Voltage-dependent STM image of a charge density wave

William Sacks, Dmitri Roditchev, and Jean Klein
Groupe de Physique des Solides, Universités Paris 7 et Paris 6, Laboratoire Associé au CNRS, 2 place Jussieu, 75251 Paris Cédex 5, France
(Received 17 November 1997)

In the present work we write a general expression for the local density of states (LDOS) due to a commensurate charge density wave (CDW). The main goal is to investigate the voltage dependence of the contrast in the scanning tunneling microscope (STM) images of materials showing CDW's. For layered materials, having nearly two-dimensional electronic structures, the problem is complicated by the many-band situation near the Fermi level, and by the incomplete band gapping. Nevertheless, a simple perturbation method allows one to relate the amplitude and phase of the CDW to features of the band structure. We emphasize the role of particular characteristic energies, at which the CDW has a large contribution from special **k** points of the surface Brillouin zone, leading to different modulations in the STM image. In a second part of the paper, we consider the voltage-dependent contrast of $NbSe_2$. For this material, we find that the amplitude and the phase of the CDW change significantly as a function of energy (or voltage), resulting in a number of different possible motifs. For example, the direct comparison between occupied and empty states reveals that new states on the order of $E_F \pm \Delta$, giving a dominant contribution to the LDOS, have different phases. As a result, in the corresponding STM images, the maxima of the corrugation have shifted positions along the diagonal of the conventional unit cell. [S0163-1829(98)00320-8]

I. INTRODUCTION

While a large number of STM experiments exploit its unmatched atomic resolution, in many instances it is the additional voltage dependence of the tunneling current that gives full information on the local electronic structure of the sample. This fact was recognized early on, and was applied to the visualization of different localized electronic states on semiconductor surfaces, for example Si(1,1,1), at selected bias voltages.[1–5] This voltage (or equivalently energy) dependence of the images was in some cases decisive in the determination of the surface structure.[6] Since then, there have been other cases in which the STM was shown to be sensitive to the local surface electrons (or holes) rather than to the ''topography.'' Among the earlier works, one can cite the screening charge due to oxygen on GaAs,[7] or a small molecule on a surface,[8] and the direct imaging of the charge density wave (CDW) in transition-metal dichalcogenides.[9–12]

The interpretation of such experiments follows qualitatively from Lang's extension[13] of the theory of Tersoff and Hamann[14] expressing the bias-dependent tunneling current. If $\psi_k(\mathbf{r})$ is the surface wave function at the position $\mathbf{r}$ of the tip, then the local density of states (LDOS) at the energy E is defined as

$$\rho(\mathbf{r},E)=\sum_k |\psi_k(\mathbf{r})|^2 \delta(E-E_k). \tag{1}$$

When the junction is biased with a voltage V (sample grounded), the (zero temperature) tunnel current is

$$I=\frac{e}{h}\int_{E_F}^{E_F+eV} T(V,E)\rho(\mathbf{r},E)\rho_t(E-eV)dE, \tag{2}$$

where ρ_t is the tip state density and $T(V,E)$ is a barrier coefficient. Thus for very small bias, the Ohmic part of Eq. (2) gives back Tersoff and Hamann's result.

The energy window for tunneling at zero temperature is the range E_F to E_F+eV. By merely choosing the sign of V, one can image either the occupied states (electrons tunneling from the sample to the tip) or the unoccupied states (electrons tunneling to the sample). Using simultaneous imaging $(+V,-V)$ in the study of GaAs and Si(1,1,1) 2×1, Feenstra found the direct *spatial* difference between the filled and empty states.[4–6] In the latter case, the π-bonded chain model accounts well for the phase opposition of the surface states across the band gap. On Si(1,1,1)7×7, selecting different bias voltages reveals electronic states having maxima at different sites within the surface unit cell,[2,3] which can be identified on the basis of the structure and bonding at this surface.

The question that arises is, for which other samples does one expect the STM image to deviate from the atomic topography, and to be strongly influenced by the electronic structure? In a general way, for anisotropic materials that show quasi-two-dimensional (or one-dimensional) electronic structures, the wave-vector sum in the LDOS [Eq. (1)] is restricted due to the Fermi surface topology. For example, in an ideal layered structure, a constant energy surface consists of open cylinders running perpendicular to the layers. Then the wave vector $\mathbf{k}_\parallel$ takes on a value along a curve in the surface Brillouin zone. We thus expect stronger electronic effects in the STM image to occur in materials such as graphite, MoS_2, and even high-T_c superconductors (BiSrCaCuO, etc.). Graphite is actually an extreme case since, near the Fermi level, states are confined to nearly a point at the zone edge. Tersoff has shown[15] that this leads to an unusually large corrugation of the STM image.

Charge density wave materials, such as the transition-metal dichalcogenides ($NbSe_2$, $NbSe_3$, $TaSe_2$, etc.), fall into this class of anisotropic properties. Our particular motivation in the study of $2H$-$NbSe_2$ is to investigate the electronic structure effects in the STM images when the material is in the CDW state. As is well known, $NbSe_2$ has both a CDW transition ($T_{CDW} \approx 35$ K) as well as a superconducting one, at $T_c = 7.2$ K. In Ref. 16, we described low-temperature STM observations of the charge density wave at 4.2 K. The specific problem we addressed was the possibility of a relative change in contrast for states above, or below, the CDW energy gap. In a simultaneous and reversed-bias mode, we showed that there was a phase shift occurring in the LDOS corresponding to states at opposite energies with respect to the Fermi level, and even for energies within the gap. This was not a simple contrast reversal, as one might have expected, but an unknown phase shift in the LDOS. In addition, since the total density is a superposition of the atomic corrugation with the CDW one, the STM image results in an intricate moiré pattern. Some theoretical questions naturally arise: What phase shift in the corresponding electron (hole) wave function does one expect? What is the dependence of this phase on the energy, and hence the tunneling voltage? Despite the voluminous literature on CDW materials and previous STM investigations,[9–12,17–24] these questions were almost never asked, and a simple theory is still needed.

The full description of the electronic structure in the CDW state is a challenging problem: The CDW involves not only the static modulation of the charge density, but also the detailed periodic lattice distortion (PLD) involving the displacement of often many atoms per unit cell (54 for $2H$-$NbSe_2$). Our approach to the problem is to notice that (within the Tersoff and Hamann framework) the STM measures the local density of states at some distance from the surface, and hence is quite insensitive to the precise, and small, atomic displacements. We assume that the CDW is exactly commensurate to the lattice, for simplicity. Then by perturbation theory, the Bloch functions are formed by the coupling of states labeled by $\mathbf{k}$, $\mathbf{k}-\mathbf{g}$, $\mathbf{k}-\mathbf{g}'$,… where $\mathbf{g}$ is a reciprocal-lattice vector of the CDW system. The STM then responds to the lowest nontrivial terms of the relevant Bloch function, together with the usual atomic corrugation.

It is then possible to write a quite general expression for the LDOS, and to investigate the physical parameters, its amplitude and phase, at different characteristic energies. In spite of the many-band situation near the Fermi level, some of them overlapping, we show that the largest contribution to the CDW modulation arises from special $\mathbf{k}$ points of the surface Brillouin zone, in particular at the zone boundaries. As a result, the problem becomes tractable, and it is this aspect we hope could be useful in the study of different surfaces.

The paper is then presented in the following order: Using the approximate Bloch functions for the CDW, in Sec. II we write a general expression for the CDW component of the LDOS. The textbook example of a one-dimensional atomic chain is used to clarify many aspects prior to treating the two-dimensional CDW. In Secs. III and IV we apply our method to the problem of $NbSe_2$. Beginning with the question of the Fermi surface nesting, and the gapping of the band structure due to the CDW, we identify the new states which arise on the order of Δ, the gap parameter. In Sec. III we show that these states, having characteristic phases, must lead to a definite motif in the STM image. There we emphasize the $\mathbf{k}$ dependence of the modulation, and show which $\mathbf{k}$ points give the major contribution. In Sec. IV, we discuss the amplitude and phase of the CDW as a function of energy, for a wide range of values, from within to beyond the gap. Examining the modulation in real space, we find that the maxima of the LDOS have different positions within the conventional unit cell, depending on the energy. This is used to interpret the final results for the voltage dependence of the STM images.

II. DESCRIPTION OF THE METHOD: CDW COMPONENT OF THE LDOS

Electronic structure calculations of CDW systems are usually quite complex, even in a tight-binding approach.[17,18] Prior to the STM, the main focus was on the energy of the system and the displacement of the atoms. Our concern here is to extract the main features of the real-space charge modulation, relevant to the STM image. The principal simplification in the theoretical treatment is a consequence of the barrier penetration damping on the surface Bloch functions. As detailed by Tersoff and Hamann,[14] a key aspect is that the STM measures the LDOS at some distance z from the surface, on a scale larger than $\kappa^{-1} = \sqrt{\hbar/2m\varphi} \approx 1$ Å, the decay length of the wave function into the vacuum, where φ is the work function. In this regime, the tunneling is exponentially small: $I \propto e^{-2\kappa z}$, and the local density of states defined in Eq. (1) follows the same asymptotic decay.

Exploiting this property, we begin by writing the Bloch wave function associated with a CDW, which is assumed periodic in the plane of the surface. It is straightforward to show[14] that it must have the form

$$\psi_{\mathbf{k}}(\mathbf{x},z) = \sum_{\mathbf{g}} c_{\mathbf{k}-\mathbf{g}} e^{i(\mathbf{k}-\mathbf{g})\cdot\mathbf{x}} e^{-\kappa_{\mathbf{k}-\mathbf{g}} z}, \tag{3}$$

where $\mathbf{k} = (k_x, k_y)$ and $\mathbf{g}$ is the two-dimensional (2D) reciprocal lattice. The inverse decay length is

$$\kappa_{\mathbf{k}-\mathbf{g}} = \left((\mathbf{k}-\mathbf{g})^2 + \frac{2m}{\hbar^2} |E| \right)^{1/2},$$

where E is measured from the vacuum level. Therefore, each plane wave in the sum, $e^{i(\mathbf{k}-\mathbf{g})\cdot\mathbf{x}}$, has an amplitude that decays very fast as $\mathbf{g}$ increases, and this property determines the limiting resolution of the instrument.[14,25]

To extract from Eq. (3) the dominant term owing to the CDW, we can consider at the outset a number of simplifying assumptions: First, supposing that $\kappa z \gg 1$, then only the smallest wave vectors $\mathbf{g}_n$ will contribute significantly to the Bloch sum. If $\mathbf{k}$ is restricted to the first Brillouin zone, then the $\mathbf{g}_n$ can be the nearest reciprocal lattice points. Most importantly, the $\mathbf{g}_n$ then coincide with the generating vectors of the CDW, i.e., $\mathbf{g}_n = \mathbf{Q}_n$. Keeping only two terms as a first example, thus ignoring the atomic corrugation, we have

$$\psi_{\mathbf{k}}(\mathbf{x},z) \approx e^{i\mathbf{k}\cdot\mathbf{x}} (c_{\mathbf{k}} e^{-\kappa_{\mathbf{k}} z} + c_{\mathbf{k}-\mathbf{g}} e^{-i\mathbf{g}\cdot\mathbf{x}} e^{-\kappa_{\mathbf{k}-\mathbf{g}} z}). \tag{4}$$

Second, as will be justified in the specific case of a CDW, the wave vector $\mathbf{k}$ is on or near the Brillouin zone boundary,

for energies near the Fermi level. Since the exponential damping factors are slowly varying functions, we can simplify further by setting $\mathbf{k}=\mathbf{g}/2$ in their arguments:

$$\psi_{\mathbf{k}}(\mathbf{x},z)\approx e^{i\mathbf{k}\cdot\mathbf{x}}\varphi(z)(c_{\mathbf{k}}+c_{\mathbf{k}-\mathbf{g}}e^{-i\mathbf{g}\cdot\mathbf{x}}), \tag{5}$$

where $\varphi(z)=e^{-\kappa_{g/2}z}$. The physical interpretation is then straightforward: quite far from the surface, the wave function associated with the CDW modulation appears as if only a single Bragg reflection occurs. Once the complex amplitudes $c_{\mathbf{k}}$ and $c_{\mathbf{k}-\mathbf{g}}$ are determined from the particular band structure, using a variety of possible methods, it is then a relatively direct calculation for the corresponding LDOS. One can use either forms (4) or (5) for the wave function, since the accuracy depends on the distance z and the variation of $\mathbf{k}$ in the surface Brillouin zone. Finally, we note that the approximation (5) leads to the factoring of the LDOS into perpendicular and parallel parts:

$$\rho(\mathbf{r},E)\approx\varphi(z)^2\sum_{\mathbf{k}}|\psi_{\mathbf{k}}(\mathbf{x})|^2\delta(E-E_{\mathbf{k}}), \tag{6}$$

which is often assumed to be valid. This approximation is actually quite correct for graphite, since the Fermi surface lies very close to the zone boundary, at the K corner point.[15] We will show explicitly in Sec. III that it is still reasonable in the case of a CDW.

In the following we pursue the idea of single Bragg reflection, and its consequence on the CDW modulation through the amplitudes $c_{\mathbf{k}}$ and $c_{\mathbf{k}-\mathbf{g}}$. Given the complexity of the general problem, we give a very brief summary of the 1D case, but stressing those aspects that are relevant to the energy dependence of the 2D modulation and LDOS.

A. One-dimensional modulation

As is well known, from the electronic point of view the one-dimensional metal having a band dispersion,

$$E_k=\varepsilon_0+2\beta\cos(ka), \tag{7}$$

is unstable to a period doubling, or dimerization.[19] This is because the Fermi wave vector k_F lies at $\frac{1}{2}\Gamma M$, the band being half-filled [Fig. 1(a)]. Period doubling, with new reciprocal vector $g=G/2$, lowers the total electron energy by opening a gap at the new zone boundary, the m point. The Fermi ''surface,'' which consisted of only two points (k_F and $-k_F$) for the unperturbed chain, is completely removed in the dimerized chain [Fig. 1(b)]. This metal-insulator transition is characterized by the perfect nesting of the Fermi surface: the wave vector $\mathbf{Q}=\mathbf{g}$ translates one Fermi point onto the other. A general discussion of nesting in low-dimensional conductors can be found in Ref. 20.

A simple model to describe the CDW modulation, hence the coefficients in the Bloch expansion (5), uses degenerate perturbation theory. If the system has a new period $2a$, then the bands E_k and E_{k-g} become degenerate at the m point of the new Brillouin zone. The corresponding states, $|k\rangle$ and $|k-g\rangle$, are coupled through the periodic potential, $V(x)=V^{\mathrm{CDW}}(x)$, with fundamental Fourier component $V_g^{\mathrm{CDW}}=\Delta$, which we associate with the static charge-lattice interaction. Then, using the linear combination

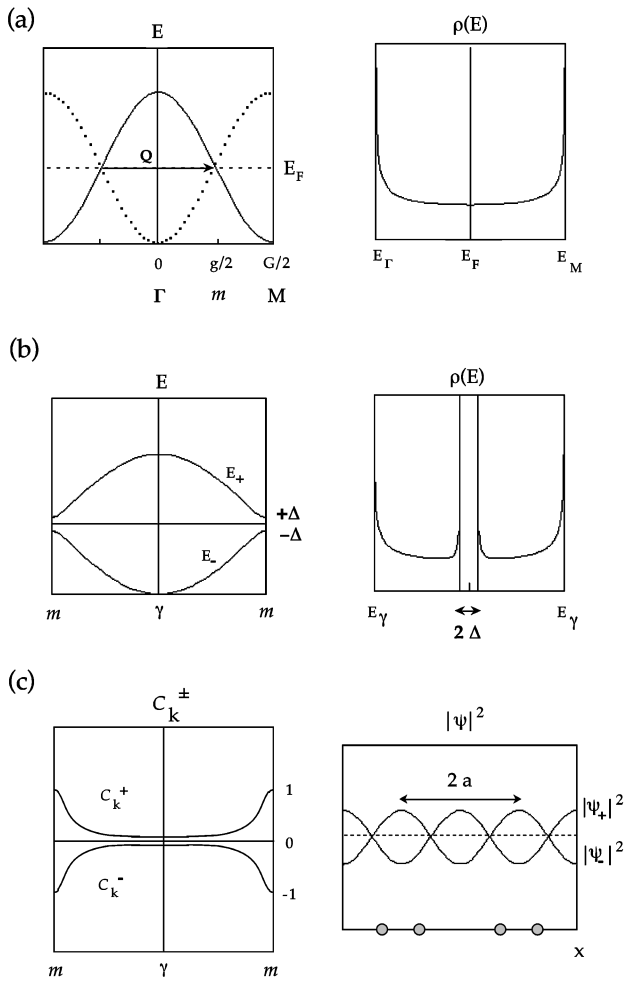

FIG. 1. Prior to treating the 2D case, it is useful to consider the gap opening in a one-dimensional conductor with a half-filled band, assuming period doubling. (a) The unperturbed band (solid line) and the folded band (dotted line) are doubly degenerate at k_F (or $\Gamma M/2$). (b) A gap of magnitude 2Δ at the new Brillouin zone boundary (m point) separates the two bands E_+ and E_-. The corresponding DOS is shown in the right panels of (a) and (b). (c) The k-dependent CDW amplitude $C_k^{\pm}$ has two branches, that tend to ±1 at the m point, the standing wave conditions. The corresponding modulations $|\Psi_k^{\pm}(x)|^2$ of wave vector $Q=g$ are in phase opposition.

$$\psi_k=N_k(|k\rangle+\alpha_k|k-g\rangle), \tag{8}$$

one obtains two bands, E^+ and E^-, separated by a gap 2Δ at the m point, and the corresponding solutions for α_k:

$$\alpha_k^{\pm}=\frac{E^{\pm}-E_k}{\Delta}=\frac{\Delta}{E^{\pm}-E_{k-g}}. \tag{9}$$

There are also two different CDW modulations, for the occupied and unoccupied bands:

$$|\psi_k^{\pm}(x)|^2=1+C_k^{\pm}\cos(gx), \tag{10}$$

where the new amplitude is $C_k^{\pm}=2N_k^2\alpha_k^{\pm}$. This expression for the modulation, which neglects the atomic contribution, has a period $2a$ and a wave vector $g=Q$, as expected.

Although quite standard, we anticipate some important features of the 2D modulation by considering these relations.

In particular, the CDW amplitude $C_k^{\pm}$ is k-dependent and, using the normalization factor $N_k=1/\sqrt{1+\alpha_k^2}$, it is related to the band structure via

$$C_k^{\pm}=\frac{2\alpha_k^{\pm}}{1+(\alpha_k^{\pm})^2}. \tag{11}$$

It is quite small except for points near the zone boundary [Fig. 1(c)]. Indeed, at the m point we obtain the extremal values $C_k^{\pm}=\pm 1$. Thus at the energies corresponding to the band edges, $E^{\pm}=\pm\Delta$, we have the standing wave condition, $|C_k^{\pm}|=1$, and the CDW modulation (10) has nodes. The change in sign of $C_k^{\pm}$, or a phase change of π in the LDOS, would appear as a relative contrast inversion of the STM image between the filled and the empty states.

These arguments are nearly identical to the Pandey π-bonded chain model that Feenstra used to explain the contrast inversion observed on the Si(1,1,1)2×1 surface states.[4,5] It is an open question whether a phase shift in the density can occur across a CDW gap, by analogy. In the general 2D problem, we define an analogous k-dependent amplitude, which becomes a complex function. Its argument will have an important influence on the phase of the LDOS associated with the charge modulation.

B. CDW modulation in two dimensions

The details of the Fermi surface nesting in $NbSe_2$, its band structure in the CDW state, and explicit calculations of the wave function, will be treated in the following section. In the expansion of a 2D modulation, the approximate Bloch function (5) should contain at least the nearest reciprocal vectors $\mathbf{g}_n$. The price to pay for the reduced zone scheme is the inevitable multiple band situation. Indeed, even if the normal spectrum is only a single band, in the reduced zone scheme of the CDW state, there are a number of sometimes overlapping bands. Still, this seems to be the simplest description.

The CDW in materials such as $NbSe_2$ has ternary symmetry and, if we assume it is exactly commensurate to the underlying lattice, the hexagonal reciprocal lattice is generated by the three vectors $\{\mathbf{g}_0,\mathbf{g}_1,\mathbf{g}_2\}$:

$$\mathbf{g}_0=g(\sqrt{3}/2,1/2),$$

$$\mathbf{g}_1=g(0,1),$$

$$\mathbf{g}_2=g(-\sqrt{3}/2,1/2),$$

where g is the reciprocal-lattice constant. Including only three plane waves as a first approximation, the wave function is then

$$\psi_{\mathbf{k}}=N_{\mathbf{k}}(|\mathbf{k}\rangle+\alpha_{\mathbf{k}}|k-\mathbf{g}_0\rangle+\beta_{\mathbf{k}}|k-\mathbf{g}_1\rangle), \tag{12}$$

with the normalization factor $N_{\mathbf{k}}=\sqrt{1+|\alpha_{\mathbf{k}}|^2+|\beta_{\mathbf{k}}|^2}$. We stress that the electronic structure enters the problem through the complex coefficients $\alpha_{\mathbf{k}}$ and $\beta_{\mathbf{k}}$, as in the one-dimensional case above. In all expressions, we omit the notation of the band index. Using the obvious property $\mathbf{g}_2=\mathbf{g}_1-\mathbf{g}_0$, the modulation density is

$$|\psi_{\mathbf{k}}(\mathbf{x})|^2=1+2N_k^2\mathrm{Re}[\alpha_{\mathbf{k}}e^{-i\mathbf{g}_0\cdot\mathbf{x}}+\beta_{\mathbf{k}}e^{-i\mathbf{g}_1\cdot\mathbf{x}}+\alpha_k^*\beta_{\mathbf{k}}e^{-i\mathbf{g}_2\cdot\mathbf{x}}], \tag{13}$$

from which a useful form can be written if we define three new complex amplitudes $C_{\mathbf{k},n}$, one for each of the three directions $\{\mathbf{g}_0,\mathbf{g}_1,\mathbf{g}_2\}$:

$$C_{\mathbf{k},n}=\{C_{\mathbf{k},0},C_{\mathbf{k},1},C_{\mathbf{k},2}\}=2N_{\mathbf{k}}^2\{\alpha_{\mathbf{k}},\beta_{\mathbf{k}},\alpha_{\mathbf{k}}^*\beta_{\mathbf{k}}\}. \tag{14}$$

We finally obtain

$$|\psi_{\mathbf{k}}(\mathbf{x})|^2=1+\sum_{n=0}^{2}|C_{\mathbf{k},n}|\cos(\mathbf{g}_n\cdot\mathbf{x}-\theta_{\mathbf{k},n}), \tag{15}$$

where $\theta_{\mathbf{k},n}=\arg(C_{\mathbf{k},n})$. For a given $\mathbf{k}$, this modulation has quite a simple interpretation: It is a hexagonal standing wave, due to the lowest-order Bragg reflection associated with the new superlattice. This expression could also be extended to a larger number of plane waves, or modified to include the z dependence of the amplitude, if higher precision is required. The atomic corrugation gives an additional term similar to Eq. (15), but with a larger wave vector.

As a consequence of extending to two dimensions, $C_{\mathbf{k},n}$ is a function of $\mathbf{k}$ throughout the new Brillouin zone. The amplitude and the phase of the CDW can be computed once the complex functions $C_{\mathbf{k},n}$ are known, these being directly related to the band structure via Eq. (14). We find that for $NbSe_2$, the amplitudes $|C_{\mathbf{k},n}|$ retain some aspects of the 1D problem. More significantly, since $C_{\mathbf{k},n}$ has a complex value, it introduces a new phase $\theta_{\mathbf{k},n}$ in the modulation (15). The motif in the STM image then becomes energy dependent.

C. General expression for the LDOS

We conclude this section by deriving a simple form for the CDW part of the LDOS. Using the cosine form of the modulation (15), together with the definition (1), we must determine

$$\rho(\mathbf{x},E)=\sum_{\mathbf{k}}\left\{1+\sum_{n=0}^{2}|C_{\mathbf{k},n}|\cos(\mathbf{g}_n\cdot\mathbf{x}-\theta_{\mathbf{k},n})\right\}\delta(E-E_{\mathbf{k}}), \tag{16}$$

where we omit the sums over spin and band index, for simplicity. In two dimensions, the above sum (or integral) is over all values of $\mathbf{k}$ on the constant energy contour, $l(\mathbf{k})$, implied by the equation $E=E_{\mathbf{k}}$. If $E=E_F$, then the integration is along the cross section of the Fermi surface. Thus, the exact evaluation of the LDOS, except for the most simple contours $l(\mathbf{k})$, is quite difficult, and one usually must resort to subtle numerical techniques.

For the purposes of interpreting the STM images, we notice that the first term in Eq. (16), which is independent of $\mathbf{x}$, is just the average DOS of the system, $\rho(E)$, but in the CDW state:

$$\rho(E)=\sum_{\mathbf{k}}\delta(E-E_{\mathbf{k}}). \tag{17}$$

The second term, $\delta\rho(\mathbf{x},E)$, is the corrugation, which can be understood as the excess density transferred to the CDW, or removed: $\rho(\mathbf{x},E)=\rho(E)+\delta\rho(\mathbf{x},E)$. By expanding the co-

sine, one eliminates the undesirable **x** argument in the sum over **k** in Eq. (16). Then, by elementary means one obtains

$$\delta\rho(\mathbf{x},E)=\sum_{n=0}^{2} a_n\cos(\mathbf{g}_n\cdot\mathbf{x})+b_n\sin(\mathbf{g}_n\cdot\mathbf{x}) \tag{18}$$

with the coefficients

$$a_n=\sum_{\mathbf{k}} \mathrm{Re}(C_{\mathbf{k},n})\,\delta(E-E_{\mathbf{k}}), \tag{19a}$$

$$b_n=\sum_{\mathbf{k}} \mathrm{Im}(C_{\mathbf{k},n})\,\delta(E-E_{\mathbf{k}}). \tag{19b}$$

These amplitudes (a_n,b_n) involve the sum of the complex coefficients $C_{\mathbf{k},n}$ along the constant energy contour $l(\mathbf{k})$. This summation is analogous to the usual DOS but with $C_{\mathbf{k},n}$ as a weighting factor at the relevant **k** points. Its real and imaginary parts thus contribute to the amplitude of the symmetric and antisymmetric parts of the density, respectively.

In order to interpret the phase observed in the STM images, it is more convenient to write Eq. (18) in terms of the cosine again. Obviously Eqs. (19a) and (19b) can be combined into the single complex equation:

$$a_n+ib_n=\sum_{\mathbf{k}} C_{\mathbf{k},n}\,\delta(E-E_{\mathbf{k}}), \tag{20}$$

which leads to the simple final result for the density:

$$\rho(\mathbf{x},E)=\rho(E)+\sum_{n=0}^{2} A_n(E)\cos[\mathbf{g}_n\cdot\mathbf{x}-\varphi_n(E)], \tag{21}$$

with

$$A_n(E)=|a_n+ib_n|=\left|\sum_{\mathbf{k}} C_{\mathbf{k},n}\,\delta(E-E_{\mathbf{k}})\right|, \tag{22a}$$

$$\varphi_n(E)=\arg(a_n+ib_n)=\arg\left(\sum_{\mathbf{k}} C_{\mathbf{k},n}\,\delta(E-E_{\mathbf{k}})\right). \tag{22b}$$

The latter is a quite general form of the LDOS for a CDW involving only the smallest nontrivial Bragg reflection terms. In particular, the argument $\varphi_n(E)$ appears, which could be responsible for any phase changes in the bias-dependent STM images. For example, if $\varphi_n(E)$ were to change from $\{0,0,0\}$ to $\{\pi,-\pi,\pi\}$, for two different energies, this would be a complete contrast inversion. A definite phase shift has been found on the Pb/Ge(1,1,1) surface, which, although attributed to the formation of a CDW, is not a classic system.[26]

In reality, due to the **k** summation in Eq. (20), the phase of the LDOS $\varphi_n(E)$ is not directly the argument of $C_{\mathbf{k},n}$ (i.e., $\theta_{\mathbf{k},n}$), as it would be for a one-dimensional problem. Nevertheless, we shall find that in many instances, a particular **k** point dominates the sum due to the intensity of Bragg reflection there: A straightforward physical interpretation is then possible. On the other hand, the correct energy-dependent amplitudes $A_n(E)$ can only be obtained by evaluating the integration implied by Eq. (22a). We now focus on the particular example of $NbSe_2$, on which voltage-dependent STM studies have been done, and calculate explicitly the quantities derived here.

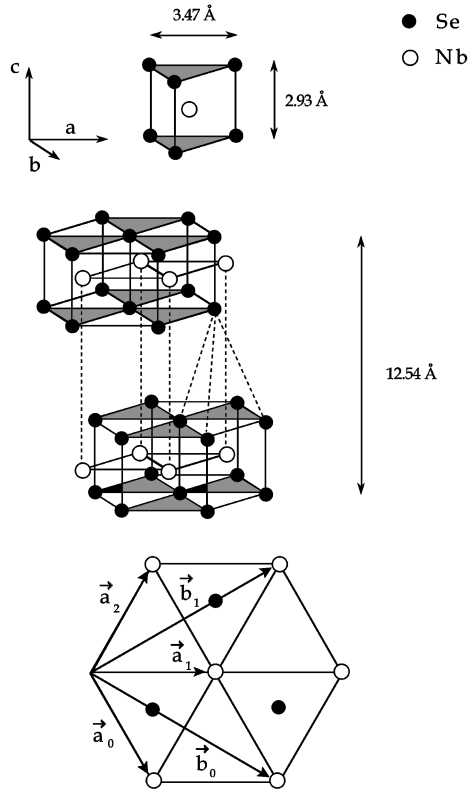

FIG. 2. Structure of $NbSe_2$ showing basic unit (upper panel) of a layer, a perspective view of two layers (middle panel), and the surface view of a single layer (lower panel). Black circles indicate Se sites, and open circles indicate the Nb sites. In the surface view, the Se atoms are in the first plane, the Nb atoms lying in the second plane, 1.47 Å below.

III. BAND STRUCTURE AND WAVE FUNCTIONS FOR THE CDW IN $NbSe_2$

The mechanisms behind the contrast in STM images becomes interesting for materials having very anisotropic electronic structures. In these materials, such as layered structures or chains, the band structure and Fermi surface have unique properties, and electronic instabilities are typical.[19,20] In some instances, there are still debated questions concerning the nesting, gapping, and Fermi surface reduction,[21,22] as is the case for $NbSe_2$. Wangbo *et al.*[23] have recently aimed to clarify the picture for the transition-metal dichalcogenides, and also the STM and AFM images.[24] In this section, we use a perturbation model for the band structure of $NbSe_2$, in both the normal and the CDW states. As a result we suggest a simple interpretation for the nesting in $NbSe_2$.

A. Unperturbed band structure and the nesting question

In Fig. 2 we show both the bulk and surface structures of $2H$-$NbSe_2$. Band-structure calculations of this material, or the related $2H$-$TaSe_2$, show that qualitatively the Fermi level is crossed by a band associated with the transition metal. The recent *ab initio* calculation of Kikushi and

Tsukada (KT) (Ref. 22) shows that the width of this band is about 2 eV, and is derived from the Nb $4d_{z^2}$ orbital. The narrow $4p$ Se bands are all occupied, and the remaining niobium bands are empty. We assume that this Nb band is due to a single conduction electron per unit cell: the band is therefore half-filled. Since only a very small portion near the Fermi level is involved in the CDW,[21] this should be a reasonable starting point for determining the CDW band structure. KT do suggest a small Nb-Se hybridization in their paper.

Since the coupling between the layers is small, our focus will be only on a single sandwich consisting of three atomic planes (Se-Nb-Se). This is analogous to the graphene model of graphite, i.e., the consideration of only a single layer. Using t_1 and t_2 as free parameters, and ε_0 the Nb d orbital energy, the 2D band can be approximated by

$$E_{\mathbf{k}}=\varepsilon_0+t_1\sum_{a_n}\cos(\mathbf{k}\cdot\mathbf{a}_n)+t_2\sum_{b_n}\cos(\mathbf{k}\cdot\mathbf{b}_n), \qquad (23)$$

where the $\mathbf{a}_n$ and the $\mathbf{b}_n$ are nearest neighbors, and next nearest neighbors, respectively. The three parameters are fit to the band calculation of KT. In Fig. 3(a) we show the Fermi surface given by Eq. (23), consisting of a central cylinder S_I around the Γ point, and six cylinders S_{II} centered on the K corner points of the hexagonal Brillouin zone. For future reference, we note that for a lattice constant a, the values of G and K are $G=4\pi/\sqrt{3}a$, and $K=4\pi/3a$, respectively. In Fig. 3(b) we show the band dispersion along the high-symmetry directions, together with the total DOS [Fig. 3(c)]. The main features within the band are the relative maximum, at the K point, and the saddle point along the ΓK direction. These give rise to a discontinuity and a singularity in the DOS, as expected for a 2D system. The vanishing of the group velocity at the relevant $\mathbf{k}$ points accounts for this behavior. Interlayer coupling will tend to round off the singularities. Nevertheless, one might have expected the saddle point to play a role in the CDW formation. Wilson argues[21] that, in view of its position in energy, and its aspect with or without the CDW gap, the saddle point is probably not involved in the mechanism.

We believe that it is the S_{II} portions of the Fermi surface, i.e., those centered around the K point, that are involved in the formation of the CDW, leaving S_I approximately intact. While being far from a proof, we will show below that this is consistent with a $3a$ superlattice formation. In Ref. 22, KT propose the nesting of the S_I surface, which naturally leads to a period *doubling* CDW. It is clear from the STM images, and previous neutron diffraction,[27] that the $3a$ superlattice is observed. Still, the electronic structure in the CDW state is not well established: Wilson[21] gives some detail on the possible Fermi surface of the similar material, $2H$-$TaSe_2$. He uses the band structure of Wexler and Woolley,[17] together with a Harrison construction for a period tripling, to deduce four distinct pieces of the Fermi surface. Later work on the same material is largely based on the calculation of Doran and Woolley,[18] who also assume the exact $3a$ condition, so the ''nesting question'' is still not settled.

The main characteristic of nesting in one dimension, as discussed previously, is that the entire Fermi surface is completely removed by the CDW gap, resulting in an insulator.

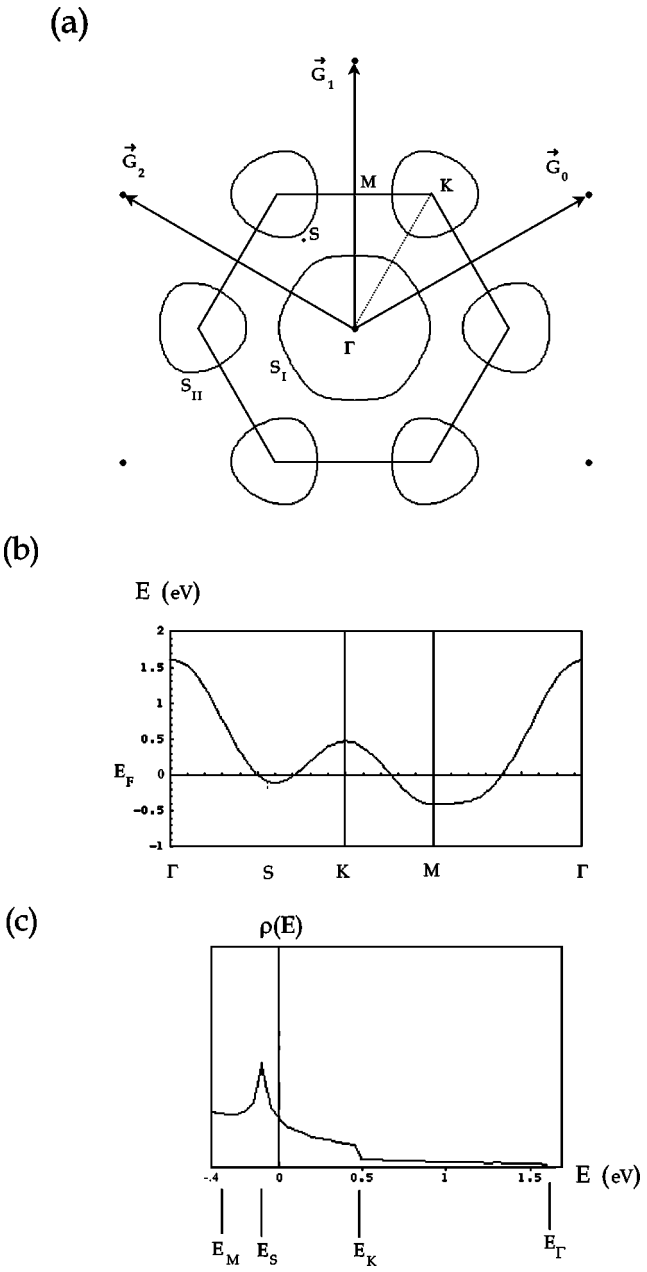

FIG. 3. (a) Reciprocal lattice of single layer $NbSe_2$ (black dots). High-symmetry points of the hexagonal Brillouin zone are labeled (Γ,K,M). The Fermi surface has two parts, S_I being the portion around Γ and S_{II} the portion centered on K. We also indicate with S the saddle point in the band dispersion. (b) Band dispersion for single layer $NbSe_2$ using a tight-binding approximation. The characteristic minimum at M, and a relative maximum at K, give the discontinuities in the DOS. (c) The saddle point S leads to a sharp singularity in the DOS, characteristic of a 2D system.

Indeed, the nesting vector $\mathbf{Q}$ maps the Fermi ''surface'' point at $-\mathbf{k}_F$ onto the symmetric point at $\mathbf{k}_F$. Then the density modulation has a wave vector $\mathbf{Q}=2\mathbf{k}_F=\mathbf{G}/2$, giving period doubling. While Fermi surface nesting is perfect in one dimension, in two dimensions (and still worse three) there is no exact nesting. In Ref. 20, Pouget gives examples of nesting in two or three dimensions, where portions of the Fermi surface have similar curvature, and can thus overlap significantly upon translation by a given $\mathbf{Q}$. The same assumption was apparently made by KT in their work.

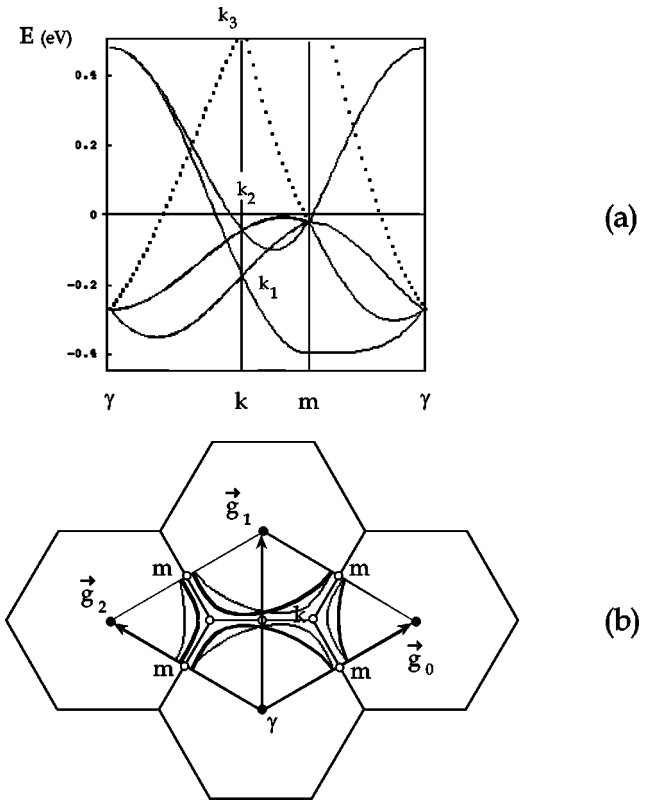

FIG. 4. (a) Folded band structure into the new hexagonal Brillouin zone assuming the period tripling condition: $\gamma k = \Gamma K/3$. High degeneracy occurs at the k point and the m point just under the Fermi level (i.e., within 50 meV). (b) Folding of the S_{II} portions of the Fermi surface into the Brillioun zone. Note the overlap particularly near the m point, whose high symmetry is also evident. More significantly, the Fermi surface is very near the boundary, particularly at the k and m points. Here we prefer to display the conventional cell, since the cutting up of the Fermi surface, due to the CDW interaction, is easier to visualize (cf. Fig. 6).

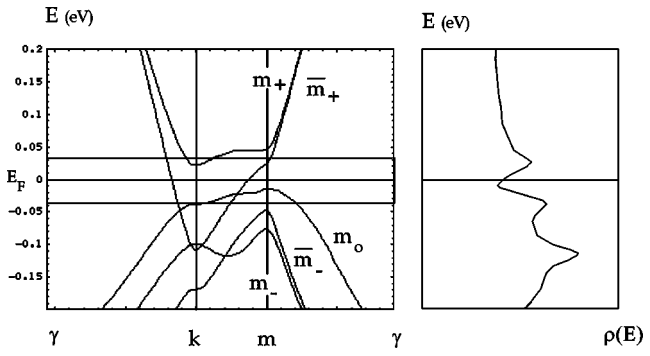

FIG. 5. Gapped band structure for the two-dimensional CDW (left panel) and the computed DOS (right panel). The new bands are labeled according to their position at the m point. The $\bar{m}_+$ band crosses the Fermi level, and gives a Fermi surface portion centered on the k point. Two new bands (m_+ and m_0) associated with the CDW appear on either side of the Fermi level, which lead to a different phase in the LDOS. The energy gap $E_F \pm \Delta$ nearly coincides with saddle points of the m_+ and m_0 bands, and not the band edges.

Our approach for $NbSe_2$, which we prefer to term quasinesting, is to notice that even if the Fermi surface does not overlap completely in the $3a$ condition, it may be *close enough* so that gap opening reduces the electron energy by the required amount. Although the full analysis is quite intricate, let us assume that the superlattice is exactly commensurate to $3a$; hence it is generated by the new reciprocal vectors: $\mathbf{g}_n = \mathbf{G}_n/3$. The new Brillouin zone is exactly $\frac{1}{9}$ the parent zone, and the corner point is reduced by three: $\gamma k = \Gamma K/3$ (small letters shall refer to the new zone). Now consider the folding of the single band of Fig. 3 into the new reduced zone, Fig. 4(a), prior to any charge-lattice interaction, which gives a total of nine bands. Then the $3a$ periodic system has highly degenerate bands, as expected, at the k corner points and the m points. Most importantly, the degenerate points k_2 and m of Fig. 4(b) are within 50 meV of the Fermi level, i.e., within the CDW interaction range. If gaps are opened in the band structure near the Fermi level due to this interaction, then the DOS will be significantly lowered there.

The idea of quasinesting can also be viewed another way by the folding of the S_{II} portions of the Fermi surface into the new reduced zone of Fig. 4(b). For reasons that will become apparent, we prefer to display the conventional unit cell defined by $\{\mathbf{g}_0, \mathbf{g}_2\}$. In the figure, we see the near overlap particularly near the m point, whose symmetry is also evident. More significantly, the S_{II} portions of the Fermi surface are near the boundary (or Bragg planes), particularly at the k and m points. We deduce that the Fermi surface will be reduced significantly, being cut up into new pieces, when the gapping potential V^{CDW} intervenes.

A remarkable feature of Figs. 4(a) and 4(b) is the symmetry of the k and m points in the vicinity of the Fermi surface: All the m points are equivalent, and there are two types of k points labeled k_1 and k_2 in Fig. 4(a). The folding of the bands reveals that the k points are each threefold degenerate, as expected, but the m point is sixfold degenerate. In summary, we find that in the reduced zone, the folded bands have a significantly different band topology in the $3a$ and $2a$ superlattice conditions, in addition to their multiplicity (nine bands compared to four). Why the system prefers the former case cannot be decided here, and must depend on the phonon part of the total energy, i.e., the low-frequency modes of the lattice.

B. Gapped band structure in the $3a$ condition

The potential V^{CDW} therefore pushes bands away from the Fermi level, as shown in Fig. 5. In addition to the new band structure, the figure shows the calculated total DOS in the 200 meV range. As a direct consequence of the gapping, the DOS at the Fermi level is lowered, as expected, while *new* states appear on the order of $E_F \pm \Delta$, at or near the high symmetry points (k and m) of the Brillouin zone boundary. The main aspect is that these new states, highly degenerate at k and m in the unperturbed problem, have a new phase. Recall that in the 1D problem the lifting of twofold degeneracy led to phase opposition in the period doubling modulation. As shown in Sec. II, the new phases are precisely determined by the argument of the complex amplitude $C_{\mathbf{k},n}$ of the modulation. In the following paragraphs, we explicitly calculated $C_{\mathbf{k},n}$ for the single layer model of the CDW in $NbSe_2$, which leads to the band structure of Fig. 5. In this

simple model, the complex amplitudes can be related to the band structure at any **k** point, but also their values at the k and m points can be determined analytically. The idea of quasinesting, while not being strictly quantitative, is a useful starting point to determine the amplitude and phase of the new CDW states.

Focusing on only one of the k points of Fig. 4(a), for example k_2, we can calculate analytically the lifting of the threefold degeneracy if we neglect all but the lowest Fourier coefficients of the potential, as well as accidental band crossing. Since the wave-function expansion is exactly as Eq. (12), the complex CDW amplitudes $C_{\mathbf{k},n}$ depend on $\alpha_{\mathbf{k}}$ and $\beta_{\mathbf{k}}$ through the established relation

$$C_{\mathbf{k},n}=2N_{\mathbf{k}}^2\{\alpha_{\mathbf{k}},\beta_{\mathbf{k}},\alpha_{\mathbf{k}}^*\beta_{\mathbf{k}}\}.$$

In turn, $\alpha_{\mathbf{k}}$ and $\beta_{\mathbf{k}}$ are solutions to

$$\begin{pmatrix} \lambda_{\mathbf{k}} & -V_{\mathbf{g}_0} & -V_{\mathbf{g}_1} \\ -V_{\mathbf{g}_0}^* & \lambda_{\mathbf{k}-\mathbf{g}_0} & -V_{\mathbf{g}_2} \\ -V_{\mathbf{g}_1}^* & -V_{\mathbf{g}_2}^* & \lambda_{\mathbf{k}-\mathbf{g}_1} \end{pmatrix} \begin{pmatrix} 1 \\ \alpha_{\mathbf{k}} \\ \beta_{\mathbf{k}} \end{pmatrix} = 0, \tag{24}$$

where λ is the energy difference $\lambda_{\mathbf{k}-\mathbf{g}}=E-E_{\mathbf{k}-\mathbf{g}}$, and where we have used the matrix elements $\langle\mathbf{k}|V^{\mathrm{CDW}}|\mathbf{k}-\mathbf{g}\rangle = V_{-\mathbf{g}}=V_{\mathbf{g}}^*$. This potential is not known explicitly, but we assume that $V_{\mathbf{g}}$ has the form

$$V_{\mathbf{g}_n}=\{\Delta_0 e^{i\varphi_0},\Delta_0 e^{-i\varphi_0},\Delta_0 e^{i\varphi_0}\},$$

corresponding to the directions $\mathbf{g}_n=\{\mathbf{g}_0,\mathbf{g}_1,\mathbf{g}_2\}$. Here Δ_0 is a free parameter, and φ_0 is a phase that restricts $V^{\mathrm{CDW}}(\mathbf{x})$ to the symmetry of the underlying lattice. The lattice is symmetric under a rotation of 120°, but also a combined rotation of 60° and a reflection. It is notably *not* symmetric under a 60° rotation alone (or sixfold symmetry), which excludes the possible value $\varphi_0=0$. There is then still a choice for a value of φ_0 consistent with these symmetries, and we admit this is a weakness of the model. However, we choose the value of $\varphi_0=\pi/2$, which is consistent with the results of Ref. 18 for the related material $TaSe_2$.

The eigenvalue problem differs significantly from the 1D case since Eq. (24) leads to a cubic equation in E, which gives a splitting into three distinct bands, instead of two. From Eq. (24), we find for the eigenvalue equation

$$\lambda_{\mathbf{k}}\lambda_{\mathbf{k}-\mathbf{g}_0}\lambda_{\mathbf{k}-\mathbf{g}_1}=\Delta_0^2(\lambda_{\mathbf{k}}+\lambda_{\mathbf{k}-\mathbf{g}_0}+\lambda_{\mathbf{k}-\mathbf{g}_1}), \tag{25}$$

and the coefficients are

$$\alpha_{\mathbf{k}}=-\left(\frac{\lambda_{\mathbf{k}}+i\Delta_0}{\lambda_{\mathbf{k}-\mathbf{g}_0}-i\Delta_0}\right), \tag{26a}$$

$$\beta_{\mathbf{k}}=-\left(\frac{\lambda_{\mathbf{k}}-i\Delta_0}{\lambda_{\mathbf{k}-\mathbf{g}_1}+i\Delta_0}\right). \tag{26b}$$

The procedure is to solve the eigenvalue problem varying Δ_0 to reasonably fit the experimental DOS. The result in Fig. 5 shows the five principal bands arising from k_1 and k_2 in the vicinity of the Fermi level that are the most relevant to the CDW modulation. Each band is labeled according to its position at the new m point. The k_3 point being out of range, the bands shown with the dotted lines in Fig. 4(a) will be ignored for the present discussion.

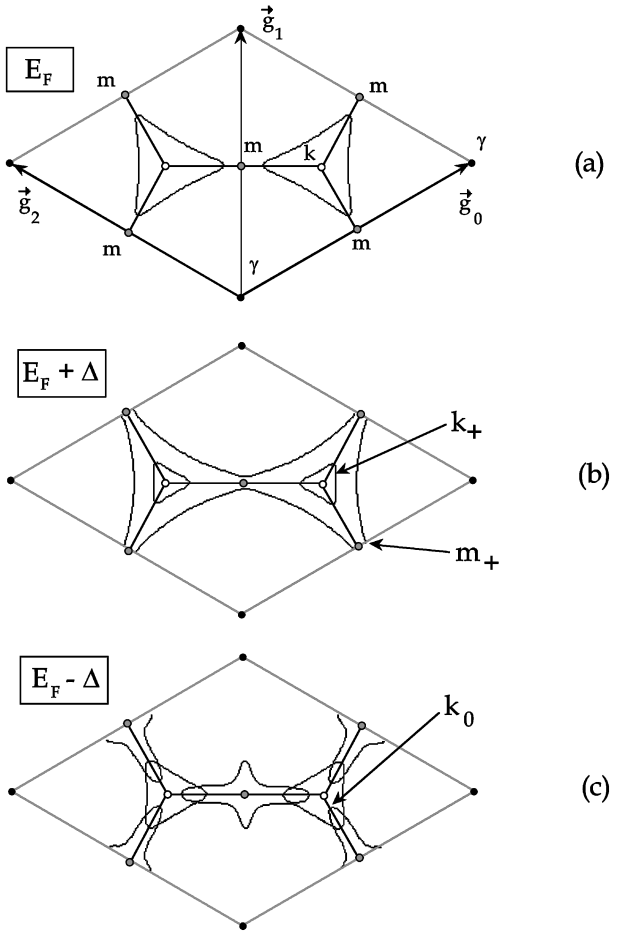

FIG. 6. Constant energy contours plotted in the conventional unit cell for the characteristic energies E_F and $E_F\pm\Delta$. The CDW contributes to the LDOS even for very small energies, since in (a) the Fermi surface piece crosses the zone boundary (or Bragg plane). The contours of gapped states, due to the CDW are allowed to come very close to or on the k and m points. In (b) both the CDW amplitude and phase are dominated by the new states near k_+ and m_+, while in (c) the states near k_0 dominate (cf. Fig. 9).

Here we note the characteristic curvature near the high symmetry points, k and m, where in addition to new band edges, there are new saddle points. The Fermi level lies in a gap between the m_0 and m_+ bands, but still crosses the $\overline{m}_+$ band. Thus the gapping leaves a triangular shaped Fermi surface piece centered on the k point, Fig. 6(a). This has two important consequences: It is clear that there is no real ''gap'' in the CDW band structure, and the material remains metallic. This is consistent with the tunneling spectroscopy measurements.[10] Second and contrary to the one-dimensional problem, when tunneling involves states within the gap, i.e., for small voltages, we expect to see the CDW modulation in the STM image. Indeed, Fig. 6(a) shows that the remaining Fermi surface piece comes close to or even crosses the zone boundary near k.

In Figs. 6(b) and 6(c) we show the constant energy contours at the gap limits, $E=E_F\pm\Delta$, where new states arise: At a given energy, there are *two* contours, each giving a characteristic phase to the wave function. As evident at a

glance at Fig. 5, the 2D problem has the complication that tunneling will occur simultaneously to more than one band. Fortunately, the energy contours at $F_F \pm \Delta$, which arise from gapped bands, are then allowed to come very close to either the *m* point or the *k* point. We shall show in the following paragraphs that these points dominate the value for the phase of the complex amplitude $C_{\mathbf{k},n}$.

C. CDW modulation amplitude near the *k* and *m* points

Although we shall look at the full **k** dependence of the CDW amplitude, it is instructive to first determine $C_{\mathbf{k},n}$ right at the *k* point. There the degeneracy is exactly threefold, and thus

$$\lambda_{\mathbf{k}} = \lambda_{\mathbf{k}-\mathbf{g}_0} = \lambda_{\mathbf{k}-\mathbf{g}_1} = \lambda,$$

and Eq. (25) then gives $\lambda^3 = 3\Delta_0^2 \lambda$, or the three distinct eigenvalues

$$\lambda_{\pm} = \pm\sqrt{3}\Delta_0,$$

and

$$\lambda_0 = 0.$$

At the *k* point, the theoretical value for the energy splitting between the m_0 and m_+ bands, i.e., across the Fermi level, is $\sqrt{3}\Delta_0$.

The complex coefficients are easily found using the results (26a) and (26b), and we have the following solutions for $C_{\mathbf{k},n}$, for each eigenvalue:

$$\lambda_+ = \sqrt{3}\Delta_0, \quad C_{\mathbf{k},n} = \tfrac{2}{3}\{e^{-i2\pi/3}, e^{i2\pi/3}, e^{-i2\pi/3}\}, \quad (27a)$$

$$\lambda_0 = 0, \quad C_{\mathbf{k},n} = \tfrac{2}{3}\{1,1,1\}, \quad (27b)$$

$$\lambda_- = -\sqrt{3}\Delta_0, \quad C_{\mathbf{k},n} = \tfrac{2}{3}\{e^{i2\pi/3}, e^{-i2\pi/3}, e^{i2\pi/3}\}. \quad (27c)$$

The CDW modulation has remarkably the same amplitude $|C_{\mathbf{k},n}| = \frac{2}{3}$ in each of the three directions, and we can write

$$|\psi_{\mathbf{k}}(\mathbf{x})|^2 = 1 + \tfrac{2}{3}\sum_{n=0}^{2} \cos(\mathbf{g}_0 \cdot \mathbf{x} - \theta_n), \quad (28)$$

where the phase angles, or $\arg(C_{\mathbf{k},n})$, take on the characteristic values

$$\lambda_+ = \sqrt{3}\Delta_0, \quad \theta_n^+ = \left\{-\frac{2\pi}{3}, +\frac{2\pi}{3}, -\frac{2\pi}{3}\right\}, \quad (29a)$$

$$\lambda_0 = 0, \quad \theta_n^0 = \{0,0,0\}, \quad (29b)$$

$$\lambda_- = -\sqrt{3}\Delta_0, \quad \theta_n^- = \left\{+\frac{2\pi}{3}, -\frac{2\pi}{3}, +\frac{2\pi}{3}\right\}. \quad (29c)$$

Consider the modulation $|\psi_{\mathbf{k}}|^2$ for the middle eigenvalue $\lambda_0 = 0$. It has maxima on points of a triangular lattice (of spacing $3a$) and sixfold rotational symmetry. Chen[28] refers to a similar function, $\Phi^6(k\mathbf{x})$, as a ''hexagonal cosine.'' Here the functions (28), having the three phases θ_n^+, θ_n^0, and θ_n^- are all hexagonal cosines, but cases (a) and (c) have shifted maxima with respect to a *fixed* unit cell: θ_n^+ and θ_n^- select the maxima at the positions $\frac{1}{3}$ and $\frac{2}{3}$ of the unit cell

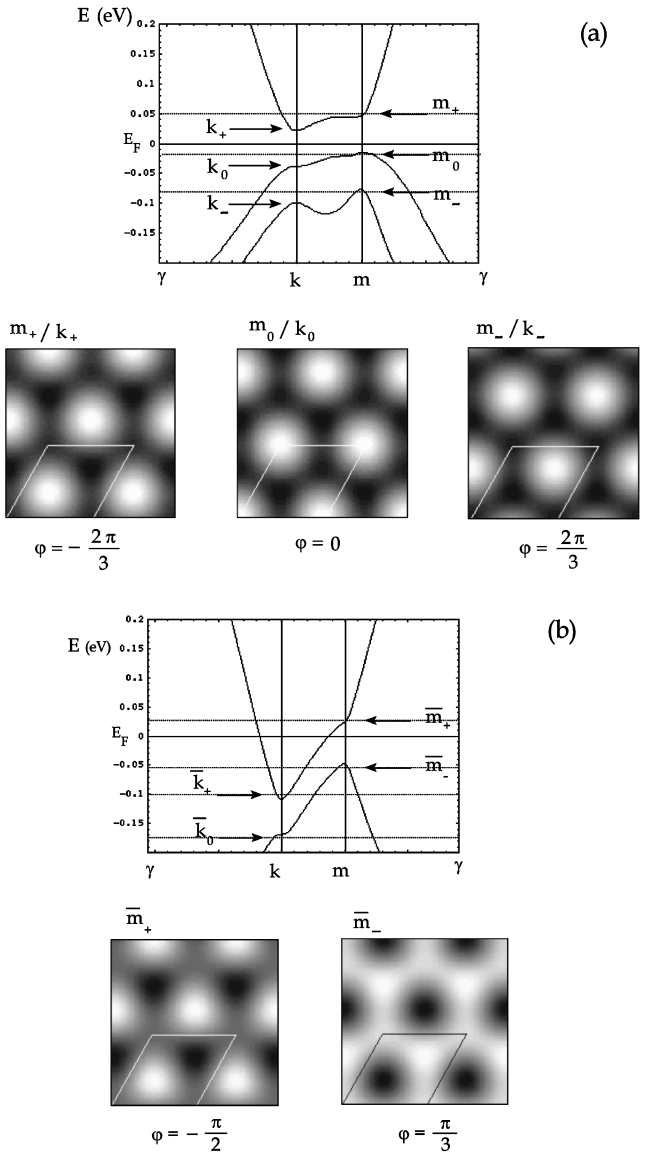

FIG. 7. LDOS, for a given band, at characteristic energies coinciding with points of high symmetry of the Brillouin zone. We separate the bands into two sets, (a) the bands arising from the k_2 point and (b) from the k_1 point, of Fig. 5. In (a), the lifting of threefold degeneracy gives three new states that are $2\pi/3$ out of phase. In (b) the motifs of the LDOS at the *m* point are quite different, showing threefold rotation symmetry, and the contrast is nearly inverted. The Fermi level lies in a gap in (a) but not in (b), which explains the reduction of the Fermi surface, and the spatial separation between occupied states and empty states.

diagonal, respectively [Fig. 7(a)]. We therefore expect to see an identical contrast change in the LDOS, if the energy is shifted from one eigenvalue to the next. The eigenvalues at the *m* point can be calculated in quite a similar fashion, and the splitting of the $\bar{m}$ bands gives two new motifs as those shown in Fig. 7(b). There, the occupied state at *m* gives a modulation that, in this case, is practically inverted with respect to the empty state at the same **k** point. In summary, the contrast changes described here result from the breaking of the initially symmetric state, into a set of new orthogonal states, such as Eqs. (27) above, due to the charge-lattice interaction.

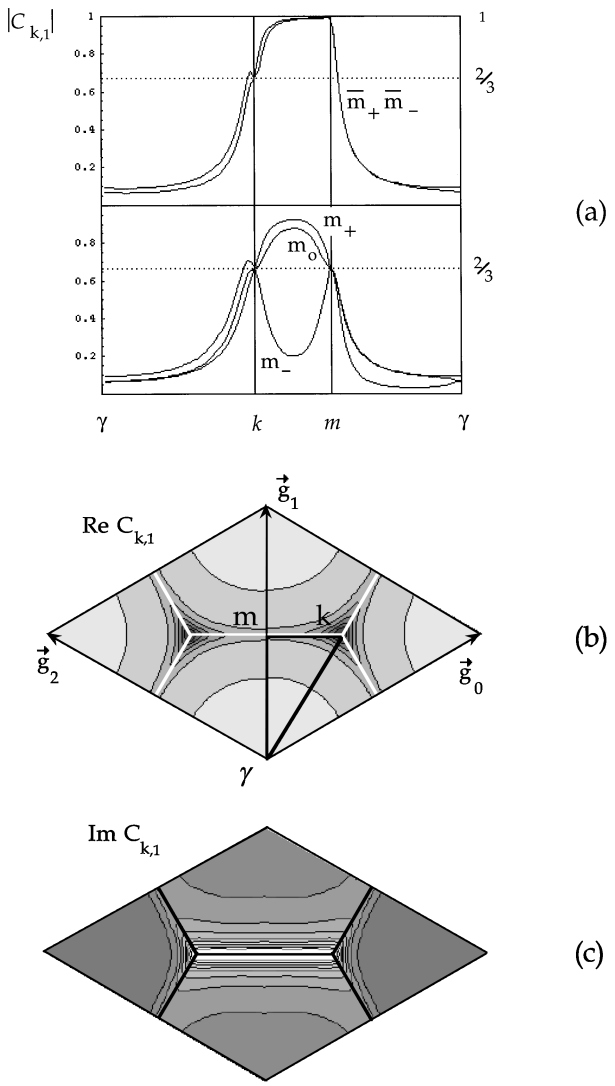

FIG. 8. The complex CDW coefficient $C_{\mathbf{k},n}$ as a function of $\mathbf{k}$ for the $\mathbf{g}_1$ direction only, and for the $\overline{m}_-$ band: In (a) $|C_{\mathbf{k},1}|$ is plotted along the high symmetry directions [dark line of (b)] for each of the five relevant bands. The characteristic values of 2/3 and 1 are due to the lifting of threefold and twofold degeneracies, respectively. The amplitude decays away from the zone boundary, as in the 1D problem (Fig. 1). In (b) and (c) the real and imaginary parts, which contribute to the even and odd parts of the LDOS, are shown as contour plots in the conventional cell. The real part is significant along the Bragg planes, while the imaginary part highlights the Bragg plane orthogonal to the $\mathbf{g}_1$ direction.

D. k dependence of the complex amplitude $C_{\mathbf{k},n}$

Recalling that the LDOS, by Eq. (21), involves the sum of $C_{\mathbf{k},n}$ over all $\mathbf{k}$ points on the constant energy contours, its functional dependence on $\mathbf{k}$ is relevant. This is particularly true for contours that have portions curving away from the new zone boundary, such as in Fig. 6. We also need to show that not all the $\mathbf{k}$ points of the Brillouin zone contribute to the amplitude and phase in an equal manner. If this were the case, then a direct physical interpretation would be quite difficult, due to the $\mathbf{k}$ summation.

In Fig. 8(a) we show typical plots of $|C_{\mathbf{k},n}|$ for each of the five dominant bands, and only for the $n=1$ case, i.e., corresponding to the $\mathbf{g}_1$ direction. By comparison to the 1D case, they have the expected property of being large when $\mathbf{k}$ is near or on the zone boundary. Note that all the curves pass through the value of $\frac{2}{3}$ at the k point and either $\frac{2}{3}$ or 1 at the m point. This is the two-dimensional analog of the standing wave condition: The former value ($\frac{2}{3}$) results from lifting the threefold degeneracy, leading to the three phases described above, while the latter resembles the case of twofold degeneracy, and leads to contrast inversion in the modulation.

The real and imaginary parts of $C_{\mathbf{k},n}$ (but only for the $\overline{m}_-$ band) are illustrated in Figs. 8(b) and 8(c) as contour plots. Recall from the preceding section that the real and imaginary parts of $C_{\mathbf{k},n}$ contribute, respectively, to the symmetric and antisymmetric parts of the real-space modulation. Only the case for $n=1$ is shown, i.e., for the $\mathbf{g}_1$ direction, the others are qualitatively similar (for nine bands, and three directions, there would be 27 such figures to display). The real part, $\mathrm{Re}(C_{\mathbf{k},n})$, has a characteristic star shape centered on the k point and decreasing in the radial direction, except along the Bragg planes. Im $C_{\mathbf{k},n}$ is particular in that it has a maximum value along the portion of the Bragg plane that is *orthogonal* to the $\mathbf{g}_n$ direction, and decreases rapidly elsewhere, Fig. 8(c). The $\mathbf{k}$ analysis of this model therefore provides a direct visualization of the process of Bragg reflection in a two-dimensional system: The $C_{\mathbf{k},n}$ plots tell us what $\mathbf{k}$ points are significantly contributing to both the amplitude and the phase of the CDW.

It would be tempting, *a priori*, to interpret the STM images based only on the gap splitting at the k or m points, together with symmetry arguments. However, a few aspects hinder this simple approach. First, the correct form of the LDOS, even at the characteristic energies, depends on the details of the relevant $\mathbf{k}$ integration. One must also take into account the competition arising from overlapping bands. These aspects, and the interpretation of the STM images, we leave to the final section.

IV. ENERGY AND VOLTAGE DEPENDENCE OF THE STM IMAGE

A. Energy dependence of the CDW component

As derived in Sec. II, the LDOS is determined as a function of energy by the summation over $C_{\mathbf{k},n}$ using Eq. (20). For example, for a given band (i.e., one among a total of nine bands in the gapped state), recall that the relevant complex sum is

$$a_n + ib_n = \sum_{\mathbf{k}} C_{\mathbf{k},n}\,\delta(E-E_{\mathbf{k}}), \tag{30}$$

where its modulus and argument give the amplitude $A_n(E)$ and phase $\varphi_n(E)$ of the CDW component of the LDOS, for the direction n.

Applying fairly standard numerical methods[29] for the LDOS of $NbSe_2$ in the CDW state, we find that the excess density has the same amplitude in all three directions, i.e.,

$$\delta\rho(\mathbf{x},E) = A(E)\sum_{n=0}^{2} \cos[\mathbf{g}_n\cdot\mathbf{x} - \varphi_n(E)], \tag{31}$$

and there is only a *single* independent angle to specify:

$$\varphi_n(E)=\{\varphi(E),-\varphi(E),\varphi(E)\}. \quad (32)$$

This property is actually to be expected given the symmetry of the charge-lattice potential, chosen at the outset. Particular values of $\varphi(E)$ are summarized in Fig. 7 for the characteristic energies where the bands cross the high-symmetry points, with the corresponding LDOS, as discussed previously. Here we prove that, for a given band, the k integration *conserves the phase* derived in the preceding section, i.e., by considering the phase of the CDW only in the vicinity of the k and m points. This result is somewhat surprising, given the explicit contours of Fig. 6, which have portions that move quite away from the zone boundary. This means that $|C_{\mathbf{k},n}|$ decays fast enough along a contour, away from the boundary points, such that the phase in the integral (30) is conserved. This is confirmed in Fig. 8(a), where we show $|C_{\mathbf{k},n}|$ decaying rapidly away from the k point, for example in the γk direction.

We therefore recover in the LDOS the five distinct motifs: three motifs involving a shift in the maxima of the CDW, and two a contrast inversion. These results are satisfactory since, given the position of the Fermi level, there is a distinct spatial separation between occupied and unoccupied states. A closer look at the LDOS, however, reveals that the states at the m point lead to a motif having threefold rotational symmetry, whereas those at the k point have exclusively sixfold symmetry.

As we have noted, these solutions for the LDOS at the characteristic points of Fig. 7 are insufficient to completely interpret the STM images. At a given energy, tunneling occurs to more than one band simultaneously. For example, at the gap edges, $E=E_F\pm\Delta$, there are two bands involved simultaneously, as shown explicitly by the constant E contours of Fig. 6. There is yet another contour, not shown in the figure, associated with the S_I portion of the Fermi surface. In our model it is essentially ungapped, and contributes little to the phase. Nevertheless it is included in all computations of the total density.

At a given energy E, we then write the excess LDOS as the explicit sum over the contributions arising from all bands:

$$\delta\rho(\mathbf{x},E)=\sum_{\lambda}\ \delta\rho_{\lambda}(\mathbf{x},E). \quad (33)$$

This multiple band situation, characteristic of a 2D system, is well illustrated in the plots in Fig. 9(a) of the density $\delta\rho_{\lambda}(\mathbf{x},E)$ arising from each band individually. Note that in this figure, we give the *relative* corrugation of the CDW to the background density, and Fig. 9(b) shows the corresponding phase. In each case, we label the important characteristic energies of Fig. 7, i.e., those energies that are at band extrema (solid lines) and saddle points (dotted lines). In the lower panel of Fig. 9(a) the only ''real'' gap occurs at the Fermi level, between the energies E_{k_+} and E_{m_0}. Within this gap, the CDW still contributes to the LDOS through the $\bar{m}_+$ band, and one expects to see the CDW in an STM image even for very small voltages. These properties are easily read from the gapped band structure of Fig. 5.

A number of points can be drawn from the $\delta\rho_{\lambda}(\mathbf{x},E)$ plots, first concerning the amplitude of the CDW component,

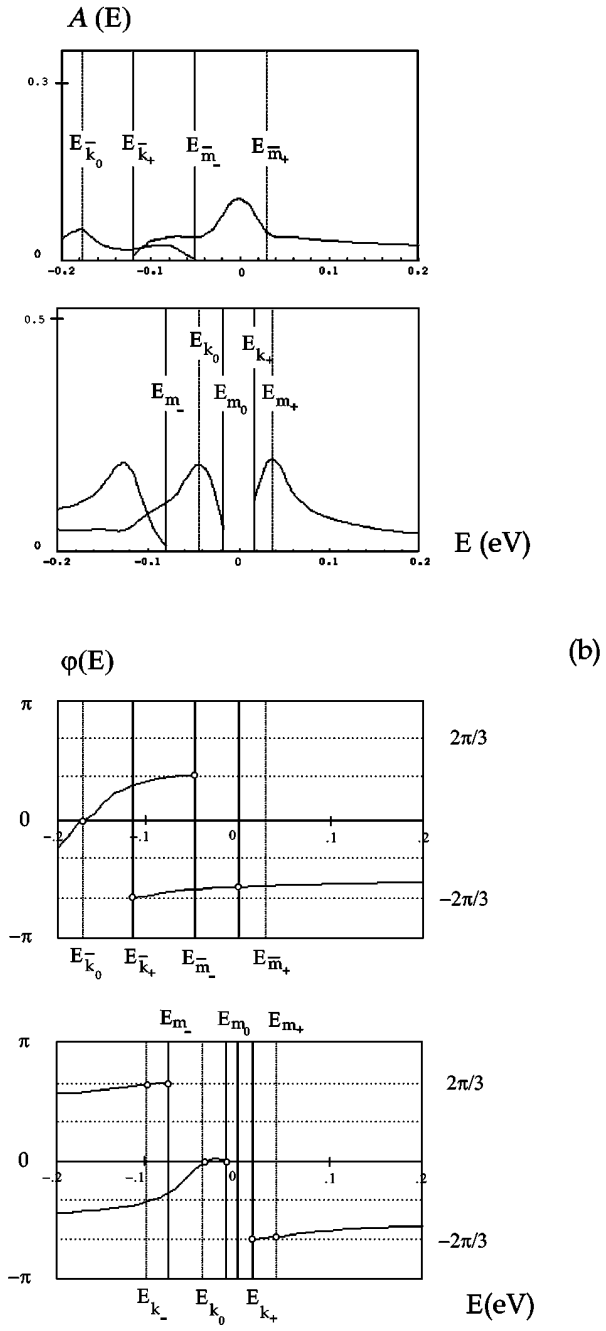

FIG. 9. Relative corrugation of the CDW density $\delta\rho_{\lambda}(\mathbf{x},E)$ as a function of energy for the five principal bands near the Fermi level: In (a) its relative amplitude $A(E)$ and in (b) its phase $\varphi(E)$. In each, the lower panel refers to the three bands of Fig. 7(a), and the upper panel refers to the two bands of Fig. 7(b). We indicate the characteristic energies corresponding to band extrema (solid lines) and saddle points (dotted lines). At these energies, special values for the phase are found: In the lower panel, there is a phase shift of $2\pi/3$, due to the lifting of threefold degeneracy at the k and m points. Notice the gap in $A(E)$ in the lower panel of (a), separated by maxima occurring at the energies E_{k_0} and E_{m_+}, which are not band edges, but saddle points. The maximum CDW amplitude is about 20%. The corresponding phase difference shown in (b) is $2\pi/3$, which shifts the maxima in the LDOS. Within the gap, the residual amplitude is about 10%, thus for small voltage one expects to still see the modulation in the STM image.

and then the phase. As Fig. 9(a) shows, the overall contribution of the CDW is quite small, for the 200 meV range given, and reaches a maximum value of about 20% of the total DOS. Even at the energies ± 200 meV, the CDW accounts for 5% of the DOS, which may be surprising, but is understandable since, even for these energies, the constant energy contours still come sufficiently close to the zone boundary.

The most important feature of the $\delta\rho_\lambda(\mathbf{x},E)$ plots is that the maxima in the CDW amplitude do not occur at the band-edge energies, but rather occur at the saddle points, at E_{k_0} and E_{m_0}. In the 1D problem it is the *band extrema* that give the largest amplitude, and a singular density of states at the gap edges (as in Fig. 1). On the contrary, for the 2D problem, the group velocity vanishes at the saddle points, and these give rise to a logarithmic singularity in the DOS. Looking again at these energies in the gapped band structure of Fig. 5, we notice that $E_F-\Delta$ lies very close to the saddle point of the m_0 band, at k_0. Indeed, this is confirmed by the constant energy contour in Fig. 6, as indicated by the arrow. In a similar way, the energy $E_F+\Delta$ is close to the saddle point at m_+ (or at $\overline{m}_+$). In view of this coincidence between Δ and the characteristic energies described above, we speculate that the broadened peaks in the measured DOS, i.e., using dI/dV spectroscopy, are due to the saddle points and not to the band extrema. Indeed, the DOS calculated from our CDW band structure closely resembles the experimental spectra of Refs. 10 and 30.

The motif of the LDOS depends on the phase angle $\varphi(E)$, as discussed above. In Fig. 9(b), we plot the CDW phase for the five bands that are near the Fermi level. Again we highlight the band extrema and the saddle points at which characteristic phase values occur: In the lower panel, there is a phase shift of $2\pi/3$ in the density for the three band edges. This is in agreement with the lifting of threefold degeneracy, as derived previously, and the corresponding LDOS is shown in Fig. 7(a). The upper panel shows the phase as a function of energy for the m gap, i.e., between the points $\overline{m}_+$ and $\overline{m}_-$. We find an abrupt change in $\varphi(E)$, from $\varphi=\pi/3$ for the occupied states to $\varphi=-\pi/2$ for the empty states. The gray scale plots of Fig. 7(b) show the apparent contrast inversion between these two densities. Here we emphasize that the LDOS shows, in this case, threefold rotational symmetry, and not sixfold. Put another way, the motif contains both white spots and dark spots, which is a general feature of the STM images.[16,30]

These remarks then allow us to understand the possible changes in the motif observed in the LDOS at the particular energies corresponding to the gap edges, i.e., at $E=E_F\pm\Delta$. For instance, at $E_F-\Delta$, the LDOS is dominated by the phase at k_0 since, as Fig. 9(a) shows, the CDW amplitude is largest there owing to the saddle point at E_{k_0}. The corresponding constant energy contour (at $E=E_F-\Delta$) of Fig. 6(c) shows the same result: the contour comes very close to the k_0 point, as indicated by the arrow. We then conclude from Fig. 9(b) that the net phase is $\varphi(E_F-\Delta)\approx 0$. Reasoning in a similar fashion for the unoccupied states at the energy $E_F+\Delta$, the phase is dominated by states at m_+ with some small contribution from k_+, and we find $\varphi(E_F+\Delta)\approx -2\pi/3$. We therefore have the result that the overall phase change across the gap is $\Delta\varphi\approx -2\pi/3$. The maxima, or white spots, in the CDW should shift by $\frac{1}{3}$ along the unit cell diagonal. These arguments show that it is not at all trivial to extract the information on the CDW motif for a two-dimensional system, where bands are overlapping.

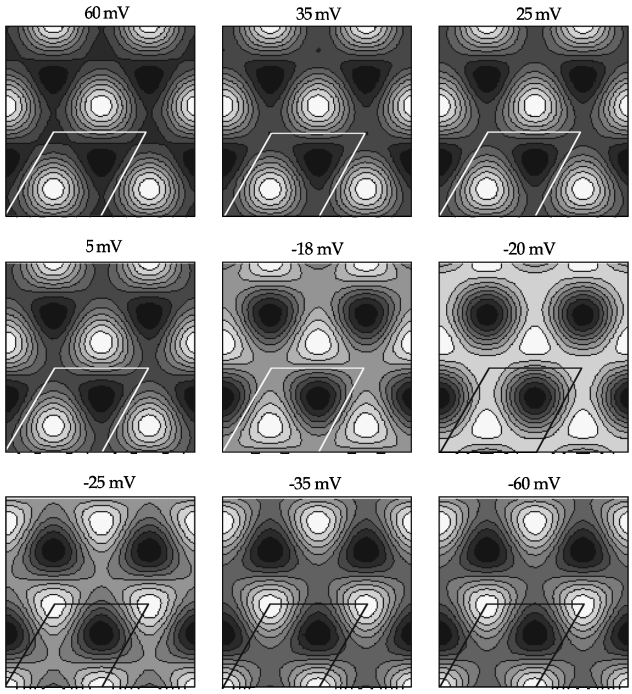

FIG. 10. STM image simulations for a sequence of voltages ranging from -60 to 60 mV showing only the CDW component. The unit cell of lattice constant $3a$ is indicated. Thus the image appears as a hexagonal array of alternating dark spots (minima) and white spots (maxima). The motif changes gradually: at -60 mV the maxima are at the corners of the unit cell, and at $+60$ mV the maxima are shifted to $\frac{1}{3}$ the unit cell diagonal. In all images, the dark spots are located at the same sites, but change intensity. At the particular voltage -20 mV, within the gap energy, the motif is very nearly a hexagonal array of holes.

B. Voltage dependence of the images

In order to compare with the STM images, we use the simplest expression for the current, which neglects the electric field in the barrier, and the DOS of the tip. For the small voltage range considered here, the simple expression for the zero temperature current,

$$I(\mathbf{x},V)\propto\int_{E_F}^{E_F+eV}\rho(\mathbf{x},E)dE, \tag{34}$$

should suffice. There are two remarks concerning expression (34) to be made here. First, that experiments performed at the temperature of liquid He, 4.2 K, will avoid the problem of thermal smearing of the electronic effects we are interested in. Second, the variation of the LDOS within the tunneling energy window implies that a current image, $I(\mathbf{x},V)$, would differ from a conductance image, or $dI(\mathbf{x},V)/dV$. Therefore, in view of the energy integration in expression (34), one cannot make a simple correspondence between energy and voltage: In the former case, the CDW amplitude, and phase, for the whole energy range (E_F to E_F+eV) will

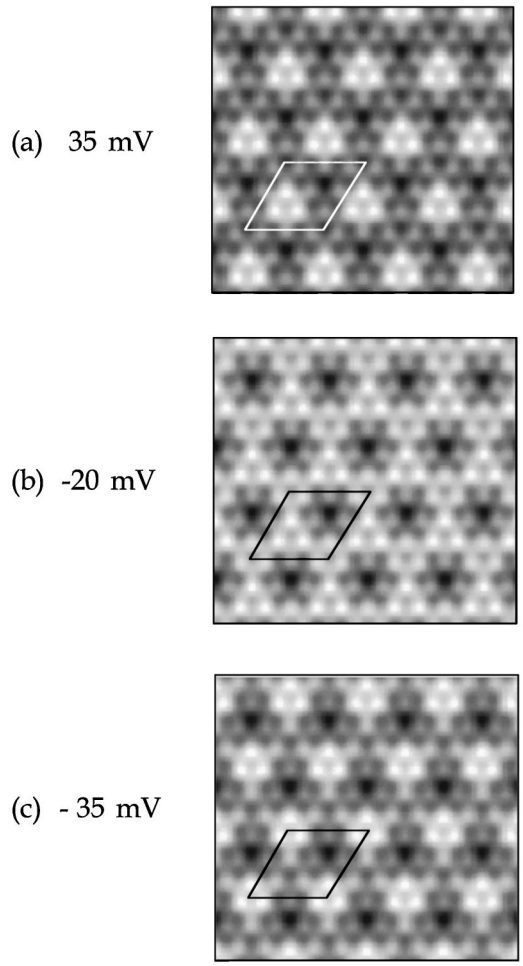

FIG. 11. STM simulation showing the CDW together with the atomic corrugation for three selected voltages. In images (a) and (c), the tunneling window encompasses the gap: $V=\pm\Delta/e$ and the relative shift in the maxima of the CDW component is seen. For a voltage within the gap, the STM simulation can either appear similar to (a) or (c), or even as an array of dark spots, as in (b). Due to the superposition of the atomic and CDW corrugation, the distinctive moiré patterns are found.

contribute, whereas in the latter case, the states at eV will dominate. Still, the general pattern of the LDOS, as previously outlined, are confirmed by the STM simulations as a function of voltage.

In the current image, $I(\mathbf{x},V)$, the superposition of the CDW signature over the energy interval leads to a continuous variation of the motif as a function of the bias. The sequence of STM simulations of Fig. 10, with the voltage ranging from -60 to 60 mV, illustrates the point. The major overall effect follows the previous conclusion concerning the LDOS, i.e., a phase change of $\Delta\varphi\approx-2\pi/3$. Thus the white spots shift position by moving $\frac{1}{3}$ of the diagonal of the unit cell. Although the motif gradually changes, the position of the dark spots (minima of the CDW) remains unchanged, i.e., at $\frac{2}{3}$ the diagonal of the unit cell. Notice the particular case at $V=-20$ mV, where the maxima become secondary, giving a hexagonal array of holes. The phase angle in this case is $\varphi\approx\pi$. These results correspond surprisingly well with the experimental STM images we have previously obtained, but the hexagonal array of dark spots was observed for a smaller bias ($V=-10$ mV).

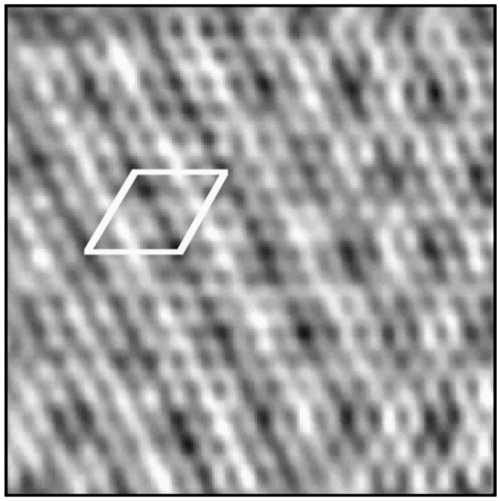

FIG. 12. A selected STM 30×30 Å^2 image of $2H$-$NbSe_2$ taken at 4.2 K ($R=100$ MΩ, $V=-10$ mV) to be compared to the simulation Fig. 11(b). Despite a small difference in bias voltage, the similarity is striking. The image has been filtered to remove noise.

Adding to the CDW component of the LDOS the atomic corrugation of the $NbSe_2$ lattice, we obtain the distinctive moiré patterns shown in Fig. 11. In (a) and (c) we give the STM images that could be seen for the voltage set at the gap edges $V=\pm\Delta/e$, while in (b) we selected the interesting case ($V=-20$ mV) where the CDW is a hexagonal array of holes. In the latter case, the moiré pattern gives rise to a distinctive motif consisting of ''flowers'' of seven dark spots (one central dark spot, surrounded by six others). In (a) and (c) the motif is dominated by three equally intense white spots. A closer look at these simulations reveals that, while case (b) is very nearly sixfold symmetric, cases (a) and (c) have threefold rotational symmetry. In Fig. 12 we give the real STM image,[30] taken at $V=-10$ mV, to be compared with Fig. 11(b).

V. CONCLUSION

We have presented a simple theory to explain the voltage or energy dependence of the STM images of a CDW in two dimensions. Our main attention was focused on two physical parameters, its amplitude and phase. Motivated by the success in the domain of semiconductors, to relate the spectroscopic content of STM imaging to the electronic structure, we have recently taken a closer look at the voltage dependence of the CDW in $2H$-$NbSe_2$, using low-temperature STM. The main question is whether or not there is indeed a change in contrast for energies within or larger than the CDW gap energy Δ. Using a simplified analysis of the band gapping in the case of $NbSe_2$, we showed that new states arise on the order of Δ, contributing to the STM image at selected bias voltages.

A study of the gapped band structure, and the related constant energy contours, revealed that for characteristic energies near the Fermi level, the LDOS involves the points of high symmetry of the new Brillouin zone. For these energies, we found special values for the phase and the amplitude of

the CDW modulation. Specifically, due to the lifting of a threefold degeneracy at the k point of the new Brillouin zone, we showed that the phase of the CDW component of the LDOS changes by $2\pi/3$ at each eigenvalue. A more careful analysis of the tunneling to the different bands, at the gap energies $E_F \pm \Delta$, gave the same result: the maxima of the LDOS thus shift a distance of $\frac{1}{3}$ the diagonal of the unit cell.

A second aspect of the problem was investigated, namely the energy dependence of the CDW amplitude arising from each band. In this analysis, we can conclude that the CDW should be seen in the STM image, even for very small voltage, $V \ll \Delta/e$. Indeed at the Fermi level, the CDW component accounts for approximately 10% of the total DOS. Moreover, we found that the maximum of the amplitude occurs at energies corresponding to the saddle points of the band structure, and not the band edges. It should be of some interest to investigate both the images and spectra of other CDW systems to compare to the results we have given here for $NbSe_2$.

Although our approach may be considered somewhat naïve, the goal is to understand, in a general way, how the physical parameters of the STM measurement relate to the band structure, without *ab initio* calculations. On the other hand, it is true that a number of aspects of the problem we have simplified may have some important effects: The perpendicular dispersion of the bands, the Nb-Se coupling (and thus the exact position of the Fermi level), the details of the charge-lattice interaction, and the nesting question. These problems existed prior to the STM, and further work is necessary to refine these details.

There have been few investigations of the energy or voltage dependence of the CDW amplitude or phase. A quite detailed study of the dI/dV spectra, as well as the distance dependence of the corrugation, has been done in Ref. 10. In this area one could also investigate the *total* charge amplitude, i.e., the amplitude $A(E)$ of our work, but integrated over all occupied states. This perhaps could be achieved by measuring the CDW amplitude for a sufficiently large negative bias. The 1D chain model gives the well-known total amplitude A as a function of the gap and the electron parameters:[19]

$$A = \frac{2\Delta}{\pi \hbar v_F} \ln\left(\frac{2E_F}{\Delta}\right).$$

It would be of interest to investigate whether the STM image could verify this result, or an analogous one for the two-dimensional system.

[1] R. Becker, J. Golovchenko, D. Hamann, and B. Schwartzentruber, Phys. Rev. Lett. **55**, 2032 (1985).

[2] R. Hamers, R. Tromp, and J. Demuth, Phys. Rev. Lett. **56**, 1972 (1986).

[3] R. Tromp, R. Hamers, and J. Demuth, Phys. Rev. B **34**, 1388 (1986).

[4] R. Feenstra, J. Stroscio, J. Tersoff, and A. Fein, Phys. Rev. Lett. **58**, 1192 (1987).

[5] R. Feenstra, J. Stroscio, and A. Fein, Surf. Sci. **181**, 295 (1987).

[6] R. Feenstra, in *Scanning Tunneling Microscopy and Related Methods*, edited by R. Behm, N. Garcia, and H. Rohrer (Kluwer, Dordrecht, 1990).

[7] R. Feenstra, J. Stroscio, and A. Fein, Phys. Rev. Lett. **58**, 1668 (1987).

[8] Ph. Avouris and R. Wolkow, Phys. Rev. B **39**, 5091 (1989).

[9] R. Coleman, B. Drake, P. K. Hansma, and G. Slough, Phys. Rev. Lett. **55**, 394 (1985).

[10] B. Giambattista, A. Johnson, W. McNairy, C. G. Slough, and R. Coleman, Phys. Rev. B **38**, 3545 (1988); C. Wang, B. Giambattista, C. Slough, R. Coleman, and M. Subramanian, *ibid.* **42**, 8890 (1990).

[11] R. Thomson, U. Walter, E. Ganz, J. Clarke, A. Zettl, P. Rauch, and F. DiSalvo, Phys. Rev. B **38**, 10 734 (1988).

[12] M. Whangbo, J. Ren, S. Magonov, H. Bengel, B. Parkinson, and A. Suna, Surf. Sci. **326**, 311 (1995).

[13] N. D. Lang, Phys. Rev. B **34**, 5947 (1986).

[14] J. Tersoff and D. R. Hamann, Phys. Rev. Lett. **50**, 1998 (1983); Phys. Rev. B **31**, 805 (1985).

[15] J. Tersoff, Phys. Rev. Lett. **57**, 440 (1986).

[16] P. Mallet, W. Sacks, D. Roditchev, D. Défourneau, and J. Klein, J. Vac. Sci. Technol. B **14**, 1070 (1996).

[17] G. Wexler and W. Woolley, J. Phys. C **9**, 1185 (1976).

[18] N. Doran and A. Woolley, J. Phys. C **14**, 4257 (1981).

[19] G. Gruner and A. Zettl, Phys. Rep. **119**, 117 (1985).

[20] J. P. Pouget, in *Low-Dimensional Electronic Properties of Molybdenum Bronzes and Oxides*, edited by C. Schlenker (Kluwer Academic, Dordrecht, 1989).

[21] J. A. Wilson, Phys. Rev. B **15**, 5748 (1977).

[22] A. Kikuchi and M. Tsukada, Surf. Sci. **326**, 195 (1995).

[23] M. Whangbo, D. Seo, and E. Canadell, in *Physics and Chemistry of Low-Dimensional Inorganic Conductors*, edited by C. Schlenker *et al.* (Plenum Press, New York, 1996).

[24] M. Whangbo, J. Ren, S. Magonov, and H. Bengel, in *Physics and Chemistry of Low-Dimensional Inorganic Conductors*, edited by C. Schlenker *et al.* (Plenum Press, New York, 1996).

[25] J. Tersoff, Phys. Rev. B **39**, 1052 (1989).

[26] J. Carpinelli, H. Weitering, E. W. Plummer, and R. Stumpg, Nature (London) **381**, 398 (1996).

[27] D. Moncton, J. Axe, and F. Disalvo, Phys. Rev. Lett. **34**, 734 (1975).

[28] C. J. Chen, Phys. Rev. Lett. **69**, 1656 (1992).

[29] We have used two methods to evaluate the **k** summation. In addition to standard numerical integration, we also use a second and more direct approach, which consists of dividing up the Brillouin zone into a mesh of a suitably large number of points. If $\{E_{\mathbf{k}}\}$ is the set of eigenvalues evaluated on the mesh points, then, up to a numerical constant, the DOS as a function of energy is equivalent to the histogram of E. The CDW amplitude is then, in view of Eq. (30), a weighted histogram. While being simple to implement, it is admittedly rather slow to converge when the band becomes flat, for example near saddle points. The histogram method is particularly useful when the amplitudes $C_{\mathbf{k},n}$ are displayed as a 2D plot, such as in Fig. 8, along with a given constant energy contour $l(\mathbf{k})$, providing a visualization of the LDOS computation.

[30] Pierre Mallet, doctoral thesis, University of Paris 7 (1996).