Ab initio molecular-dynamics techniques extended to large-length-scale systems

T. A. Arias
Physics Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

M. C. Payne
Cavendish Laboratory, Madingley Road, Cambridge, CB30HE, United Kingdom

J. D. Joannopoulos
Physics Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
(Received 27 February 1991; revised manuscript received 8 July 1991)

The Born-Oppenheimer approximation divides the problem of quantum molecular dynamics into two familiar problems: (1) solution for the electronic wave functions for a given *instantaneous* arrangement of ions and (2) the motion of the atomic cores under the influence of those wave functions. A combination of conjugate-gradient methods to solve (1) with standard molecular dynamics to solve (2) results in a scheme that is at least two orders of magnitude more accurate than previously possible, thus allowing accurate calculation of dynamic correlation functions while maintaining tolerable energy conservation for microcanonical averages of those correlation functions over picosecond time scales. By employing conjugate-gradient techniques, this method is used to extend the applicability of finite-temperature *ab initio* techniques to systems with large length scales.

I. INTRODUCTION

Car and Parrinello's introduction of their approach to *ab initio* total-energy calculations made tractable *ab initio* molecular dynamics of medium-sized quantum systems ($\sim$10–100 atoms). The basic idea in this method is to perform molecular dynamics on the combined system of atomic coordinates *and electronic wave-function coefficients,* where the forces on the atoms are calculated using the Hellmann-Feynman theorem, and the forces on the electronic coefficients are taken to be the derivatives of the electronic energy functional with respect to these coefficients. Generally, the wave-function coefficients are all assigned the same fictitious mass, judicious choice of which (typically 200–300 a.u.) limits the transfer of energy from the atomic to the electronic degrees of freedom and keeps the electronic system in the vicinity of the Born-Oppenheimer (BO) surface by approximate adiabatic isolation of the atomic and electronic subsystems.[1–5] BO tolerances reported in the literature for nontrivial semiconducting systems, those with more than one *active* ionic degree of freedom, are on the order of a few tenths of a meV/ion,[1] while those for gapless systems tend to be much worse, $\sim$10 meV/ion,[5] for reasons discussed below.

A major contribution to this BO deviation in semiconductors is in the form of periodic energy transfers between the ions and electrons. These transfers involve several tenths of a meV/ion and have frequencies in the range $(E_g/\mu)^{1/2}$ on up to $(E_{\rm cut}/\mu)^{1/2}$ (and even higher in systems with long length scales[6]). Here, E_g is the gap, $E_{\rm cut}$ is the electronic plane-wave cutoff energy, μ is the fictitious electron mass, and all physical quantities are expressed in atomic units. In a typical semiconductor calculation the lowest of these frequencies is a few times the optic phonon frequency; in a metallic situation it is zero, indicating a breakdown in the scheme for this class of systems with consequences outlined below. Unfortunately, in the combined molecular-dynamics scheme it is impractical to limit the fictitious kinetic energy of the electrons to less than some fraction[7] of the energy in the ionic system. This excess energy is needed as a store to be converted back and forth into fictitious kinetic energy of the electronic coordinates to allow them to adjust continuously as the ions move. Because the total fictitious Hamiltonian is conserved, the periodic transfer of energy to and from the electronic system results in a periodic slowing and speeding of the ionic system and thus adds a fictitious vibrational motion to the ions which *tends* to average out over time and keep the atomic and electronic subsystems adiabatically isolated, as long as the energy-transfer time scale is much quicker than the time scale for the atomic motion. One can increase the frequency of the fictitious modes and weaken this coupling by decreasing the fictitious electronic mass, but then one must decrease the time step for integrating the equations of motion, and the computational costs increase. Also, there is no reason why in principle such superposed fictitious vibrations cannot affect the calculations of thermal properties.

The thermal-expansion coefficient of Si (Ref. 4) and proton diffusion constant in Si (Ref. 5) have both been calculated using this technique with fair quantitative agreement with experiment. However, in these calculations it is unclear how much of the discrepancy is due to effects from the fictitious dynamics or from the flaws pointed out by the authors of these works in the physical models, such as neglecting the quantum nature of the Si lattice in the case of the thermal expansion[4] coefficient or not considering defects and possible molecular H_2 formation in the case of proton diffusion,[5] or simply due to

inaccuracies in the local-density approximation itself. It is important in such situations to be able to sort these issues out, which can only be accomplished with a scheme giving direct control over the sources of calculational error. Finally, the extent of the effects of errors introduced by the fictitious dynamics has so far only been determined with *a posteriori* comparisons with experiment, limiting the predictive *ab initio* nature of the quantum molecular-dynamics technique.

In addition to fictitious vibrational effects, the electrons drift away from the BO surface as a result of an overall average energy transfer from the (much) higher-temperature ionic subsystem to the lower-temperature electronic one. This is especially true in calculations involving gapless systems, which include metals and molten semiconductors. As mentioned above, in these systems the fictitious electronic vibrational spectrum goes soft, the adiabatic separation is lost, and the forces experienced by the ions are no longer correct even in an average sense. Energy then flows freely from the atomic to the electronic subsystem. One must then "quench" the electronic system, hold the ionic coordinates fixed, and relax the electronic coordinates back to the BO surface, as often as 15 times per picosecond (about once every optic phonon period) of simulation time.[8] The excess energy removed from the electronic system in this process ultimately has its origin in the ionic system, which must then be compensated for its loss by pumping energy back into it. This is accomplished in various gross ways, but it is clear that without detailed knowledge of which ionic modes are being depleted of energy, one cannot expect such a procedure to *robustly* reproduce the dynamic properties of the system. Under these circumstances it is safer not to consider dynamic correlations in the ionic system but to regard the "trajectories" that one has generated as just a large number of configurations at a particular total energy and content oneself with calculating ensemble averages rather than time-dependent correlation functions.

Finally, it has been pointed out that as the longest length scale in the supercell increases, integration of the fictitious equations of motion becomes dominated by the $1/G^2$ instability (charge sloshing) in the Hartree energy.[6] Here, G refers to the shortest nonzero reciprocal-lattice vector of the supercell. To overcome this instability, the time step for integration of the equations of motion must be taken to approach zero as the system size increases. While the onset of this instability probably depends on the physics of the systems under study, the effect has been observed *empirically* at 8 Ry in a silicon system of six primitive cells (~ 32 Å long).[6]

The conjugate-gradient methods of Teter, Payne, and Allan,[6] and Gillan[9] for relaxing the electronic wave functions, however, are free of the aforementioned instability and are known for fixed atomic arrangements to converge quicker than the best electronic molecular-dynamics schemes by factors of 10 or more in terms of actual computer CPU time. By combining this scheme for the electrons, standard molecular dynamics for the atoms, and a scheme for producing good trial wave functions for each atomic arrangement, we have produced an *ab initio* molecular-dynamics scheme that is free of all the above difficulties. Without having to adjust a fictitious electronic mass, perform periodic quenches, or rely on fictitious vibrations to average out errors, one can *directly* control the BO error and, for about the same CPU time investment as traditional combined molecular-dynamics schemes, keep it at the 10^{-6}-eV/atom level. Such precision allows accurate calculation of dynamic correlation functions averaged over a microcanonical distribution over picosecond time scales. Moreover, because this method does not require an *ad hoc* scheme for pumping energy back into the ionic system, it maintains the desirable *ab initio* nature of the approach.

The basic outline of our scheme is as follows. From the electronic wave functions of preceding time steps, we construct trial wave functions for the current atomic arrangement according to the scheme described in Sec. II of this paper. We then use the complete conjugate-gradient (CCG) procedure of Teter, Payne, and Allan to relax the electrons to within a set tolerance of the Born-Oppenheimer surface. From these wave functions, we derive the forces experienced by the atoms and then move them according to some standard molecular-dynamics scheme such as the Verlet algorithm. The process then repeats. In Sec. III we present the results we obtain with this scheme, and in Sec. IV we present some concluding remarks.

II. CONSTRUCTION OF TRIAL WAVE FUNCTIONS

Taking our cue from the molecular-dynamics scheme of Car and Parrinello, we start by considering the simplest possible wave function guess. Let $\Psi_{n,k}(\{r\})$ be the wave function for band n and k point when the ionic coordinates are $\{r\}$, and let $\Psi_{n,k}(\{r+dr\})$ be the wave functions when the ions are at $\{r+dr\}$. Then, a first attempt at a trial wave function for the configuration $\{r+2\,dr\}$ might be the simple linear extrapolation

$$\Psi'_{n,k}(\{r+2\,dr\}) \equiv \Psi_{n,k}(\{r+dr\}) + [\Psi_{n,k}(\{r+dr\}) - \Psi_{n,k}(\{r\})] , \tag{2.1}$$

where the prime indicates a trial wave function as opposed to the fully relaxed version, which we call $\Psi_{n,k}(\{r+2\,dr\})$. As long as $\Psi_{n,k}(\{r\})$ is a continuously differentiable function of $\{r\}$, the error in $\Psi_{n,k}(\{r+2\,dr\})$ is clearly $O((dr)^2)$. Because the total-energy functional $E[\Psi;\{r\}]$ is stationary about $\Psi(\{r+2\,dr\})$ when $r=r+2\,dr$, the BO error of the trial wave function will be $O((dr)^4)$.[10]

There are several difficulties that must be overcome before one can employ (2.1). The first is minor and easily solved. It is that (2.1) results in wave functions that are not properly orthonormalized. However, because

$$\langle \Psi'_{n,k}(\{r+2\,dr\}) | \Psi'_{m,k}(\{r+2\,dr\}) \rangle = \delta_{n,m} + O((dr)^2) , \tag{2.2}$$

one can simply carry out the Gram-Schmidt orthonor-

malization on them without disturbing their correctness to first order. In what follows, it is understood that this orthonormalization procedure is always applied as the last step in constructing trial wave functions.

The second difficulty is more serious. In practice $\Psi_{n,k}(\{r\})$ is frequently a discontinuous function of $\{r\}$. This is due mostly to the various indeterminacies in defining the $\Psi_{n,k}(\{r\})$. Specifically, one may choose any ordering for the state label n, one may multiply each state by an arbitrary phase factor, and one may form arbitrary unitary transformations in degenerate subspaces. If one chooses the label n according to the ordering of eigenenergies, then as two bands cross, their wave functions will transform discontinuously into each other, making linear extrapolation ineffective. If one extracts the Kohn-Sham eigenstates by subspace diagonalization, then the phases ascribed to the wave functions will vary seemingly randomly according to the particular matrix diagonalization algorithm employed, again introducing discontinuities in the $\Psi_{n,k}$. Finally, if the motion from $\{r\}$ to $\{r+dr\}$ splits a degeneracy, and unless the original degenerate states were mixed according to the proper irreducible representation, the states will change discontinuously when the new crystal field resolves them. As long as one does not start the simulation at a $\{r\}$ of high symmetry so that there are no degeneracies, all of these problems can be avoided by simply not subspace diagonalizing because this is the only step in the CCG method where the state ordering, mixtures in degenerate subspaces, or phases are changed nonadiabatically. In fact, all of the results in Sec. III were obtained in this way.

However, one cannot always forego subspace diagonalization. In some cases it is needed to obtain good convergence rates in the CCG scheme, and subspace diagonalization is currently the only way in the CCG scheme to obtain the Kohn-Sham eigenstates, which are requisite for computation in a metallic system. In cases requiring subspace diagonalization, the arbitrariness in the $\Psi_{n,k}(\{r\})$ must be removed in a way that makes the changes in the wave functions as small as possible. All of the above arbitrariness in the $\Psi_{n,k}$ correspond to transformations on the $\Psi_{n,k}$ which leave the charge density and total energy invariant. We propose to minimize the differences between the wave functions at two successive time steps by employing the most general such transformation, arbitrary unitary linear combination among all of the wave functions at a given time step. We have found (see Appendix A) that one can, for about the same effort as subspace diagonalization, literally minimize the differences between the two sets of wave functions

$$S \equiv \sum_{n,k} \|\Psi_{n,k}(\{r+dr\}) - \Psi_{n,k}(\{r\})\|^2 \tag{2.3}$$

under such general unitary combination. This procedure actually factors the rotation in Hilbert space which carries the subspace spanned by the $\Psi_{n,k}(\{r\})$ to the subspace spanned at $r+dr$ into a series of simple rotations in orthogonal two-dimensional subspaces, one for each band. Because the optimal S in (2.3) scales like dr, (2.1) may again be trusted to yield trial states correct to first order in r. In all of the variations we describe below, this procedure should be done before the linear extrapolations if one is subspace diagonalizing.

All of the sources of discontinuity in $\Psi_{n,k}(\{r\})$ that we have discussed so far are not physical, in the sense that they do not affect physical quantities such as the total charge density. The only source of *physical* discontinuity in $\Psi(\{r\})$ is Kohn-Sham eigenstates crossing the Fermi level because under these circumstances the charge density is affected. In a system with many bands, the effect of bands crossing the Fermi level on the single-particle Hamiltonian will be rather small (on the order of 1 over the total number of bands), and thus the effect on the Kohn-Sham eigenstates will also be small. The main difficulties in these cases then come from the fact that the states which have just crossed the Fermi level do not resemble the states which they have replaced. These $\Psi_{n,k}(\{r\})$ will change discontinuously. Subspace alignment also provides a prescription for solving this problem. By including a few empty states in the system above the Fermi level and performing the above described subspace alignment on both the filled and empty states, alignment allows continuous extrapolation between filled and empty states. Though there still may be discontinuities as bands cross into the included empty states from yet further above, subspace alignment, as described in Appendix A, elegantly isolates the resultant large changes in state vectors to the uppermost empty states. These states contribute to neither the total energy nor the self-consistent single-particle Hamiltonian, and as long as they are brought back to convergence before they can cross the Fermi level, the simulation will not be adversely affected. Insuring this is quite cheap in terms of computational effort because in the CCG scheme empty bands require far less computation, as the self-consistent single-particle Hamiltonian does not have to be updated. Finally, an *a priori* concern introduced by the introduction of insurance states into the subspace alignment scheme is that the alignment might scramble states from above and below the Fermi level to the extent that it becomes impossible to define an approximate Fermi level for the extrapolated wave functions. However, as shown in Appendix A, alignment does not strongly mix Kohn-Sham eigenstates whose energy difference exceeds the typical eigenenergy drift between successive time steps. Thus, tracking which states are to be filled or empty is in no way complicated by the alignment procedure, and hence this extrapolation procedure may be applied to both semiconducting and *metallic* systems.

The third difficulty in using (2.1) is that this would generate acceptable $[\delta E = O((dr)^4)]$ trial wave functions for the ionic coordinates $\{r(t_n)+[r(t_n)-r(t_{n-1})]\}$ rather than the correct $\{r(t_{n+1})\}$. To surmount this difficulty, we generalize the scheme as follows. First, we take

$$\begin{aligned}\Psi'_{n,k}(\{r(t_{n+1})\}) &\equiv \Psi_{n,k}(\{r(t_n)\}) \\ &\quad +\alpha[\Psi_{n,k}(\{r(t_n)\}) - \Psi_{n,k}(\{r(t_{n-1})\})] \\ &\quad +\beta[\Psi_{n,k}(\{r(t_{n-1})\}) - \Psi_{n,k}(\{r(t_{n-2})\})] ,\end{aligned} \tag{2.4}$$

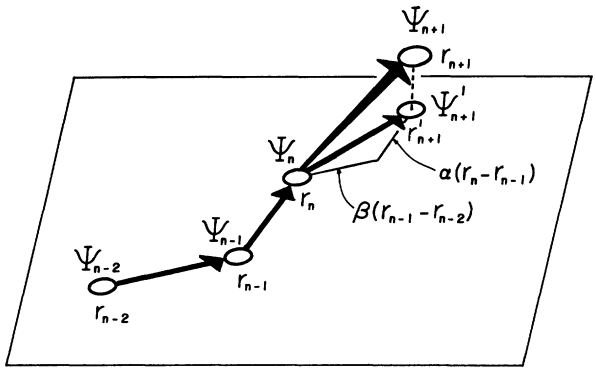

FIG. 1. Schematic illustration of the two-dimensional extrapolation scheme for trial wave functions. Taking linear combinations of Ψ_{n-2}, Ψ_{n-1}, and Ψ_n, we can only form trial wave functions Ψ'_{n+1} appropriate for ionic coordinates r'_{n+1} in the plane containing r_{n-2}, r_{n-1}, and r_n. We thus choose the linear combination, parametrized by α and β, that minimizes the difference $\|r'_{n+1} - r_{n+1}\|$.

which is a correct first-order trial wave function when the ionic coordinates are

$$\{r'\} = \{r(t_n) + \alpha[r(t_n) - r(t_{n-1})] + \beta[r(t_{n-1}) - r(t_{n-2})]\} . \quad (2.5)$$

To insure that the resulting wave functions are in as close correspondence as possible to the actual ionic locations $\{r(t_{n+1})\}$, we choose α and β by minimizing

$$\|r' - r(t_{n+1})\| = \|r(t_n) + \alpha[r(t_n) - r(t_{n-1})] + \beta[r(t_{n-1}) - r(t_{n-2})] - r(t_{n+1})\| , \quad (2.6)$$

which determines α and β uniquely. (This processes is illustrated in Fig. 1.) We have found that this two-dimensional extrapolation scheme represents a significant improvement over the corresponding one dimensional scheme, where we set $\beta = 0$ and optimize just α. One can also easily generalize such a scheme to dimensions beyond $N = 2$. If the modeled ionic motion involves N degrees of freedom, the trial wave functions become essentially exact $[r' = r(t_{n+1})]$, and the scheme will produce wave functions, and hence forces, correct to first order in the displacements, which would be sufficient to model harmonic phonon motion without refinement of the electronic wave functions beyond the guess (2.4). However, at some point r_n and r_{n-N} will become too far separated for a linear scheme such as this to work, and the benefits of matching r' closer and closer to r_{n+1} will be lost. Satisfied with the performance of the two-dimensional extrapolation scheme, we have not experimented with any of higher dimension.

III. RESULTS

As a test of our scheme, we have simulated bulk Si at 300 K in a periodic cubic supercell of 64 atoms with a Kleinman-Bylander pseudopotential at a cutoff of 8 Ry. The **k**-point integrations were approximated with just the gamma point. The initial configuration and velocities were chosen to give equal total potential and kinetic energies in directions chosen independently from the uniform process on the (3×64)-dimensional unit sphere. The equation of motion of the ions integrated using the Verlet algorithm with a time step of 3.256 fsec, which is $\frac{1}{20}$ of the $k = 0$ optic phonon period as determined from a frozen photon calculation. Note that we have somewhat arbitrarily taken the mass of the Si atoms to be the isotopic average of 28.09 amu, which in this homopolar simulation just sets the overall time scale. To update the wave functions, as described in Sec. II, we first performed the two-dimensional linear extrapolation with the optimization on α and β, and Gram-Schmidt orthonormalized the result. Because the linear extrapolation proceeds band by band, this requires no extra core memory. We then apply the CCG method. With the wave function for each band constrained in the subspace orthogonal to the other bands, we update it in successive conjugate directions until the change in the total energy is less than either 35×10^{-6} eV or 30% of the energy change of the first update. This procedure is applied to all of the bands until the energy change during an entire pass through the bands is less than 12.5×10^{-6} eV/ion. At this point, we calculate the Hellman-Feynman forces on the ions and move them again.

Our rationale for the above iterating criteria is as follows. For this system, we have found that the energy improvement from successive band updates decreases by factors of about 10, whereas the energy improvement from successive global band passes, *regardless of the number of band updates,* decreases by factors of 3 or 4. In both cases, updating individual bands and passing through all of the bands, this energy improvement proceeds in nearly perfect geometric progression. These results then indicate that the limiting factor in convergence to the BO surface is not refinement of individual bands but rather self-consistent adjustments of the bands with respect to each other. Consequently, refining the bands to more than about one-third of the energy change in their first update is not productive, and for this reason we have chosen the above 30% criteria. It is also clear that updating each band beyond where its change is 35×10^{-6} eV is unproductive because further refinement can only change the total energy of a band pass by $\sim(3.5 \times 10^{-6}$ eV/band) [2 (band/ion)] $= 7 \times 10^{-6}$ eV/ion, which is the tolerance we have set for this calculation. In a similar vein, we chose the band pass cutoff to be 12.5×10^{-6} eV/ion because in a perfect geometric sequence with decay constant one-third, the error is one-half of the last term included, placing us again within 7×10^{-6} eV/ion of the BO surface when we stop refining our wave functions.

We ran the above simulation for ~ 0.84 ps (260 time steps), which took 42.9 h of CPU time on the Pittsburgh Supercomputing Center Cray-YMP supercomputer. We caution, however, that because the simulation was started with completely converged wave functions as input, the earlier iterations required less electronic refinement than the latter. For the last half of the run, the CPU time re-

quired for each ionic iteration leveled off at a value corresponding to 66 CPU h/simulation psec. Note that although we are working with just the gamma point, our code works with fully complex wave function. Changing to real wave functions would speed the calculation by at least a factor of 3.

The results of the total and potential energies of the ionic system as a function of time are shown in Fig. 2. Over the entire 0.84-psec run, the total energy of the system drifted 5.1%, corresponding to a drift rate of 6.1%/psec. However, by the end of the run the fluctuations in U appear to be settling down to their microcanonical expectation,[11] and the total-energy drift rate settles down to well less than 1% [a least-squares fit gives $(0.30\pm0.10\%)$]. At this rate the system could run with the total-energy drift less than the microcanonical fluctuations in the potential energy ($\sim 2.5\%$) for several picoeseconds without requiring rescaling of their ionic velocities. Note that this does not represent a fundamental limit on our scheme, as energy conservation can be improved by using a better integration scheme for the ions than the Verlet algorithm or by using a smaller ionic time step. Finally, we have estimated the BO errors for this run by fitting the energies of successive band passes to geometric series. As intended, by far the most *representative* value was $\sim 4\times10^{-6}$ eV/ion, while for just a few iterations the BO error got as large as just under 8×10^{-6} eV/ion. Thus our scheme does indeed maintain the electrons extremely close to the BO surface without requiring quenches of the electronic system.[12]

As an application of the scheme and as a test of the quality of our dynamics, we have determined the photon spectrum of our supercell through the time-dependent velocity-velocity autocorrelation function. This was done by first projecting the ionic displacements from our simulation onto the allowed **k** states with no assumptions on the polarization vectors and then determining the frequencies present in the autocorrelation functions of these spatial Fourier *coefficients*. We used data from the entire 0.84 psec of the run. The transform of the autocorrelation function was effected by the maximum-entropy method (MEM).[13] (See Appendix B for a complete description of the analysis.) The *raw* results of fits with 220 poles for each class of **k** point are superposed and plotted in Fig. 3. Here, we have scaled the frequencies so that the optic phonon frequency ω_0 as calculated from a frozen phonon calculation at the same cutoff, is normalized to 1. The heights of the peaks, or the energies in them, are not meaningful in that our system has not yet equilibrated. We note that the primary caveat when working with the MEM is that it is known to produce spurious peaks when working with noisy data at a large number of poles. No such peaks were found, and the only effect of increasing the number of poles has been to sharpen the peaks, indicating very clean data. In fact, the signal to noise ratio in this plot is remarkable; although the L_2 peak is small, it is still 20 times larger than any local maximum not associated with a spectral peak. The agreement between the Γ'_{25} peak and the frozen phonon frequency ω_0 is excellent (+0.2%), and the discrepancy is on the order of what we expect due to our use of a time step of $\frac{1}{20}$ the period in a Verlet algorithm in a pure harmonic potential (+0.4%). Finally, taking the frequencies of the peak values of these spectra, we compare our results with the experimentally measured phonon frequencies,[14–17] and again find excellent agreement. These results are summarized in Fig. 4. In particular, our spectra clearly and accurately reproduce the celebrated flattening of the lower acoustic (TA) modes as one moves away from the Γ point. Also, we observe a fine splitting of the upper bands along Σ that was not reported in the initial experiments,[14] but which is present in bond charge

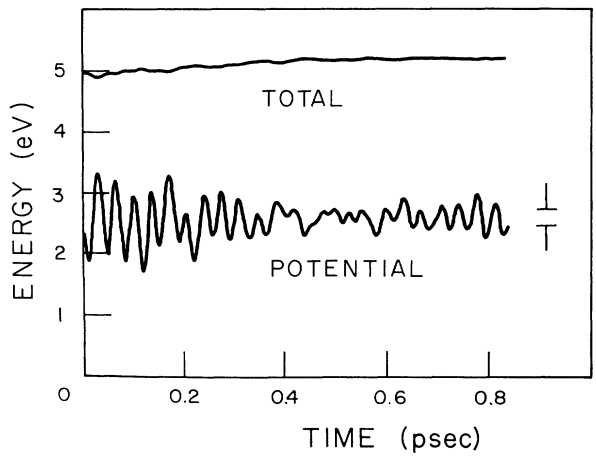

FIG. 2. Total and potential energy of the ionic system in our calculation. The bar indicates the $\pm\sigma$ variations in potential energy we would expect for our system in a microcanonical ensemble in the absence of anharmonicities. Note that though our choice of starting configuration initially results in large fluctuations, the fluctuations appear to be settling down to their expected thermal behavior by the end of the run.

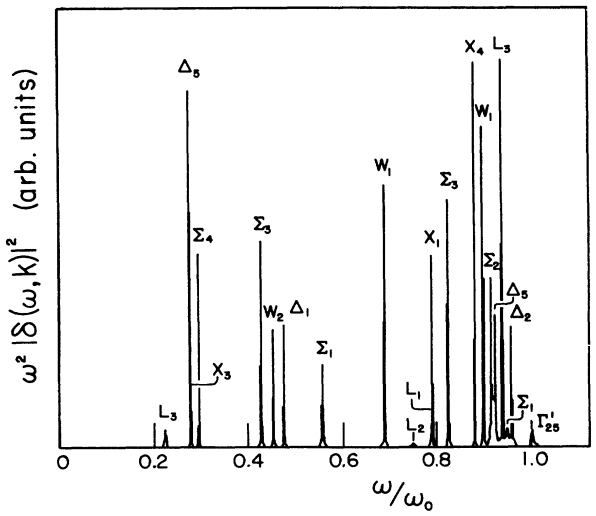

FIG. 3. Superposition of MEM spectral fits for each class of allowed phonon k state. The fits were performed with 220 poles on samples running for 260 times steps, the entire length of our run. The irreducible representations were determined from correspondence with experimental phonon data (Refs. 14–16).

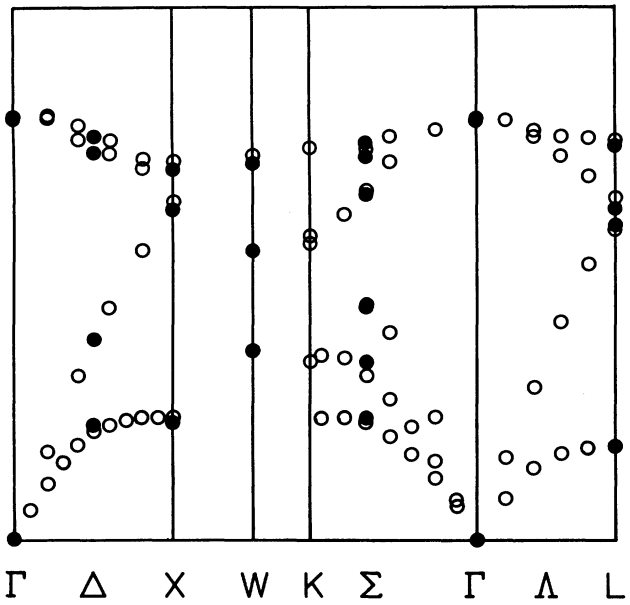

FIG. 4. Phonon spectrum as determined from peak values of data in Fig. 2. These values are completely *ab initio* with no free parameters. Open circles represent experimental data (Refs. 14–16), and solid circles represent our results.

model calculations.[18] A splitting on the same order has been observed in Ge,[19] which has a very similar phonon structure.

IV. CONCLUDING REMARKS

In conclusion, we have described a scheme that combines standard molecular dynamics for the ions, the CCG method for the electrons, and a method for producing good trial wave functions. We have verified with a realistic calculation that this scheme can maintain the electrons within a few times 10^{-6} eV of the BO surface without periodic quenching of the electronic system. Our results all show that the scheme produces accurate dynamic correlation functions with good microcanonical energy conservation over picosecond time scales without artificially scaling the ionic velocities. Finally, all of this is accomplished with about the same computational effort as traditional schemes. The present method may be further improved by employing a better molecular-dynamics scheme than the Verlet algorithm or by developing a better understanding of what band updating criteria to set in the CCG method. It also may be made more efficient by determining just what BO tolerances are necessary for a particular application. The subspace alignment formalism makes the scheme robust by providing a promising prescription for treating gapless systems that will be free of fictitious drag on the atomic system. For many systems this scheme may be viewed as available alternative to traditional combined molecular-dynamics techniques. However, if one is interested in simulating systems requiring super cells with large length scales (e.g., exceeding ~ 30 Å for Si) or systems where dynamic correlations are important, or systems that may be sensitive to fictitious vibrations in the ionic system, then this scheme might be preferred.

ACKNOWLEDGMENTS

This work was supported in part by ONR Contract No. N00014-86-K-0158, U.S. JSEP Contract No. DAAL-03-86-K-0002, and U.S. AFOSR Contract No. 87-0098. T.A.A. acknowledges support from AT&T Bell Laboratories, and M.C.P. acknowledges support from the Royal Society. We also thank M. Teter, D. Allan, K. Cho, and B. Meade for helpful discussions.

APPENDIX A: SUBSPACE ALIGNMENT

We are set with the task of minimizing $\sum_n \| |n\rangle - |n'\rangle \|^2$ by performing a separate unitary transformation in the space spanned by $\{|n\rangle\}$ and the space spanned by $\{|n'\rangle\}$ in order to "align" the two subspaces. Keeping in mind that eventually we will be concerned only with the quality of the alignment for the bands filled up to some level, we generalize our task to the minimization of a weighted sum,

$$S \equiv \sum_n f_n \| |n\rangle - |n'\rangle \|^2 . \tag{A1}$$

In this theory the quantity that plays the central role is the matrix $\underline{U}$ defined by

$$U_{nm} \equiv \langle m | n' \rangle . \tag{A2}$$

In terms of this matrix, we then are to minimize

$$S = \mathrm{Tr}[\underline{F}(2\underline{I} - \underline{U} - \underline{U}^\dagger)] , \tag{A3}$$

where $\underline{I}$ is the unit matrix and $\underline{F}$ is a diagonal matrix with f_n along its diagonal. Now consider the dual unitary transformations

$$\begin{aligned} |n^{(1)}\rangle &= \sum_m A_{nm} |m\rangle , \\ |n^{(1)'}\rangle &= \sum_m A'_{nm} |m'\rangle . \end{aligned} \tag{A4}$$

$\underline{U}$ then transforms to

$$\underline{U}^{(1)} = \underline{A}' \underline{U} \underline{A}^\dagger . \tag{A5}$$

To incorporate the unitarity constraints on $\underline{A}$ and $\underline{A}'$ we consider the Lagrange variational equation

$$0 = \delta[S^{(1)} + \mathrm{Tr}(\underline{\Lambda} \underline{A} \underline{A}^\dagger + \underline{\Lambda}' \underline{A}' \underline{A}'^\dagger)] . \tag{A6}$$

Here, $\underline{\Lambda}$ and $\underline{\Lambda}'$ are matrices of Lagrange multipliers, which must be Hermitian to incorporate the unitarity of the $\underline{A}$'s, and

$$S^{(1)} = \mathrm{Tr}[\underline{F}(2\underline{I} - \underline{A}' \underline{U} \underline{A}^\dagger - \underline{A} \underline{U}^\dagger \underline{A}'^\dagger)] . \tag{A7}$$

The above variation leads to the Lagrange-multiplier equations

$$\begin{aligned} \underline{F} \underline{A}' \underline{U} &= \underline{\Lambda} \underline{A} , \\ \underline{F} \underline{A} \underline{U}^\dagger &= \underline{\Lambda}' \underline{A}' . \end{aligned} \tag{A8}$$

The Lagrange matrices are easily related to $\underline{U}^{(1)}$ at its op-

timal value

$$\underline{F}\,\underline{U}^{(*)}=\underline{\Lambda}\ ,$$
$$\underline{F}\,\underline{U}^{(*)\dagger}=\underline{\Lambda}'\ . \tag{A9}$$

Note that all quantities transformed according to the optimal subspace alignment, will be given the superscript $(_*)$. Throughout this discussion we shall use freely the Hermitian properties of $\underline{F}$, $\underline{\Lambda}$, and $\underline{\Lambda}'$, and the unitarity of $\underline{A}$ and $\underline{A}'$. Eliminating $\underline{U}$ from (A9) gives

$$\underline{F}^2\underline{\Lambda}'=\underline{F}\,\underline{\Lambda}\,\underline{F}\ , \tag{A10}$$

and thus we conclude that $\underline{F}^2$ commutes with $\underline{\Lambda}'$. Since they commute, $\underline{F}^2$ and $\underline{\Lambda}'$ must be simultaneously diagonalizible, and thus $\underline{\Lambda}'$ must be a diagonal matrix because $\underline{F}$ is already diagonal. Note that any issues of degeneracy in $\underline{F}$'s spectrum may be trivially resolved by adding an arbitrarily small monotonic function of n to f_n. The fact that $\underline{\Lambda}'$ is diagonal is extremely important. Not only does this determine the transformation described by the $\underline{A}$'s, it also means that $\underline{U}^{(*)}$ is diagonal, which has very important geometric consequences.

To see how $\underline{\Lambda}'$ being diagonal determines the $\underline{A}$'s, we solve for $\underline{\Lambda}'$ by multiplying the second Lagrange equation in (A8) by its Hermitian conjugate

$$\underline{\Lambda}'^2=\underline{F}\,\underline{A}\,\underline{U}^\dagger\underline{U}\,\underline{A}^\dagger\underline{F}\ . \tag{A11}$$

The square root of this equation will yield a Hermitian $\underline{\Lambda}'$, as required, because the right-hand side is the product of a matrix with its Hermitian conjugate. The indeterminacy left in the sign of each eigenvalue of $\underline{\Lambda}'$ is to be resolved in the end by taking the choice which minimizes $S^{(*)}$. Because $\underline{\Lambda}'$ is diagonal so must be its square, which by (A11) means that $\underline{A}\,\underline{U}^\dagger\underline{U}\,\underline{A}^\dagger$ is diagonal because $\underline{F}$ is diagonal. Thus $\underline{A}$ is determined to be the unitary matrix diagonalizing $\underline{U}^\dagger\underline{U}$. Any further indeterminacy is resolved again by inspecting the resulting $S^{(*)}$. To determine explicit forms for the $\underline{A}$'s, we define $\underline{A}$ by

$$\underline{A}\,\underline{U}^\dagger\underline{U}\,\underline{A}^\dagger\equiv\cos^2\underline{\Theta}\ , \tag{A12}$$

where $\underline{\Theta}$ is a real diagonal matrix of elements between 0 and $\pi/2$. This choice of the form of the eigenvalues of $\underline{U}^\dagger\underline{U}$ is justified by the fact that, as shown below, its eigenvalues are all between zero and unity. With $\underline{A}$ now set, (A11) yields $\underline{\Lambda}'$,

$$\underline{\Lambda}'=\underline{F}(\pm\cos\underline{\Theta})\ , \tag{A13}$$

where the $\pm$ indicates an indeterminacy of sign in *each* diagonal element. From $\underline{\Lambda}'$ we easily get $\underline{A}'$ from the second Lagrange equation in (A8),

$$\underline{A}'=(\pm\cos\underline{\Theta})^{-1}\underline{A}\,\underline{U}^\dagger\ . \tag{A14}$$

We also can get $\underline{U}^{(*)}$,

$$\underline{U}^{(*)}=\pm\cos\underline{\Theta}\ , \tag{A15}$$

and from this we get the optimal S,

$$S^{(*)}=\sum_n 2f_n(1\mp\cos\theta_n)\ . \tag{A16}$$

From (A16) it is clear that to minimize S the correct choice of sign is the upper one, and that $\underline{A}$ should be chosen so as to list the eigenvalues of $\underline{U}^\dagger\underline{U}$ in descending (ascending) order provided that we have written the f_n in descending (ascending) order. This then determines completely and explicitly the linear combinations to be taken in each subspace to perform subspace alignment. Note that the $\underline{A}$'s are independent of the values of $\underline{F}$; this has important implications described below.

The geometric interpretation of the transformation effected by the $\underline{A}$'s is simple. After the transformation, $\underline{U}$ becomes

$$\underline{U}^{(*)}=\cos\underline{\Theta}\ , \tag{A17}$$

which means

$$\langle n^{(*)}|m'^{(*)}\rangle\equiv U^{(*)}_{mn}=\delta_{nm}\cos\theta_n\ . \tag{A18}$$

Thus it is clear that the eigenvalues of $\underline{U}^\dagger\underline{U}$, which are manifestly positive, are in fact $|\langle n|m'\rangle|^2$ for *some* alignment of the subspaces and thus must be between zero and unity as mentioned above. The fact that $\underline{U}^{(*)}$ is now diagonal indicates that the projection of each basis ket is one subspace lies along a basis ket in the other, and thus that all of these projections are orthogonal. This will not happen in general and only occurs here because we have allowed a change of basis in *both* subspaces.

This property is important because it factors the unitary transformation in Hilbert space that carries one subspace onto the other into a series of independent two-dimensional rotations. To see that this is true we shall assume that the two subspaces have already been aligned, so that $\underline{U}$ already has the form (A17), and we can drop the $(*)$ superscripts. To exhibit the factorization into independent rotations, we define a basis for the entire Hilbert space with three classes of ket: a series of kets, say $\{|\alpha\rangle\}$, that span the space orthogonal to both subspaces; those kets $|n\rangle=|n'\rangle$ for which $\theta_n=0$; and the remaining $|m\rangle$ and $|m'\rangle$ for which $\theta_m\neq0$. Clearly this forms a basis for the entire space. Orthonormalization is now trivial because $\langle n|m'\rangle\propto\delta_{nm}$. A suitable class (parametrized by η) of orthonormal bases $\{|\beta_\eta\rangle\}$ for the Hilbert space is

$$|\beta_\eta\rangle=\begin{cases}|\alpha\rangle & \text{for all } \alpha\\ |n\rangle & \text{for all } \theta_n=0\\ \cos(\eta\theta_m)|m\rangle+\sin(\eta\theta_m)\dfrac{|m'\rangle-\cos\theta_m|m\rangle}{\sin\theta_m} & \text{for all } \theta_m\neq0\\ -\sin(\eta\theta_m)|m\rangle+\cos(\eta\theta_m)\dfrac{|m'\rangle-\cos\theta_m|m\rangle}{\sin\theta_m} & \text{for all } \theta_m\neq0\ .\end{cases} \tag{A19}$$

We can then construct a corresponding series of unitary transformations over Hilbert space:

$$\mathcal{U}_\eta \equiv \sum_\beta |\beta_\eta\rangle\langle\beta_{\eta=0}| \ . \tag{A20}$$

In particular, for $\eta=1$ this transformation carries the unprimed subspace onto the primed subspace in the promised series of two-dimensional rotations. This transformation is also, by construction, the *minimal* such transformation. The geometric interpretation of the θ_n from the spectrum of $\underline{U}^\dagger\underline{U}$ is then the angles of the two-dimensional rotations in which the overall unitary transformation carrying one subspace onto the other may be factored. Finally, by adjusting η, one has an explicit formula for raising this transformation to arbitrary powers for interpolation or extrapolation of the changes between the subspaces.

The meaning of the independence of the $\underline{A}$'s from $\underline{F}$ is that the closest possible set of N pairs of states, one from each subspace, for any N up to the dimensionality of the subspaces is just those pairs of kets associated with the smallest N values of θ_n. We can see this by just setting the first N values of f_n to 1, and the rest to ϵ, where ϵ is small. (The ϵ artifice is necessary because $\underline{F}$ must be invertible.) As a result, if in going from $\{r_n\}$ to $\{r_{n+1}\}$ a few additional bands have crossed from above down into the set of bands we are considering, the procedure will *not* form linear combinations of low-lying states to try to match the new states at the expense of matching the low-lying states to each other. Thus, in such cases the unavoidable large changes in basis vectors will be isolated in the upper bands as promised in the text.

We also mentioned in the text that part of the usefulness of the subspace alignment scheme in dynamics calculations comes from the fact that alignment does not strongly mix Kohn-Sham eigenstates separated in energy by more than the typical eigenenergy drift from time step to time step. This is easy to see because $U_{nm} \equiv \langle m|n'\rangle$, where $|m\rangle$ and $|n'\rangle$ are solutions to two self-consistent single-particle Hamiltonians $\mathcal{H}_n$ and $\mathcal{H}_{n+1}$ at successive time steps. Viewed in this way, the subspace overlap matrix is given by

$$U_{nm} = \frac{\langle m|(\mathcal{H}_{n+1}-\mathcal{H}_n)|n'\rangle}{E'_n - E_m} \ , \tag{A21}$$

where the E's are exact eigenenergies as in the usual development of perturbation theory. From this it is clear that $\underline{U}$ and, hence, the $\underline{A}$'s will not strongly mix states unless $E'_n - E_m$ is $O(\mathcal{H}_{n+1}-\mathcal{H}_n)$, i.e., unless they are separated in energy by less than the typical drift in the eigenenergies between time steps. Finally, as a practical bonus, if we are solving the electronic problem with a scheme that yields the Kohn-Sham eigenstates, we can exploit the fact that $\underline{U}^\dagger\underline{U}$ is then nearly diagonal and compute the $\underline{A}$'s at little additional cost.

APPENDIX B: EXTRACTION OF FREQUENCIES—THE MAXIMUM-ENTROPY METHOD

The calculational scheme described in this paper yields $r_{Rb}(t_n)$, the displacement of each ion from its equilibrium location in the supercell at a finite set of discrete times $\{t_n\}$. Here we have labeled each ion according to its primitive lattice (in our case fcc) vector R and identity in the crystalline basis b because from this information we intend to extract the phonon spectrum $\omega(k)$ of the primitive fcc lattice. To extract the behavior at different k points, we make the following transformation:

$$r_{Rb}(t_n) \equiv \sum_k e^{ik\cdot R}\epsilon_{kb}(t_n) \ , \tag{B1}$$

$$\epsilon_{kb}(t_n) = \frac{1}{N_{\text{cell}}}\sum_R e^{-ik\cdot R} r_{Rb}(t_n) \ , \tag{B2}$$

where N_{cell} is the number of primitive cells in our supercell and k runs over the N_{cell} points of the lattice reciprocal to the superlattice that lie in the first Brillion zone of the primitive lattice. Note that because the b subscript for the polarization vectors ϵ_{kb} runs over the atoms in the basis rather than, as usual, the phonon bands, this transformation assumes nothing about the structure of the phonon system other than what we can conclude from translational symmetry. If we were to analyze *complete* discrete time sequences $\epsilon_{kb}(t_{-\infty}), \ldots, \epsilon_{kb}(t_\infty)$, then the $\omega(k)$ spectra we seek would correspond to sharp peaks in the Fourier intensities

$$P_{k^*}(\omega) \equiv \sum_{k\in k^*,b} |\epsilon_{kb}(\omega)|^2 \ . \tag{B3}$$

Here we have summed all of the spectral energy in stars of degenerate k points because this ultimately yields better statistics. Also, note that to conform with the standard practice of considering the spectrum of particle velocities rather than positions, one should use $P'_{k^*}(\omega) \equiv \omega^2 P_{k^*}(\omega)$. We have chosen the positions rather than the velocities as primary variables because the Verlet algorithm does not directly yield velocities, and the best prescriptions for estimating the velocities amount to nothing more than inverse transforming $i\omega\epsilon_{kb}(\omega)$. Finally, if we define k-star averaged autocorrelation functions

$$R^{\epsilon\epsilon}_{k^*}(t_n) \equiv \sum_m \sum_{k\in k^*,b} \epsilon_{kb}(t_m+t_n)\epsilon^*_{kb}(t_m) \ , \tag{B4}$$

then, as is well known, $R^{\epsilon\epsilon}_{k^*}(t_n)$ and $P_{k^*}(\omega)$ are transforms of each other,

$$\begin{aligned} R^{\epsilon\epsilon}_{k^*}(t_n) &= \int_{-\pi/\Delta}^{\pi/\Delta} \frac{\Delta\, d\omega}{2\pi} e^{i\omega t_n} P_{k^*}(\omega) \\ &= \frac{1}{2\pi i}\oint \frac{dz}{z} z^n P_{k^*}(z) \end{aligned} \tag{B5}$$

and

$$\begin{aligned} P_{k^*}(\omega) &= \sum_{n=-\infty}^{\infty} e^{-i\omega t_n} R^{\epsilon\epsilon}_{k^*}(t_n) \\ &= \sum_n z^{-n} R^{\epsilon\epsilon}_{k^*}(t_n) \ , \end{aligned} \tag{B6}$$

where $\Delta = t_{n+1} - t_n$ is the sampling interval, and $z \equiv e^{i\omega\Delta}$. The Fourier normalizations expressed in (B5) and (B6) shall be used throughout.

Suppose we have data for N time steps. Then we can

obtain $P_{k^*}(\omega)$ through (B6) by estimating the $R^{\epsilon\epsilon}_{k^*}$ with the so-called unbiased estimator,[20] which in our case takes the form

$$\hat{R}^{\epsilon\epsilon}_{k^*}(t_n)=\frac{1}{N-n}\sum_{m=0}^{N-n-1}\sum_{k\in k^*,b}\epsilon_{kb}(t_m+t_n)\epsilon^*_{kb}(t_m)\ . \tag{B7}$$

This though, only gives us estimates of $R^{\epsilon\epsilon}_{k^*}(t_{-(N-1)}),\ldots,R^{\epsilon\epsilon}_{k^*}(t_{N-1})$. [Recall that $R^{\epsilon\epsilon}_{k^*}(-t)=R^{\epsilon\epsilon^*}_{k^*}(t)$.] If we set the remaining $R^{\epsilon\epsilon}_{k^*}$ to zero [which would mean simply discrete Fourier transforming the $R^{\epsilon\epsilon}_{k^*}(t_n)$], our frequency resolution will be severely limited by the length of our sample. For example, in a 1-psec run of our calculation (300 time steps), where the time step is $\frac{1}{20}$ the Γ optic phonon period, the corresponding resolution is only about $\frac{1}{15}$ the optic phonon frequency.

An alternate approach is clearly needed. Through (B5) any estimator for the spectra $P_{k^*}(\omega)$ corresponds to some extrapolation scheme for the $R^{\epsilon\epsilon}_{k^*}(t_n)$ beyond those we can estimate directly. The approach that the maximum-entropy method brings to this extrapolation problem is simple: The information we have imposes a probability distribution for which sequence we are actually sampling in the space of all possible sequences $[\epsilon_{kb}(t_{-\infty}),\ldots,\epsilon_{kb}(t_{\infty})]$; a reasonable choice for the extrapolated $R^{\epsilon\epsilon}_{k^*}$ is just the mean of (B4) under this distribution. If this probability distribution is insufficiently peaked to guarantee that these mean values lay sufficiently close to the actual values we wish to extrapolate, then we must conclude that we lack sufficient information. The concept of maximum entropy enters in determining the probability distribution imposed by our information in the same way it does in statistical physics. For example, one can justify use of the macroscopic Boltzmann distribution $f(p,q)=\exp[-\beta H(p,q)]$ for a classical gas at constant energy by noting that of all macroscopic distributions with a given $E\equiv\int f(p,q)H(p,q)dp\,dq$, the Boltzmann distribution corresponds to the greatest number (in fact, the vast majority) of microstates Ω; that is, it maximizes the entropy $S\equiv\ln\Omega\propto -N_{\text{part}}\int f(p,q)\ln f(p,q)dp\,dq$, where N_{part} is the number of particles. In the same way, the vast majority of probability distribution functions $\wp(\{\epsilon_{kb}(t_n)\})$ with mean correlation functions

$$\langle R^{\epsilon\epsilon}_{k^*}(t_m)\rangle\equiv\int\wp(\{\epsilon_{kb}(t_n)\})R^{\epsilon\epsilon}_{k^*}(t_m)\prod_{n,k,b}d\epsilon_{kb}(t_n)$$

exactly equal to our estimates $\hat{R}^{\epsilon\epsilon}_{k^*}(t_m)$ yield very nearly equal $\langle R^{\epsilon\epsilon}_{k^*}(t_m)\rangle$ for the remaining m. We determine this distribution, and more importantly its mean correlation functions, in two phases. First, knowing that the distribution we seek has *some* set of $\langle R^{\epsilon\epsilon}_{k^*}(t_m)\rangle$, we find for each such set the distribution corresponding to the greatest number of "microstates" Ω as measured by the entropy

$$S=-\int\wp(\{\epsilon_{kb}(t_n)\})\ln\wp(\{\epsilon_{kb}(t_n)\})\prod_{n,k,b}d\epsilon_{kb}(t_n)\ .$$

Then, considering just *these* distributions, we pick out the one of maximum entropy which is consistent with our $\hat{R}^{\epsilon\epsilon}_{k^*}(t_m)$. This two-step process is convenient because, as we shall see shortly, the maximum entropy associated with a particular set of mean correlation functions is directly expressible in terms of the values of those mean correlations. Therefore the set of $\langle R^{\epsilon\epsilon}_{k^*}(t_m)\rangle$ we shall use in (B6) to obtain the $P_{k^*}(\omega)$ is just the one consistent with our $\hat{R}^{\epsilon\epsilon}_{k^*}(t_m)$ and maximizing this maximum-entropy expression.

We now proceed with the program, paralleling the standard development found in the literature with notation and minor generalizations appropriate to our problem. First, the distribution with maximum entropy with a fixed set of correlation functions is found by the Lagrange variation

$$0=\delta\int\wp(\{\epsilon_{kb}(t_n)\})\left[-\ln\wp(\{\epsilon_{kb}(t_n)\})-\mu-\sum_{m,k^*}\lambda_{m,k^*}R^{\epsilon\epsilon}_{k^*}(t_m)\right]\times\prod_{n,k,b}d\epsilon_{kb}(t_n)\ . \tag{B8}$$

Here, μ is a multiplier of the normalization constraint, and the λ_{m,k^*} are real and satisfy $\lambda_{m,k^*}=\lambda_{-m,k^*}$ like the $\langle R^{\epsilon\epsilon}_{k^*}(t_m)\rangle$ they constrain. Independently varying $\wp$ at each $\epsilon_{kb}(t_n)$ yields

$$\wp(\{\epsilon_{kb}(t_n)\})=Z^{-1}\exp\left[-\sum_{k^*}\sum_{k\in k^*b}\sum_{l,m}\epsilon^*_{kb}(t_m)\lambda_{l-m,k^*}\epsilon_{kb}(t_l)\right], \tag{B9}$$

where the partition function $Z\equiv e^{\mu+1}$ is set to normalize $\wp$,

$$Z=\int\exp\left[-\sum_{k^*}\sum_{k\in k^*b}\sum_{l,m}\epsilon^*_{kb}(t_m)\lambda_{l-m,k^*}\epsilon_{kb}(t_l)\right]\times\prod_{n,k,b}d\epsilon_{kb}(t_n)\ . \tag{B10}$$

Standard Gaussian integration then gives

$$S=\frac{N_{\text{dof}}}{2}+\ln Z=\frac{N_{\text{dof}}}{2}\left[1+\ln\pi-\sum_{k^*}w_{k^*}\frac{1}{2\pi i}\oint\frac{dz}{z}\ln\lambda_{k^*}(z)\right], \tag{B11}$$

where w_{k^*} is the fraction of k points in the star k^*, N_{dof}, an irrelevant constant, is the total number of degrees of freedom in the Gaussian, and $\lambda_{k^*}(z)$ is defined from the $\lambda_{k^*}(t_n)\equiv\lambda_{n,k^*}$ in analogy with (B6). To satisfy the constraints, the λ must insure

$$\langle R^{\epsilon\epsilon}_{k^*}(t_n)\rangle = -\frac{\partial}{\partial\lambda_{n,k^*}}\ln Z = w_{k^*}\frac{1}{2\pi i}\oint\frac{dz}{z}\frac{z^{-n}}{\lambda_{k^*}(z)}\ , \tag{B12}$$

which means simply

$$\langle P_{k^*}(z^*)\rangle = w_{k^*}[\lambda_{k^*}(z)]^{-1}\ . \tag{B13}$$

Hence, aside from constants which do not change with our data, we can take the entropy associated with a probability distribution whose mean spectra are $\langle P_{k^*}(z)\rangle$ to be

$$S=\sum_{k^*} w_{k^*}\frac{1}{2\pi i}\oint\frac{dz}{z}\ln\langle P_{k^*}(z)\rangle\ . \tag{B14}$$

We now maximize the entropy given in (B14) over all possible spectra that are consistent with our estimates $\hat{R}^{\epsilon\epsilon}_{k^*}$, with the result being the spectra we seek. Because we shall use the $\langle P_{k^*}(z)\rangle$ determined in this way as our estimate for the spectra, we shall now drop the expectation-value brackets for convenience. This gives

$$0=\delta\sum_{k^*} w_{k^*}\frac{1}{2\pi i}\oint\frac{dz}{z}\left[\ln P_{k^*}(z) - \sum_{m=-(N-1)}^{N-1} c_{k^*,m}P_{k^*}(z)z^m\right], \tag{B15}$$

where the c's are the multipliers for the constraints that the estimates from (B7) correspond to the actual mean correlations as we know them from (B5). The c's must be real and obey $c_{k^*,m}=c_{k^*,-m}$ as do the $R^{\epsilon\epsilon}_{k^*}(t_m)$ they constrain. The solution to (B15) is

$$P_{k^*}(z)=\frac{1}{\sum_m c_{k^*,m}z^m}\ , \tag{B16}$$

where the c's must assure

$$\hat{R}^{\epsilon\epsilon}_{k^*}(t_n)=\frac{1}{2\pi i}\oint\frac{dz}{z}\frac{z^n}{\sum_m c_{k^*,m}z^m}\ . \tag{B17}$$

Following Edward and Fitelson, we now proceed to derive the Yule-Walker equations from (B17).[21] Using Wold's method,[22] we can derive constraints on the form of the sum in (B16). Knowledge of $P_{k^*}(z)$ on the unit circle is sufficient to determine the c_m; precisely,

$$c_m=\frac{1}{2N-1}\sum_{l=-(N-1)}^{N-1} z_l^m P^{-1}_{k^*}(z_l)\ , \tag{B18}$$

where the z_l are spaced evenly around the unit circle in the usual way. From this we easily verify the Lagrange conditions that $c_m=c_m^*=c_{-m}$ because $P_{k^*}(z)$ must be real and invariant under $z\to z^*$ on the unit circle. Proceeding, we rewrite (B16) as

$$P^{-1}_{k^*}(z)=c_{N-1}z^{-(N-1)}\prod_{n=1}^{2(N-1)}(z-\zeta_n)\ , \tag{B19}$$

where the aforementioned properties of the c_n guarantee that the polynomial roots ζ_n obey the property that if ζ is a root, then so are ζ^*, ζ^{-1}, and ζ^{*-1}. These latter facts may be seen by direct substitution into the sum in (B16). Because $P_{k^*}(z)$ must also be *positive* on the unit circle, we further know that each ζ on the circle must occur with even multiplicity, otherwise $P_{k^*}(z)$ would flip sign as we pass through it. Consequently, we can divide the roots into two groups $(N-1)$ each, $\{\zeta_n\}$ and $\{\zeta'_n\}$, where, for all n, $\zeta_n=1/\zeta'^*_n$ and ζ_n does not lie inside the unit circle. This allows us to rewrite (B19) for $|z|=1$ as

$$P_{k^*}(z)^{-1}=\left[(-1)^{N-1}c_{N-1}\prod_{n=1}^{N-1}\zeta'_n\right]\times\left[\prod_{n=1}^{N-1}(z-\zeta_n)\right]\left[\prod_{n=1}^{N-1}(z-1/\zeta'^*_n)\right]^* \equiv\frac{H_{k^*}(z)H^*_{k^*}(z)}{\sigma^2}\ , \tag{B20}$$

where

$$H_{k^*}(z)\equiv\sum_{n=0}^{N-1} a_n^*(k^*)z^n \tag{B21}$$

is some $N-1$ degree polynomial with no zeros inside the unit circle, and

$$\sigma^2\equiv\left[(-1)^{N-1}c_{N-1}\prod_{n=1}^{N-1}\zeta_n\right]^{-1} \tag{B22}$$

is a real constant chosen so that $a_0(k^*)\equiv 1$. Multiplying both sides of the reciprocal of (B20) by $z^nH^*_{k^*}(z)$ and integrating around the unit circle employing (B5) yields the Yule-Walker equations

$$\sum_{l=0}^{N-1} a_l(k^*)\hat{R}^{\epsilon\epsilon}_{k^*}(t_{n-l})=\frac{1}{2\pi i}\oint\frac{dz}{z}\frac{\sigma^2z^n}{H_{k^*}(z)}=\sigma^2\delta_{n0}\ , \tag{B23}$$

where we have used the fact that $1/H_{k^*}(z)$ has no poles inside the unit circle and collapsed the contour to the origin.

Finally, the Yule-Walker equations, (B23), may be solved with the Levinson-Durbin algorithm,[23,24] which recursively solves these equations for increasing numbers of estimated $R^{\epsilon\epsilon}_{k^*}$. If we label the solution to (B23) when there are $p+1$ autocorrelation lags $\{a_{p1}(k^*),a_{p2}(k^*),\ldots,a_{pp}(k^*),\sigma_p^2(k^*)\}$, then the Levinson-Durbin algorithm gives the solution at order p in terms of the solution at order $p-1$ as

$$a_{pp}(k^*) = -\left[\hat{R}^{\epsilon\epsilon}_{k^*}(t_p) + \sum_{q=1}^{p-1} a_{p-1,q}(k^*)\hat{R}^{\epsilon\epsilon}_{k^*}(t_{p-q}) \right] \Big/ \sigma^2_{p-1}$$

$$a_{pq}(k^*) = a_{p-1,q}(k^*) + a_{pp}(k^*)a^*_{p-1,p-q}(k^*) , \quad \text{(B24)}$$

$$\sigma^2_p(k^*) = [1 - |a_{pp}(k^*)|^2]\sigma^2_{p-1}(k^*) ,$$

and the initial solutions are simply

$$a_{11}(k^*) = -\hat{R}^{\epsilon\epsilon}_{k^*}(t_1)/\hat{R}^{\epsilon\epsilon}_{k^*}(t_0) ,$$
$$\sigma^2_1(k^*) = [1 - |a_{11}(k^*)|^2]\hat{R}^{\epsilon\epsilon}_{k^*}(t_0) , \quad \text{(B25)}$$

as may be verified directly in (B23) by induction.

In summary, our procedure for determining the spectra shown in Fig. 3 is to first project out the phonon character using (B2) and compute the $\hat{R}^{\epsilon\epsilon}_{k^*}(t_n)$ from (B7). These values are then employed in (B24) and (B25) until a solution is built up to some order p. The spectra are then given directly by $\omega^2 P_{k^*}(\omega)$ with $P_{k^*}(\omega)$ given by (B20) and (B21). The only "free parameter" in this scheme is the order p, one less than the number of correlation estimates used in the fit. Many researches have observed difficulties such as spurious peaks, peak splitting, and peak shifting when using p approaching the number of sample data.[20] This is probably due to insisting that the mean correlations of the probability distribution $\wp(\{\epsilon_{kb}(t_n)\})$ be exactly the estimated correlation functions, which will tend to be most in error for larger time lags. One solution to this potential problem is to develop a scheme where the entropy is maximized with constraints that allow some differences between the $\langle R^{\epsilon\epsilon}_{k^*}(t_n)\rangle$ and the $\hat{R}^{\epsilon\epsilon}_{k^*}(t_n)$ with less weight given to the higher order $\hat{R}^{\epsilon\epsilon}_{k^*}(t_n)$.[25] However, we have found that increasing p up to as much as 90% of N, the number of sample times, only sharpens our peaks and produces none of the aforementioned affects. This suggests that our data, and hence our dynamics, are very well behaved. Another possible way to improve the scheme is to make the probability distribution $\wp$ sharper by incorporating more information. For instance, the inner product in (B4) may be replaced by an outer product resulting in matrix $R^{\epsilon\epsilon}_{k^*}(t_n)$ that incorporate polarization information. Such a scheme gives (B16) with $P_{k^*}(z)$ and $c_{k^*,n}$ replaced with matrices. The Yule-Walker equations and the Levinson-Durbin algorithm are not so easily generalized, however, and for this reason we have adopted the standard "scalar" approach in analyzing the spectra for this paper.

[1] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[2] P. Ballone, W. Andreoni, R. Car, and M. Parrinello, Phys. Rev. Lett. **60**, 271 (1988).

[3] R. Car and M. Parrinello, Phys. Rev. Lett. **60**, 204 (1988).

[4] F. Buda, R. Car, and M. Parrinello, Phys. Rev. B **41**, 1680 (1990).

[5] F. Buda, G. L. Chiarotti, R. Car, and M. Parrinello, Phys. Rev. Lett. **63**, 294 (1989).

[6] M. P. Teter, M. C. Payne, and D. C. Allan, Phys. Rev. B **40**, 12 255 (1989).

[7] Assuming the electronic wave functions rigidly follow the ions, so that $\Psi(x,t) = \psi(x - v_{\text{ion}}t)$, where Ψ is the wave function and v_{ion} is the ionic velocity, one can easily show that to adiabatically track the ions, the electronic subsystem needs a kinetic energy of $O(\mu/M) \sim 0.01$ times the total ionic kinetic energy, where μ and M are the (fictitious) electronic and ionic masses, respectively. Fully relaxed static calculations on Ge bare this out and show that the prefactor on μ/M is near unity.

[8] I. Štich, R. Car, and M. Parrinello, Phys. Rev. Lett. **63**, 2240 (1989).

[9] M. J. Gillan, J. Phys. Condens. Matter **1**, 689 (1989).

[10] The error in Born-Oppenheimer forces will be much worse, only $O((dr)^2)$. However, because the computation of $\Psi'(\{r + x\,dr\})$ is relatively cheap, one can take advantage of the $O((dr)^4)$ behavior of $E[\Psi', \{r + x\,dr\}]$ and compute by numerical differentiation the force to order $O((dr)^3)$ along $\{dr\}$, which is the important direction for energy conservation. In our work here, however, we have not taken advantage of this observation.

[11] From thermal conductivity measurements, one can estimate the phonon-phonon scattering time in room-temperature silicon to be on the order of picoseconds.

[12] A similar BO tolerance has been reported in a combined molecular-dynamics calculation; see D. K. Remler and P. A. Madden, Mol. Phys. **70**, 921 (1990), where they report $a \sim 4$-μeV/ion drift over 0.9 psec in an eight-atom germanium system at 1000 K. However, this calculation represents a very special test case, a system with only one *active* ionic degree of freedom, the Γ_{LTO} mode. Because the crucial error averaging in the combined molecular-dynamics method proceeds in time, the cancellation is much more exact when errors only have to be averaged for the force on one ionic degree of freedom. The CCG-based technique exhibits a similar effect and in a directly comparable single-mode calculation maintains the BO error at ~ 0.04 μeV/ion with 20% less CPU investment. Again, emphasize that the BO drift may be expected to be much worse in both schemes when more than one ionic degree of freedom is active, and it is for this reason that we chose to focus on a system with many (3×64) active ionic degrees of freedom.

[13] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, England, 1989), p. 431.

[14] G. Dolling, *Inelastic Scattering of Neutrons in Solids and Liquids Vol. II* (International Atomic Energy Agency, Vienna, Austria, 1963), p. 37.

[15] G. Nilsson and G. Nelin, Phys. Rev. B **6**, 3777 (1972).

[16] Paul A. Temple and C. E. Hathaway, Phys. Rev. B **7**, 3685 (1973).

[17] M. N. Wybourne, *Properties of Silicon* (The Institution of Electrical Engineers, London, 1988), p. 40.

[18] W. Weber, Phys. Rev. B **15**, 4789 (1977).

[19] G. Nelin and G. Nilsson, Phys. Rev. B **5**, 3151 (1972).

[20] S. M. Kay and S. L. Marple, Jr., Proc. IEEE **69**, 1380 (1981).

This is a very good review article discussing many kinds of spectral estimation.

[21]J. A. Edward and M. M. Fitelson, IEEE Trans. Inf. Theory **IT-19**, 232 (1973).

[22]H. O. A. Wold, *A Study in the Analysis of Stationary Time Series* (Almquist and Wiksells, Uppsala, 1938), p. 214.

[23]J. Durbin, Rev. Inst. Int. Stat. **28**, 233 (1960).

[24]N. Lenvinson, J. Math. Phys. **25**, 261 (1947).

[25]S. F. Gull and G. J. Daniell, *Image Formation from Coherence Functions in Astronomy* (Reidel, Dordrecht, 1979), p. 219.