Вестник
МОСКОВСКОГО УНИВЕРСИТЕТА
№ 5 — 1969

УДК 539.293.577

В. Л. БОНЧ-БРУЕВИЧ

БЕСПОЛЕВОЙ НАГРЕВ НОСИТЕЛЕЙ ЗАРЯДА И ЭКСИТОНОВ В ПОЛУПРОВОДНИКАХ

Рассмотрена задача о стационарной функции распределения носителей заряда или экситонов в пространственно неоднородной неравновесной системе. Исследована флуктуационная устойчивость такого неравновесного стационарного состояния.

§ 1. Введение. Основные уравнения

Как показано в работе [1], нагрев носителей заряда в полупроводниках может быть вызван не только внешним электрическим полем, но и работой сил давления, коль скоро в системе искусственно создан пространственный градиент функции распределения. Он может быть связан, например, с градиентом локальной средней энергии или просто с градиентом концентрации частиц. Очевидно, таким «бесполевым» способом можно нагревать и систему нейтральных квазичастиц — экситонов, спиновых и плазменных волн и т. д., что, по-видимому, открывает известные новые возможности в физике твердого тела.

В настоящей работе исследуются некоторые особенности пространственно неоднородных неравновесных систем, а также рассматривается задача о флуктуационной их устойчивости.

Будем исходить из обычной системы уравнений для симметричной f_s и антисимметричной f_a частей функции распределения [2].

$$\frac{\partial f_s}{\partial t} + (\vec{v},\ \nabla f_a) + e(\vec{E},\ \nabla_p f_a) = J_s[f], \tag{1}$$

$$\frac{\partial f_a}{\partial t} + (\vec{v}, \nabla f_s) + e(\vec{E},\ \nabla_p f_s) = J_a[f]. \tag{1'}$$

Здесь $\vec{F}=e\vec{E}$ — сила, действующая на частицы заряда e в электрическом поле напряженности $\vec{E}$, J — интеграл столкновений. В принципе он содержит слагаемые, ответственные как за рассеяние, так и за рекомбинацию и генерацию частиц. Последние можно опустить, коль скоро характерные рекомбинационные времена велики по сравнению со временем свободного пробега, а длина образца не слишком велика (при этом процессы рекомбинации и генерации эффективно происходят только на поверхности и на контактах и учитываются в граничных условиях, которые будут сформулированы ниже).

Ограничимся вырожденным газом и допустим, что рассеяние частиц происходит в основном на фононах, примесях и т. д., но не друг на друге. Условия применимости этой аппроксимации хорошо известны; они фактически выполняются в ряде экспериментально интересных ситуаций[1]. Тогда интеграл столкновений есть линейный функционал от f, и

$$J_s[f]=J[f_s],\quad J_a[f]=J[f_a].$$

Определим «дрейфовую» скорость $\vec{v}_d$: $\vec{j}=n\vec{v}_d$, где $\vec{j}$ — поток частиц, n — их концентрация (разумеется, v_d содержит вклады от всех компонентов тока, в том числе и от диффузионного). Будем считать, что дрейфовая скорость мала по сравнению со средней квадратичной скоростью хаотического движения частиц v_T:

$$v_d \ll v_T. \tag{2}$$

Тогда, как известно (2),

$$J[f_a]=-\frac{f_a}{\tau}, \tag{3}$$

где τ — время свободного пробега, вычисленное в предположении об упругости столкновений.

Подставляя (3) в (1′) и считая, что $\vec{E}$ не зависит от f_a, запишем

$$f_a(t)=F_a\exp\left(-\frac{t}{\tau}\right)-\int_0^t dt'\exp\left(-\frac{t-t'}{\tau}\right)\times$$
$$\times\{(\vec{v},\ \nabla f_s(t'))+(e\vec{E}\nabla_p f_s(t')\}. \tag{4}$$

Здесь $F_a(\vec{p},\vec{x})$ — начальное значение функции f_a, которое в нестационарной задаче следует считать заданным.

Уравнение (1′) теперь принимает вид (α, β — векторные индексы)

$$\frac{\partial f_s}{\partial t}+[(\vec{v},\ \nabla)+(e\vec{E}\nabla_p)]F_a\exp\left(-\frac{t}{\tau}\right)-$$
$$-\int_0^t dt'\left\{\left[v_\alpha v_\beta\frac{\partial^2 f_s}{\partial x_\alpha\partial x_\beta}+v_\alpha\frac{\partial}{\partial x_\alpha}\left(eE_\beta\frac{\partial f_s}{\partial p_\beta}\right)\right]\exp\left(-\frac{t-t'}{\tau}\right)+\right.$$
$$\left.+eE_\alpha\frac{\partial}{\partial p_\alpha}\left[v_\beta\frac{\partial f_s}{\partial x_\beta}\exp\left(-\frac{t-t'}{\tau}\right)\right]\right\}=J[f_s]. \tag{5}$$

Интересуясь бесполевым нагревом, мы вправе считать поле достаточно слабым, пренебрегая его влиянием на вид функции распределения f_s (но, разумеется, учитывая его в формуле (4) и в выражении для тока). Это оправдано [1], коль скоро выполняется неравенство

$$\frac{|F|L}{\widehat{W}}\ll 1, \tag{6}$$

---

[1] Именно в этом состоит разница между рассматриваемой системой и обычным газом нейтральных частиц. В последнем случае также, видимо, возможен бесполевой нагрев, но трактовка его усложняется благодаря нелинейности уравнения (1), (1′).

где L — характерная длина неоднородности, $\widetilde{W}$ — средняя энергия в расчете на частицу. Заметим, однако, что так может обстоять дело лишь при определенных условиях опыта, ибо электрическое поле в образце создается не только внешним источником, но и самими заряженными частицами. Неравенство (6), помимо очевидного ограничения на э. д. с. источника, предполагает в сущности, что в образце с хорошим приближением обеспечено условие квазинейтральности. В нашем случае это означает, что радиус экранирования[1] должен быть мал по сравнению с длиной свободного пробега по энергии и что надо рассматривать движение носителей заряда обоих знаков. От последнего усложнения можно избавиться, лишь интересуясь поведением неосновных носителей в резко униполярном образце. В дальнейшем мы ограничимся именно такой постановкой задачи; при этом слагаемые с напряженностью поля в (5) можно опустить. (В задаче о нагреве экситонного газа полевые слагаемые в таком виде вообще не возникают; в неоднородном поле, однако, может появиться сила F, связанная с дипольным моментом экситона. Неравенство (6) и в этом случае оправдывает пренебрежение указанными членами.)[2].

Интересующие нас особенности бесполевого нагрева мало зависят от конкретного вида закона дисперсии частиц $W(\vec{p})$. Ограничимся поэтому простейшим выражением.

$$W = \frac{p^2}{2m}. \tag{7}$$

При этом в случае экситонов под m и $\vec{p}$ следует понимать эффективную массу и импульс центра инерции. В условиях (2) и (7) можно считать, что функция f_s зависит только от энергии, $\tau=\tau(W)$, а уравнение (5) усреднено обычным образом по полярным углам вектора $\vec{p}$. Получим

$$\frac{\partial f_s}{\partial t} + \widetilde{F}(W, \vec{x}) \exp\left(-\frac{t}{\tau}\right) - \int\limits_0^t dt' \exp\left(-\frac{t-t'}{\tau}\right) \frac{2W}{3m} \times$$
$$\times \nabla^2 f_s(W, \vec{x}, t') = J[f_s]. \tag{8}$$

Здесь $\widetilde{F}$ есть результат усреднения второго слагаемого в левой части (5).

Уравнение для стационарной задачи получается из (8), если считать, что f_s не зависит от времени, и устремить t к бесконечности:

$$-\frac{2W\tau}{3m} \nabla^2 f_s = J[f_s]. \tag{9}$$

---

[1] В рассматриваемом случае неоднородной и неравновесной системы термин «радиус экранирования» нуждается в определении. Для ориентировки можно пользоваться локальным его значением, получающимся из обычной формулы Дебая заменой температуры решетки на $\widetilde{W}$ (это оправдано, если получающееся значение заметно превышает длину свободного пробега по импульсу).

[2] Несколько более тонкую аппроксимацию можно получить, отбрасывая слагаемые с полем в левой части (5), но рассматривая в дальнейшем f_s как функцию аргумента $W + e\int\limits_0^x E dx$, а не W. При этом, в частности, получается правильный предельный переход к равновесной неоднородной задаче. Все дальнейшие формулы (с указанной выше заменой аргумента) остаются в силе и в этом случае.

При этом f_s как функция аргумента W должна быть дифференцируема и интегрируема в области $0 \leqslant W < \infty$.

Граничные условия, накладываемые на f_s как функцию координат, определяются постановкой опыта.

Рассмотрим две возможные его схемы, представленные на рис. 1 и 2. В первом случае оптическая генерация неосновных носителей происходит в узком слое вблизи контакта (для простоты равномерно по всей его площади); во втором частота света отвечает только поглощению на свободных носителях заряда, освещается узкий слой в середине образца,

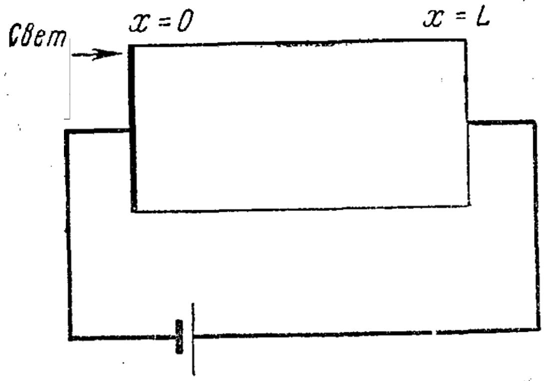

Рис. 1. Оптическая генерация. Зачернен контактный слой. Свет может падать и перпендикулярно плоскости чертежа

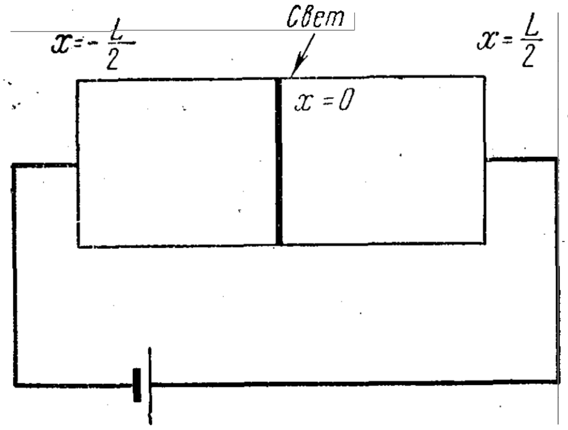

Рис. 2. Поглощение свободными носителями. Зачернен узкий освещаемый слой. Свет падает перпендикулярно чертежу

причем поглощение для простоты считается равномерным по всему объему слоя.

В схеме рис. 1 при $x=0$ (считая, что интеграл $J[f_s]$ в (1′) и (9) представляет собой разность членов «приход» — «уход») имеем

$$\lim_{\varepsilon \to +0} \frac{-2W\tau}{3m} \frac{\partial f_s}{\partial x} \Big|_{-\varepsilon}^{+\varepsilon} = g - s f_s(W, 0). \tag{10}$$

Это есть не что иное, как условие непрерывности потока частиц на границе. Функция $g(W)$ в (10) дает скорость генерации, а $s(W)$ — скорость рекомбинации неосновных носителей на контакте. Такое же условие надо написать и при $x=L$, но, вообще говоря, с другими значениями g и s. В частности, в пренебрежении тепловой генерацией $g=0$ при $x=L$.

Из дальнейшего будет видно, что явный вид граничных условий играет известную количественную, но не принципиальную роль [1]. Поэтому для ориентации будем пользоваться при $x=0$ одним из упрощенных вариантов (10), считая заданной либо саму функцию f_s $(W, 0)$, либо ее производную $\frac{\partial f_s}{\partial x}$ при $x=0$:

$$f_s(W, 0) = \Phi(W), \tag{10′}$$

либо

$$\frac{\partial f_s}{\partial x}\Big|_{x=0} = -\psi(W). \tag{10″}$$

Первое из этих условий отвечает очень большой, а второе — очень малой рекомбинации на контакте (и равновесию — за ним).

[1] Оговоримся, что граничное условие (10) с независимыми функциями g и s на разных контактах не обеспечивает постоянства полного числа рассматриваемых квазичастиц в системе (по постановке задачи этого не требуется).

При $x=L$ положим

$$f_s(W, L) = 0, \tag{11}$$

что соответствует очень большой поверхностной рекомбинации и пренебрежению тепловой генерацией.

В схеме рис. 2 процессами рекомбинации можно вообще пренебречь и (при достаточно большом L) потребовать просто ограниченности f_s при $x=\pm L/2 \to \pm\infty$. Уравнение (9) надо писать независимо для областей $x>0$ и $x<0$; граничные условия при $x=0$ имеют вид

$$\lim_{\varepsilon\to+0} \frac{\partial f_s}{\partial x}\Big|_{-\varepsilon}^{+\varepsilon} = Y(W), \tag{12}$$

$$f_s(W, \varepsilon) = f_s(W, -\varepsilon). \tag{13}$$

Здесь функция $Y(W)$ очевидным образом связана с вероятностью оптического поглощения света свободными носителями. Она интегрально содержит функцию $f_s(W, 0)$, что несколько затрудняет исследование этого случая. Заметим, однако, что принципиально он особенно интересен, ибо полное число рассматриваемых носителей в образце остается неизменным.

§ 2. Стационарная задача. Кинетические коэффициенты

Стационарная задача (9) для некоторых частных случаев рассматривалась в [1] и [3] (в последней работе, однако, не был сделан общий вывод о возможности бесполевого нагрева).

Введем собственные функции f_λ (W) оператора столкновений, полагая

$$\frac{1}{W\tau} J[f_\lambda] = -\lambda f_\lambda. \tag{14}$$

Эти функции должны быть интегрируемы в пределах $0 \leqslant W < \infty$, и все их моменты должны быть конечны. При этом — по смыслу интеграла $J[f_s]$ — собственные значения λ оказываются не отрицательными; тогда в одномерной (по координатам) задаче, соответствующей схеме рис. 1, запишем

$$f_s(W, x) = \sum_\lambda c_\lambda f_\lambda(W)\varphi_\lambda(x), \tag{15}$$

$$\frac{d^2\varphi_\lambda}{dx^2} = \lambda\varphi_\lambda, \quad \varphi_\lambda(0) = 1,$$

где c_λ — коэффициенты, очевидным образом определяемые из условия (10′) или (10″); в случае (11)

$$\varphi_\lambda = \operatorname{sh}(L-x)\sqrt{\lambda}/\operatorname{sh} L\sqrt{\lambda}.$$

Пусть, рассеяние энергии и квазиимпульса происходит на акустических фононах. Тогда (2)

$$J[f] = \frac{2\sqrt{2m}s^2}{lW^{1/2}} \frac{d}{dW}\left[W^2\left(\frac{df}{dW} + \frac{f}{T_0}\right)\right], \tag{16}$$

где T_0 — температура решетки, s — скорость звука, $l = \tau(2W/m)^{1/2}$ — длина свободного пробега, не зависящая от энергии. При этом $\tau = \tau_0 T_0^{1/2} W^{-1/2}$, где τ_0 — постоянная. Измеряя энергию в единицах T_0, а координаты — в единицах

$$x_0 = l\,(T_0/6ms^2)^{1/2}, \tag{17}$$

получим (см. [1] и [3])

$$\lambda = n,\ n = 0, 1, \ldots, f_n = e^{-W} L_n^1(W), \tag{18}$$

где L_n^1 — полином Лагерра [4].

Функция (15) с учетом (18) превращается в равновесную, если отличен от нуля только коэффициент c_0. При этом координатная ее зависимость в случае (11) становится линейной, что отвечает, в частности, постоянному градиенту концентрации (и локальной средней энергии). С другой стороны, нарушение равновесия по энергии неизбежно связано с нелинейным видом координатной зависимости f_s[1].

Одна из важных задач теории горячих электронов состоит в вычислении кинетических коэффициентов в тех или иных условиях опыта. В частности, в условиях, когда нагрев вызван пространственно однородным внешним электрическим полем напряженности $\vec{E}$, подвижность, коэффициент диффузии и другие аналогичные величины удается явно выразить через $\vec{E}$ и параметры системы. Так же обстоит дело и в условиях, близких к однородным (1). Это обстоятельство оказалось очень существенным, ибо оно обусловило возможность полуфеноменологического подхода к решению ряда задач. Так, современную теорию доменной электрической неустойчивости в полупроводниках удалось построить в предположении, что кинетические коэффициенты суть некоторые apriori заданные функции какой-то одной макроскопически наблюдаемой величины (локальной напряженности поля [5—7] или локальной потребляемой мощности [8]). Из выражения (13) явствует, что в рассматриваемом случае, который мы будем называть существенно нелокальным, дело обстоит иначе. Действительно, плотность потока частиц есть

$$\vec{j} = \frac{g}{(2\pi\hbar)^3} \int d\vec{p}\,\vec{v}\,(\vec{p})\,f_a, \tag{19}$$

где g — статистический вес (для электронов $g=2$), $\vec{v}$ — групповая скорость (в рассматриваемом случае $\vec{v}=\vec{p}/m$). В стационарной задаче равенство (4) дает

$$f_a = -\tau\,(\vec{v}, \nabla f_s) - \tau\,(e\vec{E}, \nabla_p f_s). \tag{4'}$$

Комбинируя формулы (4'), (19) и (15), получаем

$$\vec{j} = \frac{g\,(2m)^{1/2}}{3\,(2\pi^2\hbar^3)} \sum_\lambda c_\lambda \int_0^\infty dW f_\lambda(W) \left\{ e\vec{E}\,\frac{d}{dW}\,(\tau W^{3/2})\,\varphi_\lambda - W^{3/2}\tau\nabla\varphi_\lambda \right\}. \tag{20}$$

В рассматриваемом случае нет особого смысла разделять плотность тока на диффузионную и дрейфовую составляющие. Имея в виду, однако, поставленный выше вопрос, мы проделаем это, полагая

$$\vec{j}_{dr} = \frac{e}{|e|}\,\vec{E}\mu n, \quad j_{dif} = -\nabla\,(nD),$$

[1] Эти утверждения не связаны с явным видом интеграла столкновений (16), а носят вполне общий характер. Действительно, число $\lambda=0$ всегда есть собственное значение оператора J, принадлежащее собственной функции e^{-W}.

где n есть концентрация частиц:

$$n = \frac{g(2m^3)^{1/2}}{2\pi^2\hbar^3}\sum_\lambda c_\lambda\varphi_\lambda(x)\int_0^\infty W^{1/2}f_\lambda(W)\,dW,$$

а подвижность μ и коэффициент диффузии D даются выражениями

$$\mu = \frac{|e|}{m}\,\frac{\sum\limits_\lambda c_\lambda\varphi_\lambda(x)\int\limits_0^\infty dW W^{1/2}\tau f_\lambda\left(1+\frac{2}{3}\frac{d\ln\tau}{d\ln W}\right)}{\sum\limits_\lambda c_\lambda\varphi_\lambda\int\limits_0^\infty dW W^{1/2}f_\lambda}, \tag{21}$$

$$D = \frac{2}{3m}\,\frac{\sum\limits_\lambda c_\lambda\varphi_\lambda\int\limits_0^\infty dW W^{3/2}\tau f_\lambda}{\sum\limits_\lambda c_\lambda\varphi_\lambda\int\limits_0^\infty dW W^{1/2}f_\lambda}. \tag{22}$$

Аналогичные выражения легко получаются и для других кинетических коэффициентов.

В существенно нелокальном случае кинетические коэффициенты даже в отсутствие вырождения явно зависят от локальной концентрации частиц и не могут быть выражены в виде функций от одной-двух макроскопических переменных.

Первое обстоятельство не должно вызывать удивления: возможность ввести кинетические коэффициенты, не зависящие от концентрации, есть следствие представимости f_s в виде произведения двух сомножителей, из которых один зависит только от энергии, а другой — только от координат. Согласно (15), в рассматриваемой задаче свойство мультипликативности не имеет места.

Второе обстоятельство, по-видимому, исключает возможность привычного полуфеноменологического описания рассматриваемой системы: состояние последней задается только совокупностью всех коэффициентов c_λ. Это означает, в частности, что для описания свойств неравновесной нелокальной системы нужны довольно детальные сведения о граничных условиях.

§ 3. Флуктуации

Из результатов предыдущего параграфа следует, что, интересуясь временным поведением флуктуаций в существенно нелокальной системе, мы должны рассматривать непосредственно саму функцию распределения. Для этого обратимся к уравнению (3), считая, что при $t=0$ задано значение $f_s(W, x, 0) = F_s(W, x)$. Разумеется, описывая флуктуации, функция F_s не обязана совпадать со стационарным значением (15). Подвергая уравнение (8) одностороннему преобразованию Лапласа, для лапласовского образа $f_s(W, \vec{x}, p)$, получим уравнение

$$pf_s(p) - \hat{L}[f_s(p)] = F_s - \frac{\tau\tilde{F}}{1+\tau p}. \tag{23}$$

Здесь p — комплексный лапласовский параметр,

$$\widehat{L}[f_s] = \frac{2W}{3m} \frac{\tau}{1+p\tau} \nabla^2 f_s + J[f_s]. \tag{24}$$

Граничные условия, накладываемые на функцию $f_s(W, x, p)$ по переменным W, x — те же, что и для соответствующего стационарного состояния.

Введем собственные функции, ψ_μ и собственные значения $\mu(p)$ оператора $\widehat{L}$:

$$\widehat{L}[\psi_\mu] = \mu\psi_\mu. \tag{25}$$

Полагая

$$F_s = \sum_\mu b'_\mu \psi_\mu, \quad \widetilde{F} = \sum_\mu b''_\mu \psi_\mu, \tag{26}$$

находим

$$f_s(W, x, p) = \sum_\mu \left(b''_\mu - \frac{b''_\mu}{1+p\tau} \right) \frac{1}{p-\mu(p)}. \tag{27}$$

Выполняя обратное преобразование Лапласа, видим, что функция $f_s(W, x, t)$ представляет собой сумму экспонент, одна из которых, пропорциональная b''_μ, описывает затухание с постоянной времени τ, а другие имеют вид $e^{p_i t}$, где p_i — корни уравнения

$$p_i - \mu(p_i) = 0, \quad i = 1, 2, \ldots . \tag{28}$$

При этом под $\mu(p)$ понимается и аналитическое продолжение этой функции в область $Re\,p<0$. Из формулы (24) и уравнения (9) видно, что при $p=0$ оператор $\widehat{L}$ имеет собственное значение $\mu=0$, а принадлежащая ему собственная функция ψ_0 есть не что иное, как стационарная функция распределения $f_s(W, x)$. Этот результат был, разумеется, заранее очевиден, равно как и вывод о затухании флуктуаций антисимметричной функции распределения (ср. (26)) со временем τ. Не вполне тривиален, однако, вопрос о характере затухания флуктуаций симметричной части. Легко убедиться в том, что в неравновесной системе оператор L (при $p \in Re$), вообще говоря, индефинитен, т. е. может иметь и положительные собственные значения. При этом уравнение (28) в принципе могло бы иметь и положительные корни, что отвечало бы неустойчивости системы. Далее, при комплексных значениях p оператор $\widehat{L}$ становится не эрмитовским, что, с учетом (28), наводит на мысль о возможности колебательной релаксации в рассматриваемой системе. Подобное исследование этих возможностей составляет, однако, особую задачу.

Автор весьма признателен В. Л. Гуревичу и Э. И. Рашбе за обсуждение работы.

ЛИТЕРАТУРА

1. Бонч-Бруевич В. Л. «Физика твердого тела», 3, № 7, 1969.
2. Давыдов Б. И. ЖЭТФ, 7, 1069, 1937.
3. Грибников З. С., Мельников В. И. «Физика твердого тела», 7, 1997, 1965.
4. Лебедев Н. Н. Специальные функции и их приложения. М., Гостехиздат, 1953.
5. Бонч-Бруевич В. Л. «Физика твердого тела», 8, 1753; 1966; Knight B. K., Peterson C. E. Phys. Rev., 155, 393, 1967.
6. Волков А. Ф. «Физика твердого тела», 8, 3187, 1966.
8. Анисимов С. И., Мельников В. И., Рашба Э. И. Письма ЖЭТФ, 7, № 7, 1968.

Поступила в редакцию
22.1 1969 г.

Кафедра
физики полупроводников