

Kinetic lattice Monte Carlo for point defect diffusion

Andrew P. Horsfield

fecit, 2 Longwalk Road, Stockley Park, Uxbridge, Middlesex UB11 1AB

Scott T. Dunham

Boston University, Department of Electrical and Computer Engineering, 8 St. Mary's Street,

Boston, MA 02215

Hideaki Fujitani

Fujitsu Laboratories Ltd., 10-1 Morinosato-Wakamiya, Atsugi 243-01, Japan

Abstract

We present a self-contained description of a formalism for modeling point defect diffusion in crystalline systems using a Monte Carlo technique. The main approximations required to construct a practical scheme are discussed, with special emphasis on the proper treatment of charged dopants and defects, and one approach to overcoming the problem of fast, but non-diffusive, processes is proposed.

I. INTRODUCTION

There are many processes that take place in materials that occur over time scales long compared with the period of oscillation of an atom. For atomic level studies of materials by means of computer simulation this fact eliminates molecular dynamics (a very general and accurate atomistic method) as a useful tool for studying these phenomena [1]. Thus some other approach must be taken.

For those systems where the phenomena being studied are characterized by rapid single events interspersed by long times of no important activity, stochastic methods become possible. The central argument is that during the long inactive periods, atomic vibrations randomize the positions and momenta of the atoms (within a highly restricted phase space), after which a single important event then occurs. Thus we can treat separate events as independent, and occurring at random. The numerical method for studying such processes is Monte Carlo [1].

In this paper we discuss kinetic lattice Monte Carlo (KLMC) from the perspective of diffusion of point defects (including dopants and impurities) in semiconductors. There are fairly comprehensive discussions of similar methods from the point of view of other problems [2,3], but there appears to be no complete account for defect diffusion. This paper attempts to fill this gap.

There is a long history of statistical methods being applied to these systems. Simple analytic models have been developed for metals [4,5]. The theory was subsequently generalized [6], and continues to be used as an analytical tool [7]. In recent years there has been great interest in implementing full numerical schemes for studying the diffusion of point defects at high enough densities such that non-linear effects become important [8–10].

In the following sections we present: an outline of the underlying statistical mechanical equations and approximations; the additional features required to make the method kinetic; and some extra considerations required to apply this approach to the particular problem of point defect diffusion.

II. STATISTICAL MECHANICS OF KLMC

The following arguments will all be developed within the framework of classical statistical mechanics [11]. The same results can be obtained from quantum mechanical statistical mechanics. The general picture is of atoms vibrating with small amplitudes (compared with the inter-atomic spacing) about fixed sites, and then occasionally (that is, with a time scale much greater than the atomic vibrational period) hopping to new sites. We are thus free to treat the atoms within the harmonic approximation while they vibrate about their lattice sites. Hopping between sites is clearly not accurately described by the harmonic approximation. But since these events are so rare we can neglect them initially.

The classical Hamiltonian [12] (H) is a function of the momenta (represented by \vec{p}) and positions (represented by \vec{r}) of the particles. The probability density for finding the system in the neighborhood of the phase space point (\vec{p}, \vec{r}) is

$$P(\vec{p}, \vec{r}) = \frac{1}{Z} \exp(-\beta H(\vec{p}, \vec{r})), \quad (1)$$

where $\beta = 1/k_B T$, k_B is Boltzmann's constant and T is the temperature. The partition function Z is defined by

$$Z = \int d\vec{p} \int d\vec{r} \exp(-\beta H(\vec{p}, \vec{r})). \quad (2)$$

For the purposes of a Monte Carlo simulation, the momentum is considered to be randomized between hopping events. Since we are not interested in the momentum, we can integrate it out to obtain the probability of the system being in the neighborhood of some set of spatial coordinates. If we assume that the Hamiltonian separates as $H(\vec{p}, \vec{r}) = K(\vec{p}) + V(\vec{r})$, we get the simple result

$$P(\vec{r}) = \frac{1}{Q} \exp(-\beta V(\vec{r})), \quad (3)$$

where the classical configurational integral Q is defined by

$$Q = \int d\vec{r} \exp(-\beta V(\vec{r})). \quad (4)$$

We now introduce the harmonic approximation for the potential energy $V(\vec{r})$ [13]. This allows us to further simplify the equations, and to transform the problem from one involving continuous variables (\vec{r}) to one containing discrete variables. First we define the set of sites about which the atoms vibrate to be \vec{R}_i for some configuration i . These sites correspond to a local energy minimum so when the atoms occupy these sites they experience no force. Let displacements from these sites be given by \vec{u} , so that $\vec{r} = \vec{R}_i + \vec{u}$. Within the harmonic approximation the potential energy can be written as

$$V(\vec{r}) \approx V(\vec{R}_i) + \frac{1}{2} \vec{u} \mathbf{D}_i \vec{u}, \quad (5)$$

where \mathbf{D}_i is the dynamical matrix for configuration i .

The next approximation is to replace the integral over \vec{r} with the combination of a sum over i and an integral over \vec{u} . The justification for this is as follows. Values of \vec{u} that take the system far from the energy minimum lead to large values of energy for the system. These will contribute little to configurational averages provided the temperature is not too high (in which case the harmonic approximation no longer holds). Thus the original integral and the combination of sum and integral will give approximately the same weight to each configuration, and thus produce similar configurational averages quantities. Using this approximation we get

$$P(\vec{R}_i, \vec{u}) = \frac{\exp(-\beta V(\vec{R}_i)) \exp\left(-\frac{\beta}{2} \vec{u} \mathbf{D}_i \vec{u}\right)}{\sum_j \left[\exp(-\beta V(\vec{R}_j)) \int d\vec{v} \exp\left(-\frac{\beta}{2} \vec{v} \mathbf{D}_j \vec{v}\right) \right]}. \quad (6)$$

For Monte Carlo simulations consisting of transitions between discrete configurations, we need the total probability that the system lies in the neighborhood of a specific energy minimum. To obtain this, we integrate Eq. (6) over \vec{u} . If we make the final approximation that the dynamical matrix is independent of configuration (which is not guaranteed to be a good approximation in general), we get the probability that the system is in configuration i to be

$$P(\vec{R}_i) = \frac{\exp(-\beta V(\vec{R}_i))}{\sum_j \exp(-\beta V(\vec{R}_j))}. \quad (7)$$

Equation (7) is the first central result. It is important to notice two things about it. First it involves only energy minima. This results in a huge reduction in the size of the phase space to be considered, greatly simplifying expressions for the energy V . Second, the minima are discrete objects that can be indexed readily.

Two of the above approximations can be relaxed in a straightforward way. First, the partitioning of phase space can be made exact, and is not restricted to the harmonic approximation. In general, we need to define boundaries in phase space that separate neighboring stable configurations (such as saddle points). The integral over \vec{u} then no longer extends over all space, but only over the region inside the boundary for each stable configuration. Second, rather than assuming that all the harmonic terms cancel, we can evaluate

$$\exp(-\beta F_i) = \int d\vec{p} \int d\vec{u} \exp \left[-\beta \left(K(\vec{p}) + V(\vec{R}_i) + \frac{1}{2} \vec{u} \mathbf{D}_i \vec{u} \right) \right], \quad (8)$$

where F_i is a harmonic free energy for configuration i , and then use

$$P(\vec{R}_i) = \frac{\exp(-\beta F_i)}{\sum_j \exp(-\beta F_j)}. \quad (9)$$

This is more accurate, but evaluating F_i is computationally much more demanding than evaluating $V(\vec{R}_i)$. Of course we can also combine these two observations and work with

$$\exp(-\beta F_i) = \int d\vec{p} \int_{\Omega_i} d\vec{u} \exp(-\beta(K(\vec{p}) + V(\vec{R}_i + \vec{u}))), \quad (10)$$

where Ω_i is the domain neighboring configuration i over which \vec{u} is to be integrated.

III. KINETICS OF KLMC

The result obtained above allows us to obtain configurational averages (static quantities), but does not tell us anything about the time evolution of configurations. The time evolution is important for calculating dynamical quantities (such as diffusion coefficients), as well as

providing us with the history of a system (which is important for understanding the effect of annealing, for example).

One way to approach the question of the evolution of configurations is to imagine performing molecular dynamics simulations, and think about the results they would provide. First a structure would be relaxed into a local energy minimum. The atoms would then be given some kinetic energy corresponding to a particular temperature. The system would then evolve according to Newton's Laws until such time as an atom hops from one site to another. This process would be repeated many times with different velocity distributions for the atoms (but with the same temperature in each case). The transition, and time to reach that transition, would be recorded for each simulation. This would give us a table of allowed transitions with an average rate (the average of the inverse of the times) for each one.

Being equipped with such a table for every configuration would allow us to generate a history in the following manner. We start from some particular configuration (call it i). It can make transitions to its neighboring configurations (indexed by j) with rates $W(i \rightarrow j)$. The probability that any one transition will occur within any given short time (much smaller than the time required for the fastest process) is proportional to the rate for that transition. The average time required for a transition to occur is

$$\tau_i = \left[\sum_j W(i \rightarrow j) \right]^{-1}. \quad (11)$$

Thus a transition is chosen at random from the list of allowed transitions with a probability proportional to the rate. This takes us to configuration j . The total time for the simulation is incremented by $\tau_i \ln(1/\xi)$ [3] where ξ is a random number uniformly distributed over the range 0 to 1. The random term introduces fluctuations about the average time that satisfy Poisson statistics. The process is then repeated starting with configuration j , and so on until a complete history has been created.

The prescription given above for evaluating rates is prohibitively expensive in computer time in practice. Thus some much cheaper scheme is required. We can begin by applying

detailed balance [1], which states

$$P(\vec{R}_i)W(i \rightarrow j) = P(\vec{R}_j)W(j \rightarrow i). \quad (12)$$

Combining Eq. (12) with Eq. (7) we obtain

$$\frac{W(i \rightarrow j)}{W(j \rightarrow i)} = \exp\left(-\beta [V(\vec{R}_j) - V(\vec{R}_i)]\right). \quad (13)$$

To obtain absolute rates, some form must be assumed for the rates. To motivate the form, consider the following argument. Suppose a crystal of silicon contains one vacancy. When it hops, the energy of the system is the same before and after the hop (by symmetry). Suppose a number of defects are now added to the system which interact with each other. Now when the vacancy moves by hopping, its environment changes, and thus so does the energy. Let the hopping rate of the single vacancy be $\nu \exp(-\beta E)$, where E is a barrier height and ν is an attempt frequency that varies only weakly with temperature. Adding in the extra defects modifies this rate slightly. Transitions in a direction that lower the total energy of the system would be expected to be faster than those that raise the energy to ensure that the system heads towards the ground state. Thus, one natural assumption is to let the rate become $\nu \exp(-\beta[E + \alpha(E_{after} - E_{before})])$, where E_{before} and E_{after} are the energies of the system before and after the transition, as it encourages transitions that lower the energy and discourages transitions that raise the energy. This amounts to the assumption that the energy barrier to diffusion varies linearly with the energy difference between the initial and final states. Substituting this into Eq. (13) gives

$$W(i \rightarrow j) = W_0 \exp\left(-\frac{\beta}{2} [V(\vec{R}_j) - V(\vec{R}_i)]\right), \quad (14)$$

where W_0 is a rate which corresponds to a transition with no energy change, but which is also closely related to the transition whose rate is required, and $\alpha = 1/2$. It should be noted that (for example) a vacancy/silicon exchange will have a different value of W_0 from a vacancy/dopant exchange. Thus large contributions to the energy barrier height are treated carefully.

IV. TECHNICAL DETAILS

The above has been fairly general. We now discuss some technical details associated with implementing this approach that apply especially to point defects.

A. Indexing

First, consider the problem of defining a unique indexing scheme for the configurations. If the point defects are well spaced from each other (relative to the inter-atomic spacing of the underlying crystal lattice), then the indexing is straightforward. A single vacancy corresponds to an atom missing from a lattice site. Thus the configurations containing just one vacancy can be indexed by the position of the site that has lost its atom. By extension, if there are many well spaced vacancies, then the list of sites forms the index. This is easily extended further to substitutional atoms. Note that atomic relaxation is fully taken into account in this scheme. Interstitials can be included by introducing interstitial sites in addition to the lattice sites. The index then consists of a list of which crystal and interstitial sites are occupied by defects. This approach can be continued even in the presence of clusters of defects, but see below for a more detailed discussion.

For the indexing system to be *useful* we have to be able to use it to determine neighboring configurations, and evaluate configuration energies and transition rates. For well spaced point defects these criteria are easily met. Neighboring configurations are determined by a single hop of one defect to a neighboring site (or possibly a more distant site) on a lattice. The energies can be calculated according to a simple many-center perturbation expansion about the perfect lattice. For example, adding one vacancy to the crystal raises the energy by an amount $\epsilon^{(1)}(v)$. Adding N well separated vacancies will thus raise the energy by an amount equal to $N\epsilon^{(1)}(v)$. As the vacancies get closer together they begin to interact. This interaction will be pairwise initially, but higher order terms could become important as the density increases. In general we have

$$U = \sum_i \epsilon^{(1)}(d_i, \vec{x}_i) + \frac{1}{2} \sum_{i \neq j} \epsilon^{(2)}(d_i, \vec{x}_i; d_j, \vec{x}_j) + \dots \quad (15)$$

where i and j are site indices, d_i is the type of defect on site i and \vec{x}_i is the position of site i . Finally, the rates are easily defined in terms of the two atoms that swap during a transition (this defines W_0) and the change in energy.

Clusters of vacancies can be treated in precisely the same way as well spaced vacancies. The reason for this is that there is no change of topology as clusters form. That is, as we remove silicon atoms to form a vacancy cluster, the subsequent atomic relaxation does not change the structure of the lattice to the extent that we can no longer uniquely assign an atom to a lattice site [14].

Interstitials are much more complicated to treat. Even for an isolated silicon self-interstitial it may be necessary to supplement the indexing scheme with extra degrees of freedom corresponding to different interstitial types with each type making transitions to different neighboring configurations, because there are a number of local minima with similar energies [15]. Some of the transitions could even correspond to second neighbor hops on the underlying interstitial lattice. However, once an interstitial is no longer isolated indexing becomes much harder. A very important case is the $\{113\}$ defect which plays a central part in transient enhanced diffusion of boron during post-implantation annealing [16]. Here clusters with long ranged correlations form that significantly distort the lattice [17,18]. In this case, there is no unambiguous connection between interstitial atoms placed on atomic and interstitial sites and the final topology. The challenge in this and similar cases is to develop an indexing scheme which maintains a one-to-one correspondence with the set of significant stable structures.

In the case of highly disordered material (e.g., amorphized regions), the number of distinguishable stable structures becomes too large to index practically and it is necessary to fall back to other techniques such as molecular dynamics.

B. Energy expansion

In Eq. (15) is defined a many-site expansion for the energy. There are a number of contributions to the energy that can be isolated. These include: the energy associated with covalent bond formation; elastic energy from strain fields; the energy associated with the Coulomb interaction between charged defects; and the energy associated with the population of gap states. These separate contributions behave in different ways. The first (covalent bond formation) is very short ranged and is associated with energies of the order of 1eV. The next two (elastic and Coulomb energy) are longer ranged but weaker. The population of gap states is of unlimited range, and has a value of order the size of the electron energy gap.

The energy associated with covalent bonding is complicated and can best be handled by being tabulated as a function of cluster type (such as isolated defect, defect pairs with atoms on different sites and so on) in precisely the way encapsulated by the many-site expansion. A two center expansion may be insufficient. Three center contributions appear to be important for vacancies [14]. The elastic energy can be represented by a pair potential, and thus is also easily contained within the many-site expansion.

The Coulomb energy and the energy associated with the population of the gap states must be treated together in a self-consistent manner, as the charge on a defect is a result of the population of the corresponding gap state. This requires Eq. (15) to be generalized to

$$U = \sum_i \epsilon^{(1)}(d_i, q_i, \vec{x}_i) + \frac{1}{2} \sum_{i \neq j} \epsilon^{(2)}(d_i, q_i, \vec{x}_i; d_j, q_j, \vec{x}_j) + \dots \quad (16)$$

where q_i is the charge on the defect on site i . In Appendix A is given the formalism for determining the equilibrium distribution of charges needed to evaluate this expression. The key ideas are: the dopant and defect charges are distributed according to a Boltzmann factor which depends on effective single particle energies calculated from Eq. (16); the effective single particle energies of the dopants and defects are modified by the presence of the electrostatic fields produced by neighboring charged dopants and defects.

Excitations from the valence band and into the conduction band also need to be treated. We present a simple scheme here. To motivate the scheme we note the phenomena

that need to be described: the most important excitations are from the valence band into acceptor states and from donor states into the conduction band; band bending produced by a positively (negatively) charged dopant will tend to cause electrons in the conduction band (holes in the valence band) to accumulate near the dopant. These observations suggest that we can treat the valence and conduction band as a set of localized states uniformly distributed throughout the system with some known density. In the absence of dopants they are degenerate, but the degeneracy is broken by local electrostatic fields produced by charged dopants. Empty conduction states and occupied valence states are treated as neutral. The conduction states are allowed to acquire a charge of -1 when they are populated. The valence states are allowed to acquire a charge of $+1$ when they are depopulated. The density of the valence and conduction states can be found from the density of carriers for the intrinsic semiconductor. If we assume that top of the valence band (ε_v) and the bottom of the conduction band (ε_c) are parabolic, characterised by their minimum separation in energy ($E_{gap} = \varepsilon_c - \varepsilon_v$) and their effective masses (m_v and m_c respectively) it is straightforward to show that the density of carriers is [13] is given by

$$\begin{aligned}
n_c(T) &= N_c(T) \exp[-\beta(\varepsilon_c - \mu)], \\
p_v(T) &= P_v(T) \exp[-\beta(\mu - \varepsilon_v)], \\
N_c(T) &= \frac{1}{4} \left(\frac{2m_c k_B T}{\pi \hbar^2} \right)^{3/2}, \\
P_v(T) &= \frac{1}{4} \left(\frac{2m_v k_B T}{\pi \hbar^2} \right)^{3/2},
\end{aligned} \tag{17}$$

where $n_c(T)$ is the number of electrons in the conduction band, and $p_v(T)$ is the number of holes in the valence band. The density of valence and conduction states are $P_v(T)$ and $N_c(T)$ respectively. These localized states can now be treated in the same way as the dopant and defect states.

Solving the electrostatic equations directly from the point defect positions would be very time consuming. However, the defect states, although localized, still extend over many atomic sites. Thus we can approximate the true charge density by a smooth continuum

charge distribution obtained from the distributions of the charged defects. The charge distributions of the defects can then be obtained consistently with the electrostatic field. Once the charge distributions for the defects are known, the defects can be assigned charges in a random manner so as to reproduce these distributions.

C. Electron dynamics

The hopping of electrons between localised states and the conduction and valence bands is characterised by rapid processes separated by long times. Thus in principle these events can be treated stochastically, and on the same footing as the hopping of atoms between sites. However, these processes are difficult to model accurately as they involve non-radiative transitions [19]. The presence of internal fields and the dielectric response of the silicon substrate only add to the complexities. Thus a simple scheme is proposed here that does away with true electron dynamics, but which reproduces the equilibrium charge distribution. Instead of individual electron excitations, the entire set of charge assignments is updated periodically using the scheme described in the previous section. The time between updates should correspond to a typical time for electronic excitations to prevent unrealistic charge redistributions due to the hopping of charged defects and dopants.

D. Presumption of large hopping rates

The reason that molecular dynamics cannot be used in general to study diffusion in semiconductors is that most of the computational effort goes into modeling the vibration of atoms, and these are rapid processes that do not contribute to diffusion. KLMC allows us to ignore these rapid processes and focus on the much slower processes that do contribute to diffusion.

However, other unwanted rapid processes that take up most of the computer time, but which do not contribute to diffusion, can still occur within KLMC simulations. A simple example might be a vacancy bound to a substitutional dopant atom, for which the exchange

of the vacancy and the dopant is a fast process. The slow process necessary for significant diffusion is the breaking apart of the bound pair. The origin of the variety of rates is exactly the same as in molecular dynamics. A particle is trapped in a potential well, and the probability of it acquiring enough kinetic energy to escape from the well is rather low, so it spends most of the time trapped in the well.

One way to remove the fast processes is to calculate the average time it takes to move from the configuration with the fast process (such as one in which a dopant is a neighbor to a vacancy) to one where this no longer operates (such as one in which the dopant and vacancy are sufficiently well separated that they no longer interact). This allows us to “fast-forward” over the fast processes in a simulation, jumping immediately to the point where it ceases to operate.

In Appendix B we present a rigorous theory that allows us to do this accurately. There it is shown that from this theory a much simpler approximate procedure can be inferred which can be used in practical simulations. The essential point is that the fast process generates a local thermal equilibrium. This allows us to determine the probability of finding the system in a particular configuration (it is given just by the Boltzmann factor), and hence the hopping rate away from the fast process. In passing we note it is this approximation which underlies the transition from molecular dynamics to kinetic lattice Monte Carlo.

It is possible to generalize the above: a set of configurations \mathcal{K} can be replaced by a single effective configuration provided the transitions between members of the set are fast compared to transitions between a member of the set and configurations that are not members. First, we define the effective energy for this set to be

$$V_{\mathcal{K}} = -k_B T \ln \left[\sum_{k \in \mathcal{K}} \exp \left(-\beta V(\vec{R}_k) \right) \right], \quad (18)$$

where the free energy F_k can be used in place of $V(\vec{R}_k)$ if known. Second, we note that the transition rate to a configuration j outside \mathcal{K} is

$$W(\mathcal{K} \rightarrow j) = \sum_{k \in \mathcal{K}} \exp \left(-\beta [V(\vec{R}_k) - V_{\mathcal{K}}] \right) W(k \rightarrow j), \quad (19)$$

and the rate of transition into this set of configurations is simply $W(j \rightarrow \mathcal{K}) = \sum_{k \in \mathcal{K}} W(j \rightarrow k)$. These relationships follow directly from the assumption that the configurations linked by rapid transition rates remain in local equilibrium, and thus the relative probability of finding the system in one of these configurations is just given by its Boltzmann factor. Imposing detailed balance between the transitions $k \rightarrow j$ and $j \rightarrow k$ we get:

$$W(\mathcal{K} \rightarrow j) = \exp\left(-\beta [V(\vec{R}_j) - V_{\mathcal{K}}]\right) W(j \rightarrow \mathcal{K}). \quad (20)$$

Thus we see that the set of configurations does in fact behave as a single configuration with an associated energy $V_{\mathcal{K}}$.

V. CONCLUSION

From the formalism given above for kinetic lattice Monte Carlo for point defects in crystalline systems, it is clear that this approach contains many approximations. For low defect concentrations it yields accurate results. For higher concentrations where clusters form, its accuracy is limited by the ability to effectively index and estimate transition rates between configurations. However, it does allow for at least a qualitatively correct description of diffusive processes, going far beyond what is possible with molecular dynamics (because of severe time constraints) or continuum models (where introducing correlations between particles is not straightforward).

APPENDIX A: DEFECT CHARGE DISTRIBUTION FORMALISM

Since electrons relax with a lifetime much shorter than that of ions, we can assume that the distribution of charges is always at equilibrium. The equilibrium charge distribution at finite temperature is found by minimizing the Helmholtz free energy $\mathcal{F} = U - TS$, where S is the configurational entropy associated with the distribution of the charges, and which we can write as $S = k_B \ln \Omega$ (Ω is the volume of phase space accessible to the charges). It

is convenient to work with the number density of defects of type d with charge q , which is defined by $n_{dq}(\vec{x}) = \sum_i \delta_{d,d_i} \delta_{q,q_i} \delta(\vec{x} - \vec{x}_i)$. We can now rewrite Eq. (16) as

$$U = \sum_{dq} \int d\vec{x} \epsilon^{(1)}(d, q, \vec{x}) n_{dq}(\vec{x}) - \frac{1}{2} \sum_{dq} \int d\vec{x} n_{dq}(\vec{x}) \epsilon^{(2)}(d, q, \vec{x}; d, q, \vec{x}) + \frac{1}{2} \sum_{dq, d'q'} \int d\vec{x} d\vec{x}' n_{dq}(\vec{x}) \epsilon^{(2)}(d, q, \vec{x}; d', q', \vec{x}') n_{d'q'}(\vec{x}') + \dots \quad (\text{A1})$$

Charge conservation requires

$$0 = \sum_{dq} q \int d\vec{x} n_{dq}(\vec{x}). \quad (\text{A2})$$

The number of defects of type d is N_d , and it satisfies

$$N_d = \sum_q \int d\vec{x} n_{dq}(\vec{x}). \quad (\text{A3})$$

The condition for equilibrium is that the free energy is minimized with respect to the particle and charge distribution subject to the constraints of the conservation of charge and number of defects. This gives

$$\frac{\delta \ln \Omega}{\delta n_{dq}(\vec{x})} = \beta [\varepsilon_{dq}(\vec{x}) - \mu q - \mu_d], \quad (\text{A4})$$

where μ is the Lagrange multiplier that allows us to impose charge conservation, and μ_d is the Lagrange multiplier that allows us to conserve the number of defects of type d . $\varepsilon_{dq}(\vec{x}) = \delta U / \delta n_{dq}(\vec{x})$, and is an effective single particle energy. From the many center expansion we get:

$$\varepsilon_{dq}(\vec{x}) = \epsilon^{(1)}(d, q, \vec{x}) - \frac{1}{2} \epsilon^{(2)}(d, q, \vec{x}; d, q, \vec{x}) + \sum_{d'q'} \int d\vec{x}' \epsilon^{(2)}(d, q, \vec{x}; d', q', \vec{x}') n_{d'q'}(\vec{x}') + \dots \quad (\text{A5})$$

The two center term is dominated by the Coulomb interaction. If we let $\phi(\vec{x})$ be the electrostatic potential then we can simplify Eq. (A5) to

$$\varepsilon_{dq}(\vec{x}) \approx \tilde{\epsilon}^{(1)}(d, q, \vec{x}) + q\phi(\vec{x}), \quad (\text{A6})$$

where $\tilde{\epsilon}^{(1)}(d, q, \vec{x}) = \epsilon^{(1)}(d, q, \vec{x}) - \frac{1}{2} \epsilon^{(2)}(d, q, \vec{x}; d, q, \vec{x})$, and is the single defect energy with the electrostatic energy associated with its interaction with itself removed.

If we treat the charged defects as classical particles localized at lattice sites, then evaluating $\ln \Omega$ is straightforward for a homogeneous system. If we define $N_{dq} = \int d\vec{x} n_{dq}(\vec{x})$, then we get

$$\ln \Omega = N_s \ln N_s - N_s - (N_s - N) \ln(N_s - N) + (N_s - N) - \sum_{dq} (N_{dq} \ln N_{dq} - N_{dq}), \quad (\text{A7})$$

where N_s is the total number of sites and $N = \sum_d N_d = \sum_{dq} N_{dq}$ is the total number of defects. Substituting Eq. (A7) into Eq. (A4) allows us to obtain:

$$N_{dq} = N_d \frac{\exp(-\beta[\varepsilon_{dq} - \mu q])}{\sum_{q'} \exp(-\beta[\varepsilon_{dq'} - \mu q'])}. \quad (\text{A8})$$

If define the average particle densities $\langle n_{dq} \rangle = N_{dq}/V_s$ and $\langle n_d \rangle = N_d/V_s$, where V_s is the volume of the system, Eq. (A8) can be rewritten as:

$$\langle n_{dq} \rangle = \langle n_d \rangle \frac{\exp(-\beta[\varepsilon_{dq} - \mu q])}{\sum_{q'} \exp(-\beta[\varepsilon_{dq'} - \mu q'])}. \quad (\text{A9})$$

Writing down the entropy for the general case does not lead to a simple expression. However, we can extend the above result approximately to the case where the variation of $\varepsilon_{dq}(\vec{x})$ with \vec{x} is piecewise constant. In this case we can treat the entropy approximately as the sum of entropies for each uniform section, with the entropy for one section being obtained from Eq. (A7). This gives:

$$\langle n_{dq}(\vec{x}) \rangle = \langle n_d(\vec{x}) \rangle \frac{\exp(-\beta[\varepsilon_{dq}(\vec{x}) - \mu q])}{\sum_{q'} \exp(-\beta[\varepsilon_{dq'}(\vec{x}) - \mu q'])}, \quad (\text{A10})$$

which is the expression to be used for determining the distribution of charges.

APPENDIX B: PRESUMMATION FORMALISM

A problem we can solve analytically is that of a single particle moving through a lattice of points with constant hopping rates between sites. If the fast process is much faster than any competing process in the neighborhood of the particle we are interested in (the hopper), then we effectively have this arrangement. However, there is one unavoidable source of error which follows from the fact that as the hopper moves, it swaps a new particle onto its old

site. Thus the hopping rate between two sites can vary with the route the hopper takes to reach its site. However, this will not be a major source of error provided one route between two sites can be traversed much faster than the others. We will call the approximation of fixed hopping rates the *frozen lattice approximation* (FLA).

In general we need the average time ($\tau(i \rightarrow j)$) for the hopper to migrate from some starting site i to another site j . Let the system evolve in time with a very short time step δt . If the probability that the hopper migrates from site i to site j *for the first time* in n hops is $P_n(i \rightarrow j)$, then

$$\tau(i \rightarrow j) = \sum_{n=1}^{\infty} n \delta t P_n(i \rightarrow j). \quad (\text{B1})$$

The probability of arriving at site j for the first time in n hops is the product of arriving at some neighboring site k in $n - 1$ hops *without* passing through site j , times the probability of hopping from k to j , summed over all k . That is

$$P_n(i \rightarrow j) = \sum_k P_{n-1}^{(j)}(i \rightarrow k) P_1(k \rightarrow j), \quad (\text{B2})$$

where $P_n^{(j)}(i \rightarrow k)$ is the probability of migrating from i to k in n hops while avoiding j . Now we need to calculate $P_n^{(k)}(i \rightarrow j)$. It is convenient at this point to move to matrix notation. Let us define the matrices \mathbf{P} , $\mathbf{P}^{(k)}$ and $\mathbf{P}_n^{(k)}$ by $(\mathbf{P})_{ij} = P_1(i \rightarrow j)$, $(\mathbf{P}^{(k)})_{ij} = P_1^{(k)}(i \rightarrow j)$ and $(\mathbf{P}_n^{(k)})_{ij} = P_n^{(k)}(i \rightarrow j)$. Matrix $\mathbf{P}^{(k)}$ is obtained from matrix \mathbf{P} by crossing out the k^{th} row and k^{th} column since we wish to avoid site k . The matrix $\mathbf{P}_n^{(k)}$ is then calculated from $\mathbf{P}_n^{(k)} = (\mathbf{P}_1^{(k)})^n$ as this corresponds to n independent hops. Substituting these matrices into Eqs. (B1) and (B2), and carrying out the summation over n , we get

$$\tau(i \rightarrow j) = \delta t \sum_k \left([1 - \mathbf{P}^{(j)}]^{-2} \right)_{ik} P_1(k \rightarrow j). \quad (\text{B3})$$

The hopping probabilities are given by the hopping rates $W(i \rightarrow j)$:

$$P_1(i \rightarrow j) = \begin{cases} W(i \rightarrow j) \delta t & \text{if } i \neq j; \\ 1 - \delta t \sum_k W(i \rightarrow k) & \text{if } i = j. \end{cases} \quad (\text{B4})$$

The effective hopping rates for the processes that “fast forward” over the fast processes ($\overline{W}(i \rightarrow j)$) are given by

$$\overline{W}(i \rightarrow j) = \frac{1}{\tau(i \rightarrow j)}. \quad (\text{B5})$$

To use this result in a simulation it is necessary only to remove the fundamental processes for the hopper (represented by the rates $W(i \rightarrow j)$), and replace them with the effective process (represented by the rates $\overline{W}(i \rightarrow j)$).

By way of illustration, consider a simple model for a vacancy bound to a trap. The model consists of three sites in a line, i , j and k . When the vacancy is on site i it is bound to the trap. Let the hopping rates between the sites be: $W(i \rightarrow j) = w$, $W(j \rightarrow i) = W$, $W(i \rightarrow k) = W(k \rightarrow i) = 0$ and $W(j \rightarrow k) = W(k \rightarrow j) = r$. Using the procedure given above we find

$$\overline{W}(i \rightarrow k) = \frac{wr}{(W + w + r)}. \quad (\text{B6})$$

Let the energy of the system when the vacancy is on sites i , j and k be E_i , E_j and E_k respectively. From detailed balance we have $w/W = \exp(-\beta[E_j - E_i])$. Since on site i the vacancy is trapped, $E_j > E_i$, with a typical value for the difference being 1eV. Thus at room temperature $w/W \approx 10^{-18}$. If we assume $r \ll W$ also, we get the result $\overline{W}(i \rightarrow k) \approx (w/W)r = \exp(-\beta[E_j - E_i])r$. This has a simple interpretation. The effective rate is the product of the probability that the vacancy is at the top of the hill (site j) times the hopping rate from j to k . Implicit in this is the assumption that local thermal equilibrium has been reached through the fast processes.

REFERENCES

- [1] D. Frenkel and B. Smit *Understanding Molecular Simulation*, (Academic Press, London, UK) (1996).
- [2] K. A. Fichtorn and W. H. Weinberg, *J. Chem. Phys.* **95**, 1090 (1991).
- [3] C. C. Battaile, D. J. Srolovitz and J. E. Butler, *J. Appl. Phys.* **82** 6293 (1997).
- [4] J. Stiegman, W. Shockley and F. C. Nix, *Phys. Rev.* **56**, 13 (1939).
- [5] R. P. Johnson, *Phys. Rev.* **56**, 814 (1939).
- [6] A. D. LeClaire and A. B. Lidiard, *Phil. Mag.* **1**, 518 (1956).
- [7] P. M. Fahey, P. B. Griffin and J. D. Plummer, *Reviews of Modern Physics* **61**, 289 (1989).
- [8] M. Jaraiz, G. H. Gilmer, J. M. Poate and T. D. de la Rubia, *Appl. Phys. Lett.* **68**, 409 (1996).
- [9] S. T. Dunham and C. D. Wu, *J. Appl. Phys.* **78**, 2362 (1996).
- [10] L. Pelaz, M. Jaraiz, G. H. Gilmer, H.-J. Gossman, C. S. Rafferty, D. J. Eaglesham and J. M. Poate, *Appl. Phys. Lett.* **70** 2285 (1997).
- [11] D. McQuarrie *Statistical Mechanics*, (Harper and Row, New York) (1976).
- [12] H. Goldstein *Classical Mechanics* (Addison-Wesley Publishing Company, Reading, Massachusetts) (1980).
- [13] N. W. Ashcroft and N. D. Mermin *Solid State Physics* (Holt-Saunders, Tokyo, Japan) (1976).
- [14] A. La Magna, S. Coffa and L. Colombo, *Phys. Rev. Lett.*, **82**, 1720 (1999).
- [15] J. Zhu, T. D. dela Rubia, L. H. Yang, C. Mailhot and G. H. Gilmer, *Phys. rev. B* **54**, 4741 (1996).

- [16] D. J. Eaglesham, P. A. Stolk, H.-J. Gossman and J. M. Poate, *Appl. Phys. Lett.* **65**, 2305 (1994).
- [17] M. Kohyama and S. Takeda, *Phys. Rev. B* **46**, 12305 (1992).
- [18] M. Kohyama and S. Takeda, *Phys. Rev. B* **51**, 13111 (1995).
- [19] A. M. Stoneham, *Rep. Prog. Phys.* **44**, 1251 (1981).