PHYSICAL REVIEW B VOLUME 14, NUMBER 2 15 JULY 1976

Weyl's theory and the electronic structure of finite crystals

Joaquin Sitte
Physics Department, Universidad de Los Andes, Mérida, Venezuela
(Received 20 October 1975)

The Schrödinger equation for a crystalline slab of finite thickness is separated under certain assumptions into a one-dimensional equation in the coordinate perpendicular to the slab surface and a two-dimensional band-structure problem. The independent solution of both problems yields zero-order wave functions for the electrons in the slab. Weyl's theory for second-order differential equations is used to solve the one-dimensional Schrödinger equation numerically for arbitrary potentials using square integrability as the only boundary condition. The behavior of the solutions for a particular model potential is studied in detail emphasizing the effect of the potential in the surface region on the surface states.

I. INTRODUCTION

One-dimensional models for the electronic structure of finite crystals with surfaces[1-7] have had considerable importance in the development of our understanding of the electronic structure of finite *real* crystals and the effect of the surfaces. These models have mainly relied on analytical solutions to a limited number of highly simplified situations.

Although major advances have been made in the computation of the electronic structure of surfaces in three dimensions,[8,9] these calculations are generally quite involved and do not give a simple qualitative picture of the essential physics. The main sources of complication are the nonseparability of the three-dimensional one-electron Schrödinger equation and the treatment of the boundary conditions.

The purpose of this paper is to explore a model[10] in which these two main difficulties are overcome. The part of the one-electron local potential which is responsible for the nonseparability is first neglected, yielding a zero-order separation of the Schrödinger equation.[5] The zero-order solutions which could then be used in a perturbation calculation require the exact solution to a one-dimensional Schrödinger equation. This equation gives a one-dimensional model of the finite crystal and it can be solved conveniently for quite general potentials by numerical methods and the use of Weyl's theory. The zero-order solutions are capable of giving considerable qualitative information about the electronic structure of surfaces, making them worthwhile to be studied in more detail. The present paper is mainly devoted to this point.

Since the question of the electronic structure of surfaces cannot be isolated from the question of the electronic structure of the finite crystal, attention is paid to the level structure of a finite-crystal model which exhibits a minimum of *surface effects*. Then the effects of the form of the potential in the surface region are discussed.

In Sec. II the main assumptions for the separation of the Schrödinger equation are stated and the separated equations together with the zero-order solutions are discussed. Section III presents the mathematical techniques for solving the one-dimensional equation. The solutions for some model potentials are discussed in Secs. IV and V.

II. SCHRÖDINGER EQUATION FOR A SYSTEM WITH INTERFACES

The energies and wave functions for the single-electron stationary states in a general local potential are given by the solutions of the time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\vec{r})+V(\vec{r})\psi(\vec{r})=E\psi(\vec{r}) \tag{1}$$

satisfying the boundary conditions required by the physical situation. A general model of a finite crystalline solid can be defined by a potential exhibiting a repeating pattern over a certain number of unit cells which fill a given *finite volume* in space. Far away from this volume the potential approaches a constant value arbitrarily defined as zero. The boundary condition for wave functions representing bound states is square integrability. No practical way of solving this problem in general has yet been found and a number of restrictive conditions have to be specified. The first restriction concerns the geometry of the solid. The most commonly used geometrical configurations are the semi-infinite solid exhibiting a single plane boundary and the film configuration exhibiting two plane boundaries. The plane boundaries or surfaces are in both cases of infinite extension. The flat surface configurations are still quite ideal situations; nevertheless, they are reasonable approximations to a great variety of experimental situations, mainly those concerning the study of well-defined surfaces along a definite crystal plane. A second restriction concerns translational symmetry. There can be no translational invariance along the direction perpendicular to the surface, but in both the semi-infinite solid and the thin-film configurations, translational symmetry *parallel* to the surface may exist.

In most cases the primitive two-dimensional surface cell is a multiple of the one corresponding to a given crystal plane and a repeating pattern can be found which is valid for both the bulk and the surface.

The model to be considered below is that of an infinite crystalline slab of finite thickness which may include in its interior an arbitrary number of plane, parallel interfaces separating different crystalline structures, but in such a way that overall two-dimensional translation symmetry parallel to the interfaces is conserved. The interfaces, including the surfaces, are not meant to be mathematical planes but transition regions of certain thickness.

Under these assumptions the potential in the Schrödinger equation (1) has the property

$$V(\vec{r}) = V(\vec{R} + z\vec{a}_z) = V(\vec{R} + \vec{m} + z\vec{a}_z)\,, \qquad (2)$$

where the general position vector $\vec{r}$ is written as the sum of a vector $\vec{R}$ parallel to the interfaces and a vector $z\vec{a}_z$ perpendicular to the interfaces. $\vec{m}$ is a two-dimensional translation vector

$$\vec{m} = m_1\vec{a}_1 + m_2\vec{a}_2 \qquad (3)$$

of the plane lattice defined by the coplanar vectors $\vec{a}_1, \vec{a}_2$.

Using Bloch's theorem for the variable $\vec{R}$ the general form of the solution $\psi(\vec{r})$ can at once be written,

$$\psi_{n,\vec{K}+\vec{G}}(\vec{r}) = e^{i(\vec{K}+\vec{G})\cdot\vec{R}}\, u_{n,\vec{K}+\vec{G}}(\vec{r})\,, \qquad (4)$$

where n is the quantum number arising from the quantization in the z direction, $\vec{K}$ is the two-dimensional wave vector and $\vec{G}$ a two-dimensional reciprocal-lattice vector. The function $u_{n,\vec{K}+\vec{G}}(\vec{r})$ has the same translational symmetry as the potential in Eq. (2). Substitution of (4) into (1) gives the differential equation for these functions,

$$\nabla^2 u + 2i(\vec{K}+\vec{G})\cdot\nabla_{\vec{R}}\, u + \frac{2m}{\hbar^2}\left(E - V(\vec{r}) - \frac{\hbar^2(\vec{K}+\vec{G})^2}{2m}\right)u = 0\,, \qquad (5)$$

where $\nabla_{\vec{R}}$ is the gradient applied to the variable $\vec{R}$. The periodic boundary conditions have reduced the problem to finding the functions $u_{n,\vec{K}+\vec{G}}(\vec{r})$ over a cell of tubular shape extending from $z = -\infty$ to $z = +\infty$, its cross section being the unit cell of the plane lattice.

Under the additional assumption that the potential $V(\vec{r})$ can be written as a sum

$$V(\vec{r}) = V_1(\vec{R}) + V_2(z)\,, \qquad (6)$$

Eq. (5) separates into two equations:

$$\nabla_{\vec{R}}^2 u'(\vec{R}) + 2i(\vec{K}+\vec{G})\nabla\cdot u'(\vec{R}) + \frac{2m}{\hbar^2}\left(E_{\vec{K}+\vec{G}} - V_1(\vec{R}) - \frac{\hbar^2}{2m}(\vec{K}+\vec{G})^2\right)u'(\vec{R}) = 0\,,$$

$$\frac{d^2 v_n(z)}{dz^2} + \frac{2m}{\hbar^2}\,[E_n' - V_2(z)]\, v_n(z) = 0\,, \qquad (7)$$

and the solution u can be written as a product

$$u_{n,\vec{K}+\vec{G}}(\vec{r}) = u'_{\vec{K}+\vec{G}}(\vec{R})v_n(z)\,. \qquad (8)$$

The solutions $\psi(\vec{r})$ then have the general form

$$\psi_{n,\vec{K}+\vec{G}}(\vec{r}) = e^{i(\vec{K}+\vec{G})\cdot\vec{R}}\, u'_{\vec{K}+\vec{G}}(\vec{R})v_n(z)\,, \qquad (9)$$

with energies

$$E_{n,\vec{K}+\vec{G}} = E_n' + E_{\vec{K}+\vec{G}}\,. \qquad (10)$$

As has been pointed out by Statz,[3] realistic potentials cannot be written in the form (6). This is easily seen by considering the case of the solid-vacuum interface. At a sufficiently long distance into the vacuum the potential should reach a constant value,

$$V(\vec{r}) = V_1(\vec{R}) + V_2(z) = \text{const.} \qquad (11)$$

for large z and all $\vec{R}$. The only possibility is that $V_1(\vec{R})$ is constant, i.e., there can be no variation of the potential along the surface for any value of z.

However, owing to the symmetry property (2), the nonseparable potential $V(\vec{r})$ can be expanded in a Fourier series in the variable $\vec{R}$,

$$V(\vec{r}) = V_0(z) + \sum_{\vec{G}\neq 0} V_{\vec{G}}(z)e^{i\vec{G}\cdot\vec{R}}\,. \qquad (12)$$

The first term in the expansion is independent of $\vec{R}$, while all other terms involve both $\vec{R}$ and z. Formally this suggests that all the expansion terms except the first be neglected.

That this may even be physically acceptable in some cases has been shown by Schulte and Bross,[5] discussed by Kenkre,[11] and applied by García, Solana, and Cabrera[10] to the calculation of surface bands on the (111) faces of Si and Ge.

The physical idea behind this approximation is that the periodic variations in the potential should be small compared to the potential jump at the surfaces. The form of $V_0(z)$ obtained by Caruthers, Kleinman, and Alldredge[8] for the (001) surface of aluminum displays this feature nicely. Under these conditions one could consider all the terms, except the first one, in (12) as a perturbation and use the set of equations (7) to obtain zero-order wave functions. The solutions to the first of the two equations are simply

$$u'^{(0)}(\vec{R}) = \text{const.}, \quad E^{(0)}_{\vec{K}+\vec{G}} = (\hbar^2/2m)(\vec{K}+\vec{G})^2 , \tag{13}$$

since V_1 is zero. To find the functions $v_n^{(0)}(z)$ the one-dimensional Schrödinger equation

$$\frac{d^2 v_n^{(0)}(z)}{dz^2} + \frac{2m}{\hbar^2}\left[E_n'^{(0)} - V_0(z)\right] v_n^{(0)}(z) = 0 \tag{14}$$

has to be solved. This would give zero-order wave functions

$$\psi^{(0)}_{n,\vec{K}+\vec{G}}(\vec{r}) = e^{i(\vec{K}+\vec{G})\cdot\vec{R}}\, v_n^{(0)}(z) \tag{15}$$

with energies

$$E^{(0)}_{n,\vec{K}+\vec{G}} = (\hbar^2/2m)(\vec{K}+\vec{G})^2 + E_n'^{(0)} . \tag{16}$$

Rayleigh-Schrödinger perturbation theory could then be used to calculate higher-order corrections to the energies and the wave functions.

The main difficulty in carrying out this program is the calculation of the solutions to Eq. (14) for an arbitrary potential $V_0(z)$, as for example the one obtained by a superposition of atomic pseudopotentials.[12] I will devote the rest of this paper to show how this can be done numerically for most cases and discuss the nature of the solutions for a simple case.

III. MATHEMATICAL TECHNIQUES

The one-dimensional Schrödinger equation (14) is more conveniently rewritten

$$-\frac{d^2 v(z)}{dz^2} + [V(z) - E]v(z) = 0 \tag{17}$$

using rydbergs as units of energy and Bohr radii as units of length.

This equation has to be solved in the interval $z \in (-\infty, \infty)$ with the boundary condition of square integrability for bound states or the finiteness of the solution for unbound states (e.g., tunneling states). A singular boundary-value problem is thus defined. In the theory of the infinitely periodic solid the singular boundary-value problem is reduced to an ordinary boundary-value problem by means of the periodic Born–von Kármán boundary conditions. No such thing is possible here since the potential $V(z)$ in Eq. (17) will be oscillatory in some finite region of the z axis and then level off to a constant value as one moves away from it.

The theory for singular boundary-value problems of self-adjoint second-order differential equations was developed by Weyl[13] in 1910 but has found only very limited applications in physics. A short review of this theory has been given recently in this journal.[14] The theory is also conveniently described in Coddington and Levinson's textbook[15] or the references given therein.

One may assume that in the interior of the slab there is a region where the potential along the z axis is strictly periodic over a certain finite number of periods; i.e., the presence of the surface influences the potential only in the outermost lattice planes. Then the singular boundary-value problem can be split into two singular boundary-value problems on semi-infinite intervals and one or more boundary problems on finite intervals with a *periodic potential*. This separation turns out to be quite useful because the theory of differential equations with periodic coefficients can be used.

Equation (17) is a second-order self-adjoint differential equation of the standard form,

$$-(pu')' + qu = \lambda u . \tag{18}$$

The prime indicates derivation with respect to the independent variable, p and q are real functions, and λ is a complex constant. For the particular case of (17) p is a constant equal to 1.

A solution φ of (18) is uniquely defined, up to a constant factor, by its phase α at a point z_0. The phase α is defined by

$$p\varphi'(z_0)/\varphi(z_0) = -\cot\alpha(z_0) , \tag{19}$$

which has the advantage, particularly for numerical work, of specifying a real initial condition in terms of a single parameter in the interval $(0, \pi)$.

The singular boundary-value problem. To find solutions to the one-dimensional Schrödinger equation arising in the previously stated model the solution of a singular boundary-value problem on the semi-infinite interval $(0, \infty)$ is of primary interest. The main feature of Weyl's theory is the use of complex eigenvalues in the self-adjoint second-order differential equation (18). Depending on the function $q(z)$ the problem can fall in one of the two categories: the *limit-circle* case and the *limit-point* case. In the limit-circle case all solutions of (18) are square integrable, whereas in the limit-point case there is one and only one square-integrable solution for each λ. All the situations considered in this paper belong to the limit-point case. In this case the square-integrable solution is given by

$$\chi(z;\lambda) = \varphi(z;\lambda) + m\psi(z;\lambda) , \tag{20}$$

where φ and ψ are two linearly independent solutions satisfying the initial conditions

$$\varphi(0;\lambda)=\sin\alpha\,,\quad \psi(0;\lambda)=\cos\alpha\,,$$
$$p\varphi'(0;\lambda)=-\cos\alpha\,,\quad p\psi'(0;\lambda)=\sin\alpha\,. \tag{21}$$

If one asks for the value of m such that χ satisfies the boundary condition

$$p\chi'(b;\lambda)/\chi(b;\lambda)=-\cot\beta \tag{22}$$

at some finite point $z=b$, then one finds that m lies on a circle in the complex plane which can be generated by varying β from 0 to π. Furthermore for any $b'>b$ the corresponding circle is entirely included in the one for b. Letting b go to infinity, m tends to a *limit point* independent of β,

$$\lim_{b\to\infty} m(b,\lambda,\beta)=m_\infty(\lambda)\,. \tag{23}$$

It can be shown[15] that

$$m_\infty(\lambda)=-\lim_{b\to\infty}\frac{\varphi(b;\lambda)}{\psi(b;\lambda)}, \tag{24}$$

which allows the calculation of m from the knowledge of two linearly independent solutions. From the function $m(\lambda)$, in turn, the initial phase at the origin for the square-integrable solution for each λ can be calculated. The numerical procedure is then as follows: For a given energy ($\lambda=E+i\epsilon$) two linearly independent solutions can be calculated by numerical integration starting with an arbitrary initial condition; from these solutions the initial phase of the square-integrable solution is calculated via the m function. The imaginary part in the energy is taken small ($\epsilon<10^{-5}$) according to the accuracy required in the calculations. The m function is further of central importance because its imaginary part relates directly to the spectral function $\rho(\omega)$ and hence to the spectral density

$$\rho(E_\beta)-\rho(E_\alpha)=\lim_{\epsilon\to 0}\frac{1}{\pi}\int_\alpha^\beta \mathrm{Im}[m_\infty(E+i\epsilon)]\,dE\,. \tag{25}$$

For a discrete spectrum the imaginary part of the m function exhibits peaks when calculated as a function of λ, which become δ functions in the limit $\epsilon\to 0$. When the energy range is scanned using the proper initial phase in the integration, eigenvalues can be detected by looking for peaks in the m function.

Final phases for a periodic potential. Whenever the infinite interval is divided into subintervals for the periodic and aperiodic parts of the potential, the dependence of the phase on the energy is of interest, since the construction of the overall solution by joining the solutions in neighboring intervals requires the matching of the phases at the common boundary. The phases at the end point of a finite interval (final phases) in which the potential is periodic can easily be calculated by numerical integration starting with some definite initial phase. Making use of the properties of the solution to a second-order differential equation with periodic coefficients it is sufficient to integrate numerically over a single period. It turns out that final phases exhibit a quite typical behavior in their energy dependence. This was first noted by Schokley,[2] who showed that in the case of a symmetric periodic potential, of period a, the logarithmic derivatives γ and μ of the even and odd solutions satisfy the following relation at $a/2$:

$$\gamma(a/2)/\mu(a/2)=-\tan^2(ka/2)\,, \tag{26}$$

valid for both real and imaginary k. This result can be generalized for points $na/2$ by considering a new potential made up of n periods of the original one:

$$\gamma(na/2)/\mu(na/2)=-\tan^2(nka/2)\,. \tag{27}$$

This equation predicts an oscillatory dependence on k of the quotient of the phases in the bands (real k), the number of oscillations increasing with n, whereas in the band gaps the quotient of the phases is not oscillatory, converging exponentially to unity, with increasing n. The numerical results, as those shown in Fig. 2, indicate that this typical behavior is also true for other points than $na/2$ although a general proof is still lacking.

Numerical Methods. To calculate the m function and the initial phases for the singular boundary-value problem I followed the numerical technique introduced by McIntosh[16] for the Weyl theory. The integration of the complex equation (19) is done in matrix form with the sixth-order Runge-Kutta formulas given by Sarafyan[17] which proved to be quite efficient. When using Weyl's theory numerically a fundamental point has to be remembered, namely, that with arbitrary initial conditions the two linearly independent solutions will in general both be divergent (in the limit-point case). Therefore the square-integrable solution which usually goes to zero for large values of z will be given as the difference of two large numbers and in any calculation with finite precision the calculated square-integrable solution will sooner or later exhibit an exploding tail. Fortunately this diverging tail is generally out of the region of interest and appears after the solution (and its derivative) has already become very small. Therefore the tail can be neglected and the function set equal to zero thereafter without seriously affecting the accuracy of the results.

The magnitude of the imaginary part of the energy has a minor effect on the relevant quantities (real part of the wave function, initial phases and energies) and convenient values are between 10^{-2} and 10^{-4} unless extreme precision is desired.

IV. ELECTRONIC STRUCTURE OF FINITE CRYSTALS

The model. Most of the theoretical work on surface states has been done for the limiting case of a semi-infinite solid.[18,19] Surface states appear in the forbidden energy gaps giving rise to additional peaks in the density of states but leaving it unaffected in the bands. In this situation the surface states can be considered separately from the electronic structure of the *bulk*. This is not so for the *finite* crystal, where bulk and surface states have to be considered on the same level, as we shall see below, because *surface states* can be transformed continuously into *bulk states* and vice versa.

In this section we will concentrate our interest on the question of how the electronic structure depends on the crystal size. A qualitative understanding of the behavior of the energy levels with increasing size of the crystal proves to be very helpful for the understanding of the effects of the surface. For the purpose of the study we will select a definite model potential. Many results will not depend on the precise form of the potential, but whenever a feature depends on the specific form of the potential this will be stated explicitly. The model potential will be purely sinusoidal up to a point z_0, where it is matched continuously to a $-1/z$-like decay. The point z_0 is selected so as to produce a minimal distortion of the potential in the surface cells. The potential is shown in Fig. 1 and defined as follows:

$$\begin{aligned} &V(z) = C_3 - C_1\cos(2\pi z/C_2)\,, \quad |z| \leq |z_0| \\ &V(z) = -C_4/(z - z_1)\,, \quad |z| > |z_0| \qquad (28) \\ &z_0 = (n + \tfrac{1}{4})C_2\,, \quad z_1 = z_0 - C_4/C_3\,. \end{aligned}$$

The potential is symmetric, the number of potential wells is $2n+1$, and C_1–C_4 are constants which allow a wide range of variation of the potential while retaining the same functional form. The Schrödinger equation (17) with the potential (28) in the interval $(-z_0, z_0)$ reduces to the standard Mathieu equation[20] by means of the change of variable $v = \pi z/C_2$. The stability intervals, i.e., the band structure, for this equation are known and served as test cases for the numerical calculations. The connection between the parameters in the potential and the quantities a and q of Ref. 20 are given by

$$a = (E + C_3)C_2^2/\pi^2\,, \quad q = -C_1C_2^2/2\pi^2\,. \qquad (29)$$

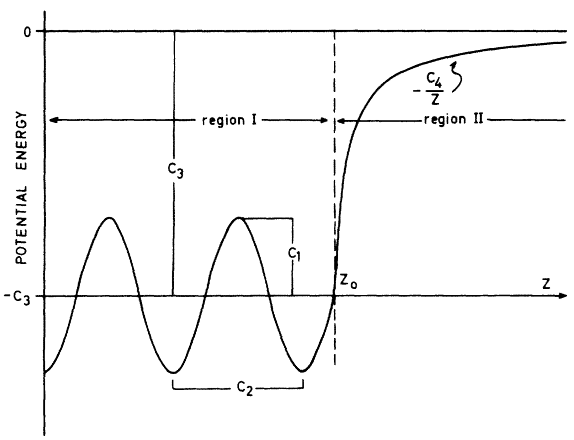

FIG. 1. Illustration of the potential defined in Eq. (28) for $n = 2$.

A sinusoidal potential was selected for the bulk for several reasons: It is known that a sinusoidal model potential reproduces qualitatively many of the characteristic features of the electronic structure of real solids. Good band structures for metals can be calculated by potentials represented by a Fourier series with few components. It further leads to a band structure which depends on the single parameter q [see Eq. (29)] and contains the tight-binding and nearly-free-electron approximations as limiting cases for large and small values of q, respectively. Finally, it has been used in previous[3,21,22] works, which allows a comparison of the results. The image potential is certainly appropriate at larger distances. Truncating it and fitting it to the sinusoidal potential is probably the simplest thing to do which still seems to resemble roughly the more complicated behavior at small distances from the surface.[23]

Exchange and correlation contribute significantly to the effective one-electron potential so that a good potential should at least be self-consistent for this case. The potential (28) probably does not satisfy this requirement, but this is not a severe limitation here because the primary interest of this study is the *dependence* of the electron energy levels on the *characteristics* of the potential.

From the symmetry of the potential it follows that the solutions of the Schrödinger equation have to be either even or odd functions of z and the left-hand side of the boundary-value problem can be substituted by the initial conditions:

$$\begin{aligned} &g(0) = 1\,, \quad g'(0) = 0 \text{ for the even solution}\,, \\ &u(0) = 0\,, \quad u'(0) = 1 \text{ for the odd solution}\,. \end{aligned} \qquad (30)$$

The problem can further be divided into two regions: Region I comprises the interval $(0, z_0)$ and region II comprises the interval (z_0, ∞). The boundary condition at z_0 is the matching of the phases of the solutions in region I with those of the solutions in region II. The methods outlined in Sec. III can be applied directly to these regions to calculate the final phases (phases at z_0) of the even and odd solutions in region I and the initial phases (phase at z_0) of the square-integrable solution for region II, all as a function of the energy. The crossing points of the curves of the initial phases vs energy with that of the final phases vs energy determine the energies where a matching is possible, and thus the energy levels. Once the energies are known the wave functions can be calculated.

The final phases at z_0 for the even and odd solutions as a function of energy are shown in Fig. 2 for $n = 1$ and $n = 4$. The parameters of the potential have been selected in such a way so as to produce an allowed and a forbidden band of approximately equal width in the range of negative energies. Many other different cases were actually calculated, but those shown are the most illustrative. The final phases display the qualitative behavior already discussed in general terms in Sec. III, where no specific assumption on the potential was made except its periodicity. The phases vary slowly outside the bands and the even and odd phases tend to the same limiting value. In the band the phases vary rapidly and tend quickly to a set of very steep almost parallel lines. Phases for higher values of n are not shown in the band because of the obvious plotting difficulties, but the behavior is clear. In region II we deal with a boundary-value problem singular at infinity. In this region the Schrödinger equation takes again a form which, in principle, can be handled analytically since it can be reduced to a confluent hypergeometric equation of a special form, the Whittaker equation. The properties of its solutions have been studied by Wannier and have been applied to the study of surface states by Cole[24] and García and Solana.[7] I will not solve the problem analytically here because numerical methods will

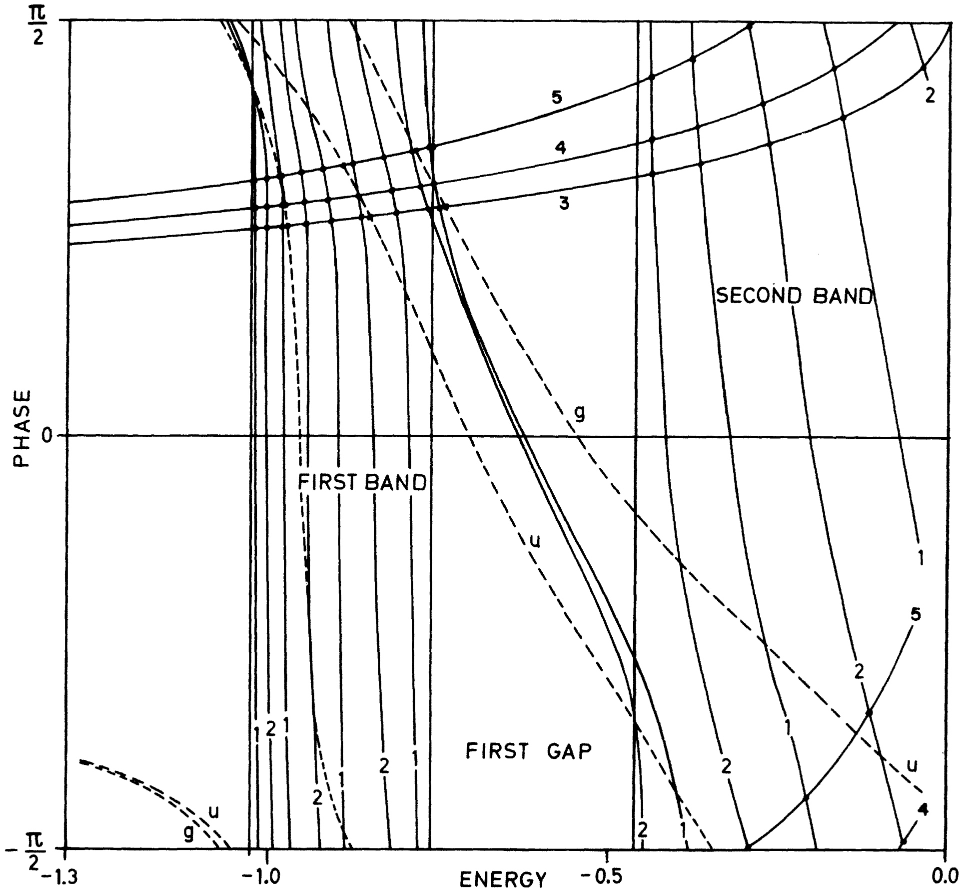

FIG. 2. Energy dependence of the final phases at z_0 of the solutions in region I and inital phases in region II. The final phases are shown for $n = 4$ (solid lines) and $n = 1$ (interrupted lines). The even solutions are labeled 1, g and the odd solutions 2, u, respectively. The potential parameters for this case are $C_1 = 0.3$, $C_2 = 5.0$, and $C_3 = 1.0$. The curves numbered 3, 4, and 5 correspond to the initial phases and to values of C_4 of 0, 0.1, and 0.5, respectively. The intersections of the initial- and final-phase curves determine the allowed energies (in rydbergs).

be needed anyway for the other cases to be discussed. The numerical procedure is sufficiently fast and accurate for the range of values of interest for C_4. The largest value of C_4 of interest is 2.0, which corresponds to the ordinary Coulomb potential. Figure 2 shows the initial phases for various values of C_4 and fixed C_3. The initial phase is a monotonically increasing function of the energy because the wave functions become more oscillatory when $V(z)-E$ is positive, or the exponential decay less strong, when $V(z)-E$ is negative. In the case of an abrupt potential step at z_0 ($C_4=0$), the square-integrable solutions are decaying exponentials and their initial phases are all positive. If the potential rises more slowly to zero then there will be an interval to the right of z_0 where $V(z)-E$ is positive, and this interval will become large for higher energies. Then it can happen that the wave function at z_0 has a diverging (negative) phase, i.e., the wave function grows away from the axis first, goes through a maximum, and then decays to zero. If the potential rise is sufficiently slow the phase may again become positive. All this means is that as the potential rises more slowly the initial phase becomes a more steeply increasing function of the energy. Note that in this discussion we have implicitly used the theorem which says that the solutions to the Schrödinger equation depend continuously on E and on the potential as long as V is a continuous function of the potential parameters. This theorem will be used implicitly throughout this study.

The energy levels. The *crossing points* of the final phases in region I with the initial phases in region II (Fig. 2) determine the energy levels. The most important feature of Fig. 2 is that the levels are grouped in energy intervals completely contained in the bands of the infinite crystal. There may be at most one crossing point for the even and one for the odd solution in the band gap since there the final phases vary by less than π over a band gap.

At this point it is convenient to make a comment on the meaning of *bands* and *band gaps*. For a finite crystal a band in the usual sense does not exist, but it could be defined by the lowest and highest energies of a group of closely spaced levels. Such a definition will, as shown below, not allow for making a distinction between bulk and surface states. Here instead the term band will be used in the sense of *stability intervals* as defined for the solutions of differential equations with periodic coefficients.[15] For a given potential these are perfectly defined energy intervals.

Returning to Fig. 2 we see that inside a band the levels are rather uniformly spaced, being of the order of

$$E_{j+1}-E_j=\Delta E/m\,, \tag{31}$$

where ΔE is the bandwidth and m the number of levels in the band. The levels are more compressed towards the border of the band. This is probably not a general feature but a potential-dependent one. Even so, the spacings are all of the same order of magnitude. If there is an even and an odd level in the band gap, they approach each other exponentially with increasing n and become degenerate for the limit of the infinite crystal. Because of the exponential behavior, for practical purposes, the states can be considered degenerate already for relatively small values of n, say between 10 and 20. The convergence to this limit is slower for levels with higher energy than with lower energies, but is always exponential. The effect of varying C_4 can be directly estimated. For a fixed energy the initial phase increases with increasing values for C_4, but since the final phases are monotonically decreasing functions of the energy the crossing points and thus the energy levels are shifted to lower energies. The final phases have steep negative slope in the bands and therefore the energy levels there are very insensitive to changes in the surface potential, whereas in the band gaps the slope is considerably less steep and the levels in the band gap will be much more sensitive to changes in the surface potential. The lowest level in the lowest band will always be even. Even and odd levels alternate in the band. If the number of potential wells $n'=2n+1$ is even then the highest level in the band will be odd; if n' is odd it will be even (there is only one level in a band for each well). The lowest level in the next higher band will be odd (even) for n' odd (even) and so on. This will still be true when there are states in the band gap since they occur in closely spaced pairs (for the symmetric crystal). When the crystal is small and or the levels in the band are very close to a band edge then it may happen that only one level is actually located in the gap. Finally we can conclude that the band structure itself, i.e., the location and width of the bands, does not depend on how the potential is terminated at the *surface*. It is entirely determined by the periodic or bulk part of the potential. The potential at the surface will, however, influence the exact position of the levels in the energy range of the band. The *existence and position* of the states with energy in the band gaps is dominated by the behavior of the potential at the surface.

To further investigate the behavior of the energy levels, in particular their dependence on the surface potential, we will study the effect of a strong systematic perturbation of the potential in the surface cells. A very illustrative way to do this is simply to vary the matching point z_0 over a whole period of the periodic potential, but leaving the potential to the right of z_0 unaltered. This has been done by Levine[22] for a semi-infinite Mathieu potential terminated by an abrupt potential step. He made an approximate analytical study of the levels in the band gaps. By means of a qualitative analysis of the finite periodic potential it is possible to interpret Levine's results in a more general way. For that purpose let us do the thought experiment of letting the crystal increase in size *continuously* (and symmetrically) from n' to $n'+2$ potential wells by moving the matching point from $(n+\frac{1}{4})C_2$ to $(n+1+\frac{1}{4})C_2$ continuously. Calculation of the curves for the final phases of the solutions show that the curves for u and g in the band gap move in a more or less parallel way over the whole range of the phase as they are evaluated at increasing values of z over one cycle. They move from positive values of the phase to negative values. The speed with which they shift downward is, however, not constant and the exact way of doing it depends on the potential. The general feature is, however, that after one period the phases have essentially the same position as before, with the difference that they have come closer to their limiting value. The crossing points with the initial phases show that the two levels close to the bottom of the second band move out into the gap, across it, and finally become included in the lower band. The levels in the lower band compress towards the bottom of the band to accommodate the two new levels. In the higher bands the levels will also shift to lower energies. There will be a certain position, in general a certain range, for z_0 where there are no states in the band gap, which means that there is a certain threshold value for the perturbation of the surface cell for producing levels in the band gap. After the threshold is achieved an increase in the perturbation makes the levels *fall* rather quickly through the band gap into the lower band. This is the situation for the lowest band gap, illustrated in Fig. 3.

For the next band gap the situation will be somewhat different because the net gain in the number of levels of each band has to be two, since two potential wells are added to the total potential. The second band, for example, loses two levels to the lowest band and thus has to receive four levels from the third band. Since there can, in general, only be two levels in a band gap, as can be seen from the behavior of the final phases, the four levels will shift through the second band gap in pairs, one pair after another. The behavior in the higher bands will be similar. From this we conclude that the nonexistence of levels in the lowest gap does not mean that there will be no levels in the second gap and so on. For the higher-order band gaps the perturbation of the surface cell required to press out some levels into the band gap becomes smaller.

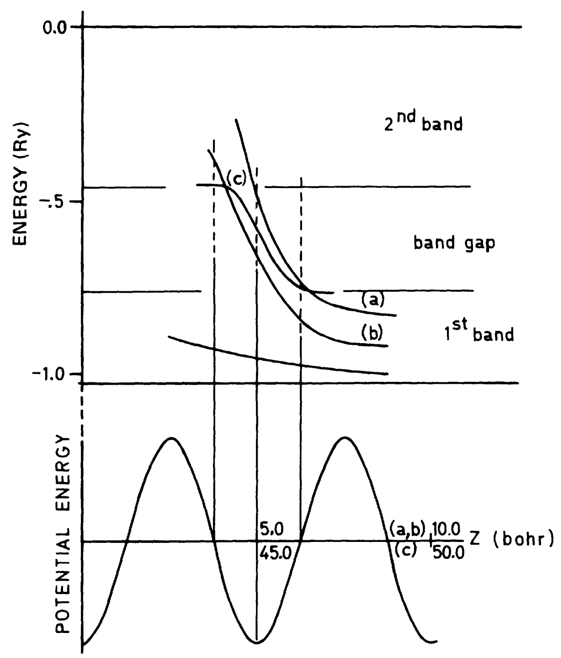

FIG. 3. Behavior of the energy levels in the first gap when the matching point z_0 is shifted over a whole period. Curves (a) and (b) correspond to the odd and even states, respectively, when z_0 is shifted across the second potential to the right of the origin. Curve (c) shows the result of moving z_0 across the tenth well. The even and odd states are practically degenerate in the last case. The potential parameters are $C_1=0.3$, $C_2=5.0$, $C_3=1.0$, and $C_4=0$.

When the bands are very broad and the surface potential has a long range out into the vacuum, such as, for example, the Coulomb potential $(C_4=2)$, the initial phase can become a rapidly growing function of the energy and it may happen that the initial phase varies over more than π in a band gap, thus crossing the final phases of the periodic part more than once. In this case the rule of having at most two states in a gap breaks down.

The wave functions. In region I and except at a band edge, the even and odd solutions can be written as linear combinations of *Floquet* solutions:

$$\varphi(z)=af_+(z)e^{ikz}+bf_-(z)e^{-ikz}, \tag{32}$$

where a, b are constants and f_+, f_- two functions periodic over a lattice spacing. The solutions in this region will be identical with those of the infinite symmetric crystal for the same energy. For the finite lattice those values of k which make

φ satisfy the boundary conditions at the origin and at z_0 are selected from the continuum of possible k values.

If there is a level in a band gap, k can be taken as purely imaginary, since the real part is constant over the whole gap, and in region I the solution φ will diverge exponentially, both in positive and negative z directions (from symmetry requirements). Near the edges of a band gap the imaginary part of k will be small since it is zero at the band edges and grows to attain its maximum near the center of the band gap. The exponential growth will thus be strongest near the middle of the band gap. The solutions in region I join smoothly to square-integrable solutions in region II which are of decaying character. The overall solution will consequently have a maximum in the amplitude around the surface (z_0) representing a particle localized in the surface region. States in the band gaps are then always surface states which decay exponentially into the crystal. The localization is strongest when the energy is near the middle of the band gap. The localization increases also strongly with the size of the crystal. In Fig. 4 the wave functions for a surface state slightly above the band top and one near the middle of a gap are illustrated. The crystal is small; one of the states is only slightly localized at the surface.

Let us now turn to the states inside a band where k is real. The exponentials in (32) are now complex and the form of the solution in region I is that of a superposition of two oscillatory functions, one with period a (the lattice constant) and the other with period $2\pi/k$. Near the bottom of the lowest band k is real and very small. For all those energies for which $z_0 \leq \pi/2k$ the wave function will have a decreasing amplitude towards the surface. The amplitude at the surface will first decrease as the energy increases away from the band gap, attaining a minimum and increasing again, going through several oscillations as we approach the center of the band. The wave functions for the lowest level in the band will, as a consequence of this, be more localized towards the center of the crystal. A similar behavior occurs for the energies close to the top of the band and at the bottom of the next band, but here because the periods of the two periodic components in the wave function are close multiples ($2\pi/k \approx 2a$) and the wave function shows "beats" of long period. In conclusion, for a finite crystal the states in the bands still extend over the whole crystal but the probability of finding a particle at a certain point will vary *both* with position and energy. Particles in states near the band edges are pushed away from the surface; in the centers of the bands the particles extend homogeneously through the crystal, while in the band gaps they become localized at the surface. In Fig. 5 all the wave functions of the first band for a finite periodic lattice are shown. The potential parameters are such that in this case no surface states are present in the first gap. It can be clearly seen how the amplitude of the wave functions near the band edges decreases near z_0.

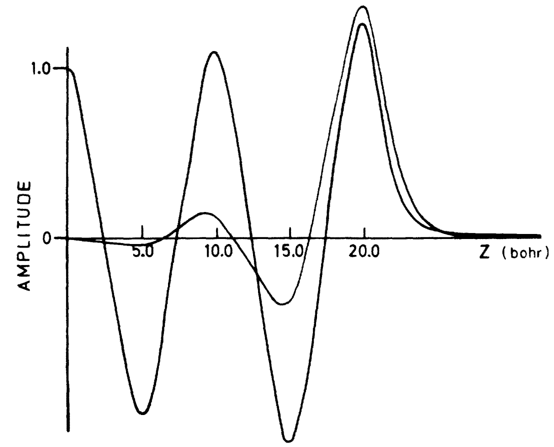

FIG. 4. Wave functions for states in the band gap. Only the part of the wave functions to the right of the origin are shown since the functions are either even or odd. The model crystal has a total of nine potential wells (atoms). The functions are normalized with respect to their initial conditions (unit matrix). The function which shows strong localization lies near the middle of the band gap and its amplitude has been scaled down by a factor 100. The other function corresponds to a state just slightly above a band edge and is thus only slightly localized.

The local density of states. The density of states $\rho(E)$ does not account for the local variations of the relative probabilities of finding a particle in a certain state. Many surface spectroscopies, such as, for example, photoemission, ion neutralization spectroscopy and others, probe the occupation of states over a small region of space. The relevant quantity in these cases is the local density of states defined as

$$\rho(E,r)=\sum_i |\psi_i(r)|^2\delta(E-E_i)\,, \tag{33}$$

which is appropriate to describe differences in *weight* of the states in different regions of an extended system.

For a finite crystal the local density of states in the bulk and at the surface shows some characteristic differences. From the discussion of the behavior of the wave functions we conclude that at the surface the local density of states is reduced compared with the regions in the center of the crystal. This effect has been observed in photoemission[25] and ion neutralization spectroscopy[26] for transi-

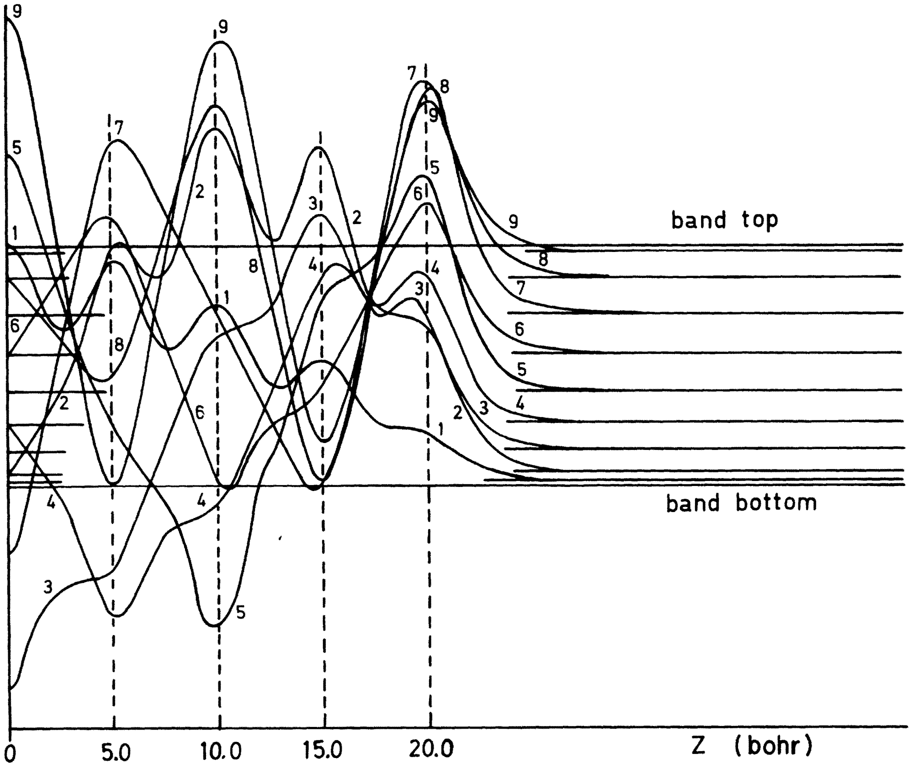

FIG. 5. Wave functions for the first band and $n=4$. The potential parameters are the same as those in Fig. 6. $C_4 = 0.5$. Normalization is again only with respect to the initial conditions. The odd solutions have been scaled down by a factor 7.5. The functions are numbered with increasing energy and their reference lines show their position in the band. The potential parameters used do not produce localized states in the first band gap.

tion metals. Calculations of this effect have been made in the tight-binding approximation[27,28] and it has been interpreted as due to the reduced coordination number of the atoms at the surface. It may be interesting to note that the effect appears in this one-dimensional model where no mention is made about neighbor interactions since all neighbor interactions are automatically included. The reduction of the amplitude of the wave functions diminishes as the size of the crystal increases. As a function of size the local density of states should also display an oscillation superimposed on it on the shoulders near the band edges because of the already discussed modulation of the wave functions in these energy regions. Such superimposed oscillations should also appear in the local density of states for a crystal of fixed size as we move away from the surface into the bulk. Such a structure in the local density of states is shown in the calculation by Haydock and Kelly.[28] The case of a crystal with surfaces is always different from the one where periodic boundary conditions are applied, even when the surfaces are translated to infinity by letting the crystal become infinite, since surface states if present do not disappear from the spectrum with increasing size. The total densities of states will differ in both cases, but because of the localization of the surface states, for large crystals the local densities of states will be indistinguishable far inside the bulk.

V. PERTURBED SURFACE

In Sec. IV it was found that a finite crystal may have *surface states* with energies in the forbidden gap whenever the potential in the cell at the surface was *sufficiently* perturbed. The perturbation used there may appear to be highly unrealistic, but from the general arguments we expect a very similar behavior for more realistic perturbations. We expect that the outermost two states in a band can be continuously transformed into surface states by a continuous variation of the potential in the surface cells or in the vacuum part.

The kind of surface perturbations we may consider as realistic are essentially two: the modifi-

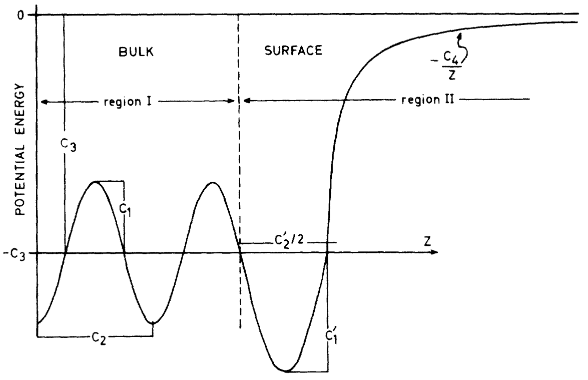

FIG. 6. Definition of regions I and II and the potential parameters C_1', C_2' for the case of surface perturbation.

cation of the strength of the potential in the last cell, which can be thought of as a very crude model of adsorption, and the modification of the lattice spacing in the last cell, which would simulate lattice relaxation. The model is essentially the same as before, but with the difference that the last half-period of the cosine potential is included in region II. The amplitude and the period of the cosine there will then be varied independently of the values in region I. The definition of regions I and II and the parameters C_1' and C_2' are shown in Fig. 6.

The effects of C_1 and C_2 are not independent. They, together, determine the band structure of the periodic potential in region I by means of the parameter q defined in (29). For this reason C_2 is given a fixed value and only C_1 is varied. The effect of C_3 is only to shift the bands as a whole up or down in energy, but it does not influence the bandwidths. (C_3 has, however, a small effect on the potential in the vacuum side.) Restricting the variation to the parameters C_1, C_1', C_2', and C_4 leaves ample space for exploring the main features of the model.

In Fig. 7 the initial phases at z_0 for region II are shown for various values of the parameters C_1' and

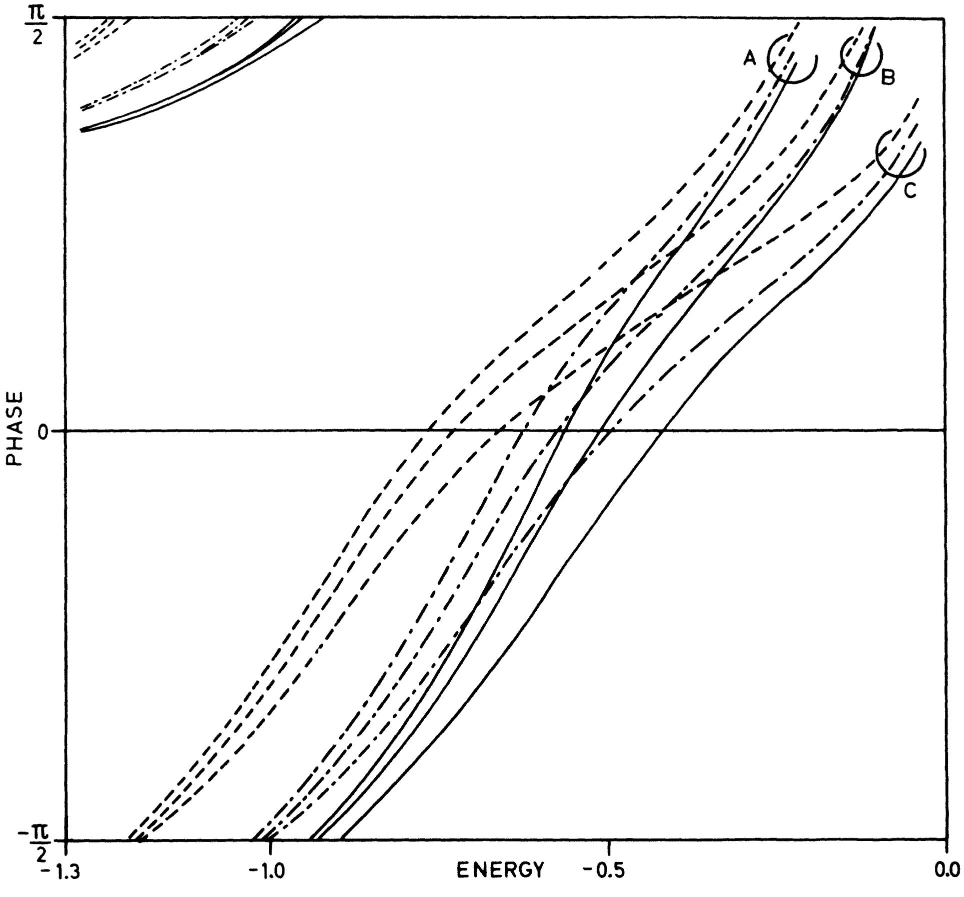

FIG. 7. Initial phases for region II for different values of the parameters C_1', C_4 ($C_2' = C_2 = 6.0, C_3 = 1.0$). The curves in each of the circles have the same value of C_4: A, 1.0; B, 0.5; C, 0.1. The values of C_1' are 0.15 (——), 0.3 (---), and 0.6 (-·-·-).

C_4. C_2' is fixed. In the high-energy region the initial phases are mainly determined by the way in which the potential approaches zero in the vacuum, whereas in the low-energy part they are dominated by the depth of the potential well at the surface. This can be understood by noting that the wave functions for low energy levels will not extend very far into the potential barrier at the surface. When C_4 is varied one has to remember that the $-1/z$ decay is adjusted such as to have always the same value at $z_0 + C_2'/2$, namely, $-C_3$ (for any value of C_4); thus the potential barriers differ little at low energies in the region where the wave function is appreciably different from zero. At high energies the potential increases very slowly and the wave functions will extend far out into the vacuum and the region over which C_1' produces a change in the potential will be relatively small, having a small effect on the wave function.

In Fig. 8 the final phases at z_0 for region I have been superimposed on the initial phases of Fig. 7. The final phases are evaluated at $z_0 = \frac{3}{4}C_2$. This is sufficient since we know that in the band gap the curves for the even and odd functions define the region where the limiting curve for large n lies. In the bands the phases become a set of regularly spaced almost vertical lines for large n. C_2 equals C_2' in all the cases shown. From Fig. 8 we can read off the level structure for 27 different situations by combining the three values of C_1 with the three values of C_1' and the three values of C_4.

We first observe that no surface states originate from the states at the bottom of the second band. We can, however, extrapolate the results and guess under what condition this would be possible. Since the initial phases shift upwards with increasing C_4 and C_1', it may be possible, by widening and deepening the potential well at the surface, to lower the levels in energy such that they fall outside the band. An increase in C_4 will be most effective

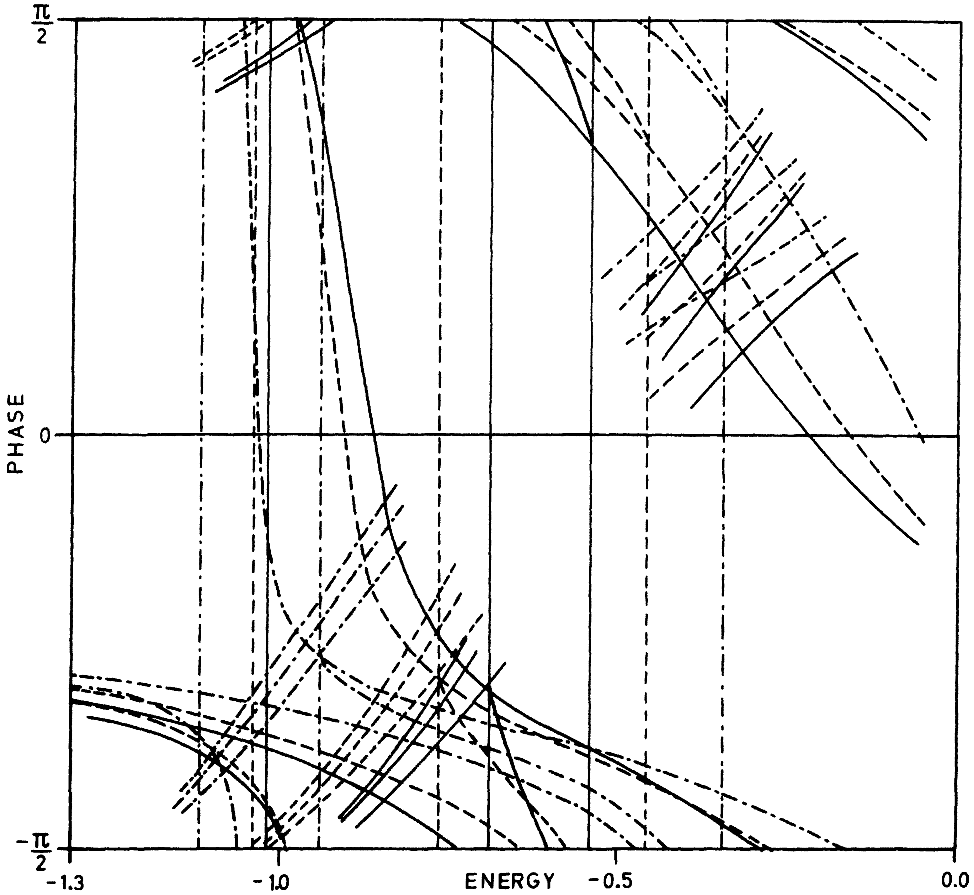

FIG. 8. Initial phases of Fig. 7 superimposed on the graph of the final phases of region I. Final phases for three different values of C_1 and $n = 1$ are shown. They are the same as the values for C_1 in Fig. 7. The band edges are indicated for each case as well as the limiting curves for large n in the band gap. From left to right the succession of bands and gaps is band-gap-band-gap.

since the levels are in the high-energy region. The effect can further be enhanced by increasing the amplitude of the bulk potential (C_1) which widens the band gap and raises the bottom of the second band.

No surface states at all occur when $C_1 = C_1'$, which means that the range of values for C_4 produces a perturbation which is not sufficiently strong, but those selected include the range of physically reasonable values for C_4, namely, small values for insulators and a maximum of 0.5 for metals. The conclusion would be that a perfect crystal surface does not necessarily have surface states. From the figure we further see that in this case there is a level very close to the band top for very small values of C_4, which may move slightly out into the gap for C_4 approaching zero. This is then an abrupt step and in such a case the sinusoidal potential is also no longer very reasonable and no significance should be given to this limit of the model.

For the case of a narrow bulk band ($C_1 = 0.6$) surface states exist in the gap. They have been lifted up in energy because the *surface potential well* is less deep than the wells in the bulk. As the amplitude of the potential variation in the bulk decreases, the bands become wider and the top of the first band rises but the levels in the band gap rise only very slowly and as the band becomes broader it finally *swallows* the surface states. Something similar happens when the surface well is deep enough to bind a level at an energy lower than the bottom of the lowest band. When the bulk potential is modified in such a way as to raise the energy of the bottom of the band, the surface states remain essentially at the same energy as the band bottom moves up and away. The levels can again be *swallowed* if the bulk potential is changed so that the band bottom moves down in energy. This general behavior suggests that the energy of a surface state is primarily determined by the potential well at the surface (and of course the potential tail in the vacuum), the complex band structure of the bulk being of less importance. Roughly speaking one might say that a surface state will exist when the surface potential well binds an electron at an energy which lies in a band gap for the bulk. Unfortunately it is not completely defined what exactly the surface potential well is. It may be that criteria can be found which allow the definition of an *equivalent surface well*, which could then be studied separately from the bulk and which may allow a first guess about the existence or nonexistence of surface states.

Finally we take a quick look at the effect of surface relaxation. Very little is known about the changes in the lattice spacing at a surface but calculations[29,30] and experimental results[31] indicate that it may be of the order of 10% or less. I have tested the effect of varying C_2' up to 20%. No new features appeared and the results could be interpreted in the same way as for the cases above. One case is, however, worth mentioning, namely, when the bulk bands are narrow and the band gap is wide, and further, the surface well is less deep than the bulk wells. In this case a pair of surface states appeared at the bottom of the band gap and another pair appeared near the top of the gap. The surface well was such as to bind two states in the range of energy of the band gap. The origin of the states can also be understood as the lower pair being pushed up in energy because the surface well was less deep than in the bulk. At the same time, because of the increase in lattice spacing, the levels decreased in energy, but not enough to fall back into the band. The widening of the well also pulled down two levels from the next higher band into the gap. This illustrates that there may be more than two states in the same band gap, but it is certainly a more exceptional case.

Wave functions. In Fig. 9 the wave functions in the lowest band and for the surface states are

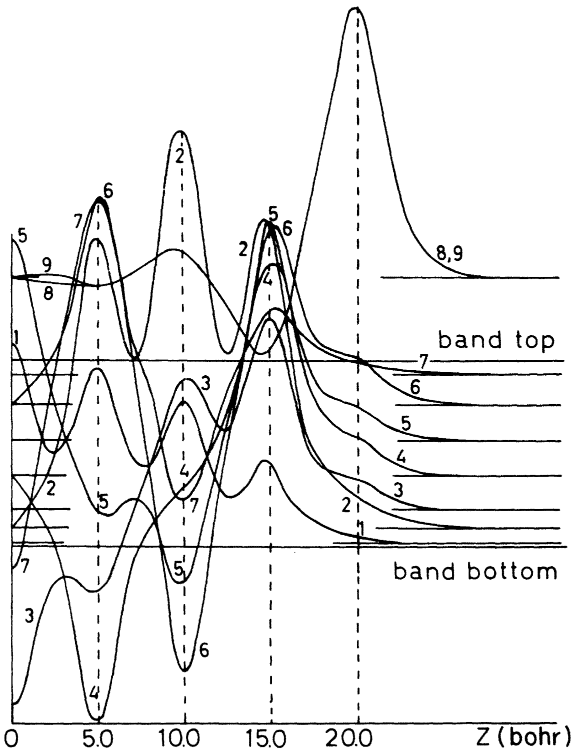

FIG. 9. Wave functions for a model crystal with nine potential wells and with a perturbed surface well. Two levels have been pushed out into the first gap producing localized states. The parameters are $C_1 = 0.6$, $C_1' = 0.15$, $C_2 = C_2'$, $C_3 = 1.0$, and $C_4 = 0.5$.

shown for the case of $C_1 = 0.6$ and $C_1' = 0.15$. In this case the surface states are nearly in the middle of the first gap. The surface states, which are practically degenerate, are strongly localized at the surface potential well. There is still a non-negligible amplitude at the next two *subsurface* atoms. More remarkable is the fact that the amplitudes of the bulk states at the surface are strongly reduced. If the levels are filled with two electrons each, the result will be a crystal displaying a strong positive surface charge. The situation will be similar for the case when the surface cell distortion is such as to produce a pair of levels below the first band. Now the surface states will be occupied producing a negative surface charge. The other cases discussed will be somewhere between these two limiting situations. It may be worth pointing out that these cases can be considered as one-dimensional analogs to charge transfer due to adsorption.

VI. CONCLUSION

Our theoretical understanding of the electronic structure of perfect crystalline solids is based primarily on the solutions of the single-particle time-independent Schrödinger equation with periodic boundary conditions. Application of these boundary conditions implies, however, the exclusion of all surface effects. The proper boundary condition for a finite crystal is the square integrability of the wave function over the whole space for bound states and the finiteness of the wave function for *free* states. These boundary conditions account naturally for surface and finite-size effects but they give rise to considerable mathematical complications. In atomic calculations the separation of the Schrödinger equation is of fundamental importance for the mathematical treatment of the problem. The zero-order solutions in the form of products of radial and angular functions have been the very basis of the qualitative understanding of the electronic structure of atoms. In this paper a similar approach to the study of finite-size and surface effects has been explored. Under certain restrictive conditions the Schrödinger equation for an infinite crystalline slab can be separated into a one-dimensional equation in the coordinate perpendicular to the slab and a two-dimensional equation in the parallel coordinates. For the equation in the parallel coordinates translation symmetry produces a two-dimensional band-structure problem. The one-dimensional equation can be solved numerically for quite general potentials by means of Weyl's theory, retaining the proper boundary conditions. It is the solutions of this equation which introduce the zero-order finite-size and surface effects. In fact the study of the solutions of a model potential for a finite one-dimensional crystal reveals many features which have been observed in real systems and predicted by more elaborate three-dimensional calculations, such as the narrowing of the bands at the surface and the rapid approach with increasing size of the level structure to that of the band structure of the infinite crystal. The appearance of surface states can be directly related to the distortion of the potential in the outer cells. By such distortion of the potential in the last cell, surface states can be made to appear anywhere in the forbidden gap, and more than a pair of levels may possibly be found in a gap. Through changes in the outermost potential well, adsorption and the possibility of charge transfer can be simulated. Besides the use of the zero-order solutions for three-dimensional perturbation calculations, the techniques used in this study can also be used to study the level structure in the presence of an electric field for applications in field emission and also in metal-semiconductor contacts.

The most serious limitation of the calculations in this paper is the lack of self-consistency of the solutions and the neglect of electron-electron interactions, but there seems to be no fundamental obstacle for using this calculation as part of a self-consistent scheme. Studies in this direction and the effect of an electric field are under way.

ACKNOWLEDGMENTS

I am much in debt to Professor Harold McIntosh for introducing me to the numerical applications of the Weyl theory. I thank Docent J. L. Calais for many critical discussions and Professor P. O. Lowdin for the hospitality I enjoyed during my stay with the Quantum Chemistry Group, Uppsala University, where this work was done. R. Sitte patiently assisted in carrying through the calculations.

---

[1] Ig. Tamm, Phys. Z. Sowjetunion 1, 733 (1932).
[2] W. Shockley, Phys. Rev. 56, 317 (1939).
[3] H. Statz, Z. Naturforsch. A 5, 534 (1950).
[4] B. A. Lippmann, Ann. Phys. (N.Y.) 2, 16 (1957).
[5] F. K. Schulte and H. Bross, Z. Phys. 224, 190 (1969).
[6] A. A. Cottey, J. Phys. C 6, 2446 (1973).
[7] N. García and J. Solana, Surf. Sci. 36, 262 (1973).
[8] E. Caruthers, L. Kleinman, and G. P. Alldredge, Phys. Rev. B 8, 4570 (1973).
[9] J. A. Appelbaum and D. R. Hamann, Phys. Rev. B 6,

2166 (1972).
[10] N. García, J. Solano, and N. Cabrera, Surf. Sci. 38, 455 (1973).
[11] V. M. Kenkre, Phys. Lett. 37A, 371 (1971).
[12] D. S. Boudreaux, Phys. Rev. B 1, 4551 (1970).
[13] H. Weyl, Math. Ann. 68, 220 (1910).
[14] M. Hehenberger, H. V. McIntosh, and E. Brändas, Phys. Rev. A 10, 1494 (1974).
[15] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955), Chap. 9.
[16] H. V. McIntosh, Seminar Notes, Quantum Chemistry Group, Uppsala University, 1972 (unpublished).
[17] D. Sarafyan, J. Math. Anal. Appl. 40, 436 (1972).
[18] R. O. Jones, in *Surface Physics of Semiconductors and Phosphors*, edited by C. G. Scott and C. E. Reed (Academic, New York, 1974), Chap. 2.
[19] S. G. Davinson and J. D. Levine, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1970), Vol. 25.
[20] G. Blanch, in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. Stegun (Dover, New York, 1965).
[21] J. C. Slater, Phys. Rev. 87, 807 (1952).
[22] J. D. Levine, Phys. Rev. 171, 701 (1968).
[23] J. Appelbaum and D. R. Hamann, Phys. Rev. B 4, 1122 (1972).
[24] M. W. Cole, Phys. Rev. B 2, 4239 (1970).
[25] D. E. Eastman and W. D. Grobman, Phys. Rev. Lett. 30, 177 (1973).
[26] H. D. Hagstrum and G. E. Becker, Phys. Rev. B 8, 1580 (1973).
[27] R. Haydock, V. Heine, M. J. Kelly, and J. B. Pendry, Phys. Rev. Lett. 29, 868 (1972).
[28] R. Haydock and M. J. Kelly, Surf. Sci. 38, 139 (1973).
[29] J. Burton and G. Jura, J. Phys. Chem. 71, 1937 (1967).
[30] G. Allen and M. Lannoo, Surf. Sci. 40, 375 (1973).
[31] S. Andersson and J. B. Pendry, J. Phys. C 6, 601 (1973).