

ГЕНЕРАЦИЯ НЕОСНОВНЫХ НОСИТЕЛЕЙ ЗАРЯДА В ЭЛЕКТРИЧЕСКИХ ПОЛЯХ

МАКРОСКОПИЧЕСКИХ НЕОДНОРОДНОСТЕЙ НА ГРАНИЦЕ РАЗДЕЛА ПОЛУПРОВОДНИК—ДИЭЛЕКТРИК

Е. И. Гольдман

Институт радиотехники и электроники Российской академии наук,
141120, Фрязино, Россия

(Получена 7.04.1992. Принята к печати 20.07.1992)

Предложена новая модель генерации неосновных носителей заряда, основанная на межзонной термоавтоэмиссии в электрических полях пятен высокой концентрации зарядов, встроенных на границе раздела полупроводник—диэлектрик. Развита теория электронных переходов между отвечающей такому потенциальной ямой и разрешенными зонами полупроводника. Построена температурная зависимость темпа рождения электронно-дырочных пар, график которой имеет необычный вид — кривой с двумя максимумами.

По современным представлениям образование канала инверсии у границы раздела (ГР) полупроводник—диэлектрик в МДП структуре происходит путем термической генерации неосновных носителей заряда (ННЗ), для определенности дырок, либо через глубокий уровень в приповерхностном слое обеднения, либо через поверхностные состояния на ГР. Из теории этих процессов [1–3] следует, что энергия активации темпа генерации ННЗ равна расстоянию от квазиуровня Ферми дырок до дна зоны проводимости на ГР при рождении электронно-дырочных пар через поверхностные состояния и максимуму между расстояниями от глубокого уровня до краев размещенных зон при генерации дырок в области объемного заряда. В обоих случаях получающаяся энергия активации превосходит половину ширины запрещенной зоны полупроводника. Поэтому, например для ГР Si—SiO₂, генерация дырок за типичные в релаксационных измерениях времена ~1 мин должна происходить при температурах вблизи или выше комнатной. Однако в некоторых экспериментах показано, что рождение ННЗ имеет место и при существенно более низких температурах. Так, в [4] представлен пик DLTS, отвечающий релаксации высокочастотной емкости Si МОП структуры в процессе образования канала инверсии, с максимумом при 160 К. В принципе ускорение генерации дырок может быть достигнуто за счет уменьшения эффективной энергии активации под влиянием электрического поля приповерхностного слоя обеднения. Но «низкотемпературная» генерация ННЗ наблюдается в том числе и в структурах с концентрациями легирующей примеси 10¹⁴ см⁻³, при которых эти электрические поля малы и не могут существенно увеличить вероятности переходов носителей заряда с локализованных электронных состояний в разрешенные зоны. В этой связи в данной работе развивается новая модель генерации ННЗ, основанная на межзонной электронной термоавтоэмиссии в электрических полях заряженных неоднородностей на ГР полупроводник—диэлектрик.

Будем рассматривать релаксацию МДП структуры к состоянию глубокой инверсии, в котором флуктуационный потенциал на ГР практически экранирован

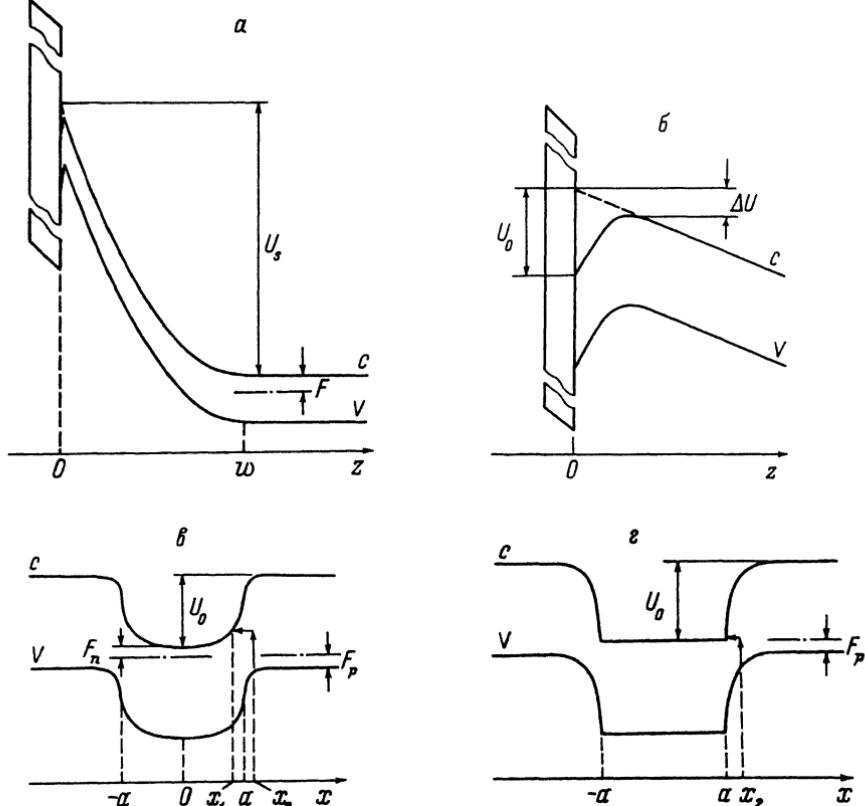


Рис. 1. Зонная диаграмма полупроводника в окрестности пятна высокой концентрации зарядов на границе с диэлектриком. а — потенциальный рельеф в слое обеднения полупроводника; б — фрагмент потенциального рельефа в слое обеднения вблизи ГР полупроводник—диэлектрик; в — потенциальный рельеф вдоль ГР ($z = 0$). Случай «незаполненной» электронами потенциальной ямы; г — потенциальный рельеф вдоль ГР ($z = 0$). Случай «заполненной» электронами потенциальной ямы. $z > 0$ — полупроводник, $z < 0$ — диэлектрик, с — зона проводимости, V — валентная зона.

свободными дырками. В отличие от [5], где генерация ННЗ была обусловлена туннельными межзонными переходами электронов между соседними максимумом и минимумом флуктуационного потенциала, в нашей ситуации главную роль играют электрические поля отдельных, остающихся не полностью экранированными потенциальных ям, отвечающих «пятнам» с высокой концентрацией встроенного положительного заряда.¹ На рис. 1 изображен потенциальный рельеф такой ямы на ГР n -полупроводник—диэлектрик МДП структуры, находящейся в состоянии глубокого неравновесного обеднения. Он складывается из кулоновской энергии притяжения U_1 зарядов, создающих яму, и потенциала слоя обеднения $U_{II} = U_s(1-z/w)^2$, где U_s — поверхностный изгиб зон в полупроводнике, z — координата по нормали к ГР, w — ширина слоя обеднения. Зависимость потенциальной энергии $U = U_1 + U_{II}$ от координаты z имеет вид, типичный для эффекта Френкеля—Пула (рис. 1, б). Высота барьера для выхода электрона из ямы глубиной U_0 под влиянием поля слоя обеднения понижается на величину ΔU , которая существенно превосходит уменьшение энергии активации однозарядного локализованного состояния в таком же (как и поле обед-

¹ Максимумы потенциала, отвечающие скоплению отрицательных зарядов, будут «заливаться» дырками в первую очередь.

ненного слоя) электрическом поле. Последнее обстоятельство обусловлено тем, что число нескомпенсированных зарядов в пятне много больше единицы.² Из-за наличия резкого края у пятна встроенного заряда изменение потенциала вдоль ГР (рис. 1, в—г) происходит круче, чем по нормали z . Так, при $x = a \partial U_0 / \partial x = \infty$, где x — координата вдоль ГР, a — радиус пятна. В зависимости от степени заполнения потенциальной ямы электронами могут реализовываться два противоположных случая: «не заполненной» (рис. 1, в) и «заполненной» (рис. 1, г) ям. Во втором случае пятно на ГР фактически можно рассматривать как металлическое.

Рождение электронно-дырочных пар происходит путем термотуннельных переходов в потенциале $U(x, z)$ (указанны стрелками на рис. 1, в и 1, г). При этом дырки растекаются по ГР, а электроны локализуются в яме, откуда затем термоэмиссионным образом через барьер высотой $U_0 - \Delta U$ (рис. 1, б) стекают в объем полупроводника. Кинетика заполнения потенциальной ямы описывается уравнениями:

$$\frac{dN}{dt} = J_p - J_n, \quad (1)$$

$$J_p = \alpha_p e^{-\frac{\epsilon}{T}} \left(1 - e^{\frac{E_g - F_p - F_n - U_0}{T}} \right), \quad (2)$$

$$J_n = \alpha_n e^{-\frac{U_0 - \Delta U + F_n}{T}} \left(1 - e^{-\frac{U_s + F - U_0 - F_n}{T}} \right). \quad (3)$$

Здесь N — число электронов в яме; J_p — темп генерации—рекомбинации электронно-дырочных пар в яме; J_n — разность потоков электронов из ямы в объем полупроводника и обратно; E_g — ширина запрещенной зоны; F_p , F_n и F — энергии Ферми дырок на ГР, электронов в яме и в электронейтральном объеме полупроводника соответственно; T — температура в энергетических единицах.

$$\epsilon = E_g - \max \left\{ [U_1(x_2) - U_1(x_1)] - \left(\frac{8mT^2}{\hbar^2} \right)^{1/2} \int_{x_1}^{x_2} [U_1(x) - U_1(x_1)]^{1/2} dx \right\},$$

$$0 < x_1 < x_2, \quad z = 0, \quad (4)$$

x_1 и x_2 — координаты точек поворота (см. рис. 1, в—г), подбором которых минимизируется величина ϵ [для случая заполненной ямы (рис. 1, г) $x_1 = a$], \hbar — постоянная Планка, m — эффективная масса; α_p и α_n — предэкспоненциальные факторы, по порядку величины α_p должен совпадать с нормированным на площадь пятна ($\sim a^2$) и деленным на элементарный заряд предэкспоненциальным множителем в полученном Л. В. Келдышем^[7] выражении, описывающем, например, полевую и температурную зависимости термотуннельного тока пробоя обратно смешенного $p-n$ -перехода, α_n — определяется произведением средней тепловой скорости электронов, эффективной плотности состояний в зоне проводимости и эффективной площади термоэмиссионного потока через перевал потенциала $U(x, z)$ (рис. 1, б). Конкретный вид величины ϵ как функции U_0 , a , T и других параметров зависит от формы потенциала $U_1(x)|_{z=0}$, который в свою очередь определяется распределениями зарядов в пятне и дырок на ГР. Эффект экранирования в двумерном дырочном газе существенно не локален^[8],

² Подобный эффект на флюктуационных потенциальных ямах во многом определяет динамику перезарядки флюктуационных поверхностных состояний^[6].

поэтому аналитический расчет зависимости $U_1(x, z)$ затруднен. Простое выражение для ϵ получается при условии

$$\left(\frac{-^2 U_0}{8mT^2 a^2} \right) \gg 1, \quad \epsilon \approx E_g - U_0. \quad (5)$$

Фактически неравенство (5) — это приближение «прямоугольной» ямы. Скорость нарастания концентрации свободных дырок на ГР удовлетворяет уравнению

$$\frac{dp_s}{dt} = \frac{J_p}{s} - R, \quad (6)$$

где s — площадь ГР, приходящаяся на одну яму (т. е. $1/s$ — концентрация пятен), R — суммарный темп локализации дырок в максимумах флюктуационного потенциала и их рекомбинации с электронами, захваченными на пограничные состояния не «флюктуационной» природы.

Проанализируем закономерности генерации ННЗ в поле пятна встроенного заряда. Потенциальная яма играет роль эффективного центра генерации с необычными свойствами: энергии активации электронов в зону проводимости и дырок в валентную зону равны соответственно $U_0 - \Delta U$ и ϵ , они зависят от электрического поля слоя обеднения, концентрация дырок p_s и заполнения ямы N ; сечения захвата электронов и дырок на центр аномально велики — по порядку величины не менее площади пятна. Как и при рождении электронно-дырочных пар через глубокие уровни в полупроводнике или на ГР [1^{-3}] генерация ННЗ происходит в квазистационарных условиях для заполнения центра генерации. Полагая в (1) $dN/dt = 0$, из выражений (2) и (3) имеем:

при

$$(\alpha_p/\alpha_n) \exp [(E_g - F_p - \Delta U - \epsilon)/T] \ll 1$$

$$U_0 + F_n = \Delta U + \epsilon + T \ln \frac{\alpha_n}{\alpha_p}, \quad (7)$$

$$J_p = \alpha_p e^{-\frac{\epsilon}{T}}; \quad (8)$$

при

$$(\alpha_p/\alpha_n) \exp [(E_g - F_p - \Delta U - \epsilon)/T] \gg 1,$$

$$U_0 + F_n = E_g - F_p, \quad (9)$$

$$J_p = \alpha_n e^{-\left(\frac{(E_g - \Delta U - F_p)}{T} \right)} \left[1 - e^{-\frac{(E_g - F_p - F - U_s)}{T}} \right]. \quad (10)$$

Уравнения (7) и (8) описывают раннюю стадию генерации, пока дырок на ГР мало (их энергия Ферми F_p велика). Рождение носителей заряда происходит за счет последовательных электронных переходов из валентной зоны в яму и из ямы в объем полупроводника. Если изначально глубина пустой ямы меньше, чем величина правой части равенства (7), то яма остается слабо заполненной, U_0 и Δ не изменяются в процессе генерации, а само уравнение (7) определяет энергию Ферми электронов F_n . В противоположном случае яма «заливается» электронами, $F_n \ll U_0$ и уравнение (7) определяет ее глубину. Пользуясь известным выражением [9] для потенциала, создаваемого тонким заряженным металлическим диском, $U_1(z)|_{x=0} = U_0[(2/\pi) \operatorname{arctg}(z/a) - 1]$, полагая, что выпол-

няется приближение прямоугольной ямы (5) и поле в ней гораздо больше, чем поверхностное поле в слое обеднения $U_0/a \gg U_s/w$, из (7) и (8) получаем

$$U_0 = \frac{E_g}{2} + \left[\frac{2aU_s}{\pi w} \left(E_g + T \ln \frac{\alpha_n}{\alpha_p} \right) \right]^{1/2} + T \ln \frac{\alpha_n}{\alpha_p}, \quad (11)$$

$$J_p = (\alpha_n \alpha_p)^{1/2} e^{-\frac{E_g}{T}}, \quad E_1 = \frac{E_g}{2} - \left[\frac{2aU_s}{\pi w} \left(E_g + T \ln \frac{\alpha_n}{\alpha_p} \right) \right]^{1/2}. \quad (12)$$

Временная зависимость концентрации дырок на ГР, а следовательно, и их энергии Ферми, могут быть найдены из уравнения (6). Поскольку рекомбинация ННЗ — «быстрый», безактивационный процесс, то $J_p \approx R$. Отсюда следует, что на ранней стадии генерации концентрация дырок на ГР практически не растет, а сама стадия продолжается по крайней мере до тех пор, пока не будут «залиты» все максимумы флуктуационного потенциала и не рекомбинируют все электроны, захваченные на пограничные состояния.

На следующем этапе рождения дырок их концентрация на ГР возрастает настолько, что между ними и электронами в яме устанавливается квазиравновесие [см. выражение (9)]. При этом для не глубокой ямы ($U_0 < E_g - F_p$) заполнение остается слабым, $U_0 \approx \text{const}$, а уравнение (9) определяет энергию Ферми электронов в ней; для заполненной ямы $F_p \ll U_0$ и $U_0 \approx E_g - F_p$. Временная зависимость концентрации ННЗ может быть найдена из уравнения (6) при подстановке в него выражения (10) и условия $R = 0$. Определим эту зависимость для термостимулированного режима релаксации МДП структуры $dT/dt = \beta = \text{const}$ в случае заполненной ямы, полагая, что концентрация дырок еще достаточно мала, так что, во-первых, эффектом экранирования ими поля пятна можно пренебречь и, во-вторых, $p_s \ll N_d W$ (N_d — концентрация полностью ионизированных доноров в полупроводнике). Тогда $\Delta U = 4(U_0 U_s a / \pi w)^{1/2}$, $N_v \exp(-F_p/T) = 2p_s U_s / T w$ и в приближении $F_p \ll E_g$ получаем

$$p_s = \left(\frac{\alpha_n t_{\text{eff2}}}{s} \right)^{1/2(1-\nu)} \left(\frac{T w N_v}{2 U_s} \right)^{1/2(1+\nu)} e^{-\frac{E_2}{T}}, \quad J_p = \frac{p_s s}{t_{\text{eff2}}},$$

$$t_{\text{eff2}} = \frac{T^2}{\beta E_2}, \quad E_2 = \frac{E_g}{2} - \frac{3}{2} \left(\frac{E_g U_s a}{\pi w} \right)^{1/2}, \quad \nu = \left(\frac{a U_s}{\pi w E_g} \right)^{1/2}, \quad (13)$$

N_v — эффективная плотность состояний в валентной зоне. Формулы (13) будут описывать генерацию ННЗ и в изотермическом режиме ($T = \text{const}$), если в них произвести замену эффективного времени t_{eff2} на реальное $t - t_1$, где t_1 — момент установления квазиравновесия электронов в яме и дырок в валентной зоне. Для не заполненной ямы релаксационные зависимости также имеют вид (13) при $\nu = 0$ и $E_2 = (E_g/2) - (U_0 U_s a / w)^{1/2}$.

Рост числа дырок на ГР приводит к уменьшению величины ΔU за счет экранирования и (или) снижения поля слоя обеднения (U_s/w) в процессе релаксации МДП структуры к равновесному состоянию глубокой инверсии. В результате, как следует из уравнений (6) и (10), генерация ННЗ замедляется, концентрация p_s медленно (логарифмически) растет со временем, а темп рождения электронно-дырочных пар

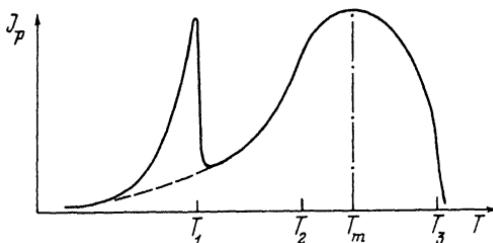


Рис. 2. Температурная зависимость темпа генерации неосновных носителей заряда.

$$J_p = \frac{J_{p2} e^{\frac{E_g}{T_2} - \frac{1}{T}}}{1 + \frac{t_{\text{eff}}}{sT} \left(-\frac{d\Delta U}{dp_s} \right) J_{p2} e^{\frac{E_g}{T_2} - \frac{1}{T}}}, \quad (14)$$

где $t_{\text{eff}} = T^2/\beta E_g$ в термостимулированном и $t_{\text{eff}} = t - t_2$ в изотермическом режимах, t_2 — время, отвечающее началу заметного проявления эффектов дырочного экранирования и (или) снижения поля в слое обеднения, $J_{p2} = J_p|_{t=t_2}$, $T_2 = T|_{t=t_2}$. Такое поведение указанных величин со временем (температурой) продолжится до тех пор, пока концентрация дырок не достигнет значений, близких к равновесному $p_{s\infty}$, соответствующему данной температуре и смещению между полевым электродом и полупроводником. После этого становится существенным отличие от единицы множителя в квадратных скобках выражения (10). В линейном приближении для темпа генерации J_p как функции разности концентраций $p_{s\infty} - p_s$ из уравнений (6) и (10) находим

$$J_p = J_{p3} e^{\frac{E_3}{T_3} - \frac{1}{T}} \exp \left[-\frac{\alpha_n t_{\text{eff}3}}{sp_*} e^{-\frac{E_3}{T}} \right],$$

$$E_3 = E_g - (\Delta U + F_p)|_{p_s=p_{s\infty}}, \quad p_* = T \left[-\frac{d(U_s + F_p)}{dp_s} \right]_{p_s=p_{s\infty}}^{-1}, \quad (15)$$

$t_{\text{eff}3} = T^2/\beta E_3$ в термостимулированном и $t_{\text{eff}3} = t - t_3$ в изотермическом режимах, t_3 — момент времени, отвечающий переходу к равновесию ($p_{s\infty} - p_s < p^*$ при $t = t_3$), $J_{p3} = J_p|_{t=t_3}$, $T_3 = T|_{t=t_3}$.³

На рис. 2 представлен качественный вид графика температурной зависимости темпа генерации ННЗ. Его форма одинакова как для термостимулированного, так и для изотермического режимов релаксации МДП структуры к состоянию глубокой инверсии (в изотермическом случае имеется в виду зависимость $J_p(T)$ при заданном времени наблюдения t). Участок характеристики, где $T < T_1$ описывается выражениями (8) или (12), соответствующими ранней стадии генерации, пока дырок на ГР мало. Температура T_1 отвечает окончанию этапа, на котором вновь рождающиеся дырки локализуются в максимумах флюктуационного потенциала или рекомбинируют с электронами, захваченными на пограничные состояния. При $T > T_1$ в узком температурном интервале энергия Ферми дырок резко уменьшается, а их концентрация увеличивается вплоть до установления

³ Формула (15) по внешнему виду совпадает с выражением для температурной зависимости тока термостимулированного опустошения глубокого уровня в полупроводнике [10]. Эта зависимость имеет форму пика, у которого температура максимума меньше, чем T_3 . Соотношение (15) справедливо при $T > T_3$ и, следовательно, описывает спадающую ветвь указанного пика.

квазиравновесия с электронами в яме. В результате энергия активации темпа генерации значительно возрастает [см. (10)], что проявляется на кривой $J_p - T$ в виде пика с максимумом вблизи $T = T_1$. Отметим, что в отсутствие локализации и рекомбинации дырок на ГР переход к квазиравновесию произошел бы плавно и на графике зависимости $J_p (T)$ (пунктирная кривая на рис. 2) в окрестности этого перехода максимума бы не было. На участке $T_1 < T < T_2$ количество дырок на ГР еще не достаточно для экранирования потенциала пятна и сужения области объемного заряда. Здесь для темпа генерации ННЗ справедливы выражения (13). Эффекты экранирования и (или) уменьшения ширины слоя обеднения в полупроводнике начинают сказываться при $T > T_2$. В этом диапазоне вплоть до перехода к равновесию в окрестности $T = T_3$ характеристика описывает относительно пологий максимум [см. (14)], форма которого и температура экстремума T_m определяются зависимостью $\Delta U [p_s, w(p_s)]$. Так как $p \ll p_{ss}$, то температурный интервал перехода к равновесию (квазиравновесию в случае $dT/dt \neq 0$) $T > T_3$ значительно уже отрезка $[T_2, T_3]$. Темп генерации ННЗ спадает по закону (15). Заметим, что если рождение электронно-дырочных пар происходит в условиях поддержания высокочастотной емкости МДП структуры постоянной за счет синхронного с накоплением ННЗ увеличения полевого напряжения, то $w = \text{const}$, равновесие принципиально не достижимо, а максимум на графике $J_p - T$ при $T > T_2$ может и не проявиться из-за электрического пробоя образца. Скорость генерации дырок зависит от начального изгиба зон в полупроводнике. С ростом U_s температуры T_1 , T_2 и T_m поникаются из-за уменьшения соответствующих энергий активации, а температура T_3 повышается из-за увеличения равновесной концентрации дырок. Пик на кривой $J_p - T$ при $T = T_1$ с увеличением U_s становится более узким, а его площадь сохраняется.

Построенная теория генерации ННЗ исходит из представлений о наличии на ГР полупроводник—диэлектрик пятен с высокой концентрацией встроенного заряда. В принципе эти пятна могут отличаться друг от друга и по числу зарядов, и по величине диаметра. Выше было показано, что наиболее действенные глубокие потенциальные ямы, отвечающие таким пятнам, заливаются электронами. При этом устанавливается единая глубина для всех ям и энергия активации темпа генерации дырок через них не зависит от числа зарядов в пятне и слабо изменяется с a . Оценим, каковы допустимые значения радиуса пятна и числа зарядов в нем N_+ , чтобы рождение ННЗ через этот центр было эффективным. Туннелирование существенно при $(\hbar^2 E_g / m T^2 a^2) \geq 1$; глубина пустой ямы должна быть больше ширины запрещенной зоны полупроводника $(N^+ g^2 / \bar{\alpha} a) > E_g$, где q — элементарный заряд, $\bar{\alpha} = (\alpha_s + \alpha_i)/2$, α_s и α_i — диэлектрические проницаемости полупроводника и диэлектрика соответственно. Из этих неравенств для ГР Si—SiO₂ при 200 К получается $a < 3.6 \cdot 10^{-6}$ см, $N_+ > 240$, что отвечает минимальной концентрации зарядов в пятне $6 \cdot 10^{12}$ см⁻². Данные величины параметров вполне реальны для Si-МОП структур, и это позволяет надеяться на экспериментальное обнаружение рассмотренных эффектов.

Автор признателен А. Г. Ждану и А. М. Сумароке за полезные дискуссии.

СПИСОК ЛИТЕРАТУРЫ

- [1] J. G. Simmons, H. A. Mar. Phys. Rev., B8, 3865 (1973).
- [2] P. U. Calzolari, S. Graffi, C. Morandi. Sol. St. Electron., 17, 1001 (1974).
- [3] H. A. Mar, J. G. Simmons. Phys. Rev., B11, 775 (1975).
- [4] D. Vuillaume, M. Lannoo, J. C. Bourgois, E. Rosenthaler. J. Appl. Phys., 66, 5920 (1989).
- [5] B. A. Гергель, В. М. Масловский, Р. А. Сурик. ЖЭТФ, 85, 1377 (1983).
- [6] B. A. Гергель. ФТП, 17, 637 (1983).
- [7] Л. В. Келдыш. ЖЭТФ, 34, 962 (1985).

- [8] Т. Андо, А. Фаулер, Ф. Стерн. Электронные свойства двумерных систем, 415. М. (1958).
- [9] Л. Д. Ландау, Е. М. Лифшиц. Теоретическая физика, т. 8. Электродинамика сплошных сред, 37. М. (1982).
- [10] А. Г. Ждан, В. Б. Сандомирский, А. Д. Ожередов. ФТП, 2, 11 (1968).

Редактор Ю. Б. Лянда-Геллер
