

©1995 г.

ФАЗОВЫЙ ПЕРЕХОД ДИЭЛЕКТРИК–МЕТАЛЛ: ТЕРМОДИНАМИЧЕСКИЙ АСПЕКТ ПРОБЛЕМЫ

Д.Н.Бычковский, О.В.Константинов, Б.В.Царенков

Физико-технический институт им. А.Ф. Иоффе Российской академии наук,
194021, Санкт-Петербург, Россия
(Получена 27 июня 1994 г. Принята к печати 10 июля 1994 г.)

В известных работах Мотта по фазовому переходу диэлектрик–металл этот переход связывался с экранировкой кулоновского поля донора или акцептора свободными носителями заряда, и считалось, что его условием является выталкивание примесного уровня в зону проводимости.

Оказывается, что фазовый переход диэлектрик–металл будет термодинамически устойчивым лишь в том случае, если произойдет достаточно сильное сужение запрещенной зоны полупроводника, обусловленное свободными носителями заряда.

Было учтено два механизма сужения зоны — корреляционный (Дебая–Хюккеля) и обменный (Вигнера–Зейтца). Показано, что эти два механизма приводят к термодинамической устойчивости перехода диэлектрик–металл. Проведено сравнение теоретических значений с экспериментальными данными в случаях GaAs и InP.

Проведен расчет энергии связи донора в GaAs при различных концентрациях электронов.

Введение

Теория фазового перехода диэлектрик–металл в системе донорных или акцепторных атомов в полупроводнике впервые была дана Моттом [1,2] и обобщена им в монографии [3]. Рассмотрим для определенности доноры. Как было предсказано Моттом и подтверждено экспериментально, переход диэлектрик–металл происходит тогда, когда среднее расстояние между донорными атомами l_d ($l_d = N_d^{-1/3}$, N_d — концентрация доноров) становится примерно равным четырем боровским радиусам a_B донорного атома:

$$l_d \simeq 4a_B, \quad (1)$$

$$a_B = \frac{\epsilon \hbar^2}{me^2}. \quad (2)$$

Здесь ϵ — диэлектрическая проницаемость полупроводника, m — эффективная масса электрона в полупроводнике. Мы рассматриваем для простоты однодолинный полупроводник с изотропным квадратичным законом дисперсии. Критерий (1) фазового перехода был получен Моттом в рамках следующей модели. Рассматривается диэлектрическая сторона перехода, при концентрации доноров меньше критической величины, определяемой формулой (1). Производится мысленный эксперимент, в котором все электроны удаляются с примесных атомов и образуют при температуре $T = 0$ вырожденный электронный газ. Потенциальная энергия взаимодействия электрона с ионом донора не является больше кулоновской, а становится дебаевской за счет экранирования иона свободными электронами:

$$U(r) = -\frac{e^2}{\epsilon r} \exp\left(-\frac{r}{r_D}\right), \quad (3)$$

где

$$r_D = \left(\frac{\epsilon \mathcal{E}_F}{6\pi n e^2}\right)$$

— радиус дебаевского экранирования в вырожденном ферми-газе с энергией Ферми \mathcal{E}_F , n — концентрация электронов в зоне проводимости. Если учесть, что при нулевой температуре энергия Ферми \mathcal{E}_F равна

$$\mathcal{E}_F = \frac{\hbar^2 k_F^2}{2m}; \quad k_F = (3\pi n^2)^{1/3}, \quad (4)$$

где k_F — фермиевский импульс, m — эффективная масса, то формула для дебаевского радиуса экранирования принимает вид

$$r_D^2 = \left(\frac{\pi}{192}\right)^{1/3} \frac{a_B}{n^{1/3}} \simeq 0.25 l a_B. \quad (5)$$

Энергия основного состояния для потенциала Дебая $\varphi(r) = U(r)/e$ (3) находится вариационным методом и оказывается тем выше, чем больше концентрация электронов в плазме. С ростом концентрации энергия связи уменьшается и обращается в нуль при некоторой критической концентрации $N_d = N_M$. Это и есть, согласно Мотту, то значение концентрации, при котором происходит фазовый переход. При этом радиус дебаевского экранирования r_D оказывается равным боровскому радиусу a_B , что соответствует среднему расстоянию между донорами

$$l_M = a_B \sqrt[3]{192/\pi} \simeq 3.94 a_B. \quad (6)$$

Это и есть упомянутый выше критерий Мотта (1).

Вскоре после первых работ Мотта, проблема была изучена Андерсоном на основе другого подхода. Рассматривалась уже не диэлектрическая, а металлическая сторона фазового перехода, и ставился вопрос: что будет происходить с электропроводностью вырожденного электронного газа, когда концентрация доноров (и электронов) не увеличивается, а уменьшается, приближаясь к критической. Экспериментально известно, что при этом электропроводность падает на несколько порядков величины. Согласно Андерсону [4] физическая причина обращения электропроводности в нуль при $N_d = N_M$ заключается

в локализации состояния сплошного спектра благодаря интенсивному рассеянию электронов. По существу андерсоновская локализация имеет много общего с захватом электрона на связанное состояние донора.

Критическая концентрация доноров $N_d = N_M$ должна быть одной и той же, в независимости от того, как мы ее теоретически описываем, т.е. рассматривая диэлектрическую или металлическую сторону перехода. Однако при подходе Мотта можно изучать теоретически различные физические явления при концентрациях меньше критической, тогда как нет возможности проводить рассмотрение при концентрациях выше критической (т.е. на металлической стороне). При подходе Андерсона, напротив, можно изучать различные физические явления при концентрациях выше критической и нет возможности перейти на диэлектрическую сторону перехода. Таким образом, эти два подхода дополняют друг друга. Отметим также, что теория Андерсона оказывается менее разработанной с количественной стороны, чем теория Мотта, поскольку столь же конкретного выражения для критической концентрации, как формула (1), в подходе Андерсона не получается.

Следует отметить, что до сих пор в литературе упоминается и другая точка зрения на фазовый переход диэлектрик-металл, согласно которой этот переход связан с расщеплением примесных состояний в примесную зону и слиянием этой зоны с зоной проводимости. Такая концепция присутствует в ряде учебников, например, в [5]. Поэтому мы считаем необходимым остановиться на этом вопросе. В упомянутой модели по существу невозможно оправдать количественно экспериментально установленное условие (1) перехода диэлектрик-металл. Согласно (1), переход происходит при столь низких концентрациях легирующей примеси, что невозможно представить себе механизм расщепления мелких донорных уровней в примесную зону заметной ширины, а тем более слияния ее с дном зоны проводимости. Условие фазового перехода (1) означает, что концентрация доноров в момент фазового перехода такова, что между двумя соседними донорными атомами возможно поместить еще один, если считать эти атомы непроницаемыми сферами радиуса a_B . Это расстояние между соседними атомами в атомных масштабах очень велико. Если рассмотреть, например, энергию притяжения, ΔU , двух атомов водорода с противоположными спинами, когда расстояние между их центрами равно четырем боровским радиусам, то из [6] имеем

$$\Delta U = -0.05Ry. \quad (7)$$

В полупроводнике следует заменить единицу энергии ридберг (Ry) на эффективный ридберг:

$$Ry^* = \frac{e^2}{2\epsilon a_B}, \quad (8)$$

где a_B дается формулой (2). Расщепление уровней двух соседних атомов в полупроводнике при моттовской концентрации оказывается ничтожным. Так, для арсенида галлия $Ry = 5.3 \text{ мэВ}$, и, следовательно, $\Delta U \simeq -0.3 \text{ мэВ}$. Опускание энергии ΔU пары доноров, заполненных электронами с противоположными спинами, можно рассматривать как оценокное значение для средней энергии «хвоста» электронных состояний в запрещенной зоне, расположенного ниже невозмущенного донорного уровня. Примерно на такую же величину поднимается вверх

энергия пары доноров с параллельными спинами. Поскольку эта энергия в двадцать раз меньше ридберга, слияния примесной зоны с зоной проводимости никак нельзя ожидать. Полная величина размытия донорного уровня в примесную зону составит удвоенное значение (7), т.е. весьма малую величину, порядка $0.1Ry^*$.

1. Термодинамическая устойчивость перехода диэлектрик-металл

Согласно идее Мотта при критической концентрации доноров ($N_d = N_M$) электроны, покинув доноры, теряют возможность вернуться назад на донорные уровни, поскольку последние сливаются с зоной проводимости вследствие электронной экранировки. Это рассуждение представляется достаточно очевидным для безграничного полупроводника, в котором все электроны одновременно претерпели ионизацию. Если же электроны покинули донорные состояния в некоторой ограниченной области полупроводника, то это рассуждение теряет свою очевидность.

Пусть в некоторой области полупроводника (будем называть ее плазменной металлической каплей) произошла полная ионизация доноров, в результате которой при $T = 0$ там образовался вырожденный электронный газ (рис. 1). Уровень Ферми вырожденных электронов в капле $E_F^{(met)}$ будет проходить выше дна зоны проводимости. Уровень Ферми в окружающей каплю диэлектрике $E_F^{(diel)}$ будет проходить ниже дна зоны проводимости, посередине между ее дном и донорным уровнем. Поэтому металлическое состояние окажется термодинамически неустойчивым, и капля должна с неизбежностью рассосаться. Конечно, это происходит не одновременно для всех электронов внутри капли, а постепенно, с участием лишь электронов из приповерхностной области вблизи границы между плазменной каплей и диэлектриком. В результате поверхность раздела будет перемещаться внутрь капли, все более стягивая ее объем.

Можно провести аналогию с каплей воды, образовавшейся по какой-то причине внутри льда при температуре ниже точки замерзания. Согласно рассуждениям, аналогичным моттовским, можно было бы думать, что молекулам воды невозможно образовать кристалл

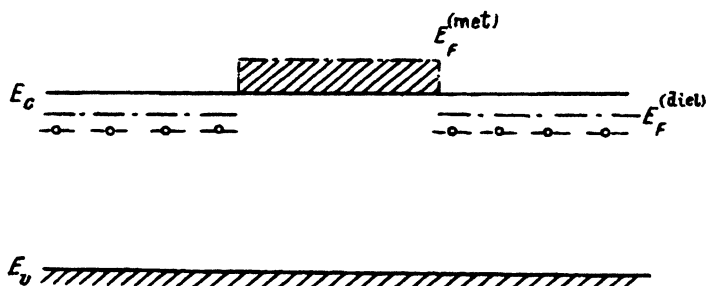


Рис. 1. Энергетическая диаграмма полупроводника, находящегося в диэлектрической фазе с плазменной каплей, внутри которой примесное донорное состояние электрона слилось с дном зоны проводимости из-за экранирования. Уровень Ферми вырожденных электронов в капле проходит выше уровня Ферми в диэлектрической фазе, и поэтому капля должна быть поглощена диэлектриком.

внутри капли, ввиду отсутствия кристаллического окружения (некуда вернуться). Однако при температуре ниже температуры замерзания термодинамический потенциал молекулы воды в жидкости будет выше, чем в окружающем кристалле. Поэтому молекулы вблизи границы капли будут переходить из жидкого состояния в кристаллическое, и диаметр капли будет сокращаться, пока она не исчезнет вовсе. Таким образом, переход диэлектрик-металл оказывается термодинамически неустойчивым по отношению к возврату в диэлектрическое состояние.

Единственная и решающая причина осуществления такого перехода — это понижение дна зоны проводимости [7], а точнее говоря, увеличение электронного сродства в области капли так, как это показано на рис. 2. Об электронном сродстве (χ_{met} , χ_{diel}) мы говорим здесь потому, что при этом автоматически предполагается постоянство вакуумного уровня ($E_{\text{vac}} = \text{const}$) во всей системе. Это предположение основано на факте электрической нейтральности как диэлектрического окружения, так и металлической капли. На рис. 2 показано не только понижение дна зоны проводимости, но также и повышение потолка валентной зоны. Таким свойством обладает один из рассмотренных далее механизмов, а именно корреляционный механизм сужения запрещенной зоны, предусматривающий одинаковую величину для опускания вниз дна зоны проводимости и подъема вверх потолка валентной зоны. Обменный механизм этим свойством не обладает. Он дает только вклад в опускание дна зоны проводимости.

Как видно из рис. 2, на котором изображено состояние термодинамически устойчивой плазменной капли (и неустойчивого диэлектрика), ее отличительное свойство состоит в том, что полная величина понижения дна зоны проводимости в плазме, $\Delta E_c^{(p)}$, должна быть больше, чем сумма энергии Ферми \mathcal{E}_F электронов в плазме, отсчитанной от дна зоны проводимости в капле, и энергии Ферми электронов на донорах, отсчитанной от дна зоны проводимости в диэлектрике (равная по модулю $Ry^*/2$). Математически это условие устойчивости плазменной капли будет даваться неравенством

$$\Delta E_c^{(p)} > \mathcal{E}_F + Ry^*/2, \quad (9)$$

где \mathcal{E}_F дается согласно (2), а Ry^* — формулой (8).

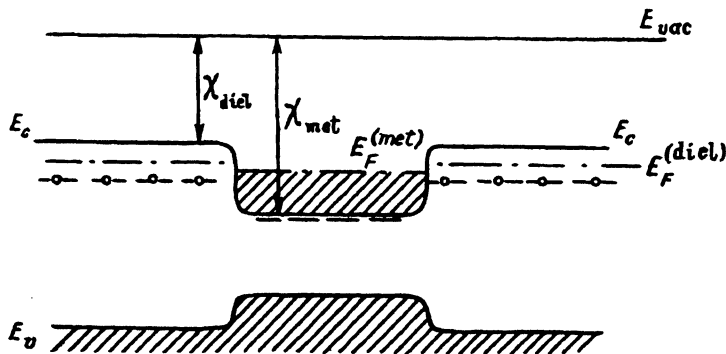


Рис. 2. Энергетическая диаграмма капли с учетом понижения в ней дна зоны проводимости. Уровень Ферми вырожденных электронов в капле проходит выше уровня Ферми в диэлектрической фазе, и поэтому капля растечется, поглотив диэлектрик.

2. Понижение дна зоны проводимости за счет корреляционного и обменного эффектов

Обсудим теперь известные из литературы механизмы понижения края зоны проводимости $\Delta E_c^{(p)}$, которое возникает благодаря существованию плазмы, состоящей из свободных электронов и положительных ионов. Мы будем учитывать только те механизмы, которые обязаны своим происхождением наличию газа свободных электронов. Таких механизмов два. Один из них — это так называемое корреляционное понижение дна зоны проводимости,

$$\Delta E_c^{(\text{corr})} = \frac{e^2}{2\epsilon r_D}, \quad (10)$$

где r_D — дебаевский радиус, определяемый формулой (3) в случае вырожденного газа. Формула (10) была впервые получена Дебаем и Хюккелем [8]. Их вывод воспроизводится в учебнике [9]. Кулоновская корреляционная энергия (10) считается в настоящее время главной причиной сужения запрещенной зоны в легированных полупроводниках [10]. Согласно [10], сужение запрещенной зоны должно быть в 1.5 раза больше величины, даваемой формулой (10). В нашей работе [11] было показано, что сужение запрещенной зоны должно быть не в полтора, а в два раза больше величины (10). К такому же выводу пришли и авторы работы [12]. Однако в работе [12] считается, что подавляюще велик не корреляционный, а обменный член. Обменное слагаемое, используемое в [12], почти в три раза превосходит общепринятую величину. Последняя впервые была получена Вигнером и Зейтцем и приводится в монографиях [13,14]. Она имеет вид

$$\Delta E_c^{(\text{exch})} = \frac{3e^2 k_F}{4\pi\epsilon}. \quad (11)$$

Здесь k_F — фермиевский волновой вектор, определяемый согласно (4). В итоге общее понижение дна зоны проводимости в плазменной капле мы берем в следующем виде:

$$\Delta E_c^{(p)} = \Delta E_c^{(\text{corr})} + \Delta E_c^{(\text{exch})}, \quad (12)$$

где слагаемые в правой части даются формулами (10) и (11). Для их нахождения вначале следует определить моттовскую концентрацию N_M с помощью (6). В табл. 1 приведены параметры, необходимые для расчета a_B и N_M , т.е. ϵ и m , для GaAs и InP. Далее приводятся расчетные значения боровского радиуса и моттовской концентрации. В последнем столбце приведены экспериментальные значения моттовской

Таблица 1.

	ϵ	m/m_0	$a_B, \text{Å}$	$N_M^{(\text{th})}, 10^{16} \text{ см}^{-3}$	$N_M^{(\text{exp})}, 10^{16} \text{ см}^{-3}$
GaAs	12.9	0.067	102	1.47	1.3
InP	12.6	0.077	87	2.4	1.5

Таблица 2.

	$\Delta E_c^{(corr)}$, мэВ	$\Delta E_c^{(exch)}$, мэВ	$\Delta E_c^{(p)}$, мэВ	\mathcal{E}_F , мэВ	Ry^* , мэВ	Σ , мэВ
GaAs	5.3	2.0	7.3	3.3	3.7	6.0
InP	6.6	2.5	9.1	3.9	3.3	7.2

концентрации из работы [15]. Эти экспериментальные значения неплохо согласуются с расчетными. При этом экспериментальные значения оказываются меньше теоретических. С помощью найденных значений $N_M^{(th)}$ были рассчитаны корреляционные и обменные энергии. Их значения помещены в табл. 2. В третьем столбце табл. 2 содержатся значения суммарного понижения дна зоны проводимости. Далее следуют значения энергии Ферми в капле, \mathcal{E}_F , и в диэлектрической фазе, $Ry^*/2$, а также их сумма Σ

$$\Sigma = \mathcal{E}_F + Ry^*/2. \quad (13)$$

Условие термодинамической реализуемости перехода диэлектрик-металл состоит в том, что $\Delta E_c^{(p)}$ должно быть больше Σ ($\Delta E_c^{(p)} > \Sigma$). Как видно из табл. 2, это условие действительно выполняется, и для его реализации необходимо учитывать не только корреляционную энергию, которая дает главный член в энергию понижения дна зоны проводимости, но также и обменное слагаемое.

3. Энергия основного состояния электрона на донорном атоме при учете экранирования

Движение электрона в поле донорной примеси при учете экранирования газом свободных электронов описывается уравнением Шредингера с изотропной эффективной массой m и потенциалом, имеющим форму потенциала Юкавы. Формулу (3) для него мы запишем в виде

$$U(r) = -U_0 \frac{r_D}{r} \exp(-r/r_D), \quad U_0 = \frac{e^2}{\epsilon r_D}, \quad (14)$$

где r_D — радиус экранирования, который определяется выражением (3). С ростом концентрации электронов радиус экранирования уменьшается, что ведет к уменьшению размера квантовой ямы, задаваемой (4). При значении концентрации электронов $n = N_M$ размер ямы достигает критического значения, при котором в ней больше не существует дискретных уровней энергии. Это соответствует тому, что все электроны переходят с донорных уровней в зону проводимости, т.е. наступает переход Мотта. Цель настоящего параграфа — решение волнового уравнения с потенциалом (14) при любой концентрации доноров.

Энергию основного состояния электрона на донорном атоме будем находить с помощью вариационной процедуры Ритца, используя пробную волновую функцию:

$$\Psi(r) = \left(\frac{\alpha^3}{\pi}\right)^{1/2} \exp(-\alpha r), \quad \int \Psi^*(r)\Psi(r)d^3r = 1, \quad (15)$$

где α — вариационный параметр Ритца. Следует заметить, что подобный выбор волновой функции является наилучшим, как это отмечается в монографии [16]. С помощью функции (15) определим среднее значение энергии $\langle E \rangle$, которое представляет собой сумму кинетической и потенциальной энергии:

$$\langle E \rangle = \frac{\hbar^2 \alpha^2}{2m^*} - U_0 \frac{4(r_D \alpha)^3}{(1 + 2r_D \alpha)^3}. \quad (16)$$

Для дальнейшего удобно ввести безразмерные переменные: безразмерный вариационный параметр p , характеризующий отношение среднего радиуса потенциальной ямы к среднему радиусу той области, где волновая функция существенно отличается от нуля,

$$p = 2r_D \alpha \simeq \alpha \sqrt{l a_B}, \quad (17)$$

и безразмерную мощность потенциальной ямы K ,

$$K = \frac{2m^* r_D^2}{\hbar^2} U_0 \simeq \sqrt{l/a_B}. \quad (18)$$

Средняя энергия как функция p имеет вид

$$\langle E \rangle = \left[\frac{p^2}{4K} - \frac{p^3}{2(1+p)^2} \right] U_0. \quad (19)$$

Оптимальное решение получим из условия $dE/dp = 0$, которое дает

$$K = \frac{(p+1)^3}{p(p+3)}. \quad (20)$$

Если подставить в левую часть этого равенства $K \simeq \sqrt{l/a_B}$, то уравнение (20) позволяет определить вариационный параметр p как функцию концентрации или среднего расстояния l между донорами. Удобно ввести безразмерную среднюю энергию связи γ , равную отношению $\langle E \rangle$ к ридбергу,

$$\gamma = \frac{\langle E \rangle}{Ry} = \frac{p^2(p-1)}{p+3} \frac{1}{K^2}. \quad (21)$$

Проведем анализ этой формулы. Если концентрация будет предельно низкой ($l \rightarrow \infty$), то величина p в соответствии с (17) становится неограниченно большой ($p \rightarrow \infty$). Тогда, согласно (20), $K \simeq p$ и $\gamma \rightarrow 1$, что вполне естественно. Другой предел соответствует переходу Мотта, когда энергия связи $\gamma = 0$. При этом из (21) имеем $p = 1$. Подставляя это значение p в выражение для K , даваемое (20), мы получим, что в точке перехода $K = 2$. Согласно формуле (18), значению $K = 2$ соответствует расстояние между донорами, даваемое формулой (1).

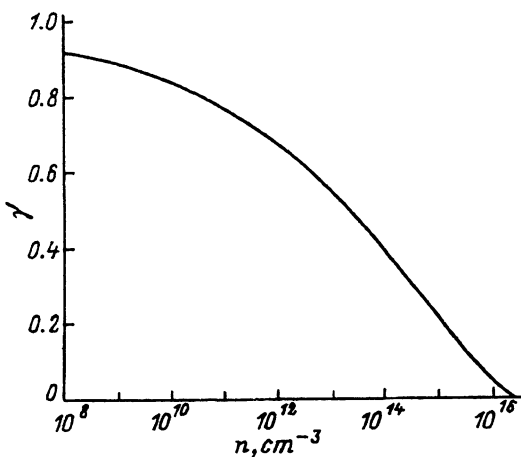


Рис. 3. Зависимость безразмерной энергии связи электрона на донорном центре с учетом экранирования от концентрации свободных электронов в арсениде галлия.

При этом в соответствии с (5) получим, что дебаевский радиус равен борновскому радиусу:

$$r_D = \left(\frac{\epsilon \mathcal{E}_F}{6\pi n_M e^2} \right)^{1/2} = a_B. \quad (22)$$

На рис. 3 изображена полученная таким образом зависимость безразмерной энергии γ от концентрации электронов в арсениде галлия. Видно, что даже в очень чистом материале с концентрацией электронов $n \approx 10^{12} \text{ см}^{-3}$ экранирование оказывается существенным. Если бы экранирования не было, то энергия основного состояния электрона на доноре была бы равна полупроводниковому ридбергу. Она составляет всего 60% от ридберга. При приближении к значению моттовской концентрации γ стремится к нулю и достигает его при $N_d = N_M = 1.5 \cdot 10^{16} \text{ см}^{-3}$.

Список литературы

- [1] N.F. Mott. Proc. Phys. Soc. A, **62**, 461 (1949).
- [2] N.F. Mott. Can. J. Phys., **34**, 1356 (1956).
- [3] Н. Мотт, Э Дэвис. *Электронные процессы в некристаллических веществах* (М., 1974).
- [4] P.W. Anderson. Phys. Rev., **109**, 1492 (1958).
- [5] В.Л. Бонч-Бруевич, С.Г. Калашников. *Физика полупроводников* (М., 1990).
- [6] Г. Бете. *Квантовая механика простейших систем* (Л., 1935).
- [7] *Handbooks on Semiconductors* (1992) v. 1.
- [8] P. Debye, E. Hückel. Phys. Zs., **24**, 185, 305 (1923).
- [9] Л.Д. Ландау, М.Е. Лифшиц. *Статистическая физика* (М., 1976).
- [10] С. Зи. *Физика полупроводниковых приборов* (М., 1984) т. 1.
- [11] Л.Н. Бычковский, О.В. Константинов, Б.В. Паренков. ФТП, **24**, 1848 (1990).
- [12] K.E. Berggren, B.E. Sernellins. Phys. Rev. B, **24**, 1971 (1981).
- [13] Ф. Зейтц. *Современная теория твердого тела* (М., 1949).
- [14] Н. Ашкрофт, Н. Мермин. *Физика твердого тела* (М., 1979).
- [15] Т.И. Воронина, А.И. Дахно, О.В. Емельяненко, Т.С. Лагунова, С.П. Старосельцева. ФТП, **22**, 1230 (1988).
- [16] З. Флюге. *Задачи по квантовой механике* (М., 1974) т. 1.

Dielectric–metal phase transition: a thermodynamic approach

D.N. Bychkovsky, O.V. Konstantinov, B.V. Tsarenkov

A.F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St.Petersburg,
Russia
