Ab Initio Molecular Dynamics: Analytically Continued Energy Functionals and Insights into Iterative Solutions

T. A. Arias,[1] M. C. Payne,[2] and J. D. Joannopoulos[1]

[1]*Physics Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*
[2]*Cavendish Laboratory, Madingley Road, Cambridge, CB30HE, United Kingdom*

(Received 26 May 1992)

We present a new method for performing finite-temperature *ab initio* total-energy calculations at long length scales, which we demonstrate with a dynamics calculation of 50-Å-long phonon modes in silicon. The method involves *both* a prescription for the analytic continuation of traditional fermionic energy functionals into the space of nonorthonormal single-particle orbitals (speeding convergence to the minimum) *and* insights into the common computational physics problem of solving by iterative refinement for the state of a complex system as a function of a continuous external parameter.

PACS numbers: 71.10.+x, 63.20.Dj

Molecular dynamics is an intuitive and direct method of computer simulation of condensed matter systems. In a milestone paper of the last decade [1], Car and Parrinello introduced an *ab initio* technique where the electronic wave functions are evolved simultaneously with the ions using a fictitious equation-of-motion approach which keeps the electrons close to the Born-Oppenheimer (BO) surface. With recent advances in highly parallel computer architecture, calculations on systems with hundreds of atoms are now tractable [2], not only permitting more comfortable extraction of the thermodynamic limit but also extending the qualitative level of complexity in accessible problems. However, with growing interest in calculations involving larger systems, it has been demonstrated that, as the longest physical length scale in a calculation increases, integration of the fictitious electronic equations of motion becomes dominated by fluctuations in the electronic Hartree energy ("charge sloshing") [3] which force the time step for integration of the equations of motion to approach zero as the system size increases. While the onset of this instability depends on the physics of the system under study, the effect has been demonstrated dramatically in a silicon system of six primitive cells (~46 Å long) at 8 Ry [3]. To fully benefit from the next generation of computer technologies, an approach free of charge sloshing is needed.

Recently, conjugate-gradient procedures introduced by Teter, Payne, and Allan [3] and Gillan [4] for relaxing the electronic wave functions at fixed atomic arrangements have been shown not to exhibit the aforementioned instability and to consume an order of magnitude less CPU time than the best fictitious electronic dynamics schemes. Despite this improvement, performing a dynamics calculation, by repeatedly taking the wave-function solution from the previous ionic time step and refining it back to the BO surface with conjugate-gradient techniques to produce the solution for the current time step, is very costly [5] because of the tight tolerances to which the electronic problem must be solved (typically 10 μeV/atom) to yield stable atomic trajectories. In this work we introduce improvements over the procedure just described both in the choice of trial wave function and in the search for the BO surface which results in an efficient *ab initio* dynamics scheme suited for calculations in long-length-scale systems.

Alternative initial conditions for refinement at each time step.—We are faced with the general problem of solving by iterative refinement for the state of a system, specified in our case by the wave functions Φ, as a function of a continuously varying external parameter, for us, the time t. At each time step t_n one may, as we described above, simply take the solution from the previous time step $\Phi(t_{n-1})$ as input to the refinement process. An improvement on this simple idea may be to take the familiar multilinear extrapolation as the initial "trial" solution,

$$\Phi'(t_n)=\sum_{m=1}^{M}(-1)^{m+1}\binom{M}{m}\Phi(t_{n-m}), \tag{1}$$

where $\binom{M}{m}\equiv M!/m!(M-m)!$ [6]. But, as we soon shall see, such extrapolations often introduce instabilities into the algorithm.

To understand the origin of these instabilities in the hope of eliminating them, let us assume that $\Phi'(t_n)$ is already fairly close to the exact solution, $\Phi^*(t_n)$, and linearize the effect of the iterative process so that we may write

$$\Phi(t_n)=\Phi^*(t_n)+F[\Phi'(t_n)-\Phi^*(t_n)], \tag{2}$$

where F is some matrix. Assuming that F may be resolved into principle axes α, (1) and (2) combine to give a series of linear difference equations for the amplitude of the error along each principle axis, $\delta_\alpha\equiv[\Phi(t_n)-\Phi^*(t_n)]_\alpha$:

$$\delta_\alpha(t_n)+f_\alpha\sum_{m=1}^{M}(-1)^m\binom{M}{m}\delta_\alpha(t_{n-m})=-f_\alpha\sum_{m=0}^{M}(-1)^m\binom{M}{m}\Phi^*(t_{n-m}), \tag{3}$$

where the f_α are the characteristic values of F. Though we know nothing *a priori* about the term on the right-hand side of (3), it serves as a driving term for the homogeneous kernel on the left, and the error along α will be unstable to the formation of exponentially growing modes if any of the roots of the kernel's characteristic polynomial lay outside the unit circle. In Fig. 1 we see that for the first proper extrapolation, $M=2$, all modes are stable (in general we expect $|f_\alpha|<1$), but for $M=3$ any error mode with $f<-\frac{1}{7}$ or $f>\frac{1}{2}$ will grow exponentially until these "difficult" modes dominate the convergence of the system from Φ' toward Φ^* and significantly degrade the performance of the algorithm. The analysis of the intermediate case where $M=2$ and $M=3$ extrapolations alternate with each time step is more difficult (we must then consider distinct F matrices for the $M=2$ and $M=3$ time steps), but the behavior will be intermediate with an *increased* range of stability.

Our calculations in the system described in [5] exhibit exactly these phenomena. Though the $M=2$, the alternating, and finally the $M=3$ extrapolation schemes all represent successive improvements at the beginning of the calculations, the alternating and the $M=3$ schemes rapidly degrade, eventually performing no better or even worse than using no extrapolation at all (see Fig. 2). Analysis of the convergence in the conjugate-gradient iterative solution for the wave functions confirms that the cause of the slowing is a reduction in the overall convergence rate. The initial benefit of these higher-order extrapolation schemes *always* will be lost to this instability unless a new iterative solution for the electronic system is able to relax the "difficult" modes more efficiently so that their convergence factors move into the range of stability.

Analytic continuation of the total-energy functional.—One class of difficult wave-function errors that we must learn to relax more efficiently are those which require simultaneous self-consistent adjustment of all the wave functions. Because the method of conjugate gradients is not a constrained minimization procedure, current application of these techniques to the electronic problem either simply ignore the effect on the gradients of the "forces of constraint" which keep the wave functions orthonormal [4] or they treat these effects properly but at the cost of cycling through the wave functions individually and minimizing each wave function in the *subspace* orthogonal to all the other states which are then held fixed [3]. In either case those modes in which the wave functions must change in response to one another are treated poorly. We see evidence of this in our calculations of the latter class, where the wave functions converge much more rapidly in their restricted subspaces (with $f_\alpha \approx 0.3$) than does the system as a whole after complete cycles through the bands (with an overall $f \approx 0.75$). This effect is also sometimes seen in systems with ionic disorder. To speed the electronic relaxation in general and eliminate the instabilities which keep us from a very efficient dynamics scheme, we now present a method for relaxing the orthonormality constraints on the wave functions. Because traditional local-density approximation (LDA) functionals are physically meaningless unless evaluated on an orthonormal set of wave functions, to give meaningful results we must introduce a new energy functional which is the composition of a traditional LDA functional with a second functional which carries a set of nonorthonormal wave functions to an orthonormal set. The condition for the practicality of this scheme is

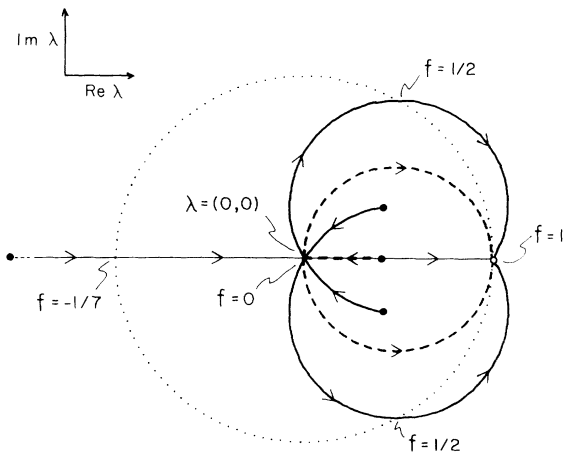

FIG. 1. Behavior of the characteristic roots of the $M=2$ (thick curves) and the $M=3$ (thin curves) in the complex plane in relation to the unit circle (dots) as a function of the parameter $-1<f_\alpha<1$. Note that both roots of the $M=2$ kernel always lay inside the unit circle and modes with all f_α's are therefore stable. For $M=3$, however, those modes with $f_\alpha<-\frac{1}{7}$ or $f_\alpha>\frac{1}{2}$ have roots outside the unit circle and therefore will show exponential growth until they dominate the convergence of the entire system.

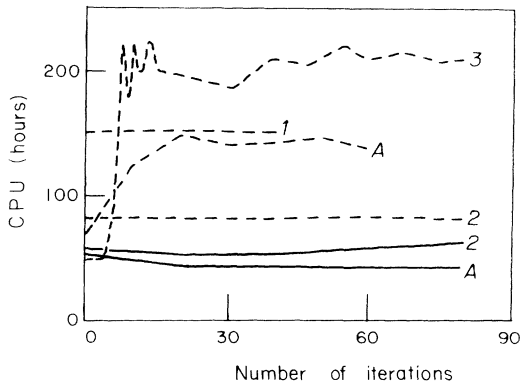

FIG. 2. Average CPU consumption (expressed in terms of the number of hours required to simulate 1 ps on a Cray YMP supercomputer) as a function of atomic iteration number for the simple nonextrapolatory scheme ($M=1$), the $M=2$, the alternating (labeled "A"), and the $M=3$ schemes when relaxing wave functions using a traditional LDA energy functional (dashed lines) and our new continued functional (solid lines).

that the gradient of the resulting composite functional be simple to evaluate.

This condition is met when we consider functionals carrying a nonorthonormal set of wave functions $\{\phi_{nk}\}$ to an orthonormal set $\{\psi_{nk}\}$ *spanning the same subspace*. The most general such transformation is of the form

$$\psi_k \equiv A_k \phi_k \equiv e^{iB_k} U_k^{-1/2} \phi_k , \tag{4}$$

where the two sets of N wave functions at each k point, $\{\psi_{nk}\}$ and $\{\phi_{nk}\}$, have been collected into the column vectors ψ_k and ϕ_k, B_k is an arbitrary $N \times N$ Hermitian matrix, and U_k is the $N \times N$ matrix with elements

$$[U_k]_{nm} \equiv \langle \phi_{mk} | \phi_{nk} \rangle . \tag{5}$$

There are several useful choices for the B_k, which just parametrize the arbitrary unitary transformation $V_k \equiv e^{iB_k}$ that we are free to apply once multiplication by $U_k^{-1/2}$ has given us orthonormal wave functions. First, the $\{[B_k]_{nm}\}$ may be treated as a set of parameters to be optimized on an equal footing along with the ϕ_{nk}, which speeds convergence over holding the B_k fixed by a factor of 2 in systems with variable fillings, in which the rotations V_k are then physically relevant. The B_k also may be chosen to diagonalize the subspace Hamiltonian, giving manifest treatment of the Kohn-Sham eigenstates, which will be a great advantage in metals. Finally, in cases like our semiconductor calculations where the fillings are constant and the V_k irrelevant, simply holding $B_k \equiv 0$ is most convenient. The new composite functional is now just

$$E(\{\phi_k^\dagger\},\{\phi_k\}) \equiv E_{\mathrm{LDA}}(\{\phi_k^\dagger A_k^\dagger(\phi_k^\dagger,\phi_k)\},\{A_k(\phi_k^\dagger,\phi_k)\phi_k\}) , \tag{6}$$

where $E_{\mathrm{LDA}}(\{\psi_k^\dagger\},\{\psi_k\})$ is the traditional LDA energy functional expressed in terms of the wave functions and their complex conjugates. The required analytic expressions for the gradients are simply

$$\frac{\partial E}{\partial \phi_k^\dagger} = A_k^\dagger F_k \hat{P}_k \hat{H}_{\mathrm{LDA}} \psi_k - Q\{V_k^\dagger [H_k, F_k] V_k\} V_k^\dagger \psi_k , \tag{7}$$

$$\frac{\partial E}{\partial [B_k^*]_{nm}} = -(R\{[H_k, F_k]\})_{nm} ,$$

where F_k is a diagonal matrix loaded with the filling factors for each band n, $\hat{P}_k$ is the projection operator

$$\hat{P}_k = \left(1 - \sum_n |\psi_{nk}\rangle\langle\psi_{nk}|\right) , \tag{8}$$

$\hat{H}_{\mathrm{LDA}}$ is the Kohn-Sham Hamiltonian operator constructed as usual from the *orthonormal* wave functions, $[H_k]_{nm} \equiv \langle \psi_{mk} | \hat{H}_{\mathrm{LDA}} | \psi_{nk} \rangle$ is the subspace Hamiltonian matrix, and Q and R are matrix to matrix functions defined by the relations

$$[W^\dagger Q\{M\} W]_{nm} \equiv [W^\dagger M W]_{nm} \frac{1}{\mu_n^{1/2} + \mu_m^{1/2}} \tag{9}$$

and

$$[Z^\dagger R\{M\} Z]_{nm} = [Z^\dagger M Z]_{nm} \frac{e^{-i(\beta_n - \beta_m)} - 1}{\beta_n - \beta_m} , \tag{10}$$

where M is an arbitrary matrix, and $\{Z,\beta\}$ and $\{W,\mu\}$ result from diagonalizing B and U,

$$Z \beta Z^\dagger \equiv B, \quad W \mu W^\dagger \equiv U . \tag{11}$$

A test of the prowess of the new dynamics technique.—To solve the electronic problem we now release the orthonormality constraints and simply apply the standard method of conjugate gradients [7] to the energy functional (6) with the preconditioning suggested by Teter, Payne, and Allan [3]. Calculations for fixed ionic coordinates with the wave functions from (1) which slowed the $M = 3$ simulations confirm that the new technique does indeed rapidly relax the previously difficult modes and can speed convergence by at least a factor of 3. When using the continued functional to perform dynamical calculations, we find that it is most efficient to input the orthonormal "images" from (4) into the extrapolation (1). Note that there are no longer any technical issues in wave-function extrapolation resulting from nonorthonormality of the extrapolated wave functions. To demonstrate the practicality of the new technique for long-length-scale systems, we have performed tests using the system described in [5] below, which is well over 50 Å long. Though the $M = 3$ scheme is still unstable (due to the presence of another, yet undetermined, set of "difficult" modes), the more robust alternating scheme is now stabilized, and requires only 44 CPU hours per simulated picosecond (Fig. 2). Given that this calculation is carried out with fully complex wave functions at four k points in the Brillouin zone and with both p and d nonlocal corrections, this is very competitive even with the CPU requirements of traditional combined-dynamics simulations, which are carried out in systems with much smaller length scales.

Finally, as a rigorous test of the quality of the trajectories that this procedure generates, even at these extended length scales, we have determined the phonon spectrum of Si through the velocity-velocity autocorrelation function of the ions in our calculation. Although there are methods for mapping out the phonon spectrum over the full Brillouin zone with calculations on a single *primitive* unit cell [8–11], calculating the phonon spectrum dynamically provides a simple, direct test of the quality of our molecular-dynamics technique in the highly (120) dimensional phase space of our supercell. After projecting the ionic displacements from our calculation onto the allowed k states and the known purely longitudinal and transverse polarization vectors, we then determine the frequencies present in the time autocorrelation functions of these spatial *Fourier coefficients* with the maximum entropy method (MEM) [12], a procedure we have discussed in detail elsewhere [5]. Taking the frequencies of

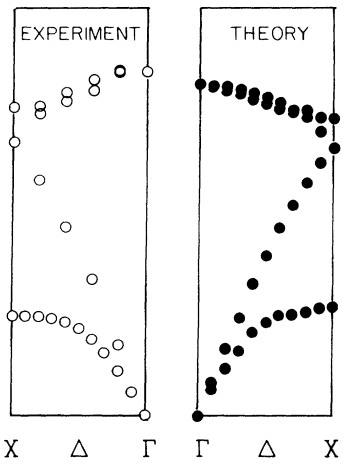

FIG. 3. Phonon spectrum as determined from maximum peak values of MEM fits. These values are completely *ab initio* with no free parameters. The left-hand panel displays experimental data [13,14], and the right-hand panel displays frequencies from our dynamically determined trajectories.

the peak values of the resulting spectra, we compare our parameter-free results with the experimentally measured phonon frequencies [13,14] in Fig. 3 and find excellent agreement [15]. Our spectra clearly and accurately reproduce not only the celebrated flattening of the lower acoustic (TA) modes as one moves away from the Γ point but also the delicate splitting of the optic modes along Δ. Note that the calculation also accurately reproduces the lowest-frequency acoustic mode despite its period of nearly 1 ps and manifests the delicate splitting of the optic modes, which beat against each other with even longer periods. This illustrates that our dynamics technique produces reliable ionic trajectories even over periods of picoseconds and is thus very precise.

Conclusions.— We have clarified several stability issues which will be of use to the general computational physics community in the problem of using iterative refinement methods to solve for the state of a system as a function of a continuously varying external parameter. This led to an understanding of the problem which prompted us to develop a technique for continuing traditional energy functionals into the space of nonorthonormal wave functions. These new functionals not only have greatly improved the performance of our conjugate-gradient-based *ab initio* molecular-dynamics scheme, which we have proven with a rigorous test in a system well over 50 Å in length, but also may be useful for practitioners of other techniques. The possibility of using these functionals as the basis for a Car-Parrinello scheme for metals where the Kohn-Sham eigenstates are obtained manifestly is intriguing. Finally, relieving the orthonormality constraints may be useful in other *ab initio* methods for fermionic systems involving orthonormal orbitals, such as configuration-interaction techniques.

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. acknowledges support from AT&T Bell Laboratories, and M.P. acknowledges support from the Royal Society.

[1] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[2] K. D. Brommer, M. Needels, B. E. Larson, and J. D. Joannopoulos, Phys. Rev. Lett. **68**, 1355 (1992).
[3] M. Teter, M. Payne, and D. Allan, Phys. Rev. B **40**, 12255 (1989).
[4] M. J. Gillan, J. Phys. Condens. Matter **1**, 689 (1989).
[5] To follow the dynamics of a 53.8-Å-long orthogonal supercell of forty silicon atoms at a time step of one-twentieth the optic-phonon period using this method requires ~150 CPU hours per simulated picosecond on a Cray YMP supercomputer. The remaining parameters of this test calculation are the same as those for the calculation described by T. A. Arias, M. C. Payne, and J. D. Joannopoulos, Phys. Rev. B **45**, 1538 (1992).
[6] Previously, we presented a refined choice for the multilinear coefficients which we have employed in the calculations in the present work [5]. In practice, however, the coefficients fluctuate narrowly about the binomial values in (1).
[7] M. R. Hestenes and E. Stiefel, J. Res. Natl. Bur. Stand. **49**, 409 (1952).
[8] S. Baroni, P. Giannozzi, and A. Testa, Phys. Rev. Lett. **58**, 1861 (1987).
[9] X. Gonze and J.-P. Vigneron, Phys. Rev. B **39**, 13120 (1989).
[10] P. Giannozzi and S. de Gironceli, Phys. Rev. B **43**, 7231 (1991).
[11] D. Kh. Blat, N. E. Zein, and V. I. Zinenko, J. Phys. Condens. Matter **3**, 5515 (1991).
[12] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, England, 1989), p. 431.
[13] G. Dolling, *Inelastic Scattering of Neutrons in Solids and Liquids* (IAEA, Trieste, 1963), Vol. II, p. 37.
[14] G. Nilsson and G. Nelin, Phys. Rev. B **6**, 3777 (1972).
[15] While in near perfect agreement with our frozen-phonon calculations, our dynamically determined frequencies are in somewhat worse agreement with experiment due to the relatively low plane-wave cutoff of 8 Ry. For instance, frozen-phonon frequencies for the X_{TO} mode yield 13.21, 13.96, and 14.13 THz at cutoffs of 8, 16, and 32 Ry, respectively, while our dynamic results give 13.15 THz and the experimental value is 13.90 THz.