xIn_{1-x}As_yP_{1-y}$ і т.п.), що також є важливими для сучасної електроніки.

1.4. Сполуки типу $A^{II}B^{VI}$, $A^{IV}B^{VI}$ та інші.

Напівпровідники $A^{II}B^{VI}$ характеризуються кристалічною структурою типу сфалериту або вюрциту (рідше – структурою «типу NaCl»). Зв'язок між атомами у ґратках носить ковалентно-іонний характер (частка іонної складової досягає 45 – 60%). Для сполук типу $A^{II}B^{VI}$ характерні *поліморфізм* і наявність політипів кубічної і гексагональної модифікацій. Напівпровідники зі

структурою вюрциту мають не тільки п'єзоелектричні, але і **піроелектричні властивості**.

Найважливіші представники цієї групи напівпровідників – кристали CdTe, CdS, ZnTe, ZnSe, ZnO та ZnS. Багато сполук типу $A^{II}B^{VI}$ утворюють між собою безперервний ряд твердих розчинів, характерними представниками яких є $Cd_xHg_{1-x}Te$, $Cd_xHg_{1-x}Se$, $CdTe_xSe_{1-x}$. Фізичні властивості сполук типу $A^{II}B^{VI}$ значною мірою визначаються характером власних точкових дефектів структури, що мають низьку енергію іонізації і виявляють високу електричну активність.

Сполуки типу $A^{IV}B^{VI}$ також мають важливе практичне значення. Вони кристалізуються в структуру «типу NaCl» або в орторомбічну структуру і теж характеризуються ковалентно-іонним типом хімічного зв'язку. Найважливіші представники – напівпровідники PbS, PbSe, PbTe і SnTe; серед твердих розчинів найбільш відомі $Pb_xSn_{1-x}Te$, $Pb_xSn_{1-x}Se$. Представляють також технічний інтерес сполуки типу $A_2^{III}B_3^{VI}$, багато з яких мають кристалічну структуру типу сфалериту з 1/3 незаповненими катіонними вузлами (типові представники: Ga_2Se_3 , Ga_2Te_3 , In_2Te_3). Серед сполук елементів VI групи з перехідними металами і рідкоземельними елементами є багато тугоплавких напівпровідників, що мають іонний характер зв'язку і володіють **феромагнітними або антиферомагнітними властивостями**.

Карбід кремнію SiC - єдина хімічна сполука, утворена тільки елементами IV групи. Карбід кремнію має напівпровідникові властивості у всіх структурних модифікаціях: як для β -SiC (структура сфалериту), так і для α -SiC (гексагональна структура), що має близько 15 різновидів. Це один з найбільш тугоплавких і широкозонних кристалів серед широко використовуваних напівпровідникових матеріалів.

1.5. Аморфні напівпровідникові матеріали.

Типовими представниками цієї групи є склоподібні напівпровідники – халькогенідні й оксидні. До *халькогенідних* відносяться сплави Tl, P, As, Sb, Bi з S, Se та Te. Вони характеризуються широким діапазоном значень електричної провідності, низькими температурами розм'якшення, стійкістю до кислот і луг. Типові представники – тверді розчини $\text{As}_2\text{Se}_3\text{-As}_2\text{Te}_3$ і $\text{Tl}_2\text{Se-As}_2\text{Se}_3$.

Оксидні склоподібні напівпровідники, наприклад, сполука типу $\text{V}_2\text{O}_5\text{-P}_2\text{O}_5\text{-RO}_x$ (R-метал I-IV групи) характеризуються електричною провідністю 10^{-4} – $10^{-5} \text{ Ом}^{-1}\text{см}^{-1}$. Усі склоподібні напівпровідники мають електронну провідність, виявляють фотопровідність і термо-електро-рушильну силу (термо-ЕРС). При повільному охолодженні вони звичайно перетворюються у кристалічні напівпровідники.

Іншим важливим класом некристалічних напівпровідникових матеріалів є тверді розчини ряду аморфних напівпровідників із воднем – так називані *гідровані некристалічні напівпровідники*: $\alpha\text{-Si:H}$, $\alpha\text{-Si}_{1-x}\text{C}_x\text{:H}$, $\alpha\text{-Si}_{1-x}\text{Ge}_x\text{:H}$, $\alpha\text{-Si}_{1-x}\text{N}_x\text{:H}$, $\alpha\text{-Si}_{1-x}\text{Sn}_x\text{:H}$. У механізмі електропровідності аморфних неорганічних і напівпровідників виявлено ряд особливостей. У цих напівпровідниках водень має високу розчинність і замикає на собі значну кількість "вільних" зв'язків, характерних для аморфних напівпровідників. У результаті різко знижується щільність енергетичних станів у забороненій зоні і з'являється можливість створення *p-n*-переходів.

1.6. Органічні напівпровідники

Такі напівпровідники являють собою тверді речовини, що мають (або набувають під впливом зовнішніх впливів) електронну або діркову провідність. Органічні напівпровідники характеризуються наявністю в молекулах системи сполучні зв'язки (валентності). Носії заряду в органічних напівпровідниках утворюються в результаті збудження електронів, делокалізованих по системі

сполучених зв'язків. Енергія активації, необхідна для утворення носіїв заряду в органічних напівпровідниках, знижується в міру збільшення числа сполучень у молекулі й у полімері може бути порядку теплової енергії.

Більшість органічних матеріалів є електроізоляторами зі значенням електричної провідності при кімнатній температурі в діапазоні $10^{-9} - 10^{-14}$ См/см. Це обумовлено двома основними причинами. Перша причина полягає в тім, що найвища зайнята молекулярна орбіталь у більшості молекул цілком заповнена електронами, а нижча незайнята молекулярна орбіталь відділена від вищої орбіталі значною енергетичною щільною. Друга причина – тверді стани звичайно являють собою молекулярні структури, що не володіють системою ковалентних зв'язків, які охоплюють макроскопічні стани. Тому квантово-механічні взаємодії між найвищими зайнятими молекулярними орбіталями суміжних молекул є слабкими, і зони валентності, утворені цими взаємодіями, дуже вузькі. Подібно цьому зона провідності, що виникає при взаємодії між орбіталями, теж дуже мала, так що енергетична щільна, власне кажучи, виявляється такою же, як і у вільних молекул.

Щоб одержати підвищену провідність і, отже, властивості напівпровідника, необхідно зменшити заборонену зону між орбіталями. Це може бути досягнуто в органічних сполуках, наприклад, включенням гетероатомів з парою валентних електронів із протилежними спінами (наприклад, поліацетилену, поліаніліну або поліароматиків). Така зменшена зона полегшує електронам переходи між зонами провідності і валентною зоною, що наближає речовину до властивостей напівпровідника.

На основі органічних напівпровідників у даний час реалізовані транзистори і фото-люменесцентні прилади, що вже застосовуються та досить перспективні для приладів, які відображають інформацію (гнучких плоских екранів). Потенційно такі напівпровідники мають широке комерційне застосування. Є багато технічних застосувань органічних напівпровідників, таких, як гнучкі дисплеї, де їхні переваги переважають недоліки. У порівнянні з

традиційними кремнієвими пристроями вони коштують дешевше, багато процесів їхнього виготовлення виконуються при кімнатних температурах, а використовувана технологія стає усе простішою.

Органічні напівпровідники знаходять також застосування як світлочутливі матеріали (наприклад, для процесів запису інформації) у мікроелектроніці та для виготовлення різного роду датчиків. Дослідження органічних напівпровідників важливо для розуміння процесів перетворення і переносу енергії в складних фізико-хімічних і особливо в біологічних тканинах. З органічними напівпровідниками, зокрема з іон-радикальними солями, зв'язана перспектива створення надпровідників з високою температурою переходу (розділ 9.5, сполука TTF-TCNQ).

До органічних напівпровідників відносяться органічні барвники (наприклад, метиленовий голубий, фталоціаніни), ароматичні сполуки (нафталін, антрацен, віолантрен і ін.), полімери зі сполученими зв'язками, деякі природні пігменти (хлорофіл, β -каротін і ін.), молекулярні комплекси з переносом заряду, а також іон-радикальні солі.

Органічні напівпровідники існують у вигляді монокристалів, полікристалічних або аморфних порошків і плівок. Величини питомого опору ρ при кімнатній температурі в органічних напівпровідників лежать у діапазоні від 10^{18} Ом \times см (нафталін, антрацен) до 10^{-2} ом \times см (іон-радикальні солі). Найбільш електропровідними органічними напівпровідниками є іон-радикальні солі, в основному на основі аніон-радикала тетраціанхінодіметана. Вони виявляють електропровідність металевого характеру. В органічних напівпровідниках з низькою електропровідністю спостерігається явище фотопровідності.

Органічні напівпровідники мають особливості, що визначаються молекулярним характером їхньої структури і слабкою міжмолекулярною взаємодією:

- поглинання світла викликає електронне збудження молекул, що може мігрувати по кристалу у виді екситонів;

- утворення носіїв заряду під дією світла зв'язано з розпадом екситонів на поверхні кристалу, біля дефектів його структури, на домішках, при взаємодії екситонів один з одним, а також з автоіонізацією сильно збуджених молекул;
- зони провідності органічних напівпровідників вузькі ($\sim 0,1$ eV) і тому рухливість носіїв заряду в них, як правило, мала (~ 1 см²/В×с);
- поряд із зонним механізмом електропровідності здійснюється стрибкових механізм;
- у кристалах іон-радикальних солей міжмолекулярна взаємодія анізотропна, що приводить до високої анізотропії оптичних і електричних властивостей і дозволяє розглядати цей клас органічних напівпровідників як квазіодновимірні системи (розділ 9.5).

Механізми провідності в органічних напівпровідниках зумовлюють фундаментальні межі рухливості носіїв заряду. Верхня межа мікроскопічної мобільності носіїв в органічних молекулярних кристалах (визначена при температурі 300 К) лежить в інтервалі від 1 до 10 см²/В×с. За цю межу можуть бути відповідальні слабкі сили міжмолекулярної взаємодії (у більшості випадків це сили Ван-дер-Ваальса з енергією розриву, меншою за 10 ккал/моль), оскільки коливальна енергія молекул досягає величини, близької до енергій міжмолекулярних зв'язків при кімнатній температурі або вище.

На противагу цьому в неорганічних напівпровідниках, таких як кремній і германій, атоми утримуються разом дуже сильними ковалентними зв'язками, що у випадку кремнію мають енергію розриву до 76 ккал/моль. У цих напівпровідниках носії заряду рухаються як делокалізовані плоскі хвилі в широкій зоні і мають дуже високу рухливість ($\mu \gg 1$ см²/В×с). Рухливість обмежується завдяки розсіюванню носіїв на коливаннях кристалічної ґратки і, таким чином, *знижується при підвищенні температури*.

У неупорядкованих органічних напівпровідниках зонний транспорт не реалізується, а транспорт носіїв (поляронів) здійснюється за допомогою *стрибків між локалізованими станами*, і носії розсіюються на кожному стрибку

(стрибкова провідність). Стрибки полегшуються взаємодією поляронів з фононами, так що *рухливість зростає з температурою*, залишаючись усе-таки дуже низької ($\mu \ll 1 \text{ см}^2/\text{В}\times\text{с}$).

Границя між зонним транспортом і стрибками визначається матеріалами, що володіють рухливістю між 0,1 і $1 \text{ см}^2/\text{В}\times\text{с}$. Високо-впорядковані органічні напівпровідники, такі як антрацен і пентацен, мають рухливість носіїв заряду в цьому проміжному діапазоні, а в деяких випадках відзначається *незалежність рухливості від температури*, у тому числі й у полікристалічних тонких плівках пентацена. Це спостереження послужило аргументом на користь того, що активований температурою механізм стрибкового транспорту зарядів може бути виключений для високоякісних тонких плівок пентацена.

При низьких температурах (менш як 250 К) у монокристалах пентацена зонний транспорт стає основним механізмом переносу носіїв. У цьому разі зафіксована досить висока рухливість – від $400 \text{ см}^2/\text{В}\times\text{с}$ до більш ніж $1000 \text{ см}^2/\text{В}\times\text{с}$. При низьких температурах енергія коливань набагато менше, ніж енергія міжмолекулярних зв'язків, і фононне розсіювання дуже слабке. З наближенням до кімнатних температур фононне розсіювання зростає, внесок зонного механізму транспорту зменшується, а внесок механізму стрибків – збільшується. Комбінація цих двох механізмів пояснює той факт, що рухливість падає з підвищенням температури від декількох градусів Кельвіна до приблизно 250 К, а після цього починає повільно зростати.

1.7. Методи отримання неорганічних напівпровідників.

Необхідною умовою досягнення високих електрофізичних характеристик напівпровідників є їхнє глибоке очищення від сторонніх домішок. У випадку Ge і Si ця проблема вирішується шляхом синтезу їхніх летучих сполук (хлоридів, гідридів) і наступного глибокого очищення цих сполук із

застосуванням методів ректифікації, сорбції, часткового гідролізу і спеціальних термічних обробок.

Хлориди особливої чистоти піддаються потім високотемпературному відновленню воднем з попереднім глибоким очищенням, з осадженням відновлених продуктів на кремнієвих або германієвих прутках. З очищених гідридів Ge і Si виділяють шляхом термічного розкладання. У результаті одержують Ge і Si із сумарним вмістом залишкових електрично-активних домішок на рівні $10^{-7} - 10^{-9}\%$. Одержання особливо чистих напівпровідникових сполук здійснюють їхнім синтезом з елементів, що пройшли глибоке очищення. Сумарний зміст залишкових домішок у вихідних матеріалах звичайно не перевищує $10^{-4} - 10^{-5}\%$.

Синтез напівпровідникових сполук, що легко розкладаються, проводять або в запаяних кварцових ампулах при контрольованому тиску парів летучого компонента в робочому об'ємі, або під шаром рідкого флюсу (наприклад, особливо чистого збезводненого B_2O_3). Синтез сполук, що мають великий тиск парів летучого компонента над розплавом, здійснюють у камерах високого тиску. Часто процес синтезу сполучають з наступним додатковим очищенням сполук шляхом спрямованої або зонної кристалізації розплав.

Найбільш розповсюджений спосіб одержання монокристалів напівпровідників – витягування з розплаву по методу Чохральського. Цим методом одержують монокристали Ge, Si, сполуки типу $A^{III}B^V$, $A^{II}B^{VI}$, $A^{IV}B^{VI}$ і ін. Витягування монокристалів напівпровідників, що *не розкладаються*, проводять в атмосфері H_2 , інертних газів або в умовах глибокого вакууму. При вирощуванні монокристалів сполук, що *розкладаються*, (InAs, GaAs, In, Ga, CdTe, PbTe і ін.) розплав герметизують шаром рідкого флюсу (B_2O_3) та витягають монокристали, занурюючи затравку у розплав через флюс і підтримуючи в робочому об'ємі над розплавом визначений тиск інертного газу. Часто процес витягування здійснюють у камерах високого тиску, сполучаючи

процес вирощування монокристалів с попереднім синтезом сполук під шаром флюсу (GaAs, InP, GaP і ін.).

Для вирощування монокристалів напівпровідників широко використовують також методи *спрямованої і зонної кристалізації* розплаву у контейнері. У випадку сполук, що розкладаються, для монокристалів необхідного стехіометричного складу процес проводять у запаяних кварцових ампулах, підтримуючи рівноважний тиск пару летучого компонента над розплавом; часто для цього потрібні камери високого тиску, у яких підтримується протитиск інертного газу. При одержанні монокристалів необхідної кристалографічної орієнтації використовують орієнтовані відповідним чином монокристалічні затравки.

Для вирощування монокристалів напівпровідників, що мають підходяще сполучення щільності і поверхневого натягу розплаву, використовують метод *безтигельної зонної плавки*. Найбільше поширення цей метод одержав у технології одержання монокристалів Si, який має порівняно невисоку щільність і досить великий поверхневий натяг розплаву. Відсутність контакту розплаву зі стінками контейнера дозволяє одержувати цим методом найбільш чисті монокристали. Звичайно процес вирощування монокристалу сполучають з попереднім додатковим очищенням напівпровідників методом зонної плавки.

Для одержання монокристалів ряду тугоплавких напівпровідникових сполук, що розкладаються (напр., Cd, Zn, Si, AlN і ін.), використовують кристалізацію з газової фази (методи сублімації і хімічних транспортних реакцій). У випадку якщо при вирощування монокристалу не вдається одержати напівпровідника необхідного стехіометричного сполучення, кристали розрізають на пластини, які піддають додатковому відпалу у парах відсутнього компонента. Найбільше часто цей прийом використовують у технології одержання монокристалів вузькозонних сполук типу $A^{II}B^{VI}$ і $A^{IV}B^{VI}$, де власні точкові дефекти сильно впливають на концентрацію і рухливість носіїв заряду, тобто виявляють високу електричну активність ($PbTe$, $Pb_xSn_{1-x}Te$, $Cd_xHg_{1-x}Te$ і

ін.). При цьому вдається знизити концентрацію носіїв заряду в кристалах на кілька порядків. Для вирощування профільованих монокристалів напівпровідників (стрічки, прутки, трубки і т. д.) використовують спеціальні методи.

Широко поширене одержання напівпровідників у виді монокристалічних плівок на різного роду монокристалічних підкладках. Такі плівки називають *епітаксіальними*, а процеси їхнього одержання – епітаксіальним нарощуванням. Якщо епітаксіальна плівка нарощується на підкладку тієї ж речовини, то одержувані структури називають *гомоепітаксіальними*; при нарощуванні на підкладку з іншого матеріалу – *гетероепітаксіальними*. Можливості одержання тонких і надтонких одношарових і багатшарових структур різноманітної геометрії із широкою варіацією сполук і електрофізичних властивостей по товщині і поверхні нарощуваного шару, з різкими границями *p-n*-переходів і гетеропереходів обумовлюють широке використання методів епітаксіального нарощування в мікроелектроніці й інтегральній оптиці, у практиці створення великих і швидкодіючих інтегральних схем, а також оптоелектронних приладів.

Для одержання епітаксіальних структур напівпровідників використовують методи рідинної, газозазної і молекулярно-пучкової епітаксії. Методом рідинної епітаксії одержують гомо- і гетероепітаксіальні структури на основі сполук типу $A^{III}B^V$, $A^{II}B^{VI}$, $A^{IV}B^{VI}$ і їхніх твердих розчинів. Як розчинник звичайно використовують розплав нелетучого компонента відповідної сполуки. Нарощування епітаксіального шару проводять або в режимі програмувального зниження температури, або з попередньо переохолодженого розплаву. Цим методом можна відтворено одержувати багатшарові структури з товщиною окремих шарів до $\sim 0,1$ мкм при товщині перехідних шарів на гетеро-границях порядку десятків нм.

Найбільш розповсюдженими методами газозазної епітаксії є: хлоридна, хлоридно-гідридна і з застосуванням металоорганічних сполук. При хлоридній

епітаксії як вихідні матеріали використовують летучі хлориди елементів, що входять до складу напівпровідників. Вихідними матеріалами при хлоридно-гідридній епітаксії є летучі хлориди і гідриди відповідних елементів, а при епітаксії із застосуванням летучих металоорганічних сполук використовують також летучі гідриди. Процеси здійснюють у реакторах проточного типу, де транспортним газом є H_2 . Усі вихідні матеріали і H_2 піддають попередньому глибокому очищенню.

Переваги епітаксiального нарощування плівок із застосуванням металоорганічних сполук такі:

- відсутність у газовій фазі хлор-утримуючих компонентів, хімічно взаємодіючих з підкладкою;
- низькі робочі температури;
- простота апаратурного оформлення;
- легкість регулювання товщини і складу епітаксiальних шарів.

Метод забезпечує створення багат шарових структур з тонкими, однорідними по товщині шарами і різкими границями розділу і дозволяє відтворено одержувати шари товщиною менш 10 нм при ширині перехідної області менш 1-5 нм. Його широко використовують для вирощування епітаксiальних структур сполук типу $\text{A}^{\text{III}}\text{B}^{\text{V}}$, $\text{A}^{\text{III}}\text{B}^{\text{VI}}$, $\text{A}^{\text{IV}}\text{B}^{\text{VI}}$ і твердих розчинів на їхній основі. Одержання епітаксiальних структур Si і Ge здійснюється в процесі водневого відновлення відповідних хлоридів або термічним розкладанням гідридів.

Перспективним є метод *молекулярно-пучкової епітаксії*. Процес здійснюють в умовах глибокого вакууму ($10^{10} - 10^{12}$ мм рт.ст.) при використанні молекулярних пучків відповідних елементів. Застосування особливо чистих вихідних речовин, створення багатокамерних установок з охолоджуваними до низьких температур і обернених підкладок дозволяють різко підвищити чистоту вирощуваних шарів і їхня однорідність. Розроблено метод одержання епітаксiальних композицій, що містять кілька летучих

компонентів. Істотно підвищується "гнучкість" процесу застосуванням при нарощуванні шарів і їхньому легуванні іонних пучків, а також летучих сполук як джерела відповідних елементів. Детальне дослідження механізмів кристалізації дозволило покращити умови травлення підкладок з одержанням атомно-дзеркальних і атомно-чистих поверхонь, збільшити швидкості росту шарів при збереженні рекордно низьких температур епітаксialного нарощування. Усе це дозволяє одержувати цим методом багатошарові епітаксialні структури з надтонкими шарами і найменшою товщиною перехідних шарів. Методом молекулярно-пучкової епітаксії вирощують епітаксialні композиції Si, сполуки типу $A^{III}B^V$, $A^{II}B^{VI}$, $A^{IV}B^{VI}$ і тверді розчини на їхній основі.

Застосування низькотемпературних методів епітаксії вирішує не тільки проблему створення гетеропереходів з різкими границями розділу і *p-n*-переходів, але і сприяє удосконалюванню створюваних епітаксialних композицій. Прогрес у розвитку низькотемпературної молекулярно-пучкової епітаксії і епітаксії із застосуванням металоорганічних сполук дозволив одержати високоякісні гетероструктури ряду сполук типу $A^{III}B^V$ на кремнієвих підкладках (GaAs/Si, GaAs-GaAlAs/Si, InP/Si, GaP/Si). Це відкриває нові можливості на шляху інтеграції електронних пристроїв і підвищення їхньої швидкодії.

Одержання однорідних по товщині та із заданим розподілом сполуки й електричних властивостей у тонких і надтонких епітаксialних шарів з різкими *p-n-переходами* і гетерограницями зробило реальним створення багатошарових композицій з надтонкими шарами, які являють собою, по суті, новий тип напівпровідників, тому що їхні властивості визначаються характером розподілу, товщиною і набором сполук у шарах. Товщини окремих шарів можуть бути доведені до величин, менших довжини вільного пробігу електронів, що дозволяє "конструювати" зонну структуру напівпровідника. Власне кажучи, мова йде про новий напрямок у техніці вирощування

монокристалів – «кристало-інженерії», створення періодичних структур з ультра-тонкими прошарками – надграток, фізичні властивості яких визначаються квантово-розмірними ефектами.

1.8. Легування неорганічних напівпровідників.

Для одержання напівпровідників електронного типу провідності (*n*-типу) з концентрацією носіїв заряду (електронів) звичайно використовують **донорні** домішки, що утворюють "мілкі" енергетичні рівні в забороненій зоні поблизу дна зони провідності (енергія іонізації $\geq 0,05$ eV). Для напівпровідника діркового типу провідності (*p*-типу) аналогічна задача вирішується шляхом введенням **акцепторних** домішок, що утворюють "мілкі" енергетичні рівні в забороненій зоні поблизу стелі валентної зони. Такі домішки при кімнатній температурі практично цілком іонізовані, так що концентрація носіїв заряду приблизно дорівнює концентрації домішок. Для Ge і Si основними донорними легуючими домішками є елементи V групи періодичної системи: P, As, Sb, а акцепторними – елементи III групи: B, Al, Ga.

Для сполук типу $A^{III}B^V$ використовуються легуючі домішки елементів VI групи (S, Se, Te), а також Sn, і елементів II групи (Be, Mg, Zn, Cd). Елементи IV групи (Si, Ge) у залежності від умов одержання кристалів і епітаксіальних шарів сполук типу $A^{III}B^V$ можуть виявляти як донорні, так і акцепторні властивості.

У сполуках типу $A^{II}B^{VI}$ і $A^{IV}B^{VI}$ вплив домішок сильно ускладнюється присутністю власних точкових структурних дефектів. Необхідний тип і величина провідності в них звичайно досягаються прецизійним регулюванням відхилення сполуки від стехіометричного, що забезпечує задану концентрацію визначеного типу власних точкових дефектів структури в кристалах

Перераховані вище легуючі домішки утворюють, як правило, **тверді розчини заміщення** і мають досить високу розчинність ($10^{18} - 10^{20}$ атомів/см³) у

широкому інтервалі температур. Максимум розчинності приходить на температурний інтервал 700 – 900°C в Ge, 1200 – 1350°C у Si і 1100 – 1200°C в GaAs. Ці домішки є малоефективними центрами рекомбінації носіїв заряду і порівняно слабо впливають на величину їхнього часу життя.

Домішки важких металів (Fe, Ni, Cr, W, Cu, Ag, Au й ін.) у більшості напівпровідників утворюють енергетично глибокі, часто багатозарядні донорні або акцепторні рівні в забороненій зоні, мають велику ймовірність перетини захоплення носіїв заряду і є ефективними центрами рекомбінації носіїв заряду, приводячи до суттєвого зниження їхнього часу життя. Ці домішки мають малу розчинність у напівпровідниках і мають дуже малі значення коефіцієнти розподілу між кристалом і розплавом. Легування ними роблять у тих випадках, коли треба одержати напівпровідники з малим часом життя носіїв заряду або з високим питомим опором, що досягається компенсацією дрібних енергетичних рівнів протилежної природи. Останнє часто використовують для одержання напівізолюючих кристалів широкозонних напівпровідників типу $A^{III} B^V$ (GaAs, Ga, In). Легуючими домішками служать Cг, Fe, Ni.

Легування напівпровідників звичайно здійснюють безпосередньо в процесах одержання монокристалів і епітаксialьних структур. Домішка вводиться в розплав або у виді елемента, або у виді сплаву з напівпровідником (лігатури). Часто легування здійснюють із газової фази даного елемента або його легколетучих сполук. Це – основний спосіб легування у процесах епітаксії при кристалізації із газової фази. При молекулярно-пучкової епітаксії джерелом легуючої добавки звичайно є сама елементарна домішка. Розрахунок необхідного вмісту легуючої домішки вимагає знання необхідної концентрації та зв'язку між концентрацією домішки і заданими властивостями напівпровідників, а також основних фізико-хімічних характеристик домішки: коефіцієнта розподілу між газовою фазою і кристалом, пружності парів і швидкості випарування у широкому інтервалі температур, розчинності у твердій фазі і т. п.

Одна з головних проблем легування – забезпечення рівномірного розподілу домішки по товщині епітаксiального шару. При спрямованій кристалізації з розплаву розподіл домішки по довжині злитка досягається або шляхом підтримки її постійної концентрації у розплаві за рахунок його підживлення з твердої, рідкої або газової фази, або шляхом програмованої зміни ефективного коефіцієнта розподілу домішки при відповідній зміні параметрів процесу росту.

При зонній перекристалізації звичайно використовують цільове завантаження домішки в початкову розплавлену зону з наступним її проходом через увесь кристал. Ефективний спосіб підвищення об'ємної однорідності монокристалів – вплив на масоперенос у розплаві накладенням магнітного поля. Однорідного розподілу домішки по товщині шару в процесі рідинно-фазній епітаксії досягають кристалізацією при постійній температурі в умовах підживлення розплаву, а при газо-фазній епітаксії – підтримкою постійної концентрації легуючої домішки в газовій фазі над підкладкою протягом усього процесу нарощування.

Легування напівпровідників може бути здійснене також шляхом **радіаційного впливу на кристал:** в результаті ядерних реакцій за участю *власних атомів* речовини утворюються електрично-активні домішки. Найбільший інтерес для радіаційного легування представляє опромінювання тепловими нейтронами, які володіють великою проникаючою здатністю, що забезпечує підвищену однорідність легування. Легування опроміненням тепловими нейтронами забезпечує строго контрольоване введення заданих концентрацій домішки і рівномірний її розподіл в об'ємі кристалу. Однак у процесі опромінення в кристалі утворюються радіаційні дефекти, для видалення яких необхідний наступний високотемпературний відпал. Крім того, може з'явитися наведена радіоактивність, що потребує витримки зразків після опромінення. Легування опроміненням тепловими нейтронами звичайно використовують для одержання однорідно легованих фосфором монокристалів

Si с високим питомим електричним опором. Цей метод перспективний також для GaAs і ряду інших напівпровідників.

При створенні структур з *p-n*-переходами для напівпровідникових приладів широко використовують легування шляхом **дифузійного введення домішки**. Профіль концентрації домішки при дифузії описується звичайно функцією похибок і має вигляд плавної кривої, характер якої визначається наступними факторами: температурою і часом проведення процесу; товщиною шару, з якого здійснюється дифузія; концентрацією і формою перебування домішки в джерелі, а також її електричним зарядом і можливістю взаємодії із супутніми домішками і дефектами у напівпровіднику.

Через малі значення коефіцієнта дифузії основних легуючих домішок дифузійне легування звичайно проводять при високих температурах (для Si, наприклад, при 1100 – 1350°C) і протягом тривалого часу; при цьому легування, як правило, супроводжується генеруванням у кристалі значної кількості структурних дефектів, зокрема дислокацій. При дифузійному легування виникають труднощі в одержанні тонких легованих шарів і досить різких *p-n*-переходів.

Для одержання тонких легованих шарів перспективні процеси іонного легування (**іонної імплантації**), при яких уведення домішкових атомів у приповерхній шар матеріалу здійснюється шляхом бомбардування відповідними іонами с енергією від декількох кіло-еВ до декількох мега-еВ. Можливість уведення практично будь-якої домішки в будь-який напівпровідник, низькі робочі температури процесу, гнучке керування концентрацією і профілем розподілу домішки – переваги іонної імплантації. Легування через діелектричні покриття з одержанням тонких, сильно легованих шарів забезпечили широке поширення цього методу в технології напівпровідникових приладів. Однак у процесі іонного легування генеруються власні точкові дефекти структури, виникають області розупорядкування ґратки, а при великих дозах навіть виникають аморфізовані прошарки. Тому для

одержання якісних легованих шарів необхідний наступний відпал уведених дефектів. Цей відпал проводять при температурах істотно більш низьких, чим при дифузії (не вище 700-800°C). Після відпалу властивості імплантованих шарів близькі до властивостей матеріалу, легованого до тих же концентрацій традиційними методами.

2. Основи зонної теорії напівпровідників

2.1. Кристалічні та аморфні структури твердих тіл

Сукупність атомів, яка утворює структуру з далеким порядком у розташуванні атомів, без дефектів і необмежену у просторі, називають *ідеальною кристалічною решіткою*.

Реальну структуру з далеким порядком, обмежену у просторі і, до певної міри, дефектну називають *монокристалом* або просто *кристалом*.

Сукупність атомів, яка утворює випадкову сітку атомів (довжина та просторова орієнтація валентних зв'язків – випадкові величини) з частково насиченими валентними зв'язками називають *аморфною структурою*.

Умовна межа між кристалічною і аморфною структурами: якщо впорядковане розташування атомів спостерігається на протязі 8 – 10 міжатомних відстаней і більше, така структура має властивості кристалу, якщо менше – це структура з ближнім порядком і властивостями аморфної речовини.

2.2. Оператор повної енергії для ідеальної кристалічної структури

Позначимо: координати ядер $\bar{R}_1, \bar{R}_2 \dots \bar{R}_\alpha \dots \bar{R}_\beta \dots$;

координати електронів $\bar{r}_1, \bar{r}_2 \dots \bar{r}_i \dots \bar{r}_j \dots$.

Оператор повної енергії системи електронів і ядер (гамільтоніан):

$\hat{H} = \hat{T}_z + \hat{T}_e + U_{\alpha\beta} + U_{ij} + U_{\alpha i}$ + енергії магнітних й інших взаємодій.

$\hat{T}_z = \sum_{\alpha} \left(-\frac{\hbar^2}{2m_{\alpha}} \nabla_{\alpha}^2 \right), \hat{T}_e = \sum_i \left(-\frac{\hbar^2}{2m_e} \nabla_i^2 \right)$ – оператори кінетичної енергії, M_{α}, m_e

– маси ядер і електронів, відповідно.

$U_{\alpha\beta} = \frac{1}{2} \sum_{\alpha} \sum_{\beta} \frac{Z_{\alpha} Z_{\beta} e^2}{\epsilon \epsilon_0 |\bar{R}_{\alpha} - \bar{R}_{\beta}|}, U_{ij} = \frac{1}{2} \sum_i \sum_j \frac{e^2}{\epsilon \epsilon_0 |\bar{r}_i - \bar{r}_j|}$ – потенціальні енергії

відштовхування ядер і електронів, відповідно.

$$U_{ai} = -\sum_{\alpha} \sum_i \frac{Z_{\alpha} e^2}{\epsilon \epsilon_0 |\bar{R}_{\alpha} - \bar{r}_i|} \quad \text{— потенціальна енергія притягання електронів і}$$

ядер, Z — номер ядра в періодичній системі, $\epsilon \epsilon_0$ — діелектрична проникність кристалу.

Магнітними взаємодіями (спін – спіновими, спін – орбітальними і т.п.) та іншими взаємодіями на цьому етапі нехтуємо.

Операторне рівняння Шредінгера для стаціонарного стану системи ядер та електронів $\hat{H}\Psi = E\Psi$ має стільки розв’язків, тобто хвильових функцій Ψ і значень енергії E , скільки ядер і електронів міститься у твердому тілі. Атомна щільність твердих тіл $10^{28} - 10^{29} \text{ м}^{-3}$, електронів ще у 5 – 50 разів більше. Точний розв’язок рівняння Шредінгера у таких умовах практично неможливий.

Загальний принцип вирішення проблеми системи великого числа взаємодіючих частинок: необхідно перетворити гамільтоніан системи до такого виду, щоб він був простою сумою гамільтоніанів окремих частинок:

$$\hat{H} = \sum_k \hat{H}_k, \quad \text{тоді } \psi = \prod_k \psi_k, \quad E = \sum_k E_k. \quad \text{Тепер досить розв’язати тільки одне}$$

рівняння $\hat{H}_k \psi_k = E_k \psi_k$, знайти весь спектр його розв’язків (тобто ψ_k і відповідні їм значення енергії E_k), і ми будемо мати повну інформацію про те, що відбуватиметься в цій системі з валентними електронами і атомами. Електрони внутрішніх оболонок зазвичай розглядаються разом з ядрами, оскільки вони безпосередньої участі у формуванні валентних зв’язків у твердих тілах не приймають.

2.3. Адіабатичне наближення

Наближення засноване на тому, що $m_e \ll M$, тобто ядра відносно важкі частинки. Електрони безінерційно відслідковують миттєве положення ядер у

процесі їхнього руху, а ядра реагують на середнє поле, створюване електронами.

Ґрунтуючись на цій сильній нерівності, домовилися вважати ядра нерухомими, їхнє положення фіксоване ($\bar{R}_\alpha = const$). Це дозволяє розділити ядерну й електронну підсистеми і незалежно розглядати процеси у цих підсистемах.

Наслідки „зупинки” ядер такі: $\hat{T}_z = 0$, $U_{\alpha\beta} = const = 0$ (початок відліку енергії), а $U_{\alpha i}$ тепер залежить тільки від положення електронів:

$U_{\alpha i}(r_i) = \sum_i \frac{Z_\alpha e^2}{\epsilon\epsilon_0 |\bar{R}_\alpha - \bar{r}_i|}$. Отже, можемо записати електронний гамільтоніан:

$$\hat{H}_e = \hat{T}_e + U_{ij} + U_{\alpha i}(\bar{r}_i). \quad (2.1)$$

У межах адіабатичного наближення проблема перетворення електронного гамільтоніана у просту суму ще не вирішена, оскільки подвійна сума зберігається у потенціальній енергії взаємодії кожного електрона з кожним U_{ij} . Крім того, у межах адіабатичного наближення втрачаємо інформацію про теплові коливання атомів (фонони) і про взаємодію електронів з тепловими коливаннями атомів (електрон - фононну взаємодію).

Енергетичний спектр теплових коливань атомів (фононний спектр) визначають при розв'язку квазікласичних рівнянь, тому що атоми важкі частинки і квантові ефекти для них виражені дуже слабо. Інформацію про електрон – фононну взаємодію відновлює теорія розсіювання електронів на фононах.

2.4. Одноелектронне наближення

У визначеній точці міжатомного простору \bar{r}_i умовно фіксується i -й електрон. Інші електрони створюють у цій точці деякий опосередкований

потенціал $\Omega_j(\vec{r}_j)$. Потім проводиться аналогічна процедура для іншої точки простору, причому тільки у межах найближчих сусідніх атомів, тому що кристалічна структура має трансляційну симетрію і конфігурація потенціального поля періодично повторюється. Таким чином послідовно обраховується потенціальне поле у міжатомному просторі, створюване всіма електронами окрім i -го.

Замість потенціальної енергії взаємодії кожного електрона з кожним U_{ij} вводиться потенціальна енергія для i -го електрона у полі $U_e = \sum_i \Omega_i(\vec{r}_i)$, яке створили усі інші електрони. Електрони нерозрізненні мікрочастинки, тому i -тим електроном може бути будь-який інший.

Гамільтоніан у адіабатичному і одно електронному наближеннях:

$$\hat{H}_e = \sum_i \left(-\frac{\hbar^2}{2m_e} \nabla_i^2 \right) + \sum_i \frac{Z_\alpha e^2}{\epsilon \epsilon_0 |\vec{R}_\alpha - \vec{r}_i|} + \sum_i \Omega_i(\vec{r}_i) = \sum_i \left[\hat{T}_i + U_i(\vec{r}_i) + \Omega_i(\vec{r}_i) \right] = \sum_i \hat{H}_i,$$

$$\text{де } \hat{H}_i = \hat{T}_i + U_i(\vec{r}_i) + \Omega_i(\vec{r}_i) - \text{гамільтоніан } i\text{-го електрона.} \quad (2.2)$$

Таким чином, основна мета досягнута – гамільтоніан системи електронів – проста сума гамільтоніанів окремих електронів.

Як вказувалось у п.2.2, тепер достатньо розв'язати лише одне рівняння $\hat{H}_i \Psi_i = E_i \Psi_i$, щоб визначити хвильові функції і енергетичний спектр усіх електронів у кристалі.

2.5. Періодичне поле кристала. Трансляційна симетрія

Структурна симетрія твердих тіл має визначальне значення для їх електрофізичних властивостей. *Точкова симетрія* (наявність чи відсутність центра симетрії, порядок вісей симетрії, дзеркальні площини і т.п.) має визначальне значення для ряду властивостей діелектриків. Для металів та напівпровідників найважливішою є *трансляційна симетрія*.

Трансляція – необмежено повторюване паралельне перенесення елементарної комірки у певному напрямку та на певну відстань, яка називається *періодом трансляції*. Довжина та напрямок вектора трансляції залежать від геометричної конфігурації атомної структури кристалу.

Трансляційна симетрія – якщо трансляція проходить на відстань кратну цілому числу періодів, то потрапляємо у *фізично еквівалентну точку простору*.

Положення будь-якого атома у структурі з далеким порядком можна визначити вектором

$$\bar{n} = n_1 \bar{a}_1 + n_2 \bar{a}_2 + n_3 \bar{a}_3, \quad (2.3)$$

де $\bar{a}_1, \bar{a}_2, \bar{a}_3$ найменші трансляційні вектори, n_1, n_2, n_3 – цілі числа.

Потенціальне поле у структурі з далеким порядком періодичне – при зміщенні у міжатомному просторі на вектор \bar{n} воно відтворюється: $U(\bar{r} + \bar{n}) = U(\bar{r})$. Для трансляції будь-якої іншої функції координат необхідно застосовувати *оператор трансляції*:

$$\hat{T}(\bar{n})\Psi(\bar{r}) = \Psi(\bar{r} + \bar{n}). \quad (2.4)$$

Запишемо $\Psi(\bar{r} + \bar{n})$ у вигляді нескінченного незгасаючого ряду по степеням \bar{n} :

$$\Psi(\bar{r} + \bar{n}) = \Psi(\bar{r}) + \frac{d\Psi(\bar{r})}{d\bar{r}} \bar{n} + \frac{1}{2!} \frac{d^2\Psi(\bar{r})}{d\bar{r}^2} \bar{n}^2 + \dots = \sum_{m=0}^{\infty} \frac{\bar{n}^m}{m!} \frac{d^m}{d\bar{r}^m} \Psi(\bar{r}).$$

Введемо позначення: $\frac{d^m}{d\bar{r}^m} = \nabla^m$ та виділимо операторну частину і

запишемо її у явному та експонентному вигляді:

$$\hat{T}(\bar{n}) = \sum_{m=0}^{\infty} \frac{\bar{n}^m}{m!} \frac{d^m}{d\bar{r}^m} = e^{\bar{n}\nabla}. \quad (2.5)$$

2.6. Трансляційна властивість хвильової функції електрона

Складемо рівняння на власні значення оператора трансляції $t(\bar{n})$ для періодичної функції $\varphi(\bar{r}) = \varphi(\bar{r} + \bar{n})$ з періодом \bar{n} :

$$\hat{T}(\bar{n})\varphi(\bar{r}) = t(\bar{n})\varphi(\bar{r}) = \varphi(\bar{r} + \bar{n}). \quad (2.6)$$

З наведеного запису випливає, що власне значення оператора трансляції $t(\bar{n}) = 1$, тобто, будь-яка періодична функція з періодом \bar{n} є власною функцією оператора $\hat{T}(\bar{n})$.

Рівняння на власні значення оператора $\hat{T}(\bar{n})$ для плоскої хвилі $\psi(\bar{r}) = Ae^{i\bar{k}\bar{r}}$:

$$\hat{T}(\bar{n})\psi(\bar{r}) = t(\bar{n})\psi(\bar{r}) = \sum_{m=0}^{\infty} \frac{\bar{n}^m}{m!} \nabla^m Ae^{i\bar{k}\bar{r}}. \quad (2.7)$$

Після виконання необхідних операцій отримаємо $t(\bar{n}) = e^{i\bar{k}\bar{n}}$ і, у кінцевому підсумку, маємо для плоскої хвилі:

$$\hat{T}(\bar{n})\psi(\bar{r}) = \psi(\bar{r} + \bar{n}) = e^{i\bar{k}\bar{n}}\psi(\bar{r}). \quad ((2.8)$$

Трансляційна властивість хвильової функції електрона: при просторовому зміщенні у періодичній структурі на вектор трансляції \bar{n} хвильова функція електрона помножується на фазовий множник $e^{i\bar{k}\bar{n}}$.

Таким чином, власними функціями оператора трансляції можуть бути дві функції – періодична $\varphi(\bar{r}) = \varphi(\bar{r} + \bar{n})$ з власним значенням $t(\bar{n}) = 1$ і плоска хвиля $\psi(\bar{r}) = Ae^{i\bar{k}\bar{r}}$ з власним значенням $e^{i\bar{k}\bar{n}}$. Об'єднаємо ці дві функції, включивши амплітуду A до складу $\varphi(\bar{r})$,:

$$\psi(\bar{r}) = \varphi(\bar{r})e^{i\bar{k}\bar{r}}. \quad (2.9)$$

Отримали модульовану плоску хвилю, яка має назву **хвиля Блоха**.

Одним із важливих наслідків трансляційної властивості хвильової функції є наявність *фізично еквівалентних точок* для електронних станів. Із квантової

механіки відомо, що фізичний зміст має квадрат модуля хвильової функції $|\psi(\vec{r})|^2$ – імовірність перебування електрона у точці з координатами \vec{r} . При зміщенні на вектор трансляції \vec{n} квадрат модуля $\psi(\vec{r} + \vec{n})$ не змінюється, у чому можна легко впевнитися: $|\psi(\vec{r})|^2 = |\psi(\vec{r} + \vec{n})|^2$. Це означає, що в точках \vec{r} і $\vec{r} + \vec{n}$ імовірність знайти електрон однакова.

Трансляційну властивість хвильова функція електрона може мати лише за умови, що **хвильовий вектор** \vec{k} має дійсні значення. У випадку уявних значень $\vec{k} = i\vec{\gamma}$, де $\vec{\gamma}$ - дійсне число, хвильова функція електрона буде характеризувати *локалізований стан електрона* біля точки \vec{r} , а квадрат її модуля – імовірність знайти електрон *лише у цій точці*.

Для спрощення записів об'єднаємо потенціали $U_i(\vec{r}_i) + \Omega_i(\vec{r}_i) = U(\vec{r})$ і запишемо одноелектронний гамільтоніан (2.2), опустивши індекс i ,:

$$\hat{H} = \hat{T} + U(\vec{r}) = -\frac{\hbar^2}{2m_e} \nabla^2 + U(\vec{r}). \quad (2.10)$$

Складемо і обрахуємо комутатор $\hat{H}\hat{T}(\vec{n}) - \hat{T}(\vec{n})\hat{H}$. Із урахуванням того, що оператори кінетичної енергії \hat{T} і трансляції $\hat{T}(\vec{n})$ однорідні за своєю дією, вони комутують, тобто: $\hat{T}\hat{T}(\vec{n}) - \hat{T}(\vec{n})\hat{T} = 0$, а також, враховуючи періодичність потенціальної енергії $U(\vec{r}) = U(\vec{r} + \vec{n})$, можна записати, що, у цілому, гамільтоніан і оператор трансляції комутують: $\hat{H}\hat{T}(\vec{n}) - \hat{T}(\vec{n})\hat{H} = 0$. Комутуючі оператори мають *спільну систему власних функцій*, отже можна стверджувати, що *власною функцією оператора повної енергії електрона у кристалі є модульована плоска хвиля (хвиля Блоха)* (2.9).

Резюме п.2.6

1. Наслідком *трансляційної симетрії* структури з далеким порядком у розташуванні атомів є *наявність фізично еквівалентних точок* у міжатомному просторі.
2. *Трансляційна властивість* хвильової функції електрона в періодичній структурі реалізується лише за умови *дійсних значень хвильового вектора*.
3. Оператори трансляції і гамільтоніан з періодичною потенціальною енергією *комутують* і, отже, мають *спільну систему власних функцій*.
4. *Розв'язки* стаціонарного рівняння Шредінгера для електронів у кристалі слід шукати у вигляді *модульованих плоских хвиль (хвиль Блоха)* (2.9).

2.7. Властивості хвильового вектора.

Обернена ґратка

Загальні положення: імпульс, хвильовий вектор і швидкість електрона пов'язані співвідношенням $\vec{p} = \hbar \vec{k} = m_e \vec{v}$; довжина електронної хвилі і модуль хвильового вектора (хвильове число): $k = 2\pi / \lambda$; розмірність хвильового вектора – обернена довжина.

Щоб встановити взаємозв'язок хвильового вектора системи електронів і хвильових векторів окремих електронів, запишемо хвильову функцію системи в одно електронному наближенні:

$$\Psi(\vec{r}) = \prod_{\alpha} \psi_{\alpha}(\vec{r}). \quad (2.11)$$

Застосуємо оператор трансляції до лівої і правої частини (2.11):

$$\hat{T}(\vec{n})\Psi(\vec{r}) = e^{i\vec{k}\vec{n}} \Psi(\vec{r}) = \prod_{\alpha} e^{i\vec{k}_{\alpha}\vec{n}} \psi_{\alpha}(\vec{r}) = e^{i\sum_{\alpha} \vec{k}_{\alpha}\vec{n}} \prod_{\alpha} \psi_{\alpha}(\vec{r}). \quad (2.12)$$

Враховуючи, що експоненти у (2.12) – гармонічні функції з періодом 2π , умовою їх рівності буде:

$$\bar{k}\bar{n} = \bar{n} \sum_{\alpha} \bar{k}_{\alpha} + 2\pi m, \quad (2.13)$$

де m – ціле число. У кінцевому підсумку, співвідношення хвильових векторів має такий вигляд:

$$\bar{k} = \sum_{\alpha} \bar{k}_{\alpha} + \frac{2\pi}{\bar{n}} m. \quad (2.14)$$

Наслідок: хвильовий вектор системи електронів у полі періодичного потенціалу неоднозначний і визначається з точністю до вектора $\frac{2\pi}{\bar{n}} m$, який називається **вектором оберненої ґратки**.

Зазвичай вектор оберненої ґратки визначають по мінімальному значенню вектора трансляції $\bar{n}_{\min} = \bar{a}$, тобто, $\frac{2\pi}{\bar{a}} m$.

Простір оберненої ґратки називають *простором хвильових векторів* або *імпульсів*.

Розглянемо взаємозв'язок прямої та оберненої ґраток. Пряма ґратка – це фізичний простір, у якому положення кожного атома визначає *вектор прямої ґратки* \bar{a} (2.3).

Обернена ґратка – імпульсний простір з розмірністю оберненої довжини, геометричну конфігурацію якого визначає *вектор оберненої ґратки*:

$$\bar{b} = m_1 \bar{b}_1 + m_2 \bar{b}_2 + m_3 \bar{b}_3, \quad (2.13)$$

де позначимо $b_i = \frac{2\pi}{\bar{a}_i} m_i$, $i = 1, 2, 3$, m_i – цілі числа. Взаємний напрямок

векторів:

$$\bar{b}_1 \perp [\bar{a}_2, \bar{a}_3], \bar{b}_2 \perp [\bar{a}_1, \bar{a}_3], \bar{b}_3 \perp [\bar{a}_1, \bar{a}_2], \quad (2.14)$$

де [...] – площини, у яких лежать вектори \bar{a}_i .

Відповідно до вказаних взаємних напрямків, для простої кубічної прямої ґратки обернена буде також простою кубічною; для прямої кубічної

гранецентрованої – обернена буде об'ємноцентрованою; для прямої кубічної об'ємноцентрованої – обернена буде гранецентрованої.

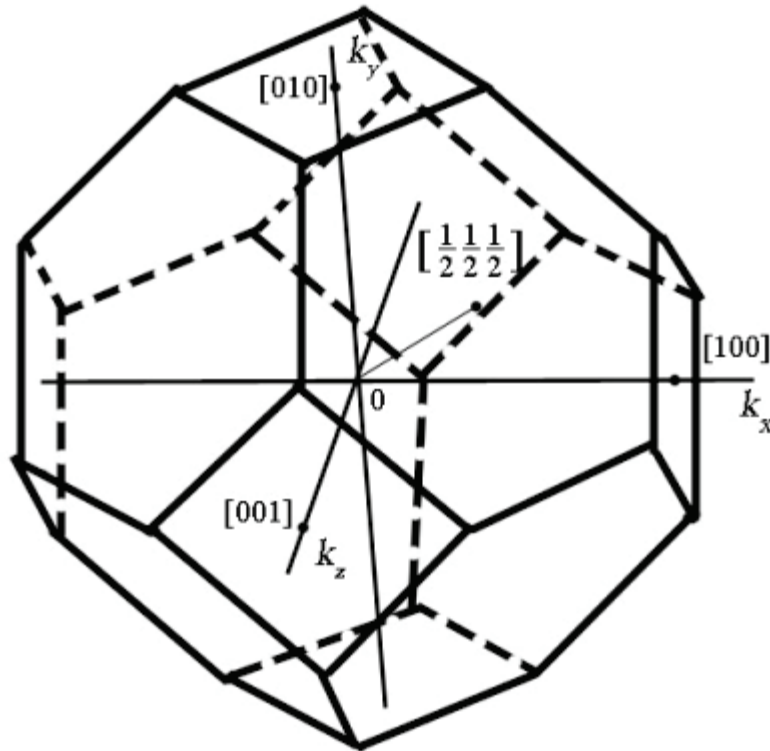


Рис.2.1. Зведена зона значень хвильових векторів у оберненому просторі структури алмазу

На цих прикладах і закінчується очевидний взаємоз'язок прямого і оберненого просторів. Для *структури алмазу* елементарна комірка оберненої ґратки, побудована на векторах $\bar{b}_1, \bar{b}_2, \bar{b}_3$, являє собою *зрізаний октаедр* (рис.2.1). Розглянемо одночасну трансляцію хвильової функції у прямому просторі на вектор \bar{a} і на вектор $2\pi/\bar{a}$ у оберненому:

$$\psi(\bar{r} + \bar{a}, \bar{k} + \frac{2\pi}{\bar{a}}) = \psi(\bar{r}) e^{i(\bar{k} + \frac{2\pi}{\bar{a}})\bar{a}} = \psi(\bar{r}) e^{i\bar{k}\bar{a}}. \quad (2.15)$$

Наслідок: *одночасна трансляція у прямому і оберненому просторах не порушила трансляційну властивість хвильової функції, тобто, хвильова функція*

має подвійну періодичність – період \bar{a} у прямій ґратці і $2\pi/\bar{a}$ у оберненій. Основний період хвильового вектора у оберненій ґратці: $0 \leq k_i \leq 2\pi/\bar{a}_i$. Для зручності цей період роблять симетричним: $-\frac{\pi}{a_i} \leq k_i \leq +\frac{\pi}{a_i}$. У межах

основного періоду знаходяться усі можливі значення хвильових векторів, тому він називається **зведеною зоною хвильових векторів**. Знаки „-” та „+” враховують те, що можуть існувати дві однакові електронні хвилі, які рухаються зустрічно.

Виділимо у просторово необмеженій періодичній структурі лінійний ланцюжок однакових атомів довжиною L , відстань між атомами a співпадає із періодом потенціального поля. Мінімальна можлива довжина електронної хвилі становить $2a$, максимальна – $2L$. Таким чином, довжина електронних хвиль лежить у межах:

$$2a \leq \lambda \leq 2L. \quad (2.16)$$

Найважливішим у (2.16) є те, що довжина електронних хвиль у кристалі *дискретна, тобто, може змінюватися лише кратно $2a$* . Дискретність хвиль означає, що і хвильовий вектор $k = 2\pi/\lambda$, і імпульс $p = h/\lambda$ також дискретні.

Дискретний спектр дає можливість обрахувати кількість дозволених значень k у межах зведеної зони. Для ланцюжка атомів $L = Na$, де N – число атомів у ланцюжку, лінійний розмір зведеної зони у оберненій ґратці $2\pi/a$. Скористаємося трансляційною властивістю хвильової функції (2.15) і поставимо умову: у просторово необмеженій періодичній структурі хвильові функції повинні „зшиватися” на зовнішніх межах ланцюжків, тобто мати однакову фазу $\psi(r + L) = \psi(r)$. Для виконання такої умови необхідно, щоб фазовий множник на кінцях ланцюжка $e^{ikL} = 1$, а це можливо, якщо $kL = 2\pi m$, де m – ціле число. Таким чином, маємо:

$$k = \frac{2\pi}{L} m = \frac{2\pi}{a} \frac{m}{N}, \text{ де } m = 0, 1, 2, \dots, N. \quad (2.17)$$

Для тривимірного простору:

$$k_i = \frac{2\pi}{a_i} \frac{m_i}{N_i}, \text{ де } i = 1, 2, 3. \quad (2.18)$$

Із наведених розрахунків випливає, що число значень k у зведеній зоні дорівнює числу атомів в об'ємі $L^3 = L_1 L_2 L_3$, тобто $N = N_1 N_2 N_3$ – кількість атомів у об'ємі L^3 . Якщо валентність атомів > 1 , відповідно збільшується і число значень k . Кожне значення k можуть мати 2 електрони з протилежними спінами (наслідок принципу Паулі).

У подальшому приймають $L^3 = 1$ і, відповідно, N – концентрація атомів.

Резюме п.2.7

1. Мінімальна можлива довжина електронної хвилі у ланцюжку однакових атомів, розташованих на відстані a , становить $\lambda_{\min} = 2a$, максимальна – $\lambda_{\max} = 2L$, де L – лінійна довжина ланцюжка атомів. Таким чином, довжина електронних хвиль лежить у межах: $2a \leq \lambda \leq 2L$.
2. Довжина електронних хвиль у кристалі *дискретна*, тобто може змінюватися лише цілократно $2a$.
3. Дискретність хвиль означає, що і *хвильовий* вектор $k = 2\pi / \lambda$, і імпульс $p = h / \lambda$ також дискретні.
4. Хвильовий вектор системи електронів *неоднозначний* і визначається з точністю до вектора $2\pi / \bar{a}$, де \bar{a} – період структури.
5. Основний діапазон значень хвильового вектора: $-\frac{\pi}{a_i} \leq k_i \leq +\frac{\pi}{a_i}$ називається *зведеною зоною* хвильових векторів. Знаки „-” та „+” враховують те, що можуть існувати дві однакові хвилі, які рухаються зустрічно.
6. У тривимірному просторі зведена зона займає певний об'єм, але, на відміну від реального фізичного простору, має *розмірність оберненої*

довжини і називається *елементарною коміркою* оберненої ґратки. Елементарні комірки щільно заповнюють увесь простір і утворюють обернену ґратку.

7. Число значень хвильового вектора \mathbf{k} у межах зведеної зони визначається числом атомів і їх валентністю, тобто кожний валентний стан кожного атома вносить одне дозволене значення хвильового вектора. Однакове значення \mathbf{k} можуть мати лише 2 електрони з протилежними спінами.

2.8. Імпульс і квазіімпульс

У ідеальній кристалічній структурі при відсутності зовнішніх силових полів повна енергія електронів не змінна у часі. Фізичні величини, які зберігаються у процесі руху називаються *інтегралами руху*. Інтегралам руху відповідають величини, оператори яких комутують із гамільтоніаном.

Наприклад, для оператора імпульсу $\hat{\mathbf{P}}$, якщо імпульс є інтегралом руху, повинна виконуватися умова:

$$\hat{H}\hat{\mathbf{P}} - \hat{\mathbf{P}}\hat{H} = 0. \quad (2.19)$$

Енергія залежить від імпульсу $E(\bar{\mathbf{P}})$, але не залежить від часу (інтеграл руху), відповідно, і імпульс повинен бути інтегралом руху, тобто $\frac{d\bar{\mathbf{P}}}{dt} = 0$.

Потенціальне поле ідеальної структури має трансляційну симетрію $U(\bar{\mathbf{r}} + \bar{\mathbf{n}}) = U(\bar{\mathbf{r}})$, але поле $U(\bar{\mathbf{r}})$ градієнтне, тобто $\nabla U(\bar{\mathbf{r}}) \neq 0$. У свою чергу, градієнту потенціалу відповідає певна внутрішня періодична сила $\bar{\mathbf{F}}_{\text{пер}}$, яка діє на електрон і повинна змінювати його імпульс. Виникає незвична ситуація при визначенні імпульсу у ідеальній періодичній структурі: *силове поле наявне, але енергія і імпульс не змінні у часі*. Наслідком цього є те, що *електронні хвилі у*

періодичній ідеальній структурі стаціонарні, незгасаючі. Імпульс \bar{P} , пов'язаний із періодичним потенціалом, називають **квазіімпульсом**.

Розглянемо ситуацію, коли у складі періодичного потенціалу $U(\bar{r})$ є додаткове внутрішнє поле $V_{вн}$, яке не має періодичності кристалічної структури. Повний гамільтоніан у цьому випадку має вигляд: $\hat{H} = \hat{T} + U(\bar{r}) + V_{вн} = \hat{H}_0 + V_{вн}$. Можна показати, що квазіімпульс буде змінюватися лише під дією внутрішньої неперіодичної сили $\bar{F}_{вн}$, зумовленої градієнтом неперіодичного поля $V_{вн}$:

$$\frac{d\bar{P}}{dt} = -\nabla V_{вн} = \bar{F}_{вн}. \quad (2.20)$$

Наслідок: *при будь-яких порушеннях періодичності потенціального поля кристалу буде змінюватися квазіімпульс \bar{P} електрона, електронні хвилі будуть розсіюватися*. Найчастіше розсіяння викликають такі дефекти, як теплові коливання атомів, нейтральні і заряджені домішкові атоми і т.п. Розсіяння є фізичною причиною обмеженої провідності твердих тіл.

Оператор звичайного імпульсу $\hat{p} = -i\hbar\nabla$ не комутує з гамільтоніаном. За наявності зовнішнього силового поля $V_{зов}$ і відповідної йому сили $\bar{F}_{зов}$:

$$\frac{d\bar{p}}{dt} = -\nabla V_{зов} = \bar{F}_{зов}, \quad (2.21)$$

квазіімпульс \bar{P} буде змінюватись як за рахунок $\bar{F}_{вн}$, так і за рахунок $\bar{F}_{зов}$:

$$\frac{d\bar{P}}{dt} = \bar{F}_{зов} + \bar{F}_{вн}. \quad (2.22)$$

Резюме п. 2.8

1. Одним із наслідків трансляційної симетрії потенціальної енергії і трансляційної властивості хвильової функції електрона є інтеграл руху – незмінний у часі *квазіімпульс*.

2. Незмінність у часі квазіімпульсу в ідеальній періодичній структурі означає, що електронні хвилі стаціонарні, тобто розсіяння відсутнє.

3. Наявність дефектів у структурі, які порушують її періодичність, викликає зміну квазіімпульсу і розсіяння електронних хвиль.

4. Якщо на кристал діє зовнішнє силове поле, квазіімпульс електрона буде змінюватись як під впливом зовнішнього поля, так і внутрішнього силового поля, пов'язаного з дефектами структури.

5. Зовнішнє силове поле виводить систему електронів із стану термодинамічної рівноваги, змінюючи модулі векторів квазіімпульсів і впорядковуючи їх напрямки. Внутрішнє силове поле повертає систему у стан термодинамічної рівноваги через механізми розсіяння.

6. Розсіяння є фізичною причиною обмеженої провідності твердих тіл.

2.9. Енергетичний спектр електронів

До утворення кристалічної структури із N атомів усі валентні енергетичні стани N – *кратно вироджені*, тобто, одному значенню енергії валентного електрона не взаємодіючих атомів E_a відповідає $\psi_{a1}, \psi_{a2}, \dots, \psi_{aN}$ хвильових функцій.

Взаємодію валентних електронів і утворення валентних зв'язків у кристалічній структурі можна розглядати як *збурення у виродженій системі*, яке повністю або частково знімає виродження. При цьому відбувається розщеплення атомних валентних рівнів на N окремих рівнів, кожному із яких відповідає своя хвильова функція: $E_1 \rightarrow \psi_1, E_2 \rightarrow \psi_2, \dots, E_N \rightarrow \psi_N$. Сукупність енергетичних рівнів E_1, E_2, \dots, E_N утворює **енергетичну зону**. Розщеплюються s і p валентні стани, незалежно від того, чи реалізовані вони електронами, чи ні. Таким чином утворюються s і p – зони а також можлива гібридизація станів у цих зонах (sp^3 - *гібриди*).

Необхідність утворення енергетичних зон у системі взаємодіючих електронів можна пояснити і таким чином: електрони – негативно заряджені мікрочастинки і між ними має місце електростатичне відштовхування. Стабільність системи валентних зв'язків забезпечується не лише притяганням електронів до позитивно заряджених ядер, але і магнітним притяганням електронів з протилежними спінами, яке компенсує їх електростатичне відштовхування. Така компенсація може мати місце лише для пари електронів із однаковою енергією, інші електронні пари повинні мати іншу енергію (принцип Паулі) і це також вимагає *зонного енергетичного спектру*.

При розв'язку рівняння Шредінгера хвильові функції мають вигляд хвиль Блоха для дійсних значень хвильового вектора \vec{k} у межах зведеної зони, а відповідні значення енергії електронів утворюють дозволені енергетичні зони: для електронів на валентних зв'язках – **валентну зону**, для першого збудженого стану валентних електронів – **зону провідності**.

Термін „дозволені енергетичні зони” означає, що тільки у цих діапазонах енергій маємо *нелокалізовані, хвильові розв'язки рівняння Шредінгера у межах кристалу*.

У напівпровідниках і діелектриках між валентною зоною і зоною провідності є *енергетична щілина* E_g (англ. energy gap) або *заборонена зона* (forbidden zone) – *енергетичний діапазон, у якому немає хвильових розв'язків і він відповідає уявним значенням хвильового вектора $\vec{k} = i\vec{\gamma}$ ($\vec{\gamma}$ -дійсне число)*.

Внаслідок цього хвильові функції втрачають комплексний характер і гармонічні властивості — стають звичайними експонентами, які спадають при віддаленні від точки \vec{r} .

З фізичної точки зору наявність забороненої зони енергій, розташованої між валентною зоною і зоною провідності (рис.2.2), означає, що *на граничних межах цих зон відбувається повне внутрішнє відбиття електронних хвиль*, тобто, електрони самі по собі не можуть вийти за межі дозволених зон.

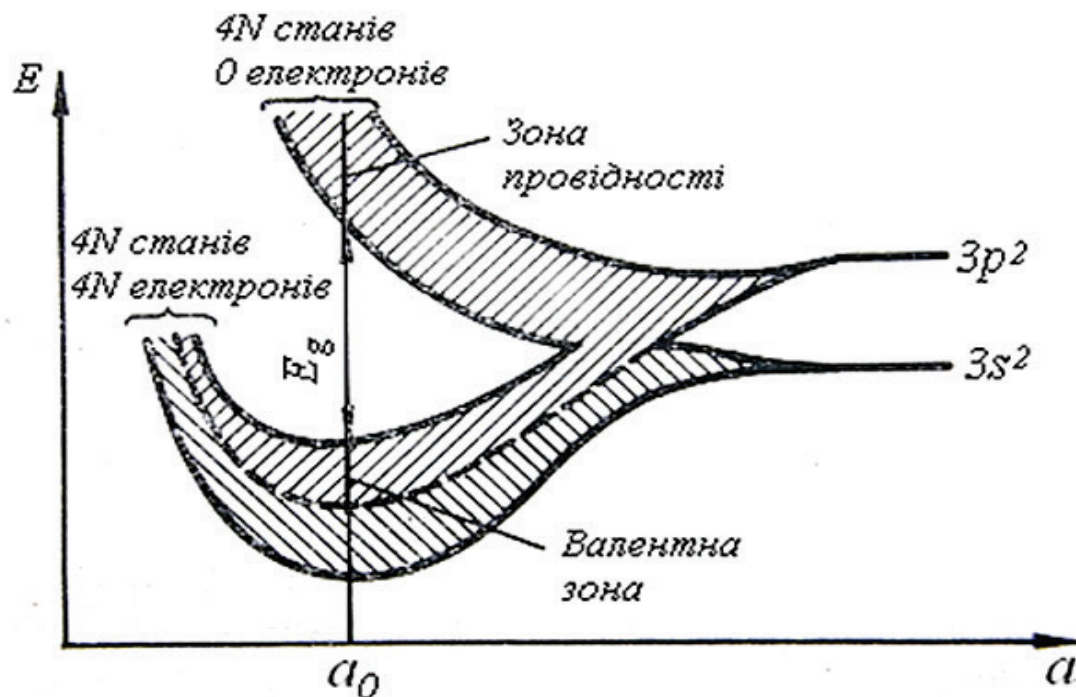


Рис.2.2. Розщеплення валентних рівнів кремнію і утворення енергетичних зон. У валентній зоні sp^3 – гібридні стани, повністю зайняті електронами. Зона провідності вільна.

Постійна структури $a_0 = 0,543$ нм

Заборонену зону можна розглядати як енергетичний бар'єр між зв'язаними на валентних рівнях станами електронів і вільними, яким відповідає зона провідності.

У забороненій зоні немає хвильових станів електрона, але можливі **локалізовані стани**. Хвильові функції локальних станів мають вигляд, подібний до радіальних функцій електрона у атомі водню: $\psi(\vec{r}) \sim r^2 e^{-\gamma r}$. Квадрат модуля такої функції — імовірність знайти електрон у точці з координатами \vec{r} має максимум у цій точці і спадає при віддаленні від неї.

Зазвичай, локальні енергетичні стани у забороненій зоні — це наслідок дефектів структури (домішкові атоми, вакансії, дислокації і т.п.).

Розглянемо загальні властивості енергії електронів у дозволених зонах як функції хвильового вектора:

– **однозначність** – електрон не може одночасно знаходитися у різних станах;

– *парність* – енергія не повинна залежати від напрямку руху електронної хвилі $E(\vec{k}) = E(-\vec{k})$;

– *періодичність* – усі можливі значення хвильового вектора знаходяться у зведеній зоні, яка має вигляд елементарної комірки оберненої ґратки. У сусідніх комітках значення хвильових векторів повторюються з періодом $\pm \frac{\pi}{a}$.

Вихід за межі елементарної комірки оберненої ґратки не дає нових значень енергії, тобто, у сусідній комірці повторяться ті ж самі значення енергії;

– *квазінеперервність* – хвильовий вектор у кристалі обмежених розмірів має дискретні значення. Відповідно і енергетичний спектр повинен бути дискретним. Число значень хвильового вектора у зведеній зоні $\geq N_a \approx 5 \cdot 10^{28} \text{ м}^{-3}$, де N_a – типове значення концентрації атомів у напівпровідниках і металах.

Розмір енергетичної зони $\approx 3 - 5 \text{ еВ}$, відповідно, різниця енергій сусідніх станів має порядок 10^{-28} еВ . Щоб краще уявити наскільки це щільний енергетичний спектр, порівняємо різницю енергій із середньою енергією теплового руху при 300 К , яка становить $2,6 \cdot 10^{-2} \text{ еВ}$. Кількість енергетичних станів, які проходять електрони у процесі теплового руху, величезна $\approx 10^{26}$. Тому і називають такий енергетичний спектр *квазінеперервним*.

2.10. Наповнення дозволених зон

У металах s – і p – валентні зони перекриваються і заборонена зона відсутня. Із урахуванням принципу Паулі ємність зон подвоюється і у металів наповнення дозволених зон часткове, вільні електрони існують за будь-яких умов. У напівпровідників при $T \approx 0 \text{ К}$ вільні електрони відсутні, тобто, валентна зона повністю заповнена, зона провідності пуста (рис.2.2). При $T > 0 \text{ К}$ з'являються вільні електрони за рахунок термоіонізації валентних зв'язків і локальних станів.

Процеси термоіонізації можуть створити у напівпровідниках не більше $10^{24} - 10^{26} \text{ м}^{-3}$ носіїв заряду, що на 2 – 4 порядки менше ємності дозволених

енергетичних зон. Незначне наповнення зон призводить до того, що вільні носії заряду зосереджені поблизу мінімуму енергії (дна) зони провідності і максимуму енергії (стелі) валентної зони, тобто, поблизу екстремальних значень енергії.

Наслідок: *немає необхідності розглядати увесь енергетичний діапазон дозволених зон, найбільший інтерес для практичного застосування напівпровідників становлять процеси поблизу екстремальних значень енергії.*

На відміну від зони провідності, у яку вільні *електрони* додаються внаслідок термо – або фотоіонізації валентних зв'язків, валентна зона, навпаки, *втрачає електрони* внаслідок цих же процесів. На місці неповних валентних зв'язків залишаються *енергетичні дірки* – не зайняті електронами енергетичні стани. Наявність вільних енергетичних станів дозволяє валентним електронам мігрувати по неповним валентним зв'язкам. Впорядковане зовнішнім полем просторове переміщення електронів зони провідності і електронам валентної зони по неповним зв'язкам – енергетичним діркам створить електричний струм. Кількість дірок у валентному спектрі незначна у порівнянні із загальною кількістю валентних електронів, тому зручно відслідковувати саме рух дірок, а не всієї маси валентних електронів. Але є особливість у поведінці дірок під дією зовнішнього поля: *негативно заряджені* електрони рухаються *проти напрямку поля* і залишають після себе

дірки, які рухаються *по напрямку поля*, тобто, як *позитивно заряджені частинки*.

Така особливість поведінки енергетичних дірок у частково наповненій валентній зоні стала основою для введення поняття *позитивно заряджених квазічастинок* – **дірок**, які мають заряд $+e$. Інші особливості дірок будуть розглянуті у подальшому.

2.11. Наближення ефективної маси

Розглянемо $E(\bar{k})$ поблизу абсолютного мінімуму енергії у зоні провідності. Позначимо точку мінімуму \bar{k}_0 і прийнемо її за початок відліку \bar{k} . Запишемо $E(\bar{k})$ у вигляді ряду Тейлора по степеням \bar{k} за умови, що $\bar{k} \ll \bar{k}_0$ (швидко згасаючий ряд):

$$E(\bar{k}) \approx E(\bar{k}_0) + \frac{1}{2} \sum_i \sum_j \frac{\partial^2 E}{\partial k_i \partial k_j} k_i k_j + \dots, \text{ де } i, j = x, y, z. \quad (2.23)$$

Перша похідна відсутня – мінімуму енергії, друга похідна має 9 компонент – тензор 2 рангу. Тензор симетричний, оскільки $\frac{\partial^2 E}{\partial k_i \partial k_j} = \frac{\partial^2 E}{\partial k_j \partial k_i}$, і число різних компонент зменшується до 6. Будь-який симетричний тензор 2 рангу відповідним поворотом координатної системи можна звести до діагонального вигляду. Нехай x, y, z – саме така система координат:

$$\sum_i \begin{bmatrix} \frac{\partial^2 E}{\partial k_x^2} & 0 & 0 \\ 0 & \frac{\partial^2 E}{\partial k_y^2} & 0 \\ 0 & 0 & \frac{\partial^2 E}{\partial k_z^2} \end{bmatrix} = \frac{\partial^2 E}{\partial k_x^2} + \frac{\partial^2 E}{\partial k_y^2} + \frac{\partial^2 E}{\partial k_z^2}. \quad (2.24)$$

Прийнемо $E(\bar{k}_0) = 0$ – початок відліку енергії і запишемо $E(\bar{k})$ поблизу абсолютного мінімуму енергії для кристалу:

$$E(k^2) \approx \frac{1}{2} \left(\frac{\partial^2 E}{\partial k_x^2} k_x^2 + \frac{\partial^2 E}{\partial k_y^2} k_y^2 + \frac{\partial^2 E}{\partial k_z^2} k_z^2 \right). \quad (2.25)$$

Запишемо $E(\bar{k})$ для електрона у вільному просторі:

$$E(k^2) = \frac{1}{2} \left(\frac{\hbar^2}{m_e} k_x^2 + \frac{\hbar^2}{m_e} k_y^2 + \frac{\hbar^2}{m_e} k_z^2 \right). \quad (2.26)$$

Співставивши (2.25) для енергії електрона у кристалі і у вільному просторі (2.26) по розмірності, отримаємо:

$$m_{ii}^* = \hbar^2 \left(\frac{\partial^2 E}{\partial k_i^2} \right)^{-1}, \text{ де } i = x, y, z. \quad (2.27)$$

Компоненти m_{ii}^* мають розмірність маси і називаються **ефективними масами**.

Енергія електрона у кристалі поблизу мінімуму зони провідності має такий вигляд у k – та p – представленнях:

$$E(k^2) = \frac{1}{2} \left(\frac{\hbar^2}{m_{xx}^*} k_x^2 + \frac{\hbar^2}{m_{yy}^*} k_y^2 + \frac{\hbar^2}{m_{zz}^*} k_z^2 \right); E(p^2) = \frac{1}{2} \left(\frac{p_x^2}{m_{xx}^*} + \frac{p_y^2}{m_{yy}^*} + \frac{p_z^2}{m_{zz}^*} \right). \quad (2.28)$$

Введена через закон дисперсії енергії $E(k^2)$ ефективна маса електрона у кристалі еквівалентна „гравітаційній масі”.

Ефективна маса обернено пропорційна другій похідній $E(k^2)$. Друга похідна – міра кривизни функції у точці екстремуму. Обернена кривизні величина – радіус кривизни. Ефективна маса прямо пропорційна радіусу кривизни $E(k^2)$.

Кривизна функції у точці екстремуму має знак: «–» у максимумі, «+» у мінімумі. З фізичної точки зору «від’ємна маса» у валентній зоні – нонсенс. Максимум енергії валентних електронів – стеля валентної зони, але ж нас цікавлять лише дірки у валентній зоні. Для дірок стеля валентної зони – мінімум енергії, тому напрямок енергетичної вісі у валентній зоні вибирають протилежним її напрямку у зоні провідності. У такій системі відліку енергії ефективна маса дірок стає позитивною.

На основі вище вказаного, можна дати таке визначення поняття **дірка**: квазічастинка із зарядом $+e$ і позитивною ефективною масою.

Щоб ефективна маса виконувала таку ж роль як і дійсна маса електрона, вона не повинна залежати від енергії і хвильового вектору, які змінюються під дією зовнішніх сил. Параболічна залежність $E \sim k^2$ якраз і дозволяє виконати

цю умову, оскільки друга похідна для такої залежності – константа. Але у глибині дозволеної зони ця умова не виконується. У середній частині зони ($k = \pm \frac{\pi}{2a}$) змінюється знак кривизни і вона проходить через нульове значення, а $m^* \rightarrow \infty$ і втрачає фізичний зміст. Таким чином, *наближення ефективної маси справедливе* лише для носіїв заряду, які знаходяться *поблизу дна зони провідності або стелі валентної зони*.

У наближенні ефективної маси можна вважати, що електрон у кристалі має властивості електрона у вакуумі за винятком величини маси. Гамільтоніан електрона у вакуумі $\hat{H}_{\text{вак}} = -\frac{\hbar^2}{2m_e} \nabla^2 + U$, де $U = \text{const}$ – не градієнтний потенціал, що не створює силові поля у вакуумі. Хвильова функція електрона у цьому випадку – плоска хвиля $\psi(\vec{r}) = Ae^{i\vec{k}\vec{r}}$. Гамільтоніан електрона у кристалі у наближенні ефективної маси $\hat{H}_{\text{крис}} = -\frac{\hbar^2}{2m^*} \nabla^2 + U$, де $U = \text{const}$. Іншими словами, при заміні градієнтного потенціалу періодичного поля $U(\vec{r})$ на не градієнтний потенціал U для збереження повної енергії довелося змінити її кінетичну складову шляхом заміни дійсної маси m_e на ефективну m^* . Важливим наслідком такої перебудови гамільтоніана є те, що *хвильова функція*

електрона у наближенні ефективної маси стає плоскою хвилею $Ae^{i\vec{k}\vec{r}}$, де $A = \text{const}$.

У наближенні ефективної маси не обов'язково знати точну конфігурацію періодичного потенціалу кристалічної структури, достатньо у рівняннях руху електронів і дірок замінити дійсну масу на ефективну.

Якщо зафіксувати певне значення енергії $E(k^2) = \text{const}$, то отримаємо рівняння замкненої поверхні другого порядку – еліпсоїда загального вигляду.

Запишемо це рівняння у канонічній формі (початок відліку \bar{k} – точка екстремуму \bar{k}_0):

$$\frac{k_x^2}{a_x^2} + \frac{k_y^2}{a_y^2} + \frac{k_z^2}{a_z^2} = 1, \quad (2.29)$$

$$\text{де } a_i = \sqrt{\frac{2m_{ii}^*}{\hbar^2} E} - \text{напіввісі еліпсоїда, } i = x, y, z. \quad (2.30)$$

Окремі випадки: 1) $a_x = a_y = a_z = a$; $m_{xx}^* = m_{yy}^* = m_{zz}^* = m^*$; $k_x^2 + k_y^2 + k_z^2 = a^2$ – рівняння сфери з радіусом $a = \sqrt{\frac{2m^*}{\hbar^2} E}$.

Наслідок: якщо ефективна маса скаляр, поверхні рівної енергії – сферичні і навпаки.

$$2) \ a_x = a_y = a_t, \ a_z = a_\ell; \ m_{xx}^* = m_{yy}^* = m_t^*, \ m_{zz}^* = m_\ell^*; \ ; \ \frac{k_x^2 + k_y^2}{a_t^2} + \frac{k_z^2}{a_\ell^2} = 1 - \text{рівняння}$$

еліпсоїду обертання, вісь обертання – z . Компонента ефективної маси m_t^* – поперечна маса (індекс t від англ. *transvers* – *поперечний*). Компонента ефективної маси m_ℓ^* – подовжня маса (індекс ℓ від англ. *longitudinal* – *подовжній*).

Наслідок: якщо тензор ефективної маси має лише дві нерівні компоненти, поверхні рівної енергії – еліпсоїди обертання і навпаки.

2.12. Зонна структура і ефективна маса

Положення абсолютного екстремуму енергії $E(\bar{k}_0)$ у зведений зоні може бути у будь-якій точці. Значення \bar{k}_0 у цьому випадку визначає найімовірнішу довжину хвилі, частоту і енергію основного стану електронів. Для кристалів з високим рівнем симетрії – кубічних структур, які найбільш характерні для напівпровідників, реалізуються такі ситуації:

1) якщо абсолютний екстремум у центрі зведеної зони $\bar{k}_0 = 0$, він єдиний, поверхні рівної енергії сферичні, ефективна маса скаляр.

2) якщо абсолютний екстремум у проміжній точці на одній із головних вісей симетрії куба ($\langle 100 \rangle, \langle 111 \rangle, \langle 110 \rangle$) у межах $0 < k_0 \leq +\frac{\pi}{a}$, поверхні рівної енергії – еліпсоїди обертання, вісь обертання – головна вісь куба. Ефективна маса має дві компоненти: m_t^* і m_ℓ^* .

Із урахуванням парності енергії $E(\bar{k}) = E(-\bar{k})$, число еквівалентних по енергії екстремумів удвічі більше числа головних вісей певного типу (6 для вісі $\langle 100 \rangle$ та $\langle 110 \rangle$, 8 для $\langle 111 \rangle$).

Базуючись на цих загальних положеннях, розглянемо зонну структуру найбільш вживаних напівпровідників.

Зони провідності кремнію і германію. Енергетичний профіль зонної структури кремнію показаний на рис.2.3. Розгортка енергетичного профілю забороненої зони (заштрихована смуга) для певних напрямків у просторі оберненої ґратки, показаної на рис.2.1.

Абсолютний мінімум енергії у кремнії на вісі $\langle 100 \rangle$, координати мінімуму $k_{\min} \approx \frac{2}{3} \frac{\pi}{a}$, число екстремумів – 6, число повних еліпсоїдів – 6,

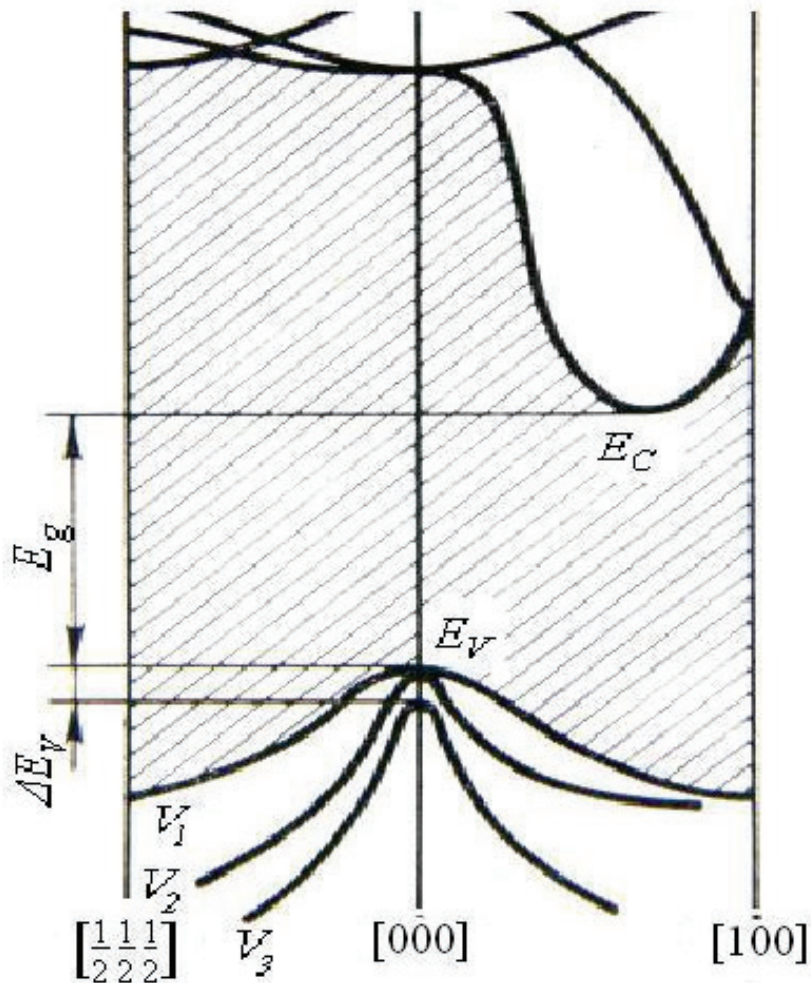


Рис.2.3. Енергетичний профіль кремнію

ефективні маси: $m_t^* = 0,19 m_e$, $m_\ell^* = 0,98 m_e$, анізотропія еліпсоїдів $m_\ell^* / m_t^* = 5,16$.

Просторове розташування екстремумів для кремнію показане на рис.2.5.

Енергетичний профіль зонної структури германію показаний на рис.2.4. Розгортка енергетичного профілю забороненої зони (заштрихована смуга) для певних напрямків у просторі оберненої ґратки, показаної на рис.2.1.

Розмір зведеної зони по $\langle 111 \rangle$ становить $\pi/2a$, точка абсолютного мінімуму енергії – центр шестигранника із координатами $\left[\frac{1}{2} \frac{1}{2} \frac{1}{2} \right]$.

Число екстремумів – 8, число повних еліпсоїдів – 4, ефективні маси: $m_{\ell}^* = 0,082 m_e$, $m_{\ell}^* = 1,58 m_e$, $m_{\ell}^* / m_t^* = 19,3$. Просторове розташування екстремумів для германію показано на рис. 2.5.

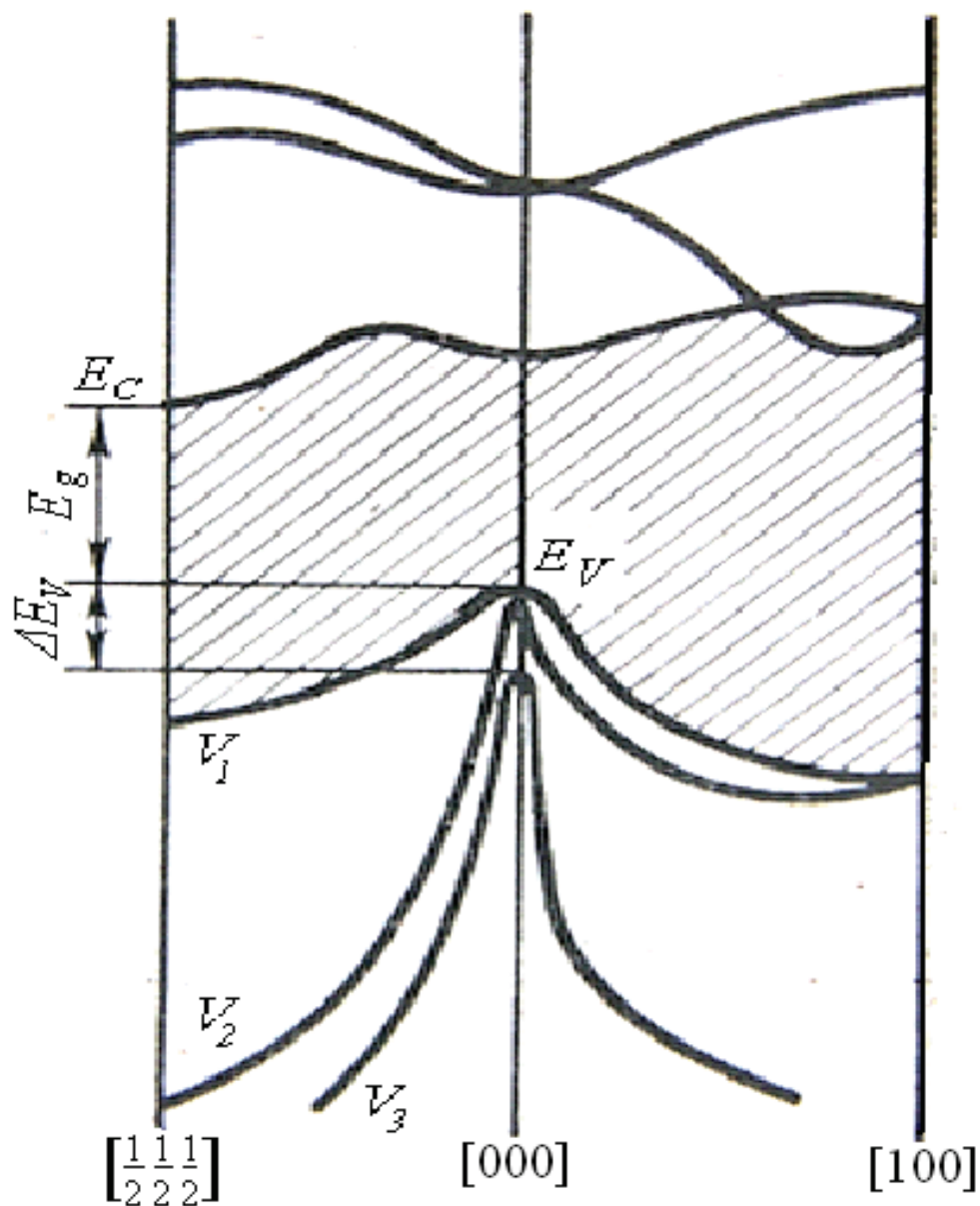


Рис.2.4. Енергетичний профіль германію

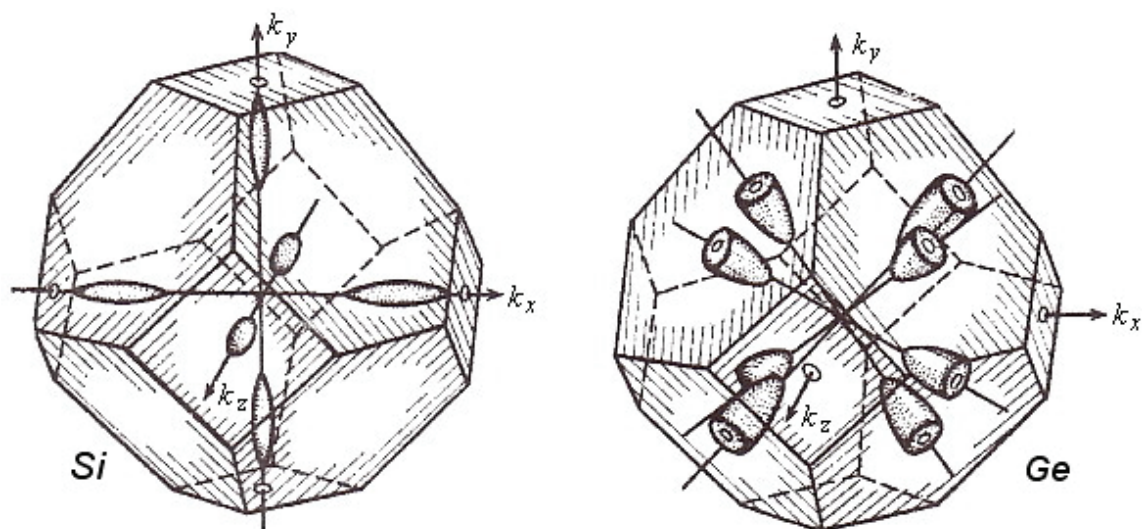


Рис.2.5. Просторове розташування поверхонь рівної енергії у зонах провідності кремнію і германію

Валентні зони кремнію і германію. Абсолютний максимум енергії валентної зони як у кремнії, так і у германії – точка $k_{\max} = 0$ (центр зведеної зони). Відповідно до загальних положень, цей максимум енергії єдиний. Валентні зв'язки – результат сильної взаємодії s і p електронних станів і утворення sp^3 гібридних зв'язків. Спін – орбітальна взаємодія призводить до розщеплення валентної зони на три підзони V_1 , V_2 , V_3 із різними значеннями ефективних мас дірок (рис.2.3, 2.4). Прийнято називати V_1 підзоною важких дірок, V_2 – легких дірок, V_3 – проміжних дірок.

Форма поверхонь рівної енергії для V_1 і V_2 – деформовані сфери (рис.2.6.)

Середні значення маси дірок у кремнії: важкі дірки $m_{pB}^* = 0,52m_e$, легкі – $m_{pL}^* = 0,16m_e$, співвідношення $m_{pB}^* / m_{pL}^* = 3,25$.

У германії: $m_{pB}^* = 0,34m_e$, $m_{pL}^* = 0,04m_e$, $m_{pB}^* / m_{pL}^* = 8,5$.

Підзона проміжних дірок V_3 у кремнії експериментально не спостерігається ($\Delta E_V = 0,04$ еВ), у германії вона опущена на $\Delta E_V = 0,3$ еВ і її

присутність виявляється лише на спектрах поглинання світла вільними дірками.

У кремнії і германії мінімуми і максимуми енергії відповідають різним значенням хвильових векторів, тобто, $k_{\min} \neq k_{\max}$. Такі напівпровідники називаються *непрямозонними*. Фізичний зміст цієї назви буде пояснений при розгляді механізмів поглинання світла.

Зонна структура сполук $A^{III}B^V$. У більшості сполук $A^{III}B^V$ $k_{\min} = k_{\max}$, тобто, вони *прямозонні*. Виняток становлять GaP, AlP.

Зона провідності із $k_{\min} = 0$ має сферичні поверхні рівної енергії.

Скалярна ефективна маса електронів лежить у досить широких межах: $0,01m_e < m_n^* < 2m_e$. Певні особливості має зонна структура GaAs, показана на рис.2.7.

У зоні провідності два мінімуми енергії: абсолютний у центрі зведеної зони з ефективною масою $m_n^* = 0,067m_e$ і неабсолютний на межі зведеної зони по напрямку $\langle 111 \rangle$, піднятий на 0,36 еВ відносно абсолютного. Для цього мінімуму поверхні рівної енергії – еліпсоїди обертання із масами $m_t^* = 0,075m_e$, $m_\ell^* = 1,9m_e$; $m_\ell^* / m_t^* = 25,3$, число повних еліпсоїдів – 4.

Увага до неабсолютного мінімуму зумовлена тим, що у сильних електричних полях відбувається його заселення електронами із одночасним суттєвим збільшенням їх ефективної маси. За певних умов це можна використати для генерації надвисокочастотних коливань (ефект Ганна).

Для усіх сполук $A^{III}B^V$ характерним є помітне відхилення від параболічного закону $E(k^2)$ у зоні провідності, що означає залежність m_n^* від квазіімпульсу електрона, який змінюється під дією зовнішнього електричного поля.

Структура валентних зон сполук $A^{III}B^V$ в цілому така ж, як у Si і Ge – максимум енергії у центрі зведеної зони, дві підзони важких і легких дірок, поверхні рівної енергії сферичні, ефективні маси скалярні.

У більшості сполук різниця мас $m_{рл}^*$ і $m_{рв}^*$ – у межах похибки визначення, середнє значення маси дірок лежить у діапазоні $0,3m_e < m_p^* < 0,6m_e$.

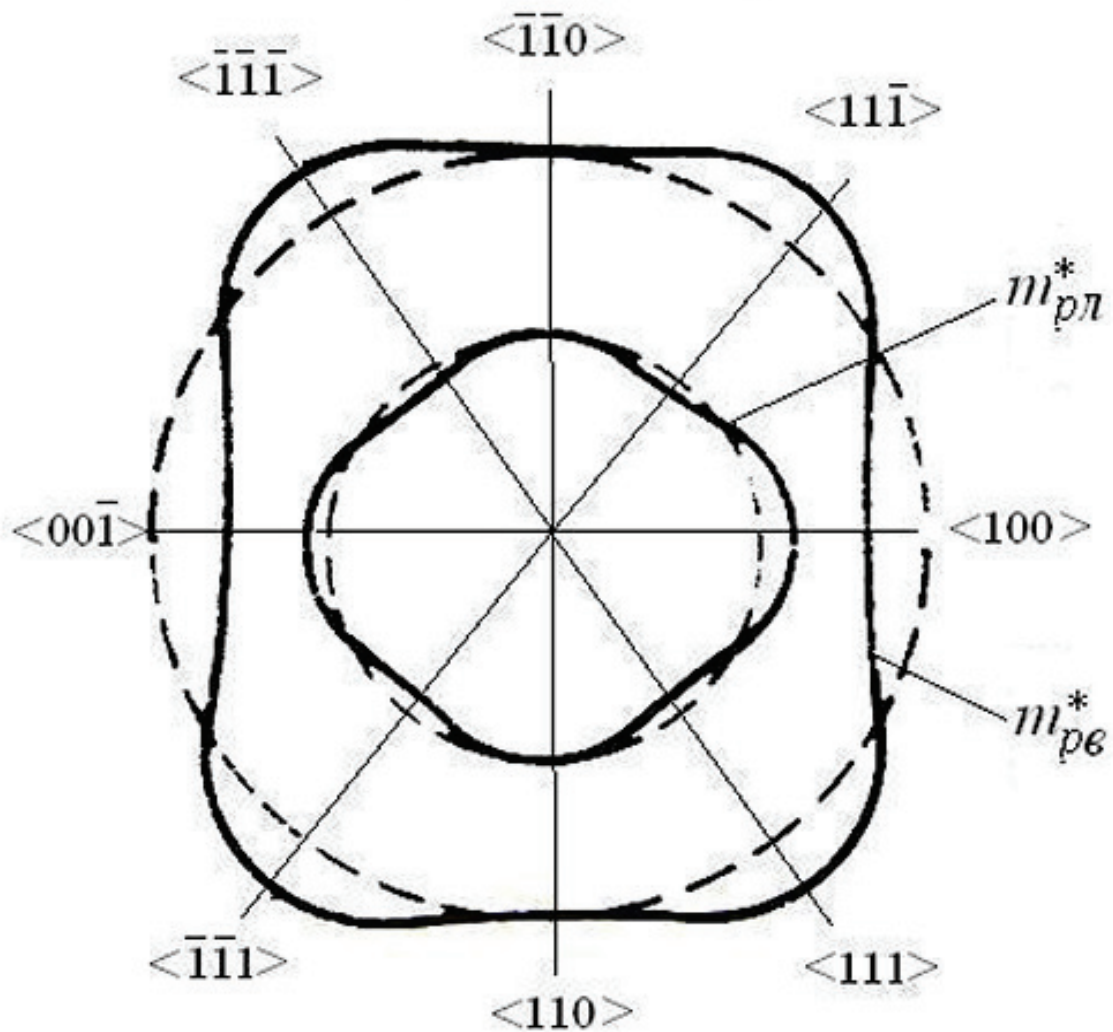


Рис.2.6. Діаметральний переріз ізоенергетичних деформованих сфер важких і легких дірок (суцільні лінії). Еквівалентні за щільністю станів сфери (штрихові лінії)

Маси важких і легких дірок надійно визначені лише для GaAs

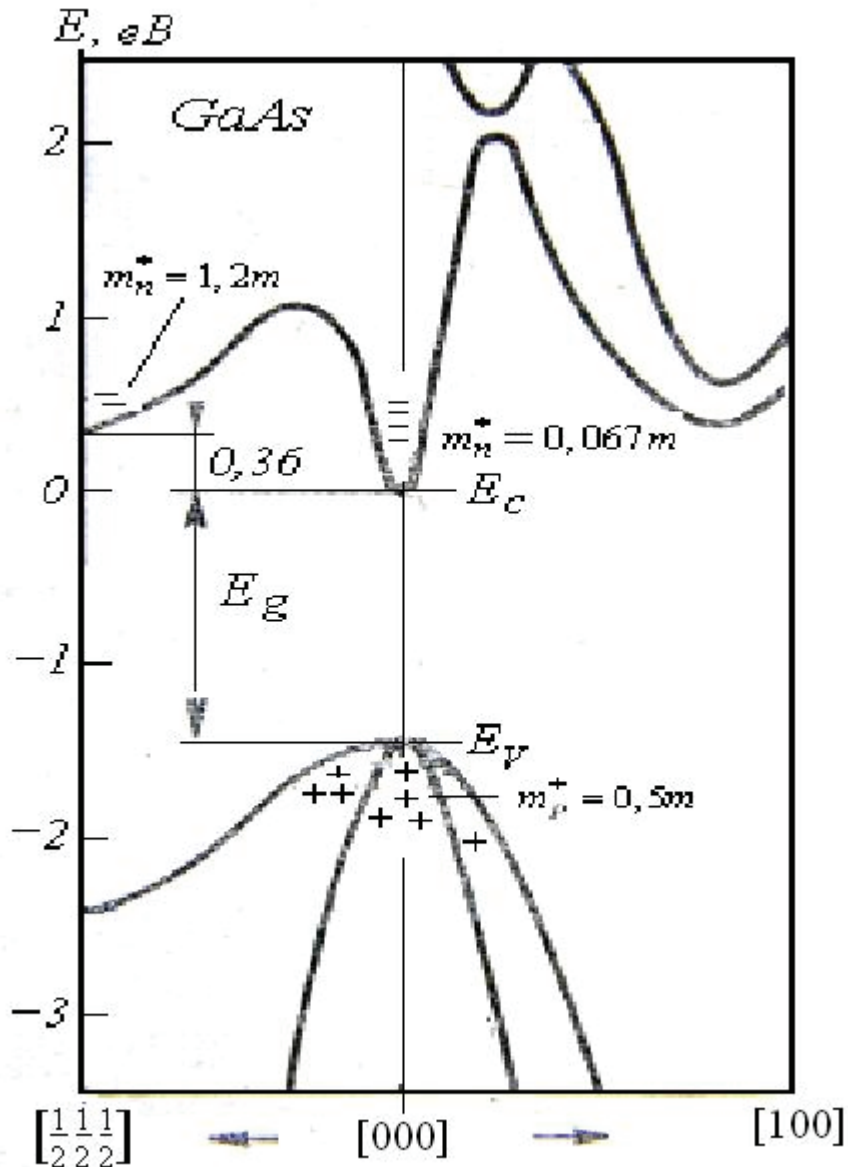


Рис.2.7. Енергетичний профіль арсеніду галію

($m_{pB}^* = 0,45m_e$, $m_{pL}^* = 0,12m_e$, $m_{pB}^* / m_{pL}^* = 3,75$) та для InSb ($m_{pB}^* = 0,6m_e$, $m_{pL}^* = 0,012m_e$, $m_{pB}^* / m_{pL}^* = 50$).

2.13. Ширина забороненої зони.

Формальне визначення: ширина забороненої зони $E_g = E_c - E_v$ – різниця енергій абсолютних екстремумів зони провідності і валентної зони.

Таке визначення застосовне як для прямозонних, так і непрямоzonних напівпровідників. Можна дати визначення і за фізичним змістом:

енергетичний бар'єр, який відділяє локалізований на валентних зв'язках стан електронів від вільного у межах зони провідності.

Ширина забороненої зони – фундаментальний параметр зонної структури кристалічних напівпровідників. Вона залежить від природи атомів, які утворили кристал, і його кристалографічної структури.

Зовнішні обставини – температура, механічна деформація – мають певний, але, у більшості випадків, незначний вплив на енергетичний розмір забороненої зони.

Термічне розширення кристалів із структурою алмазу і сфалериту при нагріванні означає, що збільшується середня відстань між атомами, а це призводить до зменшення ширини забороненої зони (рис.2.2).

У діапазоні температур $180 < T < 350$ К залежність $E_g(T)$ стає лінійною. Відносно $T = 300$ К цю закономірність, можна записати таким чином:

$$E_g(T) = E_g(300) + \gamma(T - 300), \quad (2.31)$$

де $\gamma = -dE_g(T)/dT$ – температурний коефіцієнт $E_g(T)$.

Значення $E_g(0)$, $E_g(300)$ та γ для деяких напівпровідників наведені у табл. 2.1. Там же наведений також приклад сполук групи $A^{IV}B^{VI}$ із структурою кам'яної солі, у яких термічне розширення викликає збільшення ширини забороненої зони.

Таблиця 2.1.

Параметр	InSb	Ge	Si	GaAs	PbS	PbTe
$E_g(0)$, еВ	0,23	0,74	1,17	1,52	0,286	0,19
$E_g(300)$, еВ	0,17	0,66	1,12	1,42	0,41	0,31
$\gamma \cdot 10^4$, еВ/град	- 2,9	- 3,9	- 2,4	- 4,3	$\approx + 4$	$\approx + 4$

3. Статистика електронів і дірок у напівпровідниках

3.1. Щільність енергетичних станів

Аналіз реального наповнення електронами і дірками енергетичних зон напівпровідників, наведений у п.2.10, показав, що немає необхідності розглядати увесь енергетичний діапазон дозволених зон, достатньо визначити енергетичний розподіл носіїв заряду поблизу екстремальних значень енергії.

Для початку розглянемо найпростіший випадок скалярної ефективної маси електронів – мінімум енергії у центрі зведеної зони $\bar{p}_{\min} = 0$, значення енергії $E_c(\bar{p}_{\min}) = 0$ – дно зони провідності приймаємо за початок відліку енергії (цей випадок характерний для сполук $A^{III}B^V$ (див. п. 2.12)).

У наближенні ефективної маси залежність енергії від імпульсу параболічна:

$$E(\bar{p}) = \frac{p^2}{2m_n^*} = \frac{1}{2m_n^*}(p_x^2 + p_y^2 + p_z^2). \quad (3.1)$$

Виділимо у імпульсному просторі сферичний прошарок між двома ізоенергетичними поверхнями з радіусами E і $E + dE$. Об'єм сферичного прошарку між цими сферами:

$$dV_p = 4\pi p^2 dp \quad (3.2)$$

Кожне значення імпульсу займає у імпульсному просторі об'єм h^3 і число таких значень імпульсу у об'ємі dV_p із урахуванням (3.2) і принципу Паулі буде:

$$N(p)dp = 2 \frac{dV_p}{h^3} = \frac{8\pi}{h^3} p^2 dp, \quad (3.3)$$

де $N(p)$ – щільність p – станів.

Введемо $N(E)dE$ – кількість енергетичних станів у інтервалі енергій $E, E + dE$. Із (3.1) визначимо p^2 та dp і у (3.3) перейдемо від $N(p)dp$ до $N(E)dE$:

$$N(E)dE = 4\pi(2m_n^*/h^2)^{3/2} E^{1/2} dE. \quad (3.4)$$

Фізичний зміст $N(E)$ – кількість енергетичних станів у одиничному інтервалі енергій і у одиниці об'єму кристала – щільність енергетичних станів.

Розглянемо випадок тензорної ефективної маси, характерний для германію та кремнію (див. п.2.12.). Залежність $E(p^2)$ має вигляд (2.29).

Канонічне рівняння еліпсоїдальної поверхні $E(p^2) = \text{const}$:

$$\frac{p_x^2}{a_x^2} + \frac{p_y^2}{a_y^2} + \frac{p_z^2}{a_z^2} = 1, \quad (3.6)$$

де a_x^2, a_y^2, a_z^2 – піввісі еліпсоїда:

$$a_i = (2m_{ii}^* E)^{1/2}, (i = x, y, z) . \quad (3.7)$$

Об'єм еліпсоїда із урахуванням (3.7):

$$V_p = \frac{4\pi}{3} a_x a_y a_z = \frac{8\pi}{3} (2m_{xx}^* m_{yy}^* m_{zz}^*)^{1/2} E^{3/2} . \quad (3.8)$$

Об'єм еліпсоїдального прошарку між ізоенергетичними поверхнями з енергіями E і $E + dE$ буде:

$$dV_p = 4\pi (2m_{xx}^* m_{yy}^* m_{zz}^*)^{1/2} E^{1/2} dE . \quad (3.9)$$

Далі діємо так, як і у випадку скалярної ефективної маси:

$$N(E)dE = 4\pi (2/h^2)^{3/2} M (m_{xx}^* m_{yy}^* m_{zz}^*)^{1/2} E^{1/2} , \quad (3.10)$$

де M – число повних еліпсоїдів у зоні провідності.

Формулу (3.10) можна привести до такого ж вигляду, як і (3.4), якщо ввести позначення:

$$M^2 (m_{xx}^* m_{yy}^* m_{zz}^*) = m_{dn}^{*3}; \quad m_{dn}^* = M^{2/3} (m_t^{*2} m_\ell^*)^{1/3} , \quad (3.10)$$

де m_{dn}^* – **ефективна маса щільності станів** для еліпсоїдів обертання навколо вісі Z .

Фізичний зміст введення m_{dn}^* – еліпсоїдальні ізоенергетичні поверхні усіх абсолютних мінімумів енергії замінюються на одну еквівалентну по кількості енергетичних станів сферу, радіус якої пропорційний m_{dn}^* . У межах окремо взятого мінімуму енергії ефективна маса анізотропна, але, завдяки симетричному просторовому розташуванню мінімумів енергії, середнє по всім еліпсоїдам значення маси m_{dn}^* ізотропне.

Визначаємо m_{dn}^* для зон провідності електронного кремнію і германію по (3.10) відповідно даним, наведеним у п. 2.12.1. Результати розрахунків m_{dn}^* зведені у табл. 3.1.

Таблиця 3.1

Напівпровідник	Електронний				Дірковий		
Параметр	M	m_t^*	m_ℓ^*	m_{dn}^*	m_{pv}^*	m_{pl}^*	m_{dp}^*
Кремній	6	$0,19m_e$	$0,98m_e$	$1,08m_e$	$0,52m_e$	$0,16m_e$	$0,58m_e$
Германій	4	$0,082m_e$	$1,59m_e$	$0,56m_e$	$0,34m_e$	$0,04m_e$	$0,35m_e$

Для валентних зон кремнію і германію ситуація із ефективною масою провідності дірок m_{dp}^* дещо інша через наявність підзон важких і легких дірок. Ізоенергетичні поверхні – деформовані сфери (рис.2.6), але їх можна апроксимувати звичайними сферами і відповідними значеннями ефективних мас важких і легких дірок. Загальна щільність енергетичних станів визначається сумою щільностей станів по підзонам важких і легких дірок:

$$N(E) = 4\pi(2/h^2)^{3/2}(m_{pv}^{*3/2} + m_{pl}^{*3/2})E^{1/2}. \quad (3.11)$$

Позначимо:

$$m_{dp}^* = (m_{pv}^{*3/2} + m_{pl}^{*3/2})^{2/3} \quad (3.12)$$

і (3.11) матиме вигляд:

$$N(E) = 4\pi(2m_{dp}^*/h^2)^{3/2}E^{1/2}. \quad (3.13)$$

Визначимо m_{dp}^* для валентних зон кремнію і германію по (3.12) відповідно даним, наведеним у п. 2.12. Результати розрахунків m_{dp}^* зведені у табл. 3.1.

3.2. Розподіл носіїв заряду у дозволених зонах

Виділимо у межах зони провідності енергетичний інтервал $E, E + dE$. Кількість електронів у цьому інтервалі:

$$dn = N(E)f_{0n}(E)dE, \quad (3.14)$$

де $f_{0n}(E)$ – функція розподілу Фермі – Дірака:

$$f_{0n}(E) = (\exp \frac{E - F_{\phi n}}{k_B T} + 1)^{-1}, \quad (3.15)$$

яка означає імовірність того, що енергетичний рівень E при температурі T зайнятий електроном, $F_{\phi n}$ – рівень Фермі – імовірність перебування електрона на цьому рівні дорівнює $1/2$, k_B – постійна Больцмана.

Враховуючи незначне наповнення зони провідності електронами, про що йшла мова у п.2.10, верхню її межу можна вважати нескінченною і інтегрування по енергіям робити у межах $0 \leq E < \infty$. Підставимо (3.10) у (3.4), потім (3.4) та (3.15) у (3.14) і інтегруємо:

$$n = \int_0^{\infty} \frac{4\pi}{h^3} (2m_{dn}^*)^{3/2} E^{1/2} (\exp \frac{E - F_{\phi n}}{k_B T} + 1)^{-1} dE. \quad (3.16)$$

Інтегралу (3.16) надамо безрозмірну форму шляхом таких позначень:

$\frac{E}{k_B T} = x$, $\frac{F_{\phi n}}{k_B T} = \xi_n$, винесемо за інтеграл усе, що не залежить від енергії:

$$n = 4\pi (2m_{dn}^* k_B T / h^2)^{3/2} \int_0^{\infty} x^{1/2} (\exp(x - \xi_n) + 1)^{-1} dx. \quad (3.17)$$

Інтеграл у формулі (3.17) позначимо $F_{1/2}(\xi_n)$ і він називається інтеграл Фермі порядку $1/2$. У елементарних функціях цей інтеграл не вирішується, можливі лише наближені його значення:

$$F_{1/2}(\xi_n) = \frac{\sqrt{\pi}}{2} \exp \xi_n \text{ за умови: } -\infty < \xi_n < -1; -\infty < F_{\phi n} < k_B T. \quad (3.18)$$

$$F_{1/2}(\xi_n) = \frac{\sqrt{\pi}}{2} (\exp(-\xi_n) + 1/4)^{-1} \text{ за умови: } -1 < \xi_n < 5; -k_B T < F_{\phi n} < 5k_B T. \quad (3.19)$$

$$F_{1/2}(\xi_n) = \frac{4}{3\sqrt{\pi}} \xi_n^{3/2} \text{ за умови: } 5 < \xi_n < 8; 5k_B T < F_{\phi n} < 8k_B T. \quad (3.20)$$

Від'ємне значення рівня Фермі у випадку (3.18) означає, що він розташований нижче дна зони провідності, тобто у забороненій зоні. Це свідчить про те, що мають місце або міжзонні електронні переходи, або переходи між зоною провідності і локальними рівнями у забороненій зоні. Вільні електрони з'являються у зоні із джерела поза зоною, але їх кількість набагато менша числа енергетичних рівнів (див. п.2.10). Це означає, що імовірність знайти електрон на певному рівні $f_n(E) \ll 1$, одиницею у знаменнику (3.15) можна знехтувати і перейти до функції розподілу Максвелла – Больцмана:

$$f_{0n}(E) = \exp \left(-\frac{E - F_{n\phi}}{k_B T} \right). \quad (3.20a)$$

Розподілу (3.20a) відповідає невироджений електронний газ.

Позитивне значення рівня Фермі у випадку (3.20) означає, що він розташований у зоні провідності, вище дна зони щонайменше на $5k_B T$, тобто зона частково наповнена електронами до цього рівня енергії. Така ситуація відповідає повністю виродженому електронному газу.

Випадок (3.19) проміжний – рівень Фермі поблизу дна зони провідності $F_{\phi n} \approx 0 \dots 2k_B T$ – відповідає частково виродженому електронному газу.

Невироджений напівпровідник. Замінімо інтеграл у формулі (3.17) на його значення у випадку (3.18) і для концентрації електронів у зоні провідності отримаємо:

$$n = 2(2\pi m_{dn}^* k_B T / h^2)^{3/2} \exp(F_{\phi n} / k_B T) = N_c \exp(F_{\phi n} / k_B T). \quad (3.21)$$

Передекспонентний множник N_c – ефективна щільність енергетичних станів зони провідності у діапазоні енергій $0 < E \leq k_B T$.

Функція розподілу для дірок – імовірність того, що рівень енергії E у валентній зоні при температурі T вільний, має вигляд:

$$f_p(E_p) = 1 - f_n(E_p) = (\exp(\frac{F_{\phi p} - E_p}{k_B T}) + 1)^{-1}. \quad (3.22)$$

Враховуючи, що напрямок відліку енергії у валентній зоні протилежний відносно зони провідності $E_p = -E$, початок відліку зміщений на $-E_g$, рівні Фермі для електронів $F_{\phi n}$ і дірок $F_{\phi p}$ пов'язані співвідношенням:

$$F_{\phi n} + F_{\phi p} = -E_g \rightarrow F_{\phi p} = -E_g - F_{\phi n}, \quad (3.23)$$

У рівноважному стані рівень Фермі єдиний як для зони провідності, так і для валентної зони $F_{\phi n} = F_{\phi}$. Для функції розподілу дірок у валентній зоні маємо:

$$f_p(E) = (\exp(\frac{E - E_g - F_{\phi}}{k_B T}) + 1)^{-1}, \quad (3.24)$$

або за умови $f_p(E) \ll 1$:

$$f_p(E) = \exp(-\frac{E - E_g - F_{\phi}}{k_B T}). \quad (3.24a)$$

Для визначення концентрації дірок використовуємо формулу, аналогічну (3.16):

$$p = \int_0^{\infty} \frac{4\pi}{h^3} (2m_{dp}^*)^{3/2} E^{1/2} (\exp(\frac{E - E_g - F_{\phi}}{k_B T}) + 1)^{-1} dE. \quad (3.25)$$

Після розрахунків подібних тим, які були застосовані для визначення концентрації електронів (3.21), для дірок отримаємо:

$$p = 2(2\pi m_{dp}^* k_B T / h^2)^{3/2} \exp(\frac{-E_g - F_{\phi}}{k_B T}) = N_v \exp(\frac{-E_g - F_{\phi}}{k_B T}). \quad (3.26)$$

Передекспонентний множник N_v – ефективна щільність енергетичних станів валентної зони у діапазоні енергій $-E_g < E \leq k_B T$.

Для практичних розрахунків значення N_c і N_v деяких напівпровідників наведені у таблиці 3.2.

Таблиця 3.2

Матеріал	Si	Ge	GaAs	InSb
$N_c, \text{м}^{-3}$	$2,8 \cdot 10^{25} (T/300)^{3/2}$	$1,0 \cdot 10^{25} (T/300)^{3/2}$	$0,2 \cdot 10^{24} (T/300)^{3/2}$	$3,7 \cdot 10^{22} (T/300)^{3/2}$
$N_v, \text{м}^{-3}$	$1,0 \cdot 10^{25} (T/300)^{3/2}$	$0,6 \cdot 10^{25} (T/300)^{3/2}$	$7 \cdot 10^{24} (T/300)^{3/2}$	$1,9 \cdot 10^{24} (T/300)^{3/2}$

Власний напівпровідник. Кожен збуджений валентний електрон, який переходить у зону провідності, залишає після себе дірку у енергетичному спектрі валентної зони (рис.3.1, а). Дірка, як квазічастинка із позитивним зарядом, стає вільним носієм заряду у межах валентної зони. Електронейтральність напівпровідника при цьому не порушується, що можна записати такою умовою:

$$n = p = n_i, \quad (3.27)$$

де n_i – власна концентрація (індекс i від англ. *intrinsic* – внутрішній, власний).

Підставляємо (3.21) і (3.26) у (3.27), логарифмуємо ліву і праву частини, і для рівня Фермі у власному напівпровіднику $F_{\phi i}$ маємо:

$$F_{\phi i} = -\frac{E_g}{2} + \frac{1}{2} k_B T \ln \frac{N_v}{N_c}. \quad (3.28)$$

Аналіз формули (3.28):

1) $T = 0$, рівень Фермі посередині забороненої зони: $F_{\phi i} = -E_g / 2$;

2) $T > 0$, $N_v = N_c$, рівень Фермі посередині забороненої зони: $F_{\phi i} = -E_g / 2$

і від температури не залежить;

3) $N_v \neq N_c$ другий доданок, змінюючись прямо пропорційно температурі, буде зсувати рівень Фермі до зони, у якій ефективна щільність станів менша;

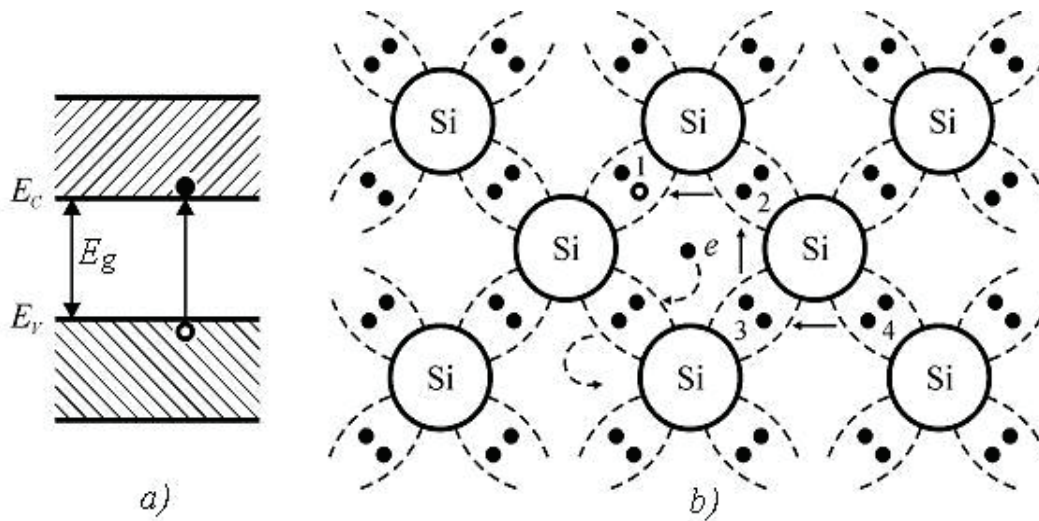


Рис.3.1. Зонна діаграма (a) та плоска модель (b) кристалічної ґратки власного напівпровідника (кремнію)

4) якщо $\ln N_v / N_c \approx 1$, як це має місце для Si і Ge (див. табл. 3.2), значення другого доданку $\approx 0,013$ eV $\ll E_g$ і для цих напівпровідників температурним зсувом рівня Фермі можна нехтувати;

5) для GaAs $\ln N_v / N_c \approx 3,7$, значення другого доданку $\approx 0,05$ eV але за рахунок значно більшої E_g і у цьому випадку температурною залежністю рівня Фермі можна нехтувати;

6) для InSb другий доданок становить + 0,08 eV, а за даними табл. 2.1 $-E_g / 2 = -0,085$ eV, тобто рівень Фермі практично співпадає із дном зони провідності при 300 K і, таким чином, InSb – *власний, частково вироджений напівпровідник при кімнатних температурах*. Цьому сприяє також зменшення ширини забороненої зі збільшенням температури.

Знайдемо концентрацію n_i , для чого підставимо (3.28) у (3.21) або (3.26):

$$n_i = (N_c N_v)^{1/2} \exp\left(-\frac{E_g}{2k_B T}\right). \quad (3.29)$$

Логарифмуємо (3.29):

$$\ln n_i = \ln(N_c N_v)^{1/2} - \frac{E_g}{2k_B} \cdot \frac{1}{T}. \quad (3.30)$$

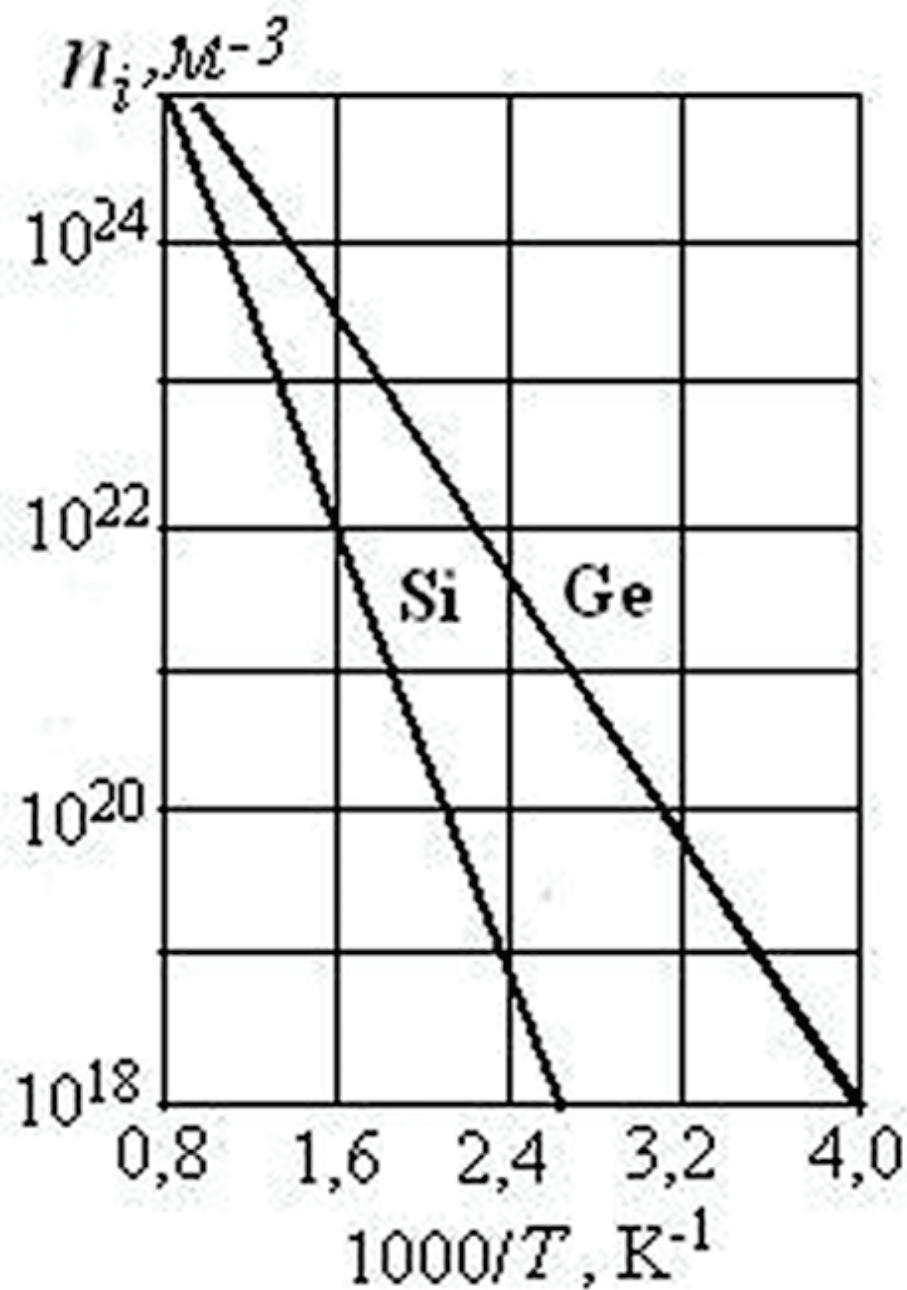


Рис.3.2. Температурна залежність власної концентрації у кремнію та германію

У координатах $\ln n_i = f\left(\frac{1}{T}\right)$ (3.30) – рівняння прямої лінії із тангенсом

кута нахилу:

$$\operatorname{tg} \alpha = -E_g / 2k_B. \quad (3.31)$$

Експериментальна залежність $n_i = f(\frac{1}{T})$, побудована у вище вказаних координатах, дає змогу визначити E_g . Приклад такої залежності поданий на рис.3.2 для кремнію і германію.

Знайдемо добуток np , перемноживши (3.21) і (3.26):

$$np = N_c N_v \exp(-\frac{E_g}{k_B T}) = n_i^2. \quad (3.32)$$

Формулу (3.32) слід розуміти таким чином: *при будь-якому співвідношенні рівноважних концентрацій електронів і дірок їх добуток завжди дорівнює n_i^2 (закон діючих мас)*.

Для прикладу, у табл.3.3 наведені значення ширини забороненої зони, власної концентрації та питомого опору при 300 К для деяких напівпровідників.

Таблиця 3.3

Напів - провідник	Ge	Si	GaAs
E_g , еВ	0,66	1,12	1,42
n_i , м ⁻³	$2,4 \cdot 10^{19}$	$1,4 \cdot 10^{16}$	$1,8 \cdot 10^{12}$
ρ_i , Ом·м	$47 \cdot 10^2$	$2,3 \cdot 10^7$	10^{10}

Леговані (домішкові) напівпровідники. Власні напівпровідники мають обмежене застосування в напівпровідниковій електроніці через те, що вони біполярні (3.27) і не мають певного типу провідності, крім того, у них сильна експонентна залежність концентрації носіїв заряду від температури (3.29).

Введення домішок донорного або акцепторного типу, які *заміщають* частину основних атомів, дозволяє вирішити такі проблеми:

1) отримати певний тип носіїв заряду – електрони або дірки;

2) отримати необхідну концентрацію носіїв заряду;

3) отримати незалежну від температури концентрацію носіїв заряду у діапазоні температур від -60 °С до $+100$ °С (так званий *технічний діапазон температур*, у якому працюють більшість напівпровідникових приладів і інтегральних мікросхем).

Для вирішення цих проблем необхідно мати джерело електронів або дірок, незалежне від власної концентрації. У напівпровідники IV групи вводять атоми, валентність яких відрізняється на одиницю, тобто атоми елементів III або V групи. У напівпровідникові сполуки типу $A^{III}B^V$ вводять атоми елементів II або VI групи. Вказані напівпровідники мають ковалентний або ковалентно – іонний тип зв'язку, тобто чотири зв'язуючі орбіталі по 2 електрони на кожній.

У атомів, валентність яких на 1 менше, не вистачає одного електрона для повноцінного зв'язку і вони доповнюють свої орбіталі шляхом локалізації на них електронів із валентної зони, перетворюючись при цьому на *негативно заряджені іони* (рис.3.3, *b*). Локалізація вимагає певних енергетичних витрат і реалізується шляхом термоіонізації валентних зв'язків (на рис.3.3, *a*,1). Таким чином, загальне число валентних електронів зменшилось і у енергетичному

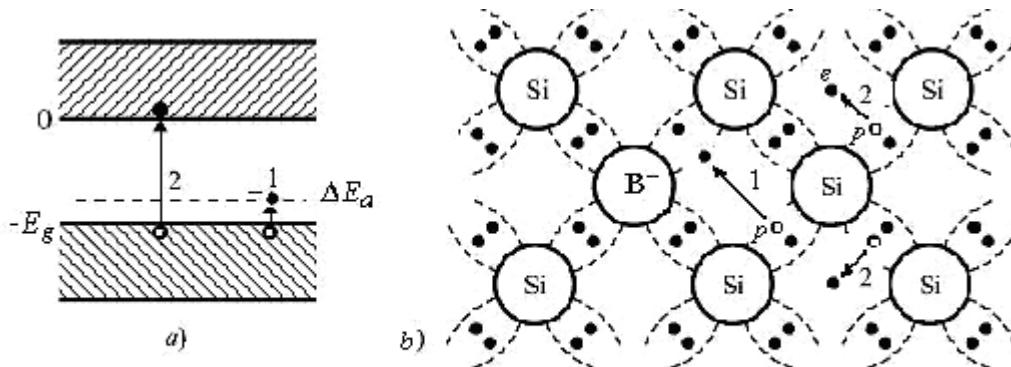


Рис. 3.3. Зонна діаграма (*a*) та спрощена модель (*b*) кристалічної ґратки напівпровідника *p* - типу (кремнію, легованого бором)

спектрі валентної зони з'являються *незаповнені енергетичні стани*, тобто *дірки* і їх стільки, скільки введено атомів домішки. Такі домішки називають *акцепторами* (від англійського *acceptor* – приймаючий на себе). Найчастіше

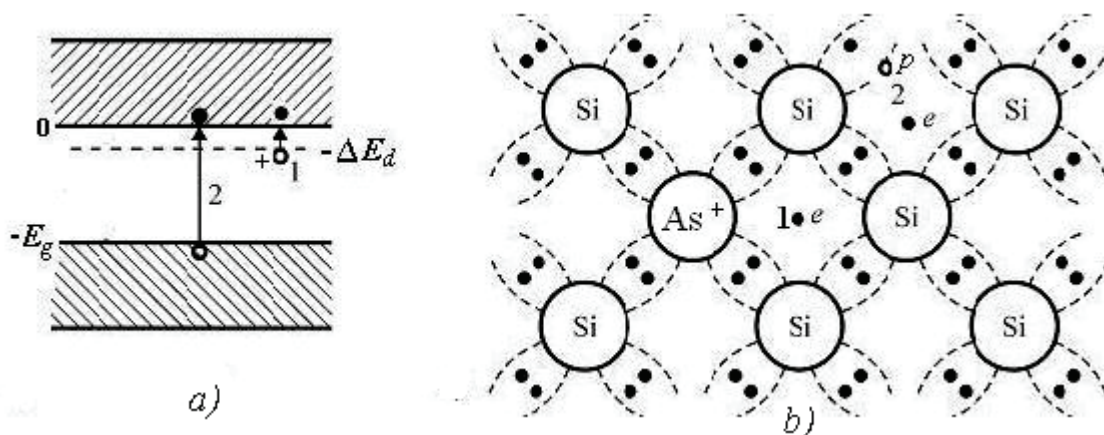


Рис. 3.4. Зонна діаграма (*a*) та спрощена модель (*b*) кристалічної ґратки напівпровідника *n* – типу (кремнію, легованого арсеном)

акцептором для Si є бор, для Ge – галій, індій, для GaAs – цинк.

У атомів, валентність яких на 1 більше, з'являються надлишкові електрони, які не приймають участі у формуванні валентних зв'язків. Оскільки такі атоми у структурі напівпровідника оточені чотирма зв'язуючими орбітами, зв'язок надлишкових електронів з ядрами суттєво послаблюється, що полегшує їх звільнення у зону провідності шляхом термоіонізації (рис.3.4, (a), 1). Таким чином утворюються вільні електрони і *позитивно заряджені іони* (рис.3.4, (b)1) і їх стільки, скільки введено атомів домішки. Такі домішки називають донорами (від англійського *donor* – той, що віддає). Найчастіше донорами для Si є фосфор, стибій, арсен, для Ge – стибій, арсен, для GaAs – телур.

Рівняння електронейтральності при наявності донорів і акцепторів можна записати такою умовою:

$$n + n_d + N_a = p + n_a + N_d, \quad (3.33)$$

де N_a, N_d – повні концентрації акцепторів і донорів, n_a, n_d – концентрації не іонізованих акцепторів і донорів, n, p – концентрації електронів і дірок.

Фізичний зміст умови (3.33) такий: у стані термодинамічної рівноваги повна кількість негативно заряджених частинок повинна дорівнювати повній кількості позитивно заряджених.

Розглянемо донорний напівпровідник в умовах, коли власною концентрацією можна нехтувати і акцептори відсутні ($N_a = n_a = p = 0$):

$$n = N_d - n_d. \quad (3.34)$$

Позначимо енергетичний рівень донорів у забороненій зоні ΔE_d (рис.3.4, a). Якщо донори не взаємодіють між собою, це значення енергії буде однаковим для усіх донорів. Енергетичний стан ΔE_d може займати лише один електрон, незалежно від напрямку спіна, тобто принцип заборони Паулі у цьому випадку не діє. Статистична вага одноелектронних донорних станів удвічі більша порівняно з двоелектронними енергетичними станами у зоні провідності. Розподіл електронів по донорним рівням має такий вигляд:

$$n_d = N_d \left(\frac{1}{2} \exp(-\Delta E_d - F_\phi) / k_B T + 1 \right)^{-1}. \quad (3.35)$$

Із (3.21) знаходимо:

$$\exp(F_\phi / k_B T) = n / N_c. \quad (3.36)$$

Підставляємо (3.35), (3.36) у (3.34) і отримуємо квадратне рівняння відносно $\exp(F_\phi / k_B T)$. Із розв'язку цього рівняння шляхом логарифмування $\exp(F_\phi / k_B T)$ знаходимо положення рівня Фермі у донорному напівпровіднику:

$$F_\phi = k_B T \ln \left\{ \frac{1}{4} \exp(-\Delta E_d / k_B T) \left[\left(1 + \frac{8N_d}{N_c} e^{(\Delta E_d / k_B T)} \right)^{1/2} - 1 \right] \right\}. \quad (3.37)$$

Формула (3.37) занадто складна для аналізу, тому спочатку вибираємо діапазон гранично низьких температур за критерієм $\exp(\Delta E_d / k_B T) \gg N_c / 8N_d$ (у кількісному значенні це діапазон $0 < T < 80 \text{ K}$) і нехтуємо одиницями у квадратних дужках:

$$F_{\phi n1} = -\Delta E_d / 2 + (k_B T / 2) \ln(N_d / 2N_c) \quad (3.38)$$

Аналіз температурної залежності рівня Фермі. У формулу (3.38) температура входить явно і неявно через $N_c(T) \sim T^{3/2}$ (див. (3.21)). За умови $T = 0$ рівень Фермі посередині між дном зони провідності і рівнем донорів $F_{\phi n1} = -\Delta E_d / 2$. Зі збільшенням температури зростає $N_c(T)$, але поки $2N_c \ll N_d$, другий доданок зростає, залишаючись > 0 , і рівень Фермі наближається до дна зони провідності (рис.3.5,(1)).

У міру подальшого зростання температури $F_{\phi n1}$ проходить на мінімальній відстані від дна зони провідності, але $N_c(T)$ продовжує невпинно зростати і другий доданок починає зменшуватися. Коли досягається рівність $2N_c = N_d$ рівень Фермі знову проходить $-\Delta E_d / 2$ і продовжує опускатися (рис.3.5,(2)). По досягненні значення $F_{\phi n1} = -\Delta E_d$ концентрація неіонізованих донорів буде, відповідно до (3.35), $n_d = 2N_d / 3$.

Можливість суміщення рівня Фермі із дном зони провідності і входу у зону має принципове значення. Справа у тому, що при $F_{\phi n1} \geq 0$ використовувати статистику Максвела – Больцмана (3.20а) недопустимо. Розрахуємо критичне значення концентрації донорів N_d^{kp} для того, щоб $F_{\phi n1} = 0$. Із умови екстремуму

$\frac{dF_{\phi n1}}{dT} = 0$ знаходимо T_{\min} і $N_c(T_{\min})$. Підставляємо ці значення у (3.38) і

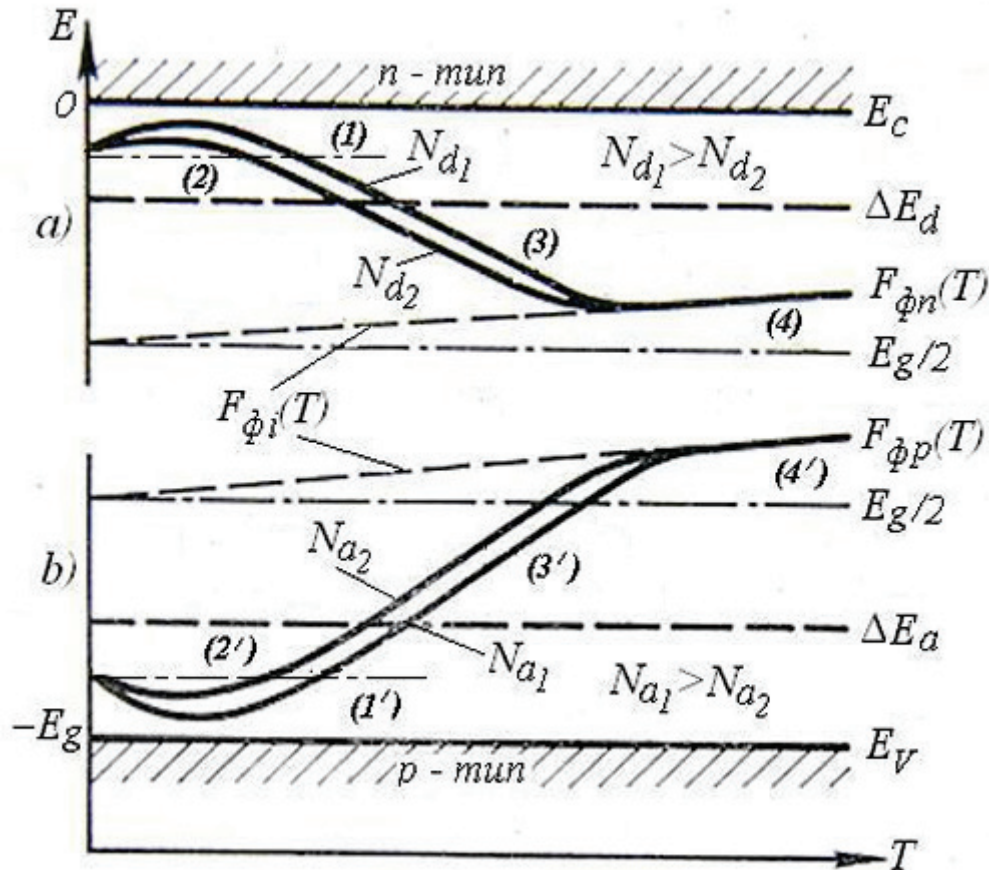


Рис.3.5. Температурні еволюції рівня Фермі у невідроджених: а) електронних, б)діркових напівпровідниках

обраховуємо N_d^{kp} по типовим для *Ge* і *Si* значенням $m_{dn}^* \approx (0,3...1)m_e$ та $\Delta E_d \approx 0,01...0,03$ eV. Результати розрахунку N_d^{kp} такі:

$$2 \cdot 10^{25} \leq N_d^{kp} \leq 1,5 \cdot 10^{26} \text{ м}^{-3}. \quad (3.38a)$$

Розрахунок показав, що N_d^{kp} співрозмірне із N_c (таб.3.2) і умова використання статистики Максвела – Больцмана $f_n(E) \ll 1$ не виконується. Результат розрахунку (3.38а) буде використаний для аналізу виродженого стану електронного газу.

Продовжимо аналіз (3.38). Знаходимо концентрацію електронів у цьому діапазоні температур, підставивши (3.38) у (3.36),:

$$n_1 = \left(\frac{N_c N_d}{2} \right)^{1/2} \exp(-\Delta E_d / 2k_B T). \quad (3.39)$$

Відповідно до (3.39), із збільшенням температури концентрація електронів монотонно зростає по складному експонентному закону. Діапазон температур, у якому діє (3.39), називається областю термоіонізації домішок або домішкової провідності (рис.3.6). Логарифмуємо (3.39):

$$\ln n_1 = \ln \left(\frac{N_c N_d}{2} \right)^{1/2} - \frac{\Delta E_d}{k_B} \cdot \frac{1}{T}. \quad (3.40)$$

У координатах $\ln n_1 = f\left(\frac{1}{T}\right)$ маємо рівняння прямої лінії із тангенсом кута нахилу:

$$\operatorname{tg} \beta = -\Delta E_d / 2k_B. \quad (3.41)$$

Експериментальна залежність $n_1 = f\left(\frac{1}{T}\right)$, побудована у вище вказаних координатах, дає змогу визначити глибину залягання донорів ΔE_d (рис.3.6).

Повернемося до аналізу (3.37) у діапазоні підвищених низьких температур ($T > 80 \text{ K}$) за критерієм $\exp(\Delta E_d / k_B T) \ll N_c / 8N_d$. Така нерівність означає, що у підкореновому виразі $\frac{8N_d}{N_c} \exp(\Delta E_d / k_B T) = x \ll 1$ і можна розкласти корінь у

швидко згасаючий ряд: $(1+x)^{1/2} \approx 1 + x/2 + \dots$ Врахувавши це, для рівня Фермі

$$\text{отримаємо:} \quad F_{\phi n2} \approx -k_B T \ln \frac{n}{N_c} \approx -k_B T \ln \frac{N_d}{N_c}. \quad (3.42)$$

Рівень Фермі $F_{\phi 2}$ опустився нижче $-\Delta E_d$ і при зростанні температури наближається до середини забороненої зони (рис. (3.5,(3))).

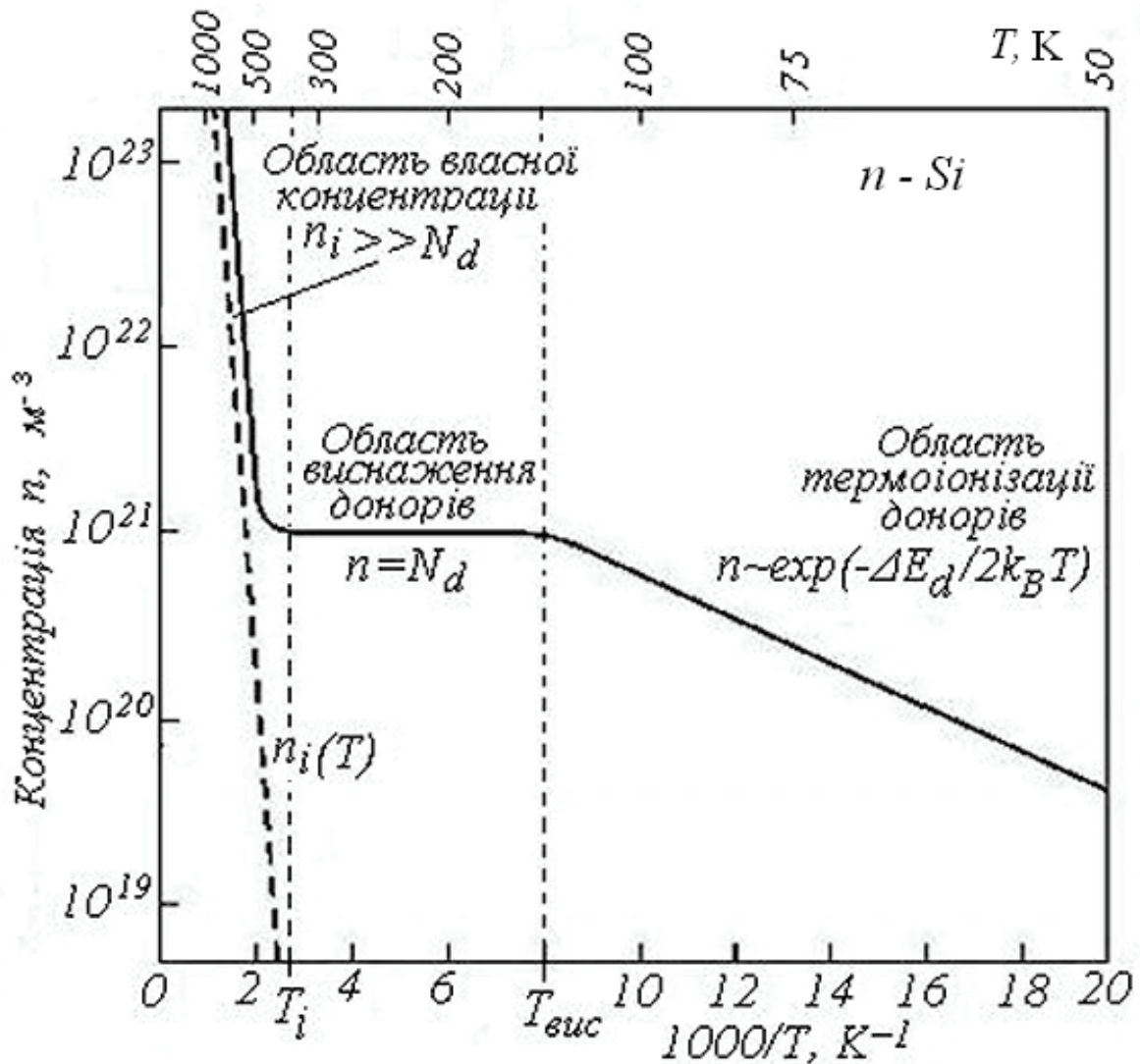


Рис.3.6. Залежність концентрації носіїв заряду у не вироджених напівпровідниках від температури. Кремній легований фосфором

Для концентрації електронів у діапазоні підвищених низьких температур маємо:

$$n_2 \approx N_d. \quad (3.43)$$

Фізичний зміст (3.43) полягає у тому, що донори повністю іонізовані, власна концентрація хоча і зростає з температурою, але умова $n_i(T) \ll N_d$ все

ще виконується. Діапазон температур, у якому діє (3.43), називається областю виснаження домішок (рис. 3.6).

Температуру переходу від області термоіонізації до виснаження домішок $T_{вис}$ можна оцінити із умови:

$$F_{\phi n2} = -k_B T_{вис} \ln \frac{N_d}{N_c} = -\Delta E_d, \quad (3.44)$$

звідки маємо:

$$T_{вис} = \Delta E_d / k_B \ln(N_c / N_d) \quad (3.45).$$

Наслідками (3.45) є такі закономірності: $T_{вис}$ тим менше, чим менші глибина залягання ΔE_d і концентрація донорів N_d .

Верхня температурна межа області виснаження визначається за умови, що власна концентрація стає співрозмірною із концентрацією донорів, а рівень Фермі $F_{\phi n2}$ суміщається із рівнем Фермі власного напівпровідника $F_{\phi i}$ у електронному напівпровіднику. Прирівнявши (3.42) до (3.28) для верхньої температурної межі області виснаження T_i отримаємо:

$$T_i = E_g / k_B \ln(N_c N_v / N_d^2). \quad (3.45a)$$

При $T > T_i$ у електронному напівпровіднику домінує власна концентрація, яка визначається (3.29) або (3.30). Наслідками (3.45a) є такі закономірності: T_i тим більша, чим більші E_g і N_d . Для прикладу, у таблиці 3.4 наведені значення $T_{вис}$ і T_i для Ge легованого стибієм у залежності від рівня легування.

Аналіз температурних залежностей рівня Фермі та концентрації у діркових напівпровідниках принципових відмінностей не має. Замість (3.34), початковим рівнянням електронейтральності буде:

$$p = N_a - n_a. \quad (3.46)$$

Подальший аналіз іде по тій же програмі, що і для електронного. Необхідно врахувати початок відліку і напрямок енергетичної вісі для

валентної зони (3.24), (3.24a). Функція розподілу електронів по акцепторним рівням:

$$n_a = N(2\exp(\Delta E_a - F_{\phi})/k_B T + 1)^{-1}. \quad (3.47)$$

На рис.3.5, *b*) показані температурні еволюції рівня Фермі $F_{\phi p}(T)$ для діркового напівпровідника. За винятком області власної провідності (рис.3.5, *b*)(4')), маємо дзеркальне відображення еволюцій $F_{\phi n}(T)$. Залежність концентрації дірок від температури має такий самий вигляд, як і електронів (рис.3.6).

Таблиця 3.4

$N_d, \text{м}^{-3}$	10^{20}	10^{21}	10^{22}	10^{23}	10^{24}
$T_{\text{вис}}, \text{К}$	15	-	30	-	80
$T_i, ^\circ\text{C}$	90	120	160	200	500

Вироджені напівпровідники. Відповідно до (3.20), формальною умовою виродженого стану електронного газу є розташування рівня Фермі у зоні провідності вище дна зони не менше ніж на $5k_B T$. З фізичної точки зору це означає, що концентрація електронів перевищує ефективну щільність станів, електронні переходи відбуваються лише у межах зони провідності. Енергетичний розподіл визначається статистикою Фермі – Дірака (3.15) і принципом Паулі.

Раніше вже розглядався випадок *температурного виродження* у деяких власних вузькодонних напівпровідниках. Окрім того, для легованих напівпровідників була проведена оцінка критичної концентрації донорів N_d^{kp} ,

за умови, що $F_{\phi n1}$ (3.38) у процесі температурної еволюції досягне дна зони провідності $F_{\phi 1} = 0$. Виявилось, що $N_d^{kp} \approx N_c$ (3.38a), тобто усі енергетичні стани N_c мають бути зайнятими електронами.

При розрахунках N_d^{kp} не враховані зміни у енергетичному спектрі електронів, які відбуваються починаючи із $N_d > 10^{24} \text{ м}^{-3}$. При таких концентраціях середня відстань між іонізованими домішками стає меншою подвоєного іонного радіуса домішок і починається їх взаємодія. Іонний радіус донорів (радіус потенціального поля) у Ge та Si становить $\approx 30...35$ ангстрем (близько 6 постійних структури). Розрахуємо концентрацію домішок по іонному радіусу N_d^{ion} , при якій середня відстань між ними становитиме подвоєний іонний радіус:

$$\frac{1}{2}(N_d^{ion})^{-1/3} = 32 \cdot 10^{-10} \text{ м} \rightarrow N_d^{ion} \approx 4 \cdot 10^{24} \text{ м}^{-3}. \quad (3.48)$$

Із порівняння результатів розрахунків (3.38a) і (3.48) можна зробити такий висновок: N_d^{ion} на 1...2 порядки менше N_d^{kp} і взаємодія донорів почнеться раніше, ніж рівень Фермі наблизиться до дна зони провідності у процесі температурної еволюції.

Основним наслідком взаємодії донорів є утворення домішкової зони, енергетичний діапазон якої збільшується при зменшенні відстані між донорами, є перекриття домішкової зони із основною зоною.

Перекриттю зон сприяє і те, що глибина залягання домішок становить лише 10 – 20 меВ. Зовнішньою ознакою цих змін є зменшення і навіть занулення енергії термоіонізації домішок $E_{ion} = 0$ при певних значеннях N_d або N_a (рис.3.7). Домішки стають іонізованими починаючи з $T \approx 0$, тобто без участі теплової енергії. Позначимо концентрацію електронів і рівень Фермі у повністю виродженому напівпровіднику n_0^+ , $F_{\phi n}^+$. Підставивши (3.20) у (3.17) для концентрації електронів і рівня Фермі отримаємо:

$$n_0^+ = \frac{8\pi}{3} \left(\frac{2m_{dn}^*}{h^2} \right)^{3/2} F_{\phi n}^{3/2}, \quad F_{\phi n}^+ = \frac{h^2}{2m_{dn}^*} \left(\frac{3n_0^+}{8\pi} \right)^{2/3}. \quad (3.49)$$

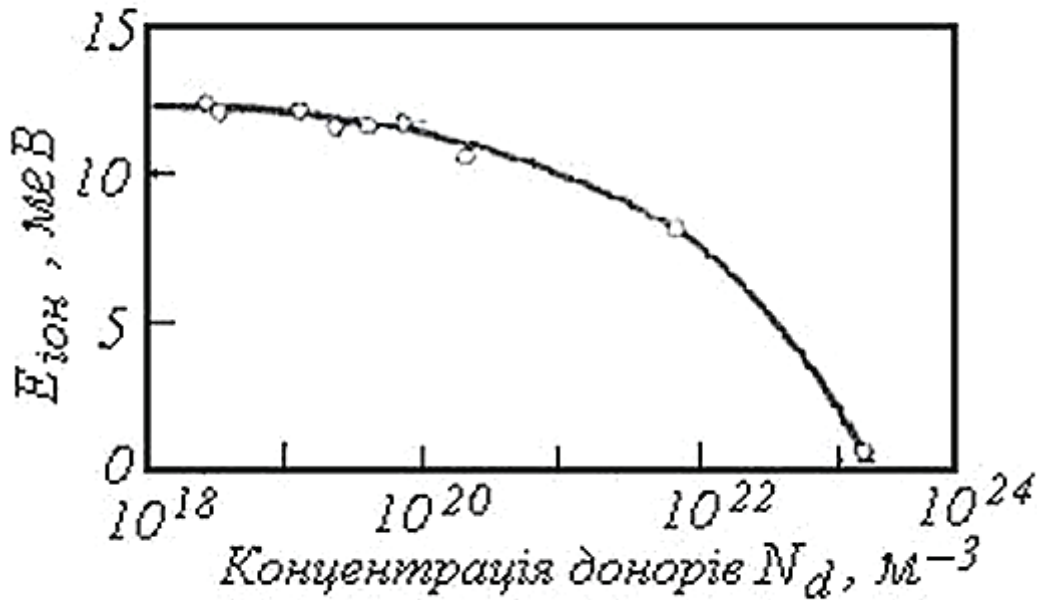


Рис.3.7. Залежність енергії іонізації донорів від їх концентрації. Германій легований арсеном

Для повністю вироджених діркових напівпровідників рівень Фермі розташований у валентній зоні на глибині $F_{\phi p}^+ \leq -(\Delta E_0 + 5k_B T)$. Формула для концентрації дірок p_0^+ аналогічна (3.49):

$$p_0^+ = \frac{8\pi}{3} \left(\frac{2m_{dp}^*}{h^2} \right)^{3/2} F_{\phi p}^{3/2}. \quad (3.49a)$$

Висновки: – концентрація носіїв заряду у вироджених напівпровідниках висока $n^+ = N_d \geq 10^{24} \text{м}^{-3}$, $p^+ = N_a \geq 10^{24} \text{м}^{-3}$ і не залежить від температури, починаючи з $T \geq 0$;

– рівень Фермі F_{ϕ}^+ розташований у межах дозволеної зони, починаючи з $T = 0$ і лежить між останнім заповненим енергетичним рівнем і першим вільним. При $T > 0$ положення F_{ϕ}^+ практично не змінюється. У цьому

відношенні вироджені напівпровідники подібні до металів і їх іноді називають напівметалами;

– для широкозонних легованих напівпровідників ($E_g \geq 0,5 \text{ eV}$) характерне концентраційне виродження. Для вузькозонних власних напівпровідників ($E_g < 0,3 \text{ eV}$) із значною (більше ніж на порядок) різницею ефективних мас електронів і дірок характерне температурне виродження.

Компенсовані напівпровідники. Матеріали, які вважаються чистими від електрично активних домішок, зазвичай мають залишковий фон донорів і акцепторів. При виготовленні напівпровідникових приладів та інтегральних мікросхем маємо справу із матеріалами, у яких необхідна одночасна присутність донорів і акцепторів. Розрізняють такі випадки по співвідношенню кількості донорів і акцепторів:

1. $N_d \gg N_a$ – електронний напівпровідник;
2. $N_d > N_a$ – частково компенсований електронний напівпровідник;
3. $N_d = N_a$ – скомпенсований напівпровідник;
4. $N_a > N_d$ – частково компенсований дірковий напівпровідник;
5. $N_a \gg N_d$ – дірковий напівпровідник.

Із фізичної точки зору явище компенсації домішок полягає у тому, що частина вільних електронів, які виникли при термоіонізації донорів, захоплюється акцепторами і вони вже не можуть створювати дірки у валентній зоні (випадок 2). Результатом цього є те, що кількість активних донорів зменшиться і становитиме $N_d - N_a$. Аналогічна ситуація має місце у випадку 4 – кількість активних акцепторів зменшиться і становитиме $N_a - N_d$.

Цікавим є випадок 3 – повної взаємної компенсації донорів і акцепторів. У цьому випадку у напівпровіднику є лише власна концентрація носіїв заряду n_i .

Відрізнити скомпенсований напівпровідник від власного можна по мобільності носіїв заряду. Донори і акцептори не створюють носіїв заряду, але як позитивно і негативно заряджені іони суттєво зменшують мобільність внаслідок іонного розсіювання. Зазвичай на практиці порівнюють питомий опір скомпенсованого напівпровідника і істинно власного: якщо концентрація у обох випадках n_i , а питомий опір виявився більшим, ніж у власного, значить має місце компенсація домішок.

При розрахунках концентрації носіїв заряду для частково компенсованих напівпровідників у температурному діапазоні області виснаження використовують такі формули:

$$\text{для електронів } n = N_d - N_a, \text{ для дірок } p = N_a - N_d. \quad (3.50)$$

3.3. Динамічні властивості носіїв заряду

Швидкість і прискорення. Проекції вектора групової швидкості електрона:

$$v_x = \frac{\partial E}{\partial k_x}, v_y = \frac{\partial E}{\partial k_y}, v_z = \frac{\partial E}{\partial k_z}.$$

У компактному записі із урахуванням $\bar{p} = \hbar \bar{k}$:

$$\bar{v} = \frac{1}{\hbar} \nabla_{\bar{k}} E = \nabla_{\bar{p}} E. \quad (3.51)$$

Швидкість електрона в кристалі – це градієнт його енергії у просторі квазіімпульсів або хвильових векторів.

В екстремальних точках $E(\bar{k})$ – дно зони провідності, стеля валентної зони – групова швидкість (3.51) $\bar{v} = 0$, електрон не може без зовнішнього впливу вийти за межі дозволеної зони.

Для прискорення необхідна певна зовнішня сила: $\bar{F}_{\text{зов}} = \frac{d\bar{p}(t)}{dt} = \frac{1}{\hbar} \frac{d\bar{k}(t)}{dt}$.

Визначимо прискорення із урахуванням того, що енергія неявна функція часу $E(\bar{k}(t))$:

$$\frac{d\bar{\omega}}{dt} = \frac{1}{\hbar^2} (\nabla_{\bar{k}} \nabla_{\bar{k}} E) \frac{d\bar{k}(t)}{dt} = \frac{1}{\hbar^2} (\nabla_{\bar{k}} \nabla_{\bar{k}} E) \bar{F}. \quad (3.51a)$$

Прискорення прямо пропорційне силі – формулювання другого закону Ньютона. Для порівняння запишемо закон Ньютона для електрона у вакуумі:

$$\frac{d\bar{\omega}}{dt} = \frac{1}{m_e} \bar{F}. \quad (3.52)$$

Коефіцієнт пропорційності між прискоренням і силою має розмірність оберненої маси. Розшифруємо подвійний градієнт енергії:

$$\nabla_{\bar{k}} \nabla_{\bar{k}} E = \sum_i \frac{\partial^2 E}{\partial k_i^2}, \text{ де } i = x, y, z. \quad (3.52a)$$

Аналіз цієї формули вже зроблений (2.24) – це компоненти діагонального тензора 2 рангу. Повертаючись до прискорення електрона у кристалі, можемо записати:

$$\frac{d\bar{\omega}}{dt} = \frac{1}{m_{ii}^*} \bar{F}, \quad \frac{1}{m_{ii}^*} = \frac{1}{\hbar^2} (\nabla_{\bar{k}} \nabla_{\bar{k}} E). \quad (3.53)$$

Таким чином, отримали формулу для ефективної маси із рівняння руху (3.53), записаного у формі другого закону Ньютона,:

$$m_{ii}^* = \hbar^2 (\nabla_{\bar{k}} \nabla_{\bar{k}} E)^{-1}. \quad (3.54)$$

Введена із рівняння руху (3.53) ефективна маса еквівалентна “інертній масі” і вона співпадає із “гравітаційною масою”, введеною із закону дисперсії $E(\bar{k})$ (2.27), що цілком узгоджується із законами механіки.

Дрейфовий рух у електричному полі. Дрейф – спрямований рух зарядів під дією електричного поля. Середня швидкість цього руху називається *швидкістю дрейфу* \bar{v}_{dp} .

Сила електричного поля, яке має напруженість \bar{E} і діє на електрони і дірки:

$$\bar{F}_n = -e\bar{E}, \quad \bar{F}_p = e\bar{E}.$$

Під дією цієї сили електрони рухаються прискорено, збільшуючи свою кінетичну енергію. Однак час дії сили обмежений процесами розсіювання енергії при взаємодії з фононами, іонізованими домішками та іншими дефектами. Середній час вільного (прискореного) руху електронів τ називається **часом вільного пробігу** або **часом релаксації**, тому що за цей час після припинення дії сили електрони втрачають (розсіюють) отриману від поля енергію і повертаються у стан термодинамічної рівноваги.

Рівняння руху під дією електричного поля для електронів із скалярною ефективною масою має такий вигляд:

$$\frac{d\bar{v}_{dp}}{dt} = -\frac{e}{m^*}\bar{E}. \quad (3.55)$$

Розділивши змінні й інтегруючи в межах $0 \leq t \leq \tau$, для дрейфової швидкості отримаємо:

$$\bar{v}_{dp} = -\frac{e\tau}{m^*}\bar{E} = -\mu_n\bar{E}. \quad (3.56)$$

Знак „-” враховує те, що електрони рухаються проти поля.

Коефіцієнт пропорційності між дрейфовою швидкістю і напруженістю електричного поля $\mu = \frac{e\tau}{m^*}$ називають **мобільністю** або **рухливістю**. Фізичний зміст мобільності: середня швидкість дрейфу в електричному полі одиничної напруженості. Розмірність мобільності $[\mu] = \text{м}^2/\text{В}\cdot\text{с}$.

Мобільність визначається зонною структурою напівпровідника через ефективну масу m^* і процесами розсіювання через час релаксації τ .

Густина дрейфових потоків електронів $\overline{W}_{n\partial p}$ і дірок $\overline{W}_{p\partial p}$:

$$\overline{W}_{n\partial p} = n\overline{v}_{n\partial p} = -n\mu_n\overline{E}, \quad \overline{W}_{p\partial p} = p\overline{v}_{p\partial p} = p\mu_p\overline{E}, \quad (3.57)$$

де n і p – концентрації електронів і дірок, відповідно.

Помітимо, що дрейфові потоки електронів і дірок направлені зустрічно.

Довжина вільного пробігу – середній шлях, який проходить носій заряду за час вільного пробігу: $L = v\tau$, де $v = v_T + v_{\partial p}$ – повна швидкість – сума теплової і дрейфової швидкостей.

Енергія, отримана електронами від електричного поля, частково іде на дрейфовий транспорт електронного заряду через провідник, частково перетворюється в теплову енергію.

Дифузійний рух носіїв заряду. Дифузія – спрямований рух зарядів у полі градієнта концентрації або хімічного потенціалу. Щільність дифузійного потоку прямо пропорційна градієнтові концентрації:

$$\overline{W}_{ndif} = -D_n \nabla n, \quad \overline{W}_{pdif} = -D_p \nabla p, \quad (3.58)$$

де D_n, D_p – коефіцієнти дифузії, $\nabla n, \nabla p$ – градієнти концентрації електронів і дірок відповідно.

Коефіцієнт дифузії $D = \frac{\overline{W}_{dif}}{\nabla n}$ – густина дифузійного потоку, яка відповідає одиничному градієнту концентрації або число носіїв, що перетинають одиницю площі, перпендикулярну потоку, за одиницю часу і за умови, що різниця концентрацій дорівнює одному носієві на одиниці довжини шляху. Розмірність коефіцієнта дифузії $[D] = \text{м}^2/\text{с}$.

Градієнт концентрації виникає у разі нерівномірного легування електрично активними домішками, нерівномірної фотогенерації носіїв заряду та інжекції носіїв заряду із джерела на поверхні напівпровідника. Інший випадок появи

градієнта концентрації – *неоднакова* по об'єму середня *теплова енергія* носіїв заряду, як *наслідок градієнта температури*.

Градiєнт концентрації, градієнт температури – окремі випадки, коли виникає дифузія. Об'єднуючою умовою для дифузії є *градієнт хімічного потенціалу (рівня Фермі)* $\nabla F_\phi \neq 0$.

За фізичним змістом F_ϕ – *вільна енергія у розрахунку на один носій заряду*. Рівноважний стан електронної системи вимагає однакового хімічного потенціалу по всьому об'єму, тобто, $\nabla F_\phi = 0$. Вирівнювання F_ϕ іде за рахунок *дифузійного потоку* носіїв заряду, направлено *проти* ∇F_ϕ .

Помітимо, що дифузійні потоки електронів і дірок, на відміну від дрейфових, *направлені однаково – проти відповідних градієнтів концентрації*.

3.4. Дифузійно – дрейфові процеси.

Співвідношення Ейнштейна.

Зважаючи на те, що мова йде про дифузію заряджених частинок – електронів або дірок, нерівномірний їх розподіл зумовлює появу внутрішнього електричного поля, яке, у свою чергу, створює дрейфовий потік.

Розглянемо електронний провідник, подовжній розмір якого d по напрямку x , значно більший за поперечні розміри по y і z . У провіднику є нерівномірний розподіл електронів по довжині $\frac{dn(x)}{dx} \neq 0$.

Дифузійний потік електронів $\overline{W}_{ndиф} = -D_n \frac{dn(x)}{dx}$ створить певний

надлишок електронів на кінці провідника $n(d)$ і між точками $x=0$ і $x=d$ виникне електричне поле $\mathcal{E}_x = -d\phi(x)/dx$, направлене від $x=0$ до $x=d$, і відповідна йому різниця потенціалів $\phi(x)$.

Розподіл електронів у полі цього потенціалу визначається формулою Больцмана $n(x) = n(0) \exp(-\frac{e\varphi(x)}{k_B T})$.

Дрейфовий потік у цих умовах:

$$W_{\text{др}} = -n(x)\mu_n \mathcal{E}_x = -n(x)\mu_n (-d\varphi(x) / dx). \quad (3.59)$$

Напруженість поля \mathcal{E}_x , зумовлена перерозподілом електронів, досягає такої величини, при якій дрейфовий потік *врівноважить* дифузійний, що є умовою *стаціонарної рівноваги*:

$$-n(x)\mu_n (-d\varphi(x) / dx) = -D_n (dn(x) / dx). \quad (3.60)$$

Після нескладних підрахунків із умови стаціонарної рівноваги (3.60) знайдемо:

$$\mu_n = \frac{k_B T}{e} D_n - \text{співвідношення Ейнштейна}. \quad (3.61)$$

Аналогічне співвідношення має місце і для дірок. Коефіцієнт $\frac{k_B T}{e} = \varphi_T$ називається *дифузійною різницею потенціалів*. Для 300 К він дорівнює $\varphi_T = 0,026$ В.

Для умови стаціонарної рівноваги у випадку дифузійно – дрейфових процесів вводять *повний або електрохімічний потенціал* $(F_\phi - e\varphi)$. Умовою стаціонарної рівноваги буде *відсутність градієнта електрохімічного потенціалу*:

$$\nabla(F_\phi - e\varphi) = 0. \quad (3.62)$$

Наслідком цієї умови є те, що *різниця потенціалів φ , яка виникає у провіднику, термодинамічно рівноважна і не може виконувати роботу у зовнішньому колі, оскільки лише підтримує рівновагу дифузійно – дрейфових процесів*.

Дифузія і дрейф змінюють градієнт хімічного потенціалу, але якщо підтримувати його ззовні незмінним (наприклад, підтримувати $n(0) = \text{const}$ за рахунок фотогенерації або $\nabla T = \text{const}$), енергія, необхідна для такої підтримки, буде трансформуватися у зовнішню *термодинамічно нерівноважну різницю потенціалів, яка може виконувати роботу* у зовнішньому колі.

Резюме по п. 3.3

1) Ефективна маса електрона введена із рівняння руху електрона в кристалі має такі ж властивості, як і ефективна маса, введена із закону дисперсії $E(\bar{k})$.

2) Під дією зовнішнього поля виникає дрейфовий рух електронів і дірок. Дрейфові потоки електронів і дірок направлені зустрічно.

3) Дрейфова швидкість прямо пропорційна напруженості електричного поля (така закономірність діє лише у слабких полях).

4) Кількісною характеристикою дрейфового руху є *мобільність* – середня дрейфова швидкість у полі одиничної напруженості, яка визначається енергетичною зонною структурою напівпровідника і механізмами розсіювання енергії носіїв заряду.

5) При наявності градієнта хімічного потенціалу виникає дифузійний потік носіїв заряду, щільність якого прямо пропорційна градієнту хімічного потенціалу. Дифузійні потоки електронів і дірок направлені однаково – проти градієнта хімічного потенціалу.

6) Кількісною характеристикою дифузійного руху є *коефіцієнт дифузії* – щільність дифузійного потоку при одиничному градієнті хімічного потенціалу.

7) Провідник із градієнтом хімічного потенціалу – термодинамічно не рівноважний. Дифузійний потік спрямований на вирівнювання хімічного потенціалу шляхом перерозподілу носіїв заряду. Такий перерозподіл заряджених частинок порушує нейтральність провідника, наслідком чого є *внутрішнє електричне поле*. Поле, у свою чергу, створює зустрічний дрейфовий потік, який врівноважує дифузійний. Таки чином встановлюється

стаціонарна рівновага в провіднику, умовою якої є відсутність градієнта електрохімічного потенціалу.

8) *Різниця потенціалів, породжена градієнтом хімічного потенціалу, досягає такої величини, яка лише підтримує стаціонарну рівновагу. Відповідно до закону збереження енергії, така різниця потенціалів не може виконувати роботу у зовнішньому колі.*

9) *У випадку, коли зовнішнє джерело енергії, незважаючи на вирівнюючу дію дифузії, підтримує незмінним градієнт рівня Фермі, виникає нерівноважна, надлишкова різниця потенціалів, яка може виконувати роботу у зовнішньому колі.*

4. Кінетичне рівняння Больцмана

4.1. Загальні положення

Фізика твердого тіла використовує три основні рівняння теоретичної фізики:

- рівняння Шредінгера для суто квантових систем;
- кінетичне рівняння Больцмана для не рівноважних систем;
- рівняння неперервності для дифузійно – дрейфових та генераційно – рекомбінаційних процесів.

Базовим є рівняння Шредінгера, кінетичне рівняння Больцмана – похідне і „працює” переважно у квазікласичному наближенні. У свою чергу, рівняння неперервності є окремим випадком кінетичного рівняння.

У кінетичному рівнянні Больцмана враховується як вплив зовнішніх силових полів, так і внутрішніх процесів на стан електронної системи твердого тіла. Розв’язком кінетичного рівняння є не рівноважна функція розподілу носіїв заряду у координатному та енергетичному просторах. Крім того, визначається також середній час переходу системи із одного рівноважного стану у інший рівноважний стан – *час релаксації* електронної системи.

Розрізняють такі можливі стани електронної системи твердого тіла.

Термодинамічна рівновага – однаковий по об’єму хімічний потенціал (рівень Фермі):

$$\nabla_{\vec{r}} F_{\phi} = \frac{dF_{\phi}}{dr} = 0. \quad (4.1)$$

Рівноважна функція розподілу електронів f_0 у цьому випадку залежить лише від повної енергії і температури:

$$\text{вироджений електронний газ } f_0(E, T) = \left(\exp\left(\frac{E - F_{\phi}}{k_B T}\right) + 1 \right)^{-1}, \quad (4.2)$$

$$\text{невироджений електронний газ } (f_0 \ll 1) \quad f_0(E, T) = \exp\left(-\frac{E - F_{\phi}}{k_B T}\right) \quad (4.3)$$

Умова $f_0 \ll 1$ означає, що електронних станів значно більше ніж електронів і знайти електрон у стані E при температури T малоймовірно.

У формулах (4.1), (4.2), (4.3) E – повна енергія у наближенні ефективної маси:

$$E = \frac{p^2}{2m^*} + U_{\text{вн}}(\vec{r}), \quad (4.4)$$

де p - квазіімпульс, $U_{\text{вн}}(\vec{r})$ – внутрішнє потенціальне поле дефектів, якому відповідає внутрішня сила:

$$\vec{F}_{\text{вн}} = -\nabla_{\vec{r}} U_{\text{вн}}(\vec{r}). \quad (4.5)$$

При наявності зовнішніх потенціальних полів і відповідних їм зовнішніх сил:

$$\vec{F}_{\text{зов}} = -\nabla_{\vec{r}} V_{\text{зов}}(\vec{r}) \quad (4.6)$$

електронна система виходить із стану термодинамічної рівноваги.

Стационарна рівновага – дія зовнішніх сил врівноважена внутрішніми процесами таким чином, що середні значення параметрів системи не залежать від часу.

Повна сила, яка діє на електрони, поділена на дві сили (ідея Больцмана):

$$\vec{F} = \vec{F}_{\text{вн}} + \vec{F}_{\text{зов}}. \quad (4.7)$$

У стані стаціонарної рівноваги повна сила:

$$\vec{F} = \vec{F}_{\text{вн}} + \vec{F}_{\text{зов}} = 0. \quad (4.8)$$

Перехід від одного стаціонарного стану до іншого при зміні зовнішніх сил відбувається протягом часу релаксації і у цей час система знаходиться у нерівноважному стані.

4.2. Нерівноважна функція розподілу

Введемо не рівноважну функцію розподілу електронів у координатному і імпульсному просторах на певний момент часу: $f(\vec{r}, \vec{p}, t)$ – число електронів у

одиниці об'єму координатного простору, одиниці об'єму імпульсного простору у момент часу t .

За фізичним змістом, $f(\bar{r}, \bar{p}, t)$ при $\bar{F}_{зov} \rightarrow 0$ повинна переходити у $f_0(E, T)$ і мати вигляд (4.2) або (4.3).

Запишемо повну похідну $f(\bar{r}, \bar{p}, t)$ по t , із урахуванням того, що $\bar{r}(t), \bar{p}(t)$:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \bar{r}} \frac{d\bar{r}}{dt} + \frac{\partial f}{\partial \bar{p}} \frac{d\bar{p}}{dt}. \quad (4.9)$$

Електронна система твердого тіла замкнена, відбувається лише перерозподіл електронів у координатному і імпульсному просторах без зміни їх загальної кількості, тому $\frac{df}{dt} = 0$. Введемо позначення: $\frac{d\bar{r}}{dt} = \bar{v}$ – швидкість; $\frac{d\bar{p}}{dt} = \bar{F} = \bar{F}_{вн} + \bar{F}_{зov}$ – сили; $\frac{\partial f}{\partial \bar{r}} = \nabla_{\bar{r}} f$, $\frac{\partial f}{\partial \bar{p}} = \nabla_{\bar{p}} f$ – градієнти у координатному та імпульсному просторах.

Із урахуванням усього:

$$-\frac{\partial f}{\partial t} = \bar{v} \nabla_{\bar{r}} f + \bar{F}_{зov} \nabla_{\bar{p}} f + \bar{F}_{вн} \nabla_{\bar{p}} f. \quad (4.10)$$

Отримано кінетичне рівняння Больцмана для нерівноважної функції розподілу. Запишемо (4.10) у компактному вигляді за допомогою позначень:

$$\begin{aligned} \bar{v} \nabla_{\bar{r}} f + \bar{F}_{зov} \nabla_{\bar{p}} f &= \left(\frac{\partial f}{\partial t} \right)_{зov}; \quad \bar{F}_{вн} \nabla_{\bar{p}} f = \left(\frac{\partial f}{\partial t} \right)_{розс}. \\ -\left(\frac{\partial f}{\partial t} \right) &= \left(\frac{\partial f}{\partial t} \right)_{зov} + \left(\frac{\partial f}{\partial t} \right)_{розс}. \end{aligned} \quad (4.11)$$

Фізичний зміст позначень: $\left(\frac{\partial f}{\partial t} \right)_{зov}$ – враховує дію зовнішніх факторів на розподіл електронів і при визначених значеннях цих факторів (градієнти концентрації і температури, електричне поле, сила Лоренця і т.п.) цю похідну завжди можна обрахувати; $\left(\frac{\partial f}{\partial t} \right)_{розс}$ – враховує процеси розсіяння, які

намагаються повернути розподіл електронів до рівноважного, залежить від природи дефектів і механізму взаємодії електронів з ними.

Стационарна рівновага означає, що $\frac{\partial f}{\partial t} = 0$ (f явно від часу не залежить) і

стационарне рівняння Больцмана має вигляд:

$$\left(\frac{\partial f}{\partial t}\right)_{\text{зов}} + \left(\frac{\partial f}{\partial t}\right)_{\text{розс}} = 0. \quad (4.12)$$

4.3. Імовірність пружного розсіювання

Дефекти структури створюють навколо себе неперіодичні потенціальні поля, які є основною причиною зміни квазіімпульсу у процесі руху електронів і розсіювання електронних хвиль. Це поле можна вважати сферично симетричним і обмеженим у просторі. Електрони, які у процесі руху потрапляють у поле розсіювання, змінюють напрям і модуль хвильового вектора.

У більшості випадків розсіювання можна вважати *пружним*. *Ідеально пружним* вважається розсіювання, при якому змінюється лише напрямок хвильового вектора при збереженні його модуля. У дійсності пружне розсіювання до певної міри не є ідеальним – при кожному зіткненні з дефектом змінюється і модуль квазіімпульсу, і енергія електрона. Якщо ці зміни у межах першого порядку малості, розсіювання вважається пружним. *Наслідком* незначних змін модуля квазіімпульсу і енергії при пружному розсіюванні *є необхідність великого числа зіткнень з дефектами* для втрати електронами нерівноважної енергії.

Для процесів розсіювання залежність нерівноважної функції від координат не суттєва, тому що час взаємодії окремого електрона з дефектом оцінюється у межах 10^{-16} с. За такий короткий час координати електрона практично не змінюються, зіткнення з дефектом можна розглядати як *пружний удар*. Основні події розгортаються у імпульсному просторі, де відбуваються квантові переходи у дискретному імпульсному спектрі.

При визначенні нерівноважної функції розподілу зазвичай використовують термін „...кількість електронів у одиниці об'єму імпульсного

простору...” (див. п.4.2), але ж імпульсний простір - це не фізичний простір, електронів там немає. Мова іде про кількість дискретних значень імпульсів, реалізованих вільними електронами.

Введемо $W(\bar{p}, \bar{p}')$ – імовірність квантового переходу $\bar{p} \rightarrow \bar{p}'$, віднесену до одиниці часу, іншими словами, *частоту переходів*. Оскільки квантові переходи мають обмінний характер, таким же чином визначаємо $W(\bar{p}', \bar{p})$ – частоту зворотних переходів $\bar{p}' \rightarrow \bar{p}$.

Виділимо у імпульсному просторі два елементарних об'єми $dV_{\bar{p}}$ і $dV_{\bar{p}'}$. Розглянемо обмін електронами між цими об'ємами $dV_{\bar{p}} \rightleftharpoons dV_{\bar{p}'}$. Число електронів у $dV_{\bar{p}}$, здатних перейти у $dV_{\bar{p}'}$: $f(\bar{p})dV_{\bar{p}}$, число вільних станів у $dV_{\bar{p}'}$, здатних прийняти ці електрони: $(1 - f(\bar{p}'))dV_{\bar{p}'}$. Аналогічні записи маємо для зворотних переходів: $f(\bar{p}')dV_{\bar{p}'}$, $(1 - f(\bar{p}))dV_{\bar{p}}$. Число переходів $\bar{p} \rightarrow \bar{p}'$ за час dt :

$$[f(\bar{p})W(\bar{p}, \bar{p}')(1 - f(\bar{p}'))]dV_{\bar{p}}dV_{\bar{p}'}dt. \quad (4.13)$$

Число переходів $\bar{p}' \rightarrow \bar{p}$ за час dt :

$$[f(\bar{p}')W(\bar{p}', \bar{p})(1 - f(\bar{p}))]dV_{\bar{p}'}dV_{\bar{p}}dt. \quad (4.14)$$

У стані термодинамічної рівноваги $f(\bar{r}, \bar{p}, t) \rightarrow f_0(E, T)$ і діє принцип детальної рівноваги Ейнштейна для обмінних квантових переходів: $W(\bar{p}, \bar{p}') = W(\bar{p}', \bar{p})$. У більшості практичних випадків відхилення від рівноважного стану незначне і можна пролонгувати дію принципу детальної рівноваги для таких систем:

$$W(\bar{p}, \bar{p}') \approx W(\bar{p}', \bar{p}). \quad (4.15)$$

Незначне відхилення від рівноважного стану реалізується у випадку пружного розсіяння, умовою якого є збереження модуля квазіімпульсу $p \approx p'$, змінюється лише напрямки векторів \bar{p} і \bar{p}' .

Допустимість такого наближення доведена практичним досвідом.

Визначимо різницю числа прямих і зворотних переходів за час dt із урахуванням (4.13), (4.14), (4.15):

$$dV_{\bar{p}} \{W(\bar{p}, \bar{p}') [f(\bar{p}') - f(\bar{p})]\} dV_{\bar{p}'} . \quad (4.16)$$

Зафіксуємо початкове значення імпульсу $\bar{p} = const$ і визначимо середнє число переходів із початкового стану \bar{p} в усі можливі кінцеві стани \bar{p}' внаслідок розсіювання:

$$\left(\frac{\partial f}{\partial t}\right)_{розс} = \int_{(\bar{p}')} \{W(\bar{p}, \bar{p}') [f(\bar{p}') - f(\bar{p})]\} dV_{\bar{p}'} . \quad (4.17)$$

Інтеграл (4.17) у теорії розсіювання отримав назву *інтеграл зіткнень*.

Запишемо кінетичне рівняння Больцмана у повному обсязі із урахуванням (4.11), (4.17):

$$-\frac{\partial f}{\partial t} = \bar{v} \nabla_{\bar{r}} f + \bar{F}_{зов} \nabla_{\bar{p}} f + \int_{(\bar{p}')} \{W(\bar{p}, \bar{p}') [f(\bar{p}') - f(\bar{p})]\} dV_{\bar{p}'} . \quad (4.18)$$

Висновок: *отримано інтегро – диференціальне рівняння для нерівноважної функції розподілу, яке не має розв'язку у загальному вигляді.*

4.4. Наближення часу релаксації

Скористаємося умовами пружного розсіювання і запишемо не рівноважну функцію у вигляді швидко згасаючого ряду:

$$f(t) \approx f_0 + f_1(t) + f_2(t) + \dots, \quad (4.19)$$

де f_0 – рівноважна функція (0 – наближення), $f_1(t), f_2(t), \dots$ – не рівноважні добавки (1-го, 2-го і т.д. порядку малості). Швидко згасання ряду забезпечується за умов: $f_0 \gg f_1(t) \gg f_2(t) \dots$

Як показав практичний досвід, найбільш суттєві для фізики твердого тіла результати можна отримати уже у першому наближенні:

$$f(t) \approx f_0 + f_1(t) . \quad (4.20)$$

Розсіяння – масовий процес за участю величезної кількості взаємодіючих частинок ($\approx 10^{28} \text{ м}^{-3}$). Спостереження масових процесів виявило таку їх закономірність у часі: зміна $f(t)$ з плином часу прямо пропорційна миттєвому значенню $f(t)$:

$$\left(\frac{\partial f(t)}{\partial t}\right)_{\text{розс}} \sim f(t). \quad (4.21)$$

Виходячи із розмірностей лівої і правої частин (4.21), позначимо коефіцієнт пропорційності $1/\tau$ і врахуємо, що f_0 від часу не залежить, а із плином часу не рівноважна добавка $f_1(t)$ зменшується:

$$\left(\frac{\partial f_1(t)}{\partial t}\right)_{\text{розс}} = -\frac{1}{\tau} f_1(t). \quad (4.22)$$

За умов: $t = 0 \rightarrow f_1(t) = f_1(0)$; $t \rightarrow \infty$, $f_1(t) \rightarrow 0$ розв'язок (4.22):

$$f_1(t) = f_1(0)e^{-t/\tau}. \quad (4.23)$$

Висновок: після припинення дії зовнішніх сил у момент $t = 0$ $f_1(t)$ спадає по експонентному закону з постійною τ . При $t = \tau$ не рівноважна добавка $f_1(t)$ має значення $f_1(t = \tau) = f_1(0)/e \approx f_1(0)/2,7 \approx 0,37 f_1(0)$. Середній час процесу релаксації системи після припинення дії зовнішніх сил τ – час, протягом якого не рівноважна добавка до функції розподілу зменшується в 2,7 рази.

Інтеграл зіткнень (4.17) із урахуванням (4.22), тобто у наближенні часу релаксації має вигляд:

$$\left(\frac{\partial f}{\partial t}\right)_{\text{розс}} = \int_{(\bar{p}')} \{W(\bar{p}, \bar{p}') [f_1(\bar{p}') - f_1(\bar{p})]\} dV_{\bar{p}'} = -\frac{f_1(\bar{p})}{\tau}. \quad (4.23)$$

Кінетичне рівняння (4.18) у наближенні часу релаксації:

$$-\frac{\partial f}{\partial t} = \bar{v} \nabla_{\bar{r}} f + \bar{F}_{\text{зов}} \nabla_{\bar{p}} f - \frac{f_1(\bar{p})}{\tau}. \quad (4.24)$$

Із (4.23) відокремимо розрахункову формулу для τ при фіксованому значенні \bar{p} :

$$\frac{1}{\tau(\bar{p})} = \int_{(\bar{p}')} W(\bar{p}, \bar{p}') \left[1 - \frac{f_1(\bar{p}')}{f_1(\bar{p})} \right] dV_{\bar{p}'}. \quad (4.25)$$

Для розв'язку кінетичного рівняння прийнято вибирати $f_1(t)$ у вигляді:

$$f_1(\bar{k}) = \left(-\frac{\partial f_0}{\partial E}\right) \bar{\chi}(E) \bar{k}, \quad (4.26)$$

де $\bar{\chi}(E)$ – векторна функція енергії, вигляд якої визначається зовнішніми силовими полями.

Для зовнішнього електричного поля $\bar{\mathcal{E}} \neq 0$ і електронного напівпровідника векторна функція енергії має такий вигляд:

$$\bar{\chi}_n(E) = -\hbar \frac{e\tau_n}{m_n^*} \bar{\mathcal{E}} = -\hbar \mu_n \bar{\mathcal{E}} = -\hbar \bar{v}_{dp}. \quad (4.26a)$$

У полі градієнта температури $\nabla T \neq 0$ векторна функція енергії:

$$\bar{\chi}_n(E) = -\frac{\hbar \tau_n}{m_n^*} \left\{ \frac{E - F_\phi}{T} \nabla T + (F_\phi - e\phi) \right\}. \quad (4.26б)$$

В умовах дії сили Лоренця і слабкого магнітного поля \bar{B} векторна функція енергії:

$$\bar{\chi}_n(E) = -\hbar \frac{e\tau_n}{m_n^*} \left\{ \bar{\mathcal{E}} + \frac{e\tau_n}{m_n^*} [\bar{B}\bar{\mathcal{E}}] \right\}. \quad (4.26в)$$

У формулі (4.25) замінімо імпульс $\bar{p} = \hbar \bar{k} / m_n^*$ на хвильовий вектор \bar{k} і

$dV_{\bar{p}'}$ на $\frac{2d^3\bar{k}'}{8\pi^3} = \frac{d^3\bar{k}'}{4\pi^3}$, де $8\pi^3$ – об'єм фазового простору на одне значення \bar{k}' ,

множник 2 враховує принцип Паулі; врахуємо також умову пружного розсіяння

$E \approx E'$ і, відповідно, $\bar{\chi}(E) \approx \bar{\chi}(E')$. У кінцевому підсумку отримаємо:

$$\frac{1}{\tau(\bar{k})} = \frac{1}{4\pi^3} \int_{(\bar{k}')} W(\bar{k}, \bar{k}') \left[1 - \frac{\bar{k}'}{\bar{k}} \right] d^3\bar{k}'. \quad (4.27)$$

Позначимо кут розсіяння між векторами $\angle \bar{k}, \bar{k}' = \theta$, врахуємо умову пружного розсіяння для модулів $k \approx k'$ і визначимо $\frac{\bar{k}'}{\bar{k}} = \cos \theta$. Повернемося до (4.27):

$$\frac{1}{\tau(\bar{k})} = \frac{1}{4\pi^3} \int_{(\bar{k}')} W(\bar{k}, \bar{k}') [1 - \cos \theta] d^3 \bar{k}'. \quad (4.28)$$

Висновок: отримали формулу усереднення імовірності розсіяння $W(\bar{k}, \bar{k}')$ при переході із початкового фіксованого стану \bar{k} в усі можливі кінцеві стани \bar{k}' із ваговим множником $[1 - \cos \theta]$, який $= 0$ при $\theta = 0$ (розсіяння відсутнє, $\bar{k} = \bar{k}'$), $= 2$ при $\theta = \pi$ (повне відбиття від дефекту, $\bar{k} = -\bar{k}'$).

Для кількісної оцінки ефективності того чи іншого механізму розсіяння використовують усереднені параметри розсіяння.

Вводиться параметр – **ефективний переріз розсіяння** σ – діаметральний переріз сферичного потенціалу центру розсіяння, перпендикулярний напрямку потоку електронів.

Позначимо N_σ – концентрація центрів розсіяння із ефективним перерізом σ . Для потоку, щільність якого n_0 , інтегральний переріз розсіяння σN_σ . Кількість розсіяних за одиницю часу електронів n_1 :

$$n_1 = n_0 \sigma N_\sigma. \quad (4.29)$$

Введемо ще один параметр – **середнє значення імовірності розсіяння** W – число зіткнень електрона із центрами розсіяння за одиницю часу, тобто, частоту зіткнень. Враховуючи, що $n_1 = nW$, знайдемо взаємозв'язок між σ і W :

$$W = v \sigma N_\sigma. \quad (4.30)$$

Час релаксації – середній час вільного пробігу між двома послідовними зіткненнями:

$$\tau = W^{-1} \quad (4.31)$$

Довжина вільного пробігу:

$$L = v\tau = vW^{-1} = (\sigma N_{\sigma})^{-1}. \quad (4.32)$$

Таким чином, формула (4.32) пов'язує між собою усі параметри, які характеризують розсіяння.

Слід зауважити, що введені параметри розсіяння (4.30), (4.31), (4.32) мають фізичний зміст *середніх значень* по ансамблю вільних електронів.

Зазвичай одночасно „працюють” декілька механізмів розсіяння. Їх дія незалежна, альтернативна – або той, або інший. Відповідно до теорії імовірностей, повна імовірність W у таких умовах є сумою імовірностей окремих механізмів розсіяння W_i :

$$W = \sum_i W_i. \quad (4.33)$$

Час релаксації, відповідно до (4.31) та (4.33):

$$\tau^{-1} = \sum_i \tau_i^{-1}. \quad (4.34)$$

Довжина вільного пробігу:

$$L^{-1} = \sum_i L_i^{-1}. \quad (4.35)$$

Фізичний зміст формул (4.34) і (4.35) можна визначити таким чином: повний час релаксації та повна довжина вільного пробігу менші найменшого значення цих параметрів для окремих механізмів розсіяння.

4.5. Класифікація механізмів розсіяння

Ідеальна кристалічна структура, як було показано у п. І.8, не розсіює електронні хвилі. Інтеграл руху у цьому випадку – квазіімпульс – не залежить від часу, незважаючи на те, що рух відбувається у градієнтному потенціальному полі. В реальних структурах присутні динамічні і статичні дефекти, які порушують періодичність потенціалу, і квазіімпульс стає залежним від часу. Таким чином, квазіімпульс електрона змінюється як під

впливом зовнішнього поля, так і внутрішнього неперіодичного поля, створеного дефектами, які розсіюють електронні хвилі.

До основних механізмів розсіяння у температурному діапазоні роботи переважної більшості напівпровідникових приладів і інтегральних мікросхем, а саме 200 – 400 К, відносяться *фононне* та *іонне*.

Фононне розсіяння – результат взаємодії носіїв заряду із тепловими коливаннями атомів, які є *динамічними дефектами*. Цей механізм „працює” у всьому діапазоні температур існування твердого тіла.

Іонне розсіяння – результат прямої електростатичної взаємодії носіїв заряду із іонізованими домішками, які є *статичними дефектами*, що заміщають основні атоми. Цей механізм починає „працювати” при температурах, достатніх для термоіонізації домішок. Зазвичай це температури $> 20 - 100$ К.

Окрім вище названих основних, існують і інші механізми розсіяння: на нейтральних дефектах, на дислокаціях, міжелектронне і т.д. Помітний вплив на динамічні характеристики носіїв заряду від цих механізмів розсіяння спостерігається лише в умовах, коли не ефективні основні механізми, тобто, при температурах < 20 К.

4.6. Розсіяння на теплових коливаннях атомів.

Теплові коливання атомів класифікуються трьома незалежними хвилями – гармонічними проекціями : одна подовжня і дві поперечні.

При *подовжніх акустичних коливаннях* гармонічно змінюється відстань між атомами, що безпосередньо впливає на валентні зв'язки і енергетичну зонну структуру. Зонна структура *виявляється модульованою* подовжніми акустичними коливаннями атомів (рис.4.1). При *поперечних коливаннях* відстань між атомами практично не змінюється і *модуляція* енергетичного спектра електронів *відсутня*.



Рис.4.1. Модуляція енергетичного спектру подовжніми акустичними тепловими коливаннями атомів. Амплітуда модуляції $w(r,t)$ – деформаційний потенціал

Найважливішим наслідком модуляції енергетичного спектру є те, що повна енергія електрона має у своєму складі енергію теплових коливань у вигляді деформаційного потенціалу із амплітудою $w(r,t)$. У реальному енергетичному масштабі $w \ll E_g$. Повний розрахунок параметрів фононного розсіяння можливий на основі теорії деформаційного потенціалу і теорії квантових переходів під дією збудження, яке гармонічно залежить від часу. Але деякі важливі для практики закономірності можна отримати і на основі спрощеної моделі електрон – фононній взаємодії.

Енергія теплових коливань *квантована*. Квант теплової коливальної енергії – **фонон** – розглядається як квазічастинка із хвильовим вектором \bar{q} , частотою ω_q і енергією $\hbar\omega_q$. Електрон – фононній взаємодія, яка відбувається на квантовому хвильовому рівні, вимагає виконання законів збереження енергії, хвильового вектора електрона і числа фононів N_q :

$$E' = E \pm \hbar\omega_q, \quad \bar{k}' = \bar{k} \pm \bar{q}, \quad N'_q = N_q \mp 1, \quad (4.36)$$

де E', \bar{k}', N'_q – енергія, хвильовий вектор електрона, число фононів після взаємодії; знаки \pm означають поглинання або народження фонона.

Запишемо закон збереження енергії (4.36) у вигляді співвідношення частот: $\omega_{k'} = \omega_k \pm \omega_q$. У хвильовій механіці таке співвідношення частот має місце при відбиванні хвиль від рухомого об'єкту і відоме як *ефект Доплера*. У нашому випадку таким рухомих об'єктом є фонон, на якому розсіюється електронна хвиля.

Закони збереження (4.36) діють в умовах *термодинамічної рівноваги* між електронною і фононною системами і мають *обмінний характер* – імовірності поглинання і народження фононів однакові (ідеально пружне розсіювання):

$$W^+ = W^-. \quad (4.37)$$

В умовах дії зовнішніх *слабких силових полів* електронна система виходить із стану термодинамічної рівноваги, енергетичний обмін стає дещо асиметричним: $W^+ < W^-$ – імовірність *народження* фононів переважає їх *поглинання*.

Кількість носіїв заряду, які приймають участь у взаємодії, постійне – це їх концентрація при певній температурі. Кількість фононів – змінна. Розрахунки, які враховують ці особливості електронної і фононної систем, показують, що:

$$W^+ \sim N_q \delta(E' - E + \hbar\omega_q), \quad W^- \sim (N_q + 1) \delta(E' - E - \hbar\omega_q), \quad (4.38)$$

де $\delta(E' - E \pm \hbar\omega_q)$ – дельта-функція, яка враховує закон збереження енергії ($\delta(E' = E \pm \hbar\omega_q) = 1$, $\delta(E' \neq E \pm \hbar\omega_q) = 0$).

Фононне розсіювання – масовий процес енергетичного обміну. Кожен електрон при одному зіткненні з фононом може або втратити, або отримати енергію $\hbar\omega_q$. Із урахуванням асиметрії енергетичного обміну, частина енергії електронів, яка трансформується у теплові коливання атомів ΔE , становить:

$$\Delta E = \hbar\omega_q \frac{W^- - W^+}{W^- + W^+} = \hbar\omega_q \frac{(N_q + 1) - N_q}{(N_q + 1) + N_q} = \hbar\omega_q \frac{1}{2N_q + 1}. \quad (4.39)$$

Із (4.39) можна зробити такий висновок: на одне зіткнення електрона і фонона з втратою енергії припадає $2N_q + 1$ обмінних зіткнень, причому, $N_q \gg 1$.

Електрон – фононна взаємодія ефективна, коли довжина електронної хвилі $\lambda_e = 2\pi/k$ і фононної $\lambda_\phi = 2\pi/q$ спів розмірні, тобто модулі хвильових векторів практично співпадають: $k \approx q$.

Зробимо оцінку пружності фононного розсіяння для 300 К. Середня теплова швидкість електронів $v_T = 2,5 \cdot 10^5$ м/с і їй відповідає модуль хвильового вектора: $k = m^* v_T / \hbar \approx 2 \cdot 10^9$ м⁻¹. Енергія фонона E_ϕ із вектором $q = 2 \cdot 10^9$ м⁻¹ (довгохвильові подовжні фонони) і швидкістю звуку $v_{зв} = 5 \cdot 10^3$ м/с:

$E_\phi = \hbar v_{зв} q = \hbar v_{зв} k \approx 6 \cdot 10^{-3}$ еВ. Енергія електрона за тих же умов :

$E_e = \hbar^2 k^2 / 2m^* = 1,2 \cdot 10^{-1}$ еВ (при розрахунку прийнято $m^* = m_e \approx 10^{-30}$ кг).

Співвідношення енергій:

$$E_\phi / E_e \approx 5 \cdot 10^{-2} \ll 1. \quad (4.40)$$

Аналіз (4.36) і (4.40) показує, що, за певних умов, непружністю фононного розсіяння можна нехтувати, тобто, вважати повну енергію електрона незмінною $E'_e \approx E_e$. На користь такого висновку можна навести ще й таке: еквівалентну “масу” фонона M_ϕ можна оцінити при рівних імпульсах фонона і електрона $M_\phi v_{зв} = m_e^* v_T$. Кількісна оцінка по вищенаведеним даним показує, що еквівалентна “маса” фонона у 40 – 50 разів більша маси електрона, але цього недостатньо для того, щоб вважати фонони нерухомими, статичними дефектами. Відносно важкі рухомі дефекти – фонони - у незначній мірі змінюють енергію електрона (4.40), але суттєво впливають на його хвильовий вектор.

Кожен атом розглядаємо як класичний осцилятор з масою M_a , що коливається з частотою ω_{\max} і амплітудою A_0 при температурі T :

$$A_0 = \sqrt{\frac{2k_B T}{M_a \omega_{\max}^2}}, \quad (4.41)$$

де M_a – маса атома, ω_{\max} – максимальна частота подовжніх акустичних коливань.

Ефективний переріз розсіяння σ визначимо як площу, яку займає атом при коливаннях за винятком площі перерізу нерухомого атома (нерухомий атом не розсіює):

$$\sigma = \pi[(A_0 + r_a)^2 - r_a^2], \quad (4.42)$$

де $r_a \approx 0,05$ нм – радіус атома.

Оцінка A_0 при 300 К для Ge, Si, GaAs дає середнє значення $A_0 \approx 0,0016$ нм. Якщо знехтувати A_0^2 , для σ отримаємо:

$$\sigma \approx 2\pi A_0 r_a. \quad (4.43)$$

Таблиця 4.1

Напівпро- відник	M_a кг 10^{-25}	ω_{\max} рад/с	$\hbar\omega_{\max}$ еВ	$v_{3\phi}$ м/с 10^3	N_a м ⁻³ 10^{28}	A_0 м 10^{-12}	σ м ² 10^{-22}
Si	0,46	8,9	0,063	9,13	4,4	1,5	4,7
Ge	1,2	5,2	0,037	5,4	5	1,6	5
GaAs	$M_{Ga}=1,1$ $M_{As}=1,2$	5	0,0355	5,28	2,2	1,7	5,3

Для розрахунків A_0 і σ використані дані, наведені в табл. 4.1. Числова оцінка по наведеним даним дає середнє значення $\sigma \approx 5 \cdot 10^{-22}$ м².

Розрахуємо інші характеристики розсіяння, прирівнявши кількість фононів N_q до атомної щільності $N_a \approx 4 \cdot 10^{28}$ м⁻³:

$$W = \sigma v_T N_a \approx 10^{13} \text{ Гц}; \quad \tau_\phi = W^{-1} \approx 10^{-13} \text{ с}; \quad L_\phi = (\sigma N_a)^{-1} \approx 0,05 \text{ мкм}. \quad (4.43a)$$

Отримані кількісні значення основних характеристик розсіяння при фіксованій температурі відповідають, у цілому, експериментальним даним.

Спрощена модель фононного розсіяння дає змогу встановити такі важливі закономірності, як залежність часу фононої релаксації τ_ϕ від температури кристалу. Із формул (4.13) і (4.15) знаходимо, що $\sigma(T) \sim T^{1/2}$. Залежність швидкості електронів від температури визначимо за умови, що $\frac{m^* v^2}{2} = k_B T$, тобто, $v(T) \sim T^{1/2}$. Середня кількість фононів N_q при рівномірному розподілі по енергіям:

$$N_q = \frac{k_B T}{\hbar \omega_q} = \frac{k_B T}{\hbar v_{\text{зв}} q}, \quad (4.44)$$

де $v_{\text{зв}}$ – швидкість звуку, яка від температури практично не залежить.

Модулі хвильових векторів електрона і фонона повинні бути співрозмірними $k \approx q$, тому при зміні k відповідно буде змінюватися і q . Температурна залежність k така сама, як і $v(T)$, тобто $k \sim T^{1/2}$ і $q \sim T^{1/2}$. Врахувавши це у (4.44), для N_q маємо: $N_q(T) \sim T^{1/2}$. Підставимо отримані температурні залежності у (4.30), (4.31) і для τ_ϕ маємо:

$$\tau_\phi \sim T^{-3/2}. \quad (4.45)$$

Температурна залежність мобільності електронів при фононному розсіянні для невироджених напівпровідників:

$$\mu_n = \frac{e \tau_\phi}{m^*} \sim T^{-3/2} \quad (4.46)$$

експериментально доведена (рис.4.4).

Температурна залежність довжини вільного пробігу L_ϕ :

$$L_\phi = v \tau_\phi \sim T^{-1}. \quad (4.47)$$

Основним недоліком спрощеної моделі фононного розсіяння є неможливість розділити залежності параметрів розсіяння від температури і енергії носіїв заряду. Спрощену модель можна використовувати, якщо у повній швидкості, яка складається із теплової і дрейфової $v = v_T + v_{dp}$, тепла компонента значно переважає дрейфову $v_T \gg v_{dp}$. Таке співвідношення швидкостей можливе лише у слабких електричних полях, коли вся енергія, отримана електронами від зовнішнього поля на шляху вільного пробігу $L_\phi e\mathcal{E}_{\text{зов}}$, трансформується у теплові коливання атомів. Із урахуванням (4.39) цю умову слабого електричного можна визначити таким чином:

$$L_\phi e\mathcal{E}_{\text{зов}} \ll \Delta E. \quad (4.48)$$

Зовнішньою ознакою слабого електричного поля за критерієм (4.48) є **закон Ома** – щільність струму прямо пропорційна напруженості поля.

Детальний аналіз впливу електричного поля на динамічні характеристики носіїв заряду та їх концентрацію поданий у п.4.13.

Квантова теорія фононного розсіяння дає такі закономірності:

$$\tau_\phi \sim m^{*-2} (k_B T)^{-1} v^{-1} \sim m^{*-1} (k_B T)^{-1} k^{-1} \sim m^{*-3/2} (k_B T)^{-1} E^{-1/2};$$

$$L_\phi = v \tau_\phi \sim T^{-1} E^0. \quad (4.49)$$

Помітимо, що довжина вільного пробігу при фононному розсіянні від енергії носіїв заряду не залежить.

Якщо у (4.49) замінити E на середню теплову енергію $k_B T$, для τ_ϕ отримаємо (4.45), тобто теж саме, що мали у спрощеній моделі розсіяння.

4.7. Розсіяння на іонізованих домішках (іонне розсіяння)

Присутність іонізованих домішок – донорів або акцепторів, вбудованих у кристалічну структуру, необхідна для отримання певного типу провідності

напівпровідника. Пряма електростатична взаємодія вільних носіїв заряду із іонами домішок і є фізичним змістом іонного розсіювання.

Пружність іонного розсіювання забезпечується тим, що маса іонів перевищує масу електронів приблизно у 10000 разів. При такій величезній різниці мас іони можна вважати нерухомими при зіткненні з електронами, електрони практично не зміщують іони і зберігають модуль свого квазіімпульсу.

Потенціальне поле іонів сферично симетричне, тому іонне розсіювання – результат руху носіїв заряду у полі центральної сили. Іонізація домішок відбувається за рахунок або втрати, або приєднання одного електрона, тобто заряд іонів $\pm e$. Потенціальне поле таких іонів у першому наближенні кулонівське:

$$U(r) = \pm \frac{e^2}{\epsilon \epsilon_0 r}, \quad (4.50)$$

ϵ – відносна діелектрична проникність, $\epsilon_0 = 8,8 \cdot 10^{-12}$ Ф/м – діелектрична постійна вакууму, r – відстань між носієм заряду і іоном, $r_{\min} = b$ – *прицільна відстань* – відстань, на якій пройшов би носій заряду без розсіювання.

Домішковий іон – це ядро атома, оточене оболонками внутрішніх електронів, які частково екранують заряд ядра і потенціальне поле спадає зі збільшенням відстані швидше, ніж поле точкового заряду (4.50). Слід врахувати також те, що носії заряду, які проходять посередині між сусідніми іонами не змінюють своєї траєкторії, тобто не розсіюються. Із урахуванням цих особливостей потенціального поля іонів, вважають це *поле просторово обмеженим на відстані b_{\max} , яка становить половину середньої відстані між іонами:*

$$b_{\max} = 0,5 N_i^{-1/3}, \quad (4.51)$$

де N_i – концентрація іонізованих домішок.

У межах $0 < b \leq b_{\max}$ поле кулонівське, але при $b > b_{\max} \rightarrow U(r) = 0$. Кут розсіяння $\angle \bar{k}, \bar{k}' = \theta$ у кулонівському полі виявляється залежним від b , відповідно і ефективний переріз розсіяння σ буде залежним від b і θ . Для цього випадку вводять поняття *диференціального ефективного перерізу розсіяння для одного центру*:

$$\sigma(\theta)d\Omega = \frac{\text{кількість електронів, відхилених на кут } \theta \text{ у межі тілесного кута } d\Omega}{\text{кількість електронів, які пройшли через кільцеплощу } 2\pi b db}. \quad (4.52)$$

Тілесний кут $d\Omega$ – кільце на сферичній поверхні, яке спирається на центральний кут $d\theta$ (рис.4.2.) .

Для $\sigma(\theta)$ Резерфорд отримав формулу, яка для однозарядових іонів має вигляд:

$$\sigma(\theta) = \left(\frac{e^2}{2\epsilon m^* v^2} \right)^2 \frac{1}{\sin^4 \theta/2}. \quad (4.53)$$

Для розрахунку часу іонної релаксації $\tau_i(\bar{k})$ у формулі (4.28) визначимо $W(\bar{k}, \bar{k}')$ через $\sigma(\theta)$ відповідно до (4.30). За умови пружного розсіяння $k \approx k'$ і фізичного змісту $\sigma(\theta)d\Omega$ (4.52): $W(\bar{k}, \bar{k}')n d\Omega = \sigma(\theta)N_i n v d\Omega$, звідки:

$$W(\bar{k}, \bar{k}') = \sigma(\theta)N_i v. \quad (4.54)$$

Відповідно до (4.54), $W(\bar{k}, \bar{k}')$ тепер залежить лише від кута між \bar{k} і \bar{k}' , тобто $W(\bar{k}, \bar{k}') = W(\theta)$. Це дає змогу перейти від інтегрування по \bar{k}' до інтегрування по θ , замінивши $\frac{d^3 \bar{k}'}{4\pi^3}$ на $d\Omega = 2\pi \sin \theta d\theta$, і для (4.28) отримаємо:

$$\frac{1}{\tau(\bar{k})} = 2\pi N_i v \int_{(\theta)} \sigma(\theta)[1 - \cos \theta] \sin \theta d\theta. \quad (4.55)$$

Інтегральна частина (4.55):

$$\sigma_c = 2\pi \int_{(\theta)} \sigma(\theta)[1 - \cos \theta] \sin \theta d\theta \quad (4.55a)$$

– *транспортний ефективний переріз* - середнє значення ефективного перерізу розсіяння по всіх можливим кутам розсіяння у межах $0 \leq \theta \leq \pi$. Якщо обрахувати цей інтеграл для $\sigma(\theta)$ у вигляді формули Резерфорда (4.53), на нижній межі $\theta = 0$ отримаємо нескінчене значення перерізу розсіяння $\sigma \rightarrow \infty$ і $\tau(\bar{k}) \rightarrow 0$. Фізичний зміст такого результату полягає у тому, що формула Резерфорда відповідає чисто кулонівському потенціалу розсіяння (4.50), який повільно спадає при зростанні відстані $U(r) \sim r^{-1}$ і на якій би відстані не знаходився носій заряду він буде розсіюватися. Введення обмеження на прицільну відстань b_{\max} (4.51) знімає цю проблему.

Кут θ і прицільна відстань b пов'язані оберненою залежністю – чим далі від іона проходить носій заряду, тим менше кут розсіяння (рис.4.2).

На верхній межі поля розсіяння $b = b_{\max}$ кут розсіяння мінімальний $\theta = \theta_{\min}$. Максимальне значення кута розсіяння $\theta_{\max} = \pi$ відповідає повному відбиттю від центру розсіяння. Таким чином, межі інтегрування по θ :

$$\theta_{\min} \leq \theta \leq \pi. \quad (4.56)$$

Аналітичний зв'язок між b і θ відомий із класичної теорії розсіяння у полі центральної сили:

$$b = \frac{e^2}{\epsilon \epsilon_0 m^* v^2} \operatorname{ctg} \frac{\theta}{2}. \quad (4.57)$$

Після підстановки (4.51) у (4.57) отримаємо:

$$\operatorname{ctg} \frac{\theta_{\min}}{2} = \frac{\epsilon \epsilon_0 m^* v^2}{2e^2 N_i^{1/3}}. \quad (4.58)$$

У інтегральну частину (4.55a) підставимо (4.53), винесемо за знак інтеграла все, що не залежить від θ і після елементарних тригонометричних перетворень і інтегрування отримаємо:

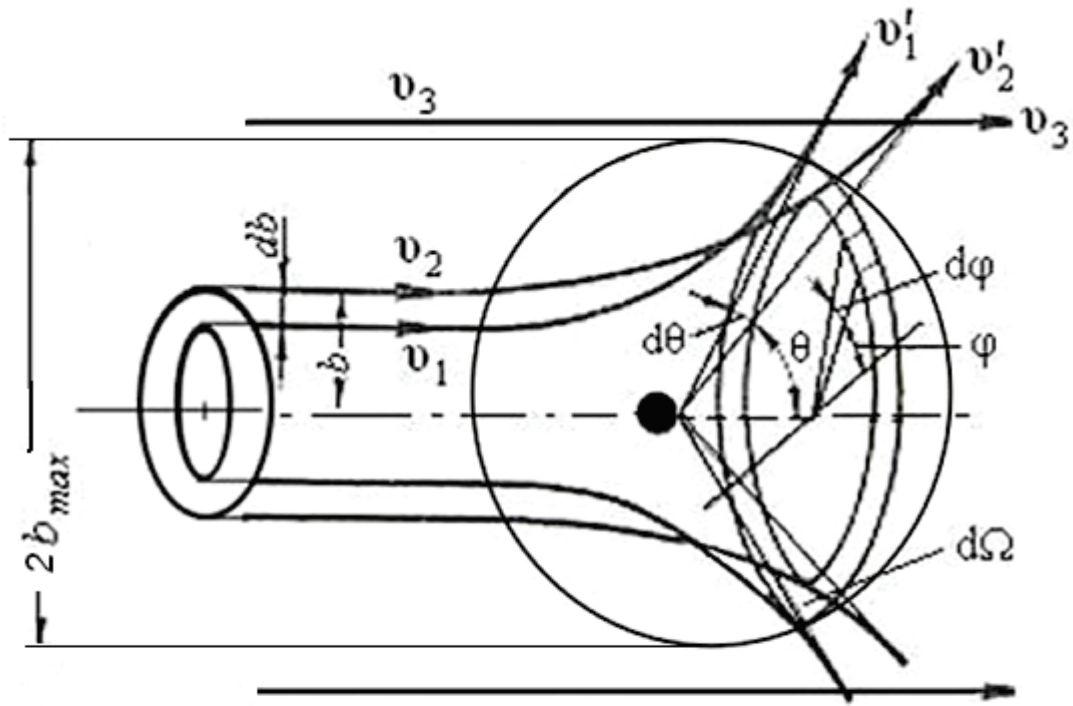


Рис.4.2. Розсіяння потоку іонізованим домішковим центром.
Носії заряду, які проходять на відстані $b > b_{max}$, не
розсіюються $\theta = 0$

$$\int_{\theta_{min}}^{\pi} \frac{(1 - \cos \theta) \sin \theta}{\sin^4 \theta / 2} d\theta = 8 \int_{\theta_{min}}^{\pi} \operatorname{ctg} \frac{\theta}{2} d\left(\frac{\theta}{2}\right) = -4 \ln \sin^2 \frac{\theta_{min}}{2}. \quad (4.59)$$

Підставимо усі проміжні результати (4.53), (4.58), (4.59) у (4.55) і для часу іонної релаксації $\tau_i(\bar{k})$ будемо мати:

$$\tau_i(\bar{k}) = \frac{(\varepsilon \varepsilon_0)^2 m^{*2} v^3}{2\pi e^4 N_i} \left\{ \ln \left[1 + \left(\frac{\varepsilon \varepsilon_0 m^* v^2}{2e^2 N_i^{1/3}} \right)^2 \right] \right\}^{-1}. \quad (4.60)$$

Співвідношення (4.60), записане для однозарядових іонів, відоме в теорії розсіяння як *формула Конуелл – Вайскопфа*.

Логарифмічна функція в (4.60) змінюється у залежності від v , N_i та m^* значно повільніше ніж v^3 , N_i^{-1} та m^{*2} , тому основна закономірність для часу іонної релаксації має такий вигляд:

$$\tau_i(\bar{k}) \sim v^3 N_i^{-1} m^{*2} \sim k^3 N_i^{-1} m^{*-1}. \quad (4.61)$$

Якщо визначити швидкість через кінетичну енергію $E = m^* v^2 / 2$, а також прийняти E як середню теплову енергію $k_B T$, отримаємо такі закономірності:

$$\tau_i(\bar{k}) \sim E^{3/2} N_i^{-1} m^{*1/2} \sim T^{3/2} N_i^{-1} m^{*1/2}. \quad (4.62)$$

Фізичний зміст закономірностей (4.61), (4.62) полягає у тому, що:

- чим більше концентрація іонізованих домішок N_i , тим імовірніше зіткнення носіїв заряду з ними і, відповідно, менше час іонної релаксації;
- чим більше швидкість, енергія, температура, тим менший час перебування носіїв заряду у потенціальному полі іонів і менший кут розсіювання θ , що збільшує час іонної релаксації;
- чим більша ефективна маса носіїв заряду, тим менший кут розсіювання θ , що також збільшує час іонної релаксації.

Визначимо довжину вільного пробігу при іонному розсіянні за допомогою (4.32) та (4.61):

$$L_i = v \tau \sim E^{1/2} E^{3/2} N_i^{-1} \sim E^2 N_i^{-1}. \quad (4.63)$$

На рис.4.3 показаний вплив концентрації іонізованих домішок на мобільність електронів і дірок при фіксованій температурі, що у цілому відповідає закономірності (4.63).

Довжина вільного пробігу при іонному розсіянні залежить від енергії носіїв заряду $L_i \sim E^2$ (4.63), на відміну від фононного розсіювання де $L_\phi \sim E^0$ (4.49). Об'єднаємо ці енергетичні залежності:

$$L \sim E^r, \quad (4.63a)$$

де $r = 0$ - фононне розсіювання, $r = 2$ - іонне розсіювання.

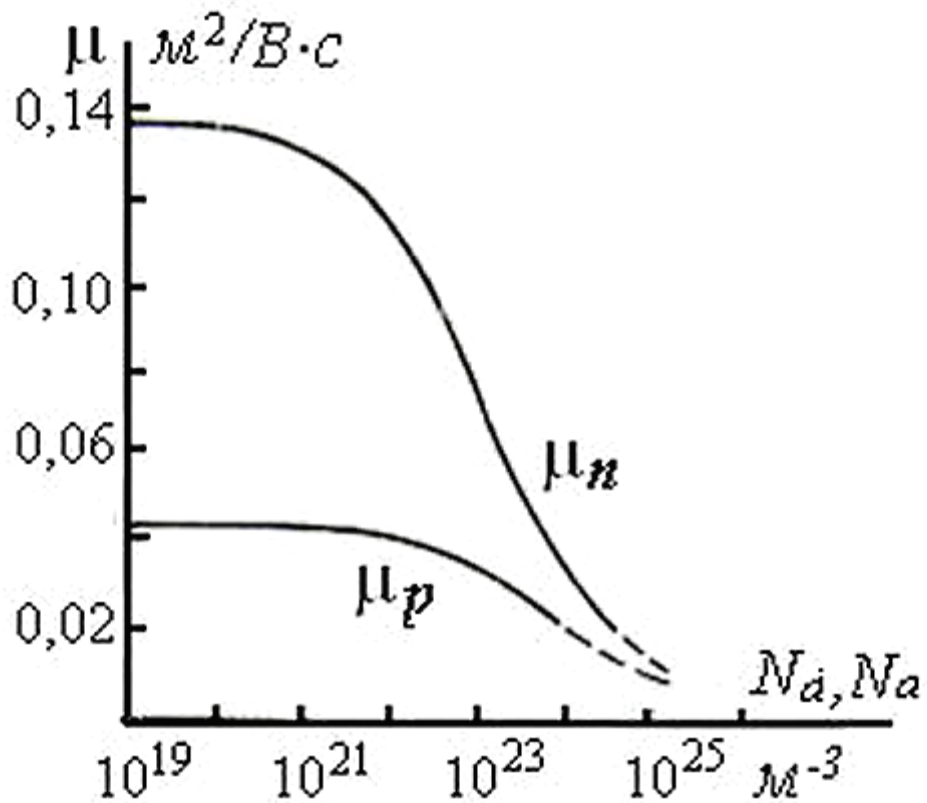


Рис.4.3. Залежність мобільності електронів і дірок у кремнії від рівня легування донорами і акцепторами ($T = 300 \text{ K}$)

4.8. Залежність мобільності від температури

Мобільність носіїв заряду – визначальна характеристика їх динамічних властивостей:

$$\mu = e\tau / m^*, \quad (4.64)$$

де m^* представляє енергетичну зонну структуру напівпровідника, τ – процеси розсіяння енергії та квазіімпульсу.

У п. 4.6 і 4.7 показано, що для основних механізмів розсіяння – фононного та іонного – залежність $\tau(T)$ має протилежний характер, а саме:

$\tau \sim T^\gamma$, де $\gamma = -3/2$ для фононного розсіяння, $\gamma = +3/2$ для іонного розсіяння.

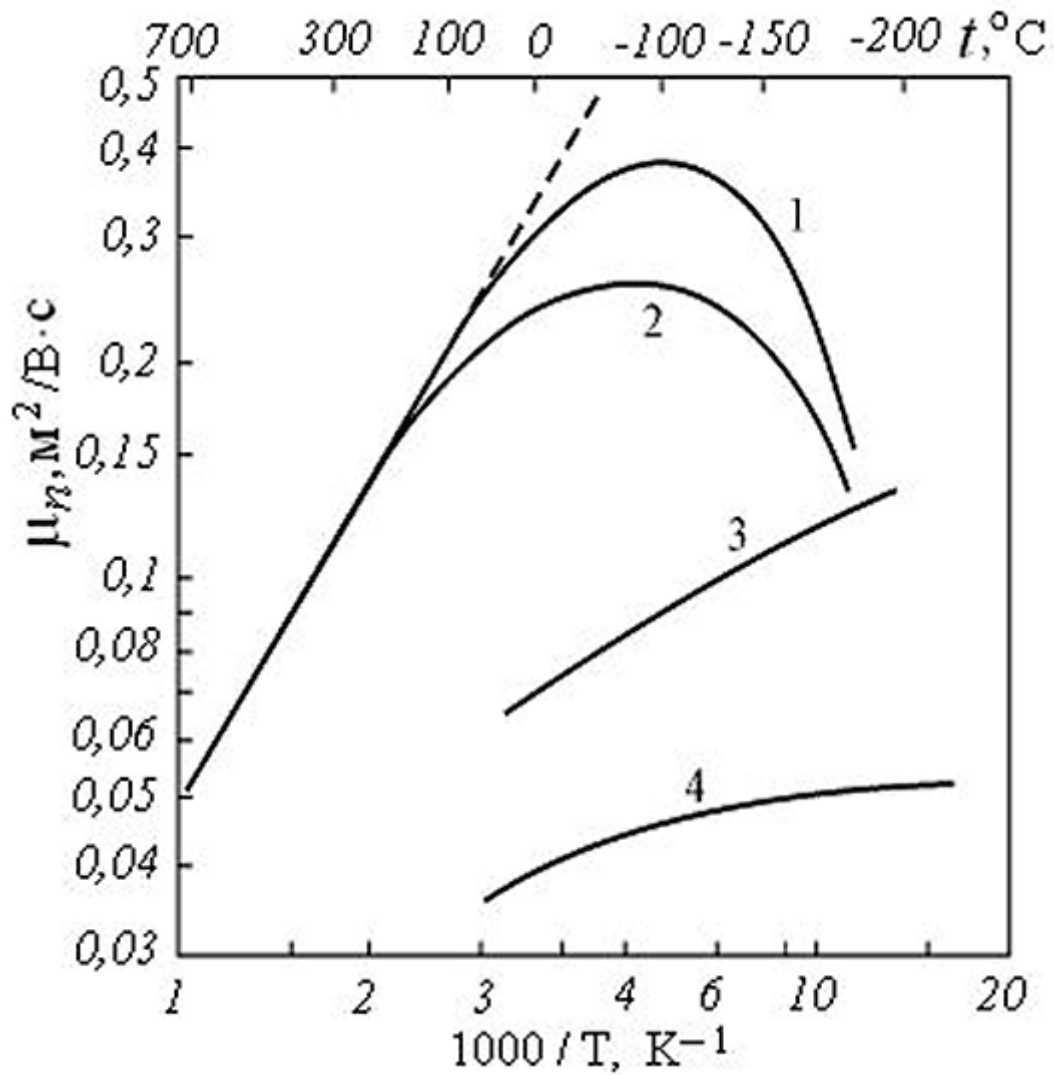


Рис.4.4. Залежність мобільності електронів у невідродженому (1,2) та відродженому (3,4) кремнії із різним вмістом фосфору від температури. Вміст атомів фосфору вказана на рис.4.5

Вказані кількісні значення γ підтверджені експериментально (у межах похибки визначення) для електронних напівпровідників і тих діркових напівпровідників, у яких незначна різниця ефективних мас важких і легких дірок. Однак для p -Ge експериментальне значення $\gamma \approx -2,3$, що пов'язується із суттєвою різницею ефективних мас важких і легких дірок.

При одночасній дії фононного і іонного механізмів розсіяння, зважаючи на протилежний характер їх температурної залежності, слід очікувати максимального значення мобільності при певній температурі T_{\max} .

На рис.4.4 наведені експериментальні залежності $\mu_n(T)$ для кремнію із різним вмістом фосфору. При концентрації донорів $N_d \leq 10^{22} \text{ м}^{-3}$ (рис.4.4, (1,2)), що відповідає невинродженому електронному газу, мобільність має максимальне значення при температурі $\approx 250 \text{ К}$.

Балістичний ефект іонного розсіяння – чим більша швидкість, енергія носіїв заряду, тим більший час релаксації і, відповідно, мобільність – має місце лише для невинродженого напівпровідника. При високих рівнях легування $N_d > 10^{23} \text{ м}^{-3}$ потенціальні поля іонів перекриваються, утворюється домішкова зона, об'єднана із зоною провідності, електронний газ стає винродженим. У межах домішкової зони енергетичні залежності часу іонної релаксації (4.61), (4.62) втрачають свою актуальність, але N_d продовжує впливати на середнє значення мобільності (рис.4.4,(3,4)), а температурну залежність мобільності зумовлює фононне розсіяння.

При температурах $> 250 \text{ К}$ фононне розсіяння стає домінуючим для залежності мобільності від температури.

4.9. Електропровідність напівпровідників

Розрахуємо щільність струму в умовах дії закону Ома $\bar{j} = \bar{j}_n + \bar{j}_p = (\sigma_n + \sigma_p)\bar{\mathcal{E}}$, де σ_n, σ_p - питомі провідності електронів і дірок, які не залежать від $\bar{\mathcal{E}}$.

Оскільки розрахунки електронного і діркового струмів проводяться за аналогічними процедурами, надалі поданий розрахунок лише електронної компоненти повного струму.

Число електронів у елементарному об'єму імпульсного простору:

$$dn = f_n \frac{d^3\bar{k}}{4\pi^3}, \quad (4.65)$$

де f_n – не рівноважна функція розподілу.

Миттєва щільність струму $d\bar{j}_n = -ef_n\bar{v} \frac{d^3\bar{k}}{4\pi^3}$, повна щільність струму:

$$\bar{j}_n = -\frac{e}{4\pi^3} \int_{-\infty}^{+\infty} \bar{v} f_n d^3\bar{k}. \quad (4.66)$$

Помітимо, що інтегрування іде по усім можливим початковим значенням \bar{k} . Вибираємо невироджений напівпровідник із скалярною ефективною масою m_n^* . Нерівноважну функцію розподілу f_n визначимо із урахуванням (4.3), (4.20), (4.26), (4.26a):

$$f_n = f_0(E, T) + f_{1n} = \exp\left(-\frac{E - F_\phi}{k_B T}\right) - \left(-\frac{\partial f_0}{\partial E}\right) \hbar \frac{e\tau_n}{m_n^*} \bar{\mathcal{E}} \bar{k}. \quad (4.67)$$

Підставимо (4.67) у (4.66), визначимо \bar{v} і E через \bar{k} : $\bar{v} = \hbar \bar{k} / m_n^*$, $E = \hbar^2 k^2 / 2m_n^*$. Розділимо інтеграл (4.66) на два інтеграли, один залежний від рівноважної функції f_0 , другій – від не рівноважної добавки f_{1n} і отримаємо:

$$\bar{j}_n = -\frac{e\hbar}{4\pi^3 m_n^*} \left[\int_{-\infty}^{+\infty} \bar{k} e^{(-\alpha k^2 + F_\phi/k_B T)} d^3\bar{k} + \int_{-\infty}^{+\infty} \bar{k} \left(-\frac{\partial f_0}{\partial E}\right) \hbar \frac{e\tau_n}{m_n^*} \bar{\mathcal{E}} \bar{k} d^3\bar{k} \right], \quad (4.67)$$

де позначено $\alpha = \hbar^2 / 2m_n^* k_B T$.

Перший інтеграл у (4.67) непарний відносно \bar{k} і при інтегруванні у симетричних межах дає 0. Фізичний зміст цього результату такий: *при рівноважному розподілі носіїв заряду рух зарядів є, але струму немає.*

Деталізуємо умови для подальших розрахунків. Положення рівня Фермі відповідає області виснаження донорів (рис.3.6). Розсіяння фононне і, відповідно до (4.49), $\tau_{n\phi}(k) = \tau_{0\phi} m_n^{*-1} (k_B T)^{-1} k^{-1}$, де у $\tau_{0\phi}$ зібрані усі величини, які не залежать від $\bar{k}, m_n^*, k_B T$.

Визначимо $(-\frac{\partial f_0}{\partial E})$ із урахуванням (3.42) і для електронного струму отримаємо:

$$\bar{j}_n = \frac{e^2 \hbar^2 n}{4\pi^3 m_n^{*2} N_c k_B T} \int_{-\infty}^{+\infty} \tau_{n\phi} e^{-\alpha k^2} (\bar{\mathcal{E}} \bar{k}) \bar{k} d^3 \bar{k}. \quad (4.68)$$

Потрійний інтеграл у формулі (4.68) можна спростити, перейшовши до полярної системи координат (полярна вісь – вектор напруженості електричного поля $\bar{\mathcal{E}}$):

$$\int_{-\infty}^{+\infty} \tau_{n\phi} e^{-\alpha k^2} (\bar{\mathcal{E}} \bar{k}) \bar{k} d^3 \bar{k} = \frac{4\pi}{3} \bar{\mathcal{E}} \int_0^{+\infty} \tau_{n\phi} e^{-\alpha k^2} k^4 dk. \quad (4.69)$$

Фізичний зміст інтегрування (4.69) – усереднення $\tau_{n\phi}$ по всім початковим значенням хвильового вектора із урахуванням певного механізму розсіяння.

Інтегральна частина (4.69) після підстановки $\tau_{n\phi}(k)$ зводиться до табличного інтегралу:

$$\int_0^{+\infty} k^3 e^{-\alpha k^2} dk = \frac{1}{2\alpha^2} = \frac{2m_n^{*2} (k_B T)^2}{\hbar^4}. \quad (4.70)$$

Після підстановки $\tau_{n\phi}(k)$, N_c із (II.21), (4.69), (4.70) у (4.68) для щільності електронного струму отримаємо:

$$\bar{j}_n = en \frac{const'}{m_n^{*5/2} T^{3/2}} \bar{\mathcal{E}} = en \mu_{n\phi} \bar{\mathcal{E}}, \quad (4.71)$$

де у $const'$ зібрані усі постійні величини, окрім ефективної маси та температури.

Виділимо питому електропровідність та мобільність в умовах фононного розсіяння із (4.71):

$$\sigma_{n\phi} = en \frac{const'}{m_n^{*5/2} T^{3/2}}, \quad (4.72)$$

$$\mu_{n\phi} = \frac{const'}{m_n^{*5/2} T^{3/2}}. \quad (4.72a)$$

В умовах переважно іонного розсіяння час релаксації, відповідно до (4.61), $\tau_{ni}(k) = \tau_{0i} k^3 m_n^{*-1} N_i^{-1}$ і після підстановки цього значення у інтегральну частину (4.69) отримаємо табличний інтеграл:

$$\int_0^{+\infty} k^7 e^{-\alpha k^2} dk = \frac{3}{\alpha^4} = \frac{48 m_n^{*4} (k_B T)^4}{\hbar^8}. \quad (4.73)$$

Далі діємо таким же чином, як і у випадку фононного розсіяння, і для щільності електронного струму отримаємо:

$$\bar{j}_n = en const'' m_n^{*3/2} T^{3/2} N_i^{-1} \bar{\mathcal{E}} = en \mu_{ni} \bar{\mathcal{E}}. \quad (4.74)$$

де у $const''$ зібрані усі постійні величини окрім ефективної маси, температури та концентрації іонізованих донорів.

Виділимо питому електропровідність та мобільність в умовах іонного розсіяння із (4.74):

$$\sigma_{ni} = en const'' m_n^{*3/2} T^{3/2} N_i^{-1}, \quad (4.75)$$

$$\mu_{ni} = const'' m_n^{*3/2} T^{3/2} N_i^{-1}. \quad (4.75a)$$

Повну мобільність визначаємо із урахуванням (4.34):

$$\mu_n^{-1} = \mu_{n\phi}^{-1} + \mu_{ni}^{-1}. \quad (4.76)$$

Залежність електропровідності домішкового напівпровідника $\sigma(T, N_{\text{дом}})$ від температури і рівня легування $N_{\text{дом}}$ є результатом суперпозиції залежностей $n(T)$ (рис.3.6), $\mu(N_{\text{дом}})$ (рис.4.3) та $\mu(T, N_{\text{дом}})$ (рис.4.4):

$$\sigma(T, N_{\text{дом}}) = e n(T, N_{\text{дом}}) \mu(T, N_{\text{дом}}). \quad (4.76)$$

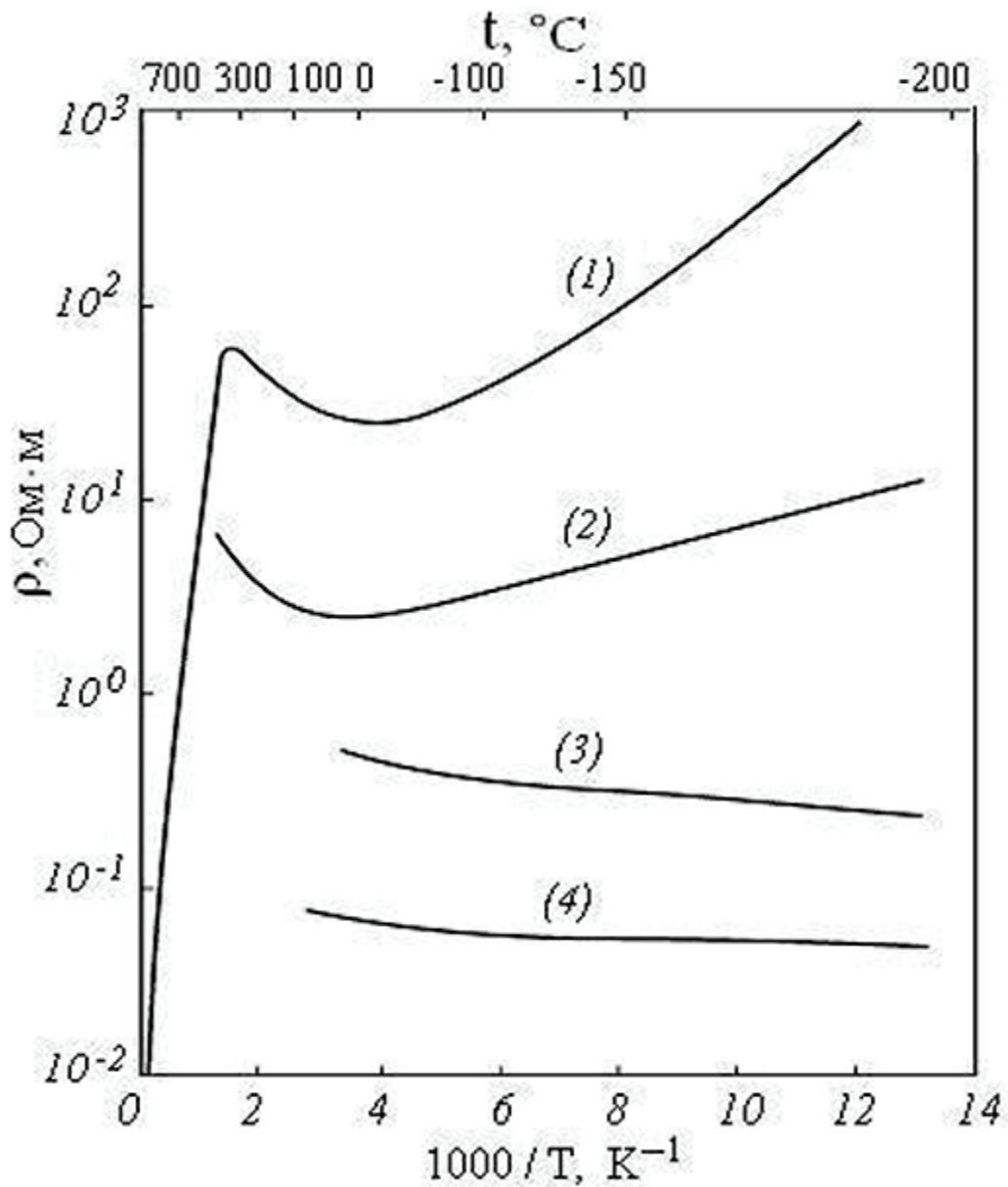


Рис.4.5. Залежність питомого опору кремнію з різним вмістом фосфору від температури. $N_d = 1 \cdot 10^{18}$ (1), $3 \cdot 10^{21}$ (2), $4,7 \cdot 10^{23}$ (3), $1 \cdot 10^{24}$ (4) м^{-3}

Питомий опір напівпровідника:

$$\rho(T, N_{\text{дом}}) = (\sigma(T, N_{\text{дом}}))^{-1} = (e n(T, N_{\text{дом}}) \mu(T, N_{\text{дом}}))^{-1}. \quad (4.76a)$$

Експериментальна залежність $\rho(T, N_{\text{дом}})$ подана на рис.4.5 у координатах $\ln \rho = f(1/T)$.

Питомий опір невиродженого напівпровідника ($N_d < 10^{24} \text{ м}^{-3}$) зменшується при збільшенні температури, але не монотонно (рис.4.5,(1,2)). У діапазоні температур області виснаження донорів ($n = N_d$) залежність $\rho(T)$ визначається залежністю $\mu(T)$ і проходить через мінімум при температурах 150 – 250 К. У діапазоні 250 – 350 К для більшості напівпровідників із $E_g > 0,5$ еВ питомий опір зростає при збільшенні температури внаслідок посилення фононного розсіювання. Починаючи із 350 К питомий опір зменшується по сильному експонентному закону (3.29) за рахунок зростання $n_i(T)$.

Питомий опір виродженого напівпровідника ($N_d > 10^{24} \text{ м}^{-3}$) монотонно зростає в усьому діапазоні температур починаючи від $T \geq 0$ (рис.4.5 ,(3,4)).

Резюме по 4.9.

1. За допомогою кінетичного рівняння Больцмана визначені щільність електронного струму, питома провідність та мобільність електронів для невиродженого напівпровідника із простою зонною структурою.
2. Отримані залежності питомої провідності, питомого опору та мобільності від температури, рівня легування та ефективної маси носіїв заряду.
3. Фізична інтерпретація залежностей $n(T)$, $\mu(T)$ та співставлення їх із експериментальними даними були подані у пунктах 4.6, 4.7, 4.8 і використані при аналізі залежності питомого опору від температури.
4. Отримана при розв'язку кінетичного рівняння Больцмана температурна залежність питомого опору у цілому відповідає експериментальним даним.
5. Використана методика розв'язку кінетичного рівняння для електричного поля є типовою і для інших зовнішніх силових полів (градієнт температури, сила Лоренця і т.п.).

4.10. Термоелектричні ефекти

Термоелектричні ефекти Зеєбека, Пельтьє та Томсона спостерігаються за таких зовнішніх умов:

- 1) однорідний напівпровідник і градієнт температури:

$$\nabla F_{\phi}|_{\nabla T=0} = 0, \nabla T \neq 0 \text{ (ефект Зеєбека);} \quad (4.77)$$

- 2) однорідний напівпровідник, градієнт температури і електричний струм:

$$\nabla F_{\phi}|_{\nabla T=0} = 0, \nabla T \neq 0, \bar{j} \neq 0 \text{ (ефект Томсона);} \quad (4.78)$$

- 3) неоднорідний напівпровідник і електричний струм:

$$\nabla F_{\phi}|_{\nabla T=0} \neq 0, \bar{j} \neq 0 \text{ (ефект Пельтьє).} \quad (4.79)$$

Ефект Зеєбека. Розглянемо однорідний лінійний напівпровідник, уздовж якого створено градієнт температури ∇T . Позначимо температури на кінцях напівпровідника T_1 і T_2 та прийнемо $T_1 > T_2$. Для зручності будемо умовно називати кінець із температурою T_1 „гарячим”, кінець із температурою T_2 – „холодним” (умовність цих термінів у тому, що і T_1 , і T_2 можуть одночасно бути і більше, і менше 0°C). Різниця температур $T_1 - T_2$ лежить у області виснаження донорів і акцепторів $n = N_d, p = N_a$. Залежність рівня Фермі від температури визначається (3.42) і, відповідно, $F_{\phi}(T_1) > F_{\phi}(T_2)$. Наслідком градієнта вільної енергії буде термодифузійний потік, направлений проти градієнта температури. Відтік електронів з „гарячого” кінця залишить некомпенсованим позитивний заряд донорів і створить надлишковий негативний заряд на „холодному” кінці.

У випадку діркового напівпровідника за тих же умов на „гарячому” кінці буде некомпенсований негативний заряд акцепторів і надлишковий позитивний

заряд на „холодному” . Різниця потенціалів між „гарячим” і „холодним” кінцями викличе дрейфові потоки електронів і дірок, направлені проти дифузійних.

Ситуація із дифузійно – дрейфовими процесами була розглянута у п.3.4. Стаціонарна рівновага для цього випадку визначається рівнянням (3.60) і умовою (3.62).

Ефект Зеебека полягає у тому, що при наявності різниці температур уздовж лінійного провідника (метала або напівпровідника) на його кінцях виникає термоелектрична різниця потенціалів, прямо пропорційна різниці температур, яка може виконувати роботу у зовнішньому колі. Джерело енергії для виконання роботи – градієнт температури.

Визначимо термоелектричну різницю потенціалів за допомогою кінетичного рівняння для однорідного невиродженого електронного напівпровідника із скалярною ефективною масою в умовах фонованого розсіяння.

При складанні нерівноважної функції f_n використаємо той факт, що при фонованому розсіянні довжина вільного пробігу електронів $L_{n\phi} = v\tau_{n\phi}$ не залежить від енергії електронів:

$$\tau_{n\phi} = \frac{L_{n\phi}}{v} = \frac{m_n^* L_{n\phi}}{\hbar} k^{-1}. \quad (4.80)$$

Повний струм \bar{j}_n має дві складові – дифузійну і дрейфову. При відсутності зовнішнього кола $\bar{j}_n = \bar{j}_{n\text{диф}} + \bar{j}_{n\text{др}} = 0$. Із цього рівняння будемо визначати термоелектричну різницю потенціалів $\nabla(\phi - F_\phi / e)$.

Запишемо повний струм:

$$\bar{j}_n = -\frac{e}{4\pi^3} \int_{-\infty}^{+\infty} v f_{1n} d^3 \bar{k}. \quad (4.81)$$

Швидкість \bar{v} визначаємо через хвильовий вектор $\bar{v} = \frac{\hbar}{m_n^*} \bar{k}$. Нерівноважну

добавку до функції розподілу f_{1n} беремо із (4.26), векторну функцію енергії $\bar{\chi}(E)$ – із (4.26б), рівноважну функцію розподілу $f_{0n}(E)$ із (3.20а). Після підстановки цих величин у (4.81) отримаємо:

$$\bar{j}_n = -\frac{e\hbar^2 n}{4\pi^3 k_B T N_c m_n^*} \int_{-\infty}^{+\infty} \tau_{n\phi} \left\{ \frac{E - F_\phi}{T} \nabla T + \nabla(F_\phi - e\varphi) \right\} e^{-\alpha k^2} \bar{k} \cdot \bar{k} d^3 \bar{k}. \quad (4.82)$$

Підставляємо $\tau_{n\phi}$ із (4.80), переходимо в полярну систему координат (полярна вісь ∇T), розділяємо \bar{j}_n на дві складові – $\bar{j}'_{n\text{diff}}$ (пропорційну ∇T) і $\bar{j}'_{n\text{dr}}$ (пропорційну $\nabla(F_\phi - e\varphi)$), доводимо інтеграли до табличного вигляду і у кінцевому підсумку отримаємо:

$$\bar{j}_n = n\mu_{n\phi} \left[\left(2k_B - \frac{F_\phi}{T} \right) \nabla T + \nabla(F_\phi - e\varphi) \right], \quad (4.83)$$

де $\mu_{n\phi}$ - мобільність електронів при фононному розсіянні (4.72а).

Із умови $\bar{j}_n = 0$ знаходимо $\nabla(F_\phi - e\varphi)$ - градієнт електрохімічного потенціалу:

$$\nabla(F_\phi - e\varphi) = -\left(2k_B - \frac{F_\phi}{T} \right) \nabla T. \quad (4.84)$$

Основна характеристика ефекту Зеебека – коефіцієнт термоелектрорушійної сили (термо-ЕРС) α – визначається як термо-ЕРС відповідна одиничному градієнту температури:

$$\alpha \equiv \frac{|\nabla(\varphi - F_\phi/e)|}{|\nabla T|}. \quad (4.85)$$

Розмірність $[\alpha]$ – вольт/на 1 градус різниці температур – $V/\text{град}$.

Для електронного напівпровідника:

$$\alpha_n = -\frac{k_B}{e} \left(2 - \frac{F_\phi}{k_B T} \right) = -\frac{k_B}{e} \left(2 - \ln \frac{n}{N_c} \right) = -\frac{k_B}{e} \left(2 + \ln \frac{N_c}{N_d} \right), \quad (4.86)$$

де F_ϕ взятий із формули (3.42). Від'ємне значення α_n пов'язане з тим, що термоелектричне поле і градієнт температури направлені протилежно.

Аналогічний розрахунок для діркового напівпровідника α_p дає:

$$\alpha_p = \frac{k_B}{e} \left(2 + \frac{F_\phi + E_g}{k_B T} \right) = \frac{k_B}{e} \left(2 - \ln \frac{p}{N_v} \right) = \frac{k_B}{e} \left(2 + \ln \frac{N_c}{N_a} \right). \quad (4.87)$$

У випадку змішаної провідності повний струм є сумою електронного і діркового струмів:

$$\begin{aligned} \bar{j} = n\mu_n \left[\nabla(F_\phi - e\phi) + \left(2k_B - \frac{F_\phi}{T} \right) \nabla T \right] + \\ p\mu_p \left[\nabla(F_\phi - e\phi) - \left(2k_B - \frac{F_\phi + E_g}{T} \right) \nabla T \right]. \end{aligned} \quad (4.88)$$

Прирівнявши повний струм (4.88) нулю, визначимо α для змішаної провідності:

$$\alpha = -\frac{k_B}{e} \left[\frac{(2 + \ln \frac{N_c}{N_d})n\mu_n - (2 - \ln \frac{N_v}{N_a})p\mu_p}{n\mu_n + p\mu_p} \right]. \quad (4.89)$$

Помножимо і поділимо (4.89) на заряд електрона e :

$$\alpha = \frac{\alpha_p \sigma_p - \alpha_n \sigma_n}{\sigma_p + \sigma_n}. \quad (4.90)$$

Для власного напівпровідника ($n = p = n_i$, $F_{\phi i} = -E_g/2$, $b = \mu_n/\mu_p$) (4.90)

має вигляд:

$$\alpha_i = -\frac{k_B}{e} \frac{b-1}{b+1} (2 + E_g/2k_B T). \quad (4.91)$$

Фізичний зміст (4.91) полягає у тому, що при одночасній термодифузії електронів і дірок термоелектричні напруги компенсуються і обов'язковою для ефекту Зеєбека у власному напівпровіднику є умова $b = \mu_n/\mu_p \neq 1$. Зазвичай $\mu_n > \mu_p$, електрони у процесі дифузії випереджають дірки і тому $\alpha_i < 0$.

Для прикладу у табл. 4.2 наведені результати розрахунків α_i при 300 К для деяких напівпровідників.

Таблиця 4.2

Напівпровідник	<i>i</i> -Ge	<i>i</i> -Si	<i>i</i> -GaAs
E_g , eV	0,66	1,12	1,42
b	2	2,5	21,25
α_i , мВ/град	– 0,057	– 0,97	– 2,1
α_n , мВ/град	– 1,1	– 1,8	– 2,1
α_p , мВ/град	+ 1,0	+ 1,7	+2,5
$ \alpha_0 $, мВ/град	0,172		

Величина α залежить від концентрації носіїв заряду. Загальну закономірність такої залежності можна визначити із умови стаціонарності – термодифузійний потік врівноважений зустрічним дрейфовим потоком, зумовленим термоелектричним полем $\bar{\mathcal{E}}_{терм} = \alpha \nabla T$. Відповідно до (3.57), (3.58) умова стаціонарності для електронного напівпровідника буде така:

$$D_n \nabla n = n \mu_n \bar{\mathcal{E}}_{терм}. \quad (4.92)$$

При фіксованому градієнті температури i , відповідно, градієнті вільної енергії електронів ліва частина (4.92) буде фіксованою. Права частина також фіксована при певному значенні $n \bar{\mathcal{E}}_{терм}$. Наслідком (4.92) є загальна закономірність: при зростанні n повинно зменшуватися $\bar{\mathcal{E}}_{терм}$ і навпаки. Аналогічна закономірність має місце і для діркового напівпровідника.

Якщо концентрація носіїв заряду визначається рівнями легування донорами або акцепторами, то на верхній межі невинродженого стану напівпровідника ($N_d \approx N_c$, $N_a \approx N_v$) логарифмічні доданки у (4.86), (4.87) близькі до 0 і α_0 має значення $|\alpha_0| \approx 2 \frac{k_B}{e} = 0,172$ мВ/град. Нижню межу

визначимо таким чином, щоб зберігався певний тип провідності: $N_d \approx 10n_i$, $N_a \approx 10n_i$. Результати розрахунків α_n і α_p по (4.86), (4.87) наведені у табл.4.2. Оскільки різниця між α_n і α_p у цьому випадку незначна, об'єднаємо їх у одне значення $\alpha_n \approx \alpha_p = |\alpha_{\max}|$.

Порівнявши α_i із $|\alpha_{\max}|$ можна помітити, що суттєва різниця між ними має місце лише для германію, у якого відносно малі значення E_g (відносно висока n_i) і b (незначна перевага дифузійних параметрів електронів над дірками). Термоелектронні поля електронів і дірок у германії частково компенсовані. У арсеніду галію майже удвічі більша E_g і майже у 10 разів більше b . Наслідком цього є те, що $\alpha_i \approx |\alpha_{\max}|$ і вплив неосновних носіїв заряду на термоелектричні процеси незначний. Кремній займає проміжне становище.

При підвищенні температури до верхньої межі області насичення T_i (3.45) і подальшого переходу до власної провідності у дірковому напівпровіднику повинен змінитися знак термо-ЕРС, α_p має пройти через 0. Це, у свою чергу, означає повну компенсацію термоелектронних полів $\alpha_n = \alpha_p$. Підставивши цю умову у (4.90) отримаємо додаткову умову $p\mu_p = n\mu_n$. Фізичний зміст цієї умови такий: менша мобільність дірок $\mu_p < \mu_n$ повинна компенсуватися більшою їх концентрацією $p > n$ при переході α_p через 0.

Зміна знаку термо-ЕРС при підвищенні температури – процес, до певної міри, керований рівнем легування, тому що T_i зростає зі збільшенням N_a . Цей ефект може мати технічне застосування у ключових датчиках температури.

У металів концентрація електронів на декілька порядків більша, ніж у навіть вироджених напівпровідників, лише за рахунок цього $\alpha_{мет}$ повинен бути

значно меншим у порівнянні із напівпровідниками. Але є ще дві обставини, які негативно впливають на коефіцієнт термо–ЕРС металів: перша – рівень Фермі у металів дуже слабо змінюється з температурою відносно свого значення при $T=0$ і створити помітний градієнт вільної енергії не вдається, друга – висока електронна теплопровідність металів, яка не дає змоги створити помітний стаціонарний градієнт температури. *Наслідком перерахованих особливостей металів є те, що $\alpha_{мет} \approx 1...10$ мкВ/град, в той час як для напівпровідників він досягає значень $\alpha \approx 100...2000$ мкВ/град (див. табл. 4.2).*

Завдяки такій великій різниці коефіцієнтів термо – ЕРС металеві контакти до напівпровідника практично не впливають на α напівпровідників.

Формули (4.87), (4.88), (4.89) визначають коефіцієнт термо-ЕРС в умовах переважно фононного розсіяння. Розрахунок α в умовах переважно іонного розсіяння при $T < 200$ К необхідно проводити із урахуванням залежності довжини вільного пробігу від енергії носіїв заряду (4.63а). Зміну механізму розсіяння можна врахувати у формулах для α введенням додаткового числового доданку у дужках, наприклад:

$$\alpha_n = -\frac{k_B}{e} \left(2 + r + \ln \frac{N_c}{N_d} \right), \quad (4.92)$$

де r – показник степеня у залежності довжини вільного пробігу від енергії (4.63а).

Ефекти Пельтьє і Томсона. Відповідно до (4.78) і (4.79), напівпровідник знаходиться під дією як поля термо–ЕРС, так і зовнішнього поля. Ці поля можуть бути паралельними або антипаралельними. Результатом взаємодії полів буде або посилення, або послаблення дрейфової складової повного струму за рахунок енергії зовнішнього поля. У полі градієнта температури буде відбуватися одночасне перенесення і заряду, і енергії.

Визначимо повну енергію у розрахунку на один електрон як суму кінетичної $E = \hbar^2 k^2 / 2m_n^*$ і потенціальної енергії $-e\phi$, де ϕ тепер включає

потенціали зовнішнього поля і поля термо–ЕРС. Потужність потоку енергії \bar{W}_n , яку переносять усі електрони,:

$$\bar{W}_n = \frac{1}{4\pi^3} \int_{-\infty}^{+\infty} f_{1n}(E - e\phi) \bar{v} d^3 \bar{k}. \quad (4.93)$$

Розділимо (4.93) на дві складові:

$$\bar{W}_n = \frac{1}{4\pi^3} \int_{-\infty}^{+\infty} f_{1n} E \bar{v} d^3 \bar{k} - \frac{e\phi}{4\pi^3} \int_{-\infty}^{+\infty} f_{1n} \bar{v} d^3 \bar{k}. \quad (4.94)$$

Другий інтеграл у (4.94), із урахуванням (4.81), дорівнює $\phi \bar{j}_n$. Перший інтеграл у (4.94) відрізняється від (4.81) лише додатковим множником k^2 і зводиться до табличного інтегралу. Провівши відповідне інтегрування, для щільності потоку енергії отримаємо:

$$\bar{W}_n = \phi \bar{j}_n - \frac{2k_B T}{e} n \mu_{n\phi} \left\{ \frac{3k_B T - F_\phi}{T} \nabla T + \nabla(F_\phi - e\phi) \right\}. \quad (4.95)$$

Знайдемо $\nabla(F_\phi - e\phi)$ із (4.83) і підставимо у (4.95):

$$\bar{W}_n = \left(\phi - \frac{F_\phi}{e} \right) \bar{j}_n - \frac{2k_B T - F_\phi}{e} \bar{j}_n - \frac{2k_B^2 T}{e} n \mu_{n\phi} \nabla T. \quad (4.96)$$

Потужність потоку (4.96) має три складові:

1) потужність, яка зумовлена роботою зовнішнього поля і термоелектричного поля по переміщенню електронного заряду (теплота Джоуля);

2) потужність, яка зумовлена додатковим полем, пов'язаним із різницею теплової енергії електронів по об'єму провідника (теплота Пельтьє);

3) складова, пропорційна градієнту температури – теплопровідність.

Перша складова потоку існує незалежно від того чи є градієнт температури, чи його немає, чи рівномірно по об'єму легований провідник, чи ні. Цей енергетичний потік супроводжується виділенням теплоти Джоуля та нагріванням провідника, що є безпосереднім наслідком електрон – фононного розсіяння.

Друга складова потоку прямо пропорційна струму. Коефіцієнт пропорційності Π називається **коефіцієнтом Пельтьє**. Із урахуванням (4.86), його можна записати таким чином:

$$\Pi = -2 \frac{k_B}{e} T + \frac{F_\phi}{e} = -\frac{k_B}{e} \left(2 - \frac{F_\phi}{k_B T} \right) T = \alpha_n T. \quad (4.97)$$

Розмірність коефіцієнта Пельтьє $[\Pi] = V$. Взаємозв'язок між Π і α_n , а саме $\Pi = \alpha_n T$, отримав назву **друге термодинамічне співвідношення Томсона**.

Третя складова – електронна теплопровідність – характеризується **коефіцієнтом електронної теплопровідності χ_e** :

$$\chi_e = \frac{2k_B^2 T}{e} n \mu_{n\phi} = 2 \left(\frac{k_B}{e} \right)^2 \sigma_n T. \quad (4.98)$$

Із (4.98) знайдемо χ_e / σ_n :

$$\frac{\chi_e}{\sigma_n} = 2 \left(\frac{k_B}{e} \right)^2 T. \quad (4.99)$$

Формула (4.99) відома як **закон Відемана – Франца**: відношення електронної теплопровідності до електропровідності при температурі T величина постійна. У металів теплопровідність переважно електронна і закон Відемана – Франца діє у повну силу. Для напівпровідників теплопровідність переважно фоновна і закон Відемана – Франца стосується лише електронної компоненти повної теплопровідності.

У кінцевому підсумку, формула для потоку енергії (4.96) має такий вигляд:

$$\bar{W}_n = \left[(\varphi - F_\phi / e) - \Pi \right] \bar{j}_n - \chi_e \nabla T. \quad (4.100)$$

Визначимо потужність, що виділяється в одиниці об'єму за одиницю часу як дивергенцію від потоку, врахувавши при цьому стаціонарність процесу $\text{div } \bar{W}_n = \text{div } \bar{j}_n = 0$. Із векторної алгебри відома формула для дивергенції від

добутку скаляра на вектор: $\text{div}(\phi \bar{a}) = \phi \text{div} \bar{a} + \bar{a} \nabla \phi$. У (4.100) скалярні величини $\phi \rightarrow \varphi, F_\phi, \chi_e, \Pi$; векторні $\rightarrow \bar{j}_n, \nabla T$. Потужність тепловиділення:

$$\text{div}(-\chi_e \nabla T) = \bar{j}_n \left[\nabla \Pi + \nabla \left(\frac{F_\phi}{e} - \varphi \right) \right]. \quad (4.101)$$

У формулу повного струму \bar{j}_n (4.83) внесемо α_n (4.86), $\sigma_n = e n \mu_n$, а також Π (4.86):

$$\bar{j}_n = \sigma_n \left[\alpha_n \nabla T + \nabla \left(\frac{F_\phi}{e} - \varphi \right) \right] = \sigma_n \frac{\Pi}{T} \nabla T + \sigma_n \nabla \left(\frac{F_\phi}{e} - \varphi \right). \quad (4.102)$$

Із (4.102) визначаємо градієнт електрохімічного потенціалу $\nabla \left(\frac{F_\phi}{e} - \varphi \right)$ і підставляємо у (4.101). Після простих перетворень маємо:

$$\text{div}(-\chi_e \nabla T) = \frac{j_n^2}{\sigma_n} \pm T \frac{d}{dT} \left(\frac{\Pi}{T} \right) (\bar{j}_n \nabla T) = \frac{j_n^2}{\sigma_n} \pm T \frac{d\alpha_n}{dT} (\bar{j}_n \nabla T). \quad (4.103)$$

*Фізичний зміст (4.103): якщо у однорідному провіднику створений градієнт температури і тече електричний струм, то, окрім теплоти Джоуля j_n^2 / σ_n , буде виділятися або поглинатися теплота Томсона, пропорційна струму і градієнту температури – **ефект Томсона**. Максимальне значення теплота Томсона має, коли \bar{j}_n і ∇T паралельні або антипаралельні. Якщо вимушений рух носіїв заряду, створений зовнішньою різницею потенціалів, відбувається проти градієнта температури – від „гарячого” кінця до „холодного” ($\bar{j}_n \uparrow \uparrow \nabla T$), носії заряду віддають надлишкову теплову енергію тепловим коливанням атомів, підвищуючи тим самим температуру „холодного” кінця (знак теплоти Томсона „+”). При інверсії знаку зовнішньої різниці потенціалів рух носіїв заряду буде співпадати із напрямком градієнта ($\bar{j}_n \downarrow \uparrow \nabla T$) і вони будуть підвищувати свою теплову енергію за рахунок теплових коливань атомів на „гарячому” кінці, тим самим охолоджуючи його*

(знак теплоти Томсона „–“). Зовнішнім проявом ефекту Томсона, за будь-яких умов, є зміна градієнта температури уздовж провідника за рахунок теплоти Томсона.

Коефіцієнт пропорційності τ_T :

$$\tau_T = T \frac{d\alpha_n}{dT} \quad (4.103)$$

називається **коефіцієнтом Томсона**, а його зв'язок із α_n (4.103) – **першим термодинамічним співвідношенням Томсона**.

Зміна градієнту температури, яка супроводжує ефект Томсона, умовно кажучи, „розмита” по всьому об'єму провідника. Виділити теплоту Томсона на фоні теплоти Джоуля у цьому випадку однорідного провідника практично неможливо.

Розглянемо ситуацію, коли струм тече через контакт двох послідовно з'єднаних однорідних провідників. Провідники різняться рівнями Фермі (наприклад, $F_{\phi 1} > F_{\phi 2}$) і, відповідно до (4.97), коефіцієнтами Пельтьє ($\Pi_1 > \Pi_2$) та коефіцієнтами теплопровідності ($\chi_{e1} \neq \chi_{e2}$). Зовнішній градієнт температури відсутній, тому різниця рівнів Фермі і коефіцієнтів Пельтьє зумовлена, у більшості випадків, різними рівнями легування напівпровідників або різною роботою виходу контактуючих матеріалів.

У стані термодинамічної рівноваги рівні Фермі $F_{\phi 1}$ і $F_{\phi 2}$ у зоні контакту зрівнюються контактною різницею потенціалів таким чином, що електрохімічний потенціал $(F_{\phi} - e\varphi)$ стає однаковим для обох провідників, тобто $\nabla(F_{\phi} - e\varphi) = 0$. Струм \bar{j}_n також однаковий у колі послідовного з'єднаних провідників і лише коефіцієнт Пельтьє стрибком змінюється в зоні контакту

($\nabla \Pi = \Pi_1 - \Pi_2$). Із урахуванням цих обставин формула для потужності (4.101) набуває такого вигляду:

$$\text{div}(-\chi_e \nabla T) = \pm(\Pi_1 - \Pi_2) \bar{j}_n. \quad (4.104)$$

Фізичний зміст (4.104): при протіканні струму через контакт двох провідників із різними рівнями Фермі у зоні контакту буде виділятися або поглинатися теплота Пельтьє, пропорційна різниці коефіцієнтів Пельтьє і струму – **ефект Пельтьє**. Носії заряду, які перетинають зону контакту, повинні змінювати свою вільну енергію – або зменшувати, або збільшувати у залежності від напрямку струму. Якщо при перетині контакту вільна енергія виявиться надлишковою, вона буде нагрівати зону контакту, трансформуючись у енергію теплових коливань атомів. Якщо ж при перетині контакту вільну енергію необхідно підвищувати, носії заряду будуть відбирати її від енергії теплових коливань атомів і зона контакту буде охолоджуватися. Таким чином, протяжність зони контакту, у якій відбувається зміна вільної енергії, відповідає вільному пробігу носії заряду при фононному розсіянні.

На відміну від теплоти Томсона, теплота Пельтьє зосереджена у зоні контакту і це дає змогу практично використати ефект Пельтьє, у першу чергу, для термоелектричного охолодження.

Протяжність зони контакту по напрямку струму – довжина вільного пробігу при фононному розсіянні (4.43а) – близько 0,05 мкм. Розходження теплового потоку (ліва частина (4.104)) на такій малій відстані можна записати як різницю електронних теплових потоків:

$$-\chi_{e1}(\nabla T)_1 + \chi_{e2}(\nabla T)_2 = \pm(\Pi_1 - \Pi_2)\bar{j}_n. \quad (4.105)$$

Фізичний зміст (4.105): теплова енергія, яка поглинається або виділяється в зоні контакту, двома протилежно направленими тепловими потоками відводиться від зони контакту при охолодженні або поповнюється при нагріванні за рахунок електронної теплопровідності.

Існування цих потоків необхідне для стаціонарності теплових процесів і це вимагає підтримання стабільного значення $(\nabla T)_1$ та $(\nabla T)_2$. Теплота Джоуля виділяється у всьому об'ємі неоднорідного провідника і зменшує градієнти

температури, створені ефектом Пельтьє, і її необхідно включити у рівняння теплового балансу:

$$Q = (j_n^2 / \sigma_n) \pm (P_1 - P_2) \bar{j}_n, \quad (4.106)$$

де Q – загальна потужність теплоти Джоуля і теплоти Пельтьє, яка пов’язана із струмом у неоднорідному провіднику і яку необхідно розсіювати у зовнішнє середовище.

Зважаючи на те, що загальна потужність (4.106) має різну функціональну залежність своїх складових від струму, існує оптимальне значення струму, при якому теплота Пельтьє перевищує теплоту Джоуля у зоні контакту. Значення оптимального струму значною мірою залежить і від фононної теплопровідності напівпровідника, оскільки у полі градієнта температури, створеного ефектом Пельтьє, виникає потік теплоти Джоуля до зони контакту. Особливо небажаним цей потік стає у випадку, коли у зоні контакту відбувається поглинання теплоти, тобто у режимі охолодження контакту.

Для врахування вищевказаних обставин при виборі напівпровідникового матеріалу введено поняття *термоелектричної якості (добротності) Z* :

$$Z = \frac{\text{питома термоелектрична потужність}}{\text{питома потужність теплопровідності}} = \frac{\alpha^2 \sigma}{\chi_\phi + \chi_e}. \quad (4.107)$$

Розмірність $[Z]$ – 1/градус різниці температур.

Фізичний зміст (4.107): термоелектрична складова загальної потужності повинна бути якомога більшою по відношенню до потужності теплопровідності, яка зумовлена фононною χ_ϕ і електронною χ_e складовими.

Розглянемо чисельник (4.107): коефіцієнт термо – ЕРС α залежить від рівня легування (4.86), (4.87) – чим більше N_d або N_a , тим менше α . У той же час, σ навпаки – чим більші N_d або N_a , тим більша σ . Із умови екстремуму $\partial(\alpha^2 \sigma) / \partial N_{\text{дом}} = 0$ знаходимо компромісні значення рівнів

легування: $N_d = N_c$, $N_a = N_v$. У цьому випадку, відповідно до таб.4.2, $\alpha = \alpha_0 = 172$ мкВ/град.

У знаменнику (4.107) основною складовою є фононна теплопровідність $\chi_\phi \gg \chi_e$. Таким чином, високо добротними матеріалами будуть такі, що мають $\alpha \approx 172$ мкВ/град і невисоку фононну теплопровідність. За інших однакових обставин, фононна теплопровідність напівпровідникового матеріалу тим менша, чим більша атомна маса елементів, які входять до його складу.

Найкращими термоелектричними матеріалами виявилися потрійні сполуки (тверді розчини) на основі на основі Bi_2Te_3 легovanого Se для n – типу провідності, Sb – для p – типу. У таб.4.3 наведені типові термоелектричні параметри таких сполук.

Дані таб.4.3 свідчать про те, що вказані матеріали мають оптимальне значення коефіцієнта термо–ЕРС α . Якщо порівняти їх питому теплопровідність χ із теплопровідністю „класичних” напівпровідників, яка має значення 0,64 Вт/см·град для Ge, 1,45 Вт/см·град для Si, 0,46 Вт/см·град для GaAs, можна зробити висновок, що вона на 1...2 порядки менша. Наслідком такого співвідношення теплопровідностей є те, що „класичні” напівпровідники не знайшли широкого застосування в якості термоелектричних матеріалів.

Таблиця 4.3

Матеріал	Тип провідн.	α , мкВ/град.	$\sigma \cdot 10^{-3}$, Ом ⁻¹ ·см ⁻¹	χ , Вт/см·град	$Z \cdot 10^3$, град ⁻¹
$\text{Bi}_2\text{Te}_3(85)+\text{Bi}_2\text{Se}_3(15)$	n	170	1,0	0,012	3,0
$\text{Bi}_2\text{Te}_3(20)+\text{Sb}_2\text{Te}_3(80)$	p	174	1,54	0,016	2,8

Термоелемент Пельтьє. Для практичного застосування використовують термоелемент Пельтьє: p – і n – гілки у електричному колі з’єднують металевими пластинами послідовно, а по тепловим потокам – паралельно.

Якщо з'єднують таким же чином декілька термоелементів Пельтьє, утворюється *термоелектричний модуль*. Кількість термоелементів у модулі може бути від 50 до 1000.

Слід звернути увагу на універсальність елементів Пельтьє як перетворювачів енергії, а саме:

–якщо створити різницю температур $|T_1 - T_2| = \Delta T \neq 0$, між спаями утвориться термодинамічно нерівноважна різниця потенціалів, здатна виконувати роботу у зовнішньому електричному колі (режим термоелектричного генератора);

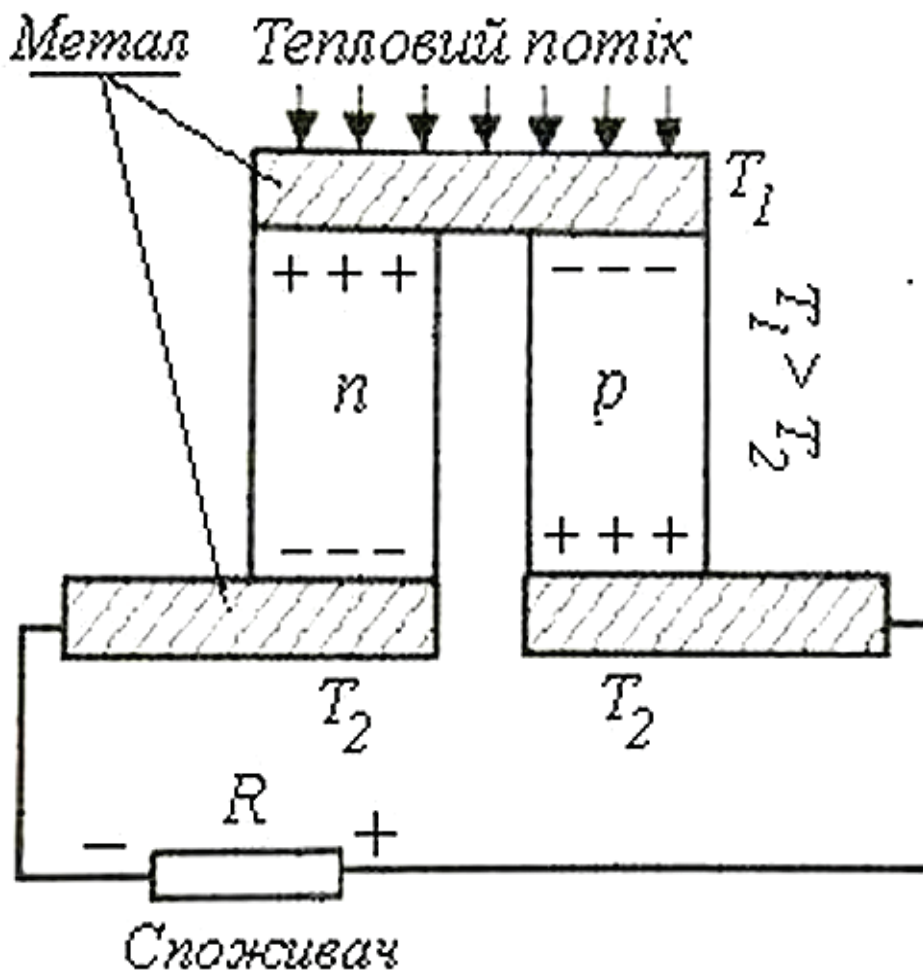


Рис.4.6. Елемент Пельтьє у режимі термоелектричного перетворювача теплової енергії у електричну енергію

– при інверсії напрямку градієнта температури інвертується і знак різниці потенціалів;

– у режимі термоелектричного генератора, при інших рівних умовах, ефект Пельтьє дає приблизно удвічі більшу різницю потенціалів $V_{II} = (\alpha_n + \alpha_p)(T_1 - T_2)$, ніж ефект Зеєбека $V_Z = \alpha_n(T_1 - T_2)$;

– за наявності зовнішнього джерела електричного струму елемент

Пельтьє перетворює енергію струму у теплову енергію у вигляді різниці температур між умовно "холодним" T_2 і "гарячим" T_1 спаями $|T_1 - T_2| = \Delta T$ (режим термоелектричного охолоджувача);

– при інверсії напрямку струму "гарячий" T_1 і "холодний" T_2 спаї міняються місцями (режим термоелектричного нагрівача).

Основна проблема термоелектричного охолоджувача – утримання стабільної температури "гарячого" спаю $T_1 = const$ за рахунок розсіяння тепла у зовнішнє середовище.

Сучасні термоелементи здатні створити різницю температур $\Delta T_{\max} = 60 \dots 70$ °С. Якщо T_1 підтримувати на рівні $20 \dots 25$ °С, на "холодному" спаї будемо мати $-40 \dots -50$ °С. Але є обмеження для мінімальної від'ємної температури на "холодному" спаї на рівні $-80 \dots -90$ °С і пов'язане воно, головним чином, із тим, що при зниженні температури суттєво зростає фононна теплопровідність, зменшується термоелектрична добротність Z (див. (4.107)).

Енергетичний коефіцієнт корисної дії (ККД) термоелементів Пельтьє визначається термоелектричною добротністю Z і перепадом температур між спаями. Для прикладу, для $\Delta T = 70$ °С ККД $\approx 3\%$.

4.11. Гальваномагнітні ефекти

Загальні положення. Вільний носій заряду (електрон або дірка) під впливом постійного магнітного поля B_0 у відсутності розсіяння сформує замкнену траєкторію у площині, перпендикулярній B_0 . Параметри цієї траєкторії: циклотронна частота обертання $\omega_c = eB_0 / m^*$, радіус орбіти $r = v / \omega_c$ (v - середня лінійна швидкість), період обертання $T = 2\pi / \omega_c$.

У стані термодинамічної рівноваги усі напрямки лінійної швидкості рівно ймовірні, магнітні моменти кільцевих струмів замкнених орбіт компенсуються. Крім того, при утворенні ковалентних зв'язків спіни електронів паруються і сумарний спіновий момент практично дорівнює нулю. Дія зовнішнього магнітного поля може, при достатній величині індукції, порушити спінову компенсацію. Це явище справді спостерігається як парамагнетизм електронного газу в металах і напівпровідниках. Однак зміна енергії електронів у зовнішньому магнітному полі при цьому незначна.

З урахуванням вище вказаного, можна вважати, що дія магнітного поля зводиться до повороту проекцій вектора лінійної швидкості \vec{v} у площині, перпендикулярній B_0 без зміни середнього значення модуля швидкості.

Якщо урахувати процеси розсіяння (час вільного пробігу обмежений), то можна виділити такі ситуації:

- а) час вільного пробігу τ (час релаксації) більший періоду обертання $\tau \gg T$,
- б) час релаксації менший періоду обертання $\tau \ll T$.

При фіксованому середньому значенні τ випадок а) відповідає сильним магнітним полям, випадок б) – слабким.

Критерії сильного і слабого магнітного поля можна записати у такому вигляді (із урахуванням (3.56)):

$$\omega_c \tau = \mu B_0 \gg 1 - \text{сильне магнітне поле;} \quad (4.108)$$

$$\omega_c \tau = \mu B_0 \ll 1 - \text{слабке магнітне поле.} \quad (4.109)$$

З фізичної точки зору критерій сильного магнітного поля (4.108) означає, що за час вільного пробігу носії заряду встигає зробити декілька повних обертів у магнітному полі, орбіта замкнена.

У випадку слабого магнітного поля (4.109) за час релаксації формується лише частина повного оберту, тобто орбіта виявляється розімкненою.

Виберемо реально досяжне значення індукції $B_0 = 1$ Тл. У цьому випадку для критичного значення поля $\mu B_0 = 1$ необхідна мобільність $\mu = 1 \text{ м}^2/\text{В} \cdot \text{с}$. У діапазоні кімнатних температур таке значення мобільності можуть мати лише електрони у деяких сполуках $A^{\text{III}}B^{\text{V}}$ (наприклад: антимоніди індію, арсеніди індію, арсеніди галію). У інших напівпровідників мобільність $\mu \leq 0.4 \text{ м}^2/\text{В} \cdot \text{с}$. У металів мобільність електронів $\mu \leq 0.006 \text{ м}^2/\text{В} \cdot \text{с}$. Таким чином, виконати умови критерію сильного поля (4.108) при кімнатних температурах можна лише для деяких напівпровідників із високою мобільністю електронів. У інших напівпровідників і, тим більше, металів критерій сильного поля при кімнатних температурах не виконується.

Розглянемо рух носіїв заряду в умовах одночасної дії на них постійних електричного і магнітного полів. В прямокутній системі координат направимо електричне поле \mathcal{E}_x по вісі x , магнітне поле B_{0z} - по вісі z . Магнітне поле слабе. Розмір напівпровідника по вісі y необмежений. Відповідна ситуація показана на рис.4.7.

За відсутності розсіяння траєкторія носія заряду формується складним рухом – обертанням в магнітному полі у площині xu і дрейфом центру кола уздовж вісі x зі швидкістю $v_x = \mu \mathcal{E}_x$. Траєкторія цього складного руху – циклоїда (означена пунктиром на рис.4.7).

В умовах розсіяння і слабого магнітного поля формується лише частина циклоїди, далі відбувається розсіяння, втрачається дрейфова швидкість і все починається спочатку. Кінцевий підсумок – траєкторія складається з

криволінійних відрізків, носії рухаються під деяким кутом до напрямку поля \mathcal{E}_x , повний струм має дві компоненти J_x і J_y (тут і далі мається на увазі щільність струму). Поява компоненти J_y зумовлена дією магнітної сили:

$$F_y = e[\bar{v}\bar{B}_0] = e v_x B_{0z}, \quad (4.110)$$

Силі F_y відповідає деяка швидкість дрейфу v_y , яку знаходимо із рівняння руху:

$$\frac{\partial v_y}{\partial t} m^* = F_y = e v_x B_{0z}. \quad (4.111)$$

Розділивши змінні і інтегруючи в межах $0 \leq t \leq \tau$, отримаємо:

$$v_y = v_x \mu B_{0z}. \quad (4.112)$$

Із (4.112) випливає геометричний зміст μB_{0z} :

$$\mu B_{0z} = \operatorname{tg} \theta, \quad (4.112a)$$

де θ – кут Холла.

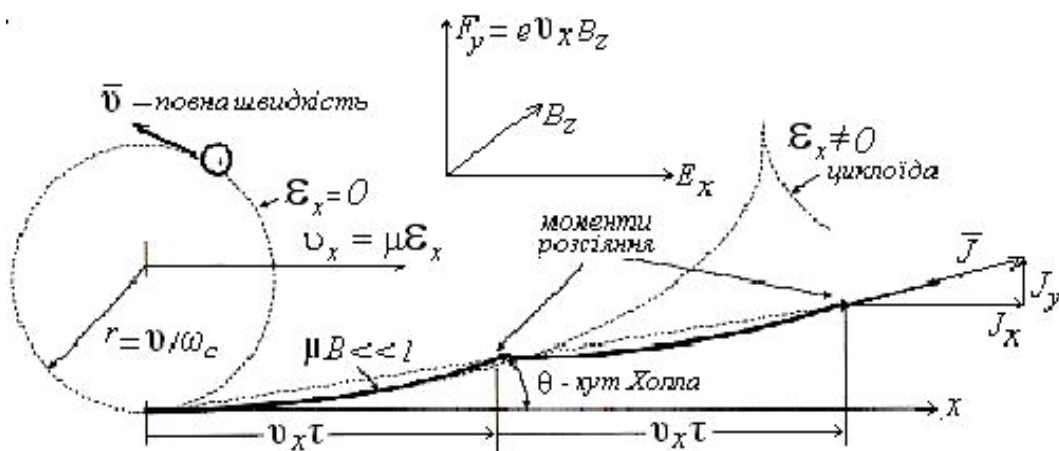


Рис.4.7. Носії заряду у скрещених магнітному та електричному полях. У відсутності електричного поля траєкторія у сильному магнітному полі - замкнене коло у площині XU . За наявності електричного поля і слабого магнітного поля траєкторія – циклоїдальні відрізки. Повний струм у необмеженому по висі Y зразку не співпадає із напрямком зовнішнього поля. Виникає Y компонента струму

Проекція струму J_x на напрямок зовнішнього електричного поля менша повного струму \bar{J} . Це положення можна уявити як дію деякої “гальмівної” сили F_x , що зменшує цей струм, і рівняння руху буде:

$$F_x = e v_y B_z = - \frac{\partial v'_x}{\partial t} m^* . \quad (4.113)$$

Введенням сили F_x , незалежної від швидкості носіїв заряду, спрощено враховується дія процесів розсіяння.

Розв'язуючи (4.113) таким же чином, як і (4.111), отримаємо:

$$v'_x = -v_x (\mu B_z)^2 \quad (4.114)$$

Для струму J_x (саме цей струм вимірюється в експериментах) отримаємо:

$$J_x = en(v_x - v'_x) = en v_x (1 - (\mu B_z)^2), \quad (4.115)$$

де n – концентрація носіїв заряду.

Розглянемо тепер зразок, що має обмежені розміри по вісі y . Якщо зовнішнє електричне коло розімкнено, тобто $J_y = 0$, дія магнітної сили F_y (4.110) призводить до порушення нейтральності за рахунок зміщення носіїв заряду відносно домішкових іонів у напівпровідниках або відносно іонів металу. На зовнішніх електродах з'явиться термодинамічно нерівноважна різниця потенціалів V_y такої величини, щоб сила електричного поля $-e\mathcal{E}_y$ врівноважувала магнітну силу:

$$e v_x B_z = -e\mathcal{E}_y = e V_y / a_y, \quad (4.116)$$

де a_y – розмір зразка по вісі y , \mathcal{E}_y – напруженість електричного поля.

При компенсації магнітної сили електричною (4.116) траєкторії носіїв заряду не викривлюється (якщо всі носії мають однакову швидкість), тобто напрямок струму в зразку співпадає із напрямком зовнішнього поля \mathcal{E}_x , але електричне поле $\bar{\mathcal{E}}$ в об'ємі не співпадає із зовнішнім за рахунок поперечної компоненти \mathcal{E}_y .

Для спрощення прийнято, що усі носії заряду мають однакову середню швидкість \bar{v} і, відповідно, на них діє однакова магнітна сила $\bar{F}_B = e[\bar{v}\bar{B}]$. Таке спрощення справедливе лише для металів і вироджених напівпровідників.

Для невідроджених напівпровідників діє максвеллівський розподіл носіїв заряду по швидкостям у потоці:

$$f(v) = \left(\frac{\beta}{\pi}\right)^{3/2} v^2 \exp(-\alpha v^2), \text{ де } \beta = m^* / 2k_B T. \quad (4.117)$$

Графічно розподіл (4.117) представлений на рис.4.8. Розрізняють три характеристичні швидкості цього розподілу: v_{\max} найвірогідніша, $\langle v \rangle$ – середня, $\sqrt{\langle v^2 \rangle}$ – середня квадратична швидкість. Співвідношення характеристичних швидкостей:

$$v_{\max} / v_{\max} : \langle v \rangle / v_{\max} : \sqrt{\langle v^2 \rangle} / v_{\max} = 1 : (4/\pi)^{1/2} : (3/2)^{1/2} = 1 : 1,13 : 1,22. \quad (4.118)$$

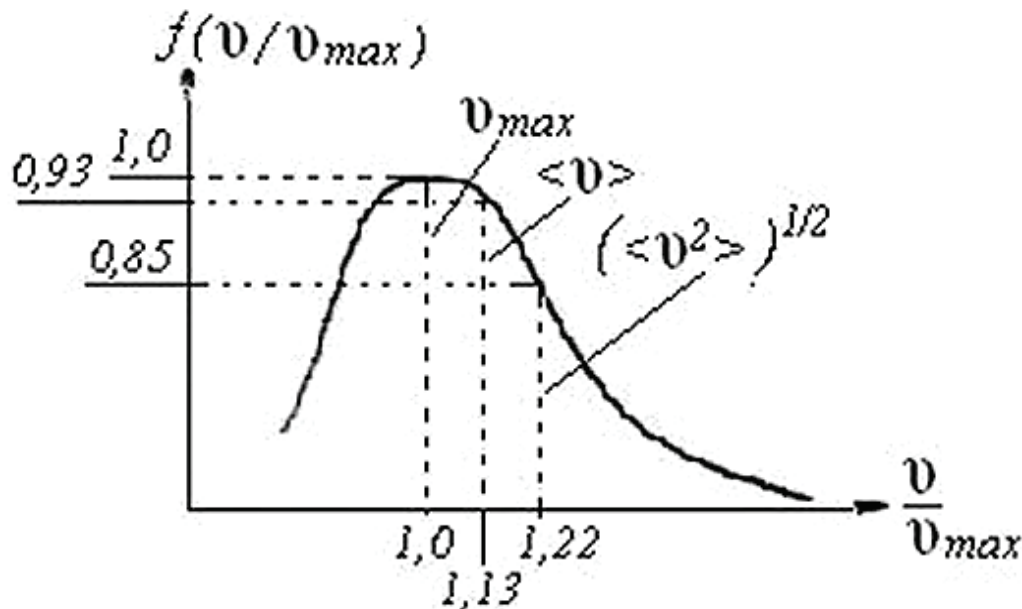


Рис.4.8. Максвеллівський розподіл носіїв заряду у невідроджених напівпровідниках по швидкостям. Магнітна сила встановлюється по середньоквадратичній швидкості, електрична – по середній.

Відносний розподіл носіїв заряду по характеристичним швидкостям:

$$f(v_{\max}) / f(v_{\max}) : f(\langle v \rangle) / f(v_{\max}) : f((\langle v^2 \rangle)^{1/2}) / f(v_{\max}) = 1 : 0,93 : 0,85 \quad (4.119)$$

Аналіз максвеллівського розподілу носіїв заряду по характеристичним швидкостям у потоці:

- розподіл несиметричний відносно v_{\max} ;
- відносна кількість носіїв заряду при збільшенні швидкості (4.119) зменшується повільніше, ніж зростає $\langle v \rangle$ і $\sqrt{\langle v^2 \rangle}$ (4.118);
- умова компенсації сил (4.116) буде різною для характеристичних швидкостей;
- виконання умови (4.116) ускладнюється ще і тим, що час релаксації залежить від швидкості носіїв заряду (4.49), (4.61).

Врахування вищевказаних особливостей розподілу по швидкостям можливе лише при розв'язку кінетичного рівняння Больцмана.

Ефект Холла у слабких полях. Розглядається однорідно легований невідроджений електронний напівпровідник ($n \gg p$) із простою зонною структурою та обмежених розмірів, вміщений у постійне слабе електричне поле з напруженістю \mathcal{E}_x (виконується закон Ома $J_x = \sigma_n \mathcal{E}_x$) і постійне слабе магнітне поле з напруженістю B_z ($\mu_n B_z \ll 1$).

Деталізуємо умови для подальших розрахунків. Положення рівня Фермі відповідає області виснаження донорів (3.42). Розсіяння фононне і, відповідно до (4.49), $\tau_n(k) = \tau_0 m_n^{*-1} (k_B T)^{-1} k^{-1}$, де у τ_0 зібрані усі величини, які не залежать від $\bar{k}, m_n^*, k_B T$.

Векторну функцію енергії $\bar{\chi}_n(E)$ (4.26в) ділимо на дві складові:

$$\bar{\chi}'_n(E) = -\hbar \frac{e\tau_n}{m_n^*} \bar{\mathcal{E}}; \quad (4.120a)$$

$$\bar{\chi}''_n(E) = -\hbar \left(\frac{e\tau_n}{m_n^*} \right)^2 [\overline{B\mathcal{E}}]. \quad (4.120б)$$

Відповідно до цього поділу, визначаємо дві компоненти повного струму $\bar{J}_n = \bar{J}'_n + \bar{J}''_n$. Компонента \bar{J}'_n вже визначена у (4.71). За такою ж методикою визначається струм \bar{J}''_n :

$$\bar{J}''_n = -e \int_{-\infty}^{+\infty} \left(-\frac{\partial f_0}{\partial E}\right) \bar{\chi}''_n(E) \bar{k} \frac{\hbar \bar{k}}{m_n^*} \frac{d^3 \bar{k}}{4\pi^3} = \left(\frac{e\hbar}{m_n^*}\right)^3 \frac{n\tau_0}{4\pi^3 k_B T N_c} [\overline{B\mathcal{E}}] \int_0^\infty k^2 e^{-\alpha k^2} dk. \quad (4.121)$$

Табличний інтеграл має значення $\sqrt{\pi}/4\alpha^{3/2}$, де $\alpha = \hbar^2/2m_n^*k_B T$. Із врахуванням (4.72), (4.72а) для \bar{J}''_n отримаємо:

$$\bar{J}''_n = \frac{3\pi}{8} en(\mu_n)^2 [\overline{B\mathcal{E}}] = \frac{3\pi}{8} \sigma_n \mu_n [\overline{B\mathcal{E}}]. \quad (4.122)$$

Повний електронний струм:

$$\bar{J}_n = \bar{J}'_n + \bar{J}''_n = \sigma_n \bar{\mathcal{E}} + \frac{3\pi}{8} \sigma_n \mu_n [\overline{B\mathcal{E}}]. \quad (4.123)$$

Аналогічні розрахунки для повного діркового струму дають:

$$\bar{J}_p = \bar{J}'_p + \bar{J}''_p = \sigma_p \bar{\mathcal{E}} - \frac{3\pi}{8} \sigma_p \mu_p [\overline{B\mathcal{E}}]. \quad (4.124)$$

Перейдемо до прямокутної системи координат і направимо зовнішнє електричне поле \mathcal{E}_x по вісі x , магнітну індукцію \bar{B}_z по z , скалярний добуток $\mu_n B_z \mathcal{E}_x$ має розмірність напруженості поля \mathcal{E}_{yn} . Повні струми у схрещених полях:

$$J_n = \sigma_n \mathcal{E}_x + \frac{3\pi}{8} \sigma_n \mu_n B_z \mathcal{E}_x, \quad (4.125a)$$

$$J_p = \sigma_p \mathcal{E}_x - \frac{3\pi}{8} \sigma_p \mu_p B_z \mathcal{E}_x. \quad (4.125б)$$

Повний електронний струм має дві складові – перша J_{nx} зумовлена полем \mathcal{E}_x , друга J_{ny} зумовлена полем $\mathcal{E}_{yn} = \mu_n B_z \mathcal{E}_x$. Рівняння (4.125а) розпадається на два рівняння:

$$\left. \begin{aligned} J_{nx} &= \sigma_n \mathcal{E}_x - \frac{3\pi}{8} \sigma_n \mu_n B_z \mathcal{E}_{yn} \\ J_{ny} &= \sigma_n \mathcal{E}_{yn} + \frac{3\pi}{8} \sigma_n \mu_n B_z \mathcal{E}_x \end{aligned} \right\}. \quad (4.126)$$

Напівпровідник має обмежений розмір по вісі y і струм по цьому напрямку відсутній $J_{ny} = 0$. Розв'язуємо систему (4.126) відносно \mathcal{E}_{yn} і \mathcal{E}_x :

$$\mathcal{E}_{yn} = -\frac{3\pi}{8} \mu_n B_z \mathcal{E}_x; \quad \mathcal{E}_x = J_{nx} / \sigma_n + \frac{3\pi}{8} \mu_n B_z \mathcal{E}_{yn}. \quad (4.127)$$

Помножимо $\mathcal{E}_x = J_{nx} / \sigma_n$ скалярно на B_z і результат підставимо у \mathcal{E}_{yn} :

$$\mathcal{E}_{yn} = -\frac{3\pi}{8} \frac{\sigma_n \mu_n^2}{\sigma_n^2} J_{nx} B_z = -\frac{3\pi}{8} \frac{1}{en} J_{nx} B_z = R_{Hn} J_{nx} B_z. \quad (4.128)$$

Наслідки отриманого результату (4.128):

- у провіднику обмежених розмірів, вміщеному у схрещені електричне \mathcal{E}_x і магнітне B_z поля, виникає поперечне поле \mathcal{E}_{yn} , яке називається **полем Холла**, а саме явище – **ефектом Холла**;
- поле Холла прямо пропорційне щільності струму і магнітній індукції;
- коефіцієнт пропорційності R_H називається **коефіцієнтом або сталою Холла** (у слабких магнітних полях не залежить від магнітної індукції);
- числова константа у R_H позначається r і називається **холл-фактором**;
- отримане значення $r = 3\pi/8 \approx 1,18$ відповідає домінуючому фононному розсіюванню. Аналогічний розрахунок для домінуючого іонного розсіяння із використанням (4.61) замість (4.49) дає значення $r = 315\pi/512 \approx 1,93$;
- у металах та вироджених напівпровідниках максвеллівський розподіл по швидкостям (4.117) не діє – усі електрони у потоці мають практично однакові характеристичні швидкості $v_n \approx \langle v \rangle \approx \sqrt{\langle v^2 \rangle}$. Селекція електронів по швидкості і часу релаксації у магнітному полі у таких умовах відсутня, тому $r = 1$.

У випадку діркового напівпровідника ($p \gg n$) подібні до вище наведених розрахунки дають для поля Холла \mathcal{E}_{yp} :

$$\mathcal{E}_{yp} = \frac{3\pi}{8} \frac{1}{e\rho} J_{px} B_z = R_{Hp} J_{px} B_z. \quad (4.129)$$

Якщо провідність напівпровідника змішана, поле Холла \mathcal{E}_y формується як електронами, так і дірками:

$$\mathcal{E}_y = \frac{r}{e} \frac{p\mu_p^2 - n\mu_n^2}{(p\mu_p + n\mu_n)^2} j_x B_z = R_H j_x B_z. \quad (4.130)$$

Для власного напівпровідника ($n = p = n_i$) :

$$R_{Hi} = \frac{r}{en_i} \frac{\mu_p - \mu_n}{\mu_p + \mu_n} = \frac{r}{en_i} \frac{1-b}{1+b}, \quad (4.131)$$

де $b = \mu_n / \mu_p$. У більшості випадків $b > 1$ і $R_{Hi} < 0$.

У власних напівпровідниках ефект Холла можливий лише за умови $b \neq 1$. Фізичний зміст цього положення такий: магнітні сили (4.110) направлені однаково для електронів і дірок (заряди протилежні і напрямки v_x також протилежні), наслідком чого є часткова ($b \neq 1$) або повна ($b = 1$) компенсація поля Холла (4.130).

Найбільші значення поля Холла \mathcal{E}_y можна отримати у невідроджених домішкових напівпровідниках за умови, що концентрація основних носіїв заряду перевищує не менше ніж на порядок концентрацію неосновних. Співставляючи ефект Холла у напівпровідниках і металах необхідно врахувати, що концентрація електронів у металах на 5...8 порядків більша ніж у напівпровідниках. У стільки ж разів холлівське поле у металах буде менше ніж у напівпровідниках за рівних інших умов.

Добуток коефіцієнта Холла на питому провідність має розмірність мобільності:

$$|R_{Hn} \sigma_n| = r \frac{en\mu_{dn}}{en} = r\mu_{dn}, \text{ де } \mu_{dn} - \text{ дрейфова мобільність.} \quad (4.132)$$

Мобільність $r\mu_d$ називається холлівською і позначається μ_H . Враховуючи значення холл-фактора r , можна записати:

-для невідроджених напівпровідників $1,18 \leq r \leq 1,93$ і $\mu_H = r\mu_d$; (4.133)

-для відроджених напівпровідників та металів $r = 1$ і $\mu_H = \mu_d$. (4.133а)

Фізичний зміст (4.133) полягає у тому, що магнітна сила встановлюється по носіям заряду, швидкість яких $\sqrt{\langle v^2 \rangle}$ і більша середньої $\langle v \rangle$ (4.118). Відносна частка цих носіїв заряду у потоці також більша, ніж частка носіїв із $\langle v \rangle$. Компенсуюча сила, яку створює поле Холла, від швидкості не залежить. Час релаксації залежить від швидкості, що і зумовлює появу холл-фактора $r > 1$.

У вказаному діапазоні значень (4.133) r змінюється немонотонно при переході від фононного до іонного механізму розсіювання. Коли у сумісному фононно – іонному розсіянні доля іонного становить близько 20%, r має мінімальне значення 1,13.

У сильних магнітних полях ($\mu B \gg 1$) постійна Холла не залежить від механізму розсіювання:

$$|R_H| = \frac{1}{en}, \text{ де } n - \text{концентрація носіїв заряду.} \quad (4.133б)$$

Магніторезистивний ефект Гаусса. При аналізі траєкторії руху носіїв заряду у схрещених полях було встановлено, що в прямокутному зразку, необмеженому по напрямку y , струм J_x в напрямку зовнішнього електричного поля E_x зменшується за рахунок викривлення траєкторій у магнітному полі і була отримана відповідна формула (4.115) для струму J_x за наявності магнітного поля B_z .

Зовнішнім проявом вище вказаного ефекту є збільшення електричного опору провідника у схрещених електричному і магнітному полях –

магніторезистивний ефект Гаусса. Кількісною характеристикою ефекту Гаусса є відносна зміна опору (магнітоопір) провідника із струмом, вміщеного

у поперечне магнітне поле $\frac{\rho_B - \rho_0}{\rho_0} = \frac{\Delta\rho}{\rho_0}$, де ρ_0 – питомий опір у відсутності магнітного поля, ρ_B – питомий опір у магнітному полі.

Розглянемо ефект Гаусса кількісно. За основу берем формулу (4.115). Замінивши v_x на $\mu\mathcal{E}_x$, отримаємо:

$$J_{nx} = en\mu(1 - (\mu B_z)^2)\mathcal{E}_x = \sigma_0(1 - (\mu B_z)^2)\mathcal{E}_x = \sigma_B\mathcal{E}_x, \quad (4.134)$$

де σ_0 – провідність у відсутності магнітного поля, σ_B – провідність у магнітному полі.

Відносні зміни провідності та опору у слабких магнітних полях ($\mu B_z \ll 1$):

$$\frac{\sigma_B}{\sigma_0} - 1 = -(\mu B_z)^2; \quad \frac{\rho_B}{\rho_0} - 1 = \frac{(\mu B_z)^2}{1 - (\mu B_z)^2} \approx (\mu B_z)^2. \quad (4.135)$$

Формула (4.135) для магнітоопору $\Delta\rho/\rho_0$ отримана в умовах спрощеного врахування розсіювання – ігнорувалася залежність часу релаксації від швидкості носіїв заряду. Якщо врахувати таку залежність за допомогою кінетичного рівняння Больцмана, для магнітоопору отримаємо:

$$\frac{\Delta\rho}{\rho_0} = \frac{3}{2}r\left(1 - \frac{\pi}{4}\right)(\mu B_z)^2, \quad \text{де } r - \text{холл-фактор.} \quad (4.136)$$

Числовий коефіцієнт у формулі (4.136), із урахуванням (4.133), має значення 0,38 для фононного розсіювання.

Розрахунки магнітоопору $\frac{\Delta\rho}{\rho_0}$ зроблені на основі формули (4.115), яка передбачає необмежений розмір прямокутного зразка по вісі y , тобто відсутність поля Холла \mathcal{E}_x .

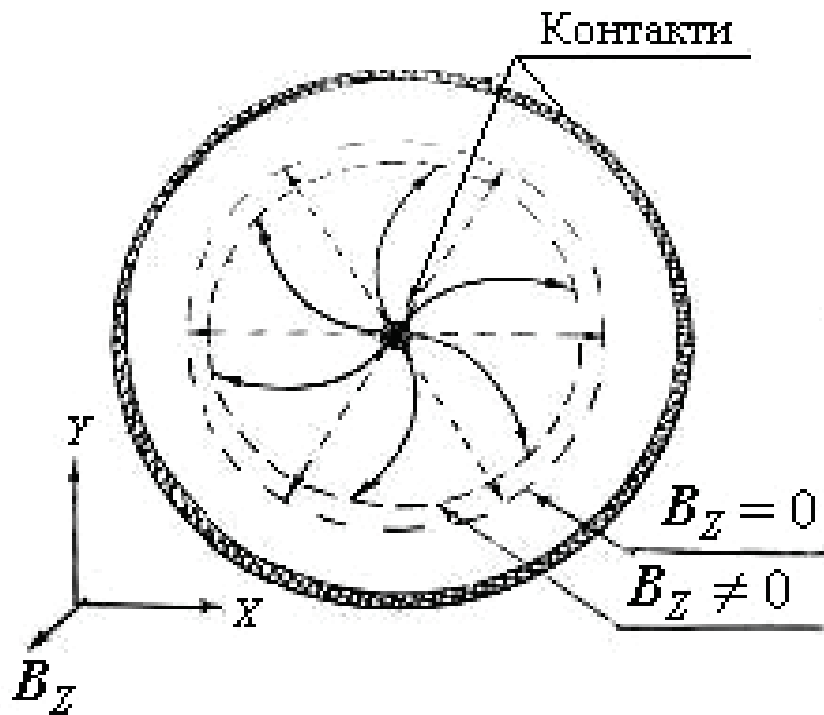


Рис.4.9.Магніторезистор
у формі диска Корбіно

Усунути поле Холла у реального зразка обмежених розмірів можна таким чином:

- а) збільшувати геометричний розмір зразка a_y і, завдяки цьому, зменшувати напруженість поля \mathcal{E}_x (див. (4.116));
- б) встановити режим короткого замикання по холлівським контактам, тобто $\mathcal{E}_x=0$;
- в) взагалі виключити утворення поля Холла за рахунок кругової поляризації струму в площині xy (диск Корбіно).

Збільшення поперечного розміру a_y послаблює компенсуючу дію поля Холла, але не усуває її. Режим короткого замикання холлівських зондів в

значній мірі зменшує поле Холла, але лише у зоні омичних контактів, в об'ємі зразка поле Холла частково зберігається.

У випадку кругової поляризації струму (рис.4.9) поле Холла принципово не може виникати і зразок у вигляді диска Корбіно є ідеальною моделлю фізично необмеженого зразка у схрещених полях.

Порівняння магнітоопору прямокутних пластин із різним відношенням довжини резистора l до ширини w та диска Корбіно подано на рис.4.10.

У сильних магнітних полях носії заряду формують кругові траєкторії, радіус яких зменшується у міру зростанні магнітної індукції. Наслідком стає згасання дрейфового руху у площині xu і відповідне зростання магнітоопору у десятки і навіть сотні разів.

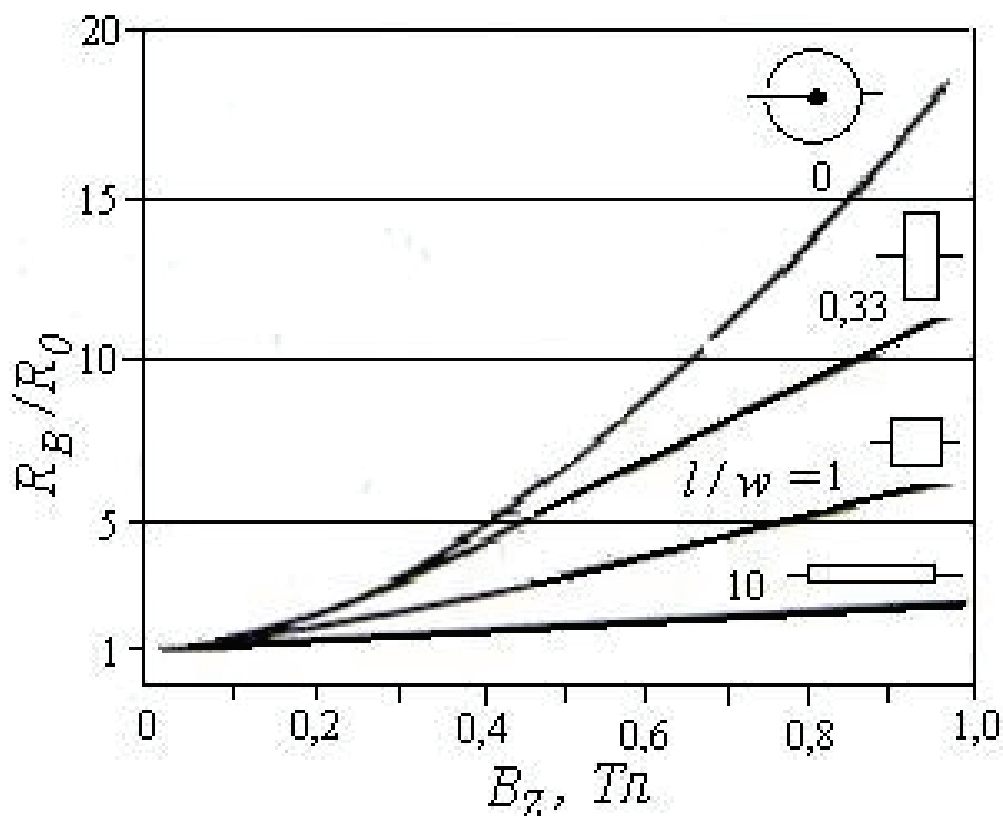


Рис.4.10. Зміна магнітоопору у залежності від відношення довжини резистора l до його ширини w

У сильних магнітних полях носії заряду формують кругові траєкторії, радіус яких зменшується у міру зростанні магнітної індукції. Наслідком стає згасання дрейфового руху у площині xy і відповідне зростання магнітоопору у десятки і навіть сотні разів.

Порівняння гальваномагнітних ефектів. Найпоширенішим технічним застосуванням ефектів Холла і Гаусса є визначення магнітної індукції.

Кількісною характеристикою ефекту Холла є холлівська різниця потенціалів V_y , яку на основі формул (4.128) і (4.129) можна записати таким чином:

$$|V_y| = \frac{a_y r}{en} j_x B_z = \frac{r}{en a_z} I_x B_z, \quad (4.137)$$

де a_y, a_z – розміри пластини по вісям y та z , n – концентрація електронів або дірок, I_x – струм, J_x – щільність струму.

Із (4.137) випливає, що холлівська різниця потенціалів V_y прямо пропорційна магнітній індукції B_z і це дуже важливо при градуюванні магнітометрів, які використовують ефект Холла. Оптимальне значення n уже обговорювалося. Товщина пластини по напрямку магнітного поля a_z повинна бути якомога меншою, співвідношення поперечних розмірів $a_y/a_x \geq 2 \dots 3$. Різниця потенціалів V_y прямо пропорційна струму I_x , але збільшувати довільно струм не можна, його величина обмежена теплотою Джоуля, яка може викликати підвищення температури напівпровідникової пластини. У режимі вимірювання магнітної індукції $I_y = 0$, тому мобільність носіїв заряду не має суттєвого значення. Не має значення і тип провідності – електронний чи дірковий, тільки не власний.

Для ефекту Гаусса, навпаки, вирішальне значення має мобільність (див.(4.136)). Магнітоопір пропорційний μ^2 і B^2 . Чутливість магніtoresистора цілком залежить від мобільності носіїв заряду. Зазвичай $\mu_n > \mu_p$ і для

магніторезисторів використовують переважно електронні напівпровідники. Параболічна залежність магнітоопору від індукції ускладнює градування магнітометрів, які використовують ефект Гаусса.

З фізичної точки зору, ефекти Холла і Гаусса є взаємовиключаючими один одного за умовами реалізації (наявність чи відсутність поля Холла), тому універсальний зразок для цих ефектів виготовити неможливо.

Для напівпровідникових матеріалів з малою мобільністю, коли $\mu < 0.5 \text{ м}^2/\text{В}\cdot\text{с}$ краще реалізувати ефект Холла, для матеріалів з високою мобільністю, коли $\mu > 1 \text{ м}^2/\text{В}\cdot\text{с}$ більше підходить ефект Гаусса.

У металів концентрація електронів на 5 – 8 порядків більша, а мобільність на 2 – 3 порядки менша ніж у напівпровідників. Відповідно, ефекти Холла і Гаусса у металів виявляються відносно дуже слабкими.

4.12. Тензорезистивний ефект у твердих тілах

Загальні положення.

Опір лінійного провідника електричного струму:

$$R = \rho \frac{\ell}{s}, \quad (4.138)$$

де: ρ – питомий опір матеріалу; ℓ – довжина провідника; s – поперечний перетин провідника.

При деформації стиску або розтягу будуть змінюватися довжина, перетин провідника і питомий опір. Відносна зміна опору при деформації:

$$\frac{\Delta R}{R} = \frac{\Delta \rho}{\rho} \pm \frac{\Delta \ell}{\ell} \mp \frac{\Delta s}{s}, \quad (4.139)$$

де „+” для розтягу, „–” для стиску.

Розглянемо відносну зміну опору, обумовлену зміною тільки геометричних розмірів провідника ($\Delta \rho / \rho = 0$). Для пружних твердих тіл існує зв'язок між відносними змінами поперечних і подовжніх розмірів при деформації:

$$\frac{\Delta r}{r} = \mu_{II} \frac{\Delta \ell}{\ell}, \quad (4.140)$$

де r – радіус поперечного перетину; μ_{II} – коефіцієнт Пуассона.

Перерахувавши Δs і s через Δr і r , для відносної зміни опору за рахунок зміни розмірів провідника при одновимірній деформації маємо:

$$\frac{\Delta R}{R} = \pm (1 + 2\mu_{II}) \frac{\Delta \ell}{\ell}. \quad (4.141)$$

Для більшості пружних твердих тіл $\mu_{II} \approx 0,3$, отже:

$$\frac{\Delta R}{R} \approx \pm 1,6 \frac{\Delta \ell}{\ell}. \quad (4.142)$$

Зміну опору провідника за рахунок зміни його розмірів при деформації (4.142) будемо називати *геометричним тензорезистивним ефектом*.

Кількісною характеристикою тензорезистивного ефекту є **коефіцієнт тензочутливості** – відносна зміна опору, яка відповідає $\Delta\ell/\ell = 1$, тобто, коли довжина провідника при одновимірній деформації змінюється удвічі,:

$$K = \frac{\Delta R}{R} / \frac{\Delta\ell}{\ell}. \quad (4.143)$$

Для геометричного тензоефекту, відповідно до (4.142), $K_G \approx 1,6$ і завжди позитивний.

У загальному випадку, із урахуванням можливої зміни питомого опору при деформації:

$$K = K_G + \frac{\Delta\rho}{\rho} / \frac{\Delta\ell}{\ell} = 1,6 + \frac{\Delta\rho}{\rho} / \frac{\Delta\ell}{\ell}. \quad (4.144)$$

Для металів експеримент дає $K_{Me} = 1,6 \dots 2,4$, тобто має місце переважно геометричний тензоефект і лише для нікелю $K_{Me} \approx 10$.

Для домішкових напівпровідників при одновимірній деформації коефіцієнт тензочутливості досягає значень 80...150. Більше того, він може бути як позитивним, так і негативним.

Висновок: у металів тензорезистивний ефект в основному геометричний, у напівпровідників визначальною є зміна питомого опору при деформації.

Подальший розгляд тензорезистивного ефекту буде проведений на прикладі кремнію, оскільки саме кремній знайшов найширше застосування у напівпровідниковій тензометрії.

Тензорезистивний ефект у електронному кремнії. У енергетичному просторі зони провідності кремнію є 6 еквівалентних мінімумів енергії (рис.4.11), розташованих на головних вісях оберненої ґратки $\langle 100 \rangle$. У силу енергетичної еквівалентності носії заряду розподіляються по 6 мінімумам

рівномірно, тобто на кожний мінімум припадає 1/6 повного числа носіїв заряду в зоні провідності.

Ефективна маса електронів для окремо узятото мінімуму анізотропна, тобто поверхні рівної енергії несферичні. Для кремнію це еліпсоїди обертання з подовжньою ефективною масою $m_{\ell}^* = 0,97m$ і поперечною ефективною масою $m_t^* = 0,19m$. Анізотропія ефективних мас $m_{\ell}^* / m_t^* \approx 5$.

Якщо електричний струм паралельний одному з напрямків $\langle 100 \rangle$, то 1/3 електронів буде брати участь у провідності з ефективною масою m_{ℓ}^* , а 2/3 – із масою m_t^* . Такий розподіл електронів по ефективних масах буде зберігатися в усіх випадках, поки не порушена енергетична еквівалентність мінімумів.

На підставі вищесказаного, запишемо питому провідність недеформованого електронного кремнію:

$$\sigma_0 = e \left[\frac{2}{3} n \mu_{\perp} + \frac{1}{3} n \mu_{\parallel} \right], \quad (4.145)$$

де μ_{\parallel} – мобільність, що відповідає m_{ℓ}^* ; μ_{\perp} – мобільність, що відповідає m_t^* .

Розглянемо детальніше процеси при одновимірній деформації уздовж напрямку $\langle 010 \rangle$. На рис.4.12(a) показане положення 2-х із 6 мінімумів при деформації стиску. Еквівалентність мінімумів порушується: два мінімуми на вісі стиску $\langle 010 \rangle$ (рис.4.12,(b)) опускаються відносно дна зони провідності E_c , чотири, у площині, перпендикулярній вісі стиску – піднімаються.

Енергетичне зміщення мінімумів у зоні провідності n -Si прямо пропорційне деформуючій механічній напрузі $P_F = F/s$, де F – сила, s – площа перерізу резистора. Для напрямку P_F по вісям $\langle 100 \rangle$ коефіцієнти пропорційності мають такі значення:

$$\Delta E_1 = \pm 4,72 \cdot 10^{-11} \text{ eV} / \text{Па}; \quad \Delta E_{2,3} = \mp 3,57 \cdot 10^{-11} \text{ eV} / \text{Па}. \quad (4.146)$$

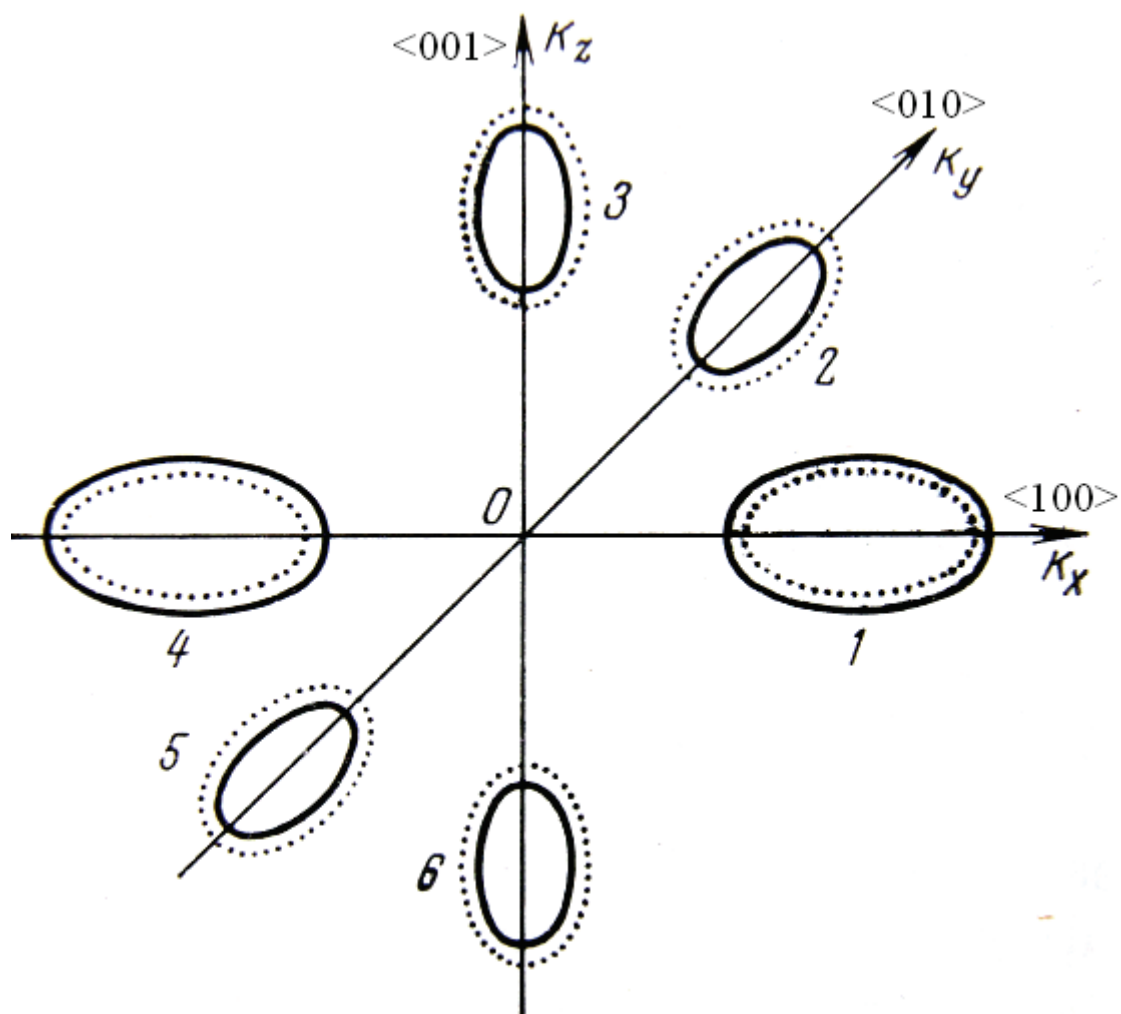


Рис.4.11. Просторове розташування мінімумів енергії у зоні провідності кремнію. Деформація стиску уздовж $\langle 100 \rangle$ збільшить об'єми 1 – го і 4 – го еліпсоїдів, у перпендикулярних напрямках об'єми 2,3,5,6 зменшаться. Пунктири – без деформації, суцільні лінії – деформація

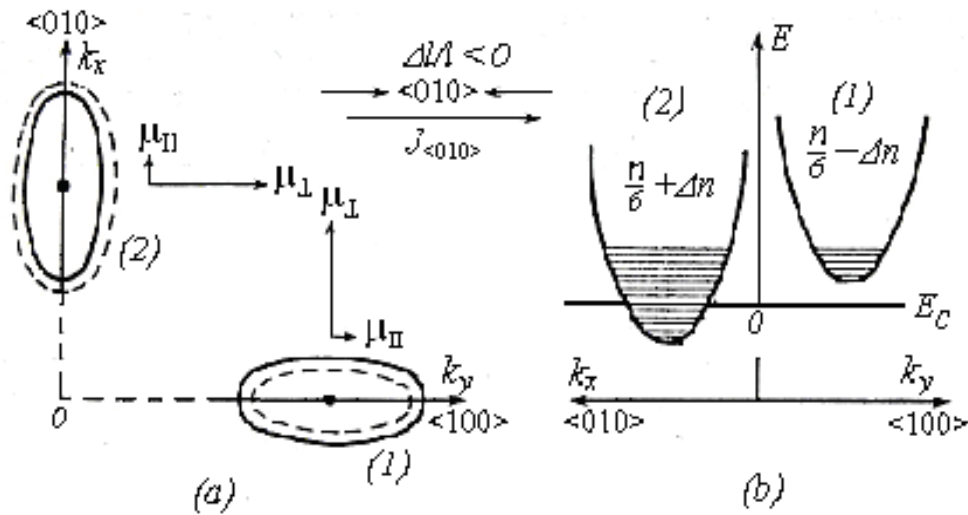


Рис.4.12. Перерозподіл електронів для 2-х із 6 мінімумів енергії у зоні провідності **n-Si** при стиску у напрямку $\langle 010 \rangle$. (a) – екстремуми у \mathbf{k} – просторі, суцільні лінії – еліпсоїди до деформації, штрихові – після стиску. (b) – порушення еквівалентності мінімумів і перерозподіл електронів. Струм тече паралельно вісі деформації.

У результаті порушення енергетичної еквівалентності мінімумів енергії відбувається перерозподіл електронів між ними. При зазначених умовах, число електронів у двох мінімумах на вісі стиску збільшиться, в інших 4-х - зменшиться. Позначимо Δn зміну числа носіїв в одному мінімумі через порушення енергетичної еквівалентності. Складемо нову формулу для провідності, вводячи Δn у (4.145):

$$\sigma' = e \left[\left(\frac{2}{3}n - 4\Delta n \right) \mu_{\perp} + \left(\frac{1}{3}n + 4\Delta n \right) \mu_{\parallel} \right]. \quad (4.146a)$$

Після елементарних перетворень (4.146) отримаємо:

$$\sigma' = \sigma_0 - 4e\Delta n(\mu_{\perp} - \mu_{\parallel}). \quad (4.147)$$

Аналогічний розрахунок для випадку розтягу уздовж $\langle 010 \rangle$ дає:

$$\sigma' = \sigma_0 + 2e\Delta n(\mu_{\perp} - \mu_{\parallel}). \quad (4.148)$$

Висновок: зважаючи на те, що $\mu_{\parallel} < \mu_{\perp}$, провідність електронного кремнію при стиску уздовж вісі обертання поверхні рівної енергії $\langle 010 \rangle$

зменшується, а опір зростає ($\Delta R / R > 0$). Коефіцієнт тензочутливості (4.143)
 $K_{n-Si} < 0$.

Аналізуючи (4.147) і (4.148) дійдемо висновку, що умовами для тензорезистивного ефекту в електронному кремнії при одновимірній деформації є:

- 1) анізотропія мобільностей $\mu_{\parallel} \neq \mu_{\perp}$;
- 2) перерозподіл електронів $\Delta n \neq 0$.

Що стосується першої умови, то для електронного кремнію вона виконується завжди, тому що поверхні рівної енергії для зони провідності неферичні.

Друга умова потребує порушення еквівалентності мінімумів. Це, у свою чергу, може відбутися лише при певній орієнтації деформуючої сили відносно кристалографічних вісей кремнію.

Для електронного кремнію максимальний коефіцієнт тензочутливості має місце при деформації уздовж будь – якої з вісей $\langle 100 \rangle$, мінімальний – при деформації уздовж вісі $\langle 111 \rangle$, тому що ця вісь складає рівні кути з вісями $\langle 100 \rangle$. Положення екстремумів у цьому випадку змінюється у енергетичному просторі однаковою чином – усі піднімаються або опускаються, але еквівалентність їх не порушується. Для вісі $\langle 110 \rangle$ коефіцієнт тензочутливості має проміжне значення (рис.4.13).

Що стосується зміни провідності, то вона, відповідно до (4.147) і (4.148), однозначно визначається числом $\Delta n / n$, тому що μ_{\parallel} , μ_{\perp} мають фіксовані значення. Зміна провідності при деформації буде тим більше, чим:

- більша різниця енергій для нееквівалентних мінімумів, тобто більше деформація (рис. 4.12,(b));
- більша різниця енергій для нееквівалентних мінімумів, тобто більше деформація (рис. 4.12,(b));

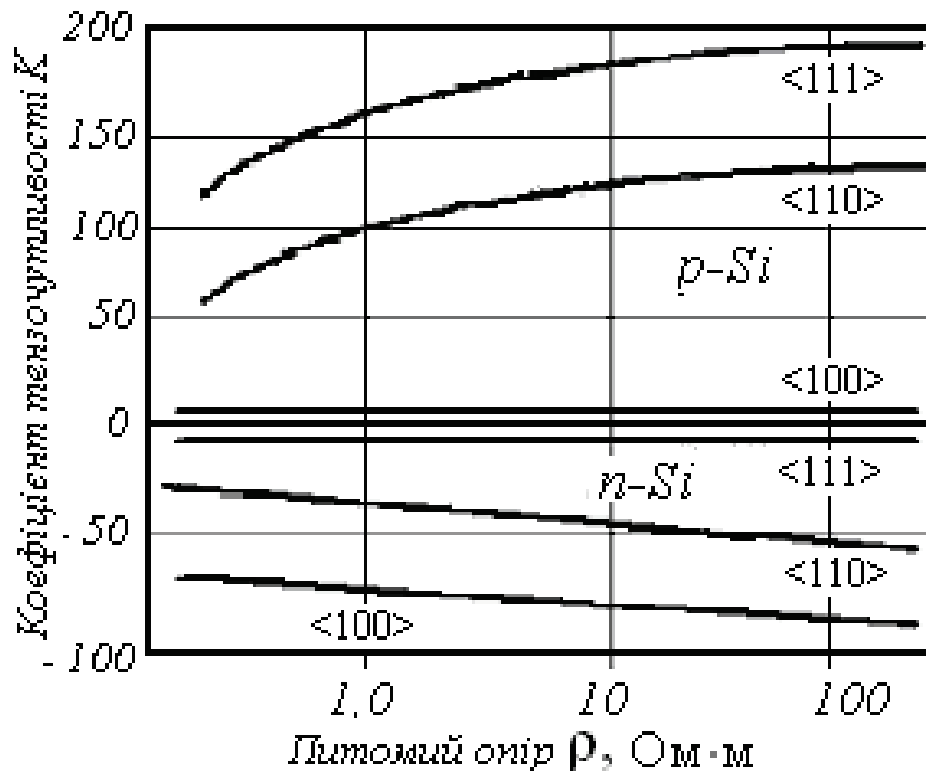


Рис.4.13. Залежність тензочутливості кремнію від питомого опору, типу провідності та кристалографічної орієнтації

- менша температура, тому що теплові перекидання електронів між мінімумами (міждолинне розсіяння) перешкоджають їхньому нерівномірному розподілу при порушенні еквівалентності мінімумів;
- більше відношення $\Delta n/n$, тобто чим більший питомий опір(див. рис.4.13).

Тензорезистивний ефект у дірковому кремнії. У енергетичному просторі валентної зони кремнію є єдиний максимум енергії, розташований у центрі зведеної зони. У центрі зведеної зони сходяться дві енергетичні гілки з різною кривизною, отже, різною ефективною масою дірок (рис.4.14). Гілка V1 має малу кривизну і дірки, що займають енергетичні стани цієї гілки, називаються важкими. Гілка V2 має значно більшу кривизну і, відповідно, дірки для цих станів називаються легкими.

Поверхні рівної енергії для V1 і V2 несферичні і найбільша різниця ефективних мас важких $m_{pв}^*$ і легких $m_{pл}^*$ дірок відповідає кристалографічній вісі <111>. У напрямку <110> це розходження менше, а в <100> ефективні маси однакові (рис.2.6).

У недеформованому кремнії дірки розподіляються між V1 і V2 відповідно до щільності станів для цих гілок, яка пропорційна $m_p^{*3/2}$ (2.11). Кількість легких і важких дірок :

$$p_l = p \left(m_{pл}^* / m_{pв}^* \right)^{3/2}, \quad p_v = p \left[1 - \left(m_{pл}^* / m_{pв}^* \right)^{3/2} \right], \quad (4.149)$$

де p – повна кількість дірок у валентній зоні.

Провідність недеформованого діркового кремнію:

$$\sigma_0 = e(p_l \mu_l + p_v \mu_v). \quad (4.150)$$

При деформації кремнію в напрямку <111> рівноважне положення гілок V1 і V2 порушується: при стиску V1 стає вище V2, при розтягу – навпаки (рис.4.13).

Енергетичне зміщення гілок V1 і V2 у валентній зоні p -Si, як і у випадку n -Si, прямо пропорційне деформуючій механічній напрузі. Для напрямку P_F по вісі <111> коефіцієнти пропорційності мають такі значення:

$$\Delta E_{\pm} = \pm 2,24 \cdot 10^{-11} \text{ eV} / \text{Па}. \quad (4.151)$$

Особливістю деформаційного зміщення V1 і V2 у p -Si є те, що обидві гілки зміщуються в одному напрямку, але на різну величину.

Наслідком взаємного енергетичного зміщення гілок є перерозподіл дірок між V1 і V2 тепер уже відповідно до їхнього нового енергетичного положення:

при стиску деяка частина дірок Δp із гілки V1 переходить на гілку V2 і вони стають „легшими” , при розтягу – навпаки.

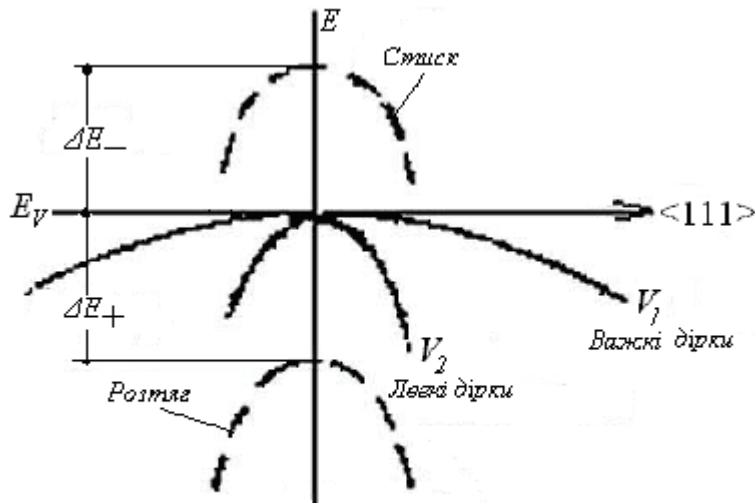


Рис. 4.14. Енергетична діаграма впливу деформації на взаємне розташування підзон легких та важких дірок

Провідність у цих умовах змінюється:

$$\sigma' = \sigma_0 \pm e\Delta p(\mu_l - \mu_v), \quad (4.152)$$

де “+” відповідає стиску, “-” – розтягу.

Висновок: зважаючи на те, що $\mu_l > \mu_v$, провідність діркового кремнію при стиску ($\frac{\Delta \ell}{\ell} < 0$) уздовж вісі $\langle 111 \rangle$ збільшується, а опір зменшується ($\Delta R / R > 0$), при розтягу – навпаки. Коефіцієнт тензочутливості (4.143) $K_{p-Si} > 0$.

Залежність тензочутливості від орієнтації деформуючої сили і рівня легування для діркового кремнію показана на рис. 4.12.

Формула (4.152) принципово не відрізняється від (4.147) і (4.148). Відповідно, основні положення щодо умов виникнення тензоефекту такі ж, як і для електронного кремнію.

Коефіцієнт тензочутливості кремнію. Відповідно до рис. 4.12, коефіцієнт тензочутливості кремнію має значення у межах 50 – 200. Відносна зміна опору $\Delta R / R$, яку можна одержати при одномірній деформації кремнієвого тензорезистора, залежить як від коефіцієнта тензочутливості K , так і відносної деформації $\Delta \ell / \ell$. Припустимо значення відносної деформації

$\Delta\ell/\ell$ не повинно перевершувати 10^{-3} . При більших деформаціях можливі нелінійні ефекти і незворотні зміни в матеріалах, які з'єднують тензорезистор із тілом, що деформується. Таким чином, реальне значення $\Delta R/R$ не перевершує $\pm 10\%$ від номінального значення опору тензорезистора.

Зменшення рівня легування підвищує коефіцієнт тензочутливості (рис.4.12), але посилює температурну залежність тензочутливості за рахунок температурної залежності рівноважної концентрації електронів і дірок.

При зниженні температури до 77K (температура рідкого азоту) коефіцієнт тензочутливості зростає в десятки разів за рахунок послаблення міждолинного розсіяння і зменшення рівноважної концентрації носіїв заряду.

Резюме по п.4.12

1. Деформація стиску або розтягу лінійного пружного провідника струму супроводжується зміною його опору внаслідок зміни довжини і поперечних розмірів. При розтягу збільшується довжина, але зменшується поперечний переріз і опір зростає, при стиску, навпаки, – опір зменшується. Це явище отримало назву *геометричний тензорезистивний ефект*.

2. У напівпровідниках і деяких металах деформація супроводжується не лише зміною розмірів, але і питомого опору матеріалу провідника.

3. Кількісною характеристикою тензорезистивного ефекту є *коефіцієнт тензочутливості* – відносна зміна опору відповідна одиничній відносній лінійній деформації.

4. Коефіцієнт тензочутливості для геометричного тензорезистивного ефекту $\approx 1,6$ і завжди позитивний. У більшості металів і сплавів спостерігаються значення коефіцієнта 1,6 – 2,5 і лише у нікелю він близько 10.

5. Коефіцієнт тензочутливості напівпровідників має значення у діапазоні 50 – 200. Окрім того, він може бути як позитивним, так і негативним.

6. Для напівпровідників визначальною є зміна питомого опору, а не геометричних розмірів.

7. У електронному кремнії тензоефект зумовлений порушенням енергетичної еквівалентності мінімумів енергії зони провідності, що супроводжується перерозподілом електронів. Тензорний характер ефективної маси електронів і їх мобільності, просторове розташування мінімумів енергії та перерозподіл електронів між ними зумовлюють *негативний коефіцієнт тензочутливості $n - \text{Si}$* .

8. Зміна опору $p - \text{Si}$ є наслідком перерозподілу між підзонами важких і легких дірок. *Коефіцієнт тензочутливості $p - \text{Si}$ позитивний*, тобто, при розтягу опір тензорезистора зростає.

9. Коефіцієнт тензочутливості суттєво залежить від орієнтації деформуючої сили відносно вісей симетрії кремнію. Максимальне значення коефіцієнта у $n - \text{Si}$ відповідає деформації по вісі $\langle 100 \rangle$, у $p - \text{Si}$ – по вісі $\langle 111 \rangle$.

10. Тензочутливість залежна від температури – чим більша температура, тим менша тензочутливість за інших рівних умов.

4.13. Електропровідність у сильних полях.

У сильних зовнішніх електричних полях спостерігається залежність ефективної маси, мобільності, часу релаксації, концентрації носіїв заряду від напруженості електричного поля. У будь-якому із цих випадків зовнішнім проявом сильного електричного поля є відхилення від закону Ома – порушується пряма пропорційність у залежності щільності струму від напруженості електричного поля.

Для побудови зрозумілої фізичної моделі процесів у сильних електричних полях у першу чергу необхідно виключити *тепловий ефект струму* – підвищення температури напівпровідника за рахунок виділення теплоти Джоуля, який стає особливо помітним у сильних полях. Виходячи з цього, фізичні моделі процесів розглядають у ізотермічних умовах – теплота Джоуля повністю розсіюється у зовнішнє середовище і температура кристалічної ґратки залишається постійною (температура фононного газу $T_\phi = const$), незалежно від енергії, яку отримують носії заряду від електричного поля.

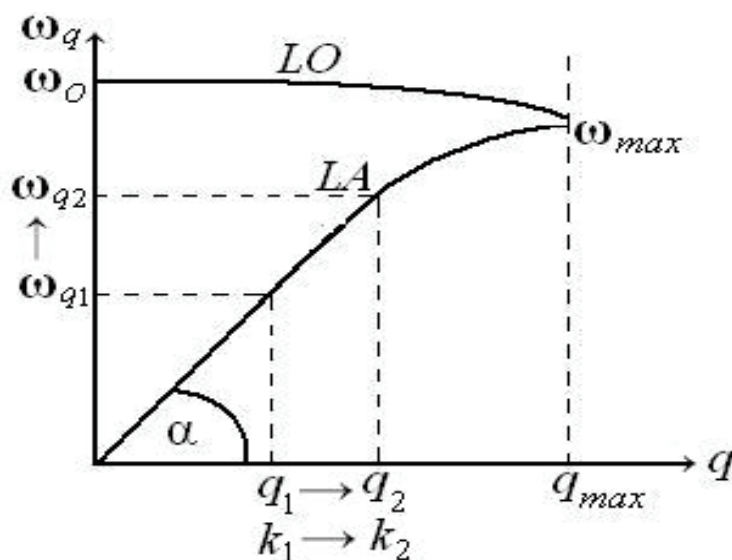


Рис.4.15. Фононний спектр подовжніх коливань типів LA та LO для ковалентних напівпровідників

У сильних електричних полях фононне розсіяння стає домінуючим механізмом енергетичного обміну між електронним і фононним газами. Іонне розсіяння в умовах високих енергій стає мало ефективним (див. (4.62)).

Фононне розсіяння у слабких електричних полях було вже розглянуто раніше (п. 4.6). Електрони розсіюють нерівноважну енергію на подовжніх довгохвильових акустичних фононах (LA). Розглянемо більш детально особливості фононного спектру ковалентних напівпровідників (рис. 4.15).

Важливим для подальшого аналізу є те, що середні значення модулів хвильових векторів електронів і фононів у процесах розсіювання зберігаються практично однаковими $k \approx q$. У разі збільшення кінетичної енергії електрона і модуля хвильового вектора ($k_1 \rightarrow k_2$ на рис.4.15) зростає модуль хвильового вектора $q_1 \rightarrow q_2$, частота $\omega_{q_1} \rightarrow \omega_{q_2}$ і енергія фонона $\hbar\omega_{q_2}$, який приймає участь у розсіянні. Відповідно зростають і енергетичні втрати ΔE (4.39).

Для подальшого аналізу зручно перевести (4.39) у тригонометричну форму за допомогою (І.9), позначивши $x = \hbar\omega_q / k_B T_\phi$:

$$\Delta E = \hbar\omega_q \operatorname{th} \frac{x}{2} = \hbar\omega_q \operatorname{th} \frac{\hbar\omega_q}{2k_B T_\phi}. \quad (4.153)$$

Деякі критичні значення гіперболічного тангенсу:

$$\left\{ \begin{array}{ll} x \rightarrow 0 & \operatorname{th} x \approx x, \quad \operatorname{th} x/2 \approx x/2 \\ x \rightarrow 1 & \operatorname{th} x \approx 0,92, \quad \operatorname{th} x/2 \approx 0,46 \end{array} \right\}. \quad (4.154)$$

Оцінка співвідношення енергій електрона і фонона із урахуванням непружності фононного розсіяння (4.40) показала, що за умов, коли теплова складова повної енергії $E \approx k_B T_\phi$ є домінуючою ($k_B T_\phi \gg \hbar\omega_q$), зміною енергії електрона можна нехтувати.

У загальному ж випадку, непружність розсіяння є вирішальним фактором для досягнення стаціонарної рівноваги в умовах дії зовнішніх силових полів.

Слабкі електричні поля. На шляху пробігу L_ϕ у зовнішньому електричному полі із напруженістю \mathcal{E} електрон набуває енергію

$$e\mathcal{E} L_\phi. \quad (4.155)$$

Рівновага між електронним та фононним газами суттєво не порушується, якщо втрата енергії за рахунок фононного розсіювання значно перевищує (4.155), тобто,

$$\Delta E \gg e\mathcal{E} L_\phi. \quad (4.156)$$

У таких умовах можна вважати, що еквівалентна середній енергії електронного газу температура T_e практично співпадає із $T_\phi = const$ (ізотермічні умови)

$$T_e \approx T_\phi = const. \quad (4.157)$$

Наслідки (4.157) такі: закономірності (4.49) не «працюють»; внесок дрейфової компоненти у повну швидкість електронів незначний:

$$v \approx v_T \gg v_{dp}; \quad (4.158)$$

мобільність і концентрація електронів не залежать від напруженості електричного поля:

$$\mu = \frac{e\tau_\phi}{m^*} \neq f(\mathcal{E}), n \neq f(\mathcal{E}). \quad (4.159)$$

Загальним наслідком (4.159) є незалежність питомої провідності від напруженості поля:

$$\sigma = en\mu \neq f(\mathcal{E}). \quad (4.160)$$

Формула (4.160) є умовою застосування закону Ома у диференціальній формі:

$$\bar{j} = \sigma \bar{\mathcal{E}}. \quad (4.161)$$

Для подальшого аналізу зручно ввести безрозмірний параметр:

$$\xi = \Delta E / E = \Delta E / k_B T_e. \quad (4.162)$$

Електричні поля, напруженість яких відповідає умові $\mathcal{E}_0 \ll \xi \frac{k_B T_e}{e L_\phi}$, не порушують закон Ома у ізотермічних умовах і називаються **слабкими** (рис.4.16, слабкі поля).

Розігрів електронного газу. Перше критичне поле. У міру зростання напруженості електричного поля енергія дрейфового руху при певній напруженості поля досягне рівня енергетичних втрат ΔE і умову (4.156) виконати неможливо буде виконати.

Критична напруженість електричного поля \mathcal{E}_1 визначається із балансу енергій:

$$e \mathcal{E}_1 L_\phi = \Delta E. \quad (4.163)$$

Запишемо (4.163) як баланс потужностей:

$$e \mathcal{E}_1 \frac{L_\phi}{\tau_\phi} = e \mathcal{E}_1 v_{dp} = \xi \frac{k_B T_e}{\tau_\phi}. \quad (4.164)$$

Визначимо v_{dp} із (4.164):

$$v_{dp} = \xi \frac{k_B T_e}{e \mathcal{E}_1 L_\phi} v_T, \quad (4.165)$$

де $v_T = L_\phi / \tau_\phi$ – теплова швидкість електронів.

За стандартним визначенням мобільності $\mu = e \tau_\phi / m^*$ можна скласти ще одне рівняння для дрейфової швидкості:

$$v_{dp} = \frac{e}{m^*} \tau_\phi \mathcal{E}_1 = \frac{e L_\phi}{m^* v_T} \mathcal{E}_1. \quad (4.166)$$

Маємо два рівняння (4.165), (4.166) з двома невідомими v_{dp} та \mathcal{E}_1 .

Розв'язок відносно \mathcal{E}_1 :

$$\mathcal{E}_1 = \frac{k_B T_e}{e L_\phi} (2\xi)^{1/2}. \quad (4.167)$$

Розв'язок відносно v_{dp} :

$$v_{dp} = (\xi / 2)^{1/2} v_T. \quad (4.168)$$

У процесі розв'язку рівнянь (4.165), (4.166) зроблена така заміна:
 $m^* v_T^2 / 2 = k_B T_\phi$.

Співвідношення (4.165) між дрейфовою і тепловою швидкостями електронів отримано за умови (4.157), але воно справедливе і у випадку, коли середня температура електронного газу перевищує середню температуру фононного.

Числова оцінка напруженості критичного поля \mathcal{E}_1 за типовими для напівпровідників даними дає

$$\mathcal{E}_1 \approx 5 \cdot 10^2 - 10^3 \text{ В/см}. \quad (4.169)$$

Електричні поля, напруженість яких перевищує перше критичне поле

$$\mathcal{E} > \mathcal{E}_1 = \frac{k_B T_e}{e L_\phi} (2\xi)^{1/2}, \text{ називаються } \textbf{сильними} \text{ (рис.4.16, "теплі" електрони).}$$

Пряма залежність \mathcal{E}_1 від температури напівпровідника T_ϕ (4.167)

підтверджується експериментально (рис.4.16).

Надлишок енергії електронів, отриманий від сильного поля, порушує енергетичну рівновагу між фононним та електронним газами, еквівалентна середня температура електронного газу T_e стає більшою за T_ϕ , яка у ізотермічних умовах не залежить від напруженості поля. *Загальним наслідком цього буде розігрів електронного газу.*

Повна швидкість електронів тепер зростає як за рахунок теплової

$$v_{T_e} = \sqrt{2k_B T_e / m^*}, \quad (4.170)$$

так і за рахунок дрейфової

$$v_{dp} = \frac{e}{m^*} \tau_\phi \mathcal{E}. \quad (4.171)$$

Вступають у дію закономірності (4.49). Зростають модулі хвильових векторів електрона і фонона. Для акустичних коливань типу LA це означає збільшення енергії $\hbar\omega_q$ фононів (рис.4.15). Зростають енергетичні втрати ΔE , *фононне розсіяння посилюється.*

Посилення фононного розсіяння триватиме поки не встановиться нова стаціонарна рівновага

$$e\mathcal{E}v_{dp} = \xi \frac{k_B T_e}{\tau_\phi} \quad (4.171)$$

для сильного поля з напруженістю $\mathcal{E} > \mathcal{E}_1$.

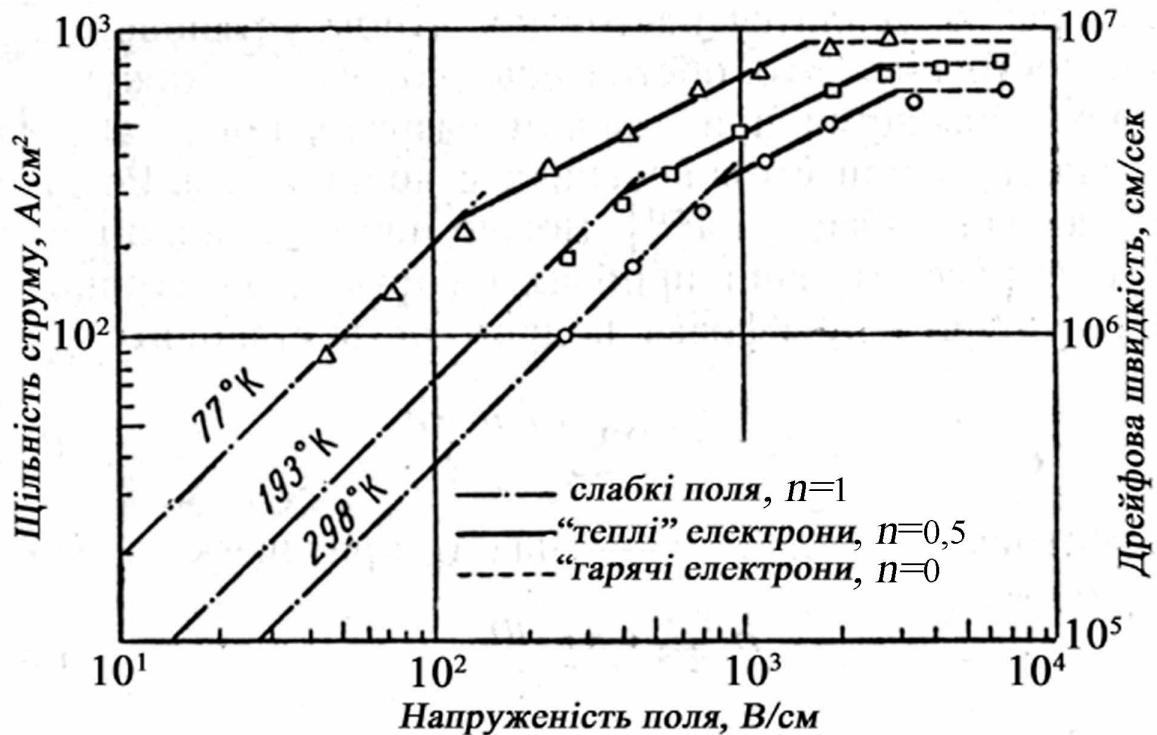


Рис.4.15. Залежність щільності струму та дрейфової швидкості від напруженості поля. Число n – показник степеню у залежностях $v_{dp} \sim \mathcal{E}^n$

Дрейфова швидкість електронів у розігрітому електронному газі:

$$v_{dp} = \frac{e}{m^*} \tau_\phi \mathcal{E} = \frac{e}{m^*} \cdot \frac{L_\phi}{v_{Te} + v_{dp}} \cdot \mathcal{E}. \quad (4.172)$$

Замінивши v_T на v_{T_e} у (4.168), знайдемо v_{T_e} , підставимо у (4.172) і запишемо рівняння для v_{dp} :

$$v_{dp}^2 = \frac{eL_\phi}{m^*} \cdot \frac{(\xi/2)^{1/2}}{1 + (\xi/2)^{1/2}} \cdot \mathcal{E}. \quad (4.173)$$

Оцінка параметра ξ із використанням (4.153), (4.162):

$$\xi = \frac{\Delta E}{k_B T_e} = \frac{\hbar\omega_q \, th \frac{\hbar\omega_q}{2k_B T_\phi}}{k_B T_e}. \quad (4.174)$$

Зростання температури електронного газу T_e зменшує ξ . Одночасно зростає і $\hbar\omega_q$. Енергія акустичних фононів обмежена їх максимальною частотою ω_{\max} (рис.4.14) і має величину у межах $(\hbar\omega_q)_{\max} \approx 0,03 - 0,05$ еВ. Середня енергія теплових коливань для 300 К $k_B T_\phi = 0,026$ еВ і значення аргументу гіперболічного тангенсу лежать у межах $0,5 < \frac{\hbar\omega_q}{2k_B T_\phi} < 1$, а відповідні значення $th x$ у межах $0,45 - 0,8$ (4.154).

Наслідок проведених оцінок – $(\xi/2)^{1/2} \ll 1$, що дозволяє спростити (4.173):

$$v_{dp} = \left[\frac{eL_\phi}{m^*} \cdot (\xi/2)^{1/2} \right]^{1/2} \cdot \mathcal{E}^{1/2}. \quad (4.175)$$

Висновок: у розігрітому електронному газі ($T_e > T_\phi$) дрейфова швидкість зростає пропорційно $\mathcal{E}^{1/2}$ (рис.4.14, “теплі” електрони).

Мобільність – коефіцієнт пропорційності між дрейфовою швидкістю і напруженістю електричного поля – визначимо із (4.175), помноживши і поділивши праву частину на \mathcal{E} :

$$v_{dp} = \left[\frac{eL_\phi}{m^*} \cdot (\xi/2)^{1/2} \right]^{1/2} \cdot \mathcal{E}^{-1/2} \cdot \mathcal{E} = \mu_0 \mathcal{E}^{-1/2} \cdot \mathcal{E}. \quad (4.176)$$

Висновок: у розігрітому електронному газі ($T_e > T_\phi$) мобільність електронів спадає пропорційно $\mathcal{E}^{-1/2}$.

Питома провідність із урахуванням (4.176):

$$\sigma = en\mu = en\mu_0 \mathcal{E}^{-1/2} = . \quad (4.177)$$

Висновок: у розігрітому електронному газі ($T_e > T_\phi$) питома провідність зменшується пропорційно $\mathcal{E}^{-1/2}$ (рис.4.16, $n = -0,5$).

Щільність струму:

$$j = \sigma_0 \mathcal{E}^{-1/2} \cdot \mathcal{E} = \sigma_0 \mathcal{E}^{1/2} . \quad (4.178)$$



Рис.4.16. Залежність провідності від напруженості поля.

Число n – показник степеню у залежностях $\sigma \sim \mathcal{E}^n$.

Висновок: у розігрітому електронному газі ($T_e > T_\phi$) щільність струму зростає пропорційно $\mathcal{E}^{1/2}$ (рис.4.15), що означає **порушення закону Ома**.

Отримані закономірності базуються на тому, що у розігрітому електронному газі («теплі» електрони) частина енергії, отримана від зовнішнього поля, витрачається на підтримку підвищеної середньої температури електронного газу відносно температури напівпровідника. Фононне розсіяння посилюється за рахунок зменшення часу фононної релаксації і саме це дає змогу встановити стаціонарну рівновагу у сильному полі. На дрейфовий рух електронів, тобто на електричний струм, витрачається

лише частина енергії зовнішнього поля, тому пряма пропорційність між щільністю струму і напруженістю поля порушується.

Розсіяння на оптичних фононах. Друге критичне поле. Можливості досягнення стаціонарної рівноваги для акустичного фононного розсіяння закінчуються, коли енергія фононів досягає значення $(\hbar\omega_q)_{\max}$. У разі перевищення цього значення стаціонарна рівновага може встановлюватися лише за рахунок розсіяння на оптичних фононах LO (рис.4.14).

Значення енергії оптичних фононів $\hbar\omega_o$ та температури Дебая θ_D для деяких напівпровідників наведені у таблиці 4.4.

Таблиця 4.4

Напівпровідник	$\hbar\omega_o$, eB	θ_D , K
GaAs	0,035	408
Ge	0,037	430
Si	0,055	730

Дані, наведені у таб.4.4, свідчать про те, що енергія оптичних фононів у 1.5 – 2 рази більша за $k_B T_\phi = 0,026$ eB, тому збудження оптичних фононів мало імовірне. Температура Дебая θ_D , яка за фізичним змістом є температурою, вище якої збуджуються теплові коливання атомів усіх можливих для даного твердого тіла частот, виявляється вищою за 300 K. Зважаючи на це, висновок такий: *у діапазоні кімнатних температур оптичні фонони практично відсутні.*

Електрони будуть збуджувати оптичні фонони, якщо їх енергія перевищить $\hbar\omega_o$, але обмінної взаємодії не буде. Закон збереження енергії у таких умовах має вигляд:

$$E' = E - \hbar\omega_0. \quad (4.179)$$

Закон збереження (4.179) характеризує *абсолютно непружне розсіяння на оптичних фононах.*

Оптичні фонони мають ще одну особливість – це протифазні коливання атомів у межах елементарних комірок, тобто, *стоячі хвилі*. Енергія таких

коливань практично не залежить від хвильового вектора, групова швидкість відсутня.

Складемо баланс потужностей для другого критичного поля \mathcal{E}_2 :

$$e\mathcal{E}_2 v_{dp} = \frac{\Delta E}{\tau_\phi} = \frac{v_{T\phi}}{L_\phi} \hbar\omega_0 th \frac{\hbar\omega_0}{2k_B T_\phi}. \quad (4.180)$$

Дрейфова швидкість у полі \mathcal{E}_2 :

$$v_{dp} = \frac{e}{m^*} \tau_\phi \mathcal{E}_2 = \frac{e}{m^*} \cdot \frac{L_\phi}{v_{T\phi}} \cdot \mathcal{E}_2. \quad (4.181)$$

Розв'язок рівнянь (4.180), (4.181) відносно \mathcal{E}_2 :

$$\mathcal{E}_2 = \frac{v_{T\phi}}{eL_\phi} (m^* \hbar\omega_0 th \frac{\hbar\omega_0}{2k_B T_\phi})^{1/2}. \quad (4.182)$$

Розв'язок рівнянь (4.180), (4.181) відносно v_{dp} :

$$v_{dp} = \left(\frac{\hbar\omega_0}{m^*} th \frac{\hbar\omega_0}{2k_B T_\phi} \right)^{1/2}. \quad (4.183)$$

Числова оцінка аргументу гіперболічного тангенсу у формулах (4.182), (4.183) із використання даних таб.4.4 та (4.154) для 300 К дає:

$$\frac{\hbar\omega_0}{2k_B T_\phi} \approx 1, \quad th \frac{\hbar\omega_0}{2k_B T_\phi} \approx 1. \quad (4.184)$$

Наслідком (4.184) для дрейфової швидкості є те, що вона насичується на рівні

$$v_{dp} = v_s = \left(\frac{\hbar\omega_0}{m^*} \right)^{1/2} \quad (4.185)$$

і перестане залежати від напруженості поля (рис.4.15, “гарячі” електрони, $n=0$).

Із урахуванням (4.184), напруженість другого критичного поля

$$\mathcal{E}_2 = \frac{v_{T\phi}}{eL_\phi} (m^* \hbar\omega_0)^{1/2}. \quad (4.186)$$

Числові оцінки другого критичного поля \mathcal{E}_2 (4.186) та дрейфової швидкості (4.185) за типовими для ковалентних напівпровідників параметрами дає значення поля у межах $10^3 - 10^4$ В/см, дрейфової швидкості $v_s \approx (5 - 7) \cdot 10^6$ см/с.

Мобільність електронів в умовах розсіяння на оптичних фононах:

$$\mu = v_s / \mathcal{E}; \quad \mu \sim \mathcal{E}^{-1}. \quad (4.187)$$

Висновок: в умовах розсіяння на оптичних фононах мобільність у полях, напруженість яких більша другого критичного, спадає обернено пропорційно напруженості поля.

Питома провідність в умовах розсіяння на оптичних фононах:

$$\sigma = en \frac{v_s}{\mathcal{E}}; \quad \sigma \sim \mathcal{E}^{-1}. \quad (4.188)$$

Висновок: в умовах розсіяння на оптичних фононах питома провідність спадає обернено пропорційно напруженості поля (рис.4.16, значення показника $n = -0,8$ замість теоретичного $n = -1$ пов'язане із наявністю у p -Ge дірок різної ефективної маси).

Щільність струму:

$$j = ev_s = j_s = en(\hbar\omega_o / m^*)^{1/2} \sim \mathcal{E}^0. \quad (4.189)$$

Висновок: в умовах розсіяння на оптичних фононах щільність струму насичується і перестає залежати від напруженості поля (рис.4.15, “гарячі” електрони, $n = 0$).

Експериментальні залежності динамічних параметрів у сильних полях, наведені на рисунках 4.15, 4.16 у цілому підтверджують результати теоретичних розрахунків.

Концентраційні ефекти сильного поля. У попередньому розділі розглянутий вплив сильного електричного поля на динамічні характеристики носіїв заряду у напівпровідниках. Поряд із цим, сильне електричне поле може змінювати і концентрацію носіїв заряду у ізотермічних умовах.

Ударна іонізація. Спостерігається у електричних полях $10^5 - 10^6$ В/см (температура 300 К). Такі поля можна створити лише у межах просторового заряду $p - n$ – переходу, де незначна кількість вільних носіїв заряду.

Електрон, кінетична енергія якого сягає рівня

$$E_i \geq 1,5E_g, \quad (4.190)$$

здатний ударно іонізувати валентний зв'язок, створюючи електронно – діркову пару. Первинний електрон, енергія якого складається із декількох квантів $1,5E_g$, може створити відповідну кількість вторинних електронно – діркових пар (лавинна ударна іонізація).

У найпростішому випадку – один електрон створює одну електронно – діркову пару (знак рівності в умові (4.190)) – енергія первинного електрона E_i рівномірно розподіляється між трьома носіями по $E_g / 2$ на кожного.

Таблиця 4.5		
Напівпро відник	E_i , еВ	L_{of} , Å
$n - Ge$	1	100
$n - Si$	3	75
$p - Si$	5	55
$n - GaAs$ $p - GaAs$	4	58

Можливі два механізми набору енергії E_i – дрейфовий і дифузійний. Який із механізмів буде домінуючим залежить від довжини вільного пробігу електронів під час розсіяння на оптичних фононах L_{of} . У таб.4.5 наведені значення E_i та L_{of} для деяких напівпровідників.

У відносно слабких полях ($\mathcal{E}_1 < \mathcal{E} < \mathcal{E}_2$), але значній довжині вільного пробігу (15 – 20 постійних кристалічної структури), енергія ударної іонізації набирається на шляху $L_i < L_{of}$, тобто, розсіяння на оптичних фононах «не працює» (рис. 4.17).

У відносно слабких полях ($\mathcal{E}_1 < \mathcal{E} < \mathcal{E}_2$), але значній довжині вільного пробігу (15 – 20 постійних кристалічної структури), енергія ударної іонізації набирається на шляху $L_i < L_{of}$, розсіяння на оптичних фононах «не працює» (рис. 4.17).

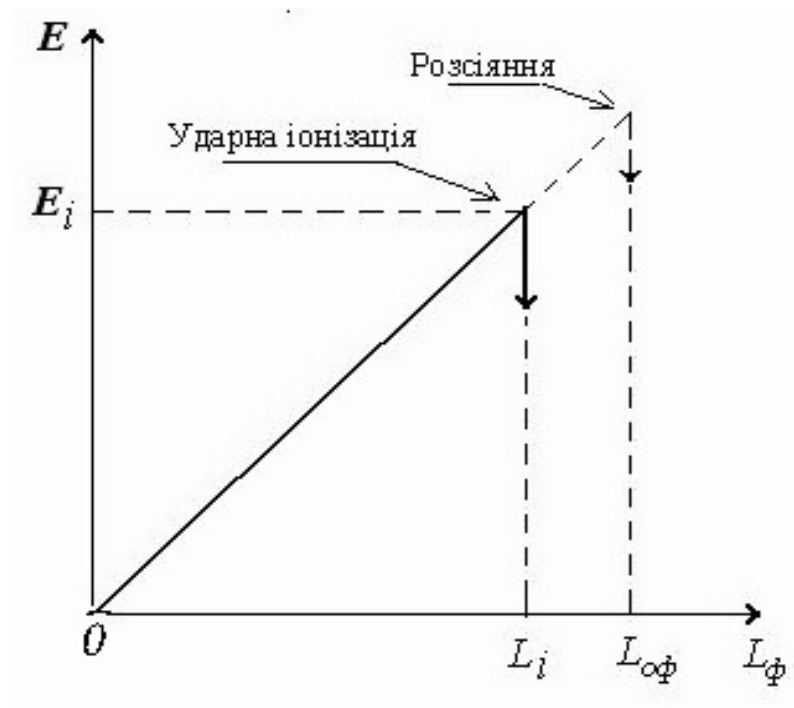


Рис.4.17. Дрейфовий механізм набору енергії для ударної іонізації. Необхідне значення енергії досягається раніше ніж народжуються оптичні фони

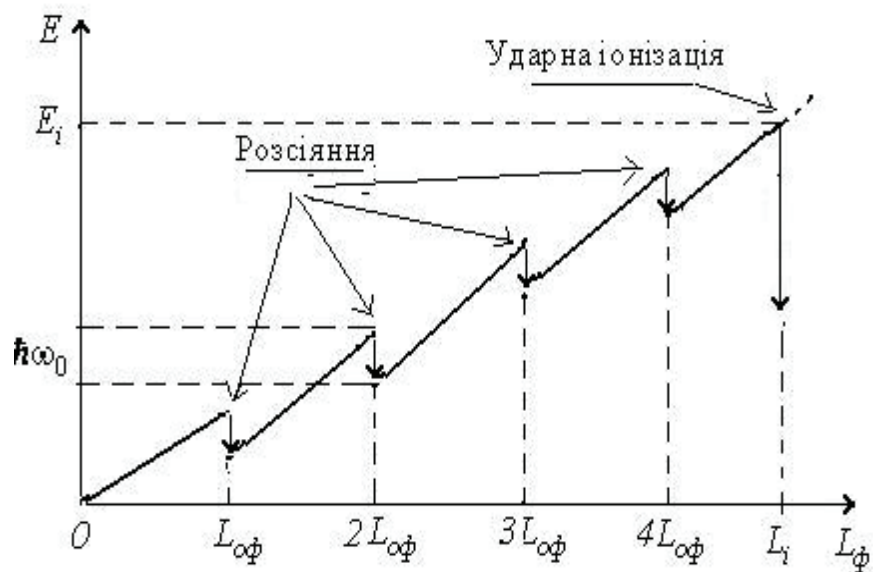


Рис.4.18. Дифузійний механізм набору енергії для ударної іонізації. Енергія електрона зростає незважаючи на те, що розсіяння на оптичних фонах кожного разу зменшує її на $\hbar\omega_0$.

У сильних полях ($\mathcal{E} \gg \mathcal{E}_2$, $e\mathcal{E}L_{оф} \gg \hbar\omega_0$) енергія електронів зростає

незважаючи на багаторазову втрату $\hbar\omega_0$ у момент народження оптичного фонона (рис.4.18) аж до досягнення значення E_i .

У *Ge* і *Si* в основному дрейфовий механізм набору енергії іонізації. У n^+ – та p^+ – *Si* спостерігається дифузійний механізм ($L_i > L_{of}$, $\mathcal{E} \geq 5 \cdot 10^5$ В/см).

Фононне та іонне розсіяння у діапазоні криогенних температур (5 – 20 К) настільки послаблюється, що ударна іонізація донорів і акцепторів спостерігається навіть у полях 5 – 10 В/см. Цьому сприяє мала енергія іонізації домішок ($E_i \leq 0,01$ еВ), висока мобільність – майже на 2 порядки вища мобільності при 300 К. Для прикладу, на рис.4.19 показана залежність щільності струму від напруженості поля у n – *Ge* ($T = 4,2$ К). У електричному полі > 5 В/см щільність струму зростає на 8 – 10 порядків. Такий значний стрибок струму можна називати переходом діелектрик – напівпровідник, стимульованим електричним полем. У фізичних дослідженнях ефект використовується для збільшення кількості носіїв заряду без участі термоіонізації.

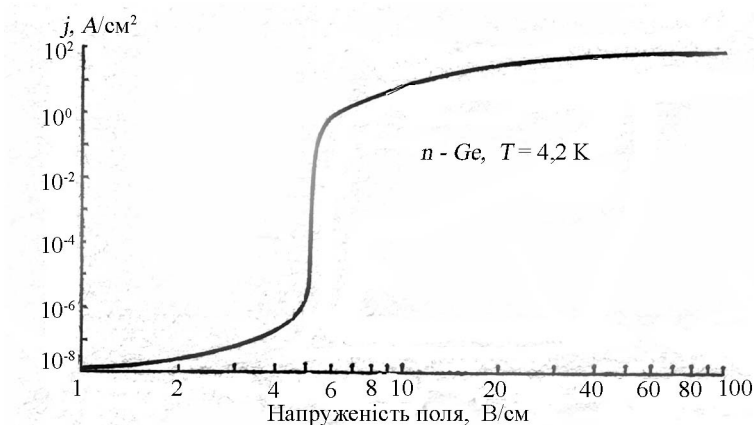


Рис.4.19. Ударна іонізація донорів у германії (криогенні температури). Енергія іонізації донорів на 2 – 3 порядки менша ширини забороненої зони, мобільність електронів майже на 2 порядки більша у цьому діапазоні температур. Достатньо напруженості поля 5 – 6 В/см для початку ударної іонізації домішок.

Тунельний ефект. Ізоенергетичне тунелювання через заборонену зону можна спостерігати у дуже сильних полях $\mathcal{E} \geq 10^6$ В/см за рахунок нахилу енергетичних зон.

У відсутності зовнішнього поля заборонена зона – енергетичний бар’єр, що має висоту E_g і необмежений по ширині $\Delta x \rightarrow \infty$. Тунельний перехід через такий бар’єр неможливий (рис.4.20(a)).

У зовнішньому електричному полі $\mathcal{E} = -\frac{d\phi(x)}{dx}$ до повної енергії електронів E додається потенціальна енергія поля $-e\phi(x)$. Наслідками цього є нахил енергетичних зон і трикутний бар’єр обмеженої ширини Δx і висотою E_g (рис.4.20(b)).

Із умови балансу енергій $e\mathcal{E}\Delta x = E_g$ знаходимо залежність ширини бар’єру Δx від напруженості поля:

$$\Delta x = E_g / e\mathcal{E}. \quad (4.191)$$

Квантова механіка для імовірності тунелювання (прозорості) у залежності від висоти, ширини бар’єру та ефективної маси електрона дає таку закономірність:

$$W \sim \exp(-A\sqrt{2m^*E_g} \cdot \Delta x), \quad (4.193)$$

де у A зібрані числові константи.

Підставивши (4.192) у (4.193), отримаємо:

$$W \sim \exp(-A\sqrt{2m^*} \cdot E_g^{3/2} / \mathcal{E}). \quad (4.194)$$

Імовірності тунельних переходів (кількість переходів за одиницю часу) із валентної зони у зону провідності і навпаки однакові, але кількість електронів у валентній зоні набагато більша, ніж у зоні провідності, тому домінуючими будуть переходи у зону провідності.

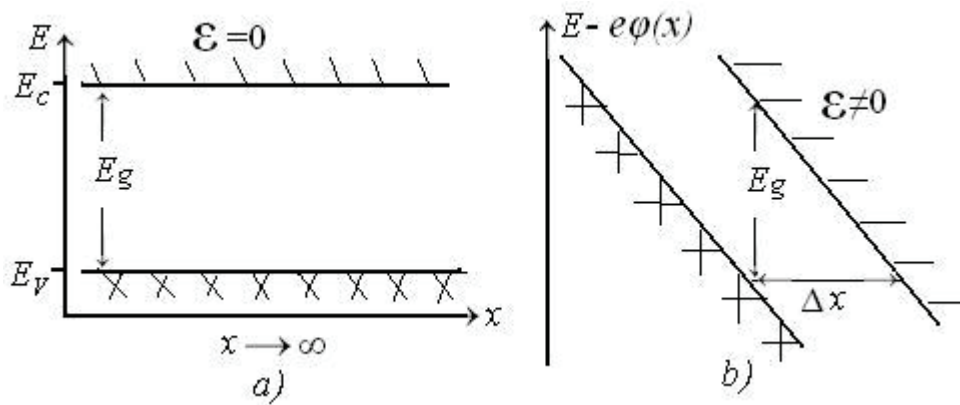


Рис.4.20. Енергетична діаграма напівпровідника у відсутності зовнішнього поля (а). Утворення потенціального бар'єру кінцевої ширини Δx (b), наслідком чого стають можливими тунельні переходи із валентної зони у зону провідності

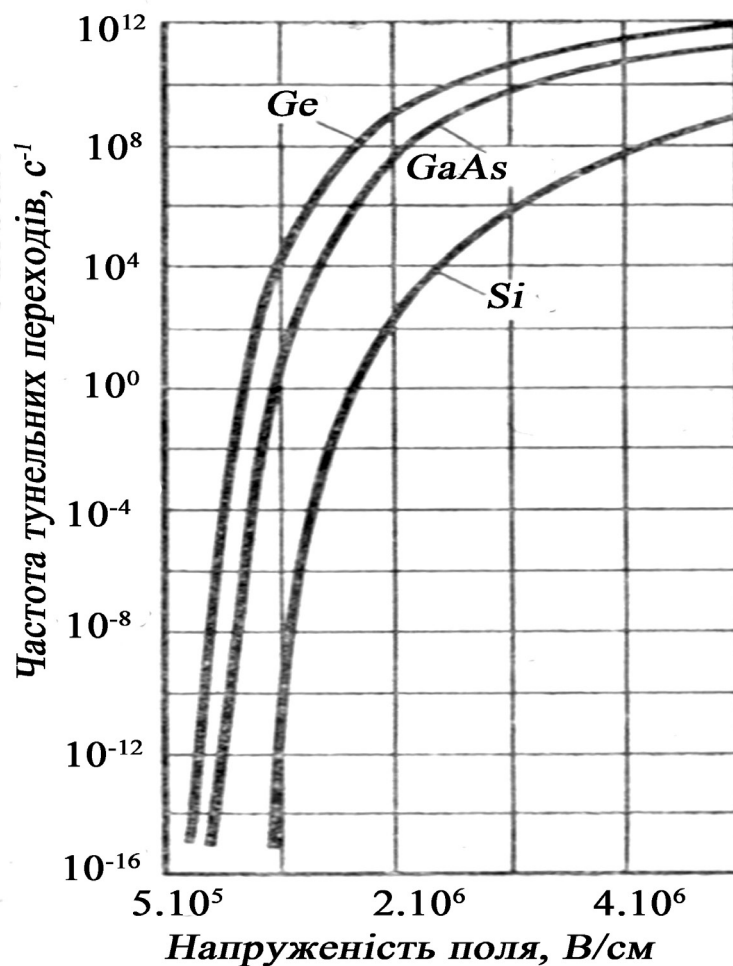


Рис.4.21. Залежність частоти тунельних переходів від напруженості поля для деяких напівпровідників

Імовірності тунельних переходів (кількість переходів за одиницю часу) із валентної зони у зону провідності і навпаки однакові, але кількість електронів у валентній зоні набагато більша, ніж у зоні провідності, тому домінуючими будуть переходи у зону провідності.

На рис.4.21 співставлені частоти тунельних переходів для *Ge*, *Si* і *GaAs* у залежності від напруженості поля. Спираючись на (4.194) в умовах однакової для вказаних матеріалів напруженості поля, робимо порівняльний аналіз: у *Ge* частота переходів більша ніж у *Si* за рахунок відносно малої ширини забороненої зони і ефективної маси електронів; у *GaAs* частота переходів більша ніж у *Si* за рахунок майже на порядок меншої ефективної маси електронів; у *Ge* частота переходів більша ніж у *GaAs* за рахунок удвічі меншої ширини забороненої зони.

Електричні поля із напруженістю $\mathcal{E} \geq 10^6$ В/см утворюються у зоні просторового заряду $p^+ - n^+$ – переходів.

Термоелектронна іонізація (ефекти Френкеля і Пуля). У випадках часткової іонізації донорів (низькі температури і "мілкі" донори або високі температури і "глибокі" донори) зовнішнє електричне поле може збільшувати імовірність термоіонізації донорів.

На рис.4.22(пунктирні лінії) зображена потенціальна діаграма у зоні неіонізованого донора без зовнішнього електричного поля. Енергія теплового руху $k_B T < \Delta E_d$ і недостатня для звільнення електрона. Умовою звільнення електрона є $U(x) > 0$. Потенціальне поле донора можна вважати квазікулонівським.

Зовнішнє електричне поле змінює потенціальну діаграму у зоні донора, як це показано на рис.4.22(суцільні лінії). Спільне потенціальне поле тепер має вигляд:

$$U(x) = -e\mathcal{E}x - e^2 / (\epsilon\epsilon_0 x)_{\mathcal{E}=0}, \quad (4.195)$$

де $e\mathcal{E}x$ - потенціальна енергія, внесена зовнішнім полем \mathcal{E} ; $\varepsilon\varepsilon_0$ - діелектрична проникність напівпровідника.

У точці x_m маємо екстремум, умовою якого є

$$dU(x) / dx = -e\mathcal{E} + e^2 / (\varepsilon\varepsilon_0 x_m)^2 = 0 . \quad (4.196)$$

Із (П.6) знаходимо координати екстремуму:

$$x_m = \sqrt{e / \varepsilon\varepsilon_0 \mathcal{E}} . \quad (4.197)$$

Після підстановки (4.195) у (4.197) можна визначити зменшення кулонівського потенціалу $\Delta U(x_m)$:

$$\Delta U(x_m) = U(x_m) - e^2 / (\varepsilon\varepsilon_0 x_m \sqrt{e / \varepsilon\varepsilon_0 \mathcal{E}}) = -\sqrt{e^3 \mathcal{E} / \varepsilon\varepsilon_0} . \quad (4.198)$$

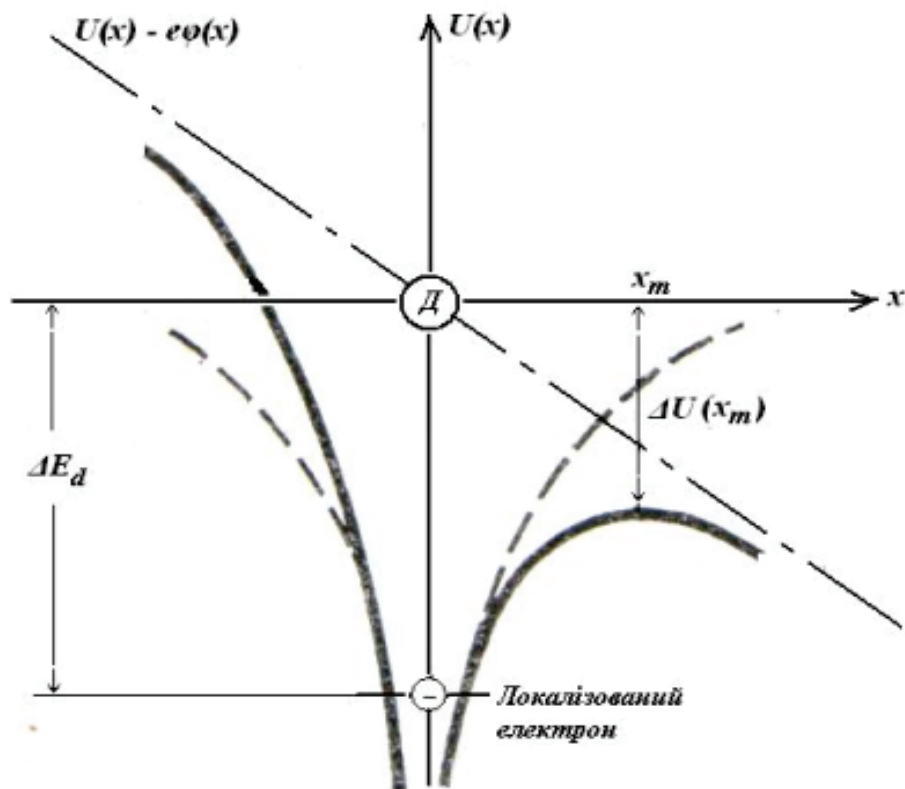


Рис.4.22. Потенціальна енергія у зоні неіонізованого донора без поля (пунктир). За наявності поля (суцільні лінії) енергія іонізації донора зменшується на $\Delta U(x_m)$

Висновок: зменшення потенціального бар'єру $\Delta U(x_m)$ пропорційне $\sqrt{\mathcal{E}}$.

Імовірність термоіонізації W_i , і питома провідність σ тепер стають залежними від напруженості поля:

$$W_i \sim \exp(-\Delta U(x_m) / k_B T); \sigma \sim \exp(-A\sqrt{\mathcal{E}} / k_B T), \quad (4.199)$$

де A – константи, які є у (4.198).

Висновки:

– однією із необхідних умов для ефекту Френкеля є та, що кількість неіонізованих донорів повинна суттєво перевищувати кількість вільних електронів;

– закономірності типу (4.199) (ефект Френкеля) та $\sigma \sim \exp(-A' \frac{\mathcal{E}}{k_B T})$ (ефект Пуля)

спостерігаються у напівпровідникових сполуках (переважно оксидах та халькогенідах). Для прикладу, на рис.4.23 наведені залежності питомої провідності від напруженості електричного поля для сполук V_2O_5 (ефект Пуля) та Sb_2S_3 (ефект Френкеля);

– основним джерелом енергії для подолання потенціального бар'єру між локалізованим станом електрона і вільним є теплова, електричне поле лише зменшує цей бар'єр.

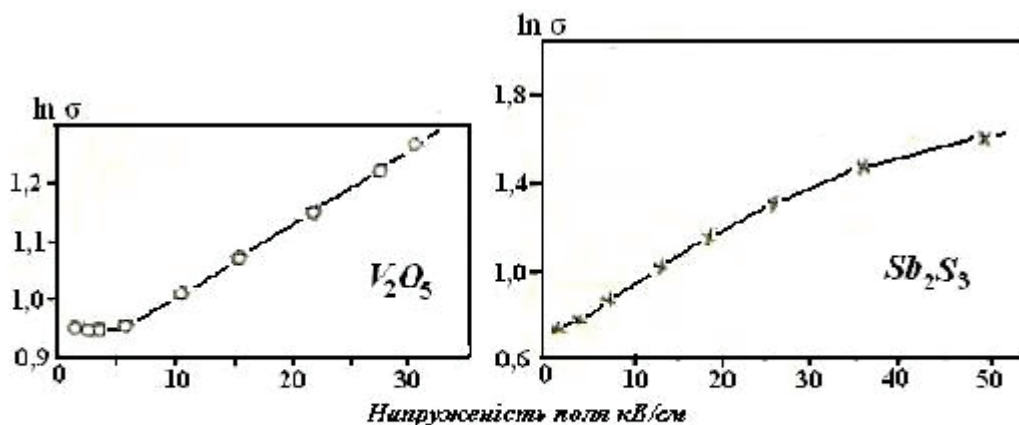


Рис. П.7. Термоелектронна іонізація. Залежність провідності від напруженості поля у напівлогарифмічному масштабі (розмірність σ у $(\text{Ом}\cdot\text{см})^{-1}$). Ефект Пуля – V_2O_5 ефект Френкеля – Sb_2S_3

Резюме по розд. 4.13

1. *Вплив електричного поля* на динамічні характеристики носіїв заряду та їх концентрацію у напівпровідниках можна *спостерігати і вивчати лише в ізотермічних умовах* – теплота Джоуля, яка супроводжує струм, повинна повністю розсіюватися у зовнішнє середовище.
2. *Зовнішній критерій* сильного електричного поля – *відхилення від закону Ома*.
3. **Слабкими** вважаються електричні поля, для яких енергія, передана носіям заряду, значно менша енергетичних втрат при розсіянні на подовжніх акустичних фонах. Електричний струм у таких умовах прямо пропорційний напруженості поля. Діє закон Ома.
4. Відхилення від закону Ома починаються з моменту, коли енергія, передана носіям заряду від поля, починає перевищувати енергетичні втрати при розсіянні. Такі поля називаються **сильними**. Напруженість поля, з якої починаються зміни динамічних параметрів носіїв заряду, отримала назву **перше критичне поле**. У числовому вимірі для ковалентних і близьких до них напівпровідників напруженості першого критичного поля лежать у межах $10^4 - 10^5$ В/см.
5. Стаціонарна рівновага у полях, більших першого критичного, означає рівність отриманої від поля енергії і розсіяної на фонах. Досягається така рівність за рахунок того, що енергетичні втрати при розсіянні на акустичних фонах зростають в міру зростання енергії електронів, час фонної релаксації і мобільність зменшуються і, тим самим, компенсують отриману від поля енергію.
6. Енергія сильного поля розділяється на дві частини – одна підтримує посилене фонне розсіяння, інша – дрейфовий рух, тобто електричний струм.

7. Зменшення енергії дрейфового руху порушує пряму пропорційність між струмом і напруженістю поля. Залежності мобільності і провідності від поля тепер мають вигляд $\mu, \sigma \sim \mathcal{E}^{-1/2}$, струму $j \sim \mathcal{E}^{1/2}$, **закон Ома не виконується**. Електронний газ стає „теплим”.

8. Посилення акустичного фононного розсіяння обмежене максимальною енергією цих фононів. По досягненні і перевищенні цієї енергії електрони починають *народжувати оптичні фонони*, починається абсолютно непружне **розсіяння на оптичних фонах**.

9. Напруженість поля, необхідна для переходу на оптичне розсіювання, називається **другим критичним полем**. У числовому вимірі напруженості другого критичного поля лежать у межах $10^5 - 10^6$ В/см.

10. Абсолютно непружне оптичне фононне розсіяння означає, що електрони лише генерують оптичні фонони, які є стоячими хвилями, і самі втрачають групову швидкість. Залежності мобільності і провідності від поля тепер мають вигляд $\mu, \sigma \sim \mathcal{E}^{-1}$, струму $j \sim \mathcal{E}^0$, тобто струм насичується і перестає залежати від поля. Електронний газ стає „гарячим”.

11. Енергія електронів у електричному полі може досягти і перевищити енергію ударної іонізації валентних зв'язків $E_i \geq 1,5E_g$. У числовому вимірі це поля з напруженістю $> 10^5 - 10^6$ В/см, які для напівпровідників можна створити лише у області просторового заряду $p - n$ – переходу.

12. Кожен первинний електрон із енергією $E_i \geq 1,5E_g$ створює щонайменше одну електронно – діркову пару. Початкова енергія рівномірно розподіляється між трьома носіями – два електрони і дірка.

13. Сильне зовнішнє поле може створити умови для тунельних переходів електронів із валентної зони у зону провідності.

14. Потенціальна енергія зовнішнього поля, складаючись із потенціальною енергією кристала, нахилиє енергетичні зони, утворює енергетичний бар'єр між зонами висотою E_g і з кінцевою шириною, залежною від напруженості поля. Ізоенергетичне тунелювання через такий бар'єр має зовнішній прояв – збільшення концентрації вільних носіїв заряду без участі термоіонізації.

15. Ще один концентраційний ефект спостерігається у напівпровідниках із незначною іонізацією донорів за умови $k_B T < \Delta E_d$. Потенціал зовнішнього поля, складаючись із квазікулонівським потенціалом у зоні донорів, зменшує потенціальний бар'єр між локалізованими і вільними станами електронів і тим стимулює інтенсивну термоіонізацію. Зовнішній прояв ефекту – експонентна залежність провідності від напруженості поля $\sigma \sim \exp(-A\sqrt{\mathcal{E}} / k_B T)$ - ефект Френкеля.

5. Рекомбінація носіїв заряду у напівпровідниках

5.1. Теплова генерація і рекомбінація

При фіксованій температурі напівпровідника концентрація вільних носіїв заряду (електронів і дірок), які утворилися виключно внаслідок термоіонізації валентних зв'язків та електрично активних домішок, і їх середня теплова енергія $k_B T$ також фіксовані. У подальшому концентрації n_0 , p_0 , n_i будемо називати *тепловими рівноважними*, а процес її утворення – *термогенерацією*.

Характеристикою термогенерації є її *темпл* або *швидкість* G_T – число вільних носіїв заряду, які утворилися в одиниці об'єму за одиницю часу. Розмірність $[G_T] = \text{м}^{-3} \cdot \text{с}^{-1}$.

Незалежність теплової рівноважної концентрації від часу забезпечується рівністю темпів термогенерації G_T і рекомбінації R_T – число носіїв заряду, які рекомбінують в одиниці об'єму за одиницю часу.

Розмірність $[R_T] = \text{м}^{-3} \cdot \text{с}^{-1}$.

Іншими словами, у стані теплової рівноваги відбувається рівноімовірний обмін носіями заряду між зонами вільного руху і локалізованими станами на валентних зв'язках і на домішках. Такий обмін супроводжується зміною енергії носіїв заряду шляхом поглинання та народження квантів теплової коливальної енергії, тобто фононів.

Збуджені при термогенерації електрони, поглинаючи фонони, утворюють “дірки” у фононному енергетичному спектрі. При рекомбінації електрони народжують фонони, які заповнюють ці “дірки”. Незмінною для термодинамічної рівноваги є умова:

$$G_T = R_T. \quad (5.1)$$

Частота обмінних переходів у стані теплової рівноваги близько 10^{13} Гц (для температури 300 K). Середній час перебування окремо взятого носія заряду у вільному стані $10^{-13} \dots 10^{-12}$ с. За цей час електрон, рухаючись із швидкістю $\approx 2,5 \cdot 10^5$ м/с, „знаходить” дірку із відповідним у цей момент напрямком спіна (або іонізований донор, чи неіонізований акцептор) і локалізується на певний час. Наступна взаємодія із фононом (поглинання фонона) знову термоіонізує валентний зв’язок або домішку і все починається з початку.

У таких обмінних переходах є одна особливість — кількість електронів незмінна, кількість фононів – змінна.

5.2. Нетеплова генерація і рекомбінація

Вільні носії заряду у зонах можна отримати не лише за рахунок термогенерації, а і за рахунок інших джерел енергії, наприклад світлової, електричного поля, радіаційного опромінення та ін. У цих випадках говорять про *нетеплову генерація*. Повна концентрація електронів або дірок буде складатися із двох частин – теплової рівноважної n_0 , p_0 , n_i і надлишкової, не рівноважної Δn або Δp , які зумовлені нетепловими джерелами енергії,:

$$n = n_0 + \Delta n, p = p_0 + \Delta p. \quad (5.2)$$

Наявність надлишкових носіїв заряду порушує енергетичну і концентраційну рівновагу у зоні провідності або у валентній зоні. За час фононної релаксації надлишкові носії стають енергетично рівноважними у межах відповідної зони, але концентраційна рівновага залишається порушеною.

Стаціонарна рівновага (незалежність надлишкової концентрації від часу) встановлюється за умови, що темп нетеплової генерації G і темп рекомбінації надлишкових носіїв заряду R однакові:

$$G = R. \quad (5.3)$$

При зовнішній подібності умов (5.1) і (5.3) вони характеризують принципово різні процеси. Надлишкові носії заряду виникають не за рахунок поглинання фононів і не створюють “дірки” у фононному енергетичному спектрі. Енергія, яку вони повинні втратити при рекомбінації, найчастіше виходить за межі енергії фононного спектру і не може безпосередньо трансформуватися у теплові коливання атомів. Необхідний детальний аналіз рекомбінаційних процесів для надлишкових носіїв заряду.

Рекомбінація – процес відновлення концентраційної рівноваги, порушеної нетепловими механізмами генерації. За будь – яких інших умов, рекомбінація зводиться, у кінцевому підсумку, до відновлення стану термодинамічної концентраційної рівноваги.

Зміна енергії кожного окремо взятого електрона при квантових переходах відбувається миттєво, але кількість переходів за одиницю часу (*частота, імовірність переходів*) обмежена і для відновлення концентраційної рівноваги потрібен певний проміжок часу.

Темп (швидкість) рекомбінації R – число електронно – діркових пар, рекомбінуючих за одиницю часу у одиниці об’єму. Для масових процесів:

$$R = \frac{d\Delta n(t)}{dt} = -\alpha \Delta n(t), \quad (5.4)$$

де $\Delta n(t)$ – концентрація нерівноважних носіїв заряду у момент часу t ;

α – коефіцієнт пропорційності, який не залежить від Δn ;

“ – ” означає, що $\Delta n(t)$ із плином часу зменшується.

Розмірність коефіцієнта пропорційності α – обернений час. Позначимо цей час $\tau = 1/\alpha$ і для темпу рекомбінації маємо такі форми запису:

$$R = -\frac{d\Delta n}{dt} = -\frac{\Delta n}{\tau} = -\frac{n - n_0}{\tau}. \quad (5.4a)$$

У момент часу $t = 0$ нетеплова генерація припинилася, але рекомбінація продовжується. Інтегруємо рівняння (5.4) за таких початкових умов:

$$\text{при } t = 0, \Delta n(0); \quad \text{при } t \rightarrow \infty, \Delta n(t) \rightarrow 0. \quad (5.5)$$

Результат інтегрування – залежність надлишкової концентрації від часу при рекомбінації:

$$\Delta n(t) = \Delta n(0) \exp(-t/\tau). \quad (5.6)$$

Фізичний зміст параметра τ такий:

$$\Delta n(t = \tau) = \Delta n(0)/e \approx \Delta n(0)/2,7 \approx 0,37\Delta n(0), \quad (5.7)$$

тобто, τ це час, протягом якого надлишкова концентрація носіїв заряду у процесі рекомбінації зменшується у 2,7 рази. Із статистичної фізики відомо, що визначений таким чином при експонентному розподілі параметр τ є середній час процесу рекомбінації і він отримав назву **час життя надлишкових носіїв заряду**.

У подальшому розглядаються такі механізми рекомбінації – через захват, міжзонна і ударна (рекомбінація Оже).

5.3. Рекомбінація через захват

Центри захвату. Якщо енергія, яку повинні втратити надлишкові носії заряду, перевищує максимальну енергію у фононному спектрі, а умов для фотонної рекомбінації немає, основним механізмом стає фононна рекомбінація через проміжні *центри захвату (уловлювачі) – рекомбінація через захват*. Такий механізм рекомбінації є основним для ковалентних напівпровідників (германій, кремній, алмаз)

Домішкові атоми, які створюють центри захвату, розташовані між основними атомами (на відміну від донорів та акцепторів, які заміщують основні атоми), спотворюють потенціальне поле і утворюють глибокі потенціальні ями, здатні локалізувати електрони. Найбільш ефективними для утворення центрів захвату є домішки атомів Cu, Ag, Au, Fe, Ni, Mn та деяких

інших. Структурні дефекти (дислокації, радіаційні дефекти і т.п.) також створюють центри захвату.

Центр захвату *або нейтральний, або одноразово від'ємно заряджений* при захваті електрона. Енергетично центри захвату розташовані поблизу середини забороненої зони.

Електронні переходи. Темп рекомбінації. Час життя. Схема електронних переходів при рекомбінації через захват показана на рис.5.1. Переходи 1 (захват електрона) + 3 (захват дірки) – рекомбінаційні, 2, 4 – термоіонізація центрів захвату.

У загальному випадку час життя складається із часу очікування захвату електрона уловлювачем τ_{zn} і часу захвату дірки τ_{zp} цим же уловлювачем:

$$\tau = \tau_{zn} + \tau_{zp} \quad (5.8)$$

Вільні електрони і дірки незалежно рухаються у міжатомному просторі із кінцевою швидкістю, тому потрапляють у зону захвату лише через певний час. Захват стане можливим, якщо у цей момент уловлювач вільний для електрона або зайнятий електроном для дірки. Крім того, необхідний ще й певний миттєвий спіновий стан носіїв заряду. Але самі по собі захвати ще не гарантують рекомбінаційні переходи. Локалізовані на уловлювачах носії можуть повернутися у відповідні зони внаслідок термоіонізації. Все це вимагає багатьох спроб для захвату і, відповідно, певного часу для рекомбінації.

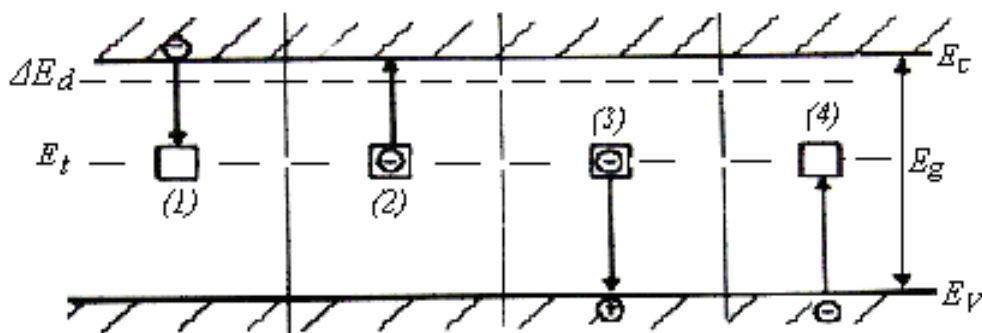


Рис. 5.1. Електронні переходи за наявності центрів захвату у забороненій зоні: (1) + (3) – рекомбінація; (2), (4) – термоіонізація

Просторові розміри потенціальних ям характеризують *перерізом захвату* S_t – площа діаметрального перерізу сфери, у центрі якої знаходиться домішковий атом або структурний дефект. *Імовірність захвату* за одиницю часу (*частота захватів*) w одним центром одного носія заряду із перерізом захвату S_t пов'язані простим співвідношенням:

$$w = v S_t, \quad (5.9)$$

де v – щільність потоку для одного носія заряду у одиниці об'єму (кількісно дорівнює швидкості носія). Розмірність $[w] = \text{м}^3 \cdot \text{с}^{-1}$.

Позначимо концентрацію центрів захвату N_t і визначимо повну імовірність захвату одного носія усіма центрами захвату за одиницю часу W :

$$W = w N_t = v S_t N_t. \quad (5.10)$$

Визначимо кількість локалізованих у одиниці об'єму за одиницю часу носіїв заряду R_z – *темп (швидкість) захвату*:

$$R_z = \Delta n W = \Delta n v S_t N_t = \Delta n / \tau_z, \quad (5.11)$$

де τ_z – середній час, необхідний для локалізації Δn носіїв заряду.

Визначимо τ_z із (5.11):

$$\tau_z = (v S_t N_t)^{-1}. \quad (5.12)$$

В умовах відсутності термоіонізації центрів захвату усі вони приймають участь у захваті і τ_z , визначений за формулою (5.12), це *мінімально можливий час захвату*.

Враховуючи те, що перерізи захвату для електронів S_{tn} і дірок S_{tp} неоднакові, визначимо окремо мінімально можливий час захвату електронів τ_{0n} і дірок τ_{0p} :

$$\tau_{0n} = (\nu S_{tn} N_t)^{-1}, \tau_{0p} = (\nu S_{tp} N_t)^{-1}. \quad (5.13)$$

Важливим параметром центрів захвату є їх енергетичне положення E_t у забороненій зоні. Одна і та ж домішка або структурний дефект можуть створювати декілька центрів із різними значеннями E_t . Наприклад, золото у кремнії створює два глибоких рівні у забороненій зоні. Рівні із $E_t = +0,35$ еВ (знак „+” означає, що відлік енергії від стелі валентної зони) мають суттєво різні імовірності захвату для електронів $w_n = \nu S_{tn} = 67 \cdot 10^{-9} \text{ см}^3 \cdot \text{с}^{-1}$ і дірок $w_p = \nu S_{tp} = 1,6 \cdot 10^{-9} \text{ см}^3 \cdot \text{с}^{-1}$ і не можуть бути ефективними центрами рекомбінації. Рівні із $E_t = -0,54$ еВ (відлік від дна зони провідності) мають співрозмірні імовірності захвату $w_n = 9,5 \cdot 10^{-9} \text{ см}^3 \cdot \text{с}^{-1}$, $w_p = 16 \cdot 10^{-9} \text{ см}^3 \cdot \text{с}^{-1}$ і, завдяки цьому, найбільш ефективні у процесах рекомбінації.

Формула Шоклі – Ріда. Статистична теорія рекомбінації через захват за наявності одного типу центрів, запропонована У. Шоклі і У. Рідом, дає універсальну формулу для часу життя :

$$\tau = \frac{n_0 + n_1 + \Delta n}{n_0 + p_0 + \Delta n} \tau_{0p} + \frac{p_0 + p_1 + \Delta p}{n_0 + p_0 + \Delta p} \tau_{0n}, \quad (5.14)$$

де n_1, p_1 – рівноважні концентрації електронів і дірок за умови, що рівень Фермі F_ϕ співпадає із $-E_t$:

$$n_1 = N_c \exp\left(\frac{-E_t}{k_B T}\right), \quad p_1 = N_v \exp\left(\frac{-\Delta E_0 + E_t}{k_B T}\right). \quad (5.15)$$

Добуток $n_1 \cdot p_1 = N_c N_v \exp(-\Delta E_0 / k_B T) = n_i^2$ від E_t не залежить і відповідає закону діючих мас для рівноважних носіїв заряду (3.32). Концентрації n_1, p_1 пропорційні імовірності термоіонізації центрів захвату.

Проведемо аналіз (5.14) для типових ситуацій.

Малий рівень інжекції: $\Delta n \ll n_0, n_1, p_0, p_1$. (5.16)

Врахуємо (5.16) у (5.14):

$$\tau_0 \approx \frac{n_0 + n_1}{n_0 + p_0} \tau_{0p} + \frac{p_0 + p_1}{n_0 + p_0} \tau_{0n}. \quad (5.17)$$

Висновок: *при малому рівні інжекції час життя не залежить від рівня інжекції.*

Малий рівень інжекції, вироджений напівпровідник:

$$n_0^+ \gg n_1, p_0, p_1, N_t. \quad (5.18)$$

Врахуємо (5.18) у (5.17):

$$\tau_0 = \tau_{0p} \approx (\nu S_{tp} N_t)^{-1}. \quad (5.19)$$

Аналогічний результат отримаємо для сильно легованого діркового напівпровідника:

$$p_0^+ \gg p_1, n_0, n_1, N_t; \quad \tau_0 = \tau_{0n} \approx (\nu S_{tn} N_t)^{-1}. \quad (5.20)$$

Висновок: *у сильно легованому напівпровіднику час життя визначається мінімально можливим часом захвату неосновних носіїв заряду.*

Фізичний зміст цього висновку полягає у тому, що рівноважні носії заряду постійно підтримують центри захвату у заповненому стані – у електронному напівпровіднику заселеними електронами, у дірковому – заселеними дірками, тобто пустими. Процес рекомбінації у таких умовах стає асиметричним, зводиться лише до очікування появи і захвату неосновних носіїв заряду – дірок у електронному, електронів у дірковому напівпровідниках.

Слід звернути увагу ще на деякі особливості рекомбінації у сильно легованих напівпровідниках – незалежність часу життя від рівноважних концентрацій n_0 , p_0 і відсутність явної залежності часу життя від температури. У формулах (5.19) і (5.20) температура входить неявно через теплову швидкість $\nu(T) \sim \sqrt{T}$, але це слабка залежність від температури і нею можна нехтувати.

Висновок: у сильно легованих (вироджених) напівпровідниках час життя не залежить від рівня легування і практично не залежить від температури.

Малий рівень інжекції, невироджений домішковий напівпровідник:

$$\left\{ \begin{array}{l} n_0 \gg p_0, p_1; \quad n_0 \cong n_1 \\ p_0 \gg n_0, n_1; \quad p_0 \cong p_1 \end{array} \right\}. \quad (5.21)$$

Врахуємо (5.21) у (5.17):

$$\tau_0 \approx (1 + \frac{n_1}{n_0})\tau_{0p}; \quad \tau_0 \approx (1 + \frac{p_1}{p_0})\tau_{0n}. \quad (5.22)$$

У формулах (5.22) з'явилася залежність часу життя від температури, оскільки n_1 і p_1 експонентні функції температури (5.15). Проведемо аналіз температурної залежності, поклавши $n_0 = N_d$, $p_0 = N_a$.

Низькі температури ($k_B T \ll E_t$). Рівень Фермі поблизу рівня донорів ΔE_d і суттєво вище рівня E_t . Термоінізація центрів захвату незначна $n_0 \gg n_1$.

Врахувавши цю нерівність і аналогічну нерівність для діркового напівпровідника $p_0 \approx p_1$ отримаємо:

$$\tau_0 \approx \tau_{0p}, \tau_0 \approx \tau_{0n}. \quad (5.23)$$

Висновок: при низьких температурах час життя неосновних носіїв заряду у невироджених напівпровідниках від температури практично не залежить і має мінімально можливе значення.

Підвищені температури. Рівень Фермі наблизився до рівня E_t . Концентрації n_1, p_1 стають співрозмірними із n_0, p_0 і навіть можуть їх перевершувати. Для спрощення записів далі продовжимо аналіз для електронного напівпровідника. Результати для діркового будуть аналогічними.

У формулі (5.22) для електронного напівпровідника замінимо n_0 на N_d , n_1 беремо із (5.15), враховуємо, що у слабо легованих напівпровідників $N_c \gg N_d$ і нехтуємо одиницею. У кінцевому підсумку маємо:

$$\tau_0 = \tau_{0p} \frac{N_c}{N_d} \exp\left(-\frac{E_t}{k_B T}\right). \quad (5.24)$$

Для наочного представлення $\tau_0(T)$ логарифмуємо (5.24):

$$\ln \tau_0 = \ln\left(\tau_{0p} \frac{N_c}{N_d}\right) - \frac{E_t}{k_B} \cdot \frac{1}{T}. \quad (5.25)$$

Отримано рівняння прямої лінії у координатах

$$\ln \tau_0(T) = f\left(\frac{1}{T}\right) \quad (5.26)$$

із тангенсом кута нахилу $-E_t/k_B$, що дає змогу розраховувати E_t по експериментальним залежностям часу життя від температури. Для підтвердження достовірності отриманої залежності (5.24) на рис.5.2 показана експериментальна залежність $\tau_p(T)$ для електронного германію.

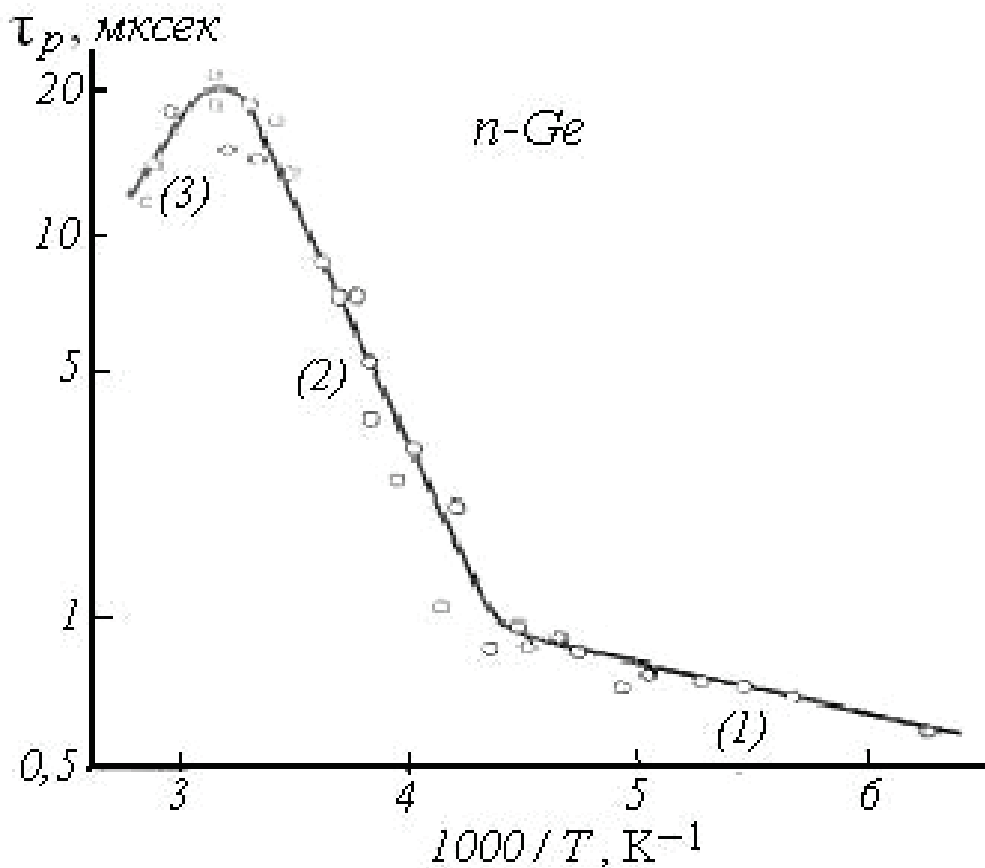


Рис.5.2. Залежність часу життя дірок у $n - \text{Ge}$ від температури

Значення глибини залягання центрів захвату E_t , розраховане на ділянці (2), дорівнює $\approx 0,25$ еВ. У діапазоні низьких температур (1) $T < 250$ К спостерігається певне зростання часу життя при збільшенні температури за рахунок того, у слабо легованих напівпровідників N_t і N_d співрозмірні і утримувати центри захвату повністю заселеними електронами не вдається, тому експериментальні результати дещо відрізняються від теоретичних (5.23).

У діапазоні підвищених температур рівень Фермі наближається до E_t і починається інтенсивна термоіонізація центрів захвату. Час життя зростає по експонентному закону (5.24). Зростання часу життя на ділянці (2) може бути

більше ніж на порядок. Верхньою температурною межею ділянки (2) є початок власної провідності T_i .

Висновок: у діапазоні проміжних температур $250 \text{ К} < T < T_i$ час життя неосновних носіїв заряду у невиродженому напівпровіднику експонентно зростає внаслідок інтенсивної термоіонізації центрів захвату. Максимального значення він досягає при наближенні до температури початку власної провідності T_i . За експериментально визначеним тангенсом кута нахилу у залежності (5.25) можна визначити E_t .

Високі температури, власна провідність. Поява власних носіїв заряду збільшує загальну кількість електронів у зоні провідності і необхідно замінити N_d на $N_d + n_i(T)$ у (5.24). Зростання $n_i(T)$ буде зменшувати час життя ((3) на рис.5.2.). Температурне зменшення часу життя при $T > T_i$ залежить від положення E_t відносно $\Delta E_0/2$, зокрема, при $E_t \approx \Delta E_0/2$ зменшення не буде, час життя насичується на певному рівні.

По досягненні ситуації, коли $n_i(T) \gg N_d$ і наповнення центрів захвату стане невизначеним, рекомбінація вимагатиме як захвату електронів, так і дірок:

$$\tau_{0i} \approx \tau_{0p} + \tau_{0n}. \quad (5.27)$$

Висновок: у температурному діапазоні власної провідності домішкового напівпровідника наповнення центрів захвату стає невизначеним і рекомбінація включає незалежний захват дірок і електронів. Час життя визначається лише перерізом захвату і концентрацією центрів захвату.

Якщо узагальнити вищенаведений аналіз температурної залежності рекомбінаційних процесів для невинроджених напівпровідників при малому рівні інжекції, то для часу життя маємо такі граничні значення:

$$\tau_{0p} \leq \tau_0 \leq \tau_{0p} + \tau_{0n}. \quad (5.28)$$

Високий рівень інжекції: $\Delta n \gg n_0, n_1, p_0, p_1. \quad (5.29)$

Нехтуємо рівноважними концентраціями у (5.14) і для часу життя, який умовно позначимо τ_∞ , отримаємо:

$$\tau_\infty = \tau_{0p} + \tau_{0n}. \quad (5.30)$$

Практично виконати умови (5.29) можливо лише для невинроджених напівпровідників. Формула (5.30) співпадає із (5.27) тому, що велика кількість нерівноважних носіїв заряду робить невизначеним заповнення центрів захвату так само, як це мало місце у випадку власної провідності при високих температурах.

Залежність часу життя від рівня легування. У попередніх пунктах аналізу розглядалися граничні випадки по рівню легування – сильно легований (винроджений) або слабо легований (невинроджений) напівпровідник. Розглянемо залежність часу життя від концентрації рівноважних носіїв заряду $n_0 = N_d$ на прикладі електронного германію (рис.5.3.).

Час життя дірок не залежить від n_0 , якщо $n_0 = N_d > 10^{23} \text{ см}^{-3}$, і це відповідає результату (5.19) для сильно легованого напівпровідника.

У діапазоні $10^{20} \leq n_0 < 10^{23} \text{ см}^{-3}$ (невироджений $n - Ge$) час життя дірок монотонно зростає. Причина зростання – наближення рівня Фермі до середини забороненої зони в міру зменшення N_d (див.(3.42)), поблизу якої і розташований E_t . Зростає імовірність термоіонізації центрів захвату і зменшується темп захвату дірок (рис.5.1). При $N_d < 10^{20} \text{ см}^{-3}$ германій стає власним ($n_i \approx 2,5 \cdot 10^{19} \text{ см}^{-3}$) і час життя дірок виходить на рівень $\tau_{0i} = \tau_{0p} + \tau_{0n}$.

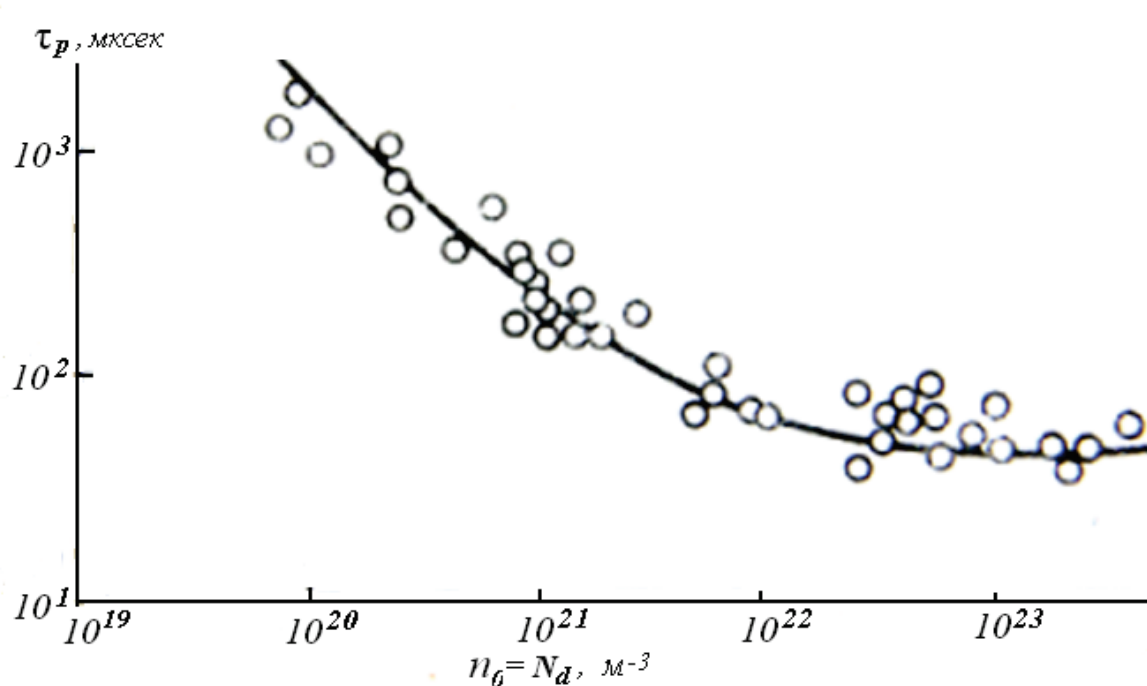


Рис.5.3. Залежність часу життя дірок у електронному германії від рівня легування донорами

Висновок: при фіксованому значенні температури і N_t час життя неосновних носіїв заряду залежить від рівня легування донорами. Діапазон зміни часу життя – від τ_{0p} у виродженому стані і до $\tau_{0p} + \tau_{0n}$ для власного напівпровідника. У числовому вимірі зростання часу життя може становити 1...2 порядки.

5.4. Поверхнева рекомбінація і мобільність

Загальні положення. Реальна поверхня напівпровідника характеризується значною кількістю різноманітних дефектів у порівнянні із об'ємом.

Частина цих дефектів посилює розсіяння носіїв заряду – мобільність у поверхневому прошарку зазвичай менше об'ємної. Розрізняють *дифузне розсіяння* на структурних дефектах (наприклад, мікротріщинах і т.п.) і *посилене іонне розсіяння* як результат неповного взаємного екранування потенціалів іонізованих домішок – двовимірна іонна ґратка. Загальний наслідок поверхневого розсіяння – зменшення мобільності у 2 – 3 рази відносно об'ємної за інших рівних обставин.

Інші дефекти (обірвані зв'язки, адсорбовані атоми і т.п.) створюють додаткові рекомбінаційні центри. Загальний наслідок – поверхневий час життя менше об'ємного.

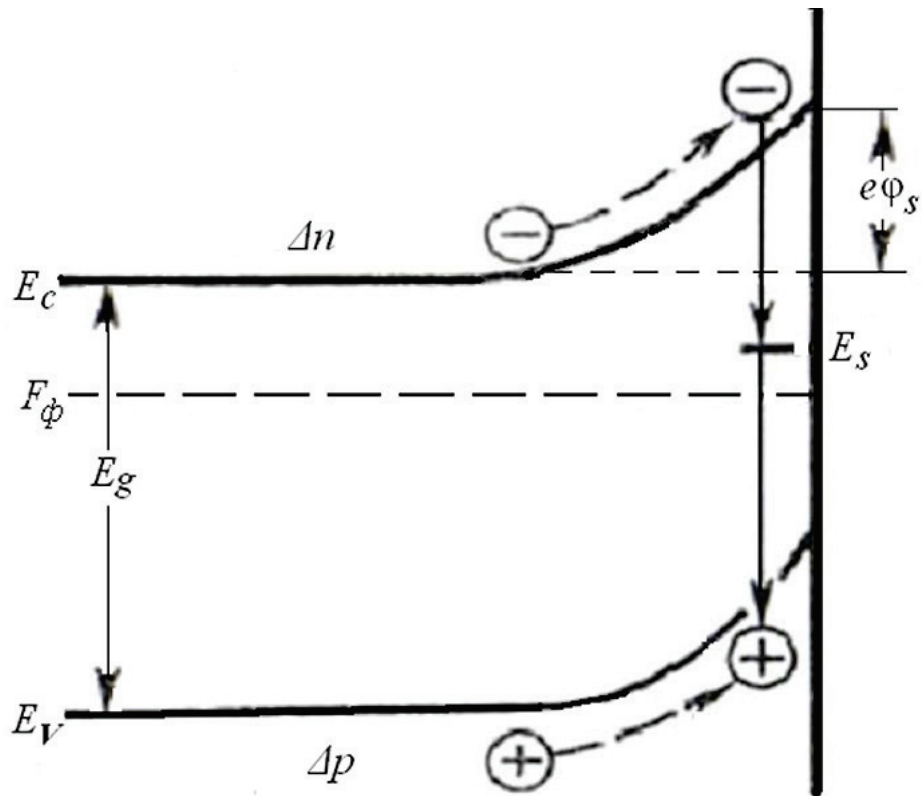


Рис.5.4. Енергетична діаграма поверхневого прошарку за наявності позитивного потенціалу $e\phi_s$ на поверхні. Рекомбінація іде через поверхневий рівень E_s .

Швидкість поверхневої рекомбінації. Наявність значної кількості рекомбінаційних центрів у поверхневому прошарку може привести до того, що темп поверхневої рекомбінації R_s буде перевершувати темп рекомбінації у об'ємі напівпровідника. Виникає потік рекомбінуючих нерівноважних носіїв заряду із об'єму до поверхні і зустрічний потік носіїв заряду, які звільняються із центрів захвату (рис. 5.4.). Різниця цих потоків і є темпом поверхневої рекомбінації R_s , який при малих рівнях інжекції пропорційний нерівноважній концентрації:

$$R_s = S\Delta n. \quad (5.31)$$

Розмірність щільності потоку $[R_s] = \text{м}^{-2}\text{с}^{-1}$, концентрації $[\Delta n] = \text{м}^{-3}$, коефіцієнта пропорційності $[S] = \text{м} / \text{с}$ – розмірність швидкості.

Коефіцієнт пропорційності $S = R_s / \Delta n$ називається швидкістю поверхневої рекомбінації. Фізичний зміст S – частота рекомбінаційних

переходів на 1 м^2 поверхні напівпровідника, нормована на концентрацію не рівноважних носіїв заряду, тобто $S = R_s$ при $\Delta n = 1 \text{ м}^{-3}$.

Швидкість поверхневої рекомбінації – основна кількісна характеристика рекомбінаційних процесів на поверхні і є еквівалентом часу життя для об’ємної рекомбінації. На рис.5.4 зображена енергетична діаграма приповерхневого прошарку за наявності електростатичного поверхневого потенціалу ϕ_s . Поверхневий потенціал утворюється або за рахунок адсорбованих на поверхні іонів, або за рахунок зовнішнього поля (ефект поля). Викривлення енергетичних зон у приповерхневому прошарку залежить від величини та знаку поверхневого потенціалу. Позитивний потенціал (межі зон викривлені угору) буде збагачувати поверхню електронами ($n_s = n \exp(e\phi_s / k_B T)$) за рахунок підтягування їх із об’єму, але відштовхувати дірки із поверхневого прошарку у об’єм ($p_s = p \exp(-e\phi_s / k_B T)$). Негативний потенціал (межі зон будуть викривлені униз), навпаки, – збагачує поверхню дірками ($p_s = p \exp(e\phi_s / k_B T)$), але збіднює на електрони ($n_s = n \exp(-e\phi_s / k_B T)$). За відсутності потенціалу ($\phi_s = 0$) викривлення зон не буде і $n_s = n = n_0 + \Delta n$, $p_s = p = p_0 + \Delta p$. Іншими словами, поверхневий потенціал впливає на рекомбінаційні потоки між поверхнею і об’ємом і, тим самим, на швидкість поверхневої рекомбінації. При певному значенні ϕ_s потік Δn або Δp може припинитися і $S \rightarrow 0$.

За будь – яких умов, стаціонарна рівновага встановлюється при рівності потоків із об’єму до поверхні і зустрічного, від поверхні у об’єм. При цьому швидкість поверхневої рекомбінації має максимальне значення. Поверхневий потенціал змінює положення рекомбінаційного рівня E_s відносно E_C і E_V , а значить і відносно рівня Фермі F_ϕ . Взаємне положення E_s і F_ϕ впливає на імовірності захвату і термоіонізації рекомбінаційних центрів.

Ефект поля використовується для вивчення кінетики поверхневої рекомбінації і визначення параметрів поверхневих центрів захвату.

Розподіл нерівноважних носіїв заряду у поверхневому прошарку.

Рекомбінаційні потоки нерівноважних носіїв заряду між об'ємом і поверхнею при малому рівні інжекції мають переважно дифузійний характер. Розглянемо електронний рівномірно легований ($dn_0/dx = 0$) напівпровідник, у об'ємі якого генеруються електронно – діркові пари $\Delta n = \Delta p$ із темпом G .

Темп об'ємної рекомбінації дірок $R_p = \Delta p / \tau_p$. Дифузійний потік дірок

$J_p / e = -D_p \frac{d\Delta p}{dx}$. Стаціонарне рівняння неперервності для цього випадку:

$$D_p \frac{d^2 \Delta p}{dx^2} - \frac{\Delta p}{\tau_p} + G = 0. \quad (5.32)$$

На поверхні ($x = 0$) дифузійний потік дірок приймаємо рівним R_s (5.31):

$$D_p \frac{d\Delta p}{dx} = S \Delta p. \quad (5.33)$$

В об'ємі, коли x значно більше дифузійної довжини дірок L_p , генерація нерівноважних дірок рівномірна ($d\Delta p/dx = 0$) і із рівняння (5.32) знаходимо:

$$\Delta p \Big|_{x \gg L_p} = G \tau_p. \quad (5.34)$$

Загальний розв'язок (5.32):

$$\Delta p(x) = C \exp(-x/L_p) + G \tau_p. \quad (5.35)$$

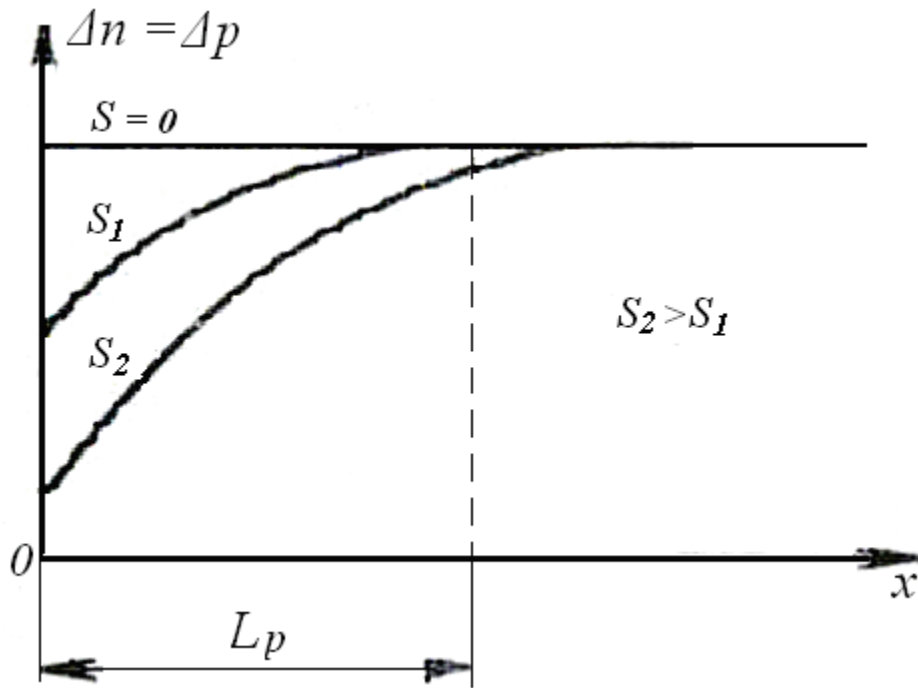


Рис.5.5. Вплив швидкості поверхневої рекомбінації на розподіл не рівноважних носіїв заряду у поверхневому прошарку

Константу C знаходимо із граничної умови (5.33) для поверхні:

$$C = -\frac{GS\tau_p^2}{L_p + S\tau_p}. \quad (5.36)$$

Із урахуванням (5.36) і дифузійної швидкості $v_D = L_p/\tau_p$, для $\Delta p(x)$ маємо:

$$\Delta p(x) = G\tau_p \left(1 - \frac{S/v_D}{1 + S/v_D} \exp(-x/L_p) \right). \quad (5.37)$$

Висновки: – якщо швидкість поверхневої рекомбінації значно менше дифузійної ($S \ll v_D$), зміна $\Delta p(x)$ у межах $0 \leq x \leq L_D$ буде незначною і $\Delta p \approx G\tau_p$;

–у протилежному випадку ($S \gg v_D$) маємо сильну експонентну

залежність $\Delta p(x) \approx G\tau_p(1 - \exp(-x/L_p))$;

–концентрація дірок на поверхні

$$\Delta p(0) = \Delta p_s = G\tau_p \frac{L_p}{L_p + S\tau_p} \quad (5.38)$$

має обернено пропорційну залежність від S ;

–поверхневий час життя

$$\tau_s = \Delta p_s / G = \tau_p \frac{L_p}{L_p + S\tau_p} \quad (5.39)$$

також має обернено пропорційну залежність від S .

На рис.5.5 показаний розподіл нерівноважних носіїв заряду, який ілюструє вище наведені закономірності.

Особливого значення поверхнева рекомбінація набуває у тонких зразках, товщиною у декілька дифузійних довжин. У таких випадках доцільно вводити поняття ефективного часу життя: $\tau_{eff}^{-1} = \tau_v^{-1} + \tau_s^{-1}$, де τ_v – об’ємний час життя.

У числовому вимірі швидкість поверхневої рекомбінації має широкі межі. Для прикладу, у германії спостерігались значення S від 50 до 10^5 см / с.

5.5. Міжзонна рекомбінація

Міжзонна рекомбінація, на відміну від рекомбінації через захват, реалізується через електронні переходи безпосередньо із зони провідності у валентну зону. Такій перехід супроводжується втратою енергії кожним надлишковим носієм заряду $\geq E_g$ і відповідною зміною хвильового вектора $\Delta \bar{k}$. Закони збереження енергії і хвильового вектора при міжзонній рекомбінації можна записати таким чином:

$$E' - E \geq E_g, \bar{k}' - \bar{k} = \Delta \bar{k}, \quad (5.40)$$

де E', \bar{k}' – енергія і хвильовий вектор після рекомбінації, E, \bar{k} – до рекомбінації.

У залежності від того, на що витрачається енергія електронно – діркової пари при міжзонній рекомбінації розрізняють такі випадки:

– енергія трансформується у енергію фотонів – **пряма міжзонна фотонна рекомбінація**

$$h\nu \geq E_g, \Delta\bar{k} = \bar{k}_{\min} - \bar{k}_{\max} + \bar{k}_{\text{фот}} \approx 0; \quad (5.41)$$

– рекомбінація відбувається при одночасній взаємодії електрона, фотона і фонона (віртуальний перехід) – **непряма міжзонна фотонна рекомбінація**

$$h\nu \geq E_g \pm \hbar\omega_q, \Delta\bar{k} = \bar{k}_{\text{фот}} + \bar{q} \approx \bar{q}; \quad (5.42)$$

– енергія передається тепловим коливанням атомів – **фононна рекомбінація**

$$\hbar\omega_q \geq E_g, \Delta\bar{k} = \bar{q}, \quad (5.43)$$

де $\hbar\omega_q$ – енергія фонона із хвильовим вектором \bar{q} ;

– рекомбінаційна енергія електронно – діркової пари передається третьому носію заряду (електрону або дірці) – **міжзонна ударна рекомбінація (Оже–рекомбінація)**.

Пряма та непряма міжзонна рекомбінація. Імпульс фотона, який народжується при міжзонній рекомбінації, приблизно на три порядки менше імпульсу електрона при однакових енергіях (5.41). Незмінність імпульсу електрона при прямій міжзонній фотонній рекомбінації означає, що відбуваються вертикальні переходи між екстремумами зони провідності і

валентної зони, яким відповідають однакові значення хвильових векторів $\bar{k}_{\min} = \bar{k}_{\max}$. Такі переходи є основними для прямозонних напівпровідників, до яких належать практично усі сполуки A^3B^5 а також деякі халькогеніди A^4B^6 (PbS, PbSe, PbTe).

Непряма міжзонна рекомбінація (5.42) спостерігається у особливо чистих по рекомбінаційним центрам кремнії і германії. У цих непрямозонних напівпровідниках $\bar{k}_{\min} \neq \bar{k}_{\max}$, тому необхідна участь фононів для зміщення у імпульсному просторі. Імпульс фонона, якій компенсує різницю між \bar{k}_{\min} і \bar{k}_{\max} , електрон може отримати як при поглинанні, так і при народженні фонона. Цим пояснюється два можливих значення енергії фотона (знак \pm у законі збереження енергії (5.42)). Імовірність віртуальних рекомбінаційних переходів у зв'язку із ускладненими законами збереження енергії і імпульсу на один – два порядки менша, ніж для прямих, вертикальних переходів.

Таблиця 5.1. Час життя для міжзонної фотонної рекомбінації (300 K)

Матеріал	Ge	Si	InSb	InAs	GaAs	PbSe	PbS
E_g , eV	0,68	1,07	0,18	0,31	1,35	0,29	0,41
τ , мкс	150	2500	0,12	0,24	0,37	0,25	0,21

Розрахунки часу життя при міжзонній фотонній рекомбінації на основі теорії Ейнштейна для рівноважного випромінювання дають значення, які приблизно на порядок більші експериментально визначених для вузькозонних напівпровідникових сполук ($E_g < 0,5$ eV). Для широкозонних напівпровідників ($\Delta E_0 > 1$ eV) розрахункові значення часу життя на три – чотири порядки більші експериментальних значень. Така велика різниця між розрахунком і експериментом пояснюється тим, що не враховується суттєвий внесок фононної рекомбінації та іонної компоненти валентних зв'язків у бінарних напівпровідникових сполуках. Для чистих непрямозонних ковалентних напівпровідників (Ge, Si) час життя непрямої фотонної рекомбінації на три – чотири порядки більший, ніж для прямозонних при співрозмірних значеннях ширини забороненої зони.

Ударна міжзонна рекомбінація. При одночасному зіткненні трьох вільних носіїв заряду – два електрони + дірка або дві дірки + електрон – рекомбінаційна

енергія електронно – діркової пари передається третьому носію заряду, тобто електрону або дірці. Носії заряду, які прийняли рекомбінаційну енергію, залишаються у межах відповідної зони, але енергетично не рівноважними, „гарячими”. Через механізми фононного та іонного розсіяння їх енергія передається тепловим коливанням атомів.

Висновок: ударна міжзонна рекомбінація проходить у два етапи. Перший етап – утворення „гарячих” носіїв заряду, які приймають на себе рекомбінаційну енергію електронно – діркової пари, але залишаються вільними у межах відповідної зони. Другий етап – розсіяння надлишкової енергії на фононах і іонах.

Імовірність одночасного зіткнення трьох носіїв заряду пропорційна їх концентрації. Для зіткнення електронно – діркової пари із вільним електроном вона пропорційна $n^2 p$, із вільною діркою – $p^2 n$ ($n = n_0 + \Delta n$, $p = p_0 + \Delta p$).

Розглянемо малий рівень інжекції $\Delta n, \Delta p \ll n_0, p_0$. Для власного напівпровідника $n_0 = p_0 = n_i$, $n_i^2 = N_C N_V \exp(-E_g / k_B T)$. Власна концентрація n_i – найменша можлива у певному напівпровіднику при температурі T . Імовірність зіткнення у цьому випадку пропорційна n_i^2 , а час життя $\tau_i^{ожс} \sim 1/n_i^2$.

Висновок: час життя для при міжзонної ударної рекомбінації у власному напівпровіднику обернено пропорційний квадрату власної концентрації і має максимальне значення. Підвищення температури буде зменшувати час життя. У вузькозонних напівпровідниках імовірність ударної рекомбінації вища, ніж у широкозонних за інших рівних умов.

Для домішкового електронного напівпровідника ($n_0 \gg p_0, n_i$) час життя $\tau_n^{ожс} \sim 1/n_0^2$, для діркового ($p_0 \gg n_0, n_i$) – $\tau_p^{ожс} \sim 1/p_0^2$.

У випадку високого рівня інжекції ($\Delta n, \Delta p \gg n_0, p_0, n_i$) імовірність ударної рекомбінації пропорційна Δn^2 , час життя обернено пропорційний Δn^2 або Δp^2 .

Висновок: у домішкових напівпровідниках спостерігається сильна залежність часу життя для міжзонної ударної рекомбінації від концентрації теплових рівноважних носіїв заряду, а для високого рівня інжекції – від рівня інжекції.

Кількість напівпровідників, у яких міжзонна ударна рекомбінація може конкурувати із іншими механізмами рекомбінації досить обмежена. Це, в основному, вузькозонні напівпровідники при високих температурах (наприклад, власний InSb, $E_g = 0,18$ eV).

Резюме по р. 5

1. Рекомбінація – процес відновлення концентраційної рівноваги, порушеної нетепловими механізмами генерації. У кінцевому підсумку, рекомбінація зводиться до відновлення стану термодинамічної концентраційної рівноваги між зонами.

2. Кількісні характеристики рекомбінації: темп рекомбінації, середній час життя нерівноважних носіїв заряду.

3. Рекомбінація через захват характерна для ковалентних напівпровідників за наявності центрів захвату (рекомбінаційні центри).

4. Час життя складається із часу очікування захвату електронів центрами і часу очікування захвату дірок цими ж центрами.

5. Центри захвату створюються структурними дефектами та домішковими атомами, які не заміщають основні, а розташовуються між ними. Найбільш активні у процесах рекомбінації домішки елементів Cu, Ag, Au, Pt, Fe, Ni і ін. Енергетичне положення центрів захвату – поблизу середини забороненої зони.

6. Час життя неосновних носіїв заряду у вироджених напівпровідниках має мінімально можливе значення, яке визначається лише концентрацією центрів і їх перерізом захвату. Від температури час життя у вироджених напівпровідниках практично не залежить.

7. У невироджених напівпровідниках за низьких температур ($T < 250 \text{ K}$) та малого рівня інжекції термоіонізація центрів захвату незначна і час життя неосновних носіїв заряду має мінімально можливе значення і від температури не залежить.

8. На підвищених температурах ($T > 250 \text{ K}$) стає помітною термоіонізація центрів захвату, що уповільнює рекомбінацію у невироджених напівпровідниках. Час життя експонентно зростає зі збільшенням температури.

9. У власних напівпровідниках і за високого рівня генерації час життя має максимальне значення, тому що є сумою часу очікування захвату електрона і часу очікування захвату дірки рекомбінаційним центром.

10. Енергія електронів і дірок, яка виділяється у процесі рекомбінації через захват, трансформується у теплові коливання атомів – «фононна рекомбінація».

11. Для напівпровідникових сполук $A^{III}B^V$ із іонно – ковалентним зв'язком і широкою забороненою зоною ($> 1 \text{ eV}$) домінуючою є міжзонна рекомбінація із випромінюванням фотонів («фотонна рекомбінація»). Енергія фотонів відповідає ширині забороненої зони.

12. Поверхня напівпровідника зазвичай збагачена центрами рекомбінації і розсіюючими дефектами відносно його об'єму. Загальним наслідком цього є те, що поверхневий час життя і поверхнева мобільність менші відповідних значень цих параметрів у об'ємі.

13. Кількісна характеристика поверхневої рекомбінації – швидкість поверхневої рекомбінації – частота рекомбінаційних переходів на 1 м^2 поверхні напівпровідника, нормована на одиничну концентрацію нерівноважних носіїв заряду.

14. Штучно створений поверхневий потенціал впливає на рекомбінаційні потоки електронів або дірок між об'ємом і поверхнею. За певних значень потенціалу надходження електронів або дірок до поверхні можна зупинити, внаслідок чого швидкість поверхневої рекомбінації збігає до нуля (ефект поля).

15. Поверхнева рекомбінація суттєво впливає на розподіл нерівноважних носіїв заряду у поверхневому прошарку. Якщо швидкість поверхневої рекомбінації набагато більша дифузійної швидкості, концентрація у поверхневому прошарку зменшується при наближенні до поверхні по експонентному закону.

16. Окрім фононної та фотонної міжзонної рекомбінації, за певних умов можлива ударна міжзонна рекомбінація (Оже – рекомбінація).

17. Процеси, які супроводжують ударну міжзонну рекомбінацію, до певної міри схожі на ударну іонізацію, розглянуту у р. 4.13, але відбуваються у зворотньому порядку. Одночасне зіткнення трьох вільних носіїв заряду – два електрони + дірка або дві дірки + електрон, супроводжується передачею рекомбінаційної енергії електронно – діркової пари третьому носію заряду, тобто електрону або дірці, які залишаються вільними у межах відповідних зон.

18. Час життя для цього механізму рекомбінації у власному напівпровіднику обернено пропорційний квадрату власної концентрації, у домішковому – квадрату концентрації основних носіїв заряду, за високого рівня інжекції – квадрату концентрації неосновних нерівноважних носіїв заряду.

6. Рівняння неперервності

6.1. Загальні положення

Виділяємо певний об'єм $d^3\vec{r} = dx \cdot dy \cdot dz$ у межах однорідного електронного напівпровідника та врахуємо усі можливі процеси, які впливають на концентрацію носіїв заряду $n(\vec{r}, t)$ у цьому об'ємі. Зміни концентрації можливі як у часі, так і у просторі. Серед таких процесів найважливішими є потік електронів, що проходить через об'єм, та процеси генерації і рекомбінації. На рис.6.1. у одновимірному зображенні показана координатно – енергетична

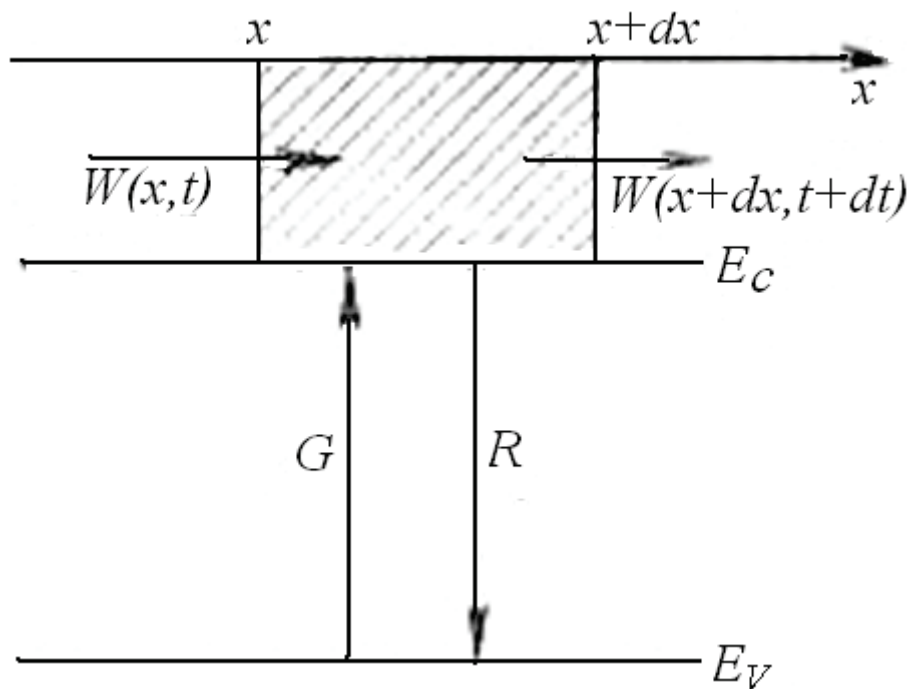


Рис.6.1. Зміна щільності потоку носіїв заряду із урахуванням генерації та рекомбінації

діаграма електронного напівпровідника із урахуванням вищевказаних процесів. Зміна концентрації електронів при наявності потоку визначається як різниця щільностей потоків електронів, які входять у об'єм $W(x, t)$ і виходять із об'єму $W(x + dx, t + dt)$ за час dt :

$$W(x,t)dt - W(x+dx,t)dt = -\frac{\partial W}{\partial x} dxdt. \quad (6.1)$$

Вважаємо, що генерація і рекомбінація рівномірні на відрізьку dx і визначаємо зміну концентрації електронів:

$$\frac{\partial n}{\partial t} dxdt = -\frac{\partial W}{\partial x} dxdt + Gdxdt - Rdxdt. \quad (6.2)$$

Узагальнимо (6.2) на випадок об'ємного розподілу електронів $n(x,y,z,t)$. Із повної концентрації виділяємо нерівноважну $\Delta n = n - n_0$ і враховуємо (5.4а), щільність потоку $\bar{W}(x,y,z)$ визначимо через електронний струм:

$$\bar{W}(x,y,z) = -\frac{\bar{J}_n(x,y,z)}{e}. \quad (6.3)$$

Рівняння (6.2) набуває такого вигляду:

$$\frac{\partial n}{\partial t} = \frac{1}{e} \left(\frac{\partial j_x}{\partial x} + \frac{\partial j_y}{\partial y} + \frac{\partial j_z}{\partial z} \right) + G - \frac{n - n_0}{\tau_n} = \frac{1}{e} \operatorname{div} \bar{J}_n + G - \frac{n - n_0}{\tau_n}. \quad (6.4)$$

Аналогічне рівняння маємо для дірок:

$$\frac{\partial p}{\partial t} = -\frac{1}{e} \operatorname{div} \bar{J}_p + G - \frac{p - p_0}{\tau_p}. \quad (6.5)$$

Рівняння (6.4) і (6.5) називаються **рівняннями неперервності** для електронів і дірок.

Розглянемо рівняннями неперервності у стані стаціонарної рівноваги $G = \partial n / \partial t = \partial p / \partial t = 0$ для одновимірного запису:

$$\frac{1}{e} \frac{\partial j_{xn}}{\partial x} = G - \frac{n - n_0}{\tau_n}; \quad (6.6)$$

$$\frac{1}{e} \frac{\partial j_{xp}}{\partial x} = G - \frac{p - p_0}{\tau_p}. \quad (6.7)$$

Висновок: фізичний зміст стаціонарних рівнянь неперервності (6.6), (6.7) полягає у тому, що потік носіїв заряду, який виходить із виділеного об'єму на межі $x + dx$, дорівнює вхідному потоку на межі $x = 0$, із урахуванням носіїв заряду, які генеруються і рекомбінують у цьому об'ємі. Іншими словами,

нерівноважна концентрація підтримується різницею струмів на вході і на виході із об'єму.

6.2. Нерівноважний заряд основних носіїв

Радіус екранування Дебая. Розглянемо напівпровідник, у якому створений нерівноважний об'ємний заряд основних носіїв $\rho(x,t) = en(x,t)$. Об'ємний заряд породжує електричне поле $\bar{\mathcal{E}}$, яке пов'язане із $\rho(x,t)$ рівнянням Пуассона:

$$\operatorname{div} \bar{\mathcal{E}} = \frac{\partial \mathcal{E}_x(x,t)}{\partial x} = \frac{\rho(x,t)}{\varepsilon \varepsilon_0}, \quad (6.8)$$

де ε - відносна діелектрична проникність, ε_0 - електрична постійна вакууму.

Електричне поле $\mathcal{E}_x(x,t)$ викликає струм $j_x = \sigma \mathcal{E}_x(x,t)$. Підставимо j_x у (6.4) і об'єднаємо із (6.8):

$$\frac{\partial \rho(x,t)}{\partial t} = -\frac{\sigma}{\varepsilon \varepsilon_0} \rho(x,t). \quad (6.9)$$

Загальний розв'язок (6.9) має вигляд $\rho(x,t) = \rho(x) \cdot \rho(t)$. Введемо позначення, яке має розмірність часу:

$$\tau_M = \frac{\varepsilon \varepsilon_0}{\sigma}, \quad (6.10)$$

і знайдемо розв'язок для $\rho(t)$ за таких початкових умов: $t = 0 \rightarrow \rho(t) = \rho(0)$; $t \rightarrow \infty \rightarrow \rho(t) \rightarrow 0$. Просте інтегрування (6.9) із урахуванням (6.10) дає:

$$\rho(t) = \rho(0) \exp(-t / \tau_M). \quad (6.11)$$

Висновок: фізичний зміст (6.11) полягає у тому, що середній час існування нерівноважного заряду основних носіїв $\tau_M = \varepsilon \varepsilon_0 / \sigma$ визначається провідністю, яка компенсує нерівноважний заряд основних носіїв, і він отримав назву **час релаксації Максвелла**. Якщо створити нерівноважну концентрацію

неосновних носіїв заряду, то їх об'ємний заряд буде скомпенсований основними носіями заряду протягом часу релаксації Максвелла τ_M , а нерівноважна концентрація неосновних носіїв заряду буде існувати протягом часу життя τ_n або τ_p .

Зробимо оцінку τ_M для невинроджених напівпровідників, прийнявши $\varepsilon = 12$, $\sigma = 5 \text{ Ом}^{-1} \cdot \text{м}^{-1}$, $\varepsilon_0 = 8,8 \cdot 10^{-12} \text{ Ф/м}$, і отримаємо $\tau_M \approx 2 \cdot 10^{-11} \text{ с}$. Оцінка для винроджених напівпровідників ($\sigma = 500 \text{ Ом}^{-1} \cdot \text{м}^{-1}$) дає $\tau_M \approx 2 \cdot 10^{-13} \text{ с}$.

Порівнявши отримане значення τ_M із часом фононної релаксації τ_ϕ (4.43а), дійдемо висновку: час релаксації Максвелла у невинроджених напівпровідниках на 1 – 2 порядки менше часу фононної релаксації, у винроджених вони співрозмірні.

Для координатної частини загального розв'язку $\rho(x)$ необхідно розглянути дифузійно – дрейфові процеси у стаціонарній рівновазі $\partial \rho(t) / \partial t = 0$.

Запишемо повний струм у напрямку x :

$$j_x = j_{x\partial p} + j_{x\partial u\phi} = en(x)\mu_n \mathcal{E}_x + eD_n \frac{dn(x)}{dx} = 0. \quad (6.12)$$

Введемо об'ємний заряд основних носіїв $\rho(x, t) = en(x, t)$ у (6.12) і визначимо \mathcal{E}_x із урахуванням співвідношення Ейнштейна:

$$E_x = -\frac{D_n}{\mu_n} \frac{1}{\rho(x)} \frac{d\rho(x)}{dx} = -\frac{k_B T}{e} \frac{1}{\rho(x)} \frac{d\rho(x)}{dx}. \quad (6.13)$$

Розглянемо незначне відхилення від теплової рівноваги:

$$\rho(x) = \rho_0 + \Delta\rho(x); \quad \Delta\rho(x) \ll \rho_0 \quad (6.14)$$

і врахуємо (6.14) у (6.13):

$$\mathcal{E}_x = -\frac{k_B T}{e} \frac{1}{\rho_0} \frac{d\rho(x)}{dx}. \quad (6.15)$$

Знайдемо похідну $d\mathcal{E}_x / dx$ і візьмемо її значення із рівняння Пуассона (6.8):

$$\frac{d^2\rho(x)}{dx^2} + \frac{e\rho_0}{\varepsilon\varepsilon_0 k_B T} \rho(x) = 0. \quad (6.16)$$

Коефіцієнт при $\rho(x)$ у формулі (6.16) має розмірність квадрата оберненої довжини, позначимо цю довжину L_e :

$$L_e^2 = \frac{\varepsilon\varepsilon_0 k_B T}{e\rho_0}. \quad (6.17)$$

Підставимо (6.17) у (6.16):

$$\frac{d^2\rho(x)}{dx^2} + \frac{1}{L_e^2} \rho(x) = 0. \quad (6.18)$$

Визначимо початкові умови для розв'язку (6.18):
 $x = 0 \rightarrow \rho(x) = \rho(0)$; $x \rightarrow \infty \rightarrow \rho(x) \rightarrow 0$. Після подвійного інтегрування отримаємо:

$$\rho(x) = \rho(0) \exp(-x/L_e). \quad (6.19)$$

Фізичний зміст відстані L_e – відстань, на якій нерівноважний об'ємний заряд основних носіїв заряду зменшується в e - раз або в 2,7 рази. Визначена таким чином відстань отримала назву **радіус екранування Дебая**.

У формулі (6.17) замінімо ρ_0 на en_0 , домножимо і поділимо вираз на μ_n , врахуємо (6.10) та співвідношення Ейнштейна (3.61) і отримаємо для L_e :

$$L_e = \sqrt{D_n \tau_M}. \quad (6.20)$$

Для числової оцінки L_e запишемо (6.20) у дещо іншому вигляді:

$$L_e = \sqrt{\frac{\varepsilon\varepsilon_0}{en_0} k_B T}, \quad (6.21)$$

де $e = 1,6 \cdot 10^{-19}$ Кул., $k_B T = 2,6 \cdot 10^{-2}$ В, $T = 300$ К, $\varepsilon\varepsilon_0 = 10^{-10}$ Ф/м.

Обрахуємо усі константи і для L_e отримаємо розрахункову формулу:

$$L_e \approx 4 \cdot 10^3 \sqrt{1/n_0}. \quad (6.22)$$

Для слабо легованого напівпровідника беремо $n_0 = 10^{20} \text{ м}^{-3}$ і для радіуса екранування Дебая маємо $L_e = 4 \cdot 10^{-7} \text{ м} = 0,4 \text{ мкм}$. У сильно легованого ($n_0 = 10^{24} \text{ м}^{-3}$) радіус екранування на два порядки менший – $L_e = 4 \cdot 10^{-9} \text{ м} = 4 \text{ нм}$.

Висновок: нерівноважний об'ємний заряд основних носіїв, створений на поверхні напівпровідника, зосереджується у тонкому прошарку (скін – прошарку) і глибоко у об'єм не проникає. Виняток становлять напівпровідникові плівки, товщина яких співрозмірна з радіусом екранування. У металів $n_0 \approx 10^{28} \text{ м}^{-3}$ і скін – прошарок має порядок декількох ангстрем (1 ангстрем = 10^{-10} м).

6.3. Дифузія і дрейф нерівноважних неосновних носіїв заряду

Вибираємо електронний напівпровідник ($n_0 \gg p_0$) і малий рівень інжекції нерівноважних дірок на зовнішній поверхні напівпровідника ($\Delta p \ll n_0$). Повна концентрація дірок $p = p_0 + \Delta p$ на поверхні більша, ніж у об'ємі ($p > p_0$), і завдяки градієнту $dp/dx \neq 0$ виникає дифузія дірок від поверхні у об'єм. Позитивний об'ємний заряд нерівноважних дірок порушує електронейтральність напівпровідника. У відповідь на це виникає перерозподіл основних носіїв заряду – електронів, які втягуються у зону дифузії і нейтралізують позитивний об'ємний заряд дірок за час релаксації Максвелла.

Висновок: дифузія нерівноважних неосновних носіїв заряду супроводжується втягуванням у цей процес такої ж кількості основних носіїв для нейтралізації об'ємного заряду неосновних. Таким чином, відбувається дифузія квазинейтральних електронно – діркових пар і одночасна їх рекомбінація в міру просування вглиб напівпровідника.

Якщо у напівпровіднику створити однорідне електричне поле, процес дифузії буде супроводжуватися ще й дрейфом. Розглянемо ці процеси за

допомогою стаціонарного рівняння неперервності (6.7). Повний дифузійно – дрейфовий дірковий струм $j_{xp} = eD_p \frac{d\Delta p(x)}{dx} - e\Delta p(x)\mu_p \mathcal{E}_x$ підставляємо у (6.7):

$$D_p \frac{d^2 \Delta p}{dx^2} - \mu_p \mathcal{E}_x \frac{d\Delta p}{dx} - \frac{\Delta p}{\tau_p} = 0. \quad (6.23)$$

Поділимо (6.23) на D_p і введемо **дифузійну довжину** L_p і **довжину дрейфу** L_E :

$$L_p^2 = D_p \tau_p, \quad (6.24)$$

$$L_E = \mu_p E_x \tau_p. \quad (6.25)$$

Фізичний зміст дифузійної довжини L_p – середній шлях нерівноважних дірок у електронному напівпровіднику, пройдений за час життя τ_p у процесі дифузії. Фізичний зміст довжини дрейфу L_E – середній шлях, пройдений дірками за час життя τ_p зі швидкістю дрейфу $\mu_p \mathcal{E}_x$.

Рівняння (6.23) у нових позначеннях набуває такого вигляду:

$$\frac{d^2 \Delta p}{dx^2} - \frac{L_E}{L_p^2} \frac{d\Delta p}{dx} - \frac{\Delta p}{L_p^2} = 0. \quad (6.26)$$

При розв'язку (6.26) слід врахувати, що напрямки електричного поля і напрямку дрейфу можуть змінюватися на протилежні. Загальний вигляд розв'язку лінійного однорідного рівняння:

$$\Delta p(x) = A \exp(\alpha x). \quad (6.27)$$

Після підстановки (6.27) у (6.26) і скорочення на $A \exp(\alpha x)$ отримаємо характеристичне рівняння

$$\alpha^2 - \frac{L_E}{L_p^2} \alpha - \frac{1}{L_p^2} = 0, \quad (6.28)$$

розв'язок якого має вигляд:

$$\alpha = \frac{1}{L_p} (L_E/2L_p \pm \sqrt{1 + L_E^2/4L_p^2}). \quad (6.29)$$

Параметр α має розмірність оберненої довжини. Введемо *довжину зтягування* $L = -1/\alpha$, яка враховує одночасну дію двох процесів – дифузії і дрейфу.

Врахуємо довжину зтягування у (6.27):

$$\Delta p(x) = A \exp(-x/L). \quad (6.30)$$

Коефіцієнт A визначаємо при $x = 0$: $A = (p - p_0)|_{x=0} = \Delta p(0)$. Таким чином, розв'язок (6.26) має такий загальний вигляд:

$$\Delta p(x) = \Delta p(0) \exp(-x/L). \quad (6.31)$$

Щоб розв'язок (6.26) мав кінцеве значення при $x \rightarrow \infty$, $\Delta p(x) \rightarrow 0$, L повинно бути > 0 за будь-яких умов. Зберігаємо знак «-» перед коренем у (6.29):

$$\frac{1}{L} = \frac{1}{L_p} (\sqrt{1 + L_\varepsilon^2 / 4L_p^2} - L_\varepsilon / 2L_p) \quad (6.32)$$

і аналізуємо (6.32) при $\mathcal{E}_x = 0$ і, відповідно до (6.25), $L_\varepsilon = 0$. У цьому випадку $L = L_p$ і (6.31) має вигляд:

$$\Delta p(x) = \Delta p(0) \exp(-x/L_p). \quad (6.33)$$

Фізичний зміст дифузійної довжини (6.24) $L_p = \sqrt{D_p \tau_p}$ визначимо при $x = L_p$:

$$\Delta p(x = L_p) = \frac{\Delta p(0)}{e} \approx \frac{\Delta p(0)}{2,7} \approx 0,37 \Delta p(0). \quad (6.34)$$

Висновок: дифузійна довжина – відстань, на якій концентрація дифундуючих нерівноважних неосновних носіїв заряду зменшується у e – раз або у 2,7 рази. У скороченому визначенні: дифузійна довжина – середня відстань пройдена шляхом дифузії за час життя.

Якщо розглянути випадок $\Delta p(0) < 0$, ($p < p_0$), отримаємо дзеркальне відображення дифузійної експоненти відносно горизонтальної вісі p_0 (рис.6.2, $\mathcal{E}_x = 0$ - дифузія).

При аналізі дифузійно – дрейфових процесів за участі неосновних носіїв заряду використовують поняття *дифузійної швидкості* $v_{\text{диф}}$ і *швидкості дрейфу* $v_{\text{др}}$:

$$v_{\text{диф}} = L_p / \tau_p, \quad v_{\text{др}} = \mu_p \mathcal{E}_x. \quad (6.35)$$

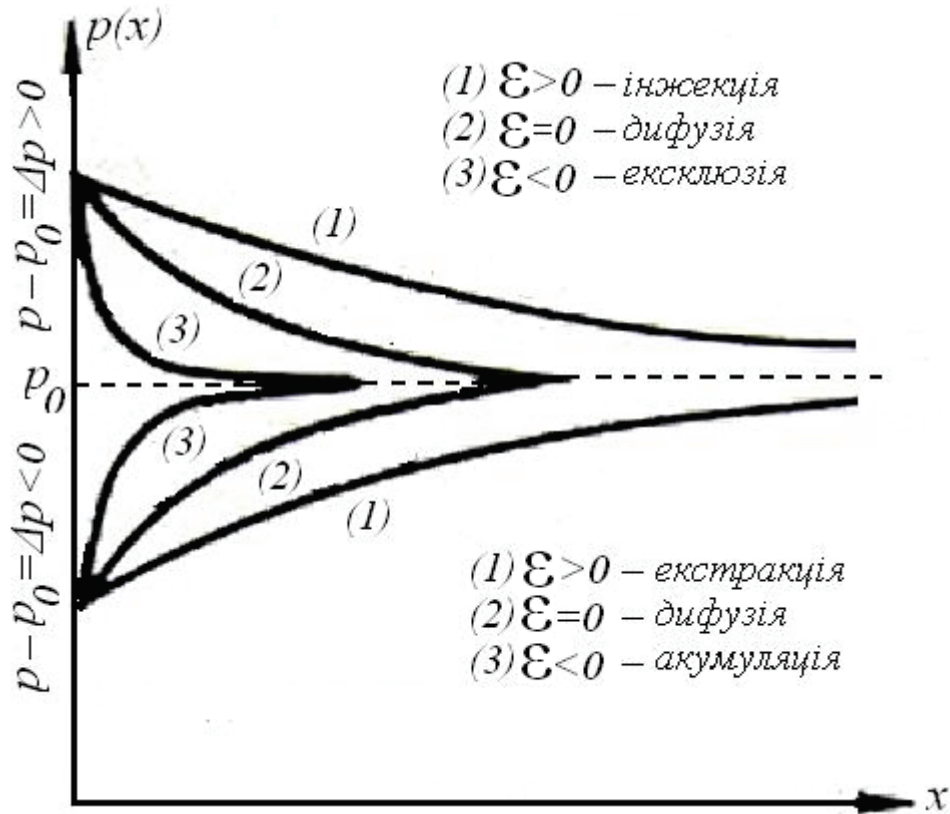


Рис.6.2. Дифузійно – дрейфові монополярні процеси нерівноважних носіїв заряду

Висновок: *дифузійна швидкість* – швидкість, з якою не рівноважні носії заряду проходять дифузійну довжину за час життя.

У випадку слабого електричного поля:

$$L_{\mathcal{E}} \ll 2L_p \quad (6.36)$$

формулу (6.32) можна спростити і для довжини затягування L отримаємо:

$$L \approx L_p. \quad (6.37)$$

Висновок: у слабких електричних полях, які відповідають критерію (6.36), довжина затягування практично співпадає із дифузійною довжиною.

Для сильного електричного поля:

$$L_{\varepsilon} \gg 2L_p; v_{dr} \gg v_{диф} . \quad (6.38)$$

необхідно врахувати не лише модуль \mathcal{E} , але і його напрямок. Розглянемо випадок $\mathcal{E} > 0, L_{\varepsilon} > 0$, тобто дрейф дірок співпадає із напрямком вісі x . Врахуємо (6.38) у (6.32) і після нескладних перетворень визначимо L :

$$\frac{1}{L} = \frac{1}{L_p} \left(\frac{L_{\varepsilon}}{2L_p} \sqrt{1 + \frac{4L_p^2}{L_{\varepsilon}^2}} - \frac{L_{\varepsilon}}{2L_p} \right) = \frac{L_{\varepsilon}}{2L_p^2} \left(1 + \frac{2L_p^2}{L_{\varepsilon}^2} + \dots - 1 \right) \approx \frac{1}{L_{\varepsilon}} . \quad (6.39)$$

Висновок: у сильних електричних полях, визначених за критерієм (6.38) і $L_{\varepsilon} > 0$, $L \approx L_{\varepsilon} = \mu_p \mathcal{E}_x \tau_p = v_{dr} \tau_p$ – довжина затягування дорівнює довжині дрейфу. Електричне поле збагачує об'єм неосновними носіями заряду і це явище отримало назву **інжекція** (рис.6.2. $p - p_0 > 0, \mathcal{E} > 0$ – інжекція).

Електричне поле сильне (6.38), але напрямок $\mathcal{E} < 0, L_{\varepsilon} < 0, |L_{\varepsilon}| > 0$ відносно вісі x протилежний. Із (6.32) отримаємо:

$$\frac{1}{L} \approx \frac{|L_{\varepsilon}|}{L_p^2}; L \approx \frac{L_p^2}{|L_{\varepsilon}|} . \quad (6.40)$$

Висновок: сильне електричне поле, яке створює дрейфовий потік дірок протилежний дифузійному, зменшує довжину затягування. Зі збільшенням напруженості поля по модулю \mathcal{E} довжина затягування $L \rightarrow 0$ і це явище отримало назву **ексклюзія**. (рис.6.2. $p - p_0 > 0, \mathcal{E} < 0$ – ексклюзія).

Аналіз дифузійно – дрейфових процесів у випадку $\Delta p(0) < 0, (p < p_0)$ аналогічний попередньому і визначає ще два можливих механізми розподілу не рівноважних дірок у електронному напівпровіднику окрім дифузії.

У випадку $\Delta p(0) < 0$ і сильного електричного поля $\mathcal{E} > 0$ довжина зтягування $L \approx L_e$, але наслідком, на відміну від інжекції, буде збіднення об'єму на дірки.

Висновок: якщо на поверхні $p - p_0 < 0$ і діє сильне поле $\mathcal{E} > 0$, довжина зтягування буде визначатися довжиною дрейфу, але об'єм буде збіднюватися на неосновні носії заряду – дірки. Це явище отримало назву **екстракція** (рис.6.2. $p - p_0 < 0, \mathcal{E} > 0$ – екстракція).

При інверсії напрямку поля $\mathcal{E} > 0$ довжина зтягування буде визначатися (6.40) і зменшуватися у міру зростання напруженості поля $|\mathcal{E}|$.

Висновок: якщо на поверхні $p - p_0 < 0$ і діє сильне поле $\mathcal{E} > 0$, довжина зтягування буде обернено пропорційною довжині дрейфу, але об'єм буде збагачуватися на неосновні носії заряду – дірки. Це явище отримало назву **аккумуляція** (рис.6.2. $p - p_0 < 0, \mathcal{E} > 0$ – аккумуляція).

У випадку діркового напівпровідника і не рівноважної концентрації електронів на поверхні $\Delta n = n - n_0$ отримаємо дифузію, інжекцію, ексклюзію, екстракцію і аккумуляцію неосновних носіїв заряду – електронів, змінивши відповідним чином напрямки електричного поля.

6.4. Біполярні дифузія і дрейф нерівноважних носіїв заряду

Розглянемо дифузійно – дрейфові процеси, які супроводжують біполярну поверхневу генерацію нерівноважних електронно – діркових пар $\Delta n = \Delta p$. Розподіл нерівноважних носіїв – градієнтний, рівноважних – рівномірний, рівень інжекції малий $\Delta n = \Delta p \ll n_0, p_0$.

Повні концентрації $n(x) = n_0 + \Delta n(x)$, $p(x) = p_0 + \Delta p(x)$.

Градiєнт концентрації нерівноважних носіїв заряду зумовлює дифузійні потоки від поверхні у об'єм напівпровідника. Якщо дифузійні швидкості електронів і дірок будуть однакові $v_{diffn} = v_{diffp}$, електронно – діркові пари не будуть розділятися у процесі дифузії і електронеутральність напівпровідника

не порушується. На практиці такий випадок не зустрічається, дифузійні швидкості не співпадають $v_{\text{дифн}} > v_{\text{дифр}}$ і електрони випереджають дірки. Просторове розділення електронно – діркових пар порушує нейтральність, виникає різниця потенціалів і відповідне їй статичне електричне поле $\mathcal{E}_{\text{ст}}$. Стаціонарна рівновага між дифузією і дрейфом встановлюється за рахунок того, що статичне поле буде гальмувати електрони і прискорювати дірки аж поки не зрівняються їх швидкості. Виникає дифузійно – дрейфовий потік квазінейтральних пар і їх рекомбінація у процесі руху.

Визначимо $\mathcal{E}_{\text{ст}}$ із умови рівності нулю повного струму:

$$J = J_{\text{xn}} + J_{\text{xp}} = e(n\mu_n + p\mu_p)\mathcal{E}_{\text{ст}} + e(-D_n \frac{dn}{dx} + D_p \frac{dp}{dx}) = 0;$$

$$\mathcal{E}_{\text{ст}} = \frac{D_p \frac{dp}{dx} - D_n \frac{dn}{dx}}{p\mu_p + n\mu_n}. \quad (6.41)$$

Врахуємо, що $dn_0/dx = dp_0/dx = 0$ і малий рівень інжекції:

$$\mathcal{E}_{\text{ст}} \approx \frac{D_p \frac{d\Delta p}{dx} - D_n \frac{d\Delta n}{dx}}{p_0\mu_p + n_0\mu_n}. \quad (6.41a)$$

Висновок: статичне поле є наслідком різниці дифузійних швидкостей електронів і дірок. Воно тим більше, чим менше концентрація рівноважних носіїв заряду n_0, p_0 . Максимального значення, за інших рівних умов, $\mathcal{E}_{\text{ст}}$ досягає у власного напівпровідника $n_0 = p_0 = n_i$. У сильно легованих (вироджених) напівпровідниках n_0^+ – або p_0^+ – типу статичне поле нейтралізується високою провідністю рівноважних носіїв заряду і практично відсутнє, нерівноважні електрони і дірки вільно дифундують нерозділеними, нейтральними парами і рекомбінують, маючи однаковий час життя $\tau = \tau_n = \tau_p$.

В умовах наявності \mathcal{E}_{cm} дифузію і дрейф не рівноважних і просторово не розділених електронно – діркових пар можна розглядати як дифузію пакету квазінейтральних частинок, у яких коефіцієнт дифузії і мобільність мають проміжні значення у порівнянні із значеннями цих же параметрів для окремо взятих електронів і дірок.

Запишемо рівняння неперервності (6.4) і (6.5) для випадку, коли напівпровідник знаходиться під дією постійного однорідного поля $\mathcal{E}_x \gg \mathcal{E}_{cm}$:

$$\frac{\partial \Delta n}{\partial t} = \frac{1}{e} \frac{\partial j_n}{\partial x} - \frac{\Delta n}{\tau}; j_n = \sigma_n \mathcal{E}_x + e D_n \frac{\partial \Delta n}{\partial x}; \sigma_n = e n \mu_n. \quad (6.42)$$

$$\frac{\partial \Delta p}{\partial t} = -\frac{1}{e} \frac{\partial j_p}{\partial x} - \frac{\Delta p}{\tau}; j_p = \sigma_p \mathcal{E}_x - e D_p \frac{\partial \Delta p}{\partial x}; \sigma_p = e p \mu_p. \quad (6.43)$$

Врахуємо, що $\Delta n = \Delta p \ll n_0, p_0$; $\sigma_n \approx \sigma_{n0} = e n_0 \mu_n$; $\sigma_p \approx \sigma_{p0} = e p_0 \mu_p$; домножимо (6.42) на σ_{p0} , а (6.43) – на σ_{n0} ; складемо обидва рівняння в умовах стаціонарної рівноваги $\frac{\partial \Delta n}{\partial t} = \frac{\partial \Delta p}{\partial t} = 0$:

$$\frac{D_n \sigma_{p0} + D_p \sigma_{n0}}{\sigma_{p0} + \sigma_{n0}} \frac{\partial^2 \Delta n}{\partial x^2} + \frac{\mu_n \sigma_{p0} - \mu_p \sigma_{n0}}{\sigma_{p0} + \sigma_{n0}} \mathcal{E}_x \frac{\partial \Delta n}{\partial x} - \frac{\Delta n}{\tau} = 0. \quad (6.44)$$

Порівняємо (6.44) із (6.23) для монополярного випадку, використаємо співвідношення Ейнштейна (3.61) ($\mu_n / D_n = \mu_p / D_p = e / k_B T$) і визначимо **коефіцієнт біполярної дифузії D_b і біполярну дрейфову мобільність μ_ϵ** :

$$D_b = \frac{D_n \sigma_{p0} + D_p \sigma_{n0}}{\sigma_{p0} + \sigma_{n0}} = \frac{n_0 + p_0}{n_0 / D_p + p_0 / D_n} = \frac{k_B T}{e} \frac{n_0 + p_0}{n_0 / \mu_p + p_0 / \mu_n}; \quad (6.45)$$

$$\mu_\epsilon = \frac{\mu_n \sigma_{p0} - \mu_p \sigma_{n0}}{\sigma_{p0} + \sigma_{n0}} = \frac{n_0 - p_0}{n_0 / \mu_p + p_0 / \mu_n}.$$

Аналіз формул для коефіцієнта біполярної дифузії D_b і біполярної дрейфової мобільності μ_ϵ показує, що їх значення лежать у межах:
 $D_p < D_b < D_n$, $\mu_p < \mu_\epsilon < \mu_n$.

Із (6.45) випливає, що D_b і μ_ϵ не пов'язані співвідношенням Ейнштейна. Коефіцієнту біполярної дифузії D_b необхідно співставити інше значення мобільності:

$$\mu_D = \frac{e}{k_B T} D_b, \quad (6.46)$$

яка отримала назву *біполярна дифузійна мобільність*. Із формули для D_b у рівнянні (6.45) для μ_D отримаємо:

$$\mu_D = \frac{n_0 + p_0}{n_0/\mu_p + p_0/\mu_n}. \quad (6.47)$$

Знак μ_ϵ , відповідно до (6.45), залежить від співвідношення n_0 і p_0 , тобто рівноважних носіїв заряду. У електронному напівпровіднику ($n_0 > p_0$) $\mu_\epsilon > 0$ і пакет квазінейтральних нерівноважних електронно – діркових пар буде дрейфувати у тому ж напрямку, що і дірки p_0 – неосновні носії заряду. У дірковому напівпровіднику ($p_0 > n_0$) $\mu_\epsilon < 0$ і пакет буде дрейфувати у тому ж напрямку, що і електрони n_0 – неосновні носії заряду.

Наслідок: *пакет квазінейтральних нерівноважних носіїв заряду буде дифундувати під впливом градієнта концентрації $\Delta n(x) = \Delta p(x)$, дрейфувати під впливом зовнішнього поля \mathcal{E}_x , просторово розширюватися (електрони і дірки дрейфують у протилежних напрямках) і рекомбінувати.*

У власних напівпровідниках ($n_0 = p_0 = n_i$) маємо:

$$D_b = 2 \frac{D_n D_p}{D_n + D_p} = 2 \frac{k_B T}{e} \frac{\mu_n \mu_p}{\mu_n + \mu_p}; \mu_D = 2 \frac{\mu_n \mu_p}{\mu_n + \mu_p}; \mu_\epsilon = 0. \quad (6.48)$$

Наслідок: *у власних напівпровідниках біполярні параметри D_b і μ_D від концентрації рівноважних носіїв заряду не залежать і визначаються лише D_n і D_p або μ_n і μ_p . Значення $\mu_\epsilon = 0$ свідчить про те, що зовнішнє поле не впливає на просторовий розподіл пакету квазінейтральних нерівноважних носіїв заряду.*

У сильно легованих напівпровідниках ($n_0 \gg p_0$ або $p_0 \gg n_0$) біполярні параметри D_b , μ_D , μ_ε дорівнюють відповідним до типу провідності параметрам неосновних носіїв заряду:

$$n\text{-тип } D_b = D_p, \mu_D = |\mu_\varepsilon| = \mu_p; \quad p\text{-тип } D_b = D_n, \mu_D = |\mu_\varepsilon| = \mu_n. \quad (6.49)$$

Наслідок: у сильно легованих напівпровідниках дифузія і дрейф не рівноважних носіїв заряду визначаються коефіцієнтом дифузії і мобільністю неосновних носіїв заряду.

Для повноти набору параметрів біполярної дифузії і дрейфу необхідно додати ще два – **біполярну дифузійну довжину** $L_D = \sqrt{D_b \tau}$, яка визначає розподіл не рівноважних носіїв у слабких зовнішніх полях ($\mathcal{E}_x \ll \mathcal{E}_{cm}$), і **біполярну довжину дрейфу** $L_\varepsilon = \mu_\varepsilon \mathcal{E}_x \tau$, яка визначає розподіл у сильних полях ($\mathcal{E}_x \gg \mathcal{E}_{cm}$).

Резюме по р. 6

1. Рівняння неперервності враховує усі можливі зміни концентрації нерівноважних носіїв заряду у координатному і часовому вимірах, як наслідки процесів *дрейфу, дифузії, генерації і рекомбінації*.
2. У стані стаціонарної рівноваги усі вищевказані процеси взаємно врівноважені таким чином, що повна концентрація носіїв заряду від часу не залежить.
3. За наявності нерівноважного заряду основних носіїв заряду його зміни у часі і просторі характеризується *часом релаксації Максвелла і радіусом екранування Дебая*. У кількісному вимірі час релаксації Максвелла і радіус екранування Дебая на декілька порядків менші, ніж дифузійна довжина і час життя нерівноважних неосновних носіїв заряду.
4. Дифузія і дрейф нерівноважних неосновних носіїв заряду характеризуються *дифузійною довжиною і довжиною дрейфу*. У випадку сумісної дії дифузії і дрейфу характеристикою процесів є *довжина*

затягування. У кількісному вимірі довжина затягування може бути як більшою, так і меншою дифузійної довжини.

5. Якщо монополярна нерівноважна концентрація створюється на поверхні напівпровідника, то розрізняють (у залежності від напрямку і величини електричного поля, знаку заряду) такі дифузійно – дрейфові процеси: *інжекція, ексклюзія, екстракція, акумуляція*.

6. У випадку біполярної дифузії та неоднакових дифузійних швидкостей електронів і дірок відбувається їх просторове розділення у процесі дифузії. Наслідком такого розділення є порушення нейтральності і статичне електричне поле. Дифузія супроводжується дрейфом.

7. Для характеристики дифузійно – дрейфових процесів у цьому випадку вводять поняття *коефіцієнт біполярної дифузії і біполярну дрейфову мобільність*. У кількісному вимірі ці параметри мають проміжні значення між коефіцієнтами монополярної дифузії електронів і дірок та монополярними мобільностями електронів і дірок.

8. У вироджених напівпровідниках біполярні параметри співпадають із відповідними по типу провідності параметрами неосновних носіїв заряду.

7. Поглинання світла у напівпровідниках

7.1. Характеристики поглинання

Поглинання фотонів – масовий процес і підпорядковується статистичним закономірностям. Зокрема, при масових процесах зміна будь – якого параметра, що характеризує процес у цілому, прямо пропорційна миттєвому значенню цього ж параметра. Оформимо це положення кількісно.

Перпендикулярно до поверхні твердого тіла падає паралельний монохроматичний світловий потік з інтенсивністю Φ_λ (індекс λ – довжина світлової хвилі – ознака монохроматичності). Інтенсивність потоку Φ_λ будемо визначати кількістю фотонів із довжиною хвилі λ , що падають на одиницю поверхні за одиницю часу. Частково потік відбивається, частково поглинається, частково проходить через речовину обмежених розмірів без поглинання.

Визначимо частину потоку Φ_λ , що поглинається у прошарку речовини $x, x + dx$ (відлік x від поверхні вглиб твердого тіла). Зміна інтенсивності потоку за рахунок поглинання:

$$d\Phi_\lambda(x) = -\alpha_\lambda \Phi_\lambda(x) dx, \quad (7.1)$$

де α_λ – коефіцієнт пропорційності, якій не залежить від Φ_λ , знак „-“, враховує зменшення інтенсивності потоку.

Інтегруючи (7.1) у межах

$$x = 0, \Phi_\lambda(x) = \Phi_\lambda(0); x \rightarrow \infty, \Phi_\lambda(x) \rightarrow 0, \text{ отримаємо:}$$

$$\Phi_\lambda(x) = \Phi_\lambda(0) \exp(-\alpha_\lambda x), \quad (7.2)$$

де $\Phi_\lambda(0)$ - інтенсивність потоку на поверхні.

Співвідношення (7.2) відоме як *закон Бугера – Ламберта*. Фізичний зміст α_λ з'ясуємо, поклавши:

$$\begin{aligned} x &= L_\lambda = 1/\alpha_\lambda \\ \Phi_\lambda(x = L_\lambda) &= \Phi_\lambda(0)/e \approx \Phi_\lambda(0)/2,7. \end{aligned} \quad (7.3)$$

Відстань L_λ , на якій інтенсивність паралельного монохроматичного потоку світла зменшується в e -раз (2,7 рази) називається пробігом фотонів.

Коефіцієнт α_λ , обернений пробігу фотонів L_λ , називається показником поглинання і визначає кількість фотонів, що поглинаються на одиниці довжини шляху їх пробігу.

Залежність показника поглинання від довжини хвилі світла $\alpha_\lambda = f(\lambda)$ або від енергії фотонів $\alpha_\lambda = f(h\nu)$ називається спектром поглинання.

Якщо товщина зразка у напрямку світлового потоку d , інтенсивність потоку на неосвітленій грані буде, відповідно до (7.2),:

$$\Phi_\lambda(d) = \Phi_\lambda(0) \exp(-x/d). \quad (7.4)$$

Добуток показника поглинання на товщину зразка $\alpha_\lambda d$ називається його оптичною товщиною.

Розрізняють два граничних випадки:

$$\alpha_\lambda d \gg 1 - \text{оптично товстий зразок}; \quad (7.5)$$

$$\alpha_\lambda d \ll 1 - \text{оптично тонкий зразок}. \quad (7.5a)$$

У оптично товстому зразку потік поглинається у тонкому поверхневому прошарку $L_\lambda \ll d$, світло у об'єм не проникає.

У оптично тонкому зразку поглинання у межах товщини d незначне і практично рівномірне.

7.2. Відбиття світла

Коефіцієнт відбиття R_r визначає частку повного потоку Φ_λ , яка відлунюється від освітленої поверхні $\Phi_{\lambda R}$:

$$R_r = \Phi_{\lambda R} / \Phi_\lambda. \quad (7.6)$$

Залежність коефіцієнта відбиття від довжини світлової хвилі $R_r = f(\lambda)$ або від енергії фотонів $R_r = f(h\nu)$ називається **спектром відбиття**.

Коефіцієнт відбиття R_λ , показник поглинання α_λ і коефіцієнт заломлення n взаємозалежні:

$$R_r \approx ((n-1)^2 + \alpha_\lambda^2) / ((n+1)^2 + \alpha_\lambda^2). \quad (7.7)$$

Із (7.7) випливає, що тверді тіла, які сильно поглинають ($\alpha_\lambda^2 \gg n^2$), сильно й відбивають ($R_r \leq 1$). Така ситуація характерна для металів у видимій і ультрафіолетовій частинах спектру (“металевий блиск”).

Для діелектриків у видимій і інфрачервоній частинах спектру ($\alpha_\lambda^2 \ll n^2$) і коефіцієнт відбиття можна визначати із формули:

$$R_r \approx (n-1)^2 / (n+1)^2. \quad (7.8)$$

Для скла $n = 1,5$ і чисто діелектричне відбиття $R_r = 0,04 = 4 \%$.

Напівпровідники мають $n = 3 \dots 4$, чисто діелектричне відбиття у них значно більше $R_r = 25 \dots 36 \%$. При фундаментальному поглинанні у напівпровідниках показник поглинання досягає значень $\alpha_\lambda \approx 10^4 \dots 10^6 \text{ м}^{-1}$ і вони стають непрозорими ($R_r \leq 0,9 \dots 1$), як і метали мають “металевий блиск”.

Із урахуванням коефіцієнта відбиття R_r , інтенсивність світлового потоку на поверхні $\Phi_\lambda(0) = \Phi_\lambda(1 - R_r)$ і закон Бугера – Ламберта має вигляд:

$$\Phi_\lambda(x) = \Phi_\lambda(1 - R_r) \exp(-\alpha_\lambda x). \quad (7.9)$$

7.3. Механізми поглинання

У металів енергія фотонів витрачається на збудження вільних електронів, енергетичний спектр яких квазінеперервний. Поглинання відбувається в межах декількох атомних прошарків на поверхні і супроводжується перевипромінюванням фотонів. Саме цим пояснюється сильна відбивна спроможність металів.

Діелектрики практично не мають вільних електронів. У видимій частині спектру більшість діелектриків поглинає дуже слабо. Електронне поглинання спостерігається у фіолетовій і ультрафіолетовій частинах спектра. Поглинання за рахунок збудження теплових коливань атомів (фононне поглинання) носить резонансний характер і спостерігається у вигляді смуг поглинання в інфрачервоній частині спектра (десятки мікрометрів).

Поглинання вільними носіями заряду у напівпровідниках. При наявності вільних носіїв заряду можливі їхні переходи у межах відповідної зони – електронів у зоні провідності (рис.7.1), дірок у валентній зоні при поглинанні фотонів.

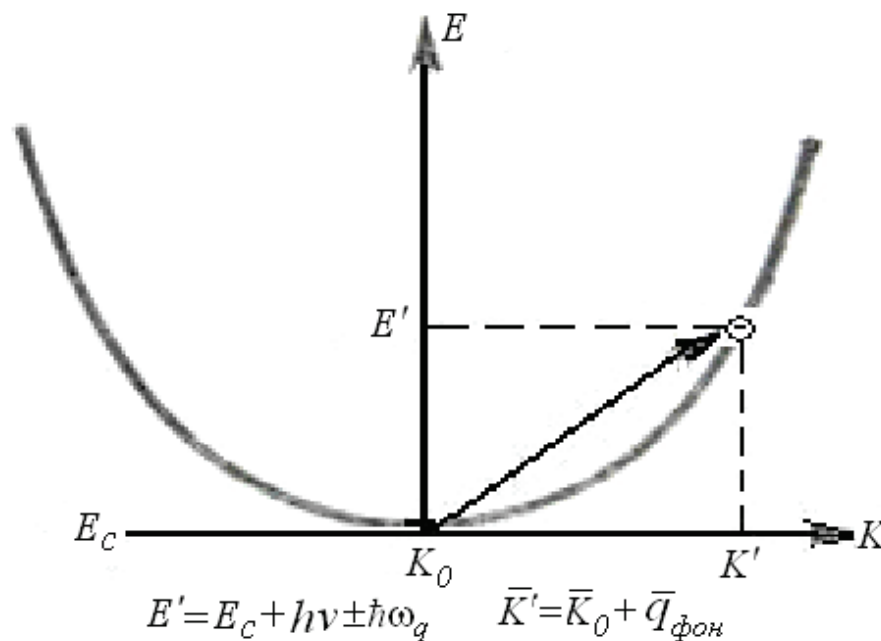


Рис.7.1. Поглинання світла вільними електронами

Закони зберігання енергії і хвильового вектора для електронів можна виконати лише при одночасній участі у процесах поглинання розсіяння на фононах та іонізованих домішках. Справа у тому, що імпульс фотона у кристалі мізерно малий і не відповідає енергії, яка передається електрону . У той же час, при розсіянні імпульс носія заряду змінюється у широких межах.

Взаємодія електрон-фотон-фонон відбувається одночасно і закони зберігання енергії й імпульсу мають такий вигляд (без урахування імпульсу фотона):

$$E' = E_c + h\nu \pm \hbar\omega_q; \bar{k}' \approx \bar{k}_0 + \bar{q}_{фон}, \quad (7.10)$$

де E_c, \bar{k}_0 – енергія і хвильовий вектор електрона до взаємодії; E', \bar{k}' – енергія і хвильовий вектор електрона після взаємодії; $h\nu = E_{фот}$ – енергія фотона; $\hbar\omega_q$ – енергія фонона; $\bar{q}_{фон}$ – хвильовий вектор фонона.

Знак „ \pm ” у (7.10) означає, що закони збереження хвильового вектора можна виконати як при поглинанні, так і при збудженні фонона.

Показник поглинання вільними електронами $\alpha_{\lambda,n}$ прямо пропорційний їх концентрації n , довжині хвилі λ^β , обернено пропорційний середньому часу релаксації τ і ефективній масі m_n^* . Таким закономірностям можна дати просте фізичне пояснення: чим більша концентрація n , тим більша імовірність електрон – фотонної взаємодії; чим більша довжина хвилі λ , тим менша енергія фотонів і, відповідно, зміна модуля хвильового вектора електрона, що полегшує виконання закону збереження імпульсу. Показник степеню β залежить від механізму розсіювання (невироджений напівпровідник): $\beta = 1,5$ – акустичне фононне розсіювання; $\beta = 2,5$ – оптичне фононне розсіювання; $\beta = 3,5$ – розсіювання на іонізованих домішках.

Узагальнимо залежність α_λ від λ :

$$\alpha_\lambda(\lambda) = A\lambda^{1,5} + B\lambda^{2,5} + C\lambda^{3,5}, \quad (7.12)$$

де A, B, C – розмірні коефіцієнти за участі акустичних, оптичних фононів і іонізованих домішок у розсіянні. Домінування того чи іншого механізму розсіювання залежить від температури і концентрації домішок. У табл. 7.1 наведені значення β для деяких напівпровідників.

Таблиця 7.1		
Напівпро- відник	N_d , 10^{23} м^{-3}	β
GaP	10	1,8
Ge	0,5–5	2
InSb	1–3	2
AlSb	0,4–4	2
InP	0,4–4	2,5
InAs	0,3–8	3
GaAs	1–5	3
GaSb	0,5	3,5

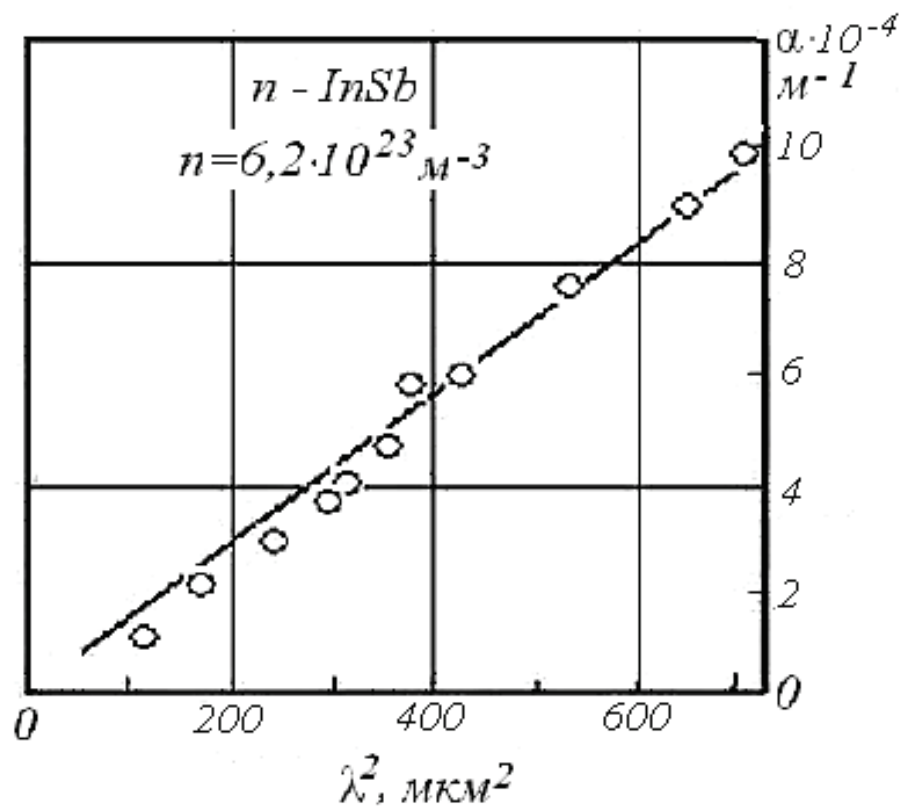


Рис.7.2. Експериментальна залежність $\alpha_\lambda = f(\lambda^2)$, яка свідчить про домінування акустичного і оптичного фононного розсіяння

На рис.7.2 подана експериментальна залежність $\alpha_\lambda = f(\lambda^2)$ для n -InSb ($n = 6,2 \cdot 10^{23} \text{ м}^{-3}$), яка відповідає закономірності $\alpha_\lambda \sim \lambda^2$, що означає домінування акустичного і оптичного фононного розсіяння.

У діркових напівпровідниках поглинання вільними дірками має ряд особливостей у порівнянні з поглинанням вільними електронами. Валентна зона складається з 3-х підзон, взаємне розташування яких у p -Ge показано на рис.2.4 (V_1 – „важкі” дірки, V_2 – „легкі” дірки, V_3 – „проміжні” дірки).

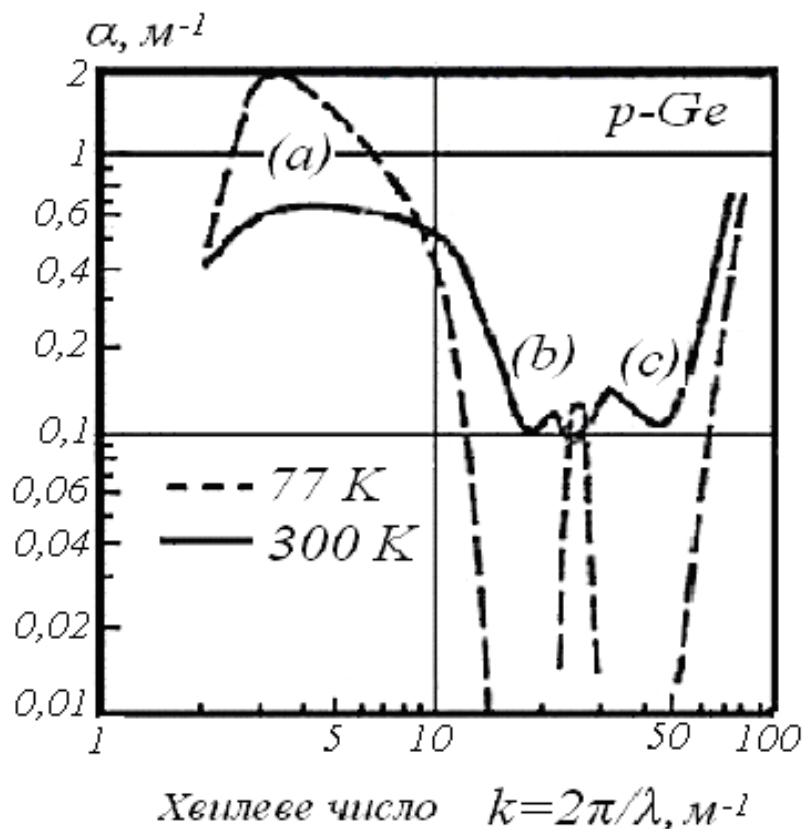


Рис.7.3. Спектр поглинання світла вільними дірками у германії при різних температурах

Світло збуджує переходи між підзонами $V_1 \rightarrow V_2$ (a), $V_1 \rightarrow V_3$ (b), $V_2 \rightarrow V_3$ (c). Відповідні спектральні максимуми для цих переходів показані на рис.7.3. На низьких температурах (77 K) максимуми поглинання для переходів $V_1 \rightarrow V_2$,

та $V_2 \rightarrow V_3$ чітко розрізняються, але переходи $V_1 \rightarrow V_3$ перекриваються із краєм власного поглинання, оскільки енергія спин – орбітального розщеплення $\Delta E_V = 0,3$ еВ одного порядку із E_g германію. Такі переходи відбуваються без зміни хвильового вектора, не вимагають участі фононів і називаються прямими (вертикальними).

Поглинання вільними електронами і дірками не змінює повного числа носіїв заряду у відповідній зоні і повинно бути електрично не активним. Однак перерозподіл дірок у межах валентної зони супроводжується зміною їхньої ефективної маси і мобільності. Зовнішнім проявом цього є ефект „несправжньої” фотопровідності – при освітленні p -Ge у довгохвильовій частині спектру фотопровідність зменшується.

Домішкове поглинання. Якщо у напівпровіднику є домішкові центри, здатні віддавати або приймати електрони при фотоіонізації, то стає можливою монополярна фотогенерація з утворенням вільних електронів або дірок.

Обов’язковими для ефективного домішкового поглинання є такі умови:

$$\Delta E_{\text{дом}} \leq h\nu < E_g; k_B T \ll \Delta E_{\text{дом}}, \quad (7.13)$$

де $\Delta E_{\text{дом}}$ – глибина залягання домішок.

Перша із умов (7.13) означає довгохвильову частину спектра для певного напівпровідника, а друга умова – відсутність термоіонізації домішкових центрів.

Кількість домішкових центрів, здатних локалізувати і утримувати електрони, на декілька порядків менша числа валентних електронів. При достатній інтенсивності світлового потоку можлива часткова або повна фотоіонізація домішок, показник домішкового поглинання стає залежним від інтенсивності потоку.

Висновки: – *ефективне домішкове поглинання можливе лише при незначній термоіонізації домішкових центрів;*

- *домішкове поглинання* має місце у довгохвильовій частині загального спектру поглинання певного напівпровідника $h\nu < E_g$;
- *показник домішкового поглинання* залежить від інтенсивності світлового потоку – спадає при зростанні інтенсивності потоку, тому закон Бугера-Ламберта (7.2) у загальному випадку не виконується для домішкового поглинання;
- ефект домішкового поглинання використовується для визначення глибини залягання домішкових центрів та у фотоприймачах для переважно інфрачервоної частини спектру.

Фундаментальне (власне) поглинання. Фундаментальне поглинання є основним механізмом поглинання в інфрачервоній і видимій частинах спектру для напівпровідників. Головна умова фундаментального поглинання – енергія фотонів достатня для фотоіонізації валентних зв'язків:

$$h\nu \geq E_g. \quad (7.14)$$

Знак рівності в умові (7.14) визначає *край власного поглинання* λ_k :

$$\lambda_k = hc / E_g, \quad (7.15)$$

де h - постійна Планка, c - швидкість світла.

Для зручності розрахунків по формулі (7.15) запишемо її у такому вигляді:

$$\lambda_k (\text{мкм}) = 1,239 / E_g (\text{eV}). \quad (7.16)$$

Частина спектру власного поглинання напівпровідника із шириною забороненої зони E_g , для якої $\lambda < \lambda_k$, називається *короткохвильовою*, для $\lambda > \lambda_k$ – *довгохвильовою*.

Власне поглинання у короткохвильовій частині спектру супроводжується біполярною генерацією нерівноважних електронно – діркових пар $\Delta n = \Delta p$.

Закони зберігання енергії і імпульса для електронних переходів при власному поглинанні визначаються енергетичною зонною структурою напівпровідника.

Власне поглинання у прямо зонних напівпровідниках. Для прямозонних напівпровідників (положення абсолютних екстремумів енергії у зоні провідності k_{\min} і валентній зоні k_{\max} у імпульсному просторі співпадають) закони зберігання енергії і імпульсу:

$$E' \geq h\nu - E_g, k_{\min} \approx k_{\max}. \quad (7.16)$$

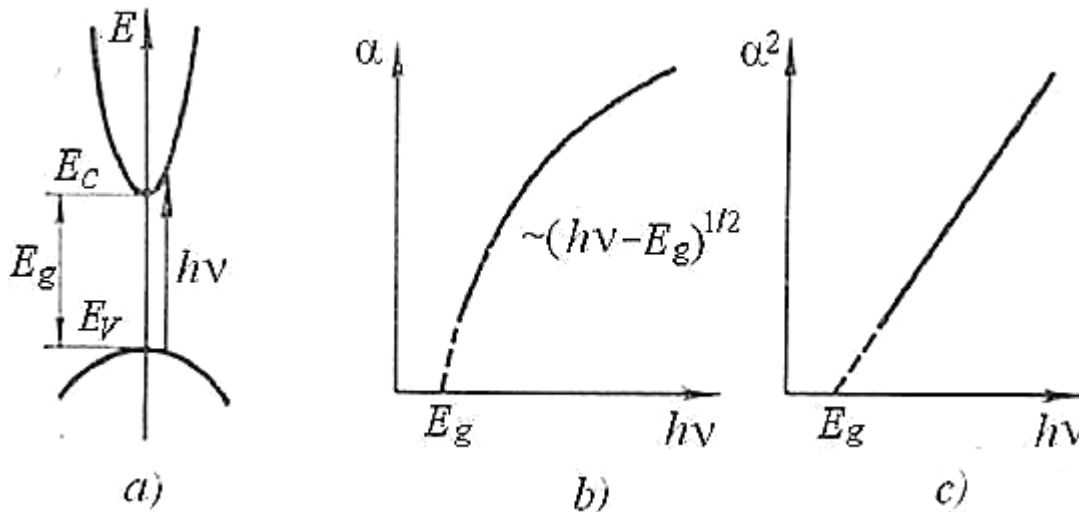


Рис.7.4. Електронні переходи у прямозонних напівпровідниках (а) та спектри поглинання світла: (b) – простий, (c) – модифікований

Переходи, що відповідають законам зберігання (7.16), називаються *прямими або вертикальними* (рис.7.4,(a)).

Наближена рівність у законі зберігання імпульсу пов'язана з тим, що нехтуємо імпульсом фотона. До прямозонних напівпровідників відносяться практично усі сполуки $A^{III}B^V$.

Спектр поглинання для прямих переходів $\alpha = f(h\nu)$ поділяється шириною забороненої зони E_g на дві частини:

$$\left. \begin{aligned} \alpha &\sim (h\nu - E_g)^{1/2} \rightarrow h\nu > E_g \\ \alpha &= 0 \rightarrow h\nu \leq E_g \end{aligned} \right\}. \quad (7.16a)$$

По краю спектру поглинання можна визначити ширину забороненої зони прямозонного напівпровідника E_g , екстраполюючи спектральну лінію на

$\alpha = 0$ (рис.7.4,(b)). З метою зменшення похибки визначення E_g спектральну лінію будують у координатах $\alpha^2 = f(h\nu)$ (рис.7.4,(c)).

Показник фундаментального поглинання для прямих переходів має значення у межах $10^6 - 10^8 \text{ м}^{-1}$, що на 3 – 5 порядків перевищує показники поглинання для інших механізмів.

Власне поглинання у непрямозонних напівпровідниках. Загальний вигляд спектра поглинання і вплив температури на нього показаний на прикладі германія, рис.7.5. Як і слід було очікувати, зі збільшенням температури ширина забороненої зони зменшується.

Для непрямозонних напівпровідників (абсолютні екстремуми в різних точках імпульсного простору), окрім прямих (вертикальних) переходів із законами зберігання енергії і хвильового вектору (7.16), можливі *непрямі, віртуальні переходи* (рис.7.6) за участю електронів, фотонів і фононів. Розподіл „ролей” для віртуальних переходів такий же, як і у випадку поглинання вільними електронами.

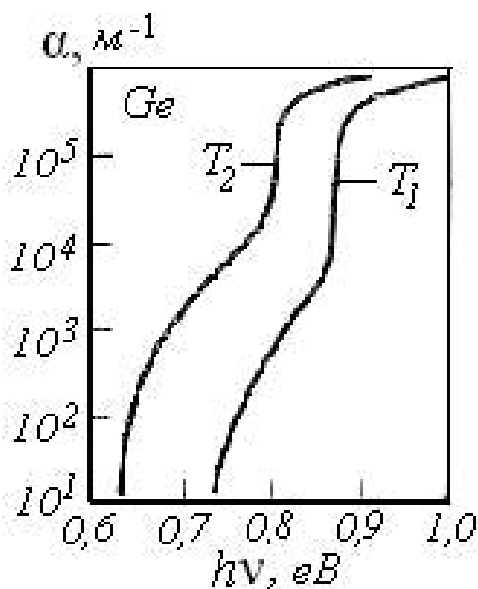


Рис.7.5. Спектр поглинання германія, $T_1=77 \text{ K}$, $T_2=300 \text{ K}$

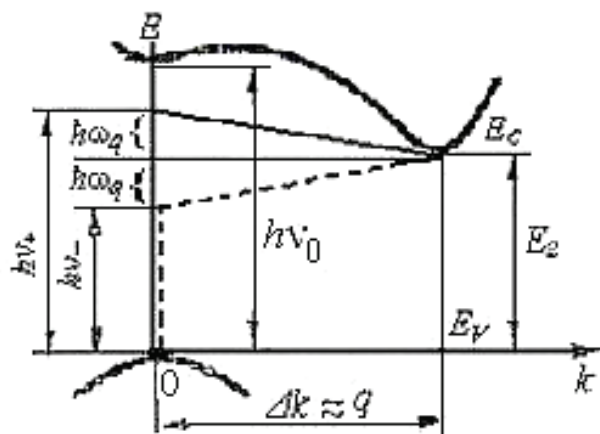


Рис.7.6. Віртуальні електронні переходи при власному поглинанні світла к непрямозонних напівпровідниках

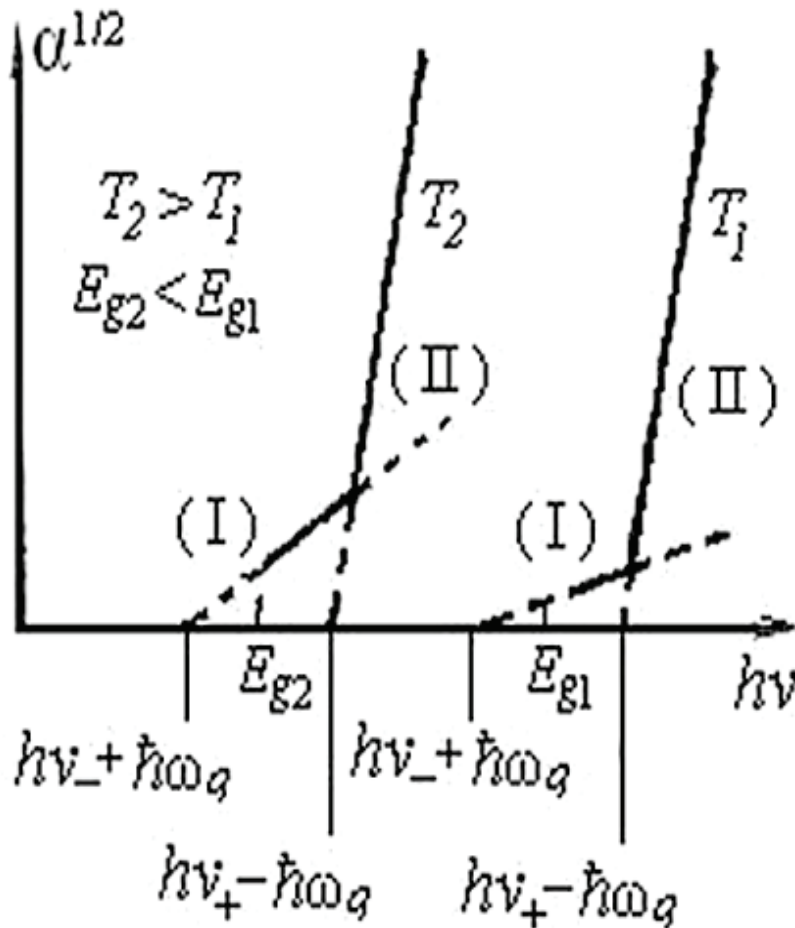


Рис.7.7. Тонка структура краю власного поглинання світла для непрямих віртуальних переходів: (I) – поглинання фононів, (II) – збудження фононів

Закони зберігання енергії і хвильового вектору для непрямих переходів:

$$\left. \begin{aligned} (I) \quad \hbar\nu_- < E_g &\rightarrow E_g = \hbar\nu_- + \hbar\omega_q; \\ (II) \quad \hbar\nu_+ > E_g &\rightarrow E_g = \hbar\nu_+ - \hbar\omega_q; \\ \bar{k}_{\min} - \bar{k}_{\max} = \Delta\bar{k} &\approx \bar{q}. \end{aligned} \right\}, \quad (7.17)$$

де $\hbar\omega_q$ – енергія фонона із хвильовим вектором \bar{q} ; \bar{k}_{\min} , \bar{k}_{\max} – хвильові вектори абсолютних екстремумів зони провідності і валентної зони, відповідно; знаки „+” та „-” у законах зберігання енергії враховують, що одна і та ж зміна хвильового вектора електрона $\Delta\bar{k}$ досягається як при поглинанні фонона (знак „+”), так і при збудженні фонона (знак „-”), для чого енергія фотона повинна

бути або менша ширини забороненої зони $h\nu_- < E_g$, або більша $h\nu_+ > E_g$; ширина забороненої зони E_g знаходиться посередині між $h\nu_-$ і $h\nu_+$.

Поділ переходів на два етапи (пунктирні лінії на рис.7.6 – поглинання фононів, суцільні – збудження фононів) має суто умовний характер, насправді ж *взаємодія фотон – електрон – фонон відбувається одночасно*.

Імовірність непрямих квантових переходів і показник поглинання квадратично залежать від енергії фотонів:

$$\alpha(h\nu) \sim \left[h\nu - (E_g \pm \hbar\omega_q) \right]^2. \quad (7.18)$$

Для непрямозонних напівпровідників (германій, кремній) структура краю поглинання у цілому досить складна (рис.7.7.) тому, що, відповідно до законів зберігання (7.16) і (7.17), є три значення краю власного поглинання: $h\nu_- < h\nu_+ < h\nu_0$. Край поглинання $h\nu_0$ – прямі (вертикальні) переходи по центру зведеної зони $k = 0$, рис.7.6.

На рис.7.7 показаний також вплив температури на положення і форму краю власного поглинання. При зменшенні температури ширина забороненої зони E_g збільшується і, як наслідок, зростає і енергія фотонів, необхідна для фотоіонізації валентних зв'язків. Окрім того, зміщується точка зламу між (I) і (II) – в умовах низьких температур (нижче температури Дебая) поглинання фононів стає менш імовірним через малу їх кількість. Закони зберігання (7.17) будуть виконуватися переважно за рахунок збудження фононів. В умовах високих температур (вище температури Дебая) у спектрі теплових коливань атомів присутні фонони усіх можливих енергій $0 < \hbar\omega_q \leq \hbar\omega_{q\max}$ і хвильових векторів $0 < \bar{q} \leq \bar{q}_{\max}$, процеси поглинання і збудження фононів у законах зберігання (7.17) стають рівноймовірними.

Для зменшення похибки визначення E_g при екстраполяції (7.18) на $\alpha = 0$, спектри поглинання будують у координатах $\alpha^{1/2} = f(h\nu)$, які у цьому випадку

будуть мати вигляд відрізків прямих ліній (рис.7.7). Злами на краю спектра власного поглинання є зовнішніми ознаками наявності непрямих переходів.

Поглинання світла у вироджених напівпровідниках. У вироджених напівпровідниках власне поглинання йде переважно за рахунок вертикальних переходів (рис.7.8,(a)). Крім того, край власного поглинання зміщений у короткохвилову частину спектра відносно краю у невиродженому стані – зсув *Буриштейна – Мосса* (рис.7.8,(b)), який характерний для вузько зонних напівпровідників із малою ефективною масою електронів (InSb, InAs і т.п.).

Зсув краю власного поглинання у InSb суттєво залежить від концентрації електронів. Для прикладу, збільшення концентрації від 10^{24} до $2 \cdot 10^{25} \text{ м}^{-3}$ зсуває край поглинання з $h\nu = E_g = 0,2 \text{ еВ}$ до $h\nu \approx 1 \text{ еВ}$, що становить $\approx 5 E_g$. Такий значний зсув краю власного поглинання характерний лише для InSb. Цьому сприяє те, що щільність енергетичних рівнів у зоні провідності на 2 – 3 порядки менша ніж у інших напівпровідників (див. табл. 3.2) та вузька заборонена зона.

Рівень Фермі F_ϕ^+ у виродженій зоні провідності піднімається дуже високо відносно дна зони E_c при наповненні її електронами.

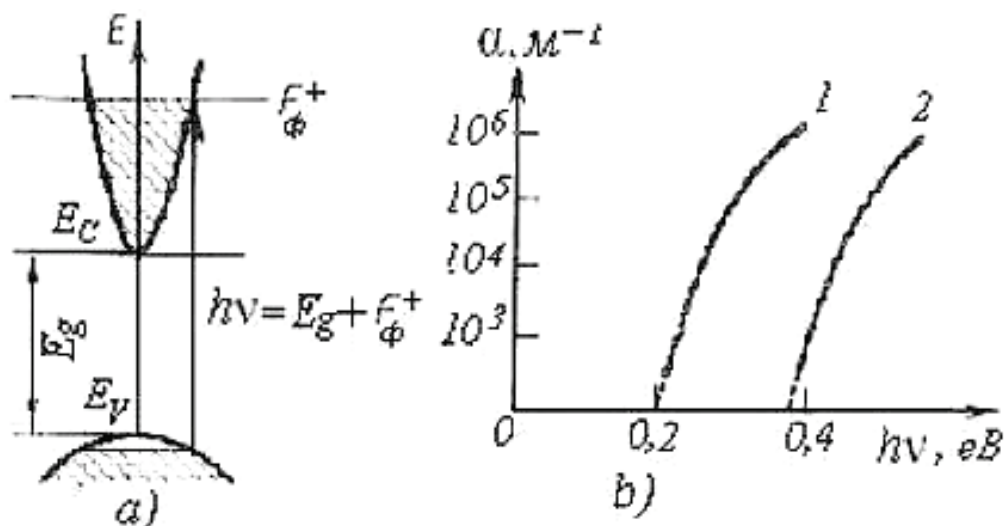


Рис.7.8. Вертикальні переходи у вироджену зону провідності (a). Спектр власного поглинання та зсув краю поглинання (b)

Концентраційний зсув Бурштейна – Мосса спостерігається і у інших вироджених напівпровідників, але він не такий значний у порівнянні із шириною забороненої зони.

Резюме по п. 7.1 – 7.3.

Основними характеристиками процесів поглинання світла слід вважати:

- просторовий розподіл щільності потоку фотонів у об'ємі напівпровідника
- закон Бугера - Ламберта (7.9);
- середнє значення пробігу фотонів L_λ (7.3);
- показник поглинання α_λ : число фотонів, що поглинаються на одиниці шляху пробігу;
- спектр поглинання: залежність показника поглинання від довжини світлової хвилі $\alpha_\lambda = f(\lambda)$ або від енергії фотонів $\alpha_\lambda = f(h\nu)$;
- оптична товщина зразка $\alpha_\lambda d$;
- коефіцієнт відбиття R_r (7.6).

До основних механізмів поглинання відносяться:

- поглинання вільними носіями заряду, якому відповідають закони зберігання (7.10);
- домішкове поглинання, яке реалізується за умов (7.13);
- власне (фундаментальне) поглинання, яке характеризується краєм поглинання λ_k (7.15). Частина спектру власного поглинання, для якої $\lambda < \lambda_k$ і $\alpha_\lambda \neq 0$, називається короткохвильовою. Частина спектру, для якої $\lambda > \lambda_k$ і $\alpha_\lambda \approx 0$, називається довгохвильовою;
- по спектру власного поглинання прямозонного напівпровідника можна визначити E_g , екстраполюючи спектральну лінію, побудовану у координатах $\alpha^2 = f(h\nu)$, на $\alpha_\lambda = 0$ (рис.7.5,в);
- у непрямозонних напівпровідниках структура спектру поглинання ускладнюється (рис.7.7.) – окрім прямих, вертикальних переходів, мають місце

непрямі, віртуальні оптичні переходи за участю фотонів, електронів і фононів із законами зберігання (7.17);

—для визначення E_g у непрямозонному напівпровіднику по віртуальним оптичним переходам їх спектр будують у координатах $\alpha_\lambda^{1/2} = f(h\nu)$, визначають два значення краю поглинання і посередині між ними відраховують E_g (рис.7.7.);

—по спектрам поглинання можна спостерігати вплив різних зовнішніх факторів на ширину забороненої зони, наприклад: *температури* (рис.7.7.); *механічного стиску або розтягу*, який змінює ширину забороненої зони і, відповідно, положення краю власного поглинання; *концентраційний зсув Бурштейна – Мосса* у вироджених напівпровідниках (край власного поглинання зсувається у короткохвильову частину спектра (рис.7.8, (a),(b))).

7.4. Фоторезистивний ефект

Стаціонарна фотопровідність. Розглянемо фотоелектрично активне власне поглинання, яке супроводжується біполярною генерацією електронно – діркових пар, з точки зору його впливу на електропровідність напівпровідника в умовах стаціонарної рівноваги (5.3).

Повну питому провідність запишемо у вигляді двох складових:

$$\sigma_{cm} = \sigma_0 + \Delta\sigma_{cm} = e(n_0\mu_n + p_0\mu_p) + e\Delta n(\mu_n + \mu_p), \quad (7.19)$$

де σ_0 – темнова рівноважна провідність, $\Delta\sigma_{cm}$ – стаціонарна фотопровідність, яка зумовлена світловою генерацією при власному поглинанні. При практичному застосуванні фоторезистивного ефекту необхідно зробити фотопровідність домінуючою у повній провідності, тобто, $\Delta\sigma_{cm} \gg \sigma_0$. Досягти цього можна у слабо легованих або власних напівпровідниках і плівкових фоторезисторах.

Темп світлової генерації визначимо таким чином:

$$G = \Delta n \tau = \eta \alpha_\lambda \Phi_\lambda (1 - R_r) \tau, \quad (7.20)$$

де τ – біполярний час життя, $\alpha_\lambda \Phi_\lambda (1 - R_r)$ – середнє число фотонів, які поглинаються на шляху пробігу фотонів за одиницю часу, η – квантовий вихід внутрішнього фотоефекту – кількість вільних носіїв заряду, яку створює один фотон при поглинанні.

Квантовий вихід спектрально залежний $\eta = f(h\nu)$, що ілюструє рис.7.10.

У германія (рис.7.9,(a)) фотогенерація починається із $h\nu \geq E_g = 0,68 \text{ eV}$, квантовий вихід у діапазоні $1 < h\nu \leq 2,7 \text{ eV}$ має незмінне значення $\eta = 1$ – один фотон, одна електронно – діркова пара. Надлишкова енергія носіїв заряду відносно E_g розсіюється на фононах та іонах. При подальшому зростанні енергії фотонів з’являються носії заряду (у першу чергу електрони, у яких $m_n^* < m_p^*$) із кінетичною енергією, достатньою для міжзонної ударної іонізації валентних зв’язків і утворення додаткових електронно – діркових пар і $\eta > 1$. На прикладі кремнію (рис.7.9,(b)) показаний вплив температури на квантовий вихід. Зі збільшенням температури енергетична межа початку ударної міжзонної іонізації зміщується у бік менших значень енергії фотонів, що можна пояснити температурним зменшенням ширини забороненої зони (див. п.2.13).

Запишемо $\Delta\sigma_{ct}$ у розгорнутому вигляді, спираючись на (7.19) і (7.20),:

$$\Delta\sigma_{ct} = e\eta\alpha_\lambda\Phi_\lambda(1 - R_r)\tau_{ct}(\mu_n + \mu_p). \quad (7.21)$$

Висновки: – один фотон при фундаментальному поглинанні створює одну електронно – діркову пару;

- надлишкова, у межах $2...3 E_g$, енергія носіїв заряду, отримана від фотона, повністю розсіюється на фононах та іонах і $\eta = 1$;

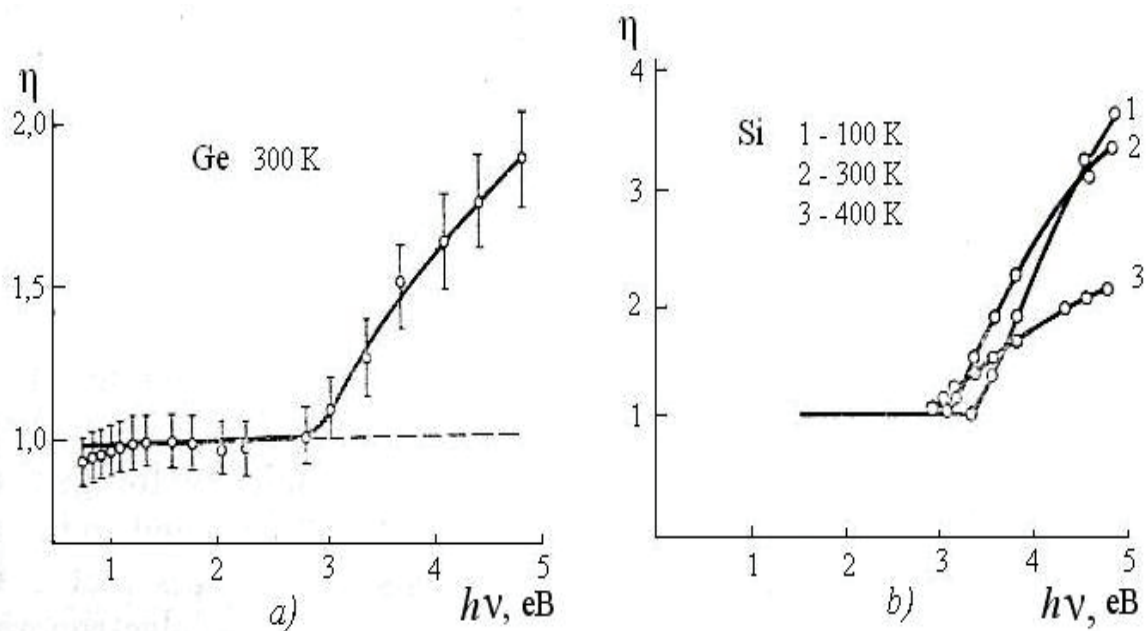


Рис.7.9. Спектральні характеристики квантового виходу внутрішнього фотоефекту у германії і кремнії

- якщо енергія носіїв заряду буде перевищувати вказані межі, стає можливою ударна іонізація валентних зв'язків з утворенням вторинних електронно – діркових пар за рахунок „гарячих” носіїв заряду і $\eta > 1$;
- температура впливає на квантовий вихід опосередковано через ширину забороненої зони, яка зменшується при підвищенні температури і зменшується енергія для початку ударної іонізації;
- енергія фотонів, при якій $\eta > 1$, відповідає, як правило, видимому та ультрафіолетовому діапазону світлових хвиль.

Спектральна характеристика фотопровідності. У формулі (7.21) явну спектральну залежність має показник власного поглинання (див. (7.12), (7.16a), (7.18)) і квантовий вихід внутрішнього фотоефекту (рис. 7.9). У будь-якому випадку, $\Delta\sigma_{cm}$ повинна зростати при збільшенні енергії фотонів (зменшенні

довжини світлової хвилі) починаючи із $h\nu \geq \Delta E_0$. Але експериментальні спектри фотопровідності дають дещо іншу закономірність, рис.7.11.

У міру зростання показника поглинання α_λ у короткохвильовій частині спектру зменшується пробіг фотонів L_λ (7.3) і поглинання зосереджується у тонкому поверхневому прошарку. Типове для напівпровідників значення $\alpha_\lambda \approx 10^6 - 10^8 \text{ м}^{-1}$, відповідно середній пробіг фотонів лежить у межах $L_\lambda \approx 10^{-8} - 10^{-6} \text{ м}$. Особливості розподілу нерівноважних носіїв заряду при поверхневій рекомбінації були розглянуті у п.5.4, основний висновок – *темплі поверхневої генерації G_s і поверхневий час життя τ_s менші відповідних об'ємних значень цих параметрів*. Окрім того, *поверхнева мобільність μ_s зазвичай менша об'ємної (п.5.4)*. Наслідком неявних залежностей $\tau(h\nu)$ та $\mu(h\nu)$ і є зменшення фотопровідності у коротко хвильовій частині спектру (рис.7.10).

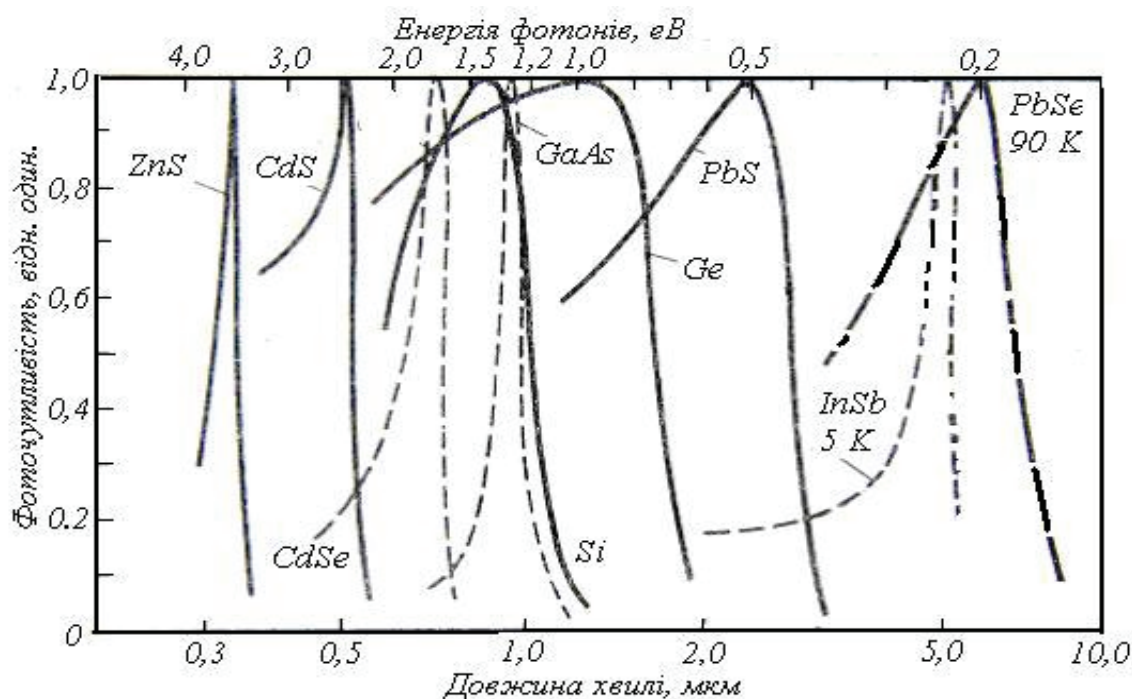


Рис.7.10. Відносна спектральна фоточутливість фоторезисторів на основі германію, кремнію та сполук $A^{III}B^V$, $A^{II}B^{VI}$, $A^{IV}B^{VI}$

Релаксація фотопровідності. Запишемо формулу для фотопровідності у скороченому вигляді, зручному для аналізу залежності фотопровідності від часу при вмиканні і вимиканні світла:

$$\Delta\sigma(t) = e\Delta n(t)\tau(\mu_n + \mu_p). \quad (7.22)$$

Поява нерівноважних носіїв заряду при фотогенерації автоматично викликає і їх рекомбінацію. У стаціонарних умовах темпи генерації і рекомбінації врівноважені. Рівняння стаціонарного процесу:

$$d\Delta n/dt = G - R = 0 \quad (7.23)$$

і розв'язком його є незалежна від часу стаціонарна концентрація Δn_{cm} . Концентрації Δn_{cm} відповідає стаціонарна фотопровідність $\Delta\sigma_{cm}$ (7.21).

Знайдемо нестаціонарну концентрацію $\Delta n(t)$, розв'язуючи рівняння

$$d\Delta n/dt = G - R = G - \Delta n/\tau \quad (7.24)$$

за початкових умов – *вмикання світла*:

$$t = 0, \Delta n(0) = 0; t \rightarrow \infty, \Delta n(\infty) \rightarrow \Delta n_{cm}. \quad (7.25)$$

Розв'язок рівняння (7.24) за умов (7.25) :

$$\Delta n(t) = \Delta n_{cm}(1 - \exp(-t/\tau)). \quad (7.26)$$

Після *вимикання світла* $G = 0$ і розв'язок (7.24) за початкових умов

$$t = 0, \Delta n(0) = \Delta n_{cm}; t \rightarrow \infty, \Delta n(\infty) \rightarrow 0 \quad (7.27)$$

має такий вигляд :

$$\Delta n(t) = \Delta n_{cm} \exp(-t/\tau). \quad (7.28)$$

Формули (7.26) і (7.28) – *релаксаційний процес* – зміна концентрації нерівноважних носіїв заряду з часом при вмиканні і вимиканні світла. Графічне зображення релаксаційного процесу у нормованих координатах $\Delta n(t)/\Delta n_{cm} = f(t/\tau)$ подано на рис.7.11 для прямокутного світлового імпульсу $\Phi(t)$ тривалістю $T \gg \tau$.

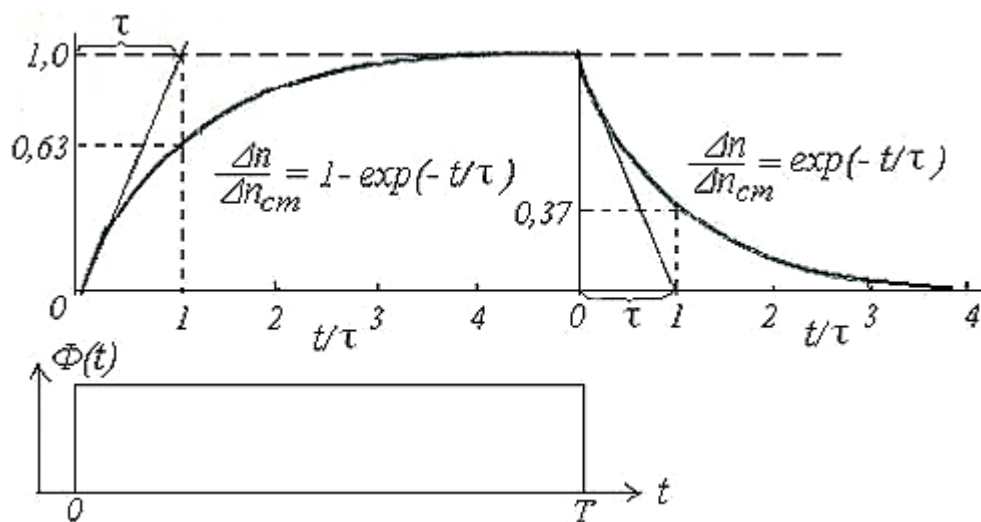


Рис.7.11. Релаксаційний процес під дією прямокутного світлового імпульсу тривалістю $T \gg \tau$

Час, необхідний для досягнення стаціонарної концентрації Δn_{cm} при вмиканні світла, повинен бути $\geq 4...5 \tau$. Наслідком цього є те, що при тривалості світлового імпульсу $T < 4 - 5 \tau$ максимальне значення концентрації Δn_{cm} і фотопровідності $\Delta \sigma_{cm}$ не досягається і чутливість фоторезистора як приймача світлового потоку зменшується.

Виникає проблема реєстрації модульованих світлових потоків – для зменшення інерційності фотоприймача необхідно зменшувати час життя τ , але ціною за малу інерційність буде зменшення його здатності реєструвати світлові потоки малої інтенсивності. Очевидно, що у кожному конкретному випадку необхідний певний компроміс між інерційністю і фоточутливістю. Релаксаційний процес широко використовується для визначення релаксаційного часу життя τ_{rel} .

На рис.7.11 показано як визначається час життя за допомогою дотичних до початкових ділянок релаксаційного процесу. Точніше релаксаційний час життя можна визначити по спадаючій концентрації після вимкнення світла, коли має місце лише рекомбінація. У цьому випадку відлік часу життя ведуть на рівні $0,37 \Delta n_{cm}$ (див. (5.7)).

Необхідно звернути увагу на співвідношення стаціонарного часу життя, розрахованого із стаціонарної фотопровідності (7.21), і релаксаційного часу життя. Нормальною є ситуація, коли $\tau_{\text{рел}} = \tau_{\text{ст}}$. На практиці спостерігаються випадки, коли $\tau_{\text{рел}} > \tau_{\text{ст}}$. Пояснюється це тим, що у релаксаційному процесі можуть приймати участь *рівні прилипання*, розташовані біля дна зони провідності (глибше ніж донорні) або стелі валентної зони (вище ніж акцепторні). У будь – якому випадку, глибина залягання ΔE_{np} цих рівнів лежить у межах $\Delta E_{d(a)} < \Delta E_{\text{np}} \ll \Delta E_0 / 2$. Саме віддаленість від середини забороненої зони призводить до того, що рекомбінаційні переходи через такі рівні мало імовірні, вони лише локалізують певну кількість носіїв заряду під час генерації. Темп генерації обмежений і час, необхідний для виходу на стаціонарну концентрації з одночасним наповненням рівнів прилипання, буде більший, ніж при відсутності цих рівнів. Після вимикання світла повинні рекомбінувати не лише надлишкові носії заряду у зоні, але і накопичені на рівнях прилипання.

Загальний наслідок участі рівнів прилипання у релаксаційному процесі – збільшення часу виходу на $\Delta n_{\text{ст}}$ і збільшення часу, необхідного для рекомбінації після вимкнення світла.

Зазвичай, рівні прилипання ефективні у поверхневому прошарку, співрозмірному із пробігом фотонів. Зовнішньою ознакою наявності рівнів прилипання є зменшення релаксаційного часу при освітленні поверхні не модульованим потоком у довгохвильовій частині спектру поглинання.

Резюме по п.7.4

- *фоторезистивний ефект* – зменшення опору напівпровідника під час фундаментального поглинання світла;
- зменшення опору є наслідком фотогенерації надлишкових, відносно теплової рівноважної концентрації, електронно – діркових пар;

- у стаціонарних умовах темп світлової генерації і темп рекомбінації врівноважені;
- повна стаціонарна провідність при освітленні має дві складові: теплову рівноважну і фотопровідність (7.19);
- для ефективного практичного застосування фоторезистивного ефекту необхідно зробити $\Delta\sigma_{ct}$ домінуючою складовою повної провідності, що досягається використанням слабо легованих або власних матеріалів і плівкових фоторезисторів;
- фотопровідність $\Delta\sigma_{ct}$ залежить від *квантового виходу внутрішнього фотоефекту* η , який, у свою чергу, є спектрально залежним, рис.7.9;
- *спектральна залежність фотопровідності* (рис.7.10) формується відповідними залежностями показника поглинання $\alpha_\lambda(h\nu)$, квантового виходу $\eta(h\nu)$, часу життя $\tau(h\nu)$ та мобільності $\mu(h\nu)$, які мають протилежний вплив на фотопровідність. Наслідком цього є те, що фотопровідність має спектральний максимум, положення якого визначається шириною забороненої зони та станом поверхні напівпровідника.

7.5. Фотовольтаїчні ефекти

Біполярна фото генерація у оптично і дифузійно тонких фото резисторах не супроводжується порушенням їх електричної нейтральності. Нейтральність напівпровідника при освітленні порушується у тих випадках, коли електронно – діркові пари просторово розділяються. При цьому виникає термодинамічно нерівноважна різниця потенціалів, яка може виконувати роботу у зовнішньому колі. Джерелом енергії для виконання роботи є енергія поглинутого світлового потоку.

Ефект Дембера. У спрощеному визначенні, ефект Дембера – порушення нейтральності оптично товстого напівпровідника із неоднаковими мобільностями електронів і дірок при власному поглинанні світла.

На однорідний ($\nabla n_0 = \nabla p_0 = 0$), оптично товстий ($\alpha_\lambda d \gg 1$, $\nabla \Delta n \neq 0$) прямокутний напівпровідник падає світловий потік із енергією фотонів $h\nu \geq \Delta E_0$. Світло поглинається у поверхневому прошарку l_s , середня товщина якого дорівнює пробігу фотонів $l_s = 1/\alpha_\lambda$. У об'єм світло не проникає. Поверхневий прошарок стає джерелом не рівноважних електронно – діркових пар і виникає їх спільна біполярна дифузія у об'єм, рис.12,(a),(b). За умови рівності $\mu_n = \mu_p$ і, відповідно, $D_n = D_p$, дифузійні швидкості електронів і дірок будуть однакові, просторового розділення електронів і дірок не буде, нейтральність напівпровідника не порушиться. Зазвичай мобільності і коефіцієнти дифузії неоднакові. Електрони, маючи більшу мобільність і коефіцієнт дифузії, випереджають дірки і їх заряди просторово розділяються. (рис.12,(b)(пунктирні лінії)). Наслідком такої динамічної поляризації є різниця потенціалів V_D і статичне електричне поле $\mathcal{E}_{ст}$, яке будемо називати *полем Дембера* E_D між освітленою та протилежною, неосвітленою гранями:

$$\mathcal{E} \equiv \mathcal{E}_D = V_D / d. \quad (7.29)$$

Напрямок поля \mathcal{E}_D такий, що воно *гальмує електрони і прискорює дірки*. Напруженість поля \mathcal{E}_D буде зростати, аж поки не зрівняються дифузійно – дрейфові швидкості електронів і дірок, рис.12,(b)(суцільна лінія). Електрони і дірки дифундують нерозділеними, нейтральними парами, але у електричному полі.

Ситуація із біполярною дифузією і дрейфом розглянута у п.6.4 та отримана формула (6.41a) для статичного поля у стаціонарних умовах і малого рівня інжекції ($\Delta n \ll n_0, p_0$).

Умовою стаціонарності для ефекту Дембера є рівність 0 повного струму, який складається із дифузійних і дрейфових струмів електронів і дірок.

Формулу (6.41a) для статичного поля запишемо із урахуванням співвідношень Ейнштейна (3.61) та відношення мобільностей $b = \mu_n / \mu_p$:

$$\mathcal{E}_D = -\frac{k_B T}{e} \frac{b-1}{n_0 b + p_0} \frac{d\Delta n}{dx}. \quad (7.30)$$

Розрахуємо різницю потенціалів V_D між освітленою та протилежною, неосвітленою поверхнями напівпровідника, відстань між якими d :

$$V_D = \int_0^d \mathcal{E}_D dx = -\frac{k_B T}{e} \frac{b-1}{n_0 b + p_0} \int_{\Delta n(0)}^0 d\Delta n = \frac{k_B T}{e} \frac{b-1}{n_0 b + p_0} \Delta n(0), \quad (7.31)$$

де $\Delta n(0) = \eta \alpha_\lambda \Phi_\lambda (1-R)\tau$ – поверхнева нерівноважна концентрація.

При визначенні меж інтегрування у (7.31) прийнято, що зразок напівпровідника не лише оптично товстий, але і дифузійно товстий, тобто, $d \gg L_d$ і нерівноважні носії заряду до неосвітленої поверхні не доходять.

Із (7.31) випливає, що обов'язковою умовою реалізації ефекту Дембера є умова $b \neq 1$ і чим більше різниця мобільностей, тим більша різниця потенціалів V_D . З іншого боку, V_D обернено пропорційна тепловій рівноважній концентрації носіїв заряду. Мінімальне значення рівноважної концентрації – власна концентрація $n_0 = p_0 = n_i$. Зробимо числову оцінку V_D для власного кремнію при 300 К, скориставшись даними табл.3.3 та табл.4.2, значенням $k_B T/e$ для 300 К 0,026 В та прийнявши $\Delta n(0) = 10^{16} \text{ м}^{-3}$. Результат розрахунку такий $V_D = 5,6 \cdot 10^{-3} \text{ В}$. У більшості випадків V_D для германію і кремнію лежать у межах від десятків мікрівольтів до одиниць мілівольтів.

Відносно невелика напруга, яку можна отримати від ефекту Дембера, не дозволяє використовувати його для енергетичного перетворення світлової енергії у електричну. Ефект знаходить застосування здебільшого для визначення відношення мобільностей b , швидкості поверхневої рекомбінації і інших параметрів.

Фотомагнітний ефект. З ефектом Дембера тісно пов'язаний фотوماгнітний ефект Кікоїна – Носкова. Основні умови тіж самі – власне поглинання, оптично і дифузійно товстий зразок, але напівпровідник вміщений

ще й у магнітне поле з індукцією B_z , перпендикулярне напрямку світлового потоку $\Phi_{\lambda x}$ (рис.12,(c)).

Магнітне поле B_z впливає на паралельні дифузійні потоки електронів і дірок, направлені по вісі x , силою F_y :

$$F_{ny} = -e v_{nx} B_z, F_{py} = +e v_{px} B_z, \quad (7.32)$$

яка має протилежні знаки для електронів і дірок. Виникає ситуація, коли електронно – діркові пари розділяються не лише за рахунок різних дифузійних швидкостей уздовж вісі x , а і протилежно направленими магнітними силами уздовж вісі y .

Наслідком дії магнітної сили буде накопичення електронів і дірок на протилежних гранях по напрямку y і виникне електричне поле \mathcal{E}_y , яке надалі будемо називати фотомагнітним і позначати $\mathcal{E}_{\phi m}$.

Повний струм \vec{j} тепер відхилений від вісі x на кут θ і має j_x – та j_y – компоненти. Виникає досить складна картина *циркуляції повного струму у площині ZX*, як це показано на рис.12,(c).

Запишемо y – компоненти струмів:

$$j_{ny} = e \mu_n (n \mathcal{E}_y + B_z j_{nx} / e), j_{py} = e \mu_p (p \mathcal{E}_y + B_z j_{px} / e). \quad (7.33)$$

Із умови стаціонарності для ефекту Дембера $j = j_{nx} + j_{px} = 0$ знаходимо:

$$j_{nx} = -j_{px}, \quad (7.34)$$

а самі x – компоненти цих струмів беремо із (6.41). За умови малого рівня інжекції ($n \approx n_0, p \approx p_0$) та формули для поля Дембера (7.30), запишемо j_{px} у розгорнутому вигляді:

$$j_{px} = -e D_p \{b(n_0 + p_0)/(b n_0 + p_0)\} d \Delta n(x) / dx. \quad (7.35)$$

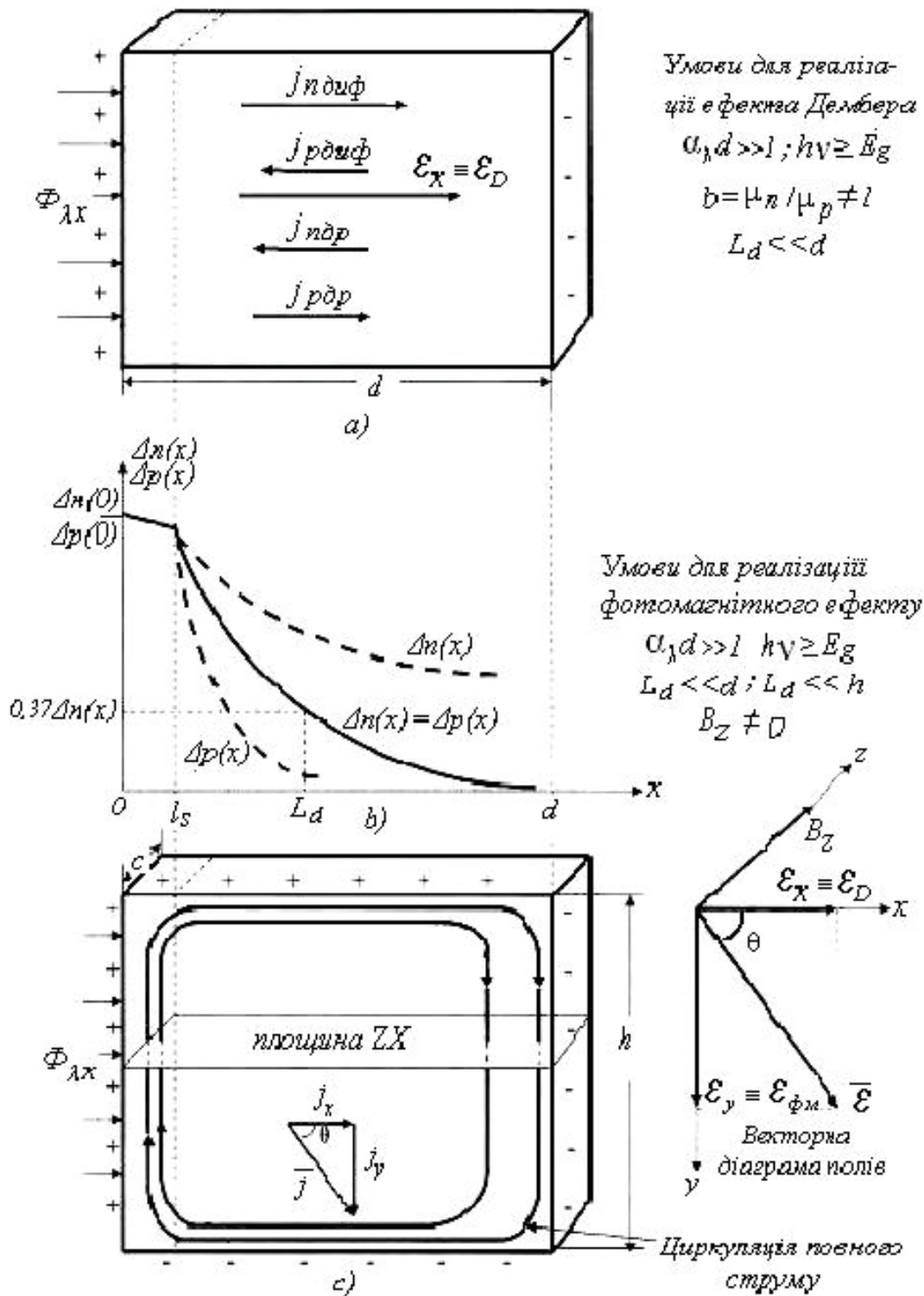


Рис.7.12. Ефекти Дембера та фотомагнітний

Введемо у (7.35) біполярний коефіцієнт дифузії (6.45) і запишемо j_{px} і j_{nx} :

$$j_{px} = -eD_b d \Delta n(x) / dx, \quad j_{nx} = eD_b d \Delta n(x) / dx. \quad (7.36)$$

Стаціонарне значення фотоманітного поля $\mathcal{E}_y \equiv \mathcal{E}_{\phi_m}$ встановиться тоді, коли його сила врівноважить магнітну силу для електронів і дірок та припинить їх рух уздовж вісі y . У площині xu повні струми \bar{j}_n та \bar{j}_p утворюють циркуляційні кола (рис.12,(с)). Тому просто прирівняти 0 повний струм $\bar{j} = \bar{j}_n + \bar{j}_p$ у таких умовах неможливо. Умовою стаціонарності для фотоманітного ефекту буде рівність 0 повного струму у площині перерізу ZX :

$$\int (j_{ny} + j_{py}) dx = 0. \quad (7.37)$$

Запишемо суму струмів, яка стоїть у дужках формули (7.37), враховуючи (7.33) і (7.36), а також те, що для струмів, які створює магнітна сила, необхідно замість звичайних мобільностей брати “холлівські” мобільності μ_{nh}, μ_{ph} (4.133):

$$j_{ny} + j_{py} = e(n_0\mu_n + p_0\mu_p)\mathcal{E}_y + e(\mu_{nh} + \mu_{ph})B_z D_b d\Delta n(x) / dx. \quad (7.38)$$

Підставимо (7.38) у (7.37) і виконаємо інтегрування у межах:

$$0 \leq x \leq L_d, \Delta n(0) \geq \Delta n(x) \geq 0.$$

Умова стаціонарності (7.37) після інтегрування має такий вигляд:

$$(n_0\mu_n + p_0\mu_p)\mathcal{E}_y L_d + (\mu_{nh} + \mu_{ph})B_z D_b \Delta n(0) = 0. \quad (7.39)$$

Розв’язуємо рівняння (7.39) відносно \mathcal{E}_y :

$$E_y = (\mu_{nh} + \mu_{ph})B_z D_b \Delta n(0) / (n_0\mu_n + p_0\mu_p)L_d. \quad (7.40)$$

Домножимо і поділимо (7.40) на $e(\mu_n + \mu_p)$ та позначимо:

$$e(n_0\mu_n + p_0\mu_p) = \sigma_0; e(\mu_n + \mu_p) \Delta n(0) = \Delta\sigma; \mu_{nh} = r\mu_n; \mu_{ph} = r\mu_p, \quad (7.41)$$

де r – холл-фактор (4.133).

Напруженість фотоманітного поля \mathcal{E}_{ϕ_m} (7.40) із урахуванням (7.41):

$$\mathcal{E}_{\phi_m} \equiv \mathcal{E}_y = r(B_z D_b / L_d)(\Delta\sigma / \sigma_0), \quad (7.42)$$

де $(\Delta\sigma / \sigma_0)$ – відносна зміна провідності при освітленні.

Знайдемо фотоманітну різницю потенціалів V_{ϕ_m} :

$$V_{\phi_m} \equiv V_y = \int_0^h \mathcal{E}_{\phi_m} dy, \quad (7.43)$$

де h – лінійний розмір зразка по вісі y . Підставляємо (7.42) у (7.43) і інтегруємо:

$$V_{\phi_m} = rB_z D_b (h/L_d) (\Delta\sigma/\sigma_0). \quad (7.44)$$

Перед тим як коментувати формулу (7.44) запишемо її у розгорнутому вигляді, розкривши $\Delta\sigma$ і σ_0 відповідно до (7.41), а також за допомогою формул

$$\Delta n(0) = \eta\alpha_\lambda \Phi_\lambda(0)\tau; \quad D_b\tau = L_d^2.$$

Після простих перетворень маємо:

$$V_{\phi_m} = rB_z D_b h L_d (b+1) \eta\alpha_\lambda \Phi_\lambda(0) / (bn_0 + p_0). \quad (7.45)$$

Висновки: – *фотомагнітна напруга* прямо пропорційна магнітній індукції і світловому потоку;

– *фотомагнітний ефект* спостерігається при будь-якому співвідношенні мобільностей b , чим суттєво відрізняється від ефекту Дембера;

– *фотомагнітна напруга* обернено пропорційна концентрації рівноважних носіїв заряду.

– за рівності інших умов, фотомагнітна напруга більша у слабо легованих і власних напівпровідниках.

У науковій літературі зустрічається ще і така назва фотомагнітного ефекту – *фото-Холл ефект*. У такий спосіб підкреслюється формальна подібність ефектів Холла і фотомагнітного за зовнішніми ознаками – виникнення поперечної різниці потенціалів V_y за наявності струму j_x і магнітного поля B_z . Принципова різниця між цими ефектами полягає у походженні струму. У ефекті Холла струм j_x – монополярний дрейфовий струм основних носіїв заряду, створений зовнішньою різницею потенціалів. Для фото-Холл ефекту струм j_x – біполярний дифузійний струм не рівноважних електронно – діркових пар, створених фотонами при власному поглинанні у оптично товстому напівпровіднику.

Бар'єрний фотоефект. У зоні контакту двох провідників із різними рівнями Фермі виникає потенціальний бар'єр як наслідок вирівнювання рівнів Фермі. Маються на увазі контакти типу метал – напівпровідник, напівпровідників n і p типів провідності т.п. Електричне поле потенціального бар'єру буде розділяти не рівноважні електронно – діркові пари, які виникають

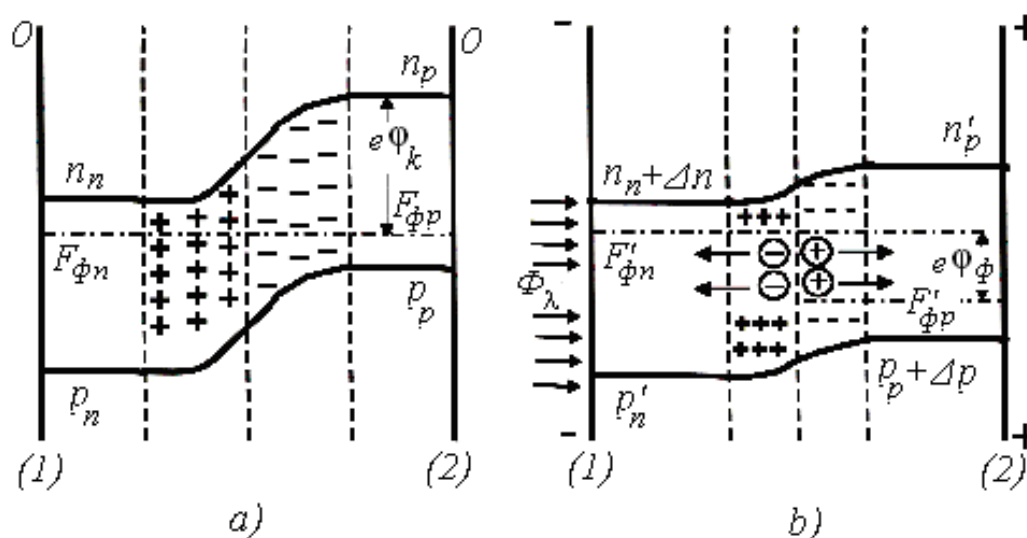


Рис.7.13. Енергетичні діаграми $p-n$ – переходу. а) у стані стаціонарної рівноваги; б) – при власному поглинанні світла. (1),(2) – зовнішні межі переходу, F_{ϕ} – рівні Фермі; F'_{ϕ} – квазірівні Фермі; Φ_{ϕ} – фотоелектрична різниця потенціалів; Φ_k – контактна різниця потенціалів (темнова різниця потенціалів)

при власному поглинанні світла безпосередньо у зоні контакту або входять у зону при дифузії.

Розглянемо фізичну модель процесів на прикладі потенціального бар'єру $p-n$ – переходу при власному поглинанні світла. Енергетична діаграма $p-n$ – переходу у стані стаціонарної рівноваги показана на рис.7.13(а). Взаємна дифузія електронів і дірок між p і n частинами переходу зупинена контактною різницею потенціалів Φ_k , яка виникла внаслідок переходу певної кількості електронів в p – і дірок в n – частини напівпровідника. Контактна різниця Φ_k має таку величину, щоб рівні Фермі були однаковими $F_{\phi n} = F_{\phi p}$, що є основною умовою стаціонарної рівноваги.

Наслідки: $-\phi_k$ –рівноважна різниця потенціалів, яка відсутня на зовнішніх контактах і не може виконувати роботу у зовнішньому колі;

– нестача певної кількості основних носіїв заряду у зоні $n - p$ контакту порушує загальну нейтралізацію зарядів донорів та акцепторів і виникає електричне поле у області просторового заряду, направлене від n до p .

Поверхня (1) електронної частини переходу (рис.7.13,(b)) опромінюється світловим потоком з енергією фотонів $h\nu \geq E_g$. Світло поглинається у оптично і дифузійно тонкому n – прошарку, товщина якого $\ll L_d$. Нерівноважні електронно – діркові пари, які утворюються при власному поглинанні світла, у області просторового заряду переходу розділяються електричним полем – електрони дрейфують проти напрямку поля, нейтралізуючи частину донорів, і створюють не рівноважний негативний заряд у n області. Освітлена поверхня (1) стає негативно зарядженою. Дірки дрейфують по напрямку поля, нейтралізуючи частину акцепторів, і створюють нерівноважний позитивний заряд у p області і на поверхні (2) (рис.7.13,(b)).

Потенціальна енергія $e\phi_k$, накопичена у просторовому заряді переходу, тепер частково трансформується у кінетичну енергію нерівноважних електронів і дірок. Потенціальний бар'єр переходу зменшиться на $e\phi_\phi$, і між контактами (1) – (2) з'явиться різниця потенціалів ϕ_ϕ .

Наслідок: *фотоелектрична різниця потенціалів* ϕ_ϕ – термодинамічно нерівноважна і може виконувати роботу у зовнішньому колі. Джерело енергії для виконання роботи – енергія фотонів, що поглинаються напівпровідником.

Зі збільшенням інтенсивності світлового потоку і нерівноважної концентрації зростає ϕ_ϕ , але одночасно зменшується поле просторового заряду внаслідок появи рухомих носіїв заряду. При достатньо високому рівні фотоінжекції ϕ_ϕ досягає рівня ϕ_k і розділення електронно – діркових пар припиняється. Проведемо відповідний аналіз цього процесу.

Скористаємося формулою (6.41), але для довільного рівня інжекції і за

таких умов:

$$\begin{aligned} p &= p_0 + \Delta n; & n &= n_0 + \Delta n; & \nabla p_0 &\neq 0, \\ \nabla n_0 &\neq 0; & \Delta n &= \Delta p; & \nabla \Delta n &= \nabla \Delta p = 0. \end{aligned} \quad (7.46)$$

Позитивний напрямок вісі x від освітленої поверхні (1) на рис. 7.13,(b), вглиб напівпровідника до поверхні (2).

Із закону діючих мас $n_0 p_0 = n_i^2$ визначимо співвідношення між градієнтами рівноважних концентрацій:

$$n_0 \frac{dp_0}{dx} + p_0 \frac{dn_0}{dx} = 0 \rightarrow \frac{dp_0}{dx} = -\frac{p_0}{n_0} \frac{dn_0}{dx}. \quad (7.47)$$

Врахуємо співвідношення Ейнштейна та (7.46), (7.47), позначимо $b = \mu_n / \mu_p$ і спочатку визначимо різницю потенціалів ϕ_k на переході без світла $\Phi_\lambda = 0$, $\Delta n = 0$:

$$\phi_k = \int \mathcal{E}_{cm} dx = \frac{k_B T}{e} \int_{n_p}^{n_n} \frac{dn_0}{n_0} = \frac{k_B T}{e} \ln \frac{n_n}{n_p}, \quad (7.48)$$

де n_n, n_p – концентрації електронів у n – та p – частинах переходу, відповідно.

Контактна (темнова) різниця потенціалів ϕ_k (7.48) має максимальне значення $\phi_{k \max}$ для певного напівпровідника за умов, що рівень Фермі $F_{\phi n}$ співпадає з дном зони провідності, а рівень $F_{\phi p}$ – зі стелею валентної зони:

$$e\phi_{k \max} \leq F_{\phi n} - F_{\phi p} \leq E_g. \quad (7.49)$$

Досягти такого суміщення рівнів Фермі із межами зон практично неможливо, оскільки завжди є певна кількість неосновних носіїв заряду p_n і n_p , які віддаляють рівні Фермі від граничних меж енергетичних зон у невідроджених напівпровідниках.

Для прикладу, у кремнії $E_g = 1,07$ еВ, а $\phi_k \approx 0,5 \dots 0,7$ В замість 1,07 В, у арсеніді галію $E_g = 1,35$ еВ, а $\phi_k \approx 1,2$ В.

По аналогії із попереднім розрахунком, визначимо φ_ϕ , врахувавши $\Delta n \neq 0$

$$\text{та } p_0 = n_i^2 / n_0: \quad \varphi_\phi = \frac{k_B T}{e} \int_{n_p}^{n_n} \frac{\Delta n (b+1)}{b n_0^2 + \Delta n (b+1) n_0 + n_i^2} d n_0. \quad (7.50)$$

У широкозонних напівпровідників ($\Delta E_0 \geq 1$ eB) $n_i \ll n_0, p_0$, (саме їх застосовують для фотоперетворювачів), тому ігноруємо доданок n_i^2 у знаменнику (7.50):

$$\varphi_\phi \approx \frac{k_B T}{e} \int_{n_p}^{n_n} \frac{\Delta n (b+1)}{b n_0^2 + \Delta n (b+1) n_0} d n_0. \quad (7.51)$$

Введемо позначення: $A = b, B = \Delta n (b+1), x = n_0$. Інтегральна частина (7.51) тепер має вигляд табличного інтеграла:

$$\varphi_\phi \approx \frac{k_B T}{e} B \int_{n_p}^{n_n} \frac{dx}{A x^2 + B x} = \frac{k_B T}{e} \ln \frac{n_n}{n_p} - \frac{k_B T}{e} \ln \frac{b n_n + (b+1) \Delta n}{b n_p + (b+1) \Delta n}. \quad (7.52)$$

Перший доданок у (7.52) – темнова різниця потенціалів φ_k (7.48), другий – залежить від рівня інжекції $\varphi = f(\Delta n)$. За умови $\Delta n = 0$ він дорівнює $\varphi = \varphi_k$, а $\varphi_\phi = 0$. Високий рівень фотогенерації $\Delta n_\infty \gg n_n, n_p$ зменшує другий доданок $\varphi \rightarrow 0$, а φ_ϕ насичується на рівні φ_k . Таким чином, фотоелектрична різниця потенціалів $\varphi_\phi = \varphi_k - \varphi$ при зміні Δn у межах $0 \leq \Delta n < \Delta n_\infty$ має значення у діапазоні:

$$0 \leq \varphi_\phi \leq \varphi_k. \quad (7.53)$$

Якщо врахувати $\varphi_{k \max}$ (7.49), то теоретичний максимум максимум $\varphi_{\phi \max \max} \leq E_g / e$ – ширина забороненої зони, визначена у вольтах.

Бар'єрний фотоефект знаходить практичне застосування у фотодіодах і сонячних фотобатареях, які є відновлюваними джерелами енергії.

Енергетичний коефіцієнт корисної дії сучасних сонячних фотобатарей сягає 14–16 %.

Резюме по п.7.5

1. Порушення нейтральності напівпровідника і поява термодинамічно нерівноважної різниці потенціалів під час власного поглинання

світла є наслідком просторового розділення електронно – діркових пар.

2. Порушення нейтральності, як наслідок розділення електронно – діркових пар, у ефекті Дембера зумовлене різницею дифузійних швидкостей електронів і дірок під час їх дифузії уздовж оптично і дифузійно товстого зразка напівпровідника. Електрони випереджають дірки ($\mu_n > \mu_p$), створюють нерівноважний негативний заряд, різницю потенціалів між освітленою і протилежною, неосвітленою поверхнями напівпровідника.
3. Відносно невелика напруга, яку можна отримати від ефекту Дембера, не дозволяє використовувати його для енергетичного перетворення світлової енергії у електричну.
4. З ефектом Дембера тісно пов'язаний фотомагнітний ефект. Розділення електронно – діркових пар відбувається не лише за рахунок різниці дифузійних параметрів носіїв заряду, але і за рахунок магнітних сил, протилежно направлених для електронів і дірок.
5. Фотомагнітний ефект спостерігається при будь-якому співвідношенні мобільностей. Максимальне значення фотомагнітна різниця потенціалів має, за рівних інших умов, для власних напівпровідників.
6. Бар'єрний фотоефект спостерігається під час появи електронно – діркових пар у зоні просторового заряду бар'єрного контакту. Електричне поле просторового заряду розділяє електрони і дірки. Нерівноважні об'ємні заряди розділених електронів і дірок зумовлюють появу фотоелектричної різниці потенціалів на зовнішніх поверхнях напівпровідника.
7. Фотоелектрична різниця потенціалів не може бути більшою темноті термодинамічно рівноважної різниці потенціалів на бар'єрному контакті. Найчастіше використовують бар'єрний контакт $p-n$ – переходу.

8. Квантово-розмірні ефекти

Залежність властивостей *твердих тіл* від їх розміру спостерігається в багатьох випадках. Наприклад, у *п'єзоелектриках* частота електромеханічного резонансу і величина діелектричної проникності залежить від розмірів кристалу. Магнітна проникність *ферромагнітних* плівок і діелектрична проникність *сегнетоелектричних* плівок залежать від їхньої товщини і т.д.

Але у найбільшій мірі **розмірний ефект проявляється для наночастинок**. Коли речовина береться в таких малих кількостях, її навіть не можна однозначно віднести до ізоляторів, провідників або напівпровідників. Наприклад, прості хімічні елементи – метали, взяті у кількості, припустимо, 20, 50 і 100 атомів, будуть послідовно проходити стадію ізолятора, напівпровідника і провідника відповідно. Справа у тому, що на відстанях 1 – 10 нм починають виявлятися **хвилові властивості електронів**.

Електроніку насамперед цікавлять електричні властивості речовини. На *електропровідність* (інакше – електроперенос заряду, або електричний транспорт) нанорозмірних структур впливає ефект **квантування** енергетичного спектру електронів, а на самих малих відстанях починають позначатися їх хвилові властивості. У твердому тілі довжина когерентності електронної хвилі при звичайній температурі складає величину декілька нанометрів. Розмірний ефект виникає, якщо довжина тіла, принаймні в одному вимірі, стає порівнянної з деякою критичною величиною $l_{\text{кр}}$. Для **класичних** розмірних ефектів $l_{\text{кр}}$ – це класична величина, наприклад дифузійна довжина, або довжина вільного пробігу електронів і т.д.

Треба відзначити, що **квантові** розмірні ефекти у нанорозмірних електронних структурах спостерігаються тоді, коли роль критичної довжини $l_{\text{кр}}$ грає **довжина хвилі де Бройля λ для електронів**, тобто коли розмір структури хоча б в одному вимірі має порядок λ . Таким чином, квантово-розмірні ефекти обумовлені **хвильовою природою електронів**.

У нанорозмірних областях поведінка електронів визначається відбиттям електронних хвиль від границь розділу таких областей, а також інтерференцією електронних хвиль або проходженням хвиль крізь потенціальні бар'єри, квантуванням енергії електронів, просторово обмежених у своїх переміщеннях, проходженням електронів крізь нанометрові діелектричні прошарки, квантуванням електроопору квантових ниток (проводів) і ін.

Новими фундаментальними явищами у нанорозмірних структурах є

квантове обмеження, тунелювання, балістичний транспорт і квантова інтерференція. Фізична природа всі цих ефектів описується квантовою механікою.

Квантове обмеження виникає тоді, коли рух електронів хоча б в одному напрямку стає обмеженим потенціальними бар'єрами, обумовленими наноструктурою. Таке обмеження впливає на спектр дозволених станів електронів і обумовлює їх рух у наноструктурі. Цей рух змінюється як у напрямку, перпендикулярному бар'єрам, так і в напрямках паралельних до бар'єрів.

Перенесення заряду *перпендикулярно* бар'єрам можливий, головним чином, шляхом *тунелювання*, яке забезпечує перехід носіїв заряду з одної ділянки наноелектронного приладу до іншої. При русі носіїв заряду *вздовж* потенціальних бар'єрів стає можливим квантова інтерференція і *балістичний транспорт* електронів.

8.1. Хвилеві властивості та енергетичний спектр електронів.

В основу квантової механіки покладений той факт, що світло має подвійну (*корпускулярно-хвильову*) природу, при чому модуль імпульсу фотона p може бути виражений через відповідну довжину хвилі λ :

$$p = h/\lambda$$

де $\lambda = c/\nu$, причому c – швидкість світла, а частота ν знаходиться для даного кванта світла з енергією E з формули Планка:

$$E = h\nu$$

Постійна Планка дорівнює: $h = 2\pi\hbar = 6,6 \cdot 10^{-27}$ ерг·с = $6,6 \cdot 10^{-34}$ Дж·с. Фізична розмірність h означає «момент кількості руху» = енергія×час = довжина×імпульс. **Момент кількості руху** – фізична величина, що має розмірність добутку енергії на час ($E \cdot t$). Вона є однією з найважливіших динамічних характеристик руху частинок і їх систем – як у класичній, так і у квантовій механіці. Саме тому постійна Планка h (або $\hbar = h/2\pi$) відноситься до фундаментальних фізичних констант так само як і постійна Больцмана k_B , як заряд e і маса m_e електрона.

Постійна Планка була прийнята фізикою при відмові від класичних уявлень – разом із квантовою (хвильовою) механікою. При цьому формальний перехід від квантової механіки до класичної механіки Ньютона завжди можна зробити, якщо покласти $h = 0$. Постійна Планка h зв'язана із широким колом

фізичних явищ у мікросвіті, для яких істотна квантованість величин.

У квантовій механіці формулу $p = h/\lambda$ звичайно записують у векторній формі $\vec{p} = \hbar \vec{k}$, де модуль $k = 2\pi/\lambda$.

Величина \vec{k} , що називається **хвильовим вектором**, широко використовується у квантовій механіці. Напрямок хвильового вектора фотона збігається з напрямком вектора імпульсу фотона, тобто хвильовий вектор спрямований убік поширення світлової хвилі. Модуль хвильового вектора називають також **хвильовим числом**.

Де Бройль встановив, що хвильові властивості притаманні не лише фотонам, але і будь-яким мікрочастинкам речовини. Для наноелектроніки дуже важливо, що електрон теж має хвильову природу.

Довжина хвилі, яка відповідає будь-якій частинці, називається **хвилею де Бройля** і визначається формулою: $\lambda = h/p = h/(mv)$, де m – маса частинки, v – її швидкість.

Хвильові властивості мікрочастинки унеможливають опис її поведінки одночасним завданням точних значень координат і швидкостей. Цей факт математично описується *співвідношенням невизначеностей Гейзенберга*:

$$\Delta p_x \Delta x \geq h/2;$$

$$\Delta p_y \Delta y \geq h/2;$$

$$\Delta p_z \Delta z \geq h/2;$$

де Δp_x , Δp_y і Δp_z – невизначеність компонентів імпульсу частинки по вісях x , y і z відповідно, а Δx , Δy і Δz – невизначеність значень координат частинки в той же момент часу.

Цікаво відзначити, що для комбінацій типу $\Delta p_x \Delta y$, $\Delta p_y \Delta x$, $\Delta p_z \Delta y$ і т.п. співвідношення невизначеностей не діє – невизначеності значень не спряжених між собою координат і компонентів імпульсу можуть незалежно мати будь-які значення. Ця особливість співвідношення невизначеностей набуває неабиякого значення при аналізі поведінки електронів у нанорозмірних структурах.

Співвідношення невизначеностей має місце також для енергетичних і часових змінних:

$$\Delta E \cdot \Delta t \geq h/2,$$

де ΔE – невизначеність енергії системи в розглянутому квантовому стані, а Δt – час перебування системи в цьому стані.

Співвідношення невизначеностей відноситься до фундаментальних

положень фізики, воно має всебічне експериментальне підтвердження.

Довжина хвилі де Бройля λ для **електрона**, що має ефективну масу m^* і рухається в кристалі зі швидкістю v має значення:

$$\lambda = \hbar/p = \hbar/m^*v = h \cdot (2m^*E_{\text{кін}})^{-1/2},$$

де $p = m^*v$ – імпульс і $E_{\text{кін}}$ – кінетична енергія. Таким чином, довжина хвилі де Бройля тим більша, чим **менша ефективна маса електрона**.

У більшості **металів** ефективна маса електронів майже така ж сама, як і у вакуумі: $m^* \sim m_e$, так що для металів довжина хвилі де Бройля електрона мала: $\lambda \leq 1$ нм. Ця величина близька до сталої кристалічної ґратки. Тому у металах при нормальних умовах (температурі близької до 300 К) практично неможливо спостерігати квантові нанорозмірні ефекти – хіба що за дуже низької температури, коли довжина хвилі де Бройля стає приблизно у 10 разів більшою у порівнянні із значенням за нормальних умов.

Однак у **напівпровідниках** ефективна маса електрона менша ніж у металах. Так, при температурі 4 К для кремнію $m^* = 0,2m_e$ і $\lambda \approx 500$ нм, а в арсеніді галію $m^* = 0,07m_e$ і $\lambda \approx 5000$ нм. При кімнатних температурах (близько 300 К) для кремнію $m^* = 0,9m_e$ і відповідно $\lambda = 8$ нм; для арсеніду галію $m^* = 0,07m_e$ і $\lambda = 30$ нм.

Таким чином, саме у напівпровідниках довжина хвилі де Бройля має нанорозміри, і у зв'язку з цим квантоворозмірні ефекти технологічно легше здійснити у напівпровідниках.

Енергетичний спектр електронів $E(k)$ і густина їх квантових станів $N(E)$ – найважливіші характеристики квантового об'єкта, що визначають його електронні властивості і реакцію на зовнішні впливи. Енергетичний спектр – це сукупність усіх можливих значень енергії частинки у визначених умовах.

Енергетичний спектр характеризує залежність енергії частинки від її імпульсу $E(p)$ або, що то ж саме, залежність енергії від хвильового вектора $E(k)$, оскільки імпульс і хвильовий вектор зв'язані простим співвідношенням: $p = \hbar k$ (модуль хвильового вектора визначається зворотною довжиною хвилі: $k = 2\pi/\lambda$). Тому:

$$E(k) = \hbar^2 k^2 / 2m^*.$$

Енергетичний спектр являє собою сукупність можливих значень енергії частинки в даних умовах. Якщо енергія квантується, то енергетичний спектр

називається *дискретним (квантовим)*; якщо ж енергія може приймати неперервний ряд значень, то спектр називається *суцільним (неперервним)*. Спектр вільного електрону у вакуумі – суцільний, а в кристалі – дискретний, хоч і із дуже високою густиною станів. Спектр електрона у наночастинці теж дискретний, але густина станів розріджена.

Густина станів $N(E)$ визначає число квантових станів електронів у одиниці об'єму, площі або довжини (у залежності від розмірності об'єкта), віднесене до одиничного інтервалу енергій. Відповідно до цього визначення, щільність станів дорівнює $dN(E)$ – числу можливих станів в інтервалі енергій від E до $E + dE$.

$$N(E) = dN(E)/dE.$$

Знання густина станів $N(E)$ і імовірності їхнього заповнення електронами $f(E)$ дозволяє встановити розподіл електронів розглянутої системи по квантових станах і описати електричні, оптичні і деякі інші властивості системи. Електрони характеризуються напівцілим спином. Тому імовірність заповнення ними квантових станів визначається статистикою Фермі – Дірака, яка базується на принципові Паулі.

Порівняння впливу різної структури доцільно починати з добре дослідженого випадку – звичайного тривимірного кристалу.

8.2. Кристал – тривимірна (3D) структура

«Необмежений» у всіх напрямках кристал – макроскопічний об'єкт – характеризується далеким упорядкуванням атомів. Зазначимо, що з погляду нанофізики навіть кристалик, що має мікронні розміри, уже виглядає «необмеженим», оскільки у всіх напрямках його регулярна кристалічна ґратка складається з тисяч атомів, що в багато разів перевищує довжину хвилі де Бройля.

При дослідженні електронного транспорту важливою моделлю є то, що у металах і напівпровідниках кристал як би наповнений об'ємним (тривимірним) **3D-електронним газом**, у якому електрони можуть вільно переміщуватися в будь-якому напрямку.

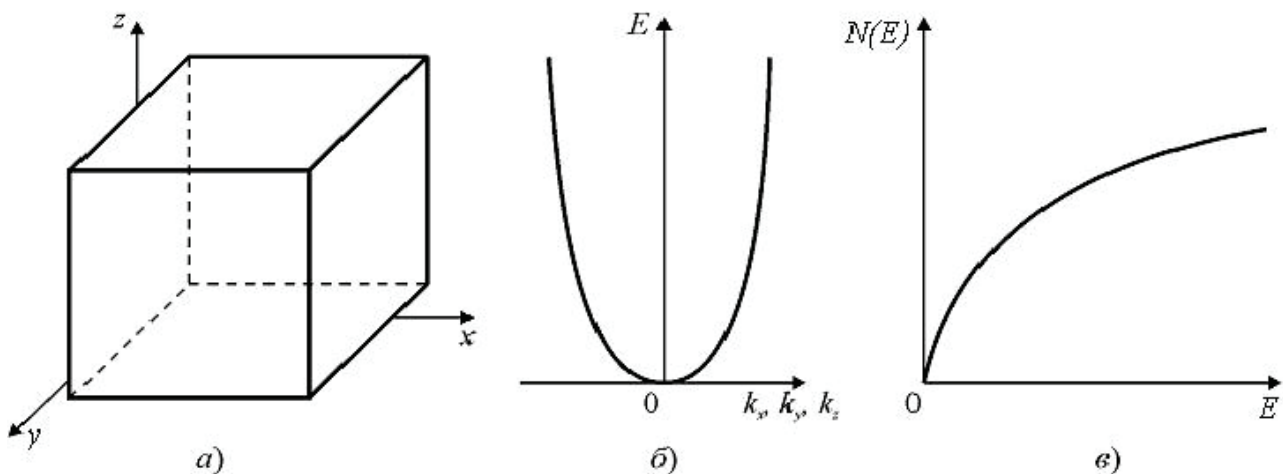


Рис. 8.1. Електронний спектр необмеженого по розмірах кристала: *a* – зразок кристала, наприклад, кубічний; *б* - залежність енергії електрона від складових квазі-хвильового вектора; *в* - залежність густина квантових станів ($g(E)$) від енергії електронів (E).

Енергія електрона в 3D-газі:

$$E = (m^*v^2)/2 = p^2/2m^* = (p_x^2 + p_y^2 + p_z^2)/2m^* = \hbar^2 (k_x^2 + k_y^2 + k_z^2)/2m^*$$

де p і k – квазіімпульс і квазіхвильовий вектор електрона відповідно, а m^* – його ефективна маса (для спрощення задачі вона вважається ізотропною).

У межах зони провідності енергія електрона може приймати практично неперервний ряд значень. Спектральна залежність $E(k)$ виглядає як квазінеперервна, оскільки рівнів енергії у зоні стільки, скільки електронів утримується в 3D кристалі (для напівпровідника $10^{14} - 10^{20} \text{ см}^{-3}$, для метала близько 10^{22} см^{-3}). Відповідно, відстань між сусідніми рівнями енергії в зоні провідності дуже мала: вона дорівнює близько $10^{-14} - 10^{-22} \text{ еВ}$. У порівнянні з енергією теплового руху 0,026 еВ за температури 300 К спектр $E(k)$ дійсно виглядає суцільним.

Залежність енергії від компонентів хвильового вектора k_x, k_y, k_z поблизу дна зони провідності параболічна (рис. 8.1, б).

Густина квантових станів електронів $N(E)$ показана на рис. 8.1, в; для необмеженого по розмірах (3D) кристала цей параметр також описується параболою: $N(E) \sim E^{1/2}$. Поблизу дна зони провідності (відкіля починається відлік), густина енергетичних станів електронів невелика, але вона плавно збільшується з ростом E , досягаючи насичення.

З рис. 8.1, в видно, що в межах однієї зони функції $E(k)$ і $N(E)$ квазінеперервні, і тому електричні властивості 3D кристалів під впливом полів змінюються неперервно.

Тривимірний електронний газ у кристалах вивчається як у фізиці металів, так і у фізиці напівпровідників. Він обумовлює різноманітні електрофізичні властивості цих матеріалів у макроскопічному (тривимірному) стані.

8.3. Квантова яма – двовимірна (2D) наноструктура

Двовимірний (2D) об'єкт – це тонкий шар кристалу, товщина якого d співрозмірна з довжиною хвилі де Бройля ($d \approx \lambda$). В різних роботах квантову яму називають «квантовою плівкою», «квантовим колодязем», тощо.

Система електронів у тонкому шарі називається двовимірним (або 2D) електронним газом. Фрагмент такого шару представлений на рис. 8.2, а. Рух електронів у цьому шарі обмежений відрізком d_x у напрямку x і не обмежений у напрямках y і z .

Рухаючись у напрямку x , електрон не здатний залишити тонкий шар, тому що його робота виходу (яка дорівнює, наприклад, в арсеніді алюмінію 4,5 еВ) набагато більша енергії теплового руху (близько 0,026 еВ при кімнатній температурі). Тому рух електрона у напрямку x правомірно розглядати як рух в одновимірній досить глибокій прямокутній потенційній ямі шириною d_x , рис. 8.2, а. Енергія такого руху E_n квантується і характеризується квантовими числами $n = 1, 2, 3, \dots$. Якби яма була нескінченно глибока, то на її ширині d_x укладалося би ціле число півхвиль $\lambda_n/2$, тобто $d_x = n\lambda_n/2$. Це означає, що стаціонарними були б тільки такі стани руху електрона, яким відповідала б стояча хвиля, утворена падаючою і відбитою від стінок ями хвилями де Бройля (пунктирні криві на рис. 8.2). Величини E_n називаються квантово-розмірними рівнями.

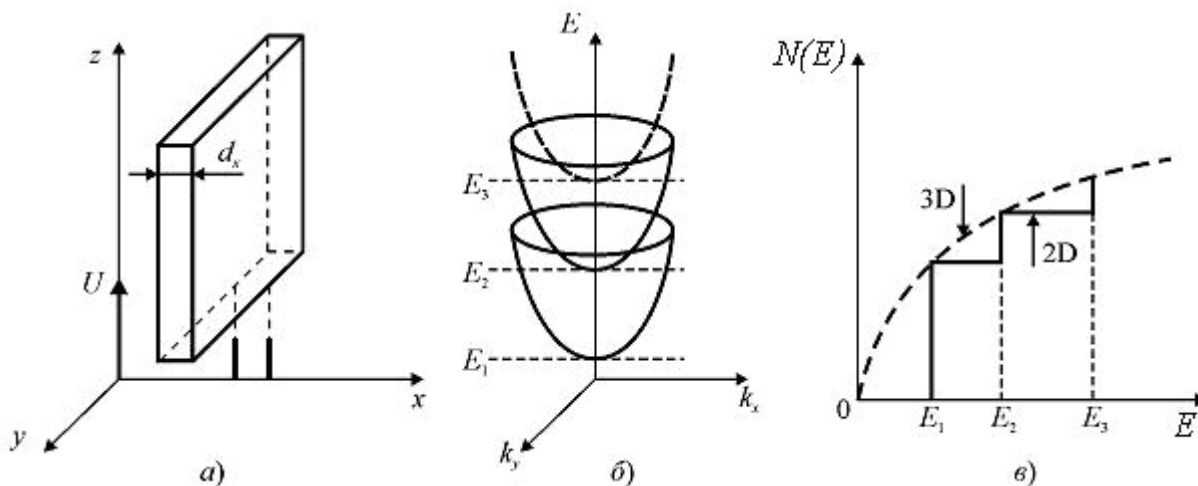


Рис. 8.2. Двовимірний (2D) нанооб'єкт (тонкий шар): а – схематичне представлення тонкого шару, обмеженого уздовж осі x ; б – енергетичний спектр електронів у тонкому шарі; в – залежність густини станів від енергії у випадку нескінченно глибокої квантової ями.

Енергія рухів уздовж осей y і z не квантується і визначається такими ж

квазінеперервними рівнями, як для вільної частинки або для об'єму. Тому повну енергію електрона за умови, що величина ефективної маси електрона m^* однакова для рухів у всіх напрямках, можна представити у вигляді:

$$E = \hbar^2(k_y^2 + k_z^2)/2m^* + E_n = \hbar^2(k_y^2 + k_z^2)/2m^* + \hbar^2\pi^2 n^2/2m^* d_x^2,$$

де квантове число $n = 1, 2, 3, \dots$

Отже, енергетичний спектр електрона у квантовій ямі двовимірного нанооб'єкту буде дискретно-неперервним. Кожному розмірному рівневі E_n відповідає безліч можливих значень енергії за рахунок вільного руху електрона уздовж осей y і z . Ця сукупність енергій E називається *двовимірною підзоною розмірного квантування*. Залежності $E(k_y, k_z)$ і $N(E)$ приведені на рис. 8.2, б, в.

Графік залежності $E(k_y, k_z)$ являє собою систему параболоїдів; дно n -го параболоїда відповідає рівневі $E = E_n$. Залежність $N(E)$ має східчастий характер. Кожна розмірна підзона вносить у величину густину станів однаковий внесок, рівний $m^*/\pi\hbar^2$.

Модель нескінченно глибокої прямокутної потенційної ями справедлива тільки для руху електрона в «ізольованій» тонкій плівці при нанорозмірній товщині d . Окрему, довільної довжини плоско-паралельну плівку нанометрової товщини навряд чи можливо реалізувати. На практиці в нанорозмірних структурах квантові ями – нанопрошарки – створюються *всередині кристала*, так щоб рух носіїв заряду був обмеженим лише в одному вимірі; тому і можна вважати, що ці носії знаходяться в одновимірній потенціальній ямі. Прикладом можуть служити електрони в нанометровому прошарку *вузькозонного* напівпровідника між двома шарами *широкозонного* напівпровідника, рис. 8.3.

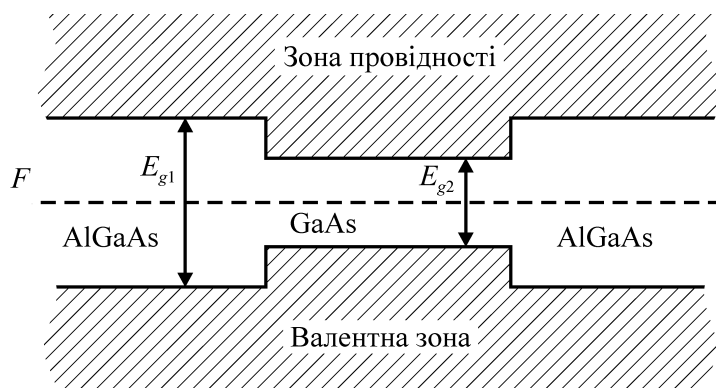


Рис. 8.3. Енергетичний спектр квантової ями – тонкого прошарку порівняно вузькозонного напівпровідника GaAs між шарами широкозонного напівпровідника AlGaAs.

Першою експериментально реалізованою і дослідженою квантовою ямою

була нанорозмірна структура на основі гетеропереходів між шарами легованого донорами напівпровідника $n\text{-Al}_x\text{Ga}_{1-x}\text{As}$ і тонким шаром власного арсеніду галію $i\text{-GaAs}$. Енергетична щілина E_{g1} (заборонена зона) твердого розчину $\text{Al}_x\text{Ga}_{1-x}\text{As}$ більша, ніж заборонена зона E_{g2} арсеніду галію. Тому спорідненість до електронів у двох контактуючих напівпровідників різна. Дно зони провідності $n\text{-Al}_x\text{Ga}_{1-x}\text{As}$ і дуже близький до нього донорний рівень розташовані вище по шкалі енергій за дно зони провідності GaAs . Тому електрони з донорного рівня $n\text{-Al}_x\text{Ga}_{1-x}\text{As}$ перетікають у зону провідності GaAs і на гетеропереходах з боку більш вузькозонного $i\text{-GaAs}$ виникають $2D$ шари, збагачені електронами (так само у $\text{Al}_x\text{Ga}_{1-x}\text{As}$ виникають збідненні шари).

Таким чином, на плоских поверхнях розділу з боку GaAs утворюється двовимірний ($2D$) електронний газ, оскільки електрони переходять через бар'єр і накопичуються у більш вузькозонному напівпровіднику.

Квантові ями (або квантові прошарки) – найбільш досліджені напівпровідникові наноструктури. Саме в них спостерігаються такі унікальні фізичні явища, як квантовий ефект Холла і висока мобільність електронів – балістичний транспорт носіїв заряду (див. розділ 10.4).

Практичними прикладами квантових ям з $2D$ -електронним газом можуть бути електропровідні канали в уніполярних транзисторах (метал-оксид-напівпровідник структури на кремнії) і вузькозонні прошарки в гетероструктурах зі сполук $\text{A}^{\text{III}}\text{B}^{\text{V}}$ (що необхідно для інжекційних лазерів). Системи близько розташованих рівнорозмірних квантових ям, між якими можливо тунелювання електронів, складають *надґратки* – гетероструктури, виготовлені з двох або більше різних матеріалів (розглянуті далі в розділі 10.3). У цих гетероструктурах важлива роль належить перехідному шарові, тобто межах розділу двох матеріалів.

Всі елементи, що входять до складу матеріалів, з яких роблять напівпровідникові гетероструктури (Zn , Cd , Hg , Al , Ga , In , Si , Ge , P , As , Sb , S , Se , Ti), розташовані в центральній частині періодичної таблиці елементів і належать до II – VI груп. У середині розташований **кремній**, що у технології електронних матеріалів займає таке ж важливе місце, як сталь у виробництві конструкційних матеріалів. Крім кремнію електроніка часто має справу з напівпровідниковими сполуками $\text{A}^{\text{III}}\text{B}^{\text{V}}$ і їх твердими розчинами, а також зі сполуками $\text{A}^{\text{II}}\text{B}^{\text{VI}}$. Зі сполук типу $\text{A}^{\text{III}}\text{B}^{\text{V}}$ найчастіше використовується арсенід галію GaAs , а з твердих розчинів – $\text{Al}_x\text{Ga}_{1-x}\text{As}$. Використання таких твердих розчинів дозволяє створювати гетероструктури з безперервною, а не

стрибоподібною зміною відносного вмісту елементів III групи. У таких гетероструктурах ширина забороненої зони теж міняється безупинно.

Для виготовлення гетероструктур дуже важливим є узгодження параметрів кристалічних ґраток двох контактуючих матеріалів. Якщо два матеріали, що мають різні постійні кристалічних ґраток, вирощуються один на одному, то при збільшенні товщини шарів у границі розділу з'являються великі механічні напруги і деформації, які здатні викликати мікротріщини у одному з шарів.

Механічні напруги і деформації з'являються незалежно від того, чи є перехід між двома шарами плавним чи ні. Для зменшення деформацій постійні ґраток двох матеріалів повинні відрізнятися якнайменше. Саме тому при вивченні гетероструктур найчастіше використовують тверді розчини системи AlAs-GaAs, оскільки арсеніди алюмінію і галію мають майже однакові параметри ґраток. У цьому випадку GaAs у вигляді монокристалів є ідеальною підкладкою для росту гетероструктур. Іншою природною підкладкою є фосфід індію InP, що застосовується в комбінації з твердими розчинами GaAs-InAs, AlAs-AlSb і іншими напівпровідниками типу $A^{III}B^V$.

Прогрес у створенні тонкошарових гетероструктур був обумовлений появою практичних технологій росту тонких шарів методами молекулярно-променевої епітаксії, газофазної епітаксії з металоорганічних сполук і рідиннофазної епітаксії. З'явилася можливість вирощувати гетероструктури з дуже різкою границею розділу.

8.4. Квантова нитка – одновимірна (1D) наноструктура

Квантові нитки (або квантовий проводи, квантові шнури) являють собою такий об'єкт, у якому рух електронів обмежений у *двох напрямках* (наприклад, уздовж вісей x і z), але не обмежений уздовж однієї вісі, наприклад, вздовж вісі y (рис. 8.4,а). Відповідно, розмірне квантування руху носіїв заряду відбувається у двох напрямках – x і z , але у напрямку y рух електронів не квантується (залишається вільним).

Квадратний перетин квантового дроту $d_x \times d_z$ – це лише спрощена модель для розрахунку; реальний профіль квантової нитки може бути й іншим, ніж це зображено на модельному рисунку. Важливо лише те, що потенціальна яма для вільних електронів у квантовій нитці *двохвимірна*.

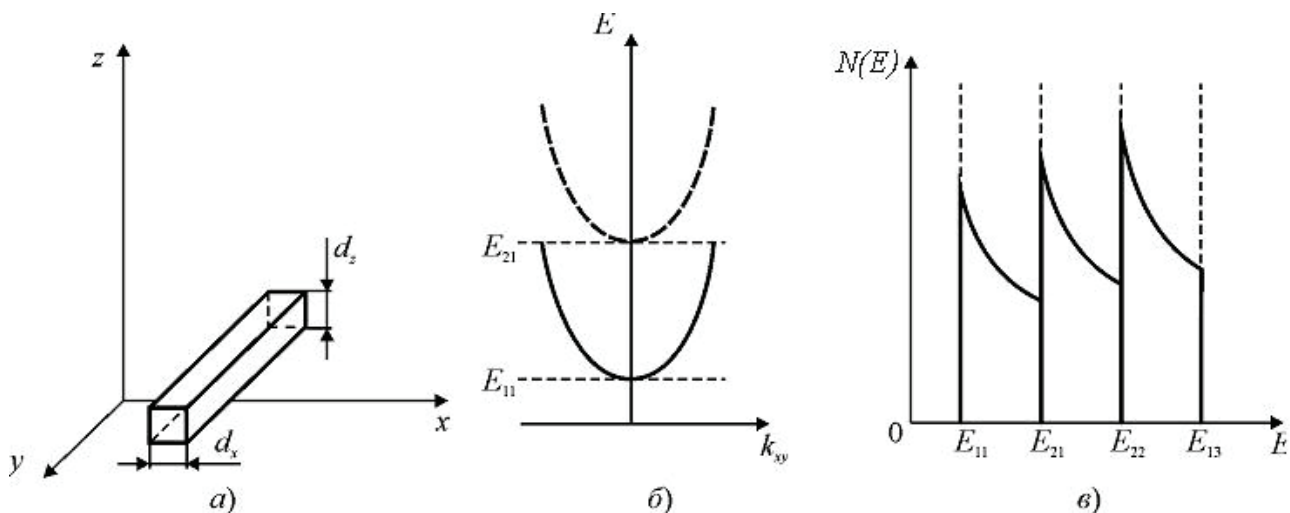


Рис 8.4. Одновимірний (1D) нанооб'єкт – квантова нитка: *а* – модельне представлення; *б* – залежність енергії від величини хвильового вектора; *в* – залежність густини станів електронів від енергії.

Енергія електрона, що зв'язана з його рухом уздовж осей y і z , повинна квантуватися так само, як і в одномірних потенціальних ямах шириною dy і dz . Повна енергія електрона дорівнює:

$$E = \hbar^2 k_y^2 / 2m^* + E_{mn} = \hbar^2 k_y^2 / 2m^* + \hbar^2 \pi^2 n^2 / 2m^* d_x^2 + \hbar^2 \pi^2 m^2 / 2m^* d_z^2,$$

де $m, n = 1, 2, 3, \dots$ і E_{mn} – енергія розмірних рівнів. Положення кожного з них залежить від двох квантових чисел m і n , а також від величин d_x, d_z , причому в напрямку протягу дроту (y) спектр є неперервним.

Зона провідності в квантовому дроті розбивається на **одновимірні підзони** (рис. 8.4, б). Щільність станів на одиницю довжини $N(E)$ має ряд різких піків (рис. 8.4, в), що відповідають розмірним рівням. Це означає, що більшість електронів у підзоні має енергію поблизу відповідного розмірного рівня.

Більшість способів виготовлення квантових ниток ґрунтуються на тому, що в системі з двовимірним електронним газом (як правило, на основі гетероструктури) тим або іншим технологічним способом обмежується рух електронів ще в одному з напрямків.

Провідність квантових ниток може мати важливі особливості. Найбільш цікавий випадок **коротких ниток** з довжиною, меншою довжини вільного пробігу електронів (остання визначається розсіюванням на домішках і дефектах нитки). Такі структури проказані на рис. 8.5. При цьому електрон, вилетівши з одного контакту квантової нитки, долітає до іншого контакту без зіткнень, як снаряд, випущений з гармати.

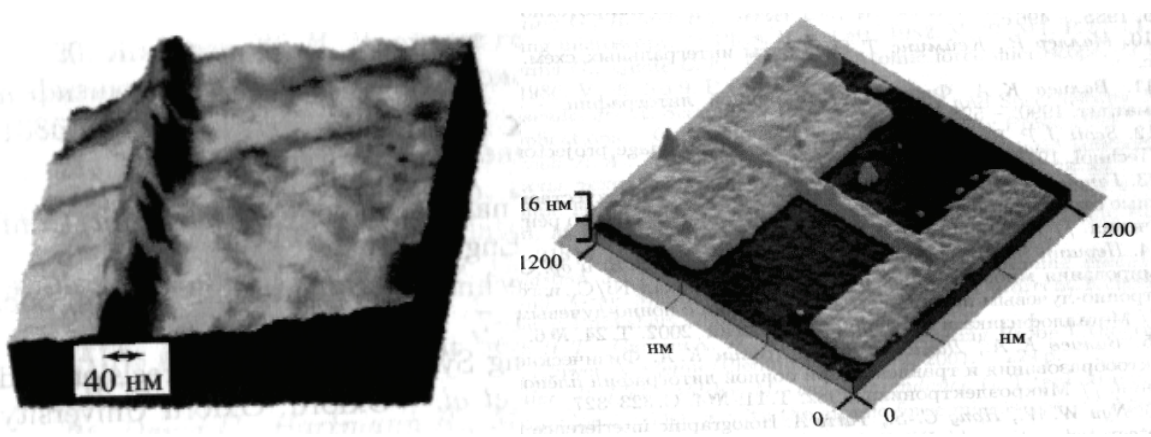


Рис. 8.5. Квантові нитки з різних напівпровідників

Така аналогія привела до того, що відомі наноструктури (котрі часто називають *балістичними*) де практично відсутнє зіткнення електронів. Однак балістичний транспорт заряду у квантових нитках не квантується (як у випадку 2D-електронного газу, п.8.3).

8.5. Квантова точка – нульвимірна (0D) наноструктура

Існує також і важливий нульвимірний об'єкт – квантова точка, яка символічно зображена на рис. 8.6,а. Рух електронів для такого об'єкту обмежений у всіх трьох напрямках точки, форма якої для спрощення позначена як куб об'ємом $d_x \times d_y \times d_z$.

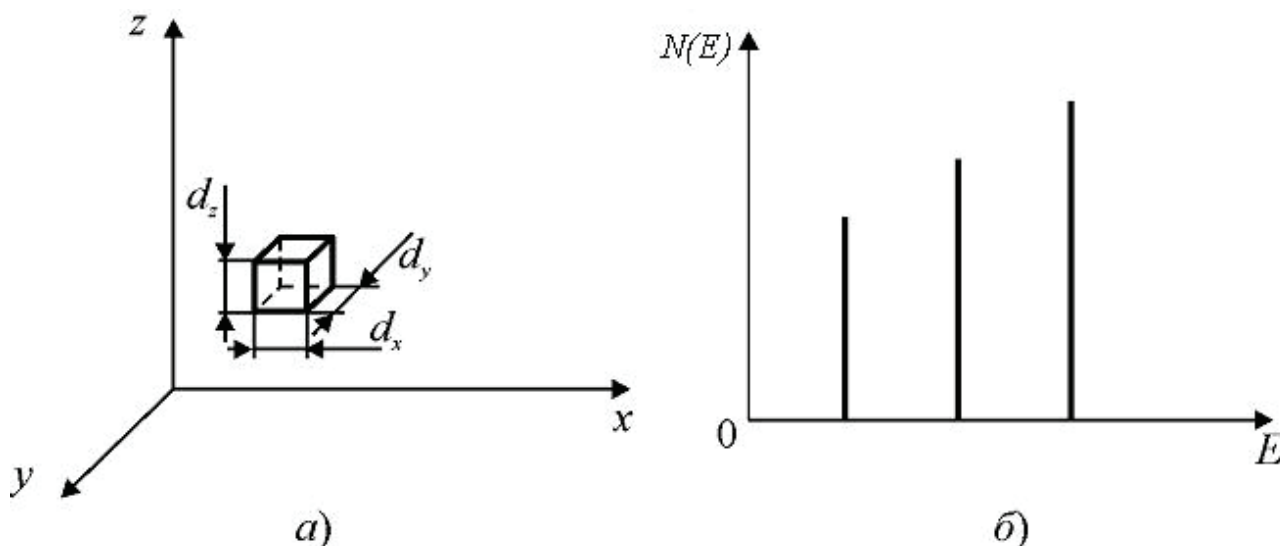


Рис. 8.6. Нуль-вимірний (0D) нанооб'єкт – квантова точка:
а – модельне представлення квантової точки
б - залежність густини станів від енергії.

Потенціальна яма для квантової точки *тривимірна*. Енергія вільних

електронів повинна квантуватися для їх рухів у всіх трьох вимірах. Енергетичний спектр електронів у квантовій точці є цілком дискретним, як і в окремому атомі. Енергія визначається таким чином:

$$E = \hbar^2 k_x^2 l^2 / 2m^* d_x + \hbar^2 \pi^2 n^2 / 2m^* d_y^2 + \hbar^2 \pi^2 m^2 / 2m^* d_z^2,$$

де $l, m, n = 1, 2, 3, \dots$; d_x, d_y, d_z – розміри області в трьох вимірах, рис. 8.6, а.

Енергетичний спектр електронів складається з окремих розмірних рівнів E_{lmn} , що нагадує спектр ізольованого атома. Величина енергії E_{lmn} залежить від трьох квантових чисел l, m, n і розмірів d_x, d_y, d_z . Графік густини станів $N(E)$ у квантовій точці має так називаний δ -подібний вид:

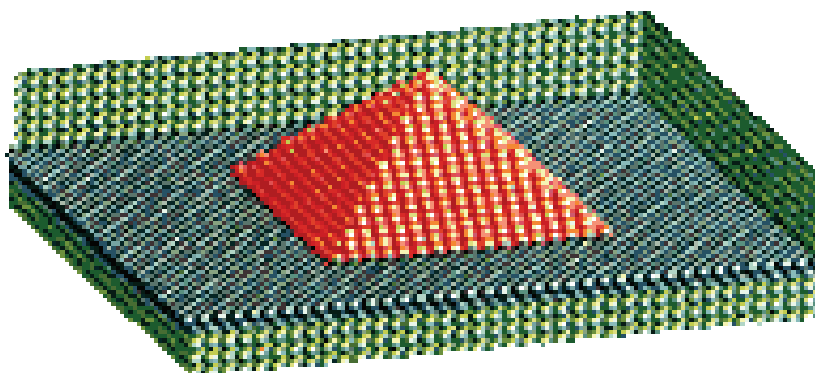
$$N(E) = \infty,$$

якщо $E = E_{lmn}$ (E збігається з розмірним рівнем) і

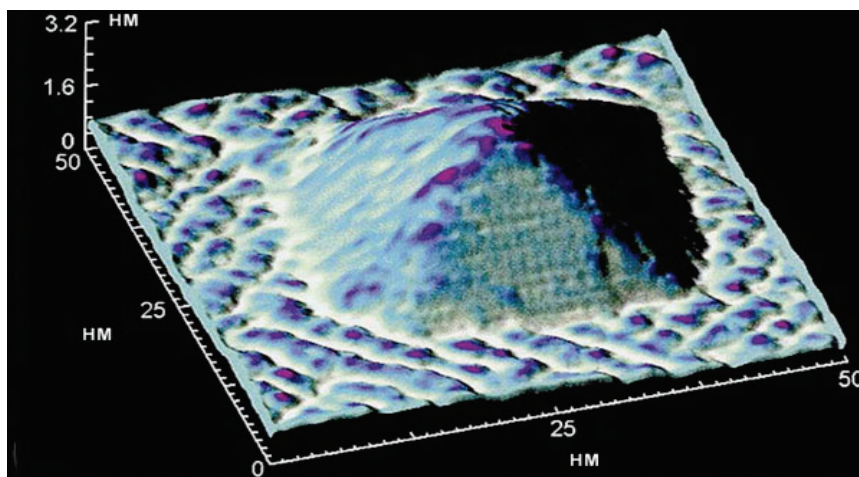
$$N(E) = 0,$$

якщо $E \neq E_{lmn}$ (тобто, E лежить у проміжку між розмірними рівнями). Графік заснований на δ -функціях – ідеалізований випадок. Насправді недосконалість границь квантової точки та її структури приводить до того, що піки густини станів розмиваються. Все ж таки залежність $N(E)$ для квантових точок характеризується **низкою вузьких максимумів**.

Прикладом квантових точок можуть служити нанокристалики одного матеріалу на поверхні зростаючого епітаксialного шару іншого матеріалу. Комп'ютерне моделювання квантової точки – острівця з атомів Ge, вирощеного на поверхні Si – а також її фактичне зображення, отримане методом скануючого електронного мікроскопу, показане на рис. 8.7.



а



б

Рис. 8.9. Об'ємне зображення квантової точки – пірамідки з атомів германію, спонтанно сформованій на кремнієвій підкладці: *а* – комп'ютерна модель, *б* – зображення зі скануючого електронного мікроскопу.

Історично першими квантовими точками були мікрокристали селеніду кадмію CdSe. Електрон у такому мікрокристалі виявляється в тривимірній потенціальній ямі, так що він має кілька стаціонарних рівнів енергії з характерною відстанню між ними (точне енергії рівнів залежить від форми і розміру квантової точки).

При переході між енергетичними рівнями квантової точки може випромінюватися фотон – аналогічно переходові електрона між рівнями енергії атома. Можна також збудити («закинути») електрон на самий високий енергетичний рівень точки, а випромінювання одержати від переходу між більш низько розташованими рівнями (люмінесценція). При цьому, на відміну від реальних атомів, частотами переходів легко керувати, змінюючи розміри наночастинки. Саме спостереження люмінесценції кристалів селеніду кадмію з частотою люмінесценції, обумовленої розміром кристала, послужило першим спостереженням квантових точок.

В даний час відомо багато експериментів, присвячених квантовим точкам, сформованим у двовимірному електронному газі. У двовимірному електронному газі рух електронів перпендикулярно площині є обмеженим, а область на площині можна виділити за допомогою обмежуючих металевих електродів, що накладаються на гетероструктуру зверху. Квантові точки в двовимірному електронному газі можна зв'язати тунельними контактами з іншими областями двовимірного газу і вивчати електропровідність через квантову точку. У такій системі спостерігається явище розглянутої далі кулонівської блокади.

Застосування квантових точок у приладах оптоелектроніки.

Найбільш широке сучасне практичне застосування квантових точок – це лазери на масивах квантових точках, створені на основі сполук $A^{III}B^V$. Експериментально підтверджені теоретично очікувані переваги таких лазерів: низький граничний струм (близько 10 А/см^2), великий коефіцієнт підсилення, температурна стабільність граничного струму.

Відомо, що лазери на квантових точках особливо важливі для телекомунікацій. Найбільш бажаний на практиці діапазон максимальної прозорості оптоволокна – довжини хвиль $1,3 - 1,5 \text{ мкм}$. В той же час давно розроблені лазери на квантових ямах на основі GaAs випромінюють на довжинах хвиль $0,85 - 0,95 \text{ мкм}$.

Необхідне для оптоволокна випромінювання дають лазери на основі складної InGaAsP-InP. Але й вони мають недоліки: низьку температурну стабільність і складну технологію виготовлення. Крім того, на підкладках InP не можна одержувати вертикально випромінюючі лазери, які мають значні переваги (лазери з вертикальним резонатором виготовляють тільки на підкладках GaAs).

Саме на довжині хвилі $1,3 \text{ мкм}$ (тобто у масимумі прозорості оптоволокна) випромінюють лазери, реалізовані на квантових точках InAs у матриці GaAs. Вони перевершують лазери на InP по граничному струму, забезпечують ефективне введення випромінювання у волокно, високу температурну стабільність і дозволяють застосовувати групову технологію виготовлення. Тому вертикальний лазер на квантових точках ***InAs у матриці GaAs є ключовим приладом для телекомунікацій.***

Розробляється також вертикальний ультрафіолетовий лазер на основі широкозонного матеріалу GaN із квантовими точками InGa. Такий лазер дає можливість підвищити щільність оптичного запису.

Відомо, що германій і кремній - непрямоzonні напівпровідники, що обмежує їхнє застосування у якості фотоприймачів і випромінювачів. Це одна з причин того, що сучасна оптоелектроніка використовує сполуки $A^{III}B^V$ і $A^{II}B^{VI}$. Але при *переході від об'ємного матеріалу до квантових точок* послабляються обмеження, пов'язані з непрямоzonністю переходів у Ge і Si. Дослідження оптичних властивостей гетероструктур Si-Ge із квантовими точками германія показали, що вони можуть успішно конкурувати з гетероструктурами на сполуках $A^{III}B^V$, які використовуються як фотоприймачі інфрачервоного

випромінювання.

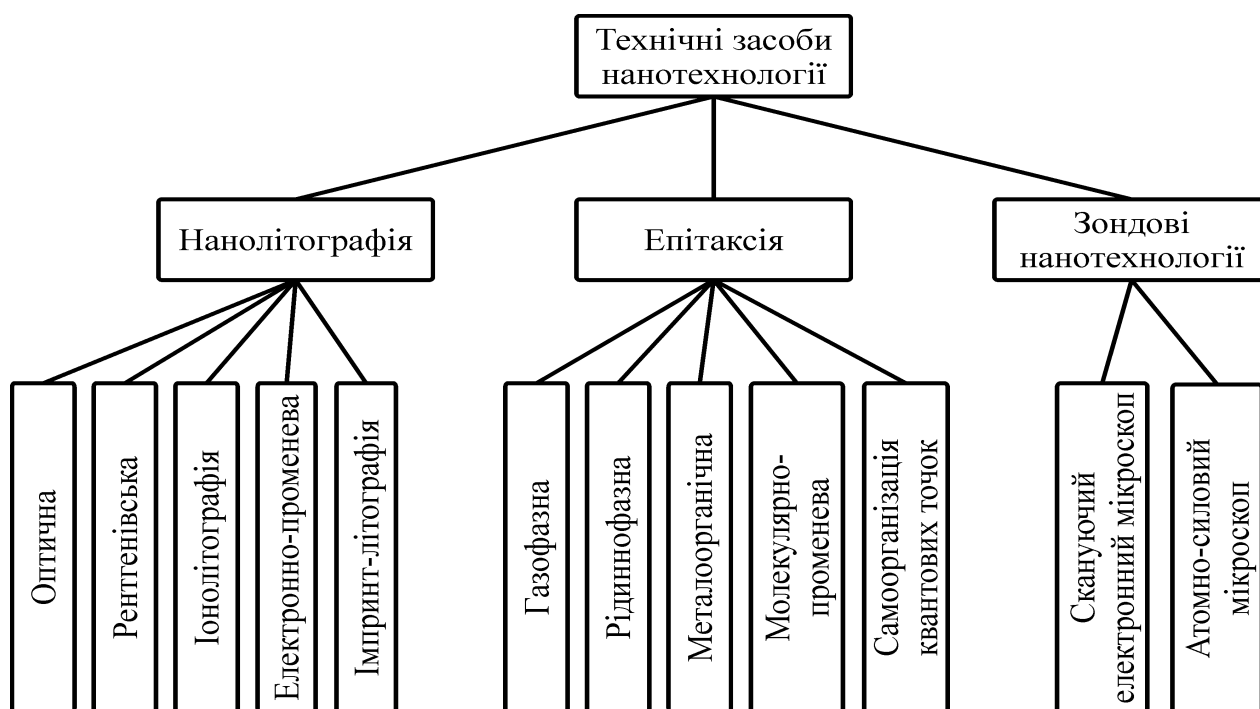
Особливо *перспективні гетероструктури Si-Ge* з погляду можливості вбудовування їх у кремнієві інтегральні мікроелектронні схеми. Напружені квантові точки германія вдається зарощувати близькими за структурою шарами кремнію, на яких потім можна формувати і інші елементи мікроелектроніки. Для волоконно-оптичних ліній зв'язку важливе створення на одному чипі всього набору компонентів, включаючи фотоприймачі. Тоді всі компоненти могли б бути інтегровані в кремнієву технологію інтегральної схеми і сформовані на кремнієвих підкладках.

9. Основні засади нанотехнології напівпровідників

Щоб опанувати технологією виготовлення наноструктур, необхідно оперувати розмірами, порівнянними з відстанню між атомами, тобто у нанометровому діапазоні, що в 1000 разів менше розмірів, звичних для сучасної мікроелектроніки. Можливість формувати нанорозмірні структури з'явилася з розвитком традиційних методів виготовлення напівпровідникових приладів, і в першу чергу, таких, як хімічне осадження з газової фази, молекулярно-променева епітаксія й електронно-променева літографія. Крім того, реальні можливості створення твердотільних наноструктур значно розширилися з застосуванням скануючих точкових зондів і саморегулюючих процесів. При цьому нанотехнологічні прийоми постійно удосконалюються, з'являються все нові й нові методи. Однак основними методами нанотехнології остаються нанолітографія, епітаксія і зондові технології, табл. 9.1.

Таблиця 9.1.

Базові технологічні засоби сучасної нанотехнології



У нанотехнології функціональні елементи і їхні комбінації створюються не так, як у мікроелектронній технології, оскільки у нанотехнологіях часто використовується властива атомам природна схильність до утворення визначених структур атомарних розмірів. У багатьох випадках нанотехнологія

синтезує елементи шляхом упорядкованої «зборки» конструкцій з окремих атомів, які необхідні для формування елементів із заданими функціями.

Як приклад, на рис. 9.1 показана напівпровідникова наноструктура.

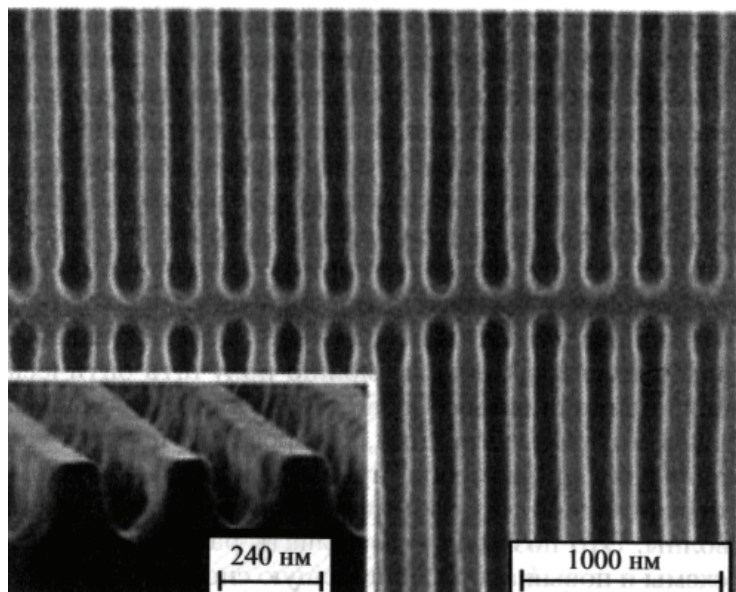


Рис. 9.1. Наноструктура на кремнію з елементами шириною 80 нм і кроком 200 нм

Існує два основних підходи до виготовлення наноструктур, що умовно називаються технологіями «зверху вниз» і «знизу нагору».

У технологіях за принципом «зверху вниз» за допомогою хімічного, механічного або іншого видів обробки з об'єктів великих розмірів одержують вироби набагато меншої величини. Приклад – створення на поверхні кремнієвої пластини схемних елементів інтегральної мікроелектроніки літографічними методами. Роздільна здатність літографії безупинно удосконалювалася протягом останніх десятиліть і в даний час наближається до 10 нм.

У технологіях за принципом «знизу нагору» виробляється збірка макроструктури з елементарних «цеглинок» – атомів, молекул, кластерів, нанотрубок, нанокристалів. Ці елементи в процесі збирання або самозбирання повинні укладатися в необхідному порядку. Приклад – поштучне переміщення й укладання атомів зондом скануючого тунельного мікроскопа. Однак збирання макроскопічного об'єкта поштучним укладанням атомів за допомогою скануючого тунельного мікроскопа або навіть нанороботів в умовах масового виробництва нереальна, це вимагало б занадто великого часу.

Тому методами створення макрооб'єктів за принципом «знизу нагору» повинні стати *кероване самозбирання* і *самоорганізація* необхідних структур,

коли атоми або молекули самі вибудовуються в заданому порядку, як це має місце, наприклад, у біологічних системах.

Рушійна сила само збирання – це прагнення системи до мінімуму енергії. Якщо енергія системи зменшується при колективізації молекул, то молекули будуть з'єднуватися потрібним чином. Якщо енергія системи зменшується при переорієнтації молекул, то вони будуть переорієнтуватися без технологічного впливу ззовні.

Процеси самозбирання йдуть під дією сил Ван-дер-Ваальса, водневих зв'язків, хімічних зв'язків і інших взаємодій. Приклад – одержання необхідної кількості монокристалічних шарів і упорядкованих масивів нанокристаліків (квантових точок) при молекулярно-променевій епітаксії.

9.1. Нанолітографія

Літографією в мікроелектроніці називають різні методи мікрогравірування металевих, діелектричних, і напівпровідникових шарів у процесах виготовлення інтегральних мікросхем. Різні літографічні процеси для створення резистивних масок можуть забезпечити розміри елементів у діапазоні 1000 – 10 нм, тобто ці методи можуть бути придатними не тільки для мікроелектроніки, але і для наноелектроніки.

Основним методом літографії у технології мікросхем в даний час залишається *фотолітографія* – фотохімічний метод мікрогравірування. Крім оптичної літографії досить широко використовуються також ультрафіолетова і рентгенолітографія. Це так звані традиційні, **променеві методи** літографії.

Променеві методи дозволяють перейти до наномасштабів у серійному виробництві інтегральних мікросхем. Водночас виявилися і граничні можливості використовувати налагоджені методи фотолітографії у наноелектроніці. Виявлені обмеження привели до пошуку нових рішень, що могли б забезпечити подальший прогрес у мікро- і нанолітографії.

З 1990-х рр. стали розвиватися **непроменеві методи** літографії, засновані на інших підходах. Це електронно- і іонолітографія, а також нанодрук (імпрінтінг). Згідно з прогнозами, ці нові методи дозволять освоїти всю наногалузь у масовому виробництві інтегральних електронних пристроїв з нанорозмірними елементами. Однак у даний час ще не ясно, які саме з багатьох розроблюваних методів нанолітографії витримають конкуренцію і стануть основними у наномасштабному виробництві.

У даному розділі розглядаються можливості традиційних променевих методів літографії в нанометровому діапазоні і деякі непроменеві методи. У наступному розділі розглянуті також і **зондові методи** нанолітографії.

Роздільна здатність літографії має принципове значення для відтворюваного формування елементів з необхідними нанометровими розмірами. Саме цей параметр є ключовим при науковому дослідженні наноструктур. Однак для масового виробництва приладів нанoeлектроніки важлива не тільки роздільна здатність літографії, але і швидкість експонування літографічного процесу.

Тому порівняння різних літографічних методів варто проводити по цих двох параметрах.

Оптична літографія (фотолітографія) має високі швидкості експонування і цілком задовольняє вимогам масового виробництва для звичайної мікроелектроніки. Однак по роздільній здатності вона має істотні обмеження на мінімальний розмір елемента. Це обумовлено порівняно великою довжиною хвилі випромінювання, використовуваного для експонування резистів.

Оптична літографія розділяється на *контактну* і *проекційну*. У першому випадку фотошаблон щільно прилягає до шару фоторезисту, а в другому методі зображення фотошаблону (маски) проектується на поверхню фоторезисту за допомогою спеціальної оптичної системи. Простота контактного способу забезпечила методів фотолітографії широке застосування в мікроелектронній галузі. Однак для нанoeлектроніки виявляється кращим саме *проекційний* варіант фотолітографії, тому що він дозволяє одержати більш високе розділення деталей.

Тому далі обговорюються можливості саме проекційного варіанта оптичної літографії.

Роздільна здатність α_{\min} залежить від довжини хвилі світла і визначає мінімальні топологічні розміри елементів, наприклад, розміри транзисторів. Чим вище ця здатність, тим більш дрібні деталі можуть бути перенесені на напівпровідникову пластину. Відповідно до емпіричного правила, при даній довжині хвилі λ можна відобразити деталі розміром до $\lambda/2$ – але тільки тоді, коли застосовуються спеціальні заходи для поліпшення якості зображення (наприклад, фазозсувні маски і висококонтрастні резисти).

Підвищення роздільної здатності та відповідне зменшення критичних розмірів формованих елементів при оптичній літографії йде по шляху зменшення довжини хвилі випромінювання. У більш простих випадках достатньо використання випромінювання парів ртуті: лінії 436 нм і лінії 365 нм.

Зараз переважно використовується випромінювання в *ультрафіолетовій* частині оптичного спектру. До 2003 р. у фотолітографії найбільше широко використовувалася довжина хвилі $\lambda = 248$ нм випромінювання KrF ексимерного лазера, чому відповідає $\alpha_{\min} \sim 120$ нм. Потім багато виробників чипів перейшли на використання випромінювання ArF ексимерного лазера з $\lambda = 193$ нм і $\alpha_{\min} \sim 90$ нм. Наступний перехід – до випромінювання ексимерного лазера на основі молекул фтору F_2 ($\lambda = 157$ нм, $\alpha_{\min} \sim 70$ нм). Однак вже для цієї довжини хвилі кварцеві лінзи стають непрозорими, і використовуються лінзи з кристалів CaF_2 – менш міцного і більш дорогого матеріалу. При подальшому зменшенні λ неминучий перехід на *дзеркальну оптику* з повною заміною літографічного устаткування.

Фотолітографія у глибокому ультрафіолеті є природним розвитком оптичної літографії з застосуванням короткохвильового випромінювання. Для цих цілей застосовується *синхротронне* випромінювання і випромінювання *плазмових* лазерних джерел. Вони добре забезпечують формування елементів розмірами до 100 нм і можуть бути вдосконалені для створення 30 нм елементів. Незважаючи на фізичні переваги, використання синхротронного випромінювання не знаходить широкого технологічного застосування в напівпровідниковій електроніці через складність, енергоємність і громіздкість синхротронів. Плазмові лазерні джерела багато в чому позбавлені цих недоліків, що робить їх більш перспективними для практичної нанолітографії.

Близько 2007 р. була досягнута роздільна здатність до 30 нм і очікується, що у 2014 – 2016 роках промислові ультрафіолетові нанолітографії досягнуть розрізняювальної здатності 20 нм, а в наступні роки – і до 10 нм. Таким чином, фотолітографія ще не вичерпала своїх можливостей. Очікується, що в найближче десятиліття напівпровідникова промисловість буде продовжувати йти по шляху удосконалення фотолітографії.

Рентгенівська літографія, яка використовує випромінювання з довжиною хвилі близько 1 нм, являє собою останню сходинка на шляху зменшення довжини хвилі *електромагнітного випромінювання* для цілей літографії. При цьому за відсутності підходящої рентгенівської оптики іноді доводиться вести пряме експонування.

Джерелами рентгенівського випромінювання є або металева мішень, яка бомбардується пучком електронів, або синхротрон (циклічний прискорювач елементарних часток), у якому по кругових траєкторіях рухаються електрони з майже світловими швидкостями. Інтенсивність рентгенівського синхротронного випромінювання в тисячі разів перевершує інтенсивність випромінювання мішені.

Принципи й основні етапи рентгенолітографії і фотолітографії однакові, але в рентгенолітографії для експонування резистів використовується рентгенівське випромінювання. Рентгенолітографія використовується для виготовлення наноструктур, але через складність і дорожнечу устаткування і технології поки не застосовується в серійному виробництві. Достоїнства рентгенолітографії – висока розрізняльна здатність і висока продуктивність, обумовлена малим часом експонування і можливістю одержання зображення на великих площах. Крім того, рентгенівське випромінювання має унікальну властивість – поглинатися без розсіювання.

Досяжні мінімальні розміри елементів складають 50 – 70 нм. Перевагою рентгенівської літографії є можливість використання одношарових резистивних масок і висока відтворюваність. Недоліки ж аналогічні тим, що відзначені вище для літографії в глибокому ультрафіолеті.

Діапазон довжин хвиль, використовуваних для експонування лежить у межах 0,5 – 5 нм. Однак більш жорстке рентгенівське випромінювання ($\lambda < 0,5$ нм) викликає утворення фотоелектронів у підкладках і їхнє проникнення назад у резист. Електронні ефекти і дифракція позначаються при розмірах деталей маски менших за 20 нм. Цими розмірами і визначається роздільна здатність рентгенолітографії.

Для рентгенівської літографії необхідна розробка спеціальних дзеркал. При їхньому виготовленні також застосовуються нанотехнології. Створення багат шарових рентгенівських дзеркал розширює можливість застосування рентгенолітографії у нанотехнологіях.

Електронно-променева літографія є найбільше підходящою основою для масового виробництва наноструктур. З використанням одиночного променя вона забезпечує також достатні швидкості експонування. Типове розділення складає 30 нм із можливістю досягти навіть 5 нм при використанні неорганічних резистів. Основним недоліком є невисока продуктивність, що визначається щільністю електронного струму і чутливістю резисту.

Підвищення густини струму в електронному промені крім технічних обмежень має і фізичні обмеження. При високій густині струму взаємодія між електронами приводить до внутрішнього розширення променя, що погіршує розділення. Для прийнятної продуктивності необхідні резисти з низьким порогом чутливості.

Таким чином, електронно-променева літографія (ЕПЛ) – це промислова технологія, що має високу роздільну здатність. Основні області її застосування: виготовлення інтегральних мікросхем із розмірами елементів менш за 100 нм, а також виготовлення шаблонів для рентгенолітографії і фотолітографії у далекому ультрафіолеті.

Принципи й основні етапи електронолітографії і фотолітографії аналогічні, відмінність у тому, що в ЕПЛ замість світлових потоків для експонування використовуються пучки електронів, прискорених електричним полем. Електронно-променева літографія в даний час розглядається як один з альтернативних методів нанолітографії, і тому інтенсивно ведуться розробки по удосконаленню і збільшенню роздільної здатності ЕПЛ.

Найважливіша перевага методу – відсутність дифракційних обмежень на роздільну здатність. Для експонування використовуються спеціальні високовольтні установки з напругою до 200 кВ, що прискорюють електрони. Але специфічна електронна літографія можлива і на низькій електричній напрузі (до 10 В) – з використанням *скануючих тунельних мікроскопів* (цей метод розглядається у наступному розділі).

Методу ЕПЛ властива велика глибина різкості переданого зображення, і не відбувається спотворення рисунку при збільшенні глибини рельєфу або неплоскостності пластини. Однак високовольтні пучки, проникаючи у підкладку, створюють в поверхневому шарі напівпровідника високу густину радіаційних дефектів. Це позначається на якості елементів мікросхем. При зниженні енергії електронів глибина проникнення в підкладку зменшується і густина радіаційних дефектів знижується.

На практиці застосовують два види ЕПЛ – скануючу і проєкційну.

Сканує електронно-променева літографія використовує керований комп'ютером звужений пучок електронів, який переміщується по поверхні підкладки, покритої резистом, і креслить заданий рисунок. Метод є серійним і застосовується в тих випадках, коли оптична літографія не задовольняє вимогам по точності відтворення і глибині фокуса.

У наноелектронному діапазоні розмірів більш високу продуктивність може дати **проекційна електронно-променева літографія**. Один з її сучасних варіантів, названий *SCALPEL*, забезпечує високе розрізнявання деталей, велике поле зображення і значну глибину фокуса. Програма *SCALPEL* – комплексна, і включає розробку літографічного устаткування, масок і резистів. Гранична здатність методу 35 нм. Вона обмежена взаємодією електронів у пучку, що приводить до його “розмиття”.

Іонно-променева літографія по своїм технологічним принципам близька до електронно-променевої літографії. Вона використовується для експонування резистів товщиною до 20 нм. Іони можуть також використовуватися для безмаскового створення рисунка елементів інтегральних схем – прямою модифікацією властивостей матеріалу підкладки. Установки для іонно-променевої обробки матеріалів мають прийнятні швидкості експонування.

Принцип і основні етапи технології такі ж, як в інших видах традиційної літографії. Експонування проводиться **іонними пучками**: широким пучком, звуженим пучком або через проекційну систему на основі промислових установок іонної імплантації.

Фокусовані іонні пучки використовуються для рисунку по резисту і можуть застосовуватися для безпосередньої модифікації підкладки. У зв'язку з тим, що іонний пучок створює дефекти в поверхневому шарі, у кілька разів змінюється швидкість розчинення підкладки в деяких розчинниках. Тому створюється можливість відмовитися від резиста, оскільки поверхневі шари матеріалів самі поведуться як резисти.

Порівнюючи ефективність експонування резисту іонами й електронами одного і того ж енергетичного діапазону (50 – 100 кеВ), слід зазначити, що іони цілком передають свою енергію резистивному шару, а електрони проходять глибше у підкладку. Це веде до істотних відмінностей у граничній чутливості резистів, які використовуються для іонного експонування.

У іонолітографії відсутнє дифракційне обмеження: через велику масу іонів довжина хвилі де Бройля для іонів на два порядки менше, ніж для електронів при тих же прискорюючих напругах. Іонні пучки не мають зворотного розсіювання, вторинні електрони у резисті мають малу енергію, а отже і малу довжину вільного пробігу. Пучок важких (у порівнянні з електронами) іонів має слабке кутове розсіювання. Завдяки цим факторам іонолітографія забезпечує розрізнявальну здатність до 10 нм.

Звичайно при іонній літографії використовують ті ж резисти, що й у електроннографії; для експонування застосовуються легкі іони H^+ або He^{++} . Довжина пробігу легких іонів з енергіями 50 – 100 кеВ приблизно дорівнює товщині резиста, використовуваного в технологічних процесах. Іони всю енергію передають резисту і не проникають у підкладку, як електрони. Тому чутливість резистів до іонного пучка на два порядки вище, ніж до електронного, а час експонування іонним пучком значно менше. У результаті вище продуктивність процесу літографії.

Проекційні установки мають високу розрізняльну здатність і продуктивність. Однак через складність створення джерел іонів, систем прискорення і фокусування іонних пучків цей метод поки не одержав поширення у серійному виробництві.

Нанодрук (імпрінт-літографія) є багатообіцяючою технологією літографії, хоча необхідні додаткові дослідження, перш ніж вона могла б повноправно увійти в промислове виробництво.

Розвиток методів традиційної променевої літографії з метою виходу в нанометровий діапазон вимагає величезних фінансових вкладень, що може бути економічно виправдано тільки при масовому виробництві продукції, наприклад мікросхем процесорів або пам'яті.

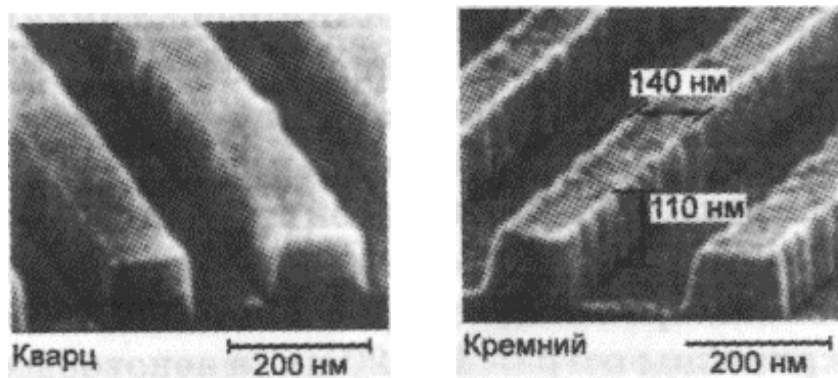


Рис. 9.2. Електронні мікрофотографії: а – кварцевий штамп, б – формована поверхня кремнію.

Імпрінт-літографія – більш проста і дешева технологія, заснована на іншому підході до виробництва нанолітографічної продукції. (Інші назви цієї технології – нанодрукowana літографія, імпрінтінг).

Дана технологія заснована на пресуванні резиста у формах з наступним переносом нанорисунка на пластину напівпровідника. При імпрінт-літографії зображення в тонкому шарі резиста створюється за рахунок *фізичної деформації резиста*, а не за рахунок модифікації хімічної структури резиста за допомогою опромінення. Роль шаблона грає прес-форма (штамп).

Електронні фотографії деяких деталей технології імпрінтинга показані на рис. 9.2.

Послідовність формування рисунка на резисті містить у собі кілька етапів. Спочатку на підкладці за допомогою центрифуги створюється площинно-паралельний шар резиста; потім композиція нагрівається до температури розм'якшення твердого резиста і підготовлений штамп, вдавлюється в шар резиста, що заповнює поглиблення на штампі. Після того система охолоджується до температури нижче температури твердіння резиста, і штамп піднімається.

Роздільна здатність імпрінтинга визначається, головним чином, мінімальними розмірами деталей на штампі. Цим методом можна одержувати структури з розмірами до 10 нм.

Як приклад на рис. 9.3 приведена електронна мікрофотографія смужок металу на підкладці після імпрінтингу і напилювання. Для смужок характерні дзеркальність і гострі кути, що недосяжно при традиційних методах.



Рис. 9.3. Електронна мікрофотографія смужок металу на підкладці після імпрінтингу і напилювання. Ширина смуг – 70 нм, висота – 200 нм

Імпрінт-літографія має ряд очевидних переваг. Тут не використовують пучки електронів, фотонів і іонів, тому немає проблем, зв'язаних з дифракцією, розсіюванням первинних і вторинних електронів, з хімічними процесами у резисті. Імпрінтинг не тільки має розділення до 10 нм, але і дозволяє наносити рисунок відразу на площі порядку декількох квадратних сантиметрів, що дає високу продуктивність. Таким чином, імпрінтинг – більш проста і дешева технологія в порівнянні з традиційною літографією у нанометровому діапазоні.

Недоліки імпрінт-літографії – зношення штампів і складність точного позиціонування структур на підкладці, коли застосовується декілька послідовно

виконуваних етапів літографії. Наприклад, у виробництві чипів потрібне позиціонування з точністю до малих часток розміру самого дрібного елемента.

Імпрінтинг може бути ефективним методом у технологіях з менш жорсткими вимогами, наприклад, при формуванні періодичних поверхневих структур, гребінчатих електродних структур, в органічній оптоелектроніці, при виробництві сенсорів, у нанобіотехнології. Такі галузі, як мікро- і нанотехніка, а також лабораторії-на-чипі також мають потребу у високопродуктивній нанотехнології з низькою вартістю, подібної до наноімпрінтингу.

9.2. Епітаксіальні методи одержання наноструктур

У попередніх розділах було розглянуто ряд пристроїв на квантових ямах і плоских гетеропереходах – лазери, надгратки й інші багатобар'єрні структури. Для нормальної роботи приладів на плоских наногетероструктурах необхідні *різкі атомарно гладкі границі* між шарами, причому товщина шарів у багатьох випадках повинна утримуватись з атомарною точністю.

Очевидно, що шарові гетероструктури можуть бути отримані тільки епітаксальним нарощуванням шарів. Але не всі епітаксіальні технології можуть дати границі з потрібними властивостями. Наприклад, епітаксія з газової фази відбувається при високих температурах, що приводить до помітної дифузії атомів в об'єм і до розмивання границь шарів. Рідиннофазна епітаксія не дає шарів, достатньо однорідних по товщині.

Для формування плоских гетероструктур і наногетероструктур використовуються в основному дві технології: молекулярно-променева епітаксія (МПЕ) і газофазна епітаксія з металоорганічних сполук.

Молекулярно-променева епітаксія дорожча за газофазну, але більше розроблена, має більш широкі можливості для контролю і керування ростом. МПЕ використовується головним чином для одержання високоякісних тонких плівок і гетероструктур з нанометровими шарами на основі напівпровідникових сполук $A^{III}B^V$, $A^{II}B^{VI}$, SiGe і інших напівпровідників.

Значення технології МПЕ для сучасної електроніки. В даний час оптоелектроніка, мікрохвильова техніка, світлотехніка, електронна техніка для систем телекомунікацій і зв'язку працюють в основному на напівпровідникових гетероструктурах і наногетероструктурах. В оптоелектроніці – це сонячні елементи, інфрачервоні фотоприймачі на надгратках і багатшарових структурах на квантових ямах, світлодіоди і фотодіоди на надгратках, лазери на надгратках, квантових ямах і квантових точках. Лазери на підкладках InP, що

випромінюють на довжинах хвиль поблизу 1,3 і 1,5 мкм (діапазон максимальної прозорості оптоволокна), забезпечують 70% комунікацій Інтернету. Лазери на основі GaAs, що випромінюють на довжині хвилі 0,85-0,95 мкм, працюють на коротких волоконно-оптичних лініях зв'язку (ВОЛЗ) і використовуються для оптичного запису.

Швидко росте ринок лазерів на основі GaAs з вертикальним резонатором на квантових ямах і квантових точках зі спеціальними багат шаровими інтерференційними дзеркалами. Такі лазери мають малі габарити (до мікронних розмірів), низький граничний струм і гарну діаграму спрямованості, що забезпечує більш ефективне введення випромінювання у волокно. Кількість шарів у лазерних структурах може складати кілька сотень, при цьому потрібно з високою точністю витримувати товщини цих шарів (близько 1%).

В останні роки освоєне виробництво яскравих (суперкандельних) світлодіодів видимого діапазону на основі гетероструктур нітриду, фосфіду й арсеніду галію для екранів, табло, сигнальних пристроїв, побутової й автомобільної світлотехніки, мобільних телефонів, фото- і відеокамер. Перевагою світлодіодів у порівнянні з лампами розжарення є зменшення енергоспоживання в 5 – 7 разів, довготривала робота (до десятків тисяч годин), керування світловим потоком і кольором.

Гетероструктурні польові транзистори використовуються в цифровому телебаченні, в електроніці сучасних систем зв'язку, включаючи супутникову і стільникову. На основі резонансно-тунельних діодів розробляється елементна база надшвидкодійних обчислювальних систем.

Наведені приклади далеко не вичерпують усіх застосувань гетеропереходів. Але і на цих прикладах можна бачити, яке велике значення технології нарощування наночарів для сучасної електронної індустрії.

Наногетероструктури забезпечують найвищі швидкості прольоту електронів у напівпровідникових приладах, мінімальні втрати енергії, можливість керування шириною забороненої зони. Для наногетероструктурних приладів необхідна атомарна дзеркальність границь шарів і достатня різкість зміни сполуки на границях. Технологія МПЕ дає можливість одержувати шари з такими властивостями. Тому напівпровідникові наноструктури виготовляються в основному за технологією МПЕ. В даний час МПЕ використовується для одержання не тільки плоских шарів, але й упорядкованих масивів квантових точок у системах на основі $A^{III}B^V$ і SiGe. Лазери на масивах квантових точок у системах типу $A^{III}B^V$ уже виробляються в промисловості.

Розробляються інфрачервоні (ІЧ) фотоприймачі на квантових точках у сполуці SiGe; прогнозується їх широке застосування.

Для телекомунікацій особливо важливі фотоприймачі і випромінювачі, що працюють у діапазоні «вікон прозорості» атмосфери: 3 – 5 мкм і 8 – 12 мкм. Отримані також ІЧ-лазери на надгратках, які можуть працювати в діапазоні 3 – 12 мкм. Це каскадні лазери, що працюють на внутрізонних і на міжзонних переходах. При виготовленні ІЧ-лазерів жорсткі вимоги ставляться до товщини шарів, які повинні мати атомарну точність. В даний час вони можуть вироблятися тільки методом МПЕ.

Основні особливості МПЕ. У методі МПЕ пучки атомів, які є компонентами сполуки (або елементарного напівпровідника), що нарощується на підкладці, направляються у надвисокому вакуумі на нагріту підкладку, де й відбувається синтез сполуки і ріст епітаксialного шару. Джерела пучків атомів – випарні комірки, у кожній з яких утримується свій матеріал. Основні елементи ростової камери схематично показані на рис. 9.4.

Схема на рис. 9.4 пояснює принцип епітаксialного нарощування. Тут 1 – підігрівач підкладки 2; 4 – випарні комірки для компонентів нарощуваної сполуки (Ga, In, As, P и т. д.); 5 – комірка для випару легуючих елементів, що визначають тип провідності і величину електропровідності. Випарних комірок у камері повинне бути стільки, скільки необхідно для одержання шарів 2-, 3-, 4- і 5-компонентних сполук (наприклад, GaAs, AlGaAs, InGaAs, InGaAsSbBi) з різними легуючими елементами. Концентрації елементів у шарі, який нарощується, визначаються інтенсивністю пучків, а інтенсивність пучків задається температурою у випарних комірках.

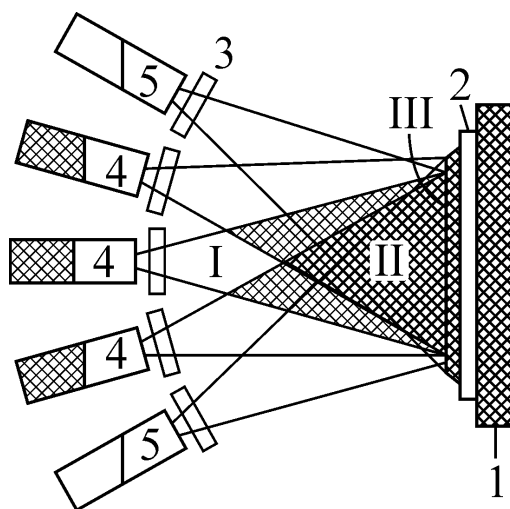


Рис. 9.4. Спрощена схема епітаксiального нарощування: I – зона генерації атомних пучків, II – зона змішування пучків, III – зона епітаксiального росту.

Керуючи температурою комірок і підкладки, можна одержувати шари потрібної сполуки. Сполуку можна змінювати після завершення формування будь-якого шару на підкладці. Комірки мають малу теплову інерцію і швидко реагують на регулювання температури, тому можна створювати різку зміну концентрації компонентів і рівня легування в гетеропереході. За допомогою заслонок 3 можна в будь-який момент вмикати і вимикати кожний з атомарних потоків, спрямованих на підкладку, що дозволяє одержувати ще більш різкий профіль зміни сполуки у гетеропереході.

На рис. 9.5 дана більш детальна принципова схема камери для епітаксiального нарощування. Камера виконана з нержавіючої сталі, при роботі в ній підтримується тиск близько 10^{-11} мм рт. ст. Кожна комірка 3 оточена кріоекранами 4, охолоджуваними рідким азотом (77 K), друга кріопанель 4 розташована по периферії камери. Ці кріоекрани діють як додаткове відкачування.

В устаткуванні використовуються керовані заслонки окремих комірок 2 і основна (загальна) заслонка 6, підкладка на обертовому столику 7, який тримає підкладку і пристрій для її підігріву. Підвищення однорідності зростаючого шару досягається обертанням підкладки; воно передається від спеціального двигуна через вісь 11. Контроль вакууму здійснюється вимірником 8. У камері маються оглядові вікна 10, вакуумний шлюз для зміни зразків 9, електронна гармата 5 і флуоресціюючий екран 1, що використовуються для контролю швидкості росту, якості поверхні, орієнтації поверхні і величини постійної ґратки (безпосередньо при епітаксії). Швидкість росту – приблизно один моношар у секунду, тобто близько 1 мкм/год.

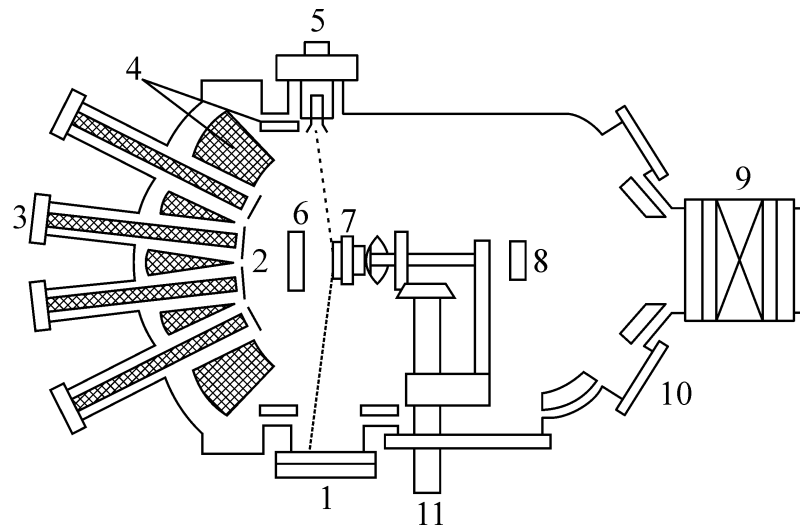


Рис. 9.5. Принципова схема установки для гетероепітаксії.

Основне досягнення технології МПЕ – це можливість формування атомарно-гладких границь шарів, що принципово важливо для наногетероструктурних приладів. На рис. 9.6 приведена мікрофотографія двох шарів AlAs товщиною близько 2 нм (6 моношарів) у матриці GaAs. Зображення ілюструє атомну точність поверхонь розділу.

Технологія МПЕ дозволяє одержувати різкі стрибки концентрації компонентів у шарах, забезпечує можливість створення структур зі складним розподілом концентрації основних і домішкових елементів. Наявність надвисокого вакууму в робочій камері виключає забруднення підкладки і зростаючих шарів.

Низька температура росту за технологією МПЕ мінімізує дифузію в об'ємі, яка розмиває границі між шарами. Важлива також можливість контролю і корекції росту безпосередньо в ході процесу, діагностика росту, точний контроль температури підкладки, комп'ютерне керування параметрами процесу.

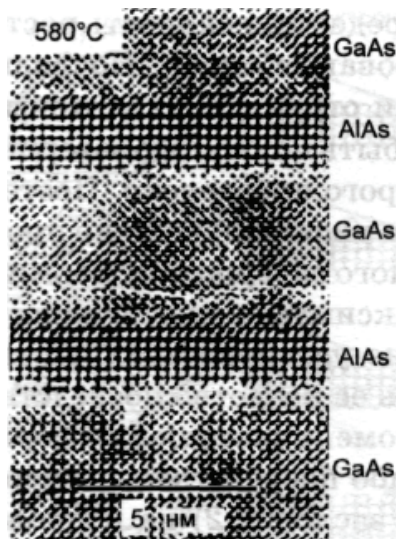


Рис. 9.6. Електронно-мікроскопічний знімок двох шарів AlAs (товщиною 6 моношарів) у матриці GaAs

Завдяки своїм можливостям, МПЕ в даний час відіграє головну роль у розвитку технології одержання напівпровідникових наногетероструктур.

9.3. Зондові методи одержання наноструктур

В даний час скануюча зондова мікроскопія (СЗМ) є одним з найбільш ефективних методів дослідження атомної структури і локальних властивостей поверхні. СЗМ заснована на скануванні поверхні твердотільним зондом з тонким вістрям. Радіус кривизни вістря 10 – 20 нм, але в деяких пристроях він може бути і менше 10 нм. Відстань від вістря зонда до поверхні в різних пристроях і при різних режимах роботи змінюється в межах 0,2 – 10 нм.

При дослідженнях поверхні *пасивними* зондовими методами зонд може взаємодіяти з поверхнею за допомогою тунельного струму, шляхом теплообміну, використовуючи міжатомні, електричні або магнітні сили. Ці взаємодії покладені в основу функціонування різних видів *скануючих зондових мікроскопів*: тунельного, атомносилового, теплового, мікроскопа на електростатичних силах, на магнітних силах і т. д. У кожному випадку локальна взаємодія зонда, що рухається над досліджуваною поверхнею, супроводжується специфічним сигналом. По сукупності сигналів, отриманих при скануванні, створюється *зображення* поверхні й одночасно може визначатися *поверхневий розподіл* досліджуваної величини, наприклад, електропровідності, діелектричної проникності, намагніченості, температури, напруженості

електричного або магнітного поля. Метод СЗМ має високу роздільну здатність (аж до атомної).

Розроблені і широко застосовуються не тільки пасивні дослідження, але й **активні методи** спрямованого і контрольованого впливу зонда на поверхню, наприклад за допомогою сильних електричних полів і досить щільних електричних струмів в області вістря, локального масопереносу, локальних хімічних реакцій і ін. На цих впливах заснований ряд **зондових нанотехнологій**, у тому числі локальна модифікація поверхні, нанолітографія, надщільний запис інформації, дослідження механічних властивостей речовини й інше. Нанолітографія досягає своєї граничної роздільної здатності, коли за допомогою зонда реалізується переміщення окремих атомів і закріплення їх у новому місці

Способи і прийоми формування, модифікації і розміщення елементів нанометрових розмірів на поверхні підкладок за допомогою зондів з вістрям досить різні. Сучасний зондовий мікроскоп дозволяє реалізувати десятки різних методик досліджень і технологічних операцій. Зондові технології мають граничну локальність, можливість такої взаємодії, що не ушкоджує поверхню при її дослідженні, візуалізацією нанотехнологічних операцій на поверхні з елементами меншими за 10 нм. Розвиток зондових методів відкриває перспективи створення елементів електроніки на основі окремих атомів і молекул. Недоліком усіх зондових технологій є поки що їх низька продуктивність.

Зондові методи дослідження і нанотехнологій застосовуються у діагностиці напівпровідникових структур, при дослідженні процесів епітаксiального росту, для надчутливої сенсорики, у медицині, біології і т. д. Далі описуються принципи роботи двох основних видів скануючих зондових мікроскопів: тунельного й атомно-силового, приводяться типові приклади їх використання в нанотехнологіях.

Скануючий тунельний мікроскоп (СТМ). В даний час СТМ і інші скануючі зондові мікроскопи є основними аналітичними засобами в нанотехнології. В основу принципу дії СТМ покладений вимір тунельного струму через тонкий вакуумний проміжок між двома провідниками. Висока чутливість СТМ обумовлена різкою залежністю тунельного струму від ширини потенційного бар'єра.

Принципова схема СТМ показана на рис. 9.6. Для переміщення зонда уздовж трьох напрямків використовуються п'єзодвигуни, що дозволяють керувати переміщеннями нанометрового масштабу. У робочому режимі відстань між зондом і зразком по вертикальній осі контролюється системою зворотного зв'язку. На двигун подається напруга зворотного зв'язку і п'єзодвигун підводить зонд до поверхні зразка на таку відстань, при якому тунельний струм J_T досягає заданої величини. Типові значення тунельного проміжку $d \sim 0,5 - 1$ нм, а значення струму $J_T \sim 1 - 10$ нА.

Поверхня має атомний рельєф, на ній є дефекти і неоднорідності. Тому при горизонтальному скануванні відстань між зондом і поверхнею міняється, отже, повинний мінятися і тунельний струм J_T . Над виступом, де зонд ближче до поверхні, тунельний струм буде більше, над западиною він менше. Залежність струму від відстані є досить різкою (експонентною), тому чутливість СТМ до змін рельєфу дуже висока.

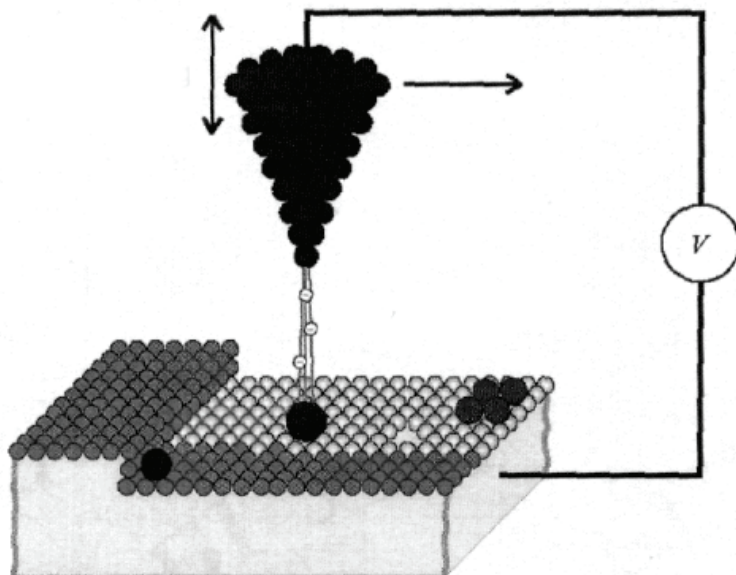


Рис. 9.7. Відносне розташування зонда і підкладки у тунельному мікроскопі

При зміні тунельного проміжку на 0,1 нм тунельний струм змінюється на порядок; тому розрізнення по нормалі до поверхні складає близько 0,05 нм. Приклад вістря зонду показаний на рис. 9.8. Роздільна здатність у площині визначається якістю вістря зонда, а також залежить від вібрацій і акустичних шумів. Наприклад, для вістря з монокристала вольфраму з віссю [111] кінчик має форму піраміди, що завершується 1 - 3 атомами. Якщо на кінці знаходиться 1 атом, СТМ реагує на виступаючий над поверхнею атом, що дозволяє «бачити» окремі атоми. Більш широке вістря має мікро виступи, і якщо хоча б

один з них ближче до поверхні усього на 0,1 нм, то весь тунельний струм іде через нього, так що розрізнення атомів також стає можливим. Слід зазначити, що істинно атомне розрізнення виходить лише при дуже низьких температурах і у високому вакуумі. У звичайних умовах досягається роздільна здатність порядку одного нанометра.

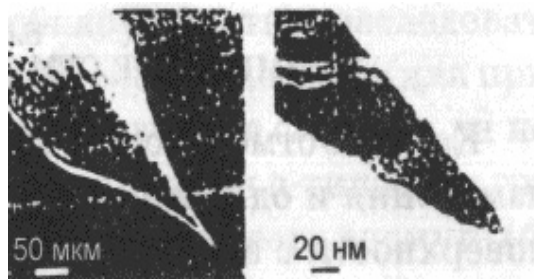


Рис. 9.8. Мікроскопічне зображення типового металевого зонда при двох різних збільшеннях

За допомогою СТМ можна досліджувати тільки електропровідні поверхні – метали, напівпровідники, надпровідники. Крім того, СТМ фактично реєструє не самі атоми, а розподіл щільності електронних станів на поверхні зразка.

В даний час скануючі електронні мікроскопи (СЕМ) високого розрізнення також дають атомну точність. Однак у СЕМ енергія електронів велика ($10^3 - 10^5$ eV), що створює дефекти на досліджуваній поверхні. У СТМ енергія тунелюючих електронів не перевищує декількох електрон-вольтів, тобто *менше енергії хімічного зв'язку*, і тому досліджувана поверхня не ушкоджується. Крім того, у СТМ немає лінз, які спотворюють зображення.

СТМ широко використовується для дослідження поверхні напівпровідників. В об'ємі ідеального кристала всі валентні зв'язки насичені. Однак в атомах поверхневого шару присутні *обірвані (ненасичені) зв'язки*, що збільшує енергію поверхневого шару. Для зменшення цієї енергії поверхневі атоми перебудовуються таким чином, щоб число обірваних зв'язків скоротилося. У результаті на поверхні змінюється структура кристалічної ґратки: відбувається так звана *реконструкція* поверхні. Існує оптимальне розташування атомів на поверхні, якому і відповідає мінімум поверхневої енергії.

За допомогою СТМ можна досліджувати поверхневі дефекти і проводити діагностику напівпровідникових епітаксціальних плівок; можна простежити процес росту тонких шарів від моменту утворення на підкладці кристалітів-

зародків до утворення шару необхідної товщини. Це дозволяє з'ясувати особливості росту при різних технологічних умовах, підібрати оптимальні режими одержання як тонких плівок, так і нанорозмірних шаруватих структур.

СТМ дозволяє також проводити маніпуляцію атомами і молекулами.

Зонд СТМ, знаходячись у безпосередній близькості від поверхні (0,5 – 1 нм), взаємодіє з її атомами. Це дозволяє маніпулювати окремими атомами і молекулами, що знаходяться на поверхні твердого тіла.

Взаємодія між зондом і зразком або адсорбованою речовиною (звичайно – водою) на поверхні зразка може здійснюватися за допомогою одного з трьох механізмів: за рахунок сил хімічного зв'язку між атомами, за рахунок сил електричного поля і за рахунок тунелювання електронів. У тунельній щілині в області вістря створюються поля до 10^8 В/см, що досить для виривання атомів електричним полем – польовий випар. Крім того, при зіткненнях з атомами поверхні тунелюючі електрони викликають електронне або коливальне зрушення молекул, що може супроводжуватися десорбцією, дисоціацією або зміною конфігурації молекул і навіть дозволяє робити синтез двох окремих молекул в одну.

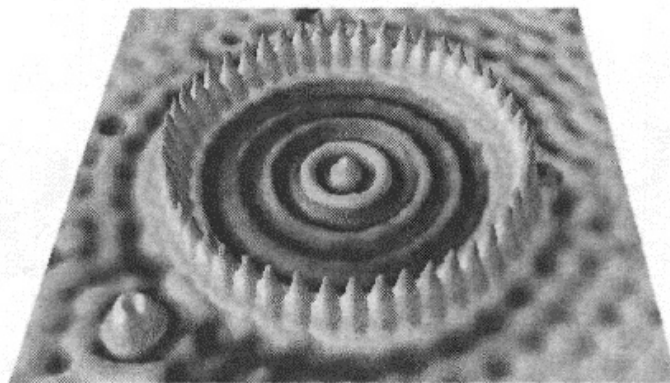


Рис. 9.9. Зображення штучного «загону» у скануючому тунельному мікроскопі

Для горизонтального переміщення атомів за допомогою СТМ використовуються сили взаємодії зонда з атомами поверхні. Зонд міститься над обраним атомом на поверхні і підводиться до нього на таку відстань, щоб сила притягання атома до зонда виявилася достатньою для *проходження атома слідом за зондом*. Далі зонд переміщається до заданого положення, і атом проходить за ним.

Прикладом такої атомної зборки може служити так називаний «квантовий загін» для електрона. На рис. 9.9 показаний такий загін з 48 атомів заліза на

поверхні пластини кремнію. Радіус кільця з атомів заліза дорівнює 7,3 нм. Концентричні кільця усередині загону – стоячі хвилі де Бройля електрона, захопленого пасткою (законом). Світлі кільця відповідають максимумам інтенсивності хвиль де Бройля (максимальної імовірності виявлення електрона).

Крім описаного горизонтального переміщення, можливі багато інших видів маніпуляцій окремими атомами і молекулами: дисоціація, синтез (утворення хімічних зв'язків між двома молекулами), зміна форми великих молекул і т.д.

Особливо слід зазначити можливість *масопереносу з вістря зонда*. Цей процес можливий тільки в умовах високого вакууму, який дозволяє зберегти чистоту підкладки і вістря. При наявності поля напруженістю $E \sim 10^9$ В/см поблизу поверхні стає можливою десорбція атомів або молекул у вигляді іонів навіть при кімнатній температурі. Такі поля створюються у зоні вістря зонда при напрузі близької до 10 В.

Атомно-силовий мікроскоп (АСМ). При використанні СТМ спостерігалися відносно великі сили, які діють між зондом і поверхнею. Саме ця силова взаємодія зонда з поверхнею і використовується для одержання сигналу в АСМ. У принципі, зондовий атомно-силовий мікроскоп може працювати на основі будь-яких взаємодій – міжатомних, електричних, магнітних, теплових і ін. У даний час існує кілька різновидів АСМ, що працюють на основі різноманітних взаємодій.

У найбільш розповсюджених АСМ використовуються міжатомні сили притягання і відштовхування. Далі розглядається саме цей варіант АСМ. Для АСМ, що діє на міжатомних силах, провідність підкладки ролі не грає, так що можна робити дослідження не тільки на електропровідних матеріалах (як у разі СТМ), але також і на діелектриках, органічних і біологічних матеріалах. Тому АСМ більш універсальний, чим СТМ, і знаходить більш широке застосування як у різних дослідженнях, так і у нанотехнологіях.

Принцип роботи АСМ пояснюється на рис. 9.10. Поверхня зразка 1 сканується зондом 2. Зонд кріпиться на дуже чуттєвій до деформацій консолі 3, що називають *кантілевером*; утримувач кантілевера 4 скріплений з трьохкоординатним п'єзосканером – у принципі, таким же як і для СТМ.

Використовуються три режими роботи кантілевера: контактний, безконтактний, напівконтактний.

У контактному режимі при зближенні вістря зонда і поверхні між ними виникає притягання, сила якого *зростає* зі зменшенням відстані. Але на деякій відстані, коли електронні хмари атомів починають перекриватися, сила електростатичного відштовхування хмар при зближенні експонентно наростає і послабляє силу далекодіючого притягання. На відстані $z_0 \sim 0,2 - 0,4$ нм обидві ці сили зрівнюються.

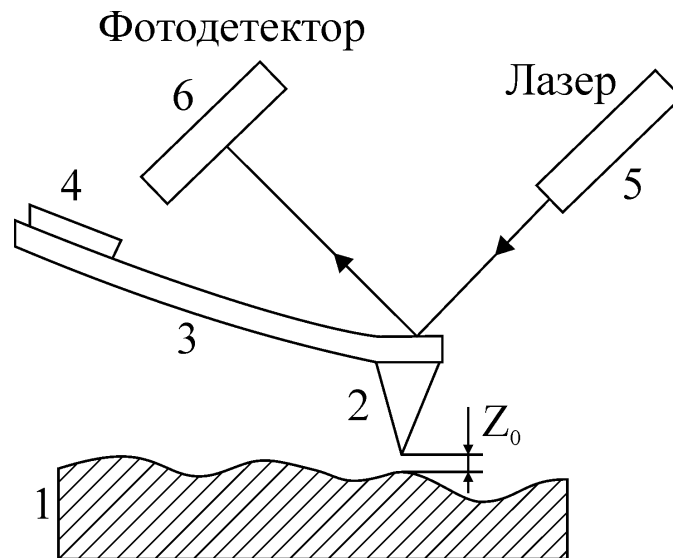


Рис. 9.10. Схема зондового датчика АСМ і оптичної реєстрації прогину кантилевера

При роботі у **контактному режимі** використовуються **сили відштовхування**, що переважають на малих відстанях (менших за 0,2 нм). У рівновазі кантилевера сила відштовхування компенсується силами, що притискають зонд до поверхні; почасти це пружна сила кантилевера. Контактний режим застосовується для дослідження поверхонь з високою твердістю і міцністю. Недоліки контактного режиму – велика імовірність ушкодження зонда і поверхні. Крім того, цей режим не дає Істро атомного розділення (за винятком окремих випадків). На практиці частіше використовуються інші режими роботи АСМ.

У **безконтактному режимі** використовуються міжатомні **сили притягання**, що домінують на відстані від вістря до поверхні близько 5 - 10 нм. В області відстаней, що відповідають безконтактному режимові, залежність сили від відстані слабкіше, ніж в області сил відштовхування, так що в цьому режимі при зміні відстані між вістрям і зразком кантилевер відхиляється значно менше, ніж у контактному режимі. Це знижує чутливість методу. Але переваги безконтактного режиму – можливість роботи з м'якими й еластичними

матеріалами (наприклад, біологічними), а також з матеріалами, властивості яких можуть змінюватися при торканні зонда (наприклад, напівпровідниковими кристалами і структурами). Слід зазначити, що, якщо на поверхні є моношари води, то у контактному режимі АСМ дасть зображення поверхні, а в безконтактному – зображення адсорбованого шару води на поверхні, так що зображення в двох режимах можуть розрізнятися.

Розрізнявальна здатність АСМ різна в різних режимах *роботи*. Оскільки сили взаємодії атомів вістря і поверхні дуже швидко зменшуються з відстанню, то вертикальне розрізнення обмежене тільки власними шумами системи детектування і тепловими флуктуаціями кантилевера. Вертикальне розрізнення може бути не гірше за 0,1 нм.

У горизонтальній площині розрізнення залежить від радіуса вістря зонда, відстані між зондом і поверхнею, сил взаємодії, пружності зразка й інших факторів. В даний час за допомогою АСМ отримані зображення кристалічних поверхонь з реальним атомним розрізненням.

Також як і СТМ, АСМ *застосовується у нанотехнологіях* – як інструмент для локальних модифікацій поверхонь і для нанолітографії. Таким чином, у даний час АСМ широко використовується як багатофункціональний аналітичний інструмент для дослідження і модифікації структури поверхонь, для вивчення розподілу приповерхніх силових і температурних полів, розподілів величин-характеристик фізичних властивостей з нанометровим або навіть з атомним розрізненням.

Для дослідження і діагностики *напівпровідникових матеріалів* і структур використовуються як СТМ, так і АСМ, але кожен прилад має свої переваги. СТМ, як правило, дає трохи кращий розрізнення, ніж АСМ, хоча атомне розрізнення досягається і за допомогою АСМ. Атомно-силовий мікроскоп дає зображення *реальної поверхні*, скануючий тунельний мікроскоп – *густину електронних станів*.

За допомогою АСМ не можна одержати інформацію про електронні властивості, що мало би велику цінність у прив'язці до топології поверхні. Для СТМ необхідні добре провідні підкладки; при дослідженні високоомних і напівізолюючих напівпровідників можуть виникнути труднощі – система зворотного зв'язку буде рухати зонд до торкання з поверхнею (щоб забезпечити необхідну для роботи мікроскопа величину тунельного струму). Тому структуру непровідних поверхонь можна досліджувати тільки за допомогою АСМ.

9.4. Самоорганізація структур у нанотехнологіях

Упорядковані самоорганізовані масиви квантових точок були отримані для якісних лазерних структур. Відповідно до теорії (див. розділ 8.5), квантові точки мають дискретний енергетичний спектр і високу щільність станів. Використання квантових точок у якості активного середовища лазерів дає великий коефіцієнт підсилення, низький граничний струм і температурну стабільність лазера.

Протягом довгого часу робилися спроби виготовлення квантових точок і приладів на їхній основі традиційними методами, наприклад літографією і селективним травленням структур із квантовими ямами. На цьому шляху відпрацьовувалася літографія з гранично високим розмежуванням, однак квантових точок з дискретним атомоподібним спектром отримати не вдалося.

Необхідні квантові точки були отримані тільки після зміни технології по типу “зверху вниз” на технологію “знизу нагору”, коли стали використовувати ефекти самоорганізації при епітаксiальному нарощуванні гетероструктур. За певних умов у епітаксiальному шарі, нарощуваному на кристалічну підкладку іншої речовини, утворюються упорядковані масиви квантових точок. Приклад такої квантової точки германія на поверхні кремнію був показаний на рис. 8.9.

На квантових точках, отриманих за допомогою методом самоорганізації, були вперше продемонстровані фізичні властивості, передбачені теоретично, зокрема, специфічна *дискретність* їх енергетичного спектру.

Механізм процесу самоорганізації. Процес самоорганізації при епітаксiальному рості є наслідком прагнення системи до зменшення енергії. Режим формування поверхневих структур визначається *неузгодженістю параметрів* кристалічної ґратки підкладки і нанесеного матеріалу, а також співвідношенням поверхневої енергії й енергії границі розділу цих матеріалів. Важливо відзначити, що всі енергетичні аргументи справедливі тільки для рівноважного стану системи.

Експериментально встановлено, що при осадженні атомів з газової фази на підкладку спостерігаються три основних механізми початкової стадії росту:

- пошаровий ріст;
- острівний ріст;
- проміжний механізм росту.

1. Постійні кристалічної ґратки підкладки і шару, що осаджується на підкладку, практично збігаються; матеріал, що осаджується, змочує підкладку,

відбувається пошаровий **двовірний** ріст (механізм Франка-Ван-дер-Мерве, рис. 9.11, *а*)

2. Матеріал, що осаджується, не змочує підкладку (матеріали, різні по властивостях, або з великим розходженням постійних ґрат). Відбувається острівний (**тривимірний**) ріст. Матеріал, що осаджується, стягається в наноострівці на поверхні підкладки (рис. 9.11, *б*, механізм Фольмера-Вебера).

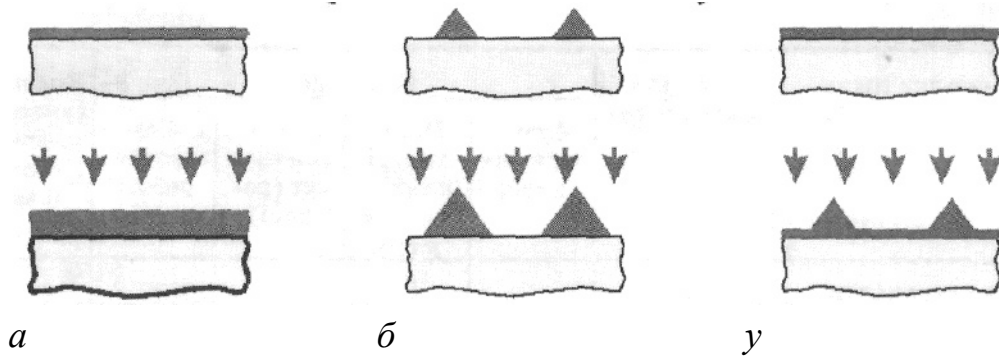


Рис. 9.11. Три механізми епітаксiального росту плівки: *а* - механізм Франка-Ван-дер-Мерве; *б* – механізм Фольмера-Вебера; *в* – механізм Странського-Крастанова

3. Матеріал, що осаджується, змочує підкладку, але спостерігається неузгодженість постійних ґраток (порядку декількох відсотків). У результаті спочатку плівка росте пошарово, а потім структура трансформується в острівкову (рис. 9.11, *в*). Це – комбінований режим росту (механізм Странського-Крастанова).

Самоорганізація квантових точок. Саме комбінований механізм росту використовується для одержання **масивів** квантових точок, наприклад квантових точок InAs у матриці GaAs (неузгодженість ґрат 7%) або квантових точок германія в кремнієвій матриці (неузгодженість 4%). На початковому етапі йде пошаровий ріст матеріалу, що осаджується, на підкладці з утворенням змочувального шару, а потім відбувається перехід до формування тривимірних острівців з матеріалу, що осаджується на частково покрити підкладку. Кожна вертикальна атомна площина підкладки продовжується в об'ємі острівця, але через розходження постійної ґратки обох матеріалів острівець стає напруженим (використовується термін “когерентно-напружений”).

Постійна ґратки InAs більше, ніж у GaAs. При осадженні InAs на підкладку GaAs спочатку формується шар InAs. Цей шар через розходження ґраток – напружений (комірки як би прагнуть вигнутися). В міру збільшення товщини шару пружна енергія росте, зв'язки між атомами шару починають

рватися, деякі атоми частково звільняються. Відбувається перерозподіл матеріалу, що осаджується, і утворюються тривимірні острівці (рис. 6). Коли острівець утворився, ґратка InAs частково розпрямляється і за рахунок цього виходить вигравш в енергії. Утворення тривимірних острівців починається після осадження приблизно двох шарів InAs.

Після осадження чотирьох моносарів створюється щільний масив острівців правильної форми. Якщо на острівці InAs знову наростити GaAs, то регулярно розташовані квантові точки InAs (вузькозонний напівпровідник) будуть вміщені у матрицю GaAs (широкозонний напівпровідник).

Аналогічне формування масивів острівців спостерігається також у системах Si-Ge, InAs-In, AlInAs-AlGaAs, PbSe-PbTe і в ряді інших.

Розмір і форма острівців. Для заданих умов росту існує визначений розмір острівців, що відповідає мінімумові енергії системи. Збільшення або зменшення розмірів острівців приводить до збільшення енергії. Більшість острівців має однаковий розмір, що залежить від умов росту. Приклад реалізації самоорганізованого масиву квантових точок наведений на електронній фотографії рис. 9.12.

Острівці мають грані – поверхні з малою поверхневою енергією (так називані *кристалографічні грані*). Наприклад, на підкладці GaAs з орієнтацією (100) рівноважна форма острівців InAs – це пірамідки з квадратною підставою.

Коли відстань між острівцями стає порівняною з розмірами острівців, починається їх пружна взаємодія, обумовлена проникненням у підкладку напруження, створюваного острівцями. Взаємодія острівців на поверхні – завжди їхнє відштовхування. Це забезпечує стійкість масиву острівців. На пружних-анізотропних поверхнях (100) острівці утворюють квадратну сітку.

Для застосування в приладах необхідна однорідність масиву острівців за формою, взаємному розташуванню і розмірам острівців, а також достатня поверхнева щільність острівців. Наприклад, для лазерних структур необхідна щільність острівців близько 10^{11} см^{-2} . Це означає, що на квадратному мікроні розміщена тисяча квантових точок.

Регулюванням умов росту можна змінювати розміри острівців і їхню поверхневу щільність. Наприклад, з ростом температури підкладки збільшується розмір основи острівця, зменшується його висота і зменшується поверхнева щільність острівців.

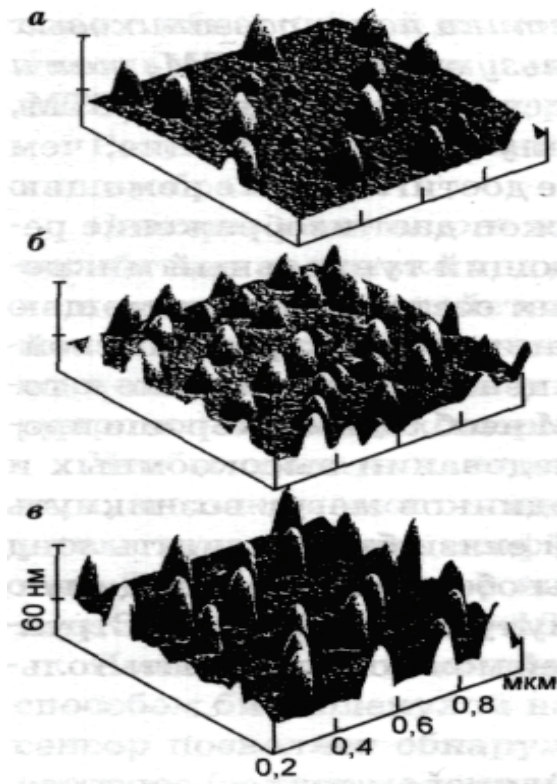


Рис. 9.12. Самоорганізовані квантові точки германію на поверхні кремнію, вирощені при температурах: *a* – 600 °С, *б* – 700 °С, *в* – 750 °С

Ці параметри залежать також від тиску пари миш'яку й інших умов росту. Підбір оптимальних умов одержання острівців робиться експериментально. Існують мінімальний і максимальний розміри острівців, коли їх можна розглядати як квантові точки. Для системи InAs-GaAs цей розмір складає від 4 нм до 20 нм. Мінімальний розмір визначається тим, що квантова точка повинна мати хоча б один енергетичний рівень; а максимальний розмір визначається необхідністю виконання умови $E_2 - E_1 \gg k_B T$.

Іншими словами, нагромадження енергії у напруженому стані в міру збільшення товщини плівки неминуче веде до утворення острівців, що знижують загальну енергію в системі. Таке перетворення відбувається при епітаксії у режимі Странського-Крастанова. При цьому утворюються самозбирані нанокристалічні епітакисальні острівці на монокристалічній підкладці. Перехід від двомірного шарового росту суцільної плівки до тривимірного росту острівців ($2D \rightarrow 3D$ перехід) має місце, коли міжатомні відстані в кристалічних ґратках матеріалу, що осаджується, більше, ніж у ґратках підкладки. Острівна структура енергетично більш сприятлива для релаксації механічної напруги, ніж шарувата.

Вертикальні масиви квантових точок. Ці структури найбільш ефективні для оптоелектроніки. Якщо перший шар квантових точок-острівців закрити декількома моношарами матричного матеріалу і знову осаджувати матеріал острівців, то нові острівці – квантові точки – будуть формуватися точно над острівцями попереднього шару. Це узгоджується з тим, що через розходження ґраток матриці і матеріалу, який осаджується, простір над острівцями виявляється напруженим. Наприклад, у матриці GaAs над острівцем InAs створюється максимальне розтягання. Оскільки атоми індію мають більший радіус, ніж атоми галію, то при осадженні атоми індію притягаються до ділянок поверхні GaAs з максимальним розтяганням, тобто до ділянок над загоєними острівцями попереднього шару.

Так одержують вертикальні масиви квантових точок для приладових структур. Регулюючи товщину проміжних шарів GaAs, можна одержати тунельно-зв'язані і електронно-зв'язані квантові точки у вертикальних масивах.

Створення квантових ниток самозборкою. У системі, утвореній матеріалами з погодженими параметрами ґраток, режим росту визначається тільки співвідношенням поверхневої енергії й енергії границі розділу. Якщо сума поверхневої енергії епітаксiальної плівки й енергії границі розділу менше, ніж поверхнева енергія підкладки (тобто матеріал, що осаджується, змочує підкладку), то має місце пошаровий ріст плівки в режимі Франка-Ван-дер-Мерве. При цьому формуються однорідні, когерентні і напружені надґратки.

Цей режим придатний також і для створення квантових ниток, що самоорганізуються, на віцінальних поверхнях кристалів.

Віцінальними називають поверхні, що не є рівноважними для даного кристала. Звичайно це поверхні, злегка розорієнтовані щодо площини кристала (на практиці найчастіше використовують розорієнтацію відносно (001) і (311) площин).

Основні етапи формування методом самозбирання убудованих квантових ниток з використанням віцінальної поверхні кристала показані на рис. 9.13.

Підготовлена віцінальна поверхня складається з однакових дистанційно розташованих терас, площа яких збігається з однією з площин кристала. Висота кожної тераси відповідає одному моношару.

Створення квантових ниток починається з нанесення матеріалу, з якого буде формуватися нитка. Температуру підкладки вибирають таким чином, щоб забезпечити достатню дифузійну рухливість атомів матеріалу нитки, який

осаджується на поверхні підкладки. Новим атомам енергетично більш вигідно прикріплюватися до сходинки тераси, ніж знаходитися на її поверхні.

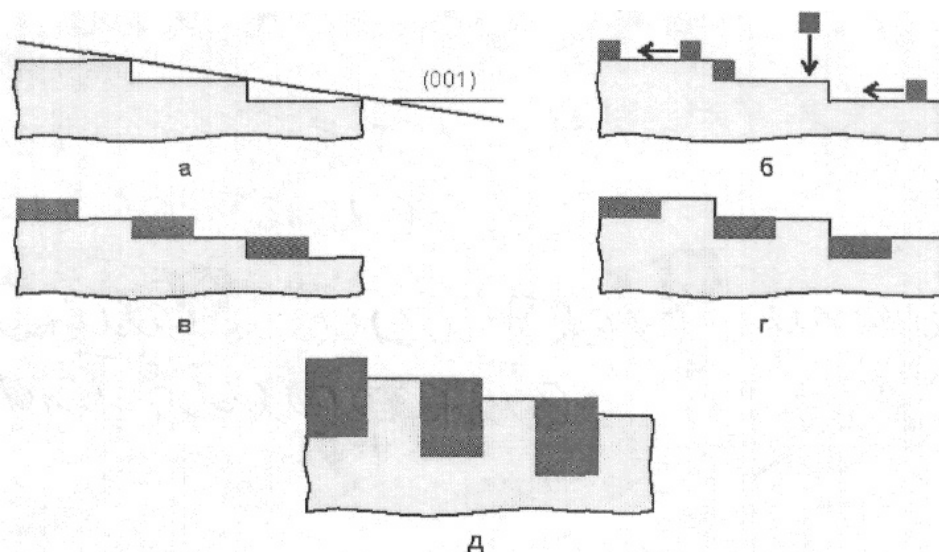


Рис. 9.13. Створення квантових ниток самозборкою в процесі епітаксiального росту на віцінальній поверхні: *а* – підготовлена віцінальна поверхня; *б* – нанесення матеріалу нитки; *в* – нанесення половини моношару матеріалу нитки; *г* – додавання матеріалу підкладки до повного моношару; *д* – вбудовані квантові нитки

Кількість матеріалу, що осаджується, складає лише частку від тієї величини, що необхідна для покриття підкладки суцільним моноатомним шаром. Це необхідно для того, щоб залишити місце на терасі для заповнення матеріалом підкладки, що і робиться після припинення подачі матеріалу нитки. Наступне нанесення матеріалу підкладки продовжують до повного відновлення терас, які при цьому стають більшими на один моноатомний шар. Нанесення матеріалу нитки, а потім підкладки повторюють кілька разів, забезпечуючи, таким чином, створення квантової нитки, вбудованої у підкладку.

9.5. Напівпровідникові гетеропереходи

Гетеропереходи відіграють провідну роль у сучасній мікро- та нанотехнології напівпровідників. Існують різні можливості створення *p-n* переходу у напівпровіднику, але усі вони звичайно реалізуються в *тому ж самому матеріалі*, легованому по-різному – акцепторами і донорами. У подальшому викладі такі переходи можна назвати *монопереходами*, наприклад, *p-Si – n-Si*.

Розглянуті далі **гетеропереходи** виникають при контакті **двох різних по хімічному складі напівпровідників**. У такому контакті відбувається не тільки зміна ширини забороненої зони, але міняються й інші фундаментальні

властивості: зонна структура, ефективні маси носіїв заряду, їхня мобільність, фізико-хімічні й оптичні властивості.

Гетеропереходи можуть бути різкі і плавні, анізотипні ($p-n$ гетеропереходи) і ізотипні ($p-p$ і $n-n$ гетеропереходи). У *різкому* гетеропереході зміна хімічного складу відбувається на відстані, меншій ширини області об'ємного заряду переходу. В ідеальному гетеропереході на границі розділу відсутні дефекти і граничні стани.

Можливість одержання монокристалічних гетеропереходів, тобто контактів різних по хімічному складі напівпровідників, здійснених **в одному монокристалі**, зв'язана з розвитком технологічних методів епітаксійного нарощування шарів напівпровідникових кристалів, тобто утворення однаково орієнтованих шарів кристалів однієї речовини на поверхні іншої речовини.

Гетероструктурою називається комбінація декількох гетеропереходів в одній монокристалічній структурі, що складає частину напівпровідникового приладу. При безлічі шарів кристалів, що чергуються, утворюється планарна періодична структура, що називається **надграткою**. Характерні розміри шарів у гетероструктурах і надгратках – нанометри, й тому напівпровідникові прилади відносяться до нового покоління приладів в електроніці – **наноелектронних**.

В даний час отримано багато напівпровідникових структур іншого типу, які також мають розміри порядку нанометрів. Тому надгратка є лише одним прикладом планарної (двовимірної, $2D$) **наноструктури**. Іншим прикладом двовимірних наноструктур є **квантова яма** (див. розділ 1.3). Ця назва розуміється в тому плані, що через малі розміри структури енергетичний спектр електронів у квантовій ямі квантується. Були успішно виготовлені і наноструктури, розміром менше двох нанометрів. Як уже згадувалося вище, отримані одномірні ($1D$) кристалічні наноструктури, називаються **квантовими нитками** (інакше – *шнурами, дротами, проводами*). Реалізовано також і «нульмірні» ($0D$) кристаліти нанометрових розмірів, що називаються **квантовими точками**. Комбінуючи різні напівпровідники, можна створити й інші цікаві структури, наприклад, структури з багатьма квантовими ямами.

Для напівпровідникових приладів мікрохвильової електроніки найбільш істотне значення мають транзистори з гетеропереходами і генераторні діоди типу ЛПД (лавинно - пролітні діоди) і діода Ганна, де використовуються саме гетеропереходи, що містять модульовано-леговані гетероструктури. Вони

дозволяють одержати високорухливий двовимірний електронний газ. На цій основі одержують польові і біполярні НВЧ транзистори.

В даний час розроблені і розробляються безліч НВЧ приладів з необхідним для високої мобільності електронів квантовим розмірним ефектом. У техніці виготовлення і контролю форми розриву зон досягнуто великого прогресу. Наприклад, у добре вивченій системі GaAs/GaAlAs товщина інтерфейсів між GaAs і GaAlAs може складати усього один моношар, що було показано експериментально за допомогою просвітлюваної електронної мікроскопії.

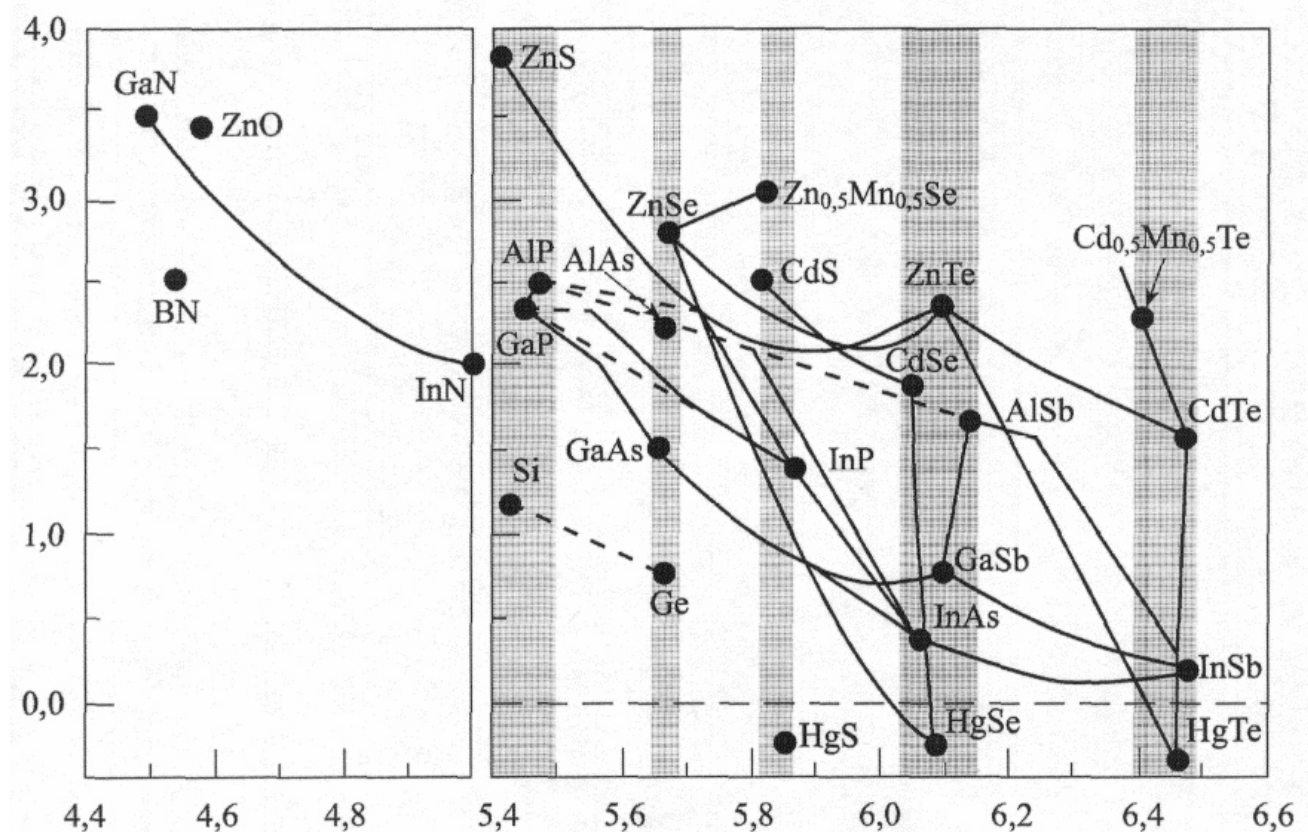


Рис. 9.14. Графік залежності енергії забороненої зони при низькій температурі від постійних ґрат для ряду напівпровідників зі структурою алмаза і цинкової обманки. Затінені області поєднують групи напівпровідників із близькими постійними ґрат. Напівпровідники, з'єднані суцільними лініями, утворюють між собою стабільні тверді розчини. Штрихові лінії вказують на непрямі заборонені зони.

Одержання різких інтерфейсів накладає жорсткі обмеження на умови росту – такі як чистота вихідних матеріалів, температура підкладки і багато інших умов, котрі тут не перелічуються. Однак, у кінцевому рахунку, якість інтерфейсу між двома різними матеріалами **A** і **B**, названого гетеропереходом, визначається їх хімічними і фізичними властивостями. Найбільш важливим параметром є розходження між постійними кристалічних ґраток. Якщо сталі

граток рівні, то всім атомам матеріалу **A** легко підбудуватися щодо всіх атомів **B**. Таке підстроювання ґрат називається *псевдоморфним ростом*, що вкрай бажано для досягнення високої якості гетеропереходів. Існує усього кілька систем, у яких постійні ґраток дуже близькі.

На рис. 9.14 приведена залежність ширини забороненої зони від параметрів ґраток – як для кристалів типу алмазу, так і для напівпровідників типу $A^{III}B^V$. Затінені вертикальні ділянки охоплюють групи напівпровідників із близькими по величині сталими ґраток, але з різною шириною заборонених зон. Це вкрай сприятливо для одержання гетеропереходів з керованою величиною розриву зон.

Вибір величини розриву зон може бути розширений за допомогою використання в технології *твердих розчинів* – подвійних (таких як SiGe), потрійних (AlGaAs) і навіть четвірних (GaInAlAs). Суцільні лінії на рис. 9.13, що з'єднують деякі напівпровідники, показують, що ці матеріали утворюють *стабільні тверді розчини* у всьому діапазоні концентрацій (наприклад, InGaAs, GaAlAs і InGa). Саме їх часто використовують при синтезі різних гетероструктур для приладів НВЧ.

Керуючись приведеним рисунком, можна планувати створення гетеропереходів “на замовлення” з бажаною величиною розриву зон або квантових ям із заданою формою потенціалу.

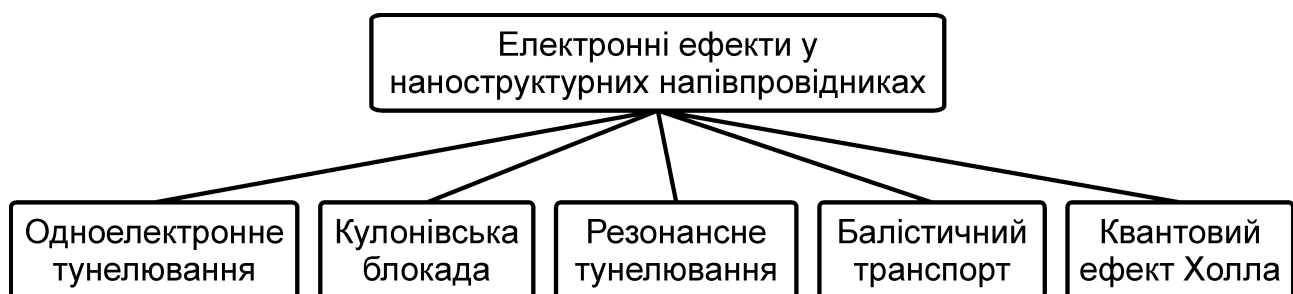
10. Властивості напівпровідникових низькорозмірних структур

Перенос носіїв заряду у напівпровідникових наноструктурах обумовлений, з одного боку, хвильовою природою електрона (безперервність, яка властива хвилі), а з іншого боку – дискретністю їх заряду і спіну. Спостерігається **квантовий режим переносу носіїв заряду**, на відміну від класичного режиму, у якому довжина вільного пробігу електронів набагато менше, ніж розміри системи, і електрони розглядаються як класичні заряджені частини.

У даному посібнику розглянуті основні фундаментальні фізичні явища, які визначають транспорт (перенос) носіїв заряду у нанорозмірних структурах. Головний акцент зроблений на те, що розміри наноструктур порівнянні з довжиною вільного пробігу електронів. Саме це приводить до того, що носії заряду можуть вільно проходити через наноструктуру без розсіювання на дефектах, домішках, фононах. В таких структурах фази невзаємодіючих електронних хвиль зберігаються протягом усього шляху і тому спостерігаються ефекти фазової інтерференції. У табл. 10.1 перераховані основні фізичні ефекти у низькорозмірних структурах, які можуть бути використані у нанoeлектроніці.

Таблиця 10.1.

Основні квантові ефекти у низькорозмірних структурах



10.1. Потенціальні ями, бар'єри і тунелювання

Для аналізу властивостей квантових систем використовуються їх енергетичні моделі. Наночастинки за їх дискретним спектром можна порівняти з атомами. Тому на рис.10.1 наведений енергетичний спектр найпростішого атома – водню. Потенціальна енергія в цьому випадку описується законом Кулона: $U(r) = -e^2/4\pi\epsilon_0 r$, де r – відстань від електрона до ядра. Точці А відповідає перша стаціонарна борівська орбіта з головним квантовим числом n

$= 1$ і енергією $E_1 = -13,6$ еВ. Радіус цієї орбіти $r_1 = 0,053$ нм. У стаціонарному стані ($n = 1$) імовірність перебування електрона на цій орбіті максимальна. Інші стани (при $n = 2, 3 \dots$) з більшою енергією ($E_2, E_3 \dots$) є збудженими, рис. 10.1, б. Якщо електронів в основному стані E_1 надати енергію, більшу або рівну $13,6$ еВ, то електрон стане вільним.

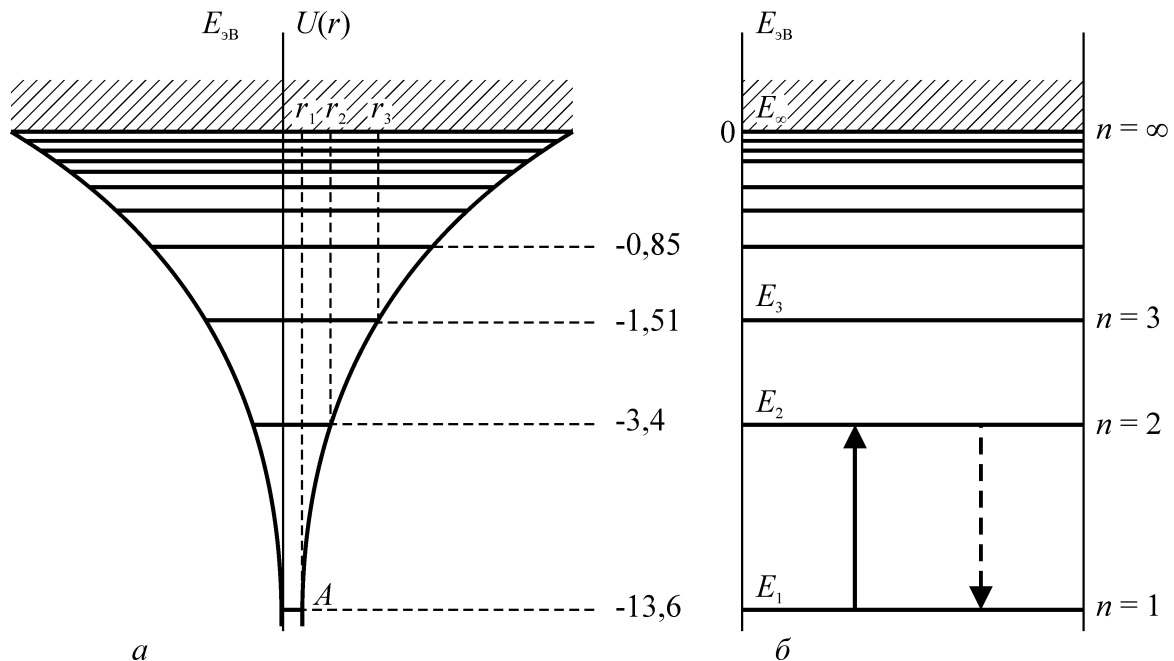


Рис. 10.1. Енергетична модель атома водню

Енергетичний підхід спирається на закон збереження енергії і тому є універсальним. Графік 10.1, а дає уявлення про потенціальну яму, висота стінок якої являє собою потенціальний бар'єр. Метод потенціальних кривих дозволяє при відомих значеннях розподілу енергії визначити динамічні параметри частинки.

Відповідно до класичної механіки тіло здатне перебороти потенціальний бар'єр тільки за умови, що його енергія перевищує потенціальну енергію на вершині бар'єра. Зовсім інакше поводить квантова частинка. Для неї існує відмінна від нуля імовірність проникнути за бар'єр навіть за умови, що енергія цієї частинки менше висоти потенціального бар'єра. Такий квантовомеханічний ефект називається *ізоенергетичним тунелюванням*.

З погляду класичної фізики при тунельному ефекті порушується закон збереження енергії. Для квантової фізики тунельний ефект можливий за умови, що “дефіцит” енергії частини ΔE переборюється протягом малого часу $\Delta t = \hbar/(2\Delta E)$ – відповідно до співвідношення невизначеностей. Тому, з погляду квантової фізики, якщо саме за такий час частина тунелює крізь бар'єр, то закон

збереження енергії *не порушується*. Справа в тому, що після тунелювання частинка *зберігає свою енергію*.

Тунельний перехід через заборонену зону напівпровідника у сильному електричному полі розглянутий у розділі 4.13. Тунельний ефект лежить в основі дії тунельного діода і широко використовується в інших приладах мікро- та наноелектроніки.

Особливість поведінки квантової частинки в потенціальній ямі нано- і атомарних розмірів полягає в тому, що її **енергетичний спектр є дискретним** (подібним до рис. 10.1, б). Саме дискретність енергетичного спектру електронів лежить в основі функціонування багатьох наноелектронних структур. Конфігурації реальних потенціальних ям і бар'єрів у мікросвіті залежать від геометричних особливостей тих полів, що утворюють ці ями (подібні до потенціальної ями для електрона, який рухається в електричному полі ядра атома водню).

Поведінку мікрочастинки у потенціальній ямі можна визначити розв'язуючи рівняння Шредінгера за відповідних початкових умов. Це рівняння є диференціальним; для його розв'язку необхідно знати, як залежить потенціальна енергія U мікрочастинок від координат, тобто повинна бути задана функція $U(x, y, z)$. Тут і далі будемо припускати, що ця функція не залежить від часу (тому що мікрочастинка знаходиться у стаціонарному стані).

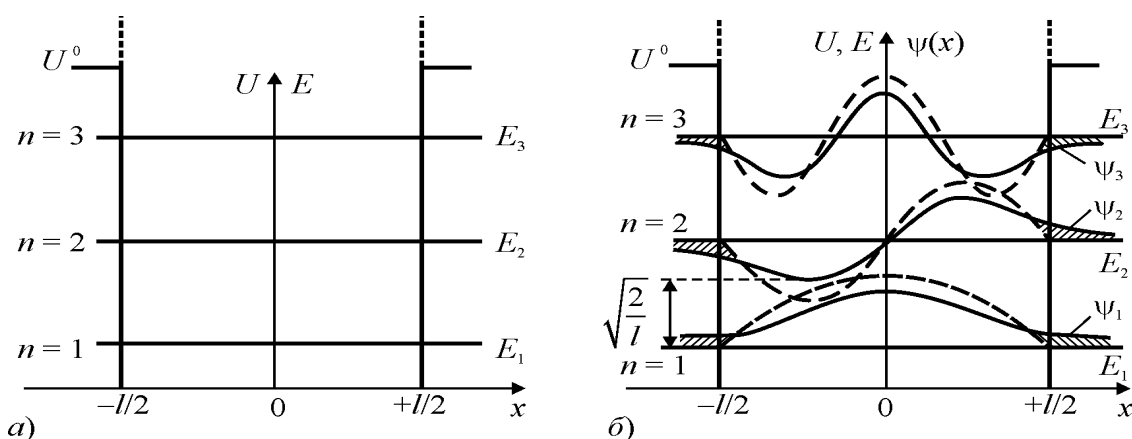


Рис. 10.2. Одновимірна потенціальна яма: а – енергетичний спектр мікрочастинки в потенціальній ямі (приведені три енергетичних рівні E_1, E_2, E_3); б – хвильові функції $\Psi(x)$ для трьох станів ($n = 1, 2, 3$) мікрочастинки у потенціальній ямі (пунктир відповідає нескінченно глибокої ями).

Для з'ясування основних особливостей поведінки мікрочастинки у потенціальній ямі досить розглянути *одновимірний* випадок $U(x)$ і потенціальну яму *прямокутної форми* (рис. 10.2, а). Ширина потенціальної ями позначена як

l , а її глибина дорівнює U^0 . У цій моделі ширина бар'єрів, які обмежують потенціальну яму праворуч і ліворуч, *нескінченна*. Глибина ями U^0 може змінюватися від деякого кінцевого значення до нескінченності (тоді яму можна вважати необмежено глибокою).

Розв'язок рівняння Шредінгера (результат якого показаний графічно на рис. 10.2, б) дає можливість визначити енергетичний спектр мікрочастинки, тобто повний набір значень її енергії E і хвильових функцій $\Psi(x)$, квадрат модуля якої $|\Psi(x)|^2$ є густиною імовірності знайти мікрочастинку у точці x .

Можна показати також, що на ширині ями l укладається **ціле число півхвиль** де Бройля:

$$l \approx n(\lambda/2)$$

Для *нескінченно глибокої ями* хвильова функція виражається через тригонометричні функції: через косинуси при непарному n ($n = 1, 3, 5$) і через синуси при парному n , що представлено на рис. 10.2, б *пунктирними кривими*. З рисунку видно, що в розглянутому випадку амплітуда хвилі де Бройля у точках $x = \pm l/2$ звертається у нуль. Отже, коли $U^0 \rightarrow \infty$ мікрочастинка не може ні проникнути усередину бар'єра, ні вийти за межі ями. Однак якщо глибина ями *кінцева*, то амплітуда хвилі де Бройля в точках $x = \pm l/2$ не звертається в нуль при будь-яких n і **має продовження за межами бар'єра**.

На рис. 10.3 цей важливий результат показаний суцільними кривими в заштрихованих областях. Іншими словами, мікрочастинка в потенціальній **ямі кінцевої глибини** U^0 може проникати за межі бар'єра при енергіях E , менших за U^0 , що суперечить законові збереження енергії класичної фізики і не спостерігається в макросвіті.

Крім того, з рис. 10.3, де показана густина імовірності при трьох різних значеннях n , видно, що енергетичний спектр квантової частинки дискретний, причому її мінімальна енергія не дорівнює нулеві ($E_1 > 0$). Ця енергія називається **нульовою** і звичайно позначається як E_0 . Наявність нульової енергії властива будь-яким квантовим системам: фізичному вакуумові, кваркам в адронах, нуклонам у ядрах атомів, електронам в атомах, атомам у молекулах і кристалах.

З рис. 10.3 випливає також, що імовірності знайти мікрочастинку в різних точках усередині ями істотно неоднакові. Спостерігаються точки, імовірність «відвідування» яких частинкою максимальна. Вони називаються *пучностями* (x_n). Є і такі точки, у яких частинка не буває, це – *вузли* (x_y). Таке поводження не

властиво **макрочастинкам**.

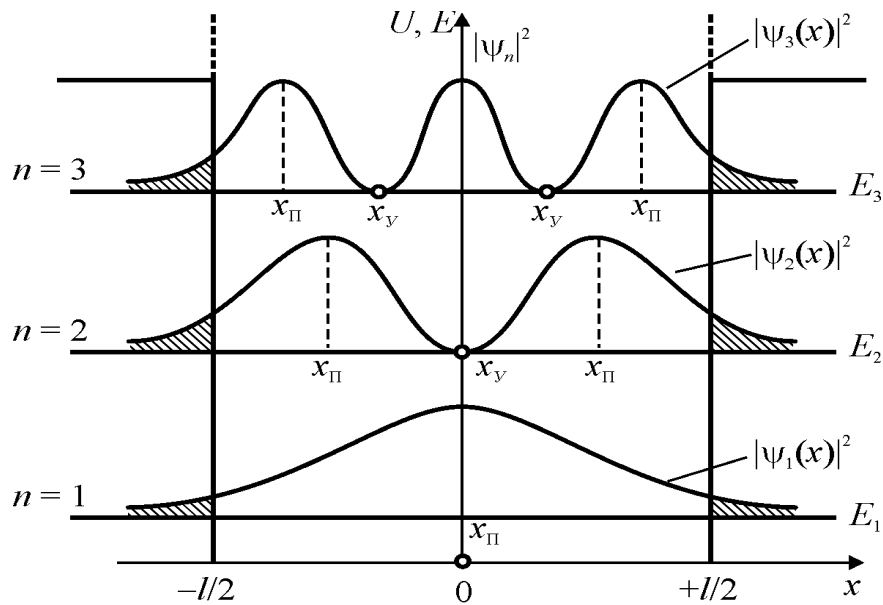


Рис. 10.3. Густина імовірності $|\Psi(x)|^2$ знайти мікрочастинку в різних точках x для випадку потенціальної ями кінцевої глибини

Для подальшого особливо важливо те, що імовірність проникнення мікрочастинки **за межі бар'єра** ($x = \pm l/2$) не дорівнює нулеві, а лише поступово знижується зі збільшенням відстані від границі бар'єра (заштриховані області на рис. 10.3). Якщо ширина бар'єра не нескінченна, то є відмінна від нуля імовірність проникнення мікрочастинки за межі бар'єра (тунельний ефект).

Тунельний ефект лежить в основі дії багатьох схемних елементів наноелектроніки. Тому особливості цього ефекту розглядаються далі більш докладно.

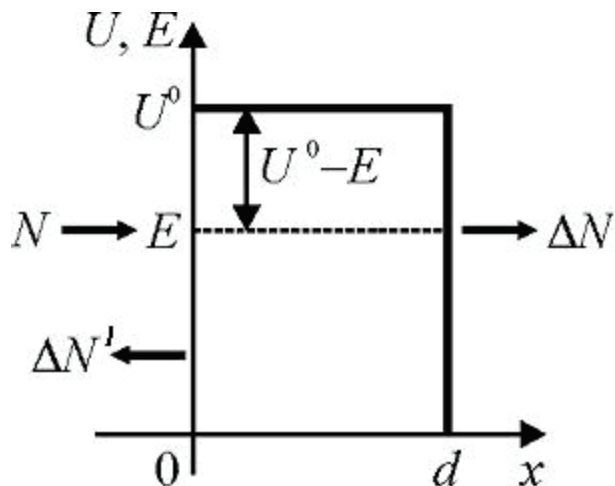


Рис. 10.4. Прямокутний потенційний бар'єр: $\Delta N'$ – число невдалих спроб мікрочастинки перебороти бар'єр ($N = \Delta N + \Delta N'$); d – ширина бар'єра і U^0 – його висота; $(U^0 - E)$ – «дефіцит» енергії частини усередині бар'єра,

Пояснення тунельного ефекту зручно дати на прикладі, показаному на рис. 10.4, коли потенціальний бар'єр має прямокутну форму.

Імовірність проникнення мікрочастинки масою m за межі потенціального

бар'єру можна виразити формулою:

$$W = W_0 \exp\left(-\frac{2}{\hbar} \sqrt{2m^*(U_0 - E)} \cdot d\right),$$

де W – **коефіцієнт прозорості** потенціального бар'єра – дорівнює відношенню числа вдалих спроб частини проникнути за бар'єр ΔN до загального числа таких спроб N ($W = \Delta N/N$). З приведеної формули випливає, що імовірність W тунельного ефекту зменшується зі збільшенням маси частини m , товщини потенціального бар'єра d і дефіциту енергії ($U^0 - E$).

10.2. Одноелектронне тунелювання і кулонівська блокада

Кулонівською блокадою називається таке явище, коли струм не проходить через тунельний перехід навіть при наявності зовнішньої електричної напруги, якщо тунелюванню електронів перешкоджає їх кулонівська взаємодія. Це фізичне явище лежить в основі досить перспективних для майбутньої електроніки **одноелектронних приладів**. Означені пристрої становлять особливий інтерес у зв'язку з наближенням сучасної технології до межі мініатюризації елементів класичних інтегральних мікросхем. Одноелектронні прилади забезпечують подальше збільшення швидкодії, зниження габаритів і енергоспоживання електронної апаратури.

У мікроелектронних приладах логічні елементи, що спрацьовують на визначену величину «зарядового пакета», який складається з деякого числа електронів (або дірок), можуть занадто часто помилятися, якщо цей «пакет» виявиться *недостатньо великим*. (У сучасних пристроях у такому «пакеті» використовується більше ніж 10^4 електронів). Так, при кодуванні логічної одиниці пакетом з десяти зарядів з порогом спрацьовування у п'ять зарядів логічний елемент буде неправильно спрацьовувати приблизно в 3% випадків, що неприйнятно для техніки. Треба зауважити, що у структурі нанометрового розміру неможливо істотно збільшити надлишковий заряд. Наприклад, у сферичному кластері радіусом 2 – 3 нм можна розмістити лише кілька зайвих електронів.

З цим пов'язана межа мініатюризації звичайної електроніки. Уже зараз мікроелектронною промисловістю створюються транзистори з розміром робочих елементів 20 – 30 нм. Вони ще здатні працювати зі звичайними електричними сигналами, однак при подальшому зменшенні розмірів дуже швидко нарастають проблеми, про які згадано вище.

Відповідно до закону Мура, повне освоєння області мезоелектроніки

очікується приблизно через десять років. Таким чином, мезотранзистори – це останній рубіж існування звичайних транзисторів, за яким піде покоління нанотранзисторів, у тому числі – одноелектронних приладів. Дослідження таких приладів активно проводиться вже зараз.

Область від 30 нм до 5 нм (область мезоструктур) варто вважати перехідною від класичної твердотільної електроніки до квантової. Промисловість упритул підійшла до цієї області і вже зіштовхнулася з низкою труднощів.

Одноелектронними називаються такі прилади, у яких контролюється переміщення одного електрона або малої їхньої кількості. У цих приладах переміщення електронів відбувається за допомогою тунельного ефекту. Оскільки час тунелювання малий, то й межа швидкодії таких приладів висока (до 10^{14} Гц). Робота, яка необхідна для переміщення одного електрона мала, і тому енергоспоживання та тепловиділення таких приладів повинні бути низькими (теоретично – біля 10^{-8} Вт для одного елемента).

У сучасних обчислювальних **пристроях перемикання** (тобто перехід між станами «0» і «1») вимагає участі $10^4 - 10^5$ електронів. Як уже говорилося, розрахунки показують, що по класичній мікроелектронній технології вже неможливо буде одержати стійке перемикання при кількості електронів близькій до 10^2 . Навпаки, по нанотехнологіях у даний час створені макети транзисторів, у яких перемикання відбувається під дією лише одного електрона, а також запам'ятовуючі пристрої, у яких носієм біта інформації є один електрон.

Тунелювання у низькорозмірних структурах. Як відомо, електричний струм у провіднику обумовлений рухом електронів щодо нерухомих іонів кристалічної ґратки. Хоча кожен електрон несе дискретний одиничний заряд e , загальний перенесений електронами заряд змінюється не стрибкоподібно, як можна було б очікувати, а безупинно, оскільки цей заряд визначається сумою зсувів всіх електронів щодо іонів ґратки. Вільні електрони у провіднику можуть зміщуватися на як завгодно малі відстані. Тому сумарний заряд, який переносять електрони, в об'ємному провіднику (або напівпровіднику) змінюється монотонно – дискретність у переносі заряду спостерігати неможливо, оскільки вона менше, ніж одиничний заряд електрона.

У низькорозмірних структурах **дискретність транспорту заряду** є типовою властивістю; більш того, саме ця дискретність і може бути фізичною основою дії деяких наноелектронних приладів.

Тунелювання електрона через потенціальний бар'єр – це типовий прояв його хвилових властивостей. У низькорозмірних структурах дане явище здобуває специфічні особливості, пов'язані не тільки з дискретністю заряду, який переносить електрон, але й з додатковим квантуванням енергетичних станів у таких структурах, викликаним квантовим обмеженням. Найбільше характерно ці особливості виявляються в одноелектронному тунелюванні й у резонансному тунелюванні.

Одноелектронне тунелювання. Спочатку уявимо просту наноструктуру, що складається з двох областей провідника, розділених тонким діелектриком. У такій структурі електричний заряд переноситься комбіновано – безперервно у провіднику і дискретно через діелектрик.

Вважаємо, що границя розділу між провідником і діелектриком електронейтральна. При наявності на зовнішніх контактах структури електричного потенціалу починається безперервна зміна заряду в провіднику, що приводить до нагромадження зарядів у провіднику на границі з діелектриком. Цей заряд виникає внаслідок невеликих безперервних зсувів електронів (щодо своїх рівноважних положень) у прикордонній області провідника.

Нагромадження заряду продовжується доти, поки його величина не виявиться достатньою для відриву і тунелювання через діелектрик *одного* електрона. Після акта тунелювання система повертається у первісний стан. При збереженні зовнішньої (прикладеної) напруги усе повторюється знову.

При тунелюванні у «прямому» напрямку електрон здобуває можливість тунелювати через діелектрик тоді, коли накопичений заряд стає більше ніж $+e/2$, оскільки тільки при цьому зменшується електростатична енергія системи (при тунелюванні у «зворотному» напрямку заряд повинний бути менше $-e/2$). У середині інтервалу $+e/2$ і $-e/2$ тунелювання неможливе через кулонівську взаємодію електрона з іншими рухливими і нерухомими зарядами в провіднику. Дане явище називають **кулонівською блокадою**.

Структура з одним тунельним бар'єром може бути представлена еквівалентною схемою, показаною на рис. 10.5, а. Її основними елементами є ємність утвореного діелектриком конденсатора C_t і тунельний опір R_t , що характеризує тунельний перехід під час відсутності зарядових ефектів. Ємність C_e є еквівалентною для електродів і джерела напруги V і враховує зміну заряду в них при протіканні електричного струму.

Зміна електростатичної енергії при тунелюванні одного електрона являє

собою різницю між енергією, накопиченою у конденсаторі, і роботою, яка виконується джерелом живлення:

$$\Delta E = e^2/(2C) - eV,$$

де $C = C_t + C_e$. Тунелювання енергетично вигідно тільки тоді, коли $\Delta E < 0$, що з урахуванням можливості тунелювання як при прямому, так і при зворотному електричному зсуві приводить до значення граничної напруги $|V_t| = e/2C$.

Величину $e^2/(2C)$ називають **кулонівською щільністю**.

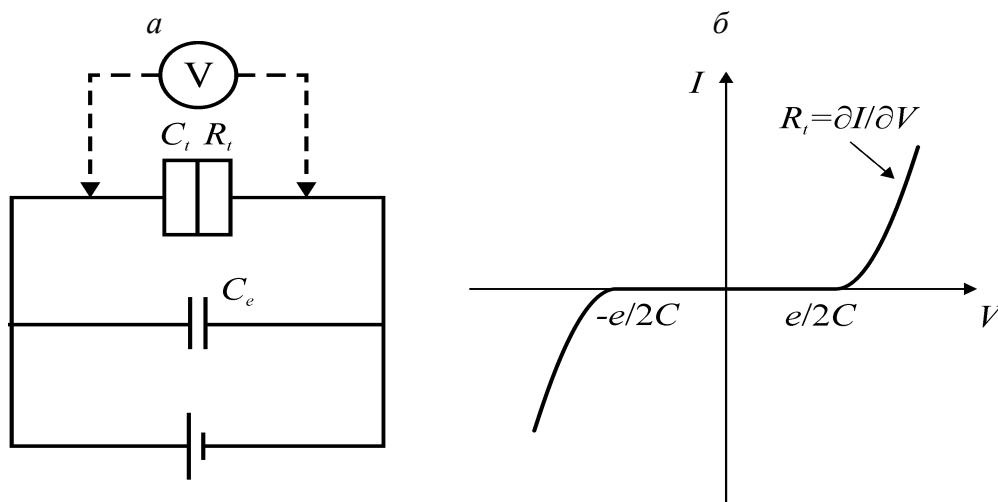


Рис. 10.5. Однобар'єрна структура: *а* - еквівалентна схема; *б* - вольт-амперна характеристика.

Вольт-амперна характеристика однобар'єрної структури показана на рис. 10.5, *б*. Електричний струм у такій структурі з'являється тільки при її електричному зсуві напругою, більшою за граничну. Між значеннями напруги $\pm e/(2C)$ структура є «закритою» – струм не протікає. На «відкритій» ділянці вольт-амперної характеристики струм визначається величиною тунельного опору R_t .

Експериментальне спостереження розглянутих закономірностей можливе тільки за умови малої величини енергії теплових флуктуацій в порівнянні з енергією перезарядження конденсатора одним електроном. Це вимагає виконання умови $k_B T \ll e^2/(2C)$. Відповідно до нього при температурі рідкого гелію (4 К) величина ємності тунельного бар'єра повинна бути менш 10^{-16} Ф, що можна здійснити тільки у нанорозмірних структурах, серед яких найбільш підходящими є квантові точки. Крім цього, для прояву ефектів, зв'язаних з одноелектронним тунелюванням в умовах кулонівської блокади, необхідно, щоб тунельний опір структури $R_t > h/e^2$. Виконання цієї умови потрібно для перевищення рівня квантових флуктуацій кількості електронів у квантовій

точці, через яку відбувається одноелектронне тунелювання. Тобто флуктуації повинні бути досить малі для зміни заряду, щоб він був стійко локалізованим у квантовій точці.

Двохбар'єрна структура. Дві топологічно сполучені, послідовно з'єднані тунельні структури і їхня еквівалентна схема показані на рис.10.6.

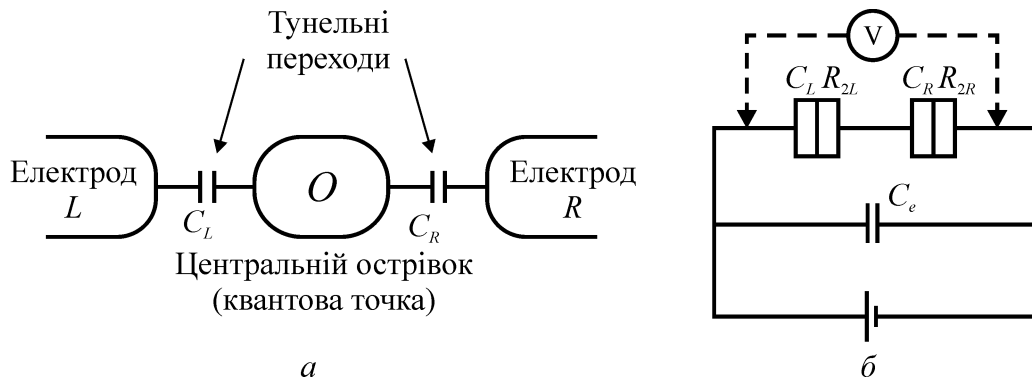


Рис. 10.6. Дві топологічно сполучені, послідовно з'єднані через квантову точку тунельні структури: *а* - модель структури; *б* - еквівалентна схема.

Такою структурою може бути металевий провідник з металевим або напівпровідниковим острівцем O у його розриві. Острівець цей за своїми розмірами і властивостями звичайно відповідає квантовій точці, у якій локалізована визначена кількість електронів. Він має ємнісний зв'язок як із правим, так і з лівим електродом, які характеризуються відповідно ємностями C_L і C_R . Сам же острівець має ємність C , що дорівнює сумі C_L і C_R .

Як і у випадку одnobар'єрної структури, для двохбар'єрної структури також існує визначений діапазон значень напруги, у якому буде відсутній електричний струм унаслідок кулонівської блокади переносу електронів. Однак симетрія вольт-амперної характеристики, типова для одnobар'єрної структури, буде порушена додатковими обмеженнями на перенос електронів, зв'язаними з дозволеними електронними станами в острівці.

На вольт-амперній характеристиці двох бар'єрної структури є діапазон напруги, у якому електричний струм у структурі не може протікати. Це залежить від електронних властивостей центрального острівця. Даний ефект можна пояснити в термінах енергії Фермі острівця й електродів. При маленьких розмірах острівця енергія Фермі у ньому відрізняється від цього параметра в електродах. Це приводить до різниці потенціалів $\Delta V_0 e$, як то показано на рис. 10.7, *а*. Причина такого розходження полягає у тім, що перерозподіл заряду в острівці може бути відповідним тільки потенціалу, кратному e/C . У

макроскопічному зразку ємність C досить велика, так що e/C стає вкрай малим, і збіг енергій Фермі у провідниках і квантовій точці при проходженні через перехід буде майже точним.

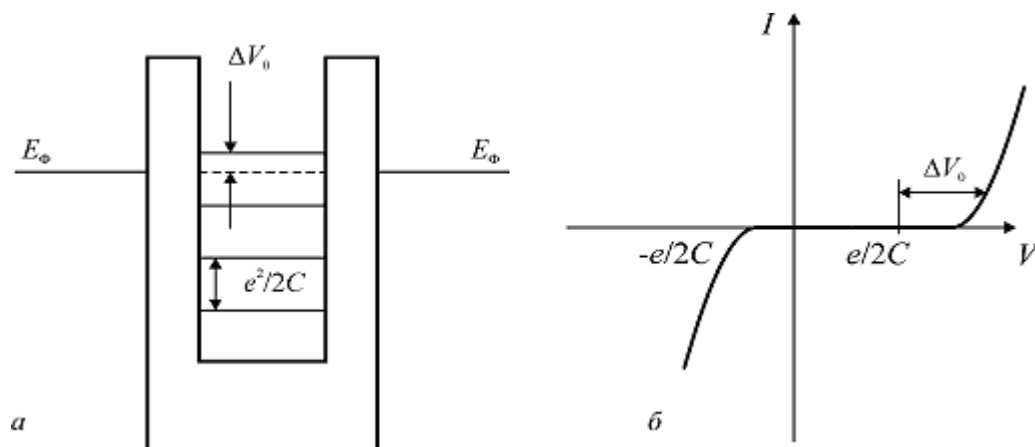


Рис. 10.7. Двохбар'єрна структура: *а* – еквівалентна схема; *б* – вольт-амперна характеристика.

У нанорозмірних структурах рівновага приводить до того, що рівні Фермі розташовуються так близько, наскільки це можливо, але ΔV_0 залишається усе ще значною величиною в діапазоні $\Delta V_0 < e/(2C)$. Ця неузгодженість енергій Фермі і веде до асиметрії вольт-амперних характеристик.

Дискретність енергетичних станів електронів в острові визначається тим, що зміна його заряду може відбуватися тільки дискретно з приходом або відходом одного електрона. Величина цієї енергетичної дискретності складає $e^2/(2C)$. У сусідніх же макроскопічних областях за границями бар'єра енергетичні стани електронів квазінеперервні. При температурі 0 К ці стани заповнені аж до рівня Фермі E_F . Це і приводить до визначеної енергетичної різниці $e/\Delta V_0$ між рівнем, з якого йде інжекція електрона, і найближчим рівнем, який електрон може зайняти в острові. Електрон, що прийшов на острів з лівого електрода, індукує поляризаційний заряд на правому бар'єрі величиною eC_R/C . Щоб її перевищити і перебороти кулонівську блокаду, прикладена зовнішня напруга повинна задовольняти умові:

$$V > C(e/2C + \Delta V_0)/C_R.$$

Це й приводить до несиметричного характеру вольт-амперної характеристики двухбар'єрної структур, рис. 10.7 , *б*. Очевидно, що при $\Delta V_0 = 0$ вольт-амперна характеристика буде симетричною – типовою для однобар'єрної

структури.

Розглянутий випадок припускає ідентичність характеристик двох з'єднаних тунельних бар'єрів. Коли ж один з бар'єрів має більш високу прозорість, то вольт-амперна характеристика здобуває специфічний, східчастий вид (рис. 10.8).

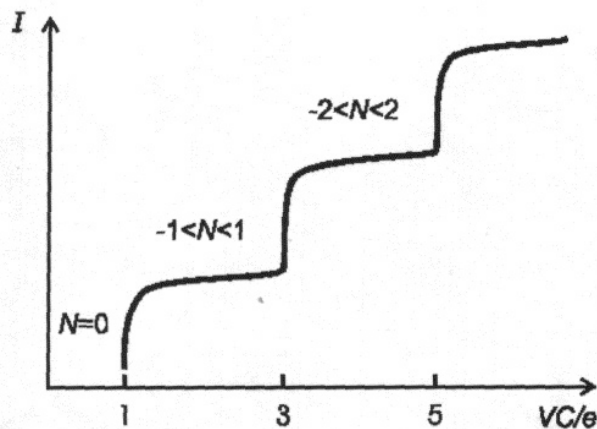


Рис. 10.8. Вольт-амперна характеристика двохбар'єрної одноелектронної структури в якій один з бар'єрів якої має більш високу прозорість.

Таку характеристику називають **кулонівськими сходинками**. Вона є наслідком розходження у швидкостях тунелювання через перший і другий бар'єр. Електрод у бар'єра з більшою прозорістю називають *витіком*, а з меншої – *стоком* (за аналогією з польовим транзистором).

При напрузі між витіком і стоком, яка перевищує поріг кулонівської блокади, електрон тунелює в острівця між електродами. Там він затримується на досить тривалий час, перш ніж відбудеться його тунелювання до стоку. Тунелювання з острівця до стоку *через меншу прозорість* бар'єра, відбувається з меншою імовірністю і відповідно обмежує перенос електронів через острівця. Заряд, накопичений на границі стокового бар'єра, стає більше заряду на витіковому бар'єрі. Це приводить до того, що підвищення напруги на зовнішніх контактах компенсується, головним чином, спаданням напруги на стоковому бар'єрі. Падіння ж напруги на витіковому бар'єрі залишається майже незмінним, а оскільки саме його величина визначає швидкість інжекції електронів в острівця, то загальний струм також залишається незмінним. Так формується перша після кулонівського зазору сходинка, на якій струм, що протікає через структуру, не залежить від прикладеної напруги.

Наступні сходинки з'являються з періодичністю, що відповідає можливості збільшувати заряд острівця збільшенням числа електронів на

острівці у міру підвищення напруги в зовнішньому ланцюзі. Таким чином, положисті ділянки на вольт-амперній характеристиці відповідають різним зарядовим станам острівця.

Двохбар'єрні структури становлять практичний інтерес для створення перемикаючих і підсилювальних приладів через можливість керування їхніми електричними характеристиками шляхом відповідного впливу на електронні стани в електродах і острівці. Таке керування досягається використанням різних матеріалів для острівця і електродів або використанням локалізованих зарядів, вбудованих у бар'єрний діелектрик. Однак найбільш ефективним з погляду практичного приладового застосування є оперативне керування вольт-амперною характеристикою структури за допомогою електричного потенціалу, подаваного через третій електрод на острівець. У цьому випадку він відіграє роль затвора польового транзистора. На цьому принципі конструюються одноелектронні транзистори.

Далі розглядається тунельний МДМ з *одним переходом* (метал-діелектрик-метал) з кулонівської блокадою (це може бути одним з принципів дії одноелектронного транзистора).

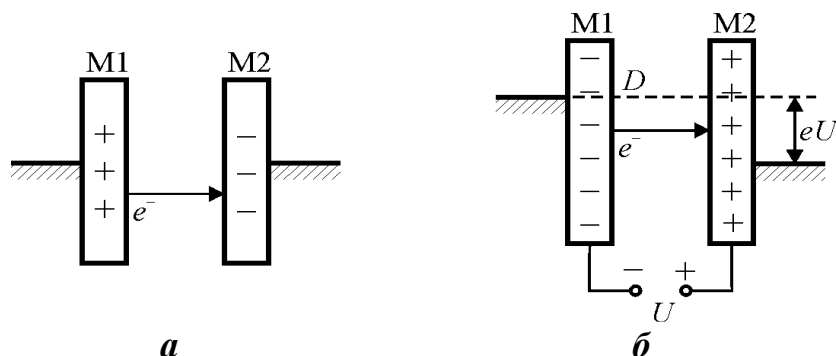


Рис. 10.9. Схема, що пояснює механізм кулонівської блокади при тунелюванні електрона

Нехай спочатку система з двох металевих провідників і з діелектриком між ними не заряджена. Цю систему можна розглядати як конденсатор з деякою ємністю C . Якщо перенести електрон з зарядом e^- з пластини M1 на пластину M2, то конденсатор виявиться зарядженим. Перенос заряду вимагає енергії ΔE , тому що цей перенос відбувається проти сил взаємодії з позитивним зарядом, що виникає на пластині M1, рис. 10.9, а. Ця енергія дорівнює енергії зарядженого конденсатора: $\Delta E = e^2/(2C)$. Величина ΔE являє собою *енергію одноелектронної зарядки*.

Якщо прикласти до пластин напругу U (як це показано на рис. 10.9, б) і поступово збільшувати її, то на пластинах стануть накопичуватися заряди. При

цьому до деякого значення $U = U_k$ тунельний струм через перехід не виникає, тому що робота джерела напруги при переміщенні електрона з пластини М1 на пластину М2 по величині залишається менше роботи проти кулонівських сил відштовхування. Іншими словами, протіканню тунельного струму перешкоджає кулонівська блокада.

Блокада ця буде «прорвана», коли напруга U досягне значення U_k , обумовленого співвідношенням: $eU_k = e^2/(2C)$, відкіль впливає й величина цієї напруги: $U_k = e/(2C)$. Напруга U_k називається напругою кулонівської блокади. Після “прориву блокади” тунелює один електрон, потім знову накопичується заряд на пластинах і т. д. Електрони тунелюють через перехід по одному. Спостереження ефекту кулонівської блокади можливо в умовах, коли енергія теплового руху електронів недостатня для подолання блокади, тобто $\Delta E \gg k_B T$. Іншими словами, умовою подолання блокади є нерівність $C \ll e^2/(2k_B T)$.

Підставивши в останню нерівність значення e і k_B , одержимо, що для спостереження ефекту кулонівської блокади при температурі 1 К необхідна ємність $C < 9 \cdot 10^{-16}$ Ф, а для 300 К – $C < 9 \cdot 10^{-18}$ Ф. Це означає, що блокаду можна спостерігати або при температурах нижче гелієвих, або при дуже малих ємностях. У звичайних умовах енергія ΔE дуже мала, тому що ємність C надто велика. Якщо температура 300 К, то пристрій типу зображеного на рис. 10.9, б повинен для прояву кулонівської блокади мати пластини М1 і М2 розміром порядку декількох нанометрів, якщо товщина діелектрика D складає приблизно 10 нм.

Ще одна умова спостереження кулонівської блокади полягає в тому, що енергія одноелектронного заряду ΔE повинна перевищувати квантові флуктуації енергії $\Delta E' = h/\tau$, де τ – час зарядки конденсатора, а R_T – опір тунельного переходу: $\Delta E > \Delta E'$.

У структурах із **двома тунельними переходами** кулонівська блокада має особливості (у приладових структурах одноелектроніки використовують саме два тунельних переходи, щоб послабити шунтуючу дію провідників).

Конденсатор із двома тунельними переходами показаний на рис. 10.10, а. Металева гранула – квантова точка, острівцець – вміщена в діелектричне середовище (заштриховане) між двома з'єднаними металевими пластинами. Острівцець і пластини утворюють дві обкладки складного конденсатора ємністю C . Так само, як і в попередньому випадку, перенесення електрона з пластин на острівок (або навпаки) вимагає енергії.

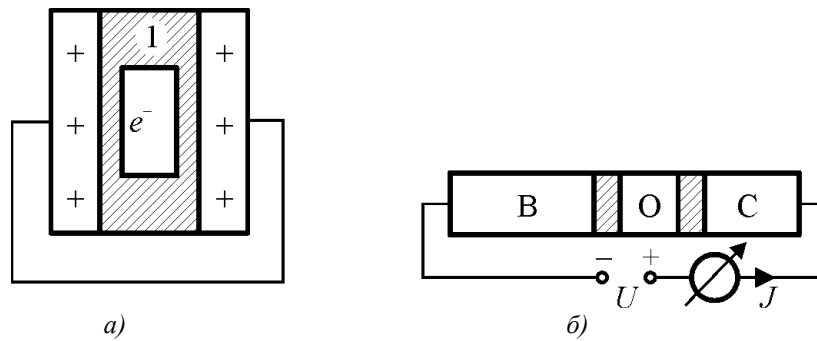


Рис. 10.10. Структура, що включає два електроди і розташований між ними в діелектричному середовищі наноострівець під час відсутності (а) і при наявності (б) різниці потенціалів між електродами.

Схема спостереження кулонівської блокади наведена на рис. 10.10, б. Металеві пластини «В» і «С» називаються «витік» і «стік»; «О» – металевий острівець (квантова точка); заштриховані тунельні бар'єри – діелектрик.

Енергетичні діаграми, що пояснюють ефект кулонівської блокади приведені на рис. 10.11.

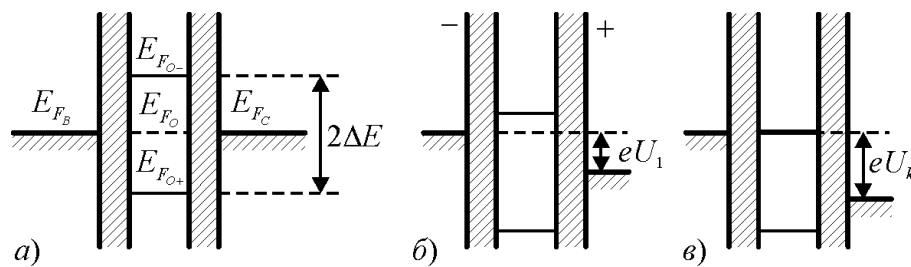


Рис. 10.11. Енергетичні діаграми, що пояснюють принцип роботи елемента на основі ефекту кулонівської блокади.

У стані рівноваги (напруга $U = 0$) рівні Фермі витіку, острівця і стоку (E_{F_B} , E_{F_O} , E_{F_C} відповідно) знаходяться на одній висоті; струму в структурі немає, рис. 10.11, а. Символами $E_{F_{O+}}$ і $E_{F_{O-}}$ позначені рівні Фермі острівця при видаленні з нього електрона і додаванні електрона відповідно; при цьому $E_{F_{O+}} - E_{F_{O-}} = 2\Delta E = e^2/C$.

Рис. 10.11, б ілюструє випадок, коли між витіком і стоком прикладена невелика напруга U_1 , яка ще не може викликати тунелювання, так що струм у ланцюзі дорівнює нулеві. Електрони починають тунелювати із витіка на острівець, а потім з острівця на стік при деякому критичному значенні напруги U_k (рис. 10.11, в). При цьому в ланцюзі з'являється струм J , що зростає при подальшому збільшенні U .

Вольт-амперна характеристика структури показана на рис. 10.12, а.

Прикладена напруга становить одиниці мілівольтів, одноелектронний тунельний струм складає десятки наноампер.

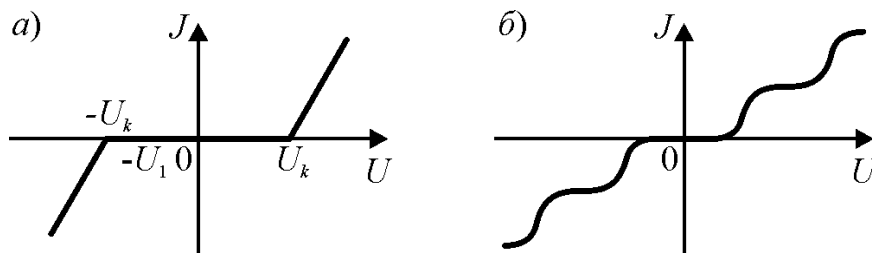


Рис.10.12. Вольт амперні характеристики елемента, що працює на основі ефекту кулонівської блокади

Коли напруга U відповідає рис. 10.11, *в*, то електрони тунелюють і проходять через острівця по одному. Поки на островці є один додатковий електрон, інший не може туди проникнути через кулонівське відштовхування.

Якщо при збільшенні U між E_{F_B} і E_{F_C} виявиться кілька зарядових станів островця, то можливі мультіелектронні переходи. Але якщо один з бар'єрів буде значно більш товстим, то електрони знову будуть проходити через острів по одному. Коли один з бар'єрів значно більший за інший, то вольт-амперна характеристика структури має вигляд, показаний на рис. 10.12, *б*, і називається **«кулонівськими сходишками»**.

Таким чином, кулонівська блокада – це блокування проходження електронів через квантову точку, включену між двома тунельними контактами. Блокада обумовлена відштовхуванням електронів у контактах від електрона на точці, а також додатковим кулонівським потенційним бар'єром який створює електрон, що помістився на точці. Експериментально кулонівська блокада виявляється як шпиляста залежність провідності квантової точки від її потенціалу, тобто від напруги на додатковому електроді (затворі). При певному підборі напруги на затворі і відносних положеннях рівнів Фермі контактів квантової точки потенційний бар'єр для переходу електрона з контакту в точку зникає. Це і спостерігається як пік електропровідності точки. Через кінцеву температуру квантової точки рівні Фермі в контактах злегка розмиті, це робить ширину піків кулонівської блокади кінцевою. Тобто звичайно ширина піка в одиницях e має порядок температури точки в одиницях $k_B T$.

10.3. Резонансне тунелювання.

Звичайний *тунельний ефект*, як уже відзначалося вище, це проходження мікрочастинок крізь потенційний бар'єр, висота якого U^0 більше енергії

частини. При цьому не всякий раз частинка з даною енергією пройде крізь бар'єр. Існує визначена імовірність її проходження крізь бар'єр, названа **коефіцієнтом прозорості**.

Значення коефіцієнта прозорості тим більше, чим менше ширина бар'єра d і різниця між його висотою U^0 і енергією частини E , тобто чим менше “дефіцит” енергії частинки усередині бар'єра ($U^0 - E$). У нанорозмірній структурі цей ефект зможе проявитися, наприклад, у проходженні електронів крізь тонкі прошарки діелектрика.

Час тунелювання дуже малий: $\hbar/(U^0 - E) \approx 10^{-15}$ с, що можна оцінити за співвідношенням невизначеностей: $\Delta E \cdot \Delta t \geq \hbar/2$. Тунельний ефект має помітну імовірність, якщо ширина бар'єра d порівнянна з довжиною хвилі де Бройля електрона. Він визначає межі функціонування елементів інтегральних мікросхем, побудованих на основі традиційних принципів. Але якщо покласти тунельний ефект в основу принципу дії приладу, то це може підвищити його швидкодію до сотень терагерців (ТГц). Наприклад, за цим принципом працюють деякі одноелектронні пристрої.

Тунелювання електронів у низькорозмірній структурі визначається не тільки характеристиками її потенціальних бар'єрів, але і дозволеними енергетичними станами для електронів усередині самої структури. При збігу рівня Фермі на інжектуючому електроді та на дискретному рівні нанорозмірної структури (обмеженої двома потенційними бар'єрами) має місце різке зростання тунельного струму, що протікає через неї. Саме це явище одержало назву *резонансне тунелювання*. Воно приводить до появи ділянки з негативним диференціальним опором на вольт-амперній характеристиці структури. Ця особливість досить важлива для створення твердотільних надвисокочастотних генераторів.

Типовими структурами для спостереження цього ефекту є квантові ями (квантові плівки). Енергетична діаграма такої структури і її вольт-амперна характеристика показані на рис. 10.13.

Як правило, двохбар'єрні резонансно-тунельні структури виготовляють на основі надгратки GaAs-AlGaAs. У якості «вузькозонного» матеріалу електродів і для самої ями використовується GaAs, а бар'єри формуються з «широкозонної» потрійної сполуки GaAlAs. Для реалізації резонансно-тунельних структур використовуються також комбінації напівпровідник-діелектрик, наприклад Si-CaF₂, Si-SiO₂.

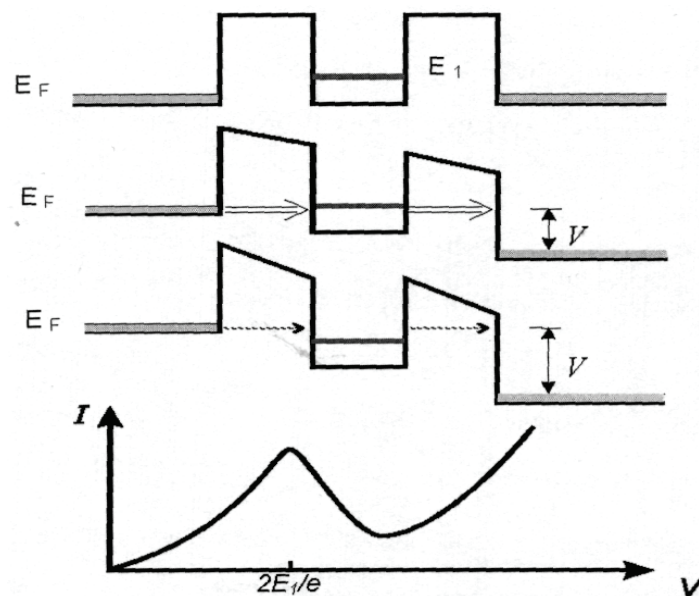


Рис. 10.13. Енергетична діаграма і вольт-амперна характеристика двохбар'єрної структури з квантовою ямою

Квантове обмеження приводить до квантування енергетичних рівнів у ямі. У двовимірній структурі – квантовій плівці – утворюється підзона з енергією, що відповідає E_1 , рис. 10.13. Під час відсутності зовнішньої напруги рівень E_1 розташовується вище рівня Фермі (E_F) у відділених бар'єрами електродах. Як звичайно, електрод, з якого здійснюється інжекція електронів, називають емітером, а електрод, що збирає електрони – колектором.

При додатку до структури невеликої зовнішньої напруги V основне падіння напруги приходить на області бар'єрів, тому що їхній електричний опір багато більше опору ями. У них відбувається найбільше скривлення енергетичних рівнів, однак електричний струм через структуру залишається невеликим. Тільки зі збільшенням прикладеної напруги рівень E_1 у квантовій ямі знижується нижче за E_F , забезпечуючи появу тунельного струму через емітерний бар'єр.

Тунелювання електронів з ями у колектор не викликає обмежень, тому що відбувається з більш високого енергетичного рівня на більш низький. Величина тунельного струму стає максимальною при збігу рівня Фермі в емітері й енергетичного рівня E_1 ями. При цьому електрони тунелюють у яму, *зберігаючи свою енергію й імпульс*. Завдяки цьому вони там не затримуються і швидко ідуть у колектор. Відбувається **резонансне проходження** електронів через квантову яму, умовою якого є співвідношення $V = 2E_1/e$ (при симетрії емітерного і колекторного бар'єрів).

При подальшому зниженні рівня E_1 електрони вже не можуть тунелювати

зі збереженням енергії і імпульсу. Яма „переповнюється” і електрони затримуються в ямі. Тому струм через структуру зменшується, створюючи ділянку з **негативним диференціальним опором** на вольт-амперній характеристиці. Подальше збільшення напруги приводить до все зростаючої, термічно активованої надбар'єрної емісії електронів і до відповідного росту струму через структуру.

Резонансний тунельний ефект, властивий для двох- або багато- бар'єрної періодичної структури. Він проявляється в різкому збільшенні імовірності проходження частини крізь бар'єри має практичний інтерес для приладів наноелектроніки.

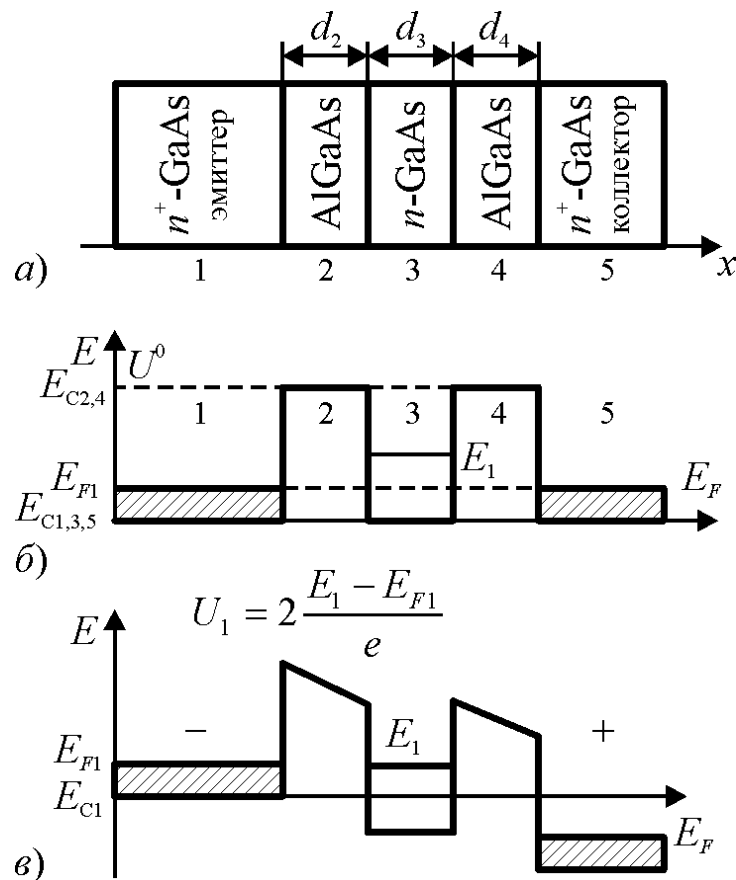


Рис. 10.14. Принцип дії резонансно тунельного діода: а – структура; б, в – енергетичні діаграми.

Резонансне тунелювання крізь ряд бар'єрів виникає тільки в тому випадку, якщо ширина ям і бар'єрів має порядок довжини хвилі де Бройля. У цьому ефекті час проходження електроном структури включає, крім часу тунелювання, час перебування електрона в ямі, що розділяє бар'єри, тобто час його життя τ на резонансному рівні.

Наприклад, для подвійної гетероструктури, що складається із шарів

$\text{Al}_{0,3}\text{Ga}_{0,7}\text{As}$ (5 нм) - $\text{Ga}_{0,7}\text{As}$ (7 нм) - $\text{Al}_{0,3}\text{Ga}_{0,7}\text{As}$ (5 нм) при висоті бар'єрів 0,2 еВ, час становить $\tau \sim 10^{-11}$ с. Таким чином, час тунелювання забезпечує роботу відповідних приладів у *терагерцевому діапазоні*. Треба зазначити, що величина τ ще зменшується при подальшому зменшенні розмірів структури.

Двохбар'єрні структури становлять великий інтерес для електроніки, тому що на їх основі можуть працювати надвисокочастотні (НВЧ) прилади в діапазоні сотень гігагерців (ГГц) і перемикачі із затримкою, меншою за одну пікосекунду. Створено прилади на основі двохбар'єрної структури – резонансно-тунельний діод і транзистор. Розроблені і набули застосування багатобар'єрні структури – надгратки.

10.4. Балістичний транспорт заряду

Транспорт носіїв заряду у низькорозмірних структурах демонструє низку нових явищ у порівнянні з поведінкою носіїв заряду у макроскопічних системах. У появному вигляді вони виявляються тоді, коли розміри наноструктур менші, ніж довжина фазової когерентності електрона. Фази невзаємодіючих електронних хвиль можуть *зберігатися* протягом усього шляху, який вони проходять у таких структурах. Коли електронні хвилі з різними фазами зустрічаються одна одну, з'являються інтерференційні ефекти.

Фазова інтерференція відбувається в структурах з розмірами порядку довжини фазової когерентності електрона в даному матеріалі, що типово для твердотільних структур з нанометровими розмірами. При таких співвідношеннях електропровідність структури визначається ефектами, пов'язаними з *фазовою інтерференцією*. Це має місце в чисто балістичних або квазібалістичних режимах переносу носіїв заряду. Останній допускає слабке розсіювання носіїв заряду, що є типовим для більшості наноструктур, розглянутих як системи з незначним розупорядкуванням. У цьому випадку критична довжина визначається середньою довжиною вільного пробігу носіїв заряду при їх непружному розсіюванні.

Електропровідність (G) звичайного дроту круглого перетину дорівнює

$$G = 1/R = \sigma S/L$$

де $S = \pi r^2$ – площа поперечного перерізу, L – довжина дроту, r – його радіус і σ – питома електропровідність. Ця формула справедлива, якщо r і L набагато більше середньої довжини вільного пробігу електрона $\lambda_{\text{в}}$. У цих умовах рух електрона по провіднику носить дифузійний характер, а траєкторія його руху –

ламана лінія (рис. 10.15, *a*). Цей рух визначає електричний опір.

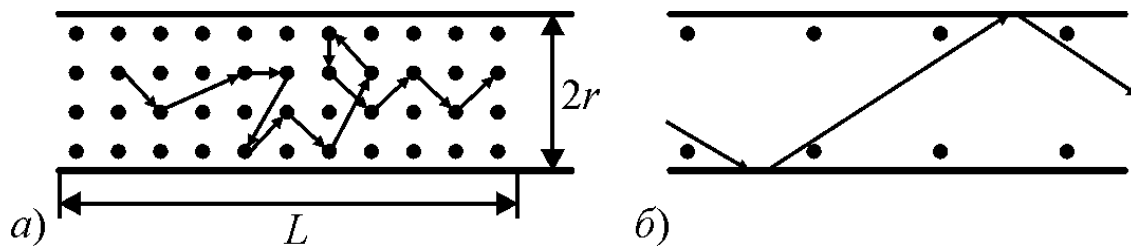


Рис. 10.15. Схематичне представлення дифузійного (*a*) і балістичного (*б*) руху електрона в провіднику

Якщо $\lambda_b > L$ і $\lambda_b > r$ (рис. 10.15, *б*), то електрон пролітає від одного контакту до іншого практично без зіткнення з атомами кристалічних ґраток. Такий режим руху називається балістичним. При цьому електрон не відчуває опору своєму рухові в об'ємі провідника.

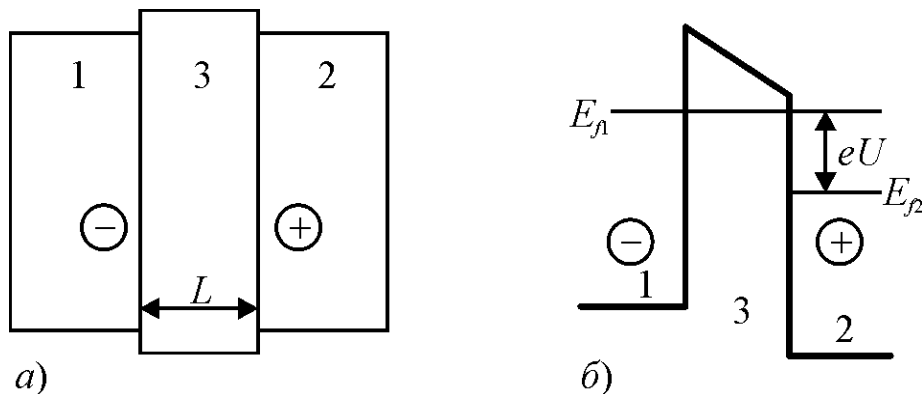


Рис. 10.16. Просторова (*a*) і енергетична (*б*) схеми балістичного провідника 3 і контактів до нього

Балістичний провідник із квантово-розмірним діаметром, вміщений між двома металевими контактами 1 і 2, схематично зображений на рис. 10.16, *a*. Припустимо, що температура має порядок декількох градусів Кельвіна і всі електрони у контактах на енергетичній діаграмі рис. 10.16, *б* розташовані нижче рівнів Фермі E_{F1} і E_{F2} . Якщо між контактами прикласти різницю потенціалів U , (рис. 10.16, *б*), то енергетичні рівні металу 2 понизяться на величину eU щодо рівнів металу 1. При цьому $E_{F1} - E_{F2} = eU$. Струм може створюватися тільки електронами, що мають енергію в інтервалі від E_{F1} до E_{F2} . Саме ці електрони з контакту 1 мають можливість переходити у контакт 2 на вільні рівні.

Якщо провідник (або напівпровідник) 3 характеризується N розмірними підзонами і є балістичним, тобто, має квантово-розмірний перетин S , то його

електрони розташовуються в розмірних підзонах. У переносі струму можуть брати участь тільки ті електрони підзон, що знаходяться в інтервалі від E_{f1} до E_{f2} . У цьому випадку провідність $G = 2e^2N/h$, а опір $R = h/2e^2N$.

Таким чином, на відміну від класичного дроту, *опір балістичного квантового дроту не залежить від його довжини L* . Число N визначається відстанню між підзонами, а ця відстань збільшується зі зменшенням перетину дроту S . Якщо поступово зменшувати діаметр дроту, то з інтервалу $(E_{f1} - E_{f2})$ будуть по черзі і по одній відходити розмірні підзони. При відході кожної підзони провідність G стрибком зменшується на величину $2e^2/h$. Коли в інтервалі $(E_{f1} - E_{f2})$ не залишиться жодної підзони, провідність G звернеться в нуль.

Величина $2e^2/h$ називається **квантом провідності**, а обернена величина $h/2e^2 = 12,9$ кОм – **квантом опору**. Фактично квантування опору обумовлене розмірним квантуванням енергії. Для спостереження ефекту квантування опору необхідні досить низькі температури (близько 1 К). При більш високих температурах стрибки провідності G розмиваються або зникають, тому що тепловий рух у контактах «закидає» електрони на рівні, де $E > E_f$. Слід зазначити, що вимірюваний у цих умовах опір – це опір у контактах. У самому балістичному нанодроті розсіювання електронів немає. Отже, він не повинний мати електричного опору.

На закінчення цього розділу слід зазначити, що функціонування багатьох приладових структур наноелектроніки визначається описаними раніше особливостями енергетичних спектрів квантово-розмірних елементів. При цьому важливим є та обставина, що розглянуте **квантування енергії** спостерігається тільки при розмірах об'єктів порядку хвилі де Бройля (хоча б в одному вимірі).

10.5. Квантовий ефект Холла.

Класичні гальваномагнітні ефекти Холла і Гауса детально розглянуті у розділі 4.11. Ці ефекти широко використовуються для дослідження електронних властивостей матеріалів і в сенсорній електроніці.

Класичні гальваномагнітні ефекти Холла і Гауса спостерігаються, коли пластику електропровідного матеріалу поміщають у магнітне поле і через провідник пропускають електричний струм, як це показано на рис. 10.17. На електрони діє сила Лоренця, яка спрямована перпендикулярно як до магнітної індукції B , так і до первісного напрямку руху електронів. Під дією сили Лоренця електрони притискаються до однієї зі сторін зразка (у залежності від

напрямку магнітного поля), викликаючи нагромадження заряду на цій стороні в порівнянні з іншою стороною.

Основний наслідок дії сили Лоренця $\vec{F}_{\text{Лор}} = e(\vec{E} + [\vec{v}\vec{B}])$ на носії заряду у провіднику обмежених розмірів по вісі y - різниця потенціалів \mathcal{E}_y у провідній пластині (рис. 10.17), яка отримала назву **холлівська різниця потенціалів**:

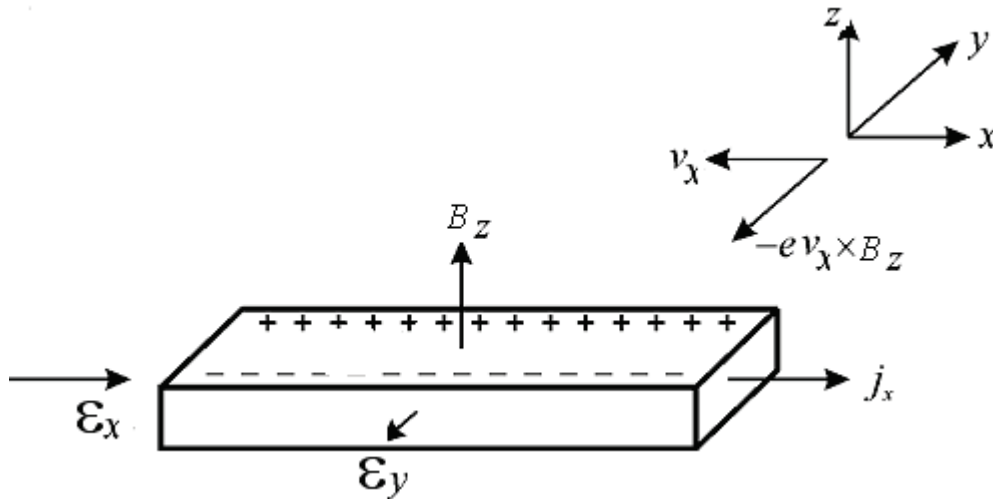


Рис. 10.17. Схема експерименту Холла.

$$\mathcal{E}_y = \frac{r}{en} j_x B_z = |R_H| j_x B_z,$$

де n - концентрація електронів або дірок у провіднику; r - Холл-фактор, залежний від механізму розсіювання та рівня легування ($r=1$ для металів та вироджених напівпровідників; неvirоджені напівпровідники: фононне розсіювання $r \approx 1,13-1,18$, іонне розсіювання $r \approx 1,93$). Помітимо, що холлівська різниця потенціалів \mathcal{E}_y *прямо пропорційна* B_z , рис.10.18,а.

Інакше кажучи, сила Лоренця приводить до нагромадження негативного заряду біля однієї грані бруска і позитивного заряду - біля протилежної грані. Таке нагромадження заряду продовжується доти, поки електричне поле зарядів \mathcal{E}_y не компенсує магнітну складову сили Лоренця: $e\mathcal{E}_y = ev_x B_z$.

Коефіцієнт пропорційності між \mathcal{E}_y і $j_x B_z$ називається **постійною Холла** $|R_H|$. Знак постійної Холла залежить від знаку носіїв заряду, що дозволяє визначати їхній тип для металів і напівпровідників.

Різниця потенціалів V_0 , обмірювана при заданому струмі I_x уздовж зразка і у відсутності магнітного поля $B_z = 0$, характеризує опір матеріалу $R_0 = V_0 / I_x$. Опір R_B , обмірюваний у присутності магнітного поля $R_B = V_B / I_x$,

зазвичай виявляється більшим за R_0 - **класичний ефект Гаусса**. Відносна зміна подовжнього опору у магнітному полі $(R_B - R_0) / R_0 = \Delta R / R_0$ називається **магнітоопором**. Відповідно до (4.135), магнітоопір $\Delta R / R_0$ і різниця потенціалів V_B пропорційні B_z^2 , рис.10.18,а.

Квантовий ефект Холла спостерігається при низьких температурах в сильних магнітних полях у квазідвовимірній (2D) електронній газі. В залежності від умов експерименту розрізняють **квантовий ефект Холла** (що приводить до квантування постійної Холла), **інтегральний квантовий ефект** і **дробовий квантовий ефект Холла** (квантування обумовлено рівнями Ландау).

Відповідно до класичного опису, вільна заряджена частина в площині, перпендикулярній до вектора магнітної індукції B , рухається по колу радіуса $r = v / \omega_c$, обертаючись із циклотронною частотою $\omega_c = \frac{e}{m^*} B$. У твердому тілі електрон має зіткнення з атомами (іонами) ґратки. Частота, імовірність зіткнень $W = \tau^{-1}$, де τ – час вільного (час релаксації) пробігу (див.(4.31)). Говорити про рух електрона по колу у магнітному полі можна у випадку, коли $\omega_c / 2\pi \ll \tau$ тобто, $\omega_c \tau = \mu B \gg 1$ (умова сильного магнітного поля). Ця умова виконується для невисоких температур (послаблене розсіювання) та потребує досить сильних магнітних полів, які називаються **квантуючими полями**.

Описувати рух в таких полях варто за квантовомеханічним підходом. Характер руху часток у квантуючих полях значно відрізняється від класичного, особливо в 2D-системах. Класична частинка рухається по коловій траєкторії. Як уже відзначалося, для квантової частинки, коли вона рухається в досить обмеженій області, поняття «траєкторії» не є застосовним. Як показують розв'язки рівняння Шредінгера, рух частинки в магнітному полі можна вважати обмеженим (як, наприклад, рух електрона у потенціальної ямі кінцевої глибини). Рух у площині, перпендикулярній до вектора магнітної індукції \vec{B} , обмежується колом радіуса $r = \sqrt{\hbar / eB}$. Енергія такого руху, як всякого обмеженого руху, квантується.

Відповідно до розв'язку рівняння Шредінгера, можливі значення енергії

$$E_n = \hbar \omega (n + \frac{1}{2})$$

де $n = 0, 1, 2 \dots$. Саме ці значення енергії називаються **рівнями Ландау**.

Якщо вектор \vec{B} спрямовано уздовж осі z , перпендикулярної до площини

2D-електронного газу, то можливі значення енергії руху вздовж вісі z – це розмірні рівні E_n . Значення енергії руху в площині xy – це рівні Ландау E_m . Повна енергія електронів 2D-газу дорівнює сумі: $E = E_n + E_m$.

Таким чином, енергія 2D-газу, тобто макроскопічної системи, у досить сильному магнітному полі *цілком квантована* (так само, як і для атомів, як для квантових точок та інших мікрооб'єктів).

Саме існуванням рівнів Ландау і пояснюється квантуванням постійної Холла. Цей ефект – один з макроскопічних проявів квантових властивостей речовини; він має важливе прикладне значення. Спостерігається ефект при дуже низьких температурах ($T \approx 1$ К) і у досить сильних полях (≈ 5 Тл). У ще більш сильних магнітних полях виявляється **дробовий квантовий ефект Холла**, що зв'язаний з кардинальною перебудовою внутрішньої структури двовимірної електронної системи.

Коли ефект Холла вимірюється при низьких температурах у зразку, що містить двовимірний електронний газ, у якому електрони здатні рухатися тільки в межах площини, виявляється, що магнітоопір відхиляється від класичного поведінки. При досить високих полях на графіку залежності напруги Холла від магнітного поля (рис. 10.18, б) спостерігається ряд плоских сходинок (плато). Саме це явище назвали **квантовим ефектом Холла**.

На самих плато (тобто на плоских сходинок) напруги Холла *подовжня напруга* V_B стає нульовою. Опір на ділянках, що відповідають плато сходинок, виявляється квантованим і рівним $R_H = h/(ae^2)$ де h – постійна Планка, a – деяке ціле число.

Такий ефект називається **інтегральним квантовим ефектом Холла**. Він незалежний від властивостей матеріалу. Квант опору h/e^2 , обмірюваний багаторазово з високою точністю в умовах квантового ефекту Холла, використовується зараз як «**стандарт опору**».

Крок квантування a може приймати дробові значення, такі, як $1/3$, $2/3$, $2/5$, $3/5$ і т.д. Взагалі параметр $a = p/q$, де p і q – цілі числа, причому q – непарні числа. Явище одержало назву **дробового квантового ефекту Холла**. Іншою ознакою квантового ефекту Холла (крім квантування опору R_H) є те, що опір R_B знижується до гранично малих величин. Обидва ефекти пов'язані між собою.

Зникнення *подовжнього* опору R_B є ознакою відсутності розсіювання енергії в системі. Це означає, що між основним станом системи і її першим

збудженим станом повинний існувати енергетична щілина – заборонені значення енергії.

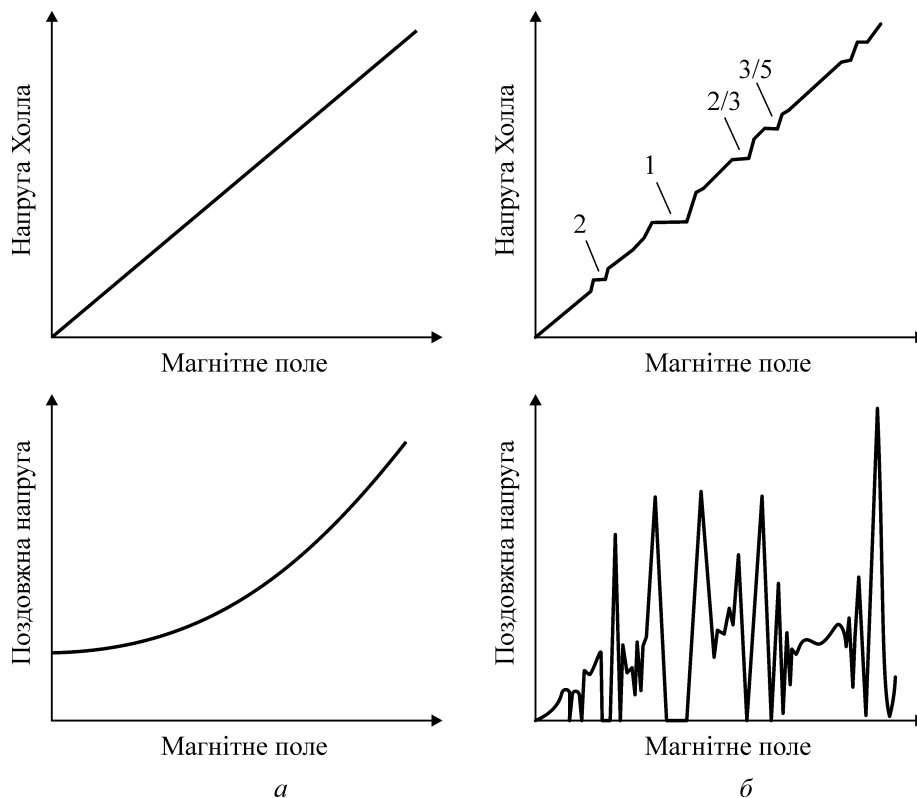


Рис. 10.18. Порівняльна характеристика класичного (а) і квантового (б) ефектів Холла

Електрони, що рухаються *перпендикулярно* магнітному полю, під дією сили Лоренця змушені змінювати траєкторії поступального руху на кругові орбіти. Вони виконують циклотронний рух із енергією $E_c = \hbar\omega_c$. Тому дозволені енергетичні стани для таких електронів і стають квантованими, дискретними.

Саме такому квантованому на кругових орбітах руху відповідають енергетичні рівні відомі як *рівні Ландау*, і вони описуються рівнянням $E_m = (m + 1/2)\hbar\omega_c$ із квантовим числом $m = 1, 2, 3, \dots$ В ідеальній системі, що містить двовимірний електронний газ, ці рівні мають форму δ -функцій, як це показано на рис. 10.19,а. Електрони можуть знаходитися тільки на енергетичних рівнях Ландау, але не в проміжках між ними. Відстань між сусідніми рівнями визначається циклотронною енергією $\hbar\omega_c$.

З підвищенням температури двовимірний електронний газ стає «не

ідеальним» і рівні Ландау розширюються (рис. 10.19,б). Тому для спостереження дискретного спектру рівнів потрібні досить низькі температури – відповідно до умови $k_B T \ll \hbar \omega_c$.

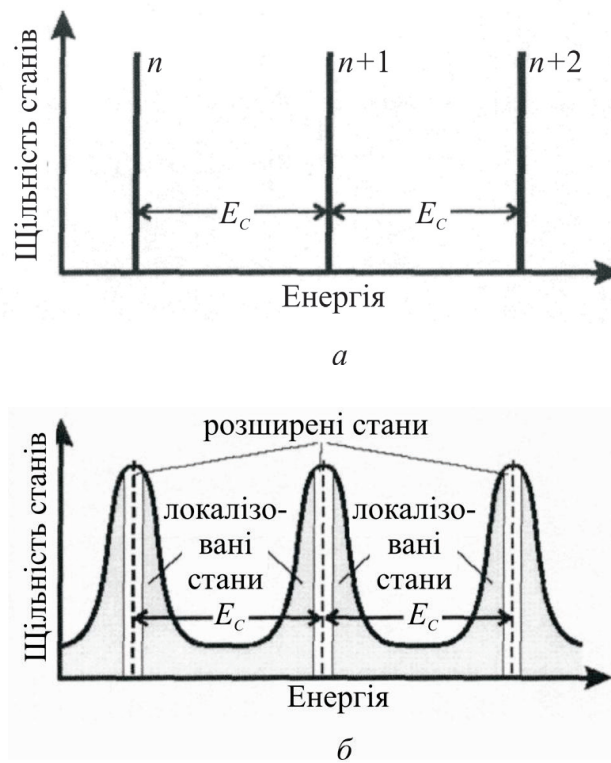


Рис. 10.19. Рівні Ландау: *а* – в ідеальній системі, що містить двовимірний електронний газ; *б* – розширення енергетичних зон біля рівнів Ландау у розупорядкованій системі.

Саме існування проміжків між енергетичними рівнями електронів у магнітному полі і є основною умовою для виникнення квантового ефекту Холла. Справа в тім, що випадок двовимірного електронного газу істотно відрізняється від тривимірного газу електронів, що вільно рухаються в трьох вимірах. Рух у третьому вимірі, уздовж магнітного поля, приводить до додавання деякої кількості енергії до енергії рівнів Ландау і до заповнення, таким чином, проміжків між ними. Тому при русі носіїв у трьох вимірах не спостерігається ніяких енергетичних зазорів, і передумови появи квантового ефекту Холла послабляються.

Крім двовимірності електронного газу необхідною умовою для експериментального спостереження квантового ефекту Холла є наявність у зразку структурних недосконалостей, домішок. Вони викликають необхідне для спостерігання ефекту розширення дозволених енергетичних зон біля рівнів Ландау, як показано на рис. 10.19, б.

Структурні дефекти перетворюють окремі тонкі рівні Ландау в смуги розширених станів, відділених смугами локалізованих станів. Зазначимо, що *електрони самих локалізованих станів не беруть участь у переносі носіїв заряду через структуру*. Тільки електрони, що знаходяться у **розширених станах** здатні брати участь у переносі заряду. Границю між локалізованими і розширеними станами називають **границею рухливості**. Розширені стани оточують дефекти і розташовуються уздовж поверхневої границі зразка, формуючи так називані **граничні стани**, що являються фактично одномірними каналами для електричного струму. Реальні зразки можуть мати складну топологію границь таких провідних каналів, яка відтворює рельєф зміни потенціальної енергії.

Оскільки енергія Фермі E_F змінюється зі зміною магнітного поля, то вона може лежати як всередині смуги розширених станів, так і в межах діапазону локалізованих станів. Коли E_F змінюється у межах діапазону енергій локалізованих станів, то кількість електронів, що беруть участь у транспорті носіїв заряду, залишається постійною. Крім того, перенос носіїв заряду відбувається *без розсіювання енергії*. У результаті звичайний (поперечний) опір є постійним, але подовжній опір R_B різко змінюється, а при певних значеннях магнітного поля може звертатися в нуль (рис. 10.19, б).

Кванти магнітного потоку – це елементарні квазічастинки, за допомогою яких магнітне поле взаємодіє із системою електронів. У цьому змісті ціле число m виглядає природним і пояснює квантовий ефект Холла як перенос носіїв заряду через ціле число *повністю заповнених рівнів Ландау*, відділених проміжками, заповненими локалізованими станами. У термінах граничних станів квантування опору R_B означає, що участь у переносі заряду беруть m одномірних каналів, причому внесок кожного каналу у провідність має величину e^2/h .

При поясненні дробового квантового ефекту Холла необхідно припустити *часткове заповнення рівнів Ландау*. Крім того, інші енергетичні зони при частковому заповненні рівня Ландау повинні мати іншу, відмінну від рівнів Ландау природу. Походження дробового квантового ефекту Холла зв'язано із взаємодією між електронами. Тому він називається ефектом «багатьох частинок», або ефектом електронної взаємозалежності: електрони, будучи ідентично зарядженими частинами, уникають розташовуватися близько один від одного. У результаті відштовхування електрони займають не всі, а лише частину рівнів Ландау у зразку, для того щоб триматися якнайдалі від

своїх сусідів. Фактор заповнення $1/3$, наприклад, означає, що тільки третина рівнів Ландау зайнята.

Якщо, однак, ще один електрон додається до одного з незаповнених рівнів, то енергія його відштовхування буде помітно більшою, тому що він буде ближче до своїх сусідів, чим всі інші електрони. Таким чином, спостерігається енергетична щілина між частково зайнятим станом і станом, розташованим енергетично вище. Через структурні дефекти локалізовані стани формуються у межах щілини. Якщо енергія Фермі доходить до цього проміжку, то опір зразка падає до нуля, як і у випадку інтегрального ефекту.

Часткове заповнення рівнів Ландау означає, що електрони в двовимірному електронному газі, поміщеному в сильне магнітне поле, конденсуються в новий колективний стан – у квантову рідину. Квант магнітного потоку й електрон співіснують як квазічастинка, що несе дробовий заряд. Такі частинки не підкоряються ні статистиці Фермі-Дірака, ні статистиці Бозе-Ейнштейна.

Література

1. Шалимова К.В. Физика полупроводников: Учебник. - М.: «Энергоатомиздат», 1985. – 392 с.
2. Киреев П.С. Физика полупроводников: Учеб. пособ. – М.: «Высшая школа», 1975. – 584 с.
3. Горбачев В.В., Спицына Л.Г. Физика полупроводников и металлов: Учебник. – М.: «Металлургия», 1982. – 336 с.
4. Ансельм А.И. Введение в теорию полупроводников: Учеб. пособ. – М.: «Наука», 1978. – 616 с.
5. Зеегер К. Физика полупроводников /Пер. с англ. – М.: «Мир», 1977. – 615 с.
6. Полякова А.Л. Деформация полупроводников и полупроводниковых приборов. – М.: «Энергия», 1979. – 168 с.
7. Ильинская Л.С., Подмарьков А.Н. Полупроводниковые тензодатчики. – М.: «Энергия», 1966. – 120 с.
8. Москалюк В.А., Синекон Ю.С., Кассинг Р. Физика электронных процессов, часть I: Учеб. пособ. – Киев, УкрИНТЭИ, 2001. – 148 с.
9. Москалюк В.А. Физика электронных процессов, часть II: Учеб. пособ. – Киев, Аверс, 2004. – 186 с.
10. Шокли У., Рид У. Статистика рекомбинации дырок и электронов// Полупроводниковые электронные приборы. – М.: ИЛ, 1953. – С. 121-140.
11. Рывкин С.М. Фотоэлектрические явления в полупроводниках: Монография. – М.: Физматгиз, 1963. – 496 с.
12. Заячук Д.М. Нанотехнології і наноструктури: Навч. посібник. – Львів: «Львівська політехніка», 2009. – 580 с.
13. Волькенштейн Ф.Ф. Электропроводность полупроводников. – М.: Гостехиздат, 1947. – 352 с.
14. Заячук Д.М. Нанорозмірні структури і надгратки: Навч. посібник. – Львів, вид. «Львівська політехніка». 2006. – 220 с.

15. Кравченко А.Ф., Овсяк В.Н. Электронные процессы в твердотельных системах пониженной размерности. – Новосибирск: Изд-во Новосиб. ун-та, 2000. – 448 с.
16. Драгунов В.П., Неизвестный И.Г., Гридчин В.А. Основы наноэлектроники. - Москва, Физматкнига, 2006. – 494 с.
17. Воронов С.А., Переверзева Л.П., Поплавко Ю.М. Физическое материаловедение. Часть 1. Перспективные направления материаловедения: Учеб. пособ. - Киев: НТУУ «КПИ», 2004. – 195 с.
18. Гусев А.И. Наноматериалы, наноструктуры, нанотехнологии. – Москва: Физматлит, 2007. – 415 с.
19. Лозовский В.Н., Константинова Г.С., Лозовский С.В. Нанотехнология в электронике. – М.: Изд. «Лань», 2008. – 327 с.
20. Waser Rainer. Nanoelectronics and information technology. - Wiley-VCH, 2005. – 1001 p.
21. Hari Singh Nalva. Nanostructured materials and nanotechnology. - Academic Press, 2002. – 834 p.
22. Свечников Г.С., Морозовская А.Н. Нанотрубки и графен – материалы электроники будущего. – Киев, Логос, 2009. – 164 с.
23. Sze S.M. Physics of Semiconductor Devices.- John Wiley, 1981. – 868 p.
24. Фистуль В.И. Сильно легированные полупроводники. – М.: Наука, 1967. – 416 с.