Challenges for Atomic Scale Modeling in Alternative Gate Stack Engineering

Atsushi Kawamoto, *Student Member, IEEE*, John Jameson, Kyeongjae Cho, and Robert W. Dutton, *Fellow, IEEE*

***Abstract*—We review the challenges for atomic scale modeling of alternative gate dielectric stacks. We begin by highlighting recent achievements of state-of-the-art atomistic simulations of the Si/SiO$_2$ system, showing how such calculations have elucidated the microscopic origins of several important experimental phenomena. For the benefit of readers who may be unfamiliar with the simulation tools, we overview and compare the relevant methods. We then describe the difficulties encountered in extending these approaches to investigate high-k dielectric stacks, pointing out exciting research directions aimed at overcoming these challenges. We conclude by presenting a roadmap of computational goals for atomic scale modeling of alternative gate dielectrics.**

***Index Terms*—Computer aided engineering, dielectric materials, semiconductor device modeling, semiconductor materials.**

I. Background

A. Motivation

THE RAPID scaling of silicon-based MOS devices has led to silicon dioxide (SiO$_2$) gate insulating films less than 2.0-nm thick, representing the most critical dimension in the modern transistor [1]. Recent work at Bell Laboratories has predicted that an SiO$_2$ gate oxide must be at least 1.2-nm thick to function as an insulator, and that given current scaling trends, this fundamental material limit will be reached by the year 2012 [2]. When more practical considerations are taken into account, such as power dissipation due to increasing electron tunneling, it is likely that SiO$_2$ will have to be replaced even sooner. According to the updated 1998 International Technology Roadmap for Semiconductors (ITRS), the 70-nm generation (year 2008) is the first generation for which there are currently no known solutions capable of meeting all gate dielectric requirements [3]. Faced with this challenge, the microelectronics community has been investigating a large number of alternative materials with dielectric constants higher than that of SiO$_2$. These materials would allow for a thicker oxide layer while maintaining sufficient gate capacitance. The rapid introduction of many novel materials resulting from this search is unprecedented in the history of semiconductor devices and presents a new challenge for gate stack engineering.

Manuscript received April 17, 2000. The review of this paper was arranged by Editor M. Lundstrom.

A. Kawamoto and R. W. Dutton are with the Department of Electrical Engineering and the Center for Integrated Systems, Stanford University, Stanford, CA 94305 USA (e-mail: kawamoto@stanford.edu).

J. Jameson is with the Department of Applied Physics, Stanford University, Stanford, CA 94305 USA.

K. Cho is with the Department of Mechanical Engineering, Stanford University, Stanford, CA 94305 USA.

Publisher Item Identifier S 0018-9383(00)07769-8.

In the development of Si/SiO$_2$ technology, conventional process simulators like SUPREM were relied upon heavily to reduce the experimental search space. These tools use partial differential equations (PDEs) to interpolate between and extrapolate beyond experimental data points. The success of this approach for Si/SiO$_2$ can largely be attributed to the extensive data and insight gained from many decades of experimental work. Extending this to the development of alternative gate dielectrics is challenging, however, since the time and cost of populating an extensive knowledge database for each novel material system are prohibitive. We believe that the traditional TCAD approach alone may prove inadequate for alternative gate stack modeling. Atomistic simulations are a promising way to fill the knowledge voids by screening potential dielectrics and predicting their microscopic behavior.

To further motivate our discussion, we briefly mention two promising dielectrics aimed at meeting the $<$15 Å equivalent t_{ox} requirement for the 70–nm generation of the ITRS. Both materials are thermodynamically stable in direct contact with silicon, so that a barrier layer is not needed. Wilk and Wallace deposited an amorphous layer of hafnium silicate (HfSi$_x$O$_y$) directly on silicon and achieved $t_{ox,\,eq}$ of 18 Å [4]. McKee *et al.* epitaxially grew a ferroelectric strontium titanate (SrTiO$_3$) crystal on silicon and achieved $t_{ox,\,eq}$ of 10 Å [5]. The distinction between the amorphous and crystalline structure of these two materials has important implications from the modeling point of view. Our intent is to elucidate such issues and to highlight the atomistic simulation techniques aimed at overcoming the challenges of alternative gate stack engineering.

B. Recent Achievements of Atomistic Simulation in Promoting Understanding of Si/SiO$_2$

Over the past decade, significant advances in so-called first principles methods (described in more detail in Section II), together with phenomenal increases in computing power, have enabled accurate, atomistic descriptions of complex materials with no empirical fitting. Naturally, the Si/SiO$_2$ system has received considerable attention, and the microscopic origins of several challenging problems have recently been explained. We briefly review the key results.

Long-term dielectric reliability continues to be a critical concern for both memory and logic devices, and many models have been proposed to explain the physical origins of time- dependent dielectric breakdown (TDDB). Charge traps are known to be precursors to oxide breakdown, but their physical origin is an intensely debated issue. Some models, such as the (1/E) thermochemical model for breakdown, assume that the oxygen vacancy is the primary precursor defect in the intrinsic breakdown

0018–9383/00$10.00 © 2000 IEEE

mode [6]. Recent first principles studies provided support for this view, confirming that the hole-trapped oxygen vacancy in bulk SiO_2 can serve as an efficient electron trap [7].

Numerous experimental studies have also focused on the role of hydrogen as a charge trap in bulk SiO_2. Yokozawa and Miyamoto quantified these observations by calculating the relative stability of charged and neutral states of hydrogen (e.g., H^+, H^0, H^-) in SiO_2 as a function of the chemical potential [8]. They determined that the charged states were always lower in energy than the neutral state. In addition, they computed two quantities which enable further modeling and experiments: 1) the energy level within the SiO_2 band gap introduced by each state, which may be input to PDE-based tunneling models that explicitly treat charge trapping and detrapping mechanisms, and 2) the vibrational frequencies of resulting Si–OH^+ and Si–H^- bonds, which were found to be sufficiently different from those of neutral Si–OH and Si–H bonds that detection of such states via Fourier transform infrared spectroscopy (FTIR) is possible. More recently, Blochl and Stathis studied the structure and energetics of various complexes of hydrogen and oxygen vacancies in silica using first principles methods [9]. By considering each defect's charge-state level and shift upon relaxation, they found that the neutral hydrogen bridge, a complex in which a hydrogen atom replaces an oxygen atom, was the most likely defect responsible for stress-induced leakage currents (SILC) in gate oxides at low voltages.

The dissociation of Si–H bonds at the Si/SiO_2 interface is also an important reliability issue, and Hess *et al.* have recently shown that dramatic improvements can be obtained by annealing with deuterium instead of hydrogen [10]. While this isotope effect was initially discovered through scanning tunneling microscope (STM) tip experiments, the underlying mechanism was explained computationally based on dynamic simulations of the Si–H and Si–D vibrational modes.

II. Computational Methods

A. Overview of Atomic Scale Modeling Methods

Conventional TCAD tools rely on a PDE-based continuum description which only depends on the average properties of materials. In contrast, for atomistic modeling, an interaction rule between all atoms in the simulated system must be established. Table I compares the three major approaches, empirical interatomic potentials, density functional theory (DFT) and tight binding (TB).

TABLE I
Simplified Classification of Atomic Scale Computational Models

	Empirical interatomic potential	Density functional theory (DFT)	Tight binding (TB)
classical or quantum model?	classical	quantum	quantum
fitted parameters?	yes	no	yes
electronic structure? (bands, gap states)	no	yes	yes
basis set in expansion of wave function	n/a	complete; plane waves	minimal; atomic orbitals
iterate to self-consistency?	n/a	yes	no
computational cost relative to interatomic potential	1	10^5	10^2-10^3

Interatomic potentials are classical models which do not explicitly take into account the role of electrons. The interaction energy between atoms is evaluated using an empirically determined expression with fitted parameters. Because they are computationally efficient, interatomic potentials have been widely used to simulate the collective behavior of large numbers of atoms, notably in ion implantation damage modeling [11].

DFT refers to a specific technique used to solve the Schroedinger equation, the governing equation of quantum mechanics. Quantum mechanics is the fundamental description of matter at the atomic scale and predicts macroscopic material properties almost exclusively in terms of the interactions of electrons. Within this framework, the electronic structure of a material is represented by a wave function Ψ. To determine Ψ, one must solve the many-electron Schroedinger equation

$$\left[\frac{h^{-2}}{2m}\nabla^2 + V_{II} + V_{Ie} + V_{ee}\right]\Psi(\vec{r}_1, \cdots, \vec{r}_n) = \varepsilon\Psi(\vec{r}_1, \cdots, \vec{r}_n). \quad (1)$$

The first term on the left represents kinetic energy and V_{II}, V_{Ie} and V_{ee} denote the potential energy arising from ion-ion, ion-electron and electron-electron interactions, respectively. Most computational techniques expand Ψ in a predefined set of basis functions. Solving for Ψ is then analogous to determining the degree to which each basis function contributes to the overall wave function. The accuracy and cost increase as the set of basis functions is made more complete.

Direct solutions of the many-body Schroedinger equation are computationally intractable due to the large number of free variables. The coupled electron-electron interactions also prevent a direct separation of variables approach. Density functional theory (DFT), which provides the foundation for the majority of modern first principles methods, provides a rigorous way to decouple the electron-electron interactions. By reformulating the problem in terms of the ground-state electron density, DFT allows one to replace (1) with an exactly equivalent set of independent one-electron Schroedinger equations (referred to as the Kohn–Sham equations) [12], [13],

$$\left[\frac{h^{-2}}{2m}\nabla^2 + V'_{II} + V'_{Ie} + V'_{ee}\right]\psi_i(\vec{r}) = \varepsilon\psi(\vec{r}); \quad i = 1, \cdots, N. \quad (2)$$

The index i refers to different eigenstates of the solution. There are now N such equations for each electron in the system, and the primed potential energy terms depend on the ground state charge density,

$$n(\vec{r}) = \sum_i |\psi_i(\vec{r})|^2. \quad (3)$$

Iterative solutions become necessary, since the eigensolution ψ_i needs to be self-consistent with the V' potentials which

depend on ψ_i through $n(\vec{r})$. The local density approximation (LDA), used to evaluate the so-called exchange-correlation energy of the electron-electron interaction, is the only approximation made. Pseudopotentials are used to evaluate the ion-electron interaction and a complete plane wave basis is used to expand the wave function. Early implementations iteratively diagonalized the set of N (2) until self-consistency was achieved, a procedure which scaled as O(N^3), making it prohibitively expensive to study more than a few tens of atoms. More recent approaches overcome the limitations of direct diagonalization by recasting the problem into an iterative search for the lowest energy configuration. Modern implementations include the Car–Parrinello simulated annealing scheme and the more efficient methods based on conjugate gradient energy minimization [14]. These methods typically scale as O(N^2), since they are dominated by the cost of computing the fast Fourier transform (FFT), which scales as Nlog(N), and by wave function orthonormalization, which scales as O(N^3).

The tight binding (TB) approach significantly reduces the cost of solving the single-electron Schroedinger equations [(2)] by making several key approximations. First, a minimal basis set of valence atomic orbitals ϕ_i is used to expand the wave function for each of N electrons in the system, so that typically 10 to 20 times fewer basis functions are used in TB compared to DFT. Second, self-consistency in the eigensolution is neglected so that iterative solutions are no longer required. Last, all interactions are parameterized. The conventional approach diagonalizes the set of N parameterized (2). Pioneering work by Harrison has led to a set of universal TB parameters for every element on the periodic table, assuming orthogonality between neighboring atomic orbitals [15]. There have been extensions to Harrison's work which use a more general nonorthogonal basis and achieve better accuracy for specific systems, at the cost of additional empirical parameters which need to be fit to experimental results [16]. TB generally scales as O(N^3).

B. *Critique of Atomic Scale Modeling Methods*

Almost every physically observable quantity is related to the total energy of a system or to differences in the total energy. The utility of atomic scale calculations lies in their ability to calculate the total energy, and the resulting accuracy is the key criterion used to compare various methods. For the gate stack application, the total energy approach enables us to determine such quantities as band structures, dielectric constants, vibrational frequencies and defect energy levels, as well as defect creation energies, impurity binding energies and activation barriers. In general, the accuracy of a method is directly related to its computational expense. This is the second key criterion for comparing simulation approaches, and it is important to consider what ratio of accuracy/cost is sufficient to meet the goals of a given simulation study.

Of the three approaches to computing the total energy of a system, DFT provides the highest level of accuracy. It is generally observed that DFT predicts basic ground state structural properties such as lattice constant and bulk modulus to within a few percent of the experimental values, with no parameter fitting [17]. One notable exception to the accuracy of DFT is prediction of excited state energies such as the bandgap, which is typically underestimated by as much as a factor of two, owing to the LDA approximation. This is an important limitation, since many of the properties we wish to investigate in the gate stack application, such as mid-gap defect energy levels or band offsets at electrode/dielectric/Si interfaces, require accurate knowledge of the bandgap. The so-called quasiparticle GW approach to the electron self-energy is the best known correction to this problem and can accurately predict the bandgap and other excited state properties for a variety of semiconductors and insulators, including transition metal oxides [18], [19]. However, since the GW calculation is considerably more expensive, it should be used in a complementary manner to correct the DFT-LDA results for the most important configurations.

The ability to predict the static (i.e., low frequency) dielectric constant of novel materials is also important for the high-k gate stack application. Some first principles approaches include: direct evaluation of the dielectric function $\varepsilon(\omega)$ based on the band-to-band optical absorption spectrum [20]; calculation of the dielectric tensor based on the modern theory of polarization [21]; and calculation of the dielectric tensor by density functional perturbation theory (DFPT), which considers the response to an external perturbation (e.g., electric field) variationally in a self-consistent manner [22], [23]. To our knowledge, variational DFPT is the only approach that has been demonstrated to accurately predict the static dielectric constant of polar insulating oxides such as SiO_2, in which there are both significant electronic and ionic responses to the external electric field. To varying degrees of accuracy relative to computational expense, these methods can be used to investigate the behavior of the dielectric constant as such factors as film composition, stress and defect states are altered in the high-k material of interest.

The major drawback of the DFT approach is its intense computational cost, which severely limits the size of the system which can be simulated. Conventional implementations of DFT theory scale as O(N^2). O(N) methods are also possible, though the difficulty of implementing such a program has prevented widespread use. A survey of the literature indicates that O(N^2) DFT simulations involving metal oxides are currently limited to about 100 atoms. Fig. 1 shows how this limit is expected to increase over the next 15 years, assuming a factor of 1.5 increase in parallel computing capacity each year [24][1].

The accuracy with which the total energy can be predicted in a TB simulation depends strongly on the formulation being used and the system being modeled. Some TB schemes achieve an accuracy comparable to that of DFT for a specific material, but do so at the cost of employing extensive parameter fitting. For example, Kwong *et al.*'s model for Si provides a particularly good description of point defect creation energies and has been used to investigate interstitial-vacancy recombination in transient enhanced diffusion (TED) studies [25], [26]. However, because it uses more than 20 fitting parameters specific to the Si system, there is no straightforward way to extend the model to describe alternative dielectrics like $HfSi_xO_y$. Others, such as Harrison's universal TB formulation and its extensions can be applied to

[1]This is a more conservative estimate than the factor of 2.15 targeted by the DOE Accelerated Strategic Computing Initiative (ASCI), which aims to drive the development of some of the world's most powerful supercomputers.

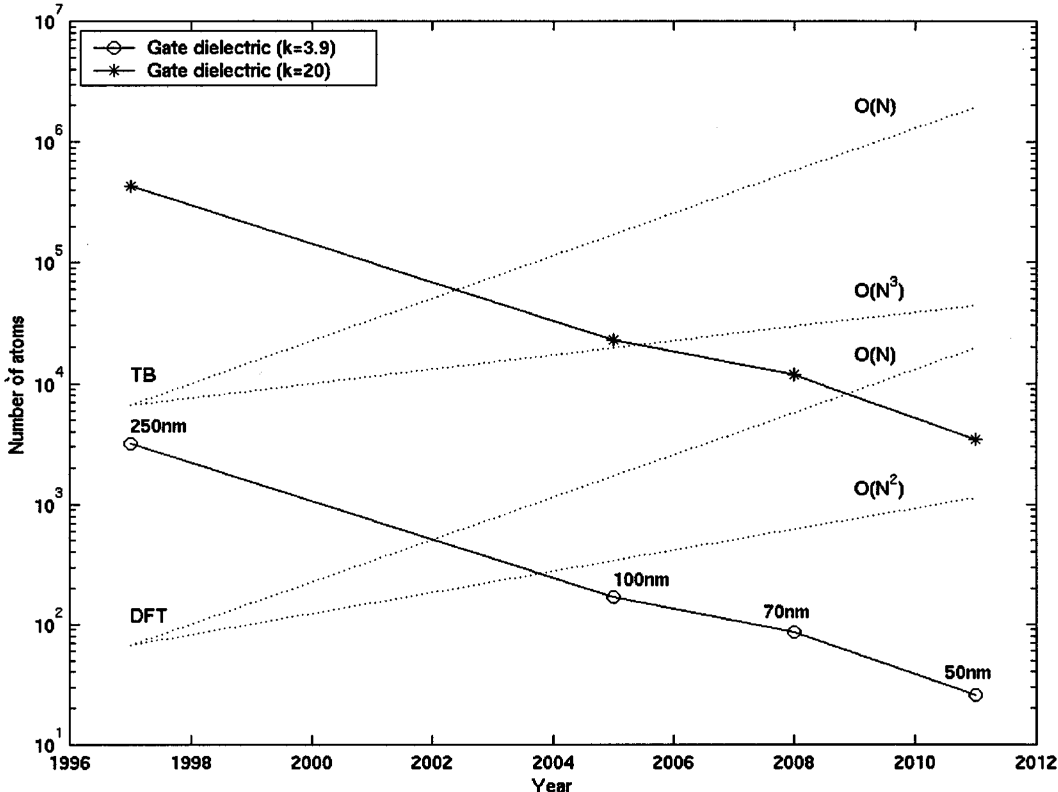

Fig. 1. Extrapolated long-term trend of DFT and TB parallel computing capacity for dielectric materials vs ITRS technology nodes.

any material system, since only a few fundamental parameters are used to model the entire periodic table. However, since it achieves its universality by avoiding extensive empirical fitting for specific systems, the universal TB method gives predictions which are generally only qualitatively accurate. Therefore, there is an important tradeoff between extensibility and accuracy in TB models. Furthermore, due to the many approximations made in basic TB theory, its predictions are generally not as accurate as those of DFT. The upshot is that TB is roughly 10^2-10^3 times less expensive than DFT (see Fig. 1). O(N) scaling methods for TB are more widely implemented than for DFT.

Due to the high computational cost of a brute-force approach using DFT-based methods alone, we believe that achieving useful results for novel, complex dielectric materials will require complementary use of DFT and less expensive methods such as TB. While not a quantitatively accurate theory, universal TB predicts the correct trends for a surprisingly large number of physical properties related to the electronic structure of a material, often at minimal computational expense [27]. It can be used to provide initial insight into the nature of bonding and other aspects of a material's electronic structure, helping to reduce the computational search space before more focused DFT calculations are performed. Indeed, many references in the literature have employed TB as an initial tool to uncover the electronic structure before carrying out more complete DFT simulations and as a test bed for exploring new electronic structure calculations (e.g., [28]–[30]).

We have avoided extended discussion of empirical interatomic potentials because we feel they have limited use in alternative dielectric studies. The extensive parameter fitting required to formulate an empirical potential is costly and time-consuming, and each model is generally only reliable for properties included in the fitting database. Three of the most popular models for Si, environment dependent interatomic potential (EDIP), Stillinger–Weber (SW), and Tersoff (T3), demonstrate this shortcoming [31]–[33]. While the EDIP model provides the best description of the unrelaxed vacancy, the T3 potential provides the best description of the unrelaxed tetrahedral interstitial. Like the extensively fitted TB potentials, there is no straightforward way to extend an empirical potential formulation and parameter set extracted for a particular system to alternative metal oxides.

III. Challenges for Gate Stack Modeling

A. Material Interfaces

The dominance of Si over other materials can ultimately be attributed to the near-perfect interface between Si and SiO_2. Any alternative dielectric material that hopes to replace SiO_2 must achieve a similarly high degree of perfection in its interface with Si. Many of the important integration concerns, such as impurity segregation and interface recrystallization, require modeling of the dielectric/Si and dielectric/electrode interfaces. Yet the computational physics community has generally focused on bulk properties of high-k materials (e.g., ferroelectrics), motivated by the novel science of such materials. From the engineering point of view, we must build on these studies by also investigating the technology concerns at material interfaces. However,

unlike bulk systems, interfaces pose significant new challenges for atomistic modeling. Many more atoms need to be included in an interface model compared to a bulk structure. Consequently, realistic modeling of high-k dielectric interfaces becomes extremely expensive, pushing the limits of state-of-the-art first principles methods.

Another difficulty is that if the dielectric is grown or deposited as an amorphous layer (e.g. $HfSi_xO_y$), the resulting bulk and interface structures are not unique. The positions of atoms within the amorphous layer and at the interface must be determined through experiment or deduced in some other manner. Instead of trying to determine a reasonable amorphous structure, most DFT studies of Si/SiO_2 have used crystalline forms of bulk oxide or have attached crystalline oxide to the silicon substrate [34], [35]. These techniques have led to close agreement between calculated results and experimental measurements of the true amorphous system (e.g., Si-O bond lengths and angles), but it remains to be seen whether a similar approach can be justified for alternative dielectric systems. We expect, on the other hand, that attaching two crystals should work well for modeling dielectrics which are grown as lattice-matched epitaxial crystals (e.g., $SrTiO_3$).

A promising way to reduce the uncertainty associated with the interface structure and to gain insight on the growth mechanism itself is to model the actual oxidation or deposition process atomistically. A recent study has used DFT molecular dynamics simulations to observe oxidation of the first few monolayers of silicon [36]. One key observation from the study was that oxidation tends to proceed via an unexpected excess of silicon atoms at the interface and a transient configuration of triply-bonded oxygen atoms, both of which help to prevent the formation of dangling bonds. As explained more completely in the following section, such dynamical simulations are very expensive, and it may still be some time before such investigations become commonplace. We believe that employing both TB and DFT methods in a complementary way may provide a tractable way to generate reasonable amorphous structures. For example, an amorphous-crystalline interface could be generated by a two-step process in which a first-order model is created using a high temperature TB anneal of the dielectric/Si system. The resulting geometry could then be accurately relaxed to a local minimum configuration using DFT. Indeed, Ng and Vanderbilt have demonstrated such a two-step approach for the Si/SiO_2 interface by first annealing an abrupt interface model with an empirical interatomic potential and then relaxing the structure by DFT [37]. Their resulting model reproduced many of the experimentally observed characteristics of the interface, such as the distribution of suboxide Si species, and provides encouragement that such an approach may also work for high-k dielectric/Si interfaces.

B. Time-Scale Problem

DFT and TB describe the interaction of atoms in a system (i.e., forces and energy) for a given static configuration. In reality, the atoms of a system are in constant motion and many quantities of interest are products of this motion. The two main computational methods used to evolve a system in time are molecular dynamics (MD) and Monte Carlo (MC). By design, the MD method reproduces the short-time behavior of the system ($<$ approx. 1 ns), while MC techniques are geared toward problems occurring on macroscopic time scales. The enormous gap that exists between these time scales, sometimes referred to as the rare event regime, poses a significant challenge to the future development of atomistic simulations.

The phenomena occurring in the rare event regime are characterized by large atomic displacements induced by the natural thermal excitations within the system. On an atomic time scale, thermally excited events occur so infrequently that it is often not feasible, even with powerful supercomputers, to witness them during the course of an MD simulation. This is the so-called time scale problem of MD. For gate stack modeling, overcoming this problem is of fundamental importance because the atomic processes that occur within the rare event regime are exactly those that underlie many of the important integration and reliability concerns, such as diffusion of transition metals into the underlying substrate or the formation of oxygen vacancy related defects. The key physical attribute of all these processes is the presence of energy barriers that must be overcome in order for the processes to occur.

While no simulation technique for overcoming the time scale problem has been perfected, the seeds for a solution have been planted. One important contribution is Voter's kinetic Monte Carlo (KMC) algorithm [38]. The central feature of this method is an event catalog which lists the important thermally activated processes in the system and the rates at which they occur. During a KMC simulation, events are randomly drawn from the event catalog and applied to the system in such a way as to produce an exact, average evolution of the system. KMC simulations are computationally inexpensive and allow systems to be simulated over very long periods of time.

The primary difficulty with the KMC method is generating the event catalog; the important thermal processes are not always clear *a priori*, and discovering them can be very difficult and time consuming. As a result, it is common to accelerate the dynamics of a system by simply increasing its temperature. Within a simulation, this has exactly the same effect on a system as a high temperature cycle does on a wafer. However, there is a fundamental problem with this approach—the rates at which different processes occur in a system do not scale linearly with temperature. As a result, if processes A and B are characterized by rates k_A and k_B at temperature T_1, such as $k_A < k_B$, then at temperature $T_2 > T_1$, we may find that $k_B < k_A$. The dynamical evolution of the system at T_1 and T_2 could thus be fundamentally different. This is a shortcoming of the DFT-MD oxidation simulation described in Section A, since an unphysically high temperature gradient was the driving mechanism behind the diffusion of oxygen.

Recently, three new accelerated dynamics (AD) methods have been presented which, on paper, avoid the problems associated with both KMC and the crude technique of temperature elevation [39]–[41]. While the expected speed-up produced by these techniques depends strongly on the material and phenomena being modeled, improvements on the order of 10^2 to 10^6 appear possible. However, AD techniques have been developed within the domain of computational physics, so, to

TABLE II
ROADMAP OF COMPUTATIONAL GOALS FOR ATOMIC SCALE MODELING OF ALTERNATIVE GATE DIELECTRICS

Modeling goals / issues	TB	DFT-LDA
improve simulation efficiency and capacity	parallelization via parallel matrix and FFT routines	
	reformulate problem to achieve linear scaling / O(N) [44]	
	accelerate time scale (e.g. Voter KMC and AD schemes) [38-41]	
process simulation	reduce computational search space via two-step approach	compute activation energies, diffusion coefficients, etc. as input to PDE process simulator
device simulation	directly investigate I-V characteristics by applying electric fields [42,43]	indirectly investigate I-V characteristics by computing band structure, effective mass, defect energy levels in bandgap, etc. as input to PDE device simulator

Legend	Shading
Solutions highly developed and utilized	white
Solutions developed but under-implemented or under-utilized	gray
Solutions still immature or unknown	dark

date, they have been applied only to rather simple problems of surface and bulk diffusion. It is still an open question as to whether their power will carry over to the complex materials and interfaces that need to be modeled for alternative gate stack engineering. Nonetheless, the potential payoff is huge in terms of our computational capability of modeling and studying basic atomic phenomena.

C. Device Simulation

Both the DFT and TB methods, as they are commonly implemented, do not take into account the effect of an externally applied electric field. For this reason, their application to device simulation is limited to the indirect role of providing fundamental material properties, such as band structure and defect energy levels, as input parameters to a PDE-solver such as PISCES. However, there exists a formal theoretical framework within which time-dependent electric and magnetic fields can be applied to any system described by a TB potential [42]. Such an approach would allow direct investigation of atomic scale effects on resulting current–voltage (*I–V*) characteristics and could provide valuable insight into the physical mechanisms underlying dynamic processes such as charge trapping/detrapping. We are aware of at least one group pursuing this approach for SiO_2 gate stack simulation [43].

D. Scaling Trends for Gate Stack Modeling

We have strong reason to be optimistic as we look forward to the role atomistic simulations will play in supporting the continued scaling of microelectronics into the next decade. We refer again to Fig. 1, where the labeled points correspond to the number of atoms in a simulation cell given by $(t_{ox,\,equivalent})^3$ for the 250-nm, 100-nm, 70-nm, and 50-nm technology generations of the ITRS, assuming 20 Å^3/atom. The two solid lines correspond to SiO_2 $(k = 3.9)$ and an alternative dielectric $(k = 20)$. Several key trends are clear. First, there are important crossover points at which the entire critical dimension of the alternative gate dielectric can be modeled atomistically. While modeling all atoms in the t_{ox} dimension may not be necessary in general, specific effects, such as the tunneling *I–V* resulting from application of an electric field across a TB gate stack or impurity diffusion from the gate electrode into the substrate, require it. We thus believe that modeling the entire critical dimension of the gate dielectric is a reasonable computational goal. At the same time, since this capability needs to be in place well before the 70-nm generation goes to production, the use of DFT alone will likely be inadequate. We also note that the O(N^2) (O(N^3)) scaling of conventional DFT (TB) methods places a significant limitation on the long-term modeling capacity. Exciting progress has been made by the physics community to improve these techniques to O(N) scaling, and we must adopt such methods to bring the crossover points as close as possible [44].

IV. CONCLUSIONS

The introduction of many novel materials for alternative gate dielectrics presents a fundamental new challenge to modeling. Because of its accuracy and extensibility, DFT has been the workhorse atomistic simulation for Si/SiO_2 technology, and it will likely remain the workhorse for the next decade. At the same time, we believe that the requirement to rapidly investigate many new materials places a limitation on brute-force approaches employing DFT-based methods alone. By using less rigorous methods such as universal TB as complementary tools for preliminary, qualitative investigations, we feel that a reasonable compromise between accuracy and cost can be achieved to meet the needs of a given study. We believe that the dependence of empirical potentials on extensive parameter fitting limits their usefulness.

We conclude by presenting a roadmap of computational goals for atomic scale modeling of alternative gate dielectrics in Table II. Increasing collaboration between the engineering and

physics communities will be a key ingredient to realizing many of the goals presented by the roadmap, such as achieving O(N) scaling, applying electric field terms and accelerating the time scale. From the application side, users will increasingly need to employ the tools in complementary ways to achieve the goals of alternative gate stack engineering. Atomic scale modeling tools rely on new domains of physics such as quantum mechanics that go beyond the foundations of traditional PDE-based TCAD tools. Yet the engineering approach which has enabled so much progress in microelectronics for the past two decades—carefully considering both the benefits and drawbacks of a given method relative to the requirements of the application—must still be employed to ensure future progress. It is our hope that atomic scale modeling approaches, used along with traditional tools, may finally evolve the role of TCAD from one of fitting and explaining experiments to actually predicting them.

REFERENCES

[1] P. Packan, "Pushing the limits," *Science*, vol. 285, p. 2079, 1999.
[2] D. Muller *et al.*, "The electronic structure at the atomic scale of ultrathin gate oxides," *Nature*, vol. 399, p. 758, 1999.
[3] Semiconduct. Res. Corp., http://www.src.org/about/itrs_update.dgw.
[4] G. Wilk and R. Wallace, "Electrical properties of hafnium silicate gate dielectrics deposited directly on silicon," *Appl. Phys. Lett.*, vol. 74, p. 2854, 1999.
[5] R. McKee, F. Walker, and M. Chisholm, "Crystalline oxides on silicon: The first five monolayers," *Phys. Rev. Lett.*, vol. 81, p. 3014, 1998.
[6] J. McPherson and H. Mogul, "Underlying physics of the thermochemical E model in describing low-field time-dependent dielectric breakdown in SiO_2 thin films," *J. Appl. Phys.*, vol. 84, p. 1513, 1998.
[7] A. Yokozawa, A. Oshiyama, Y. Miyamoto, and S. Kumashiro, "Oxygen vacancy with large lattice distortion as an origin of leakage currents in SiO_2," in *IEDM Tech. Dig.*, 1997, p. 703.
[8] A. Yokozawa and Y. Miyamoto, "First-principles calculations for charged states of hydrogen atoms in SiO_2," *Phys. Rev. B*, vol. 55, p. 13 783, 1997.
[9] P. Blöchl and J. Stathis, "Hydrogen electrochemistry and stress-induced leakage currents in silica," *Phys. Rev. Lett.*, vol. 83, p. 372, 1999.
[10] K. Hess *et al.*, "Impact of nanostructure research on conventional solid-state electronics: The giant isotope effect in hydrogen desorption and CMOS lifetime," *Physica E*, vol. 3, p. 1, 1998.
[11] G. Gilmer, T. de la Rubia, D. Stock, and M. Jaraiz, "Diffusion and interactions of point defects in silicon: Molecular dynamics simulations," *Nucl. Instrum. Meth. Phys. Res. B*, vol. 102, p. 247, 1995.
[12] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev.*, vol. 136, p. B864, 1964.
[13] W. Kohn and L. Sham, "Self-consistent equations including exchange and correlation effects," *Phys. Rev.*, vol. 140, p. A1133, 1965.
[14] M. Payne *et al.*, "Iterative minimization techniques for ab initio total-energy calculations: Molecular dynamics and conjugate gradients," *Rev. Mod. Phys.*, vol. 64, p. 1045, 1992.
[15] W. Harrison, *Electronic Structure and the Properties of Solids*. San Francisco, CA: Freeman, 1980.
[16] M. Menon and K. Subbaswamy, "Universal parameter tight-binding molecular dynamics: Application to C_{60}," *Phys. Rev. Lett.*, vol. 67, p. 3487, 1991.
[17] M. Yin and M. Cohen, "Theory of static structural properties, crystal stability and phase transformations: Application to Si and Ge," *Phys. Rev. B*, vol. 26, p. 5668.
[18] S. Louie, "Quasiparticle theory of electron excitations in solids," in *Quantum Theory of Real Materials*, J. Chelikowsky and S. Louie, Eds. Boston, MA: Kluwer, 1996, p. 83.
[19] B. Kralik, E. Chang, and S. Louie, "Structural properties and quasiparticle band structure of zirconia," *Phys. Rev. B*, vol. 57, p. 7027, 1998.
[20] Y. Xu and W. Ching, "Electronic and optical properties of all polymorphic forms of silicon dioxide," *Phys. Rev. B*, vol. 44, p. 11 048, 1991.
[21] F. Bernardini, V. Fiorentini, and D. Vanderbilt, "Polarization-based calculation of the dielectric tensor of polar crystals," *Phys. Rev. Lett.*, vol. 79, p. 3958, 1997.
[22] X. Gonze and C. Lee, "Dynamical matrices, born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory," *Phys. Rev. B*, vol. 55, p. 10 355, 1997.
[23] X. Gonze, D. Allan, and M. Teter, "Dielectric tensor, effective charges, and phonons in α-quartz by variational density-functional perturbation theory," *Phys. Rev. Lett.*, vol. 68, p. 3603, 1992.
[24] Rep., http://www.llnl.gov/asci/.
[25] I. Kwon *et al.*, "Transferable tight-binding models for silicon," *Phys. Rev. B*, vol. 49, p. 7242, 1994.
[26] M. Tang, L. Colombo, J. Zhu, and T. de la Rubia, "Intrinsic point defects in crystalline silicon: Tight-binding molecular dynamics studies of self-diffusion, interstitial-vacancy recombination, and formation volumes," *Phys. Rev. B*, vol. 55, p. 14 279, 1997.
[27] W. Harrison, *Elementary Electronic Structure*. Singapore: World Scientific, 1999.
[28] S. Ögüt and K. Rabe, "Band gap and stability in the ternary intermetallic compounds NiSnM (M = Ti, Zr, Hf): A first-principles study," *Phys. Rev. B*, vol. 51, p. 10 443, 1995.
[29] R. Nunes and D. Vanderbilt, "Real-space approach to calculation of electric polarization and dielectric constants," *Phys. Rev. Lett.*, vol. 73, p. 712, 1994.
[30] J. Bennetto and D. Vanderbilt, "Semiconductor effective charges from tight-binding theory," *Phys. Rev. B*, vol. 53, p. 15 417, 1996.
[31] J. Justo *et al.*, "Interatomic potential for silicon defects and disordered phases," *Phys. Rev. B*, vol. 58, p. 2539, 1998.
[32] J. Tersoff, "Empirical interatomic potential for silicon with improved elastic properties," *Phys. Rev. B*, vol. 38, p. 9902, 1988.
[33] F. Stillinger and T. Weber, "Computer simulation of local order in condensed phases of silicon," *Phys. Rev. B*, vol. 31, p. 5262, 1985.
[34] A. Pasquarello, M. Hybertsen, and R. Car, "Theory of Si 2p core-level shifts at the Si(001)-SiO_2 interface," *Phys. Rev. B*, vol. 53, p. 10 942, 1996.
[35] M. Hane, Y. Miyamoto, and A. Oshiyama, "Atomic and electronic structures of an interface between silicon and β-cristobalite," *Phys. Rev. B*, vol. 41, p. 12 637, 1990.
[36] A. Pasquarello, M. Hybertsen, and R. Car, "Interface structure between silicon and its oxide by first-principles molecular dynamics," *Nature*, vol. 396, p. 58, 1998.
[37] K. Ng and D. Vanderbilt, "Structure and oxidation kinetics of the Si(100)–SiO_2 interface," *Phys. Rev. B*, vol. 59, p. 10 132, 1999.
[38] A. Voter, "Classically exact overlayer dynamics: diffusion of rhodium clusters on Rh(100)," *Phys. Rev. B.*, vol. 34, p. 6819, 1986.
[39] ——, "Parallel replica method for dynamics of infrequent events," *Phys. Rev. B*, vol. 57, p. R13985, 1998.
[40] ——, "A method for accelerating the molecular dynamics simulation of infrequent events," *J. Chem. Phys.*, vol. 106, p. 4665, 1997.
[41] A. Voter and M. Sorensen, "Accelerating atomistic simulations of defect dynamics: Hyperdynamics, parallel replica dynamics and temperature-accelerated dynamics," in *Proc. Mat. Res. Soc. Symp.*, vol. 538, 1999, p. 427.
[42] M. Graf and P. Vogl, "Electromagnetic fields and dielectric response in empirical tight-binding theory," *Phys. Rev. B*, vol. 51, p. 4940, 1995.
[43] "Private communication," unpublished.
[44] S. Goedecker, "Linear scaling electronic structure methods," *Rev. Mod. Phys.*, vol. 71, p. 1085, 1999.

Atsushi Kawamoto (S'00) received the B.S. and M.S. degrees in electrical engineering from Stanford University, Stanford, CA, in 1997 and 1998, respectively. He is currently pursuing the Ph.D. degree in electrical engineering at Stanford, and is co-advised by Prof. R. W. Dutton and Dr. P. Griffin.

His current research focuses on the use of first principles methods based on density functional theory (DFT) to computationally investigate alternative high-k dielectric materials for scaled MOS gate stacks. His research is funded by the National Science Foundation through a Graduate Research Fellowship and the DesCArtES (Distributed Center for Advanced Electronics Simulations) project.

Mr. Kawamoto was elected to Phi Beta Kappa and Tau Beta Pi and received the Frederick E. Terman Award from the School of Engineering at Stanford.

John Jameson received the B.A. degree in mathematics from the College of Wooster, Wooster, OH, in 1996 and the B.S. degree in aeronautical engineering from Purdue University, West Lafayette, IN, in 1997. He entered the M.S. program in aeronautical engineering in 1997 at Stanford University, Stanford, CA, as a Stanford Graduate Fellow, and is now pursuing the Ph.D. degree in applied physics.

His current research focuses on the development of algorithms for atomistic simulation, including universal tight binding with electromagnetism and algorithms for rare-event processes (e.g., diffusion). Applications of his research include thin film mechanics and high-k dielectrics.

Kyeongjae (KJ) Cho received the B.S. degree in 1986 and the M.S. degree in 1988, both from Seoul National University, Seoul, Korea, and the Ph.D. degree from the Massachusetts Institute of Technology (MIT), Cambridge, in 1994. He pursued post-doctoral training at MIT from 1994 to 1995.

He joined the Engineering Faculty, Stanford University, Stanford, CA, in 1997 as an Assistant Professor and Frederick E. Terman Fellow with the Department of Mechanical Engineering. Prior to arriving at Stanford, he was a Research Scientist at the MIT Research Laboratory of Electronics (1995–1997) and a Research Associate with the Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA (1995–1997). His current research focuses on the realization of computational materials engineering—multiscale simulations based on the seamless integration of quantum, atomistic, and continuum mechanics. Since arriving at Stanford in 1997, he has organized the Multiscale Simulation Laboratory, which currently has two major research focus areas: high-k dielectric materials and nanotechnology. He has published about 40 journal articles.He holds two U.S. patents.

Robert W. Dutton (S'67–M'70–SM'80–F'84) received the B.S., M.S., and Ph.D. degrees from the University of California, Berkeley, in 1966, 1967, and 1970, respectively.

He is Professor of electrical engineering at Stanford University and Director of Research in the Center for Integrated Systems. He has held summer staff positions at Fairchild, Bell Telephone Laboratories, Hewlett-Packard, IBM Research, and Matsushita during 1967, 1973, 1975, 1977, and 1988, respectively. His research interests focus on IC process, device, and circuit technologies–especially the use of CAD and parallel computational methods. He has published more than 200 journal articles and graduated more than four dozen doctorate students.

Dr. Dutton was Editor of the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN (1984–1986). He was the winner of the 1987 IEEE J. J. Ebers Award, the 1988 Guggenheim Fellowship to study in Japan, was elected to the National Academy of Engineering in 1991, and received the Jack A. Morton Award for 1996.