

Graph invariants for periodic systems: Towards predicting physical properties from the hydrogen bond topology of ice

Jer-Lai Kuo and Sherwin J. Singer*

Department of Chemistry, Ohio State University, 100 W. 18th Avenue, Columbus, Ohio 43210

(Received 30 August 2002; published 28 January 2003)

Ice-Ih consists of a disordered hydrogen-bonded network. The degree of disorder in ice-Ih, and possible phase transitions to an ordered phase have been debated in recent years. The dependence of energy, free energy, and other scalar physical properties on H-bond topology is needed to understand these phenomena. Graph invariants provide a means of linking physical properties to the topology of the H-bond network. We have previously shown the effectiveness of graph invariants for finite water clusters [J.-L. Kuo, J. V. Coe, S. J. Singer, Y. B. Band, and L. Ojamäe, *J. Chem. Phys.*, **114**, 2527 (2001)]. In this work, we develop a formalism for the graph invariants of periodic systems. We demonstrate that graph invariants for small unit cells are a subset of the graph invariants of larger unit cells, providing a hierarchy of approximations by which detailed calculations for small unit cells, such as periodic *ab initio* calculations as they become available, can be used to parametrize the energy of the astronomical number of H-bond arrangements present in large unit cells. We also present graph enumeration results for ice-Ih, analyzing conflicting results that have appeared previously in the literature and furnishing information on the statistical properties of the H-bond network of ice-Ih in the large system limit.

DOI: 10.1103/PhysRevE.67.016114

PACS number(s): 05.10.-a, 64.60.Cn

Hydrogen bond order and disorder in ice-Ih is an old problem that is still the subject of controversy. Since the work of Pauling, Giauque and Stout in the 1930s [1,2], it is believed that the protons in ice-Ih are disordered, subject to the constraints of the Bernal-Fowler “ice rules” [3], that each water donates to two hydrogen bonds (H bonds) and accepts from two other H bonds. The number of allowed hydrogen bonding arrangements in an infinite periodic lattice has been well-established [1,4,5]. However, conflicting results for the number of distinct H-bond arrangements in periodically replicated units cells, candidates for a possible low-temperature ordered phase of ice, have appeared in the literature [6–8]. Small energy differences between hydrogen bond arrangements in ice may induce a phase transition to a proton-ordered crystal, although under normal conditions the transition is kinetically inaccessible. Most experimental reports center around a first-order transition at 72 K to a ferroelectric structure [9–17]. However, a substantially different transition temperature has been reported [18], and the ferroelectric nature of the low-temperature structure has been questioned as well [19,20].

The purpose of this work is to extend our graph theoretical techniques, originally developed for water clusters [21], to periodically replicated systems, providing analytic tools to address the issues mentioned above. We develop the concept of *graph invariants* for periodically replicated systems. Graph invariants are functions of hydrogen bond variables which are invariant to the symmetry operations of the system. In other words, graph invariants are symmetry-adapted functions that are appropriate for describing how the scalar properties—e.g., total energy, free energy, squared magnitude of the dipole moment, etc.—depend on the arrangement of H bonds.

Graph invariants serve two important functions.

(1) They enable the enumeration of all symmetry-distinct H-bond arrangements by changing it from an $O(N^2)$ to $O(N \ln N)$ problem [21].

(2) They provide a means to systematically parametrize physical properties of water clusters or ice configurations that differ in their H-bond topology.

In this work we use graph invariants to facilitate the listing of all the hydrogen bonding arrangements accessible in several unit cells of the ice-Ih lattice. We resolve the discrepancies in the literature [6–8] for small unit cells. We also provide results for larger unit cells containing billions of different H-bond arrangements, which are of sufficient size to serve as simulation cells for statistical calculations. With these data in hand, we analyze properties of hydrogen bonding in ice-Ih that depend exclusively on the topological constraints of the crystal lattice and satisfy the ice rules.

In a previous work on water clusters [21] we demonstrated that graph invariants furnish a very useful set of symmetry-adapted functions for capturing the dependence of scalar physical properties, such as energy, free energy, etc., on the arrangement of H bonds [21]. In that work we developed a hierarchy of invariants—first-order, second-order, and higher-order invariants (they are defined in Sec. III)—and that physical properties of water clusters could be described quite well in terms of the simplest of the invariants, the first- and second-order invariants. The first-order invariants are often identically zero, as they are for the ice-Ih lattice. Using invariants to parametrize the dependence of cluster energy on its H-bond topology, we successfully described the energy of $\sim 10^4$ – 10^5 different hydrogen bond isomers of a $(\text{H}_2\text{O})_{20}$ cluster, spanning a range of ~ 50 kcal/mol, with only seven numbers, one of which sets the zero of the energy scale [21]. In this work, we show how the same procedure may be adopted for crystal lattices. At the same level of approxima-

*Electronic address: singer@chemistry.ohio-state.edu

tion that was successful for clusters, we show here that on the order of ten numbers should parametrize the energies and other scalar physical quantities of the billions of H-bond isomers that are possible in a large unit cell of ice-Ih. The goal of this work is to provide the theoretical framework that reduces the description of large numbers of H-bond isomers to a handful of parameters. The value of those parameters must come from future detailed calculations, such as periodic *ab initio* calculations for ice, or from experiment.

The implication of these results is that a relatively small number of calculations should suffice to predict the lowest energy structure and phase transitions in ice-Ih. This is significant because different H-bond arrangements in ice are closely spaced in energy and it is difficult to predict their energetic ordering. Buch, Sandler, and Sadlej have demonstrated that empirical water potentials give inconsistent predictions of the relative stability of these arrangements [8]. This information will have to be obtained from costly periodic electronic structure calculations. It is currently not feasible to perform electronic structure calculations for every symmetry-distinct H-bond arrangement for a unit cell large enough for statistical simulations. Even such calculations for a Monte Carlo sample would be quite taxing. Instead of such expensive routes, our techniques offer the possibility of extracting parameters from calculations on small unit cells and bootstrapping to estimate the energy of hydrogen bond arrangements in much larger cells. To accomplish this task, we require the relationship between the graph invariants of small unit cells and those of large unit cells, which is developed in this work.

In Sec. I we provide a “gentle introduction” to graph invariants, using an artificial two-dimensional “square ice” lattice as an example to illustrate the basic idea with a minimum of formalism. In Sec. III we precisely define graph invariants, and how they are generated by group theoretical projection operators (Sec. III A). In Sec. III B we develop the relation between graph invariants for small and large unit cells, and the key concept that permits the physical properties of large unit cells to be parametrize by calculations on small unit cells. The concepts are illustrated in Sec. III C by returning to the example of “square ice.” The reader not interested in the details of the formalism can gather the basic ideas by reading Secs. I, II, and III C.

Section IV contains applications to ice-Ih. Graph invariants for the common eight-water molecule orthorhombic unit cell of ice-Ih are presented in Sec. IV A and the Appendix. Graph invariants also provide an efficient means for complete enumeration of all symmetry-distinct H-bond arrangements for either a cluster or periodic system. Applications to ice-Ih are presented in Sec. IV B. Graph invariants permit complete enumeration of unit cells large enough to approximate the infinite-system limit. Our largest example is a 48-molecule water hexagonal unit cells for which there are 2404 144 962 possible H-bond arrangements permitted by the ice rules, of which 8360 361 are symmetry distinct. In Sec. IV C we analyze selected statistical properties of these arrangements. Finally, we present some concluding remarks in Sec. V.

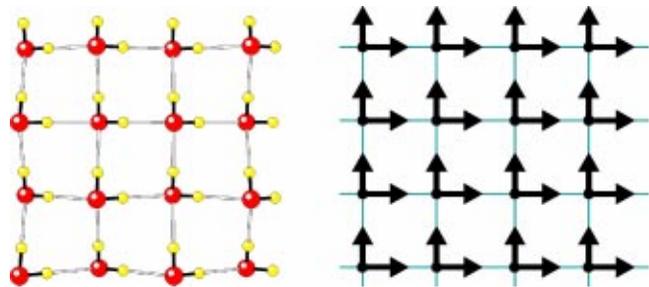


FIG. 1. A square ice lattice used to illustrate graph invariants. The molecular configuration, shown on the left, is summarized by the directed graph appearing on the right. The hydrogen bond arrangement shown here is adopted as the canonical bond orientation. Other periodic hydrogen bond arrangements are possible, as illustrated in Fig. 2.

I. A GENTLE INTRODUCTION TO ORIENTED GRAPHS AND GRAPH INVARIANTS

Each hydrogen bond in ice or water clusters consists of a hydrogen covalently bonded to one oxygen, the donor, and hydrogen bonded to a second oxygen, the acceptor. Hence H bonds are directional, and are conventionally taken to point from donor to acceptor. Proton arrangements in ice are summarized by oriented graphs, a set of vertices linked by directed edges [22–24]. The symbol b_r stands for the orientation of the r th hydrogen bond with respect to a canonical orientation, $b_r = +1$ if the H bond points in the same direction as the canonical orientation, $b_r = -1$ if the direction is opposite.

To illustrate the theory, let us take a simple example, “square ice,” which, like ordinary ice, consists of four-coordinate water molecules. (Of course, applications to the real ice-Ih lattice are presented below in Sec. IV.) Part of the square ice lattice and the direction of bonds, all in an arbitrarily chosen canonical bond orientation, are shown in Fig. 1. Six possible graphs within the 2×2 unit cell of square ice, shown in Fig. 2, when periodically replicated realize an H-bond topology in agreement with the Bernal-Fowler ice rules. The eight bonds of the 2×2 unit cell are given an arbitrary index ranging from 1 to 8, as indicated in graph A of Fig. 2. The value of the bond variables b_1, b_2, \dots, b_8 for the graphs in Fig. 2 are given in Table I.

Some of the graphs shown in Fig. 2 are related to each other by a symmetry operation. Graph D is obtained from graph A by either a C_4 rotation or reflection operation. Therefore, the energy and other scalar properties of the two configurations should be identical. The same is true for graphs B and E, and graphs C and F. If the energy depends on the topological features of the H-bond network, then it must depend upon functions of the bond variables b_r that are identical for configurations related by a symmetry operation.

Consider the combination of bond variables,

$$I_{13}^{2 \times 2} = \frac{1}{4} (b_1 b_3 + b_2 b_4 + b_5 b_6 + b_7 b_8), \quad (1)$$

which is an example of a *graph invariant*. (The origin of the notation $I_{13}^{2 \times 2}$ and the normalization constant will be ex-

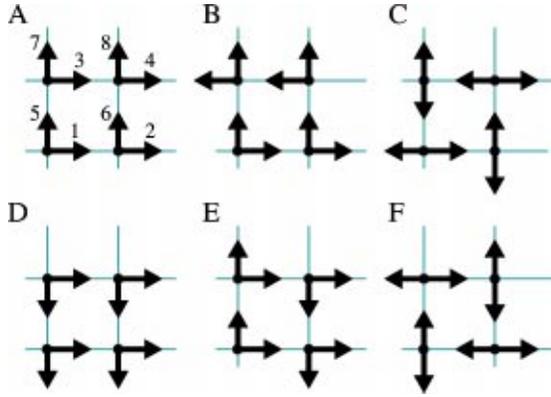


FIG. 2. Graphs that lead to periodic hydrogen bond patterns satisfying the Bernal-Fowler ice rules in the square ice lattice depicted in Fig. 1. In graph A the bonds are arranged in their canonical orientation, the same one shown in Fig. 1. The eight bonds associated with the 2×2 unit cell are numbered according to the scheme indicated in graph A. In some graphs the bonds associated with unit cells neighboring the primary unit cell are shown to make it more apparent how the orientations of complete water molecules are indicated by the graphs. For example, in graph B the periodic image of bond 4 is actually drawn to the left of bond 3. In graph B bond variables $b_1, b_2, b_5, b_6, b_7,$ and b_8 all have value $+1$, while bonds b_3 and b_4 have value -1 , all defined relative to the canonical orientations of graph A.

plained later.) Notice in Table I that $I_{13}^{2 \times 2}$ has exactly the same value among the three pairs of graphs related by symmetry operations. $I_{13}^{2 \times 2}$ also has a clear physical interpretation. It is a sum of dot products of four pairs of parallel bonds. $I_{13}^{2 \times 2}$ effectively counts the number of hydrogen bonded pairs in which nonparticipating hydrogens lie on the same side of the hydrogen bond. Bjerrum postulated that this type of bond has higher energy than those in which the nonbonded hydrogens are more distant [25,26]. Instead of the more complicated notation of Bjerrum, which is only meaningful for the three-dimensional ice-Ih lattice, we will follow

TABLE I. Value of the bond variables and graph invariants associated with each of the graphs depicted in Fig. 2.

	Graph					
	A	B	C	D	E	F
b_1	1	1	1	1	1	-1
b_2	1	1	-1	1	1	1
b_3	1	-1	-1	1	1	1
b_4	1	-1	1	1	1	-1
b_5	1	1	-1	-1	1	1
b_6	1	1	1	-1	-1	-1
b_7	1	1	1	-1	1	-1
b_8	1	1	-1	-1	-1	1
$I_{13}^{2 \times 2}$	1	0	-1	1	0	-1
$I_{23}^{2 \times 2}$	1	0	1	1	0	1
$I_{12}^{2 \times 2}$	1	1	-1	1	1	-1
$I_{15}^{2 \times 2}$	0	0	-1	0	0	-1
$I_{11}^{2 \times 2}$	1	1	1	1	1	1

Buch *et al.* [8] and refer to bonds with nonbonded hydrogens on the same side as “cis,” and others as “trans.” In Fig. 2, all four bonds of graphs A and D are cis. In graphs B and E the bonds connecting waters along the x axis are cis, while those connecting waters along the y axis are trans. None of the bonds are cis in graphs C and F.

Clearly, $n_{cis} = 2(I_{13}^{2 \times 2} + 1)$, as follows from the dot product nature of Eq. (1) and can be verified from Fig. 2. Hence, if Bjerrum’s conjecture turns out to be correct then the graph invariant $I_{13}^{2 \times 2}$ will be the appropriate link between a scalar physical property, the energy in the case of Bjerrum’s conjecture, and the topology of the H-bond network. If the conjecture is valid, we would be able to approximate the dependence of energy on the H-bond topology by a relation of the form,

$$E \approx E_0 + \alpha_{13} I_{13}^{2 \times 2}. \quad (2)$$

The validity of Bjerrum’s notion of strong and weak H bonds has been debated for many years in the literature [8,26–28]. While certainly appropriate for the water dimer [29], it is not clear that H bonds in ice-Ih fall into strong and weak groups according to their cis/trans nature. The reliable way to identify which topological features of the H-bond lattice are most relevant to its stability is to systematically identify *all* symmetry-invariant features of the H-bond topology upon which scalar physical properties may depend. For the 2×2 unit cell of our square ice example, there are four other graph invariants that depend upon pairs of bond variables,

$$I_{23}^{2 \times 2} = \frac{1}{4}(b_2 b_3 + b_1 b_4 + b_6 b_7 + b_5 b_8), \quad (3)$$

$$I_{12}^{2 \times 2} = \frac{1}{4}(b_1 b_2 + b_3 b_4 + b_5 b_7 + b_6 b_8), \quad (4)$$

$$I_{15}^{2 \times 2} = \frac{1}{16}(b_1 b_5 - b_2 b_5 - b_3 b_5 + b_4 b_5 + b_1 b_6 + b_2 b_6 + b_3 b_6 - b_4 b_6 + b_1 b_7 + b_2 b_7 + b_3 b_7 - b_4 b_7 + b_1 b_8 - b_2 b_8 - b_3 b_8 + b_4 b_8), \quad (5)$$

$$I_{11}^{2 \times 2} = \frac{1}{8}(b_1^2 + b_2^2 + b_3^2 + b_4^2 + b_5^2 + b_6^2 + b_7^2 + b_8^2). \quad (6)$$

An improvement over Eq. (2) would be given by using all the invariants that depend on bond variables,

$$E \approx E_0 + \alpha_{13} I_{13}^{2 \times 2} + \alpha_{23} I_{23}^{2 \times 2} + \alpha_{12} I_{12}^{2 \times 2} + \alpha_{15} I_{15}^{2 \times 2} + \alpha_{11} I_{11}^{2 \times 2}. \quad (7)$$

While Eqs. (2) and (7) are written for the energy, we emphasize that the dependence of any scalar physical quantity on H-bond topology can be parametrized in that fashion.

For the ice lattice, both real ice-Ih and our illustrative example square ice, all invariant linear combinations of single bond variables (first-order invariants) are identically zero. The graph invariants in Eqs. (1)–(6) are a complete set of invariant bond combinations for the 2×2 unit cell of

square ice that can be constructed from products of two bond variables. We call such combinations of pairs of bond variables *second-order invariants*. Procedures for generating graph invariants are described in Sec. III. More complicated invariants, made from products of three or more bond variables (third- and higher-order graph invariants) are possible as well, although one may hope for convergence with the respect to description of physical properties as more complicated invariants are included. We have been able to document that second-order invariants adequately describe the dependence of energy and other scalar properties on hydrogen bond topology in clusters [21].

The four additional invariants presented in Eqs. (3)–(6) can be assigned physical interpretations, just as we discussed for $I_{13}^{2 \times 2}$ with relation to Bjerrum's conjecture regarding cis and trans H bonds. For example, $I_{12}^{2 \times 2}$ measures the degree to which chains of H bonds along the x or y direction align in the same direction. Because of the constraints of the ice rules, this also measures the number of water molecules whose OH bonds are both parallel to the x or y direction. (Only graphs C and F contain such waters. All other graphs contain waters with one bond pointing along x and one pointing along y .) $I_{15}^{2 \times 2}$ can be seen to measure this same property. In fact, with regard to the graphs shown in Fig. 2, $I_{12}^{2 \times 2}$ and $I_{15}^{2 \times 2}$ are linearly dependent upon each other: $I_{12}^{2 \times 2} = 2I_{15}^{2 \times 2} + 1$. It often happens that, when evaluated for graphs that satisfy constraints such as the ice rules, invariants are linearly dependent upon each other. Relaxing the ice rules, for example, by allowing hydronium or hydroxide to appear in the lattice, will break the linear dependence of the invariants. The invariant $I_{11}^{2 \times 2}$ is rather trivial for the graphs shown in Fig. 2, merely giving the fraction of filled H bonds in a unit cell.

Let us return to Bjerrum's conjecture that the energy of different H-bond topologies can be linked to the number of cis or trans H bonds present in the lattice. The beauty of Bjerrum's simple conjecture is that it can be applied to both regular, periodic patterns of H bonds, as well as disordered arrangements. Put another way, the number of cis and trans H bonds is a topological invariant for a periodically replicated lattice of arbitrary size, for small unit cells, cells large enough for numerical simulations, or cells whose size tends toward infinity in the true thermodynamic limit. We will demonstrate in this paper that this property of cis and trans H bonds is shared by all the invariants we generate: invariants like the ones we presented in Eqs. (1)–(6) for the 2×2 unit cell of the square ice lattice are also invariants of larger unit cells.

Larger unit cells will also generate new invariants that have no counterpart in small unit cells. However, these new invariants involve bond combinations more distant from each other than in a small unit cell. As a result, one may expect that at a certain point these new, long-range invariants will not be important in capturing physical properties of the system. This sets up a strategy for describing the properties of large unit cells, those large enough for statistical simulations in terms of properties derived from small unit cells. Even though the large unit cells admit millions or billions of

H-bond topologies, the energy, free energy, or other scalar physical properties of each of these topologies, if our previous calculation for clusters [21] is any guide, depend upon the value of a handful of invariants.

II. HOW DETAILED CALCULATIONS FOR SMALL UNIT CELLS CAN APPLY TO SYSTEMS LARGE ENOUGH FOR STATISTICAL SIMULATIONS

The properties of invariants illustrated for “square ice” in the preceding section, and shown in Sec. III to hold quite generally, sets up a scheme by which information from small unit cells can be used to treat the statistical properties of cells large enough to approach the thermodynamic limit.

The potential energy surface for ice-Ih consists of a number of deep minima, each corresponding to a different hydrogen bond topology. Working within the framework of classical statistical mechanics, the classical configuration integral can be written as a sum of contributions from each of M symmetry-distinct local minima of the potential energy surface [30–38],

$$Z_{\mathbf{N}} = \int d\mathbf{r}^{\mathbf{N}} e^{-\beta V(\mathbf{r}^{\mathbf{N}})} = \sum_{i=1}^M f_i e^{-\beta E_i} \int_{\mathcal{D}_i} d\mathbf{r}^{\mathbf{N}} e^{-\beta[V(\mathbf{r}^{\mathbf{N}}) - V(\mathbf{r}_i^{\mathbf{N}})]}. \quad (8)$$

We use a boldface \mathbf{N} to stand for $(N_{\text{O}}, N_{\text{H}})$, the number of hydrogen and oxygen atoms. The position of the atoms at the i th local minimum is denoted as \mathbf{r}_i , \mathcal{D}_i is a \mathbf{N} -dimensional integration domain about the i th minimum, $E_i \equiv V(\mathbf{r}_i^{\mathbf{N}})$ is the potential energy at the i th minimum, and f_i is the number of symmetry-related configurations that are represented by one symmetry-distinct configuration. The canonical partition function of the system is given as

$$\begin{aligned} Q_{\mathbf{N}} &= \sum_{i=1}^M \frac{f_i}{\Lambda^{3\mathbf{N}\mathbf{N}!}} e^{-\beta E_i} \int_{\mathcal{D}_i} d\mathbf{r}^{\mathbf{N}} e^{-\beta[V(\mathbf{r}^{\mathbf{N}}) - V(\mathbf{r}_i^{\mathbf{N}})]} \\ &\equiv \sum_{i=1}^M f_i e^{-\beta(E_i + A_{\text{vib},i})}. \end{aligned} \quad (9)$$

In keeping with the notation of Eq. (8), $\Lambda^{3\mathbf{N}\mathbf{N}!}$ stands for $(\Lambda_{\text{O}}^{3N_{\text{O}}} \Lambda_{\text{H}}^{3N_{\text{H}}} N_{\text{O}}! N_{\text{H}}!)$, where Λ_k is the thermal de Broglie wavelength of atom k , $\Lambda_k = \sqrt{\beta \hbar^2 / 2\pi m_k}$, and m_k is the mass of atom k .

The contribution of each isomer to the partition function is determined by the potential energy E_i of the isomer, and an integral over “vibrational” or “phonon” fluctuations about the i th local minimum of the potential energy surface whose contribution we call the vibrational free energy $A_{\text{vib},i}$. At sufficiently low temperature the classical procedure could be modified to incorporate quantum effects, for example, by calculating $A_{\text{vib},i}$ quantum mechanically. Here, we have made use of classical statistical mechanics to illustrate the use of invariants with a minimum of formalism.

Calculating the E_i and $A_{vib,i}$ for the billions of hydrogen bond topologies found in a “simulation cell,” a unit cell large enough to approximate the thermodynamic limit, is not feasible, yet it is what would be needed to, say, predict proton ordering phase transitions in ice-Ih. The graph invariants we introduce in this work provide a way to circumvent the need to calculate all the E_i and $A_{vib,i}$. First consider how the scheme would work for the energy if one accepted Bjerrum’s conjecture about cis and trans H bonds. The energy difference between a cis and a trans H bond in ice-Ih could be established by *ab initio* calculations on small unit cells, for which this type of detailed calculation is feasible. Such *ab initio* calculations are not possible for billions of H-bond arrangements in large simulation cells, but we have shown that it is certainly possible to enumerate all the H-bond topologies for large cells, determining the coefficients f_i in Eq. (9) and properties such as the cis-trans bond distribution for all topologies [21]. The energies E_i for the billions of topologies possible for large unit cells would be given, relative to an all-trans configuration, by counting the number of cis bonds in each topology and multiplying by the cis-trans energy difference. This illustrates the two ingredients needed in our scheme, parameters derived from detailed calculations (such as *ab initio*) on small unit cells and enumeration results for a larger simulation cell. It also illustrates that an invariant for small unit cells, in this case the number of cis and trans H bonds, is also an invariant for large cells and that the energetic parameter obtained for small cells is applicable to a large cell.

Of course, using a single parameter, the relative number of cis and trans H bonds, is not likely to furnish an accurate description of ice-Ih. This work gives the proper generalization of this idea, showing how a hierarchy of parameters, the graph invariants, can be generated, and how graph invariants of increasing complexity can be added until convergence is attained. In our previous work, we showed how even low-order invariants provided a reasonable description of the energetics and dipole moment of the H-bond isomers in water clusters [21]. Even when using the appropriate graph invariants, the two ingredients remain detailed calculations for small unit cells and an enumeration of H-bond topologies for large cells.

To describe phase transitions, free energetics, not just energetics, are required. However, any scalar physical property can be parametrized with graph invariants. Therefore, the vibrational free energies $A_{vib,i}$ are equally amenable to parametrization with graph invariants. We briefly discuss a few methods by which $A_{vib,i}$ might be obtained, just to emphasize that graph invariants do furnish a feasible route to predict phase transitions in ice-Ih. In a harmonic approximation, $V(\mathbf{r}^N) - V(\mathbf{r}_i^N)$ would be taken as a quadratic function in deviations from \mathbf{r}_i^N and the range of integration over \mathcal{D}_i could be safely extended to all space. In the harmonic approximation, the contribution of $A_{vib,i}$ to the heat capacity C_V is identical for all H-bond isomers, and the temperature dependence of C_V is fixed by the energies E_i . Calculation of the heat capacity for a model water cluster is given in our previous work [21], showing that the invariant scheme is a fea-

sible route to thermodynamic properties. It also might be a reasonable assumption to replace $A_{vib,i}$ by an average value \bar{A}_{vib} for each of the isomers, in which case $Q_N \approx e^{-\beta \bar{A}_{vib} \sum_{i=1}^M f_i} e^{-\beta E_i}$.

III. GRAPH INVARIANTS

Graph invariants, functions of bond variables that are unchanged under any symmetry operations, can be constructed using standard group theoretical projection operators. The application of a projection operator to a single bond variable, b_r , takes the form

$$I_r = C_r \sum_{\alpha} g_{\alpha}(b_r), \quad (10)$$

where C_r is a normalization constant chosen for convenience, g_{α} is a member of the symmetry group of the system, and the sum runs over the entire symmetry group. The characters of the totally symmetric representation are identical for all symmetry operations. Therefore, to construct a linear combination that transforms according to the totally symmetric representation of the group, the terms $g_{\alpha}(b_r)$ are combined in Eq. (10) with equal coefficients. The appropriate group for a crystal lattice is the space group. We assume that the crystal is large and periodic, so the translation subgroup is of order $N_x N_y N_z$ (or obviously $N_x N_y$ for a two-dimensional lattice such as square ice). We use x, y, z to designate the crystal axes, but nothing in our formalism requires that these axes be orthogonal.

Other invariants can be constructed similarly:

$$I_{rs} = C_{rs} \sum_{\alpha} g_{\alpha}(b_r b_s), \quad (11)$$

$$I_{rst} = C_{rst} \sum_{\alpha} g_{\alpha}(b_r b_s b_t), \quad (12)$$

⋮

Throughout this work we conveniently take the normalization constant to be the inverse of the order of the group, $|G|$, making the invariants intensive quantities,

$$C_{rs\dots} = \frac{1}{|G|}. \quad (13)$$

We refer to I_r as a first-order invariant, I_{rs} as a second-order invariant, and so on. From the definition of invariants, it is obvious to see that $I_{rs} = I_{sr}$. More generally, invariants with permuted subscripts are equivalent. When all bonds are filled, all bond variables $b_r = \pm 1$. Therefore we have $I_{rr} = \text{const}$, as well as $I_{rrstu\dots} = I_{st\dots}$. Products of invariants are also invariants. For example, products of two first-order invariants can be expanded as a linear combination of second-order invariants, products of first and second are a linear combination of third-order invariants, and so on,

$$I_r I_s = \sum_{tu} c_{tu}^{r,s} I_{tu}, \quad (14)$$

$$I_r I_{st} = \sum_{uvw} c_{uvw}^{r,st} I_{uvw}, \quad (15)$$

$$\vdots = \vdots$$

We have previously shown that if a symmetry operation can bring a single bond b_r into $-b_r$, the first-order invariant of b_r is identically zero [21]. More generally, if $g_\alpha(b_r) = \pm b_s$, I_r and I_s are equivalent. Local constraints, for example, ice rules, can cause further degeneracy, as was illustrated in Sec. I.

Scalar physical properties that depend on hydrogen bond topology will be a function of graph invariants. The simplest relationship, linear dependence, is illustrated below for the energy, although any scalar physical quantity can be parametrized in a similar fashion:

$$E = E_0 + \sum_r \alpha_r I_r + \sum_{rs} \alpha_{rs} I_{rs} + \sum_{rst} \alpha_{rst} I_{rst} + \dots \quad (16)$$

The above expression will always be valid if the physical differences between H-bond arrangements are not too great. We have shown that the linear expansion can be still quite successful for water clusters [21], even when energetic differences between H-bond isomers are rather great. In Eq. (16) the graph invariants provide a vector space over H-bond topologies. In particular, the graph invariants are symmetry-adapted combinations that span the symmetry invariant subspace.

The linear expansion of Eq. (16) is not the most general relation between scalar properties and H-bond topology, and in certain situations we may expect nonlinear dependence of physical properties on the invariants. To give an example, in a simple model where the total dipole moment arises from bond dipoles μ_r , the total dipole moment could be expressed in terms of our bond variables as

$$\mu = \sum_r b_r \mu_r \quad (17)$$

and we expect the squared magnitude of the total dipole moment to be well described by a linear expansion in second-order graph invariants, $|\mu_r|^2 \approx \sum_{rs} \alpha_{rs} I_{rs}$, and indeed find this to hold nicely for H-bond topologies of the $(\text{H}_2\text{O})_6$ cage cluster [21]. This implies that a linear expansion of $|\mu_r|$ itself through second-order invariants would not be as successful, unless a series expansion of the square root of $|\mu_r|^2$ converged rapidly. Instead, the nonlinear function $\sqrt{\sum_{rs} \alpha_{rs} I_{rs}}$ would be the expansion of choice for $|\mu_r|$. (For nonlinear functions, the classification of invariants into first, second, and higher orders loses its significance.) Of course, since products of invariants are also invariants [Eqs. (14) and (15)], a linear expansion for $|\mu_r|$ in the form of Eq. (16) would eventually converge, but might require higher-order terms.

A. Graph invariants and space groups

Symmetry properties are manifested by a group of permutation operations mapping the set of vertices onto themselves. A list of adjacent vertices (vertices connected by a bond, irrespective of the bond's direction) is preserved by each of the symmetry operations. The space group of a crystal can be treated as a finite group by invoking periodic boundary conditions. Consider a lattice with possibly nonorthogonal unit cell vectors $\{a_x, a_y, a_z\}$. The full space group is designated as G . Γ , the crystallographic translational group, is generated by the elementary translation operators $\{\tau_x, \tau_y, \tau_z\}$, where $\tau_x^u \tau_y^v \tau_z^w(R) = R + ua_x + va_y + wa_z$. That is,

$$\Gamma = \{ \tau_x^u \tau_y^v \tau_z^w | u=0,1,\dots,N_x-1, v=0,1,\dots,N_y-1, w=0,1,\dots,N_z-1 \}. \quad (18)$$

We will always assume a large but finite crystal with periodic boundary conditions,

$$\tau_x^{u+N_x} = \tau_x^u, \quad \tau_y^{v+N_y} = \tau_y^v, \quad \text{and} \quad \tau_z^{w+N_z} = \tau_z^w. \quad (19)$$

Hence, Γ becomes a finite group and $|\Gamma|$, the order of Γ , is $N_x N_y N_z$.

As is well known in the theory of space groups [39], G can be decomposed into a sum of cosets of Γ ,

$$G = \Gamma p_1 \cup \Gamma p_2 \cup \Gamma p_3 \cup \dots, \quad (20)$$

where the p_β are coset representatives and \cup stands for a summation of two sets, which is the set of all objects that are contained in at least one of the sets. The set of cosets forms the factor group G/Γ . Conventionally, the coset representative p_β is chosen to be a pure point group operation if possible, or a space group operation involving a minimal translation if a screw or glide operation.

The projection operation for the totally symmetric representation of G , denoted here as \hat{G} , is generated by applying all operations of the group with coefficients proportional to the characters of the totally symmetric representation, that is, with equal coefficients. The projection operator for the totally symmetric representation of the pure translation group, denoted here as $\hat{\Gamma}$, is simply

$$\hat{\Gamma} = \sum_{u=0}^{N_x-1} \sum_{v=0}^{N_y-1} \sum_{w=0}^{N_z-1} (\tau_x)^u (\tau_y)^v (\tau_z)^w, \quad (21)$$

and for the full space group the projection operator is

$$\hat{G} \equiv \sum_{\alpha \in G} g_\alpha = \sum_{\beta \in G/\Gamma} \hat{\Gamma} p_\beta. \quad (22)$$

The first sum is over all elements in G , while the second sum is over the coset representatives. Our previous equations (10), (11) for graph invariants can be rewritten in terms of projection operators:

$$I_r = C_r \hat{G}(b_r), \quad I_{rs} = C_{rs} \hat{G}(b_r b_s), \quad I_{rst} = C_{rst} \hat{G}(b_r b_s b_t), \quad (23)$$

and so on for higher-order invariants.

The graph invariants of Eqs. (1)–(6) were generated using projection operators. Before explicitly presenting the procedure, we have to recognize that expressions like Eq. (23) are still not adapted to bond patterns consisting of periodically replicated unit cells.

B. Invariants for arbitrary unit cell choice

Practical statistical simulations of proton ordering ice-Ih require unit cells large enough to approximate the properties of an infinite system. However, cells much beyond ten water molecules allow an astronomical number of different hydrogen-bonded arrangements, seemingly making Monte Carlo sampling [40] the only feasible approach for unit cells large enough to approach the thermodynamic limit. Graph invariants provide a link between the properties of large unit cells and cells small enough to allow accurate *ab initio* studies, thereby providing an alternate to numerical simulations for larger unit cells. The key is the link between graph invariants for unit cells of different size, which we derive in this section. The derivation presented in this section is rather technical. Some readers may prefer to first see the results illustrated and confirmed for “square ice” in Sec. III C, where the relevance of these results to describing H-bond disorder is more apparent. If the main concepts are sufficiently clarified by the example in Sec. III C, this section may be skipped.

Consider the smallest unit cell, defined by the translation subgroup Γ . Since the hydrogen bonding pattern is repeated in all unit cells,

$$\forall \tau_i \in \Gamma, \text{ value of } \tau_i b_r = \text{value of } b_r. \quad (24)$$

The above equation applies to the *value* of the bonds, not the bond variables themselves. The translated bond $\tau_i b_r$ resides in a different unit cell from b_r , and the image of $\tau_i b_r$ under a symmetry operation is different from the image of b_r , even though they might share the same value of either +1 or -1. This point is illustrated with an example for “square ice” in Sec. III C, which immediately follows this section. [See discussion following Eq. (34) below.] Periodic replication of the hydrogen bonding pattern, as expressed in Eq. (24), implies that only the coset representatives need to be projected to generate invariants,

$$\begin{aligned} I_{rs\dots} &= C_{rs\dots} \hat{G}(b_r b_s \dots) = C_{rs\dots} \sum_{\beta \in G/\Gamma} \hat{\Gamma} p_{\beta}(b_r b_s \dots) \\ &= C_{rs\dots} |\Gamma| \sum_{\beta \in G/\Gamma} p_{\beta}(b_r b_s \dots) \\ &= \frac{1}{|G/\Gamma|} \sum_{\beta \in G/\Gamma} p_{\beta}(b_r b_s \dots). \end{aligned} \quad (25)$$

The second equality is a consequence of the periodicity of the lattice, as expressed in Eq. (24). The third equality is obtained by invoking the normalization condition, Eq. (13). The action of \hat{G} on bond variables $b_r, b_s \dots$ produces a linear combination of bonds spanning the entire lattice. Reduction of $\hat{G}(b_r b_s \dots)$ in the second line of Eq. (25) to a few terms over a single unit cell multiplied by $|\Gamma|$ is only true for the *value* of the invariant, given that the pattern of bond variables is periodically replicated.

For a unit cell $O(n_x \times n_y \times n_z)$ with basis $\{n_x a_x, n_y a_y, n_z a_z\}$, the translation group, denoted as $\Gamma_{n_x \times n_y \times n_z}$, can be written as

$$\Gamma_{n_x \times n_y \times n_z} = \left\{ \left(\tau_x^{n_x} \right)^u \left(\tau_y^{n_y} \right)^v \left(\tau_z^{n_z} \right)^w \mid u=0,1,\dots, \frac{N_x}{n_x} - 1, \right. \\ \left. v=0,1,\dots, \frac{N_y}{n_y} - 1, w=0,1,\dots, \frac{N_z}{n_z} - 1 \right\}. \quad (26)$$

It is elementary to see that Γ is equivalent to $\Gamma_{1 \times 1 \times 1}$ and $\Gamma_{n_x \times n_y \times n_z} \subset \Gamma_{1 \times 1 \times 1} = \Gamma$. For any $\Gamma_{n_x \times n_y \times n_z}$, we have

$$|\Gamma_{n_x \times n_y \times n_z}| = |\Gamma| / n_x n_y n_z = N_x N_y N_z / n_x n_y n_z. \quad (27)$$

For graphs satisfying the periodic boundary condition of unit cell $O(n_x \times n_y \times n_z)$, the value of a bond variable at a position translated by one of the members of $\Gamma_{n_x \times n_y \times n_z}$ is equal to the bond variable at the original position,

$$\forall \tau_i \in \Gamma_{n_x \times n_y \times n_z}, \text{ value of } \tau_i b_r = \text{value of } b_r. \quad (28)$$

Note that the above equation provides fewer constraints on the hydrogen bonds than Eq. (24) for the smaller unit cell $O(1 \times 1 \times 1)$. As the periodic cell is enlarged, a greater variety of hydrogen bonding patterns is permitted until, as the cell size approaches the thermodynamic limit, it is capable of describing all manner of disorder in an ice-Ih crystal.

The full space group can be decomposed into cosets of the translation subgroup $\Gamma_{n_x \times n_y \times n_z}$. While the the pure translation group for the crystal with unit cell $O(n_x \times n_y \times n_z)$ is smaller than for $O(1 \times 1 \times 1)$, the set of coset representatives for $O(n_x \times n_y \times n_z)$ is correspondingly enlarged by a factor of $n_x n_y n_z$. The set of coset representatives $p_{\beta}^{n_x \times n_y \times n_z}$ for the larger cell is given by

$$\{ \tau_x^u \tau_y^v \tau_z^w p_{\beta} \mid \beta \in G/\Gamma, u=0,1,\dots, n_x - 1, v=0,1,\dots, n_y - 1, \\ w=0,1,\dots, n_z - 1 \}. \quad (29)$$

The space group G may be decomposed into cosets appropriate for either the cells $O(1 \times 1 \times 1)$ or $O(n_x \times n_y \times n_z)$,

$$G = \Gamma p_1 \cup \Gamma p_2 \cup \Gamma p_3 \cup \dots = \Gamma^{n_x \times n_y \times n_z} p_1^{n_x \times n_y \times n_z} \cup \Gamma^{n_x \times n_y \times n_z} p_2^{n_x \times n_y \times n_z} \cup \Gamma^{n_x \times n_y \times n_z} p_3^{n_x \times n_y \times n_z} \cup \dots \quad (30)$$

In the above equation we have decomposed G into right cosets. For the full translation subgroup the choice between left and right cosets is irrelevant because Γ is a normal subgroup of G , for which left and right cosets are identical. However, $\Gamma^{n_x \times n_y \times n_z}$ might not be a normal subgroup of G , and the left and right cosets may be distinct. In this case, decomposition into right cosets is the most convenient choice because, according to Eq. (28), following the action of a coset representative with any member of $\Gamma^{n_x \times n_y \times n_z}$ leaves the *value* of the bond expression unchanged, as explained in the discussion accompanying Eqs. (24) and (25).

The application of \hat{G} on a product of bond variables can be simplified in several ways. Following from the unit cell conditions of $O(n_x \times n_y \times n_z)$ expressed in Eq. (28) and in analogy to Eq. (25), application of \hat{G} only needs to involve the coset representatives,

$$\begin{aligned} \hat{G}(b_r b_s \dots) &= \sum_{\beta \in G/\Gamma^{n_x \times n_y \times n_z}} \hat{\Gamma}_{n_x \times n_y \times n_z} p_{\beta}^{n_x \times n_y \times n_z}(b_r b_s \dots) \\ &= |\Gamma_{n_x \times n_y \times n_z}| \\ &\quad \times \sum_{\beta \in G/\Gamma^{n_x \times n_y \times n_z}} p_{\beta}^{n_x \times n_y \times n_z}(b_r b_s \dots). \end{aligned} \quad (31)$$

The last line of Eq. (31) applies to the *value* of the expression when evaluated for a periodically replicated system, not

the actual bond variables. [See discussion accompanying Eqs. (24) and (25).] Using the definition of the coset representatives given in Eq. (29), we can further simplify invariants $I_{rs \dots}$ in unit cell $O(n_x \times n_y \times n_z)$,

$$\begin{aligned} I_{rs \dots}^{n_x \times n_y \times n_z} &= C_{rs \dots}^{n_x \times n_y \times n_z} \sum_{u=0}^{n_x-1} \sum_{v=0}^{n_y-1} \sum_{w=0}^{n_z-1} \tau_x^u \tau_y^v \tau_z^w \left[\sum_{\beta \in G/\Gamma} p_{\beta}(b_r b_s \dots) \right] \\ &= \frac{1}{|G/\Gamma| n_x n_y n_z} \sum_{u=0}^{n_x-1} \sum_{v=0}^{n_y-1} \sum_{w=0}^{n_z-1} \tau_x^u \tau_y^v \tau_z^w \\ &\quad \times \left[\sum_{\beta \in G/\Gamma} p_{\beta}(b_r b_s \dots) \right]. \end{aligned} \quad (32)$$

Let us analyze result (32) in two different situations. First consider when all the bonds within the product $b_r b_s \dots$ lie close to each other within the cell $O(n_x \times n_y \times n_z)$, in fact, so close that they would fit within a smaller unit cell $O(1 \times 1 \times 1)$. Then, for terms of this type the quantity in square brackets in Eq. (32) for $I_{rs \dots}^{n_x \times n_y \times n_z}$ is, within a constant, an invariant $I_{rs \dots}$ for $O(1 \times 1 \times 1)$, evaluated for a portion of the larger unit cell $O(n_x \times n_y \times n_z)$. Hence each invariant for the small unit cell appears as an invariant of the larger unit cell,

$$I_{rs \dots}^{n_x \times n_y \times n_z} = \frac{1}{|G/\Gamma| n_x n_y n_z} \sum_{u=0}^{n_x-1} \sum_{v=0}^{n_y-1} \sum_{w=0}^{n_z-1} \tau_x^u \tau_y^v \tau_z^w I_{rs \dots}^{1 \times 1 \times 1}(b_r b_s \dots), \quad b_r b_s \dots \in O(1 \times 1 \times 1). \quad (33)$$

[The converse, that each invariant in Eq. (32) generated from $b_r b_s \dots$ lying within a small unit cell is an invariant of the lattice with a smaller unit cell, is not true. This is because periodicity imposes fewer constraints for the larger lattice. This point is illustrated in Sec. III C.] Of course, when the unit cell is enlarged to $O(n_x \times n_y \times n_z)$, the small cell H-bond pattern is, in general, not periodically replicated within the large cell. Therefore, the value of $I_{rs \dots}^{n_x \times n_y \times n_z}$ is not simply a multiple of the value of $I_{rs \dots}$ for a smaller unit cell.

Second, let us consider the case where the bonds within the product $b_r b_s \dots$ in Eq. (32) do not all lie within a region of the size of the small cell $O(1 \times 1 \times 1)$. The invariants generated for $O(n_x \times n_y \times n_z)$ in this case have no analog from $O(1 \times 1 \times 1)$. These new invariants arise from the greater variety of H-bond topologies permitted when the unit cell is enlarged. Since this class of invariants involves prod-

ucts of bonds separated by distances greater than the dimension of the small unit cell $O(1 \times 1 \times 1)$, these invariants will describe longer ranged physical interactions, which would be expected to be weaker.

In summary, upon enlarging the unit cell from $O(1 \times 1 \times 1)$ to $O(n_x \times n_y \times n_z)$, we find two types of invariants. The first type are simply invariants of the small cell evaluated for portions of the large cell and then summed. When decomposing the dependence of energy (or other scalar physical quantities) on H-bond topology in terms of invariants, we expect these invariants to provide the dominant contribution. The second type are fundamentally new invariants involving products of bonds separated by distances greater than the dimension of $O(1 \times 1 \times 1)$.

Our discussion of enlarging the cell from $O(1 \times 1 \times 1)$ to $O(n_x \times n_y \times n_z)$ applies equally when going from $O(n_x \times n_y \times n_z)$ to even larger unit cell dimensions

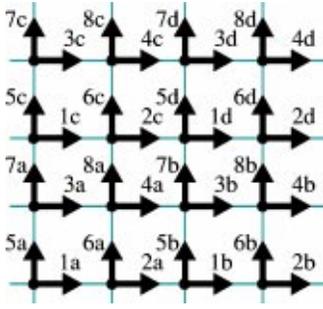


FIG. 3. Labeling scheme for bonds in the 4×4 unit cell of “square ice.”

$O(n'_x \times n'_y \times n'_z)$. This provides a natural hierarchy of approximations for decomposing the dependence of scalar physical properties on H-bond topology. The most local and dominant effects would be captured by fitting to invariants at the level of the small cell $O(1 \times 1 \times 1)$. If these effects are completely dominant, then physical properties for cell $O(n_x \times n_y \times n_z)$ would be accurately predicted in terms of invariants that are from the $O(1 \times 1 \times 1)$ cell, summed over all portions of $O(n_x \times n_y \times n_z)$. Deviations from this picture are used to parametrize physical properties in terms of the invariants for $O(n_x \times n_y \times n_z)$ that involve longer range interactions. This improved characterization could, in principle, be tested at a still larger level $O(n'_x \times n'_y \times n'_z)$ until convergence is achieved.

For simplicity, the transition from small to large unit cells has been discussed here as a mere rescaling of the unit vectors by factors of n_x , n_y , and n_z . Quite often, convenient choices of larger unit cells involve linear combinations of primitive lattice vectors. Our conclusions apply to these cases as well, and we illustrate such unit cells in our treatment of ice below. Whatever may be our choice of unit cell, the unit cell vectors are a subgroup of the full translation group Γ . The translation subgroup of the unit cell vectors can be used to decompose the full space group into right cosets, and the link made between invariants for small and large unit cells.

C. An illustration for square ice

In Sec. I we exhibited the five second-order graph invariants $I_{rs}^{2 \times 2}$ associated with the 2×2 unit cell of our “square ice” example. The formalism of Secs. III A and III B explained how those graph invariants were generated with projection operators, and exposed relations between graph invariants for unit cells of arbitrary size. The very practical consequence of these relations is that calculations feasible for only small unit cells, such as *ab initio* energetic calculations, can be applied to larger unit cells appropriate for statistical simulations. Since the formalism of Sec. III B may be forbidding at first glance, we illustrate the relationship between graph invariants for unit cells of different size for square ice. The entire set of second-order graph invariants $I_{rs}^{4 \times 4}$ for the 4×4 unit cell (Fig. 3) is presented in this section, and we discuss the connections with the graph invariants of the smaller 2×2 unit cell.

We begin by examining the result of projecting onto bonds $1a$ and $3a$ of the 4×4 unit cell,

$$I_{1a,3a}^{4 \times 4} = \frac{1}{32} \{ b_{1a}b_{3a} + b_{2a}b_{4a} + b_{5a}b_{6a} + b_{7a}b_{8a} \\ + b_{1b}b_{3b} + b_{2b}b_{4b} + b_{5b}b_{6b} + b_{7b}b_{8b} \\ + b_{1c}b_{3c} + b_{2c}b_{4c} + b_{5c}b_{6c} + b_{7c}b_{8c} \\ + b_{1d}b_{3d} + b_{2d}b_{4d} + b_{5d}b_{6d} + b_{7d}b_{8d} \\ + b_{1c}b_{3a} + b_{2c}b_{4a} + b_{5c}b_{6a} + b_{7c}b_{8a} \\ + b_{1a}b_{3c} + b_{2a}b_{4c} + b_{5a}b_{6b} + b_{7a}b_{8b} \\ + b_{1b}b_{3d} + b_{2b}b_{4d} + b_{5c}b_{6d} + b_{7c}b_{8d} \\ + b_{1d}b_{3b} + b_{2d}b_{4b} + b_{5d}b_{6c} + b_{7d}b_{8c} \}. \quad (34)$$

Each of the first four lines are clearly recognizable as $I_{13}^{2 \times 2}$ of Eq. (1) evaluated for each 2×2 sector of the 4×4 unit cell. Each of the terms represents the product of bond variables for bonds that are parallel and separated by one lattice unit in either the x or y direction, an interaction that could be estimated by a calculation for the smaller 2×2 cell. Terms like $b_{1a}b_{3c}$ may seem to violate this condition, since bonds $1a$ and $3c$ lie three lattice units from each other in the y direction. However, the term $b_{1a}b_{3c}$ actually represents the interaction of the bond $1a$ with another bond below it which lies in a neighboring unit cell. Because of lattice periodicity, that bond has the same *value* as its periodic image bond $3c$. Hence, in the term $b_{1a}b_{3c}$, the variable b_{3c} represents the *value* of another bond which is its periodic image in the lattice. This example illustrates the distinction, made immediately after Eq. (24), between bond variables and their value. In expressions like Eq. (34) above, it is most convenient to replace actual bond variables, which might be bond variables outside a primary unit cell, with other variables within the primary cell that have the same value. Returning to Eq. (34) above, we could have just as well said that the term $b_{1a}b_{3c}$ represents the interaction of a bond $3c$ with another bond one lattice unit above it whose value is the same as its periodic image, bond $1a$.

Expression (34) is an illustration of the general formulas, Eqs. (32) and (33). The terms in the last four lines would be identical in value to those of the first four lines if the lattice still had 2×2 periodicity. Put another way, if the letters were removed from the subscripts in the last four lines, thereby enforcing 2×2 periodicity, the last four lines would duplicate the first four lines. These terms are indeed part of $I_{13}^{2 \times 2}$, but they do not appear explicitly in Eq. (1) because their *value* is identical to terms already present in that expression. In the 4×4 setting these terms must be included as distinct contributions. Provided the additional invariants [Eqs. (40)–(46) below] introduced at the 4×4 level do not make a significant contribution, the contribution of an invariant like $I_{1a,3a}^{4 \times 4}$ to a scalar physical property like the energy could be estimated from *ab initio* calculations for the 2×2 unit cell.

As discussed in Sec. I, invariants like $I_{1a,3a}^{4 \times 4}$ of Eq. (34) have the physical interpretation of counting the number of

cis and trans H bonds of square ice. Therefore, if Bjerrum's conjecture was correct and the energetic difference between a cis and trans H bond was established for a 2×2 unit cell and the parameter α_{13} of Eq. (2) established, then for the 4×4 unit cell the energy would be given by

$$E \approx E_0 + \alpha_{13} I_{13}^{4 \times 4}, \quad (35)$$

where α_{13} is the same number as in Eq. (2) and has been established by detailed calculations on the smaller unit cell. Of course, an expression like Eq. (35) would only be appropriate if Bjerrum's conjecture about cis and trans H bonds was valid. Therefore, an expression using additional invariants, like Eq. (7) for the 2×2 cell, would be more accurate. In the following paragraph we illustrate that each of the invariants appearing in Eq. (7) also appears as an invariant of the 4×4 cell [as predicted, in general, by Eqs. (32) and (33)], and so the α 's of Eq. (7) determined for the smaller cell provide information about the 4×4 cell.

Just like $I_{1a,3a}^{4 \times 4}$ in Eq. (34), each of the graph invariants given below in Eqs. (36)–(39) has a counterpart among those of the 2×2 unit cell, specifically in Eqs. (3)–(6),

$$I_{2a,3a}^{4 \times 4} = \frac{1}{64} \sum_{\alpha, \beta = a, b, c, d} (b_{2\alpha} b_{3\beta} + b_{1\alpha} b_{4\beta} + b_{6\alpha} b_{7\beta} + b_{5\alpha} b_{8\beta}), \quad (36)$$

$$I_{1a,2a}^{4 \times 4} = \frac{1}{32} \left\{ \sum_{\alpha = a, b, c, d} (b_{1\alpha} b_{2\alpha} + b_{3\alpha} b_{4\alpha} + b_{5\alpha} b_{7\alpha} + b_{6\alpha} b_{8\alpha}) \right. \\ + \sum_{(\alpha, \beta) = (a, b), (c, d)} (b_{1\alpha} b_{2\beta} + b_{2\alpha} b_{1\beta} \\ + b_{3\alpha} b_{4\beta} + b_{4\alpha} b_{3\beta}) \\ + \sum_{(\alpha, \beta) = (a, c), (b, d)} (b_{5\alpha} b_{7\beta} + b_{7\alpha} b_{5\beta} + b_{6\alpha} b_{8\beta} \\ + b_{8\alpha} b_{6\beta}) \left. \right\}, \quad (37)$$

$$I_{1a,5a}^{4 \times 4} = \frac{1}{64} \left\{ \sum_{\alpha = a, b, c, d} (b_{1\alpha} b_{5\alpha} - b_{3\alpha} b_{5\alpha} - b_{1\alpha} b_{6\alpha} + b_{2\alpha} b_{6\alpha}) \right. \\ + b_{3\alpha} b_{6\alpha} - b_{4\alpha} b_{6\alpha} + b_{3\alpha} b_{7\alpha} - b_{3\alpha} b_{8\alpha} + b_{4\alpha} b_{8\alpha}) \\ + \sum_{(\alpha, \beta) = (a, c), (b, d)} (b_{1\alpha} b_{8\beta} + b_{8\alpha} b_{1\beta} - b_{2\alpha} b_{8\beta} \\ - b_{8\alpha} b_{2\beta} - b_{7\alpha} b_{1\beta} - b_{1\alpha} b_{7\beta}) \\ + \sum_{(\alpha, \beta) = (a, b), (c, d)} (b_{5\alpha} b_{4\beta} + b_{4\alpha} b_{5\beta} - b_{2\alpha} b_{5\beta} \\ - b_{5\alpha} b_{2\beta} - b_{4\alpha} b_{7\beta} - b_{7\alpha} b_{4\beta}) \\ + \sum_{(\alpha, \beta) = (a, d), (c, b)} (b_{2\alpha} b_{7\beta} + b_{7\alpha} b_{2\beta}) \left. \right\}, \quad (38)$$

$$I_{1a,1a}^{4 \times 4} = \frac{1}{32} \sum_{\alpha = a, b, c, d} (b_{1\alpha}^2 + b_{2\alpha}^2 + b_{3\alpha}^2 + b_{4\alpha}^2 \\ + b_{5\alpha}^2 + b_{6\alpha}^2 + b_{7\alpha}^2 + b_{8\alpha}^2). \quad (39)$$

Each of the invariants listed so far for the 4×4 unit cell involves products of bonds that lie sufficiently close to each other so that they also generate an invariant for the smaller 2×2 cell, and their contribution to scalar physical properties can be estimated from calculations for the smaller 2×2 cell. In other words, if the α 's in Eq. (7) were determined for the 2×2 cell, then an estimate for the properties of the larger number of H-bond isomers of the 4×4 cell would be available.

If the energy or free energy of the 2×2 unit cell was parametrized according to the value of $I_{13}^{2 \times 2}$, $I_{23}^{2 \times 2}$, $I_{12}^{2 \times 2}$, $I_{15}^{2 \times 2}$, and $I_{11}^{2 \times 2}$, then a first guess for the energy of configurations of the 4×4 cell would be in terms of the invariants in Eqs. (34)–(39). At this level of approximation, the parameters needed to describe the many H-bond isomers of the 4×4 cell, the α 's of Eq. (16), would be known from calculations for the smaller 2×2 cell, and only the enumeration of topologies required for the 4×4 cell. Perhaps comparison with more expensive, detailed calculations for the 4×4 cell would indicate reasonable convergence of the energy. If not, use of invariants involving bond pairs further separated from each other would be an option to improve the description. This would involve invariants for the 4×4 cell which have no counterpart in the 2×2 cell and are listed below.

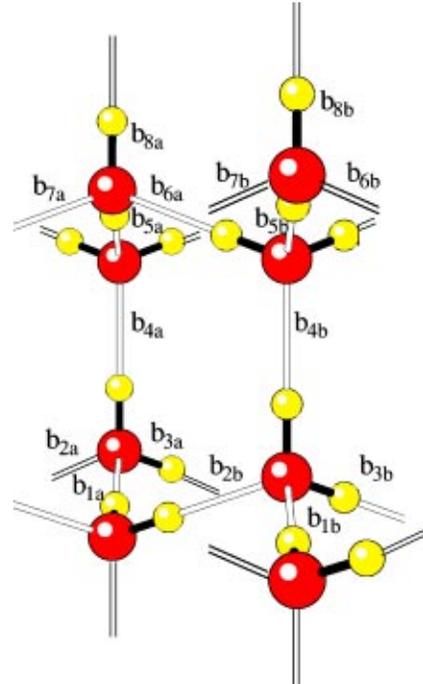


FIG. 4. The labeling of H bonds, and their canonical orientation, are shown here for the Orth($1 \times 1 \times 1$) unit cell. In the canonical orientation, all of the H bonds are cis.

$$I_{1a,3b}^{4 \times 4} = \frac{1}{32} \left\{ \sum_{(\alpha,\beta)=(a,b),(c,d),(a,d),(c,b)} (b_{1\alpha}b_{3\beta} + b_{3\alpha}b_{1\beta} + b_{4\alpha}b_{2\beta} + b_{2\alpha}b_{4\beta}) \right. \\ \left. + \sum_{(\alpha,\beta)=(a,c),(b,d),(a,d),(c,b)} (b_{6\alpha}b_{5\beta} + b_{5\alpha}b_{6\beta} + b_{8\alpha}b_{7\beta} + b_{7\alpha}b_{8\beta}) \right\}, \quad (40)$$

$$I_{1a,2c}^{4 \times 4} = \frac{1}{32} \left\{ \sum_{(\alpha,\beta)=(a,c),(c,b),(a,d),(b,d)} (b_{1\alpha}b_{2\beta} + b_{2\alpha}b_{1\beta} + b_{3\alpha}b_{4\beta} + b_{4\alpha}b_{3\beta}) \right. \\ \left. + \sum_{(\alpha,\beta)=(a,b),(c,b),(a,d),(c,d)} (b_{5\alpha}b_{7\beta} + b_{7\alpha}b_{5\beta} + b_{6\alpha}b_{8\beta} + b_{8\alpha}b_{6\beta}) \right\}, \quad (41)$$

$$I_{1a,5c}^{4 \times 4} = \frac{1}{128} \left\{ \sum_{\alpha=a,b,c,d} (b_{2\alpha}b_{5\alpha} - b_{4\alpha}b_{5\alpha} + b_{1\alpha}b_{7\alpha} + b_{4\alpha}b_{7\alpha} - b_{1\alpha}b_{8\alpha} + b_{2\alpha}b_{8\alpha}) \right. \\ + \sum_{(\alpha,\beta)=(a,c),(a,b),(b,d),(c,d)} (b_{1\alpha}b_{6\beta} + b_{6\alpha}b_{1\beta} - b_{3\alpha}b_{6\beta} - b_{6\alpha}b_{3\beta} + b_{4\alpha}b_{6\beta} + b_{6\alpha}b_{4\beta} - b_{1\alpha}b_{5\beta} - b_{5\alpha}b_{1\beta} + b_{3\alpha}b_{5\beta} \\ + b_{5\alpha}b_{3\beta} - b_{2\alpha}b_{6\beta} - b_{6\alpha}b_{2\beta} - b_{2\alpha}b_{7\beta} - b_{7\alpha}b_{2\beta} - b_{3\alpha}b_{7\beta} - b_{7\alpha}b_{3\beta} - b_{4\alpha}b_{8\beta} - b_{8\alpha}b_{4\beta} + b_{3\alpha}b_{8\beta} + b_{8\alpha}b_{3\beta}) \\ + \sum_{(\alpha,\beta)=(a,d),(c,b)} (b_{1\alpha}b_{7\beta} + b_{7\alpha}b_{1\beta} - b_{1\alpha}b_{8\beta} - b_{8\alpha}b_{1\beta} + b_{2\alpha}b_{5\beta} + b_{5\alpha}b_{2\beta} + b_{2\alpha}b_{8\beta} + b_{8\alpha}b_{2\beta} - b_{4\alpha}b_{5\beta} - b_{5\alpha}b_{4\beta} \\ \left. + b_{4\alpha}b_{7\beta} + b_{7\alpha}b_{4\beta}) \right\}, \quad (42)$$

$$I_{1b,5c}^{4 \times 4} = \frac{1}{64} \left\{ \sum_{(\alpha,\beta)=(b,c),(a,d)} (b_{1\alpha}b_{5\beta} + b_{5\alpha}b_{1\beta} - b_{1\alpha}b_{6\beta} - b_{6\alpha}b_{1\beta} + b_{2\alpha}b_{6\beta} + b_{6\alpha}b_{2\beta} + b_{4\alpha}b_{8\beta} + b_{8\alpha}b_{4\beta}) \right. \\ - b_{5\alpha}b_{3\beta} - b_{3\alpha}b_{5\beta} + b_{6\alpha}b_{3\beta} + b_{3\alpha}b_{6\beta} + b_{3\alpha}b_{7\beta} + b_{7\alpha}b_{3\beta} - b_{3\alpha}b_{8\beta} - b_{8\alpha}b_{3\beta} - b_{6\alpha}b_{4\beta} - b_{4\alpha}b_{6\beta}) \\ + \sum_{(\alpha,\beta)=(a,c),(b,d)} (b_{4\alpha}b_{5\beta} + b_{5\alpha}b_{4\beta} - b_{2\alpha}b_{5\beta} - b_{5\alpha}b_{2\beta} - b_{4\alpha}b_{7\beta} - b_{7\alpha}b_{4\beta}) \\ \left. + \sum_{(\alpha,\beta)=(a,b),(c,d)} (b_{1\alpha}b_{8\beta} + b_{8\alpha}b_{1\beta} - b_{1\alpha}b_{7\beta} - b_{7\alpha}b_{1\beta} - b_{2\alpha}b_{8\beta} - b_{8\alpha}b_{2\beta}) + \sum_{\alpha=a,b,c,d} b_{2\alpha}b_{7\alpha} \right\}, \quad (43)$$

$$I_{1b,1c}^{4 \times 4} = \frac{1}{16} \left\{ \sum_{(\alpha,\beta)=(b,c),(a,d)} (b_{1\alpha}b_{1\beta} + b_{2\alpha}b_{2\beta} + b_{3\alpha}b_{3\beta} + b_{4\alpha}b_{4\beta} + b_{5\alpha}b_{5\beta} + b_{6\alpha}b_{6\beta} + b_{7\alpha}b_{7\beta} + b_{8\alpha}b_{8\beta}) \right\}, \quad (44)$$

$$I_{1a,1c}^{4 \times 4} = \frac{1}{16} \left\{ \sum_{(\alpha,\beta)=(a,c),(b,d)} (b_{1\alpha}b_{1\beta} + b_{2\alpha}b_{2\beta} + b_{3\alpha}b_{3\beta} + b_{4\alpha}b_{4\beta}) + \sum_{(\alpha,\beta)=(a,b),(c,d)} (b_{5\alpha}b_{5\beta} + b_{6\alpha}b_{6\beta} + b_{7\alpha}b_{7\beta} + b_{8\alpha}b_{8\beta}) \right\}, \quad (45)$$

$$I_{1a,1b}^{4 \times 4} = \frac{1}{16} \left\{ \sum_{(\alpha,\beta)=(a,b),(c,d)} (b_{1\alpha}b_{1\beta} + b_{2\alpha}b_{2\beta} + b_{3\alpha}b_{3\beta} + b_{4\alpha}b_{4\beta}) + \sum_{(\alpha,\beta)=(a,c),(b,d)} (b_{5\alpha}b_{5\beta} + b_{6\alpha}b_{6\beta} + b_{7\alpha}b_{7\beta} + b_{8\alpha}b_{8\beta}) \right\}. \quad (46)$$

Equations (40)–(46), like Eqs. (34)–(39), reduce to graph invariants found for the 2×2 unit cell when the letter subscripts are removed, thereby enforcing the periodicity of the smaller cell. However, it is crucial to realize that the seven invariants in Eqs. (40)–(46) are fundamentally different in nature because they involve products of bonds further separated in the lattice than the five invariants of Eqs. (34)–(39).

IV. GRAPH INVARIANTS AND GRAPH ENUMERATION FOR ICE-Ih

In this section, we report graph invariants and graph enumeration results for ice-Ih. Historically, both orthorhombic and hexagonal unit cells have been used in the study of ice-Ih, most often the former, due to the convenience of using

orthogonal unit vectors. The symmetry of the oxygen lattice of ice-Ih has been identified as $P6_3/mmc$ since the late 1920s [41,42]. Some recent experiments reveal a transition to a low-temperature, proton-ordered phase, ice-XI, in which the hexagonal symmetry of the oxygen lattice of ice-Ih is broken by ordering of the hydrogen bonds [9–17]. The space group of ice-XI is $Cmc2_1$.

A. Invariants for the eight-molecule water orthorhombic unit cell

The series of ice-Ih unit cells formed by the cell vectors,

$$\begin{aligned} \mathbf{a} &= \sqrt{\frac{8}{3}}R\hat{\mathbf{x}}, \\ \mathbf{b} &= \sqrt{8}R\hat{\mathbf{y}}, \\ \mathbf{c} &= \frac{8}{3}R\hat{\mathbf{z}}, \end{aligned} \quad (47)$$

has been popular because these unit cell vectors are conveniently orthogonal. In Eq. (47), R is the distance between nearest neighbor oxygens, and $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are Cartesian unit vectors. The smallest orthorhombic unit cell is an eight-molecule unit cell (Fig. 4). Extending our notation to distinguish several different choices of unit cell vectors for ice-Ih, we use $\text{Orth}(n_a \times n_b \times n_c)$ in place of $O(n_x \times n_y \times n_z)$ to designate a unit cell obtained by extending the smallest orthorhombic unit cell n_a times along the a axis, n_b times along the b axis, and n_c times along the c axis. Hence the eight-water molecule orthorhombic cell is labeled $\text{Orth}(1 \times 1 \times 1)$, and there are three unit cells, $\text{Orth}(2 \times 1 \times 1)$, $\text{Orth}(1 \times 2 \times 1)$, and $\text{Orth}(1 \times 1 \times 2)$, having 16 water molecules, but their geometries are very different.

All first-order graph invariants for the ice lattice are identically zero. A set of 13 second-order graph invariants is obtained by projecting on all bond pairs from the $\text{Orth}(1 \times 1 \times 1)$ unit cell according to Eq. (31). If the projection operator acts on a pair of bonds that lie along the c axis, then all symmetry operations will produce other pairs that also lie along the c axis. Hence there is a subset of second-order invariants, three in all, that are composed totally of bond pairs along the c direction. Similarly, there is another subset of two second-order invariants composed totally of bond pairs that lie within the a - b bilayers. Finally, there is a third subset of eight invariants coupling bonds from the c and a - b bilayers.

The following set of invariants are constructed exclusively from bonds along the c direction:

$$I_{4a,4b} = \frac{1}{12} \{b_{4a}^2 + b_{4b}^2 + 4b_{4a}b_{4b} + b_{8a}^2 + b_{8b}^2 + 4b_{8a}b_{8b}\}, \quad (48)$$

$$I_{4a,8a} = \frac{1}{6} \{2b_{4a}b_{8a} + b_{4a}b_{8b} + 2b_{4b}b_{8b} + b_{4b}b_{8a}\}, \quad (49)$$

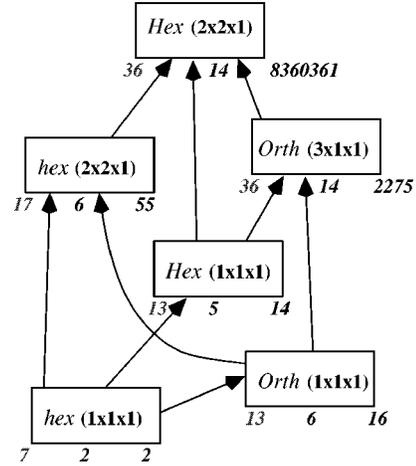


FIG. 5. Each unit cell is accompanied by three numbers, which from left to right are the number of second-order graph invariants, the number of linearly independent invariants for graphs that satisfy the ice rules for neutral water, and the number of symmetry-distinct H-bond configurations for that unit cell.

$$I_{4a,4a} = \frac{1}{4} \{b_{4a}^2 + b_{4b}^2 + b_{8a}^2 + b_{8b}^2\}. \quad (50)$$

$I_{4a,4b}$ describes interactions between c -axis bonds that are nearest neighbors above the same bilayer. (The squared term b_{4a}^2 appears because b_{4a} is nearest neighbor to its own image in an adjacent unit cell, and similarly for the other squared terms in $I_{4a,4b}$.) This invariant is what one would obtain if one were to map the energy of a two-dimensional sheet of c -axis bonds onto a two-dimensional triangular lattice. The possible dependence of energy upon ferroelectricity or antiferroelectricity of c -axis bonds will be captured by these invariants. Notice how the possible coupling between c -axis bonds between two bilayers would be described at this level by invariants $I_{4a,8a}$.

The remainder of the invariants for the $\text{Orth}(1 \times 1 \times 1)$ cell are presented in the Appendix. We have also generated invariants and determined the number of linear independent invariants for much larger unit cells. The total number of second-order invariants and the number of linearly independent invariants are reported below in Fig. 5, but we do not report the explicit form of invariants for larger unit cells in this paper. Properties that have been postulated to affect the energy of the ice lattice, such as the number of cis or trans H bonds [43] in the lattice or the degree of ferroelectricity in the lattice, can be expressed as linear combinations of these second-order invariants. In addition, the invariants must also describe other topological features that have not been discussed in the literature, but which have not been ruled out as possible factors affecting the energy. Having the full set of second-order invariants allows an unbiased analysis of which topological features are most significant.

The graph invariants can be viewed as forming a set of basis vectors in a space of H-bond configurations. For example, there are 16 symmetry-distinct H-bond arrangements possible for the $\text{Orth}(1 \times 1 \times 1)$ unit cell that we picture as forming a 16-dimensional vector space. (Further discussion

of the enumeration of the actual H-bond arrangements is given below in Sec. IV B 1.) Each invariant is mapped to a vector whose 16 components are the value of that invariant evaluated for each of the H-bond arrangements. Since there are only 13 second-order invariants, they cannot possibly span the 16-dimensional space of symmetry-distinct H-bond arrangements. In fact, for H-bond configurations that obey the ice rules; there is a high degree of linear dependence among the second-order invariants. Consequently, the space spanned by the second-order invariants is only of dimension 6. Of course, adding third- and higher-order invariants would incorporate further flexibility and more fitting parameters. Our experience to date for clusters indicates that energetics is well described at the level of second-order invariants [21], but this conclusion will have to be tested for ice. If truncation at second-order invariants is found to be a reasonable approximation for ice, this provides strong constraints on how scalar physical properties might depend on H-bond topology.

B. Enumeration of H-bond arrangements in ice-Ih

In addition to their usefulness in describing physical properties, graph invariants also provide an efficient means to generate all symmetry-distinct H-bond arrangements of a finite, or periodic, system. Eliminating symmetry equivalent configurations from a list of N graphs is nominally an $O(N^2)$ operation, because all pairs should be compared for symmetry equivalence. However, graphs with different values of any invariant must be symmetry distinct. This suggests an efficient scheme for eliminating symmetry equivalent graphs. The list of graphs is partitioned into subsets such that each subset contains graphs with unique values of one or more invariants. Hence, a graph in one subset cannot be symmetry equivalent to another graph in a different subset. As a result, symmetry equivalence need only be tested among graphs in the same subset, reducing the operation count from $O(N^2)$ to $O(N \ln N)$. Details are furnished in our previous work [21].

Results of this efficient enumeration scheme are presented here for several unit cells of ice-Ih. We provide results for small unit cells because, as mentioned in the introduction, there are conflicting results in the literature for the symmetry-distinct H-bond arrangements of cells such as the eight-member orthorhombic cell. We also provide results for large unit cells, cells whose size can be considered appropriate for statistical simulations, to identify topological properties of the ice lattice in the statistical limit, and to demonstrate the feasibility of large-scale enumeration.

1. Enumeration results for the eight-member orthorhombic cell

In 1987 Howe reported that 17 symmetry-distinct H-bond arrangements were possible for the smallest orthorhombic cell, $\text{Orth}(1 \times 1 \times 1)$ [6]. In 1998, Lekner enumerated the 114 H-bond arrangements possible for $\text{Orth}(1 \times 1 \times 1)$ before symmetry reduction, and then eliminated redundant structures according to the functional form of the Coulomb interaction [7]. This is not necessarily the same as reduction by symmetry equivalence. Lekner observed 17 distinct forms of the Coulomb potential function. Among the 17 functional

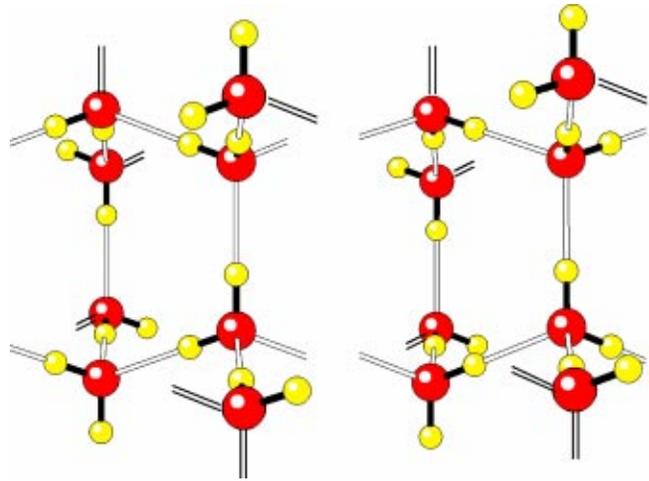


FIG. 6. These two configurations of the $\text{Orth}(1 \times 1 \times 1)$ cell are symmetry distinct, yet are related to each other by reversal of all H bonds. The left-hand structure is converted into the right-hand one by first reflecting through a horizontal plane that bisects the figure midway between the two a - b bilayers, followed by reversing all the H bonds.

forms, two pairs had the same value, leading to only 15 distinct Coulomb energies. In 1998 Buch, Sandler, and Sadlej sought to enumerate the distinct configurations of the $\text{Orth}(1 \times 1 \times 1)$ cell [8]. They employed a Monte Carlo scheme to generate H-bond configurations for $\text{Orth}(1 \times 1 \times 1)$, and then eliminated redundant configurations according to physical properties such as total energy and dipole moment. Buch and co-workers found 16 distinct arrangements.

In this work, we use symmetry properties of the H-bond topology to eliminate redundant configurations. This is the same criterion used by Howe, but we obtain different results. Using the functional form of a potential function is problematic because symmetry-distinct structures may have the same energy for certain potential functions, but the degeneracy may be lifted for other potentials. To give an elementary example originally noted by Lekner [7], the total oxygen-oxygen and oxygen-hydrogen Coulomb interaction is identical for all H-bond topologies in an idealized ice-Ih lattice where all covalent and nearest neighbor bond lengths are equal. Differences among the H-bond topologies arise exclusively from hydrogen-hydrogen Coulomb interactions. We will encounter a more subtle example below, in which two structures have identical bond lengths, and therefore are degenerate with respect to all pairwise additive potentials, yet are symmetry distinct. Finally, Monte Carlo methods may be used exhaustively for the smallest unit cells, but would be highly impractical for exhaustive enumeration of some of the larger unit cells we present below.

Following the enumeration procedure described in our earlier work [21], we obtain 16 symmetry-distinct H-bond topologies for $\text{Orth}(1 \times 1 \times 1)$. Two configurations out of the 16 are of particular note (Fig. 6). These configurations are related to each other by reversal of all H-bonds, yet are symmetry distinct. In an idealized structure where covalent and nearest neighbor oxygen-oxygen distances are the same for

all molecules, we have verified that the two structures shown in Fig. 6 have identical distributions of pair distances out to $2.9R$. From the structural information given in Lekner's work, it is clear that these two structures are the pair that Lekner found to be energetically degenerate with respect to the Coulomb potential, as they would be for any pairwise additive potential.

2. Sequences of unit cells for ice-Ih

The primitive unit cell of ice-Ih is defined by the following unit vectors:

$$\begin{aligned} \mathbf{a} &= \sqrt{\frac{8}{3}}R\hat{\mathbf{x}}, \\ \mathbf{b} &= -\sqrt{\frac{2}{3}}R\hat{\mathbf{x}} + \sqrt{2}R\hat{\mathbf{y}}, \\ \mathbf{c} &= \frac{8}{3}R\hat{\mathbf{z}}. \end{aligned} \quad (51)$$

We use $\text{hex}(n_a \times n_b \times n_c)$ to label unit cells built from multiples of the primitive unit cell.

We will also consider another hexagonal system constructed from the following unit vectors:

$$\begin{aligned} \mathbf{a} &= \sqrt{6}R\hat{\mathbf{x}} + \sqrt{2}R\hat{\mathbf{y}}, \\ \mathbf{b} &= \sqrt{8}R\hat{\mathbf{y}}, \\ \mathbf{c} &= \frac{8}{3}R\hat{\mathbf{z}}. \end{aligned} \quad (52)$$

These unit cells, designated here as $\text{Hex}(n_a \times n_b \times n_c)$, form a convenient sequence of when $n_x = n_y = n$. The $\text{hex}(n \times n \times n_c)$ and $\text{Hex}(n \times n \times n_c)$ cells can be taken to be prisms with the full hexagonal symmetry of the ice-Ih lattice. [An example of the $\text{Hex}(1 \times 1 \times 1)$ unit cell is shown in Fig. 9(b).]

The $\text{Orth}(n_a \times n_b \times n_c)$, $\text{hex}(n_a \times n_b \times n_c)$, and $\text{Hex}(n \times n \times n_c)$ cells stand in relation to each other as shown in Fig. 5. The arrows in that diagram represent a membership relation between the oriented graphs of the cells linked by an arrow. When the set of oriented graphs of a smaller unit cell is a subset of the graphs of a larger unit cell, the two cells are joined by an arrow. In effect, the arrows represent a chain by which an invariant from smaller unit cells can predict the properties of larger cells. Also shown in Fig. 5 are the number of symmetry-distinct graphs that would arise from the ice rules for neutral water, second-order graph invariants, and linearly independent second-order graph invariants for neutral water graphs. The number of linearly independent graph invariants appears to level off and remain quite small, reaching only 14 for the $\text{Orth}(3 \times 1 \times 1)$ and $\text{Hex}(2 \times 2 \times 1)$ cells.

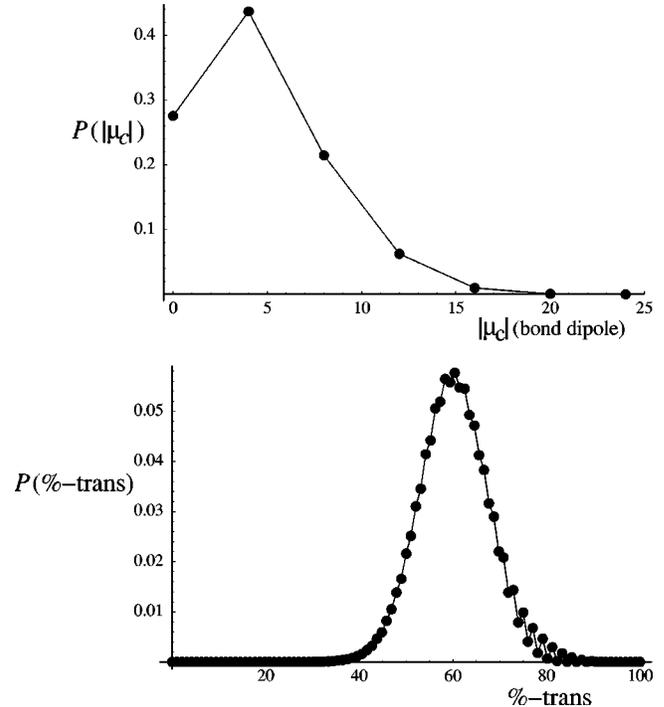


FIG. 7. The top panel shows the distribution of dipole moment magnitude arising from H bonds along the c direction in a 48-water molecule unit cell [$\text{Hex}(2 \times 2 \times 1)$] of ice-Ih. Measured in bond dipoles, the maximum dipole moment is 24, the number of H bonds along the c axis. The bottom panel shows the distribution of percent trans H bonds among the 96 H bonds of the unit cell.

C. Analysis of enumeration results

While energetic calculations are beyond the scope of this paper, constraints on ice structures, as revealed by enumeration of H-bond topologies, do provide some insights into the behavior of ice-Ih and possible low-temperature phases. Assuming a random distribution of H-bond topologies, the standard model going back to the work of Pauling [1], we are able here to predict some statistical properties of ice-Ih, such as dipole moment (in a bond dipole model). In most previous simulations of ice-Ih, the H-bond topology in the simulation cell has been chosen to have zero dipole moment, and often minimum higher multipoles [40,44,45]. Here we report on how likely or unlikely these low multipole configurations will be. Our explorations will also categorize possible candidates for the low-temperature phase of ordinary ice. Calculations of water dimer indicate that the lowest energy topology would contain maximum fraction of trans bonds, yet recent experiments have been interpreted to favor a ferroelectric structure, where the fraction of trans is 25%, far from optimal according to Bjerrum's conjecture [25,26]. We will explore the correlation between ferroelectricity and fraction of trans bonds.

We have accumulated data on H-bond geometry and dipole moments for the variety of unit cells shown in Fig. 5. The results are all qualitatively the same, both for the orthorhombic and hexagonal cells. Therefore, we present the results for the largest unit cell, $\text{Hex}(2 \times 2 \times 1)$, whose

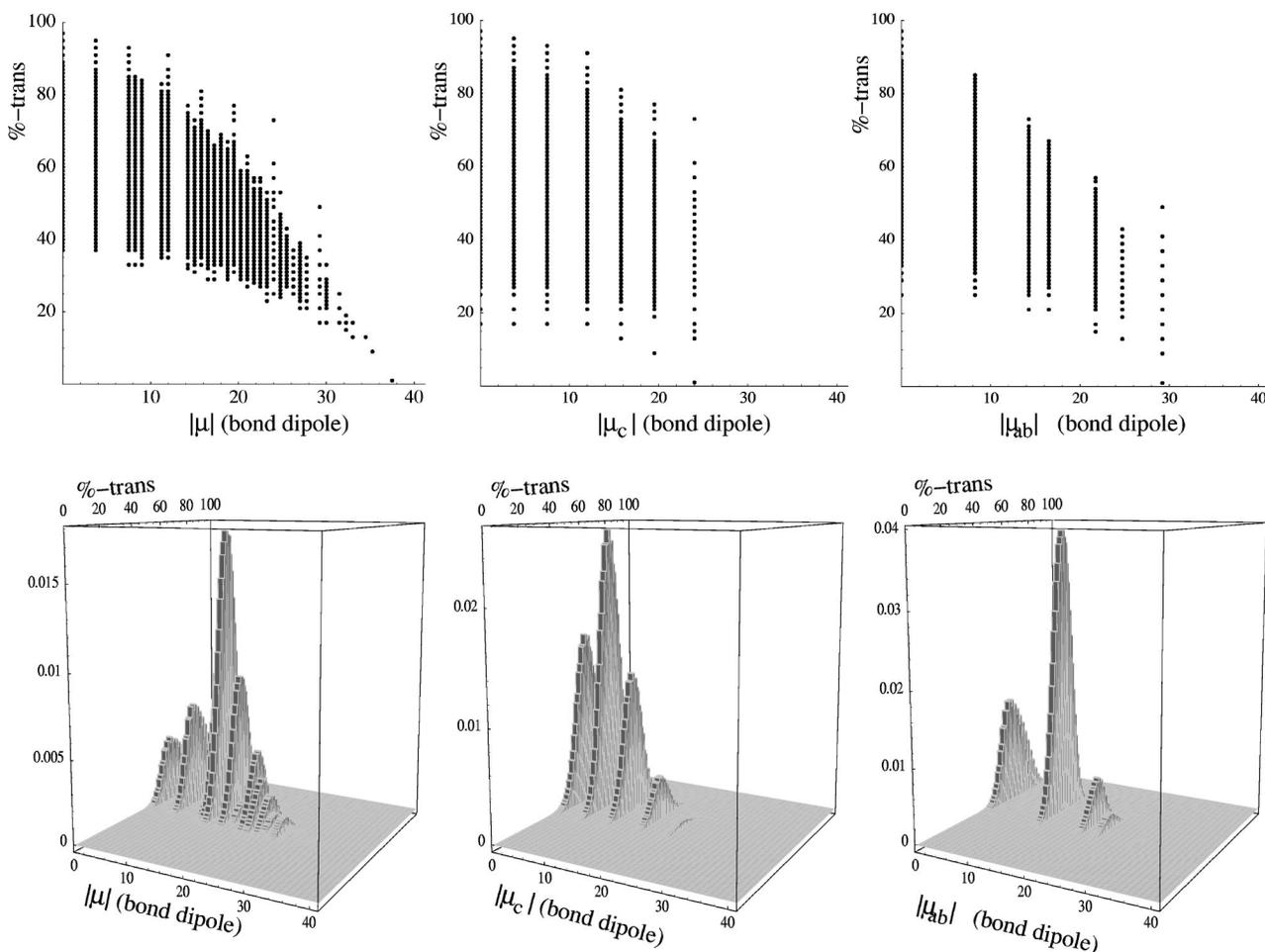


FIG. 8. Scatter plot (top row) and three-dimensional representations (bottom row) of the distribution of H-bond isomers in ice-Ih resolved according to dipole moment and percent trans H bonds. The three-dimensional plots best convey where the bulk of the distribution is located, while the scatter plots depict the locus of possible structures, regardless of their frequency. From left to right are the distribution of total dipole moment of the unit cell, dipole moment generated by H bonds along the c axis, and dipole moment generated by H bonds lying within the puckered hexagonal sheets parallel to the a and b crystallographic axes. The data were generated for the 48-water molecule $\text{Hex}(2 \times 2 \times 1)$ unit cell, for which there are 2404 144 962 isomers satisfying periodic ice rules, of which 8360 361 are symmetry distinct. The dipole moment is reported in units of OH bond dipoles.

2404 144 962 configurations (8360 361 symmetry distinct), best approximate the infinite system limit. The dipole moment is calculated in a bond dipole approximation, and the bond dipoles are assumed to be parallel to the oxygen-oxygen vector of the H bonds. We report dipoles in units of the bond dipoles. Dipole moment arising from bonds along the c axis, the origin of ferroelectricity in the proposed $Cmc2_1$ structure of ice-XI, are of particular interest. The distribution, shown in the top panel of Fig. 7, indicates that complete alignment of H bonds along the c axis is extremely rare. Zero dipole moment, as often imposed in computer simulations of ice-Ih [40,44,45], is relatively frequent, but still is not typical, only occurring in 27.5% of H-bond arrangements. The bottom panel of Fig. 7 shows that nearly all H-bond arrangements contain a percent of trans H bonds between 40% and 80%, with the maximum near 60%.

The proposed ferroelectric $Cmc2_1$ structure of ice-XI is rather unusual among ferroelectric structures in that it contains a small fraction of trans H bonds. In the

$\text{Orth}(1 \times 1 \times 1)$ unit cell, the H-bond arrangements with H bonds completely aligned along the c axis tend to have small percentage of trans bonds: one isomer has 50% trans, four (including $Cmc2_1$) have 25% trans, one has 12.5% trans,

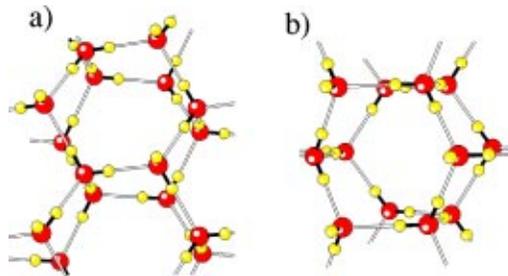


FIG. 9. Two examples of small unit cells with complete ferroelectric order along the c axis coexisting with a high percentage of trans bonds. (a) A 16-water molecule $\text{Orth}(2 \times 1 \times 1)$ cell with 62% trans bonds, (b) a 12-water molecule $\text{Hex}(1 \times 1 \times 1)$ cell with 75% trans bonds.

and one has 0% trans. This is an artifact of the small unit cell size. In larger unit cells the completely aligned structures tend to have a larger percentage of trans bonds. This property is illustrated in the two-dimensional distributions of dipole moment and fraction of trans bonds shown in Fig. 8. Looking at the top, center panel of Fig. 8, we find that the percentage of trans H bonds in arrangements with maximum ferroelectricity along the c axis extends from a minimum of 0% to a maximum of 75%.

Contrary to what one might expect by only considering the Orth($1 \times 1 \times 1$) unit cell, there exist several examples of slightly larger unit cells with complete ferroelectric order in the c direction and high percentage of trans bonds, as shown in Fig. 9. As one can see from Fig. 8, the $Cmc2_1$ structure is very atypical, at least with respect to dipole moment and fraction of trans H bonds. Optimization of a large unit cell by Monte Carlo methods [8,40] would be unlikely to uncover either the $Cmc2_1$ structure or the examples of Fig. 9. Enumeration is an important complement to Monte Carlo search.

V. CONCLUSION

Larger unit cells of ice-Ih, needed to simulate thermal properties and phase transitions, can be arranged in astronomically large numbers of H-bond configurations. Energy differences among these configurations are rather small, so accurate and expensive *ab initio* methods are likely to be required to understand the low-temperature behavior of ice-Ih. Graph invariants provide a means of describing the energy, free energy, and other scalar physical properties of the large number of configurations using only a handful of parameters. It is significant to note that, even though the number of H-bond arrangements grows exponentially with system size, the number of linear independent invariants grows quite slowly and appears to approach a finite limit. Features of the H-bond topology in ice-Ih previously suggested as determinants of the energy, such as trans and cis H bonds [25,26], emerge in our theory as some of the possible low-order invariants. However, graph invariants provide many other possible links between H-bond topology and scalar physical properties which have not been considered, but may turn out to be significant.

The hierarchy of approximations provided by graph invariants can be arranged on a two-dimensional grid. On one axis, the level of approximation is distinguished by the number of bond variables multiplied together in each term. Invariants can be constructed as linear combinations of single bond variables (first-order invariants), products of two bond variables (second-order invariants), three bond variables (third-order invariants), and so on. We hope that the expansion of scalar physical properties in terms of invariants converges with relatively low-order invariants, a property we have demonstrated for finite clusters of water molecules [21]. The second axis of the grid of approximations is unique to periodic systems. The crystal can be constructed from unit cells of different size, with large unit cells needed to describe a disordered solid such as the H-bond disordered phase of ice-Ih. Graph invariants can be ordered according to whether they are a linear combination of products of bonds from a

small unit cell, or whether the bonds only belong to a larger unit cell. We have shown that graph invariants of larger unit cells fall into two categories. The first class involves bonds that are close enough to be part of a small unit cell. The dependence of the energy on these invariants can be determined from smaller unit cells, and used to determine the energy of the much larger number of H-bond arrangements of the large cell. The second class of invariants involves products of bonds that only occur in the large cell. Eventually, as the unit cell is progressively enlarged, the contribution arising from far-away bonds will become negligible. The rate of convergence for this second of the two axes in the grid of approximations has not yet been tested, and awaits the results of periodic *ab initio* calculations.

ACKNOWLEDGMENTS

This research was supported by NSF Grant No. CHE-0109243. J.L.K. would like to thank the Ohio State University for support.

APPENDIX: INVARIANTS OF THE Orth($1 \times 1 \times 1$) CELL

There are 13 independent invariants for the orthorhombic $O(1 \times 1 \times 1)$ unit cell of ice-Ih. Three of those, Eqs. (48)–(50), are discussed in Sec. IV A. The remaining are presented in this appendix.

The following invariants involve products of bonds that lie within the same bilayer:

$$I_{1a,2b} = \frac{1}{24} (2b_{5b}b_{6a} + 2b_{5a}b_{6b} - 2b_{1b}b_{2a} - 2b_{1a}b_{2b} + b_{1a}b_{3a} + b_{1b}b_{3a} + b_{1a}b_{3b} + b_{1b}b_{3b} + b_{2a}b_{3a} + b_{2b}b_{3a} + b_{2a}b_{3b} + b_{2b}b_{3b} + b_{5a}b_{7a} + b_{5b}b_{7a} + b_{5a}b_{7b} + b_{5b}b_{7b} - b_{6a}b_{7a} - b_{6b}b_{7a} - b_{6a}b_{7b} - b_{6b}b_{7b}), \quad (A1)$$

$$I_{1a,2a} = \frac{1}{24} (2b_{5a}b_{6a} + 2b_{5b}b_{6b} - 2b_{1a}b_{2a} - 2b_{1b}b_{2b} + b_{1a}b_{3a} + b_{1b}b_{3a} + b_{1a}b_{3b} + b_{1b}b_{3b} + b_{2a}b_{3a} + b_{2b}b_{3a} + b_{2a}b_{3b} + b_{2b}b_{3b} + b_{5a}b_{7a} + b_{5b}b_{7a} + b_{5a}b_{7b} + b_{5b}b_{7b} - b_{6a}b_{7a} - b_{6b}b_{7a} - b_{6a}b_{7b} - b_{6b}b_{7b}), \quad (A2)$$

$$I_{1a,1b} = \frac{1}{24} (4b_{3a}b_{3b} + 4b_{7a}b_{7b} + 2b_{1a}b_{1b} + 2b_{2a}b_{2b} + 2b_{6a}b_{6b} + 2b_{5a}b_{5b} + b_{1a}^2 + b_{1b}^2 + b_{2a}^2 + b_{2b}^2 + b_{5a}^2 + b_{5b}^2 + b_{6a}^2 + b_{6b}^2), \quad (A3)$$

$$I_{1a,1a} = \frac{1}{12}(b_{1a}^2 + b_{1b}^2 + b_{2a}^2 + b_{2b}^2 + b_{3a}^2 + b_{3b}^2 + b_{5a}^2 + b_{5b}^2 + b_{6a}^2 + b_{6b}^2 + b_{7a}^2 + b_{7b}^2). \quad (\text{A4})$$

The next set of invariants involves bonds that lie within adjacent bilayers.

$$I_{1a,6b} = \frac{1}{24}(2b_{2b}b_{5a} + 2b_{2a}b_{5b} - 2b_{1b}b_{6a} - 2b_{1a}b_{6b} - b_{3a}b_{5a} - b_{3b}b_{5a} - b_{3a}b_{5b} - b_{3b}b_{5b} + b_{3a}b_{6a} + b_{3b}b_{6a} + b_{3a}b_{6b} + b_{3b}b_{6b} - b_{1a}b_{7a} - b_{1b}b_{7a} - b_{1a}b_{7b} - b_{1b}b_{7b} - b_{2a}b_{7a} - b_{2b}b_{7a} - b_{2a}b_{7b} - b_{2b}b_{7b}), \quad (\text{A5})$$

$$I_{1a,6a} = \frac{1}{24}(2b_{2a}b_{5a} + 2b_{2b}b_{5b} - 2b_{1a}b_{6a} - 2b_{1b}b_{6b} - b_{3a}b_{5a} - b_{3b}b_{5a} - b_{3a}b_{5b} - b_{3b}b_{5b} + b_{3a}b_{6a} + b_{3b}b_{6a} + b_{3a}b_{6b} + b_{3b}b_{6b} - b_{1a}b_{7a} - b_{1b}b_{7a} - b_{2a}b_{7a} - b_{2b}b_{7a} - b_{1a}b_{7b} - b_{1b}b_{7b} - b_{2a}b_{7b} - b_{2b}b_{7b}), \quad (\text{A6})$$

$$I_{1a,5a} = \frac{1}{6}(b_{1a}b_{5a} + b_{1b}b_{5b} - b_{2a}b_{6a} - b_{2b}b_{6b} + b_{3b}b_{7a} + b_{3a}b_{7b}), \quad (\text{A7})$$

$$I_{1a,5b} = \frac{1}{12}(2b_{3a}b_{7a} + 2b_{3b}b_{7b} + b_{1a}b_{5a} + b_{1b}b_{5a} + b_{1a}b_{5b} + b_{1b}b_{5b} - b_{2a}b_{6a} - b_{2b}b_{6a} - b_{2a}b_{6b} - b_{2b}b_{6b}). \quad (\text{A8})$$

Finally, there is a set of invariants that couple bonds that lie along the c axis with bonds in a bilayer.

$$I_{1a,4b} = \frac{1}{48}(b_{1a}b_{4a} + b_{1b}b_{4a} + b_{1a}b_{4b} + b_{1b}b_{4b} + b_{1a}b_{8a} + b_{1b}b_{8a} + b_{1a}b_{8b} + b_{1b}b_{8b} + b_{2a}b_{4a} + b_{2b}b_{4a} + b_{2a}b_{4b} + b_{2b}b_{4b} + b_{2a}b_{8a} + b_{2b}b_{8a} + b_{2a}b_{8b} + b_{2b}b_{8b} - b_{4a}b_{5a} - b_{4b}b_{5a} - b_{4a}b_{5b} - b_{4b}b_{5b} + b_{4a}b_{6a} + b_{4b}b_{6a} + b_{4a}b_{6b} + b_{4b}b_{6b} - b_{5a}b_{8a} - b_{5b}b_{8a} - b_{5a}b_{8b} - b_{5b}b_{8b} + b_{6a}b_{8b} + b_{6b}b_{8b} + b_{6a}b_{8a} + b_{6b}b_{8a} - 2b_{3b}b_{8b} - 2b_{3a}b_{8a} - 2b_{3a}b_{4b} - 2b_{3b}b_{4a} + 2b_{7b}b_{8a} + 2b_{4a}b_{7a} + 2b_{4b}b_{7b} + 2b_{7a}b_{8b}), \quad (\text{A9})$$

$$I_{1a,4a} = \frac{1}{24}(b_{1a}b_{4a} + b_{2a}b_{4a} - b_{3a}b_{4a} + b_{1b}b_{4b} + b_{2b}b_{4b} - b_{3b}b_{4b} - b_{4a}b_{5a} - b_{4b}b_{5b} + b_{4a}b_{6a} + b_{4b}b_{6b} + b_{4a}b_{7a} + b_{4a}b_{7b} + b_{1a}b_{8a} + b_{2a}b_{8a} - b_{3b}b_{8a} - b_{5a}b_{8a} + b_{6a}b_{8a} + b_{7a}b_{8a} + b_{1b}b_{8b} + b_{2b}b_{8b} - b_{3a}b_{8b} - b_{5b}b_{8b} + b_{6b}b_{8b} + b_{7b}b_{8b}). \quad (\text{A10})$$

-
- [1] L. Pauling, *J. Am. Chem. Soc.* **57**, 2680 (1935).
 [2] W.F. Giauque and J.W. Stout, *J. Am. Chem. Soc.* **58**, 1144 (1936).
 [3] J.D. Bernal and R.H. Fowler, *J. Chem. Phys.* **1**, 515 (1933).
 [4] E.A. DiMarzio and F.H. Stillinger, *J. Chem. Phys.* **40**, 1577 (1964).
 [5] J.F. Nagle, *J. Math. Phys.* **7**, 1484 (1966).
 [6] R. Howe, *J. Phys. (Paris)* **48**, C1 (1987).
 [7] J. Lekner, *Physica B* **252**, 149 (1998).
 [8] V. Buch, P. Sandler, and J. Sadlej, *J. Phys. Chem. B* **102**, 8641 (1998).
 [9] S. Kawada, *J. Phys. Soc. Jpn.* **32**, 1442 (1972).
 [10] Y. Tajima, T. Matsuo, and H. Suga, *Nature (London)* **299**, 810 (1982).
 [11] Y. Tajima, T. Matsuo, and H. Suga, *J. Phys. Chem. Solids* **45**, 1135 (1984).
 [12] T. Matsuo, Y. Tajima, and H. Suga, *J. Phys. Chem. Solids* **47**, 165 (1986).
 [13] T. Matsuo and H. Suga, *J. Phys. Colloq.* **C1**, 477 (1987).
 [14] R. Howe and R.W. Whitworth, *J. Chem. Phys.* **90**, 4450 (1989).
 [15] S.M. Jackson, V.M. Nield, R.W. Whitworth, M. Oguro, and C.C. Wilson, *J. Phys. Chem. B* **101**, 6142 (1997).
 [16] A.J. Leadbetter, R.C. Ward, J.W. Clark, P.A. Tucker, T. Matsuo, and H. Suga, *J. Chem. Phys.* **82**, 424 (1985).
 [17] C.M.B. Line and R.W. Whitworth, *J. Chem. Phys.* **104**, 10 008 (1996).
 [18] H. Fukazawa, S. Mae, S. Ikeda, and O. Watanabe, *Chem. Phys. Lett.* **294**, 554 (1998).
 [19] G.P. Johari and S.J. Jones, *J. Chem. Phys.* **62**, 4213 (1975).

- [20] M.J. Iedema, M.J. Dresser, D.L. Doering, J.B. Rowland, W.P. Hess, A.A. Tsekouras, and J.P. Cowin, *J. Phys. Chem. B* **102**, 9203 (1998).
- [21] J.-L. Kuo, J.V. Coe, S.J. Singer, Y.B. Band, and L. Ojamäe, *J. Chem. Phys.* **114**, 2527 (2001).
- [22] F. Harary, *Michigan Math. J.* **4**, 221 (1957).
- [23] F. Harary and E. Palmer, *Can. J. Math.* **18**, 853 (1966).
- [24] F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1969).
- [25] N. Bjerrum, *Science* **115**, 386 (1952).
- [26] K.S. Pitzer and J. Polissar, *J. Am. Chem. Soc.* **60**, 1140 (1956).
- [27] J.C. Li and D.K. Ross, *Nature (London)* **365**, 327 (1993).
- [28] J.S. Tse and D.D. Klug, *Phys. Lett. A* **198**, 464 (1995).
- [29] I. Morrison, J.-C. Li, S. Jenkins, S.S. Xantheas, and M.C. Payne, *J. Phys. Chem. B* **101**, 6146 (1997).
- [30] D.J. McGinty, *J. Chem. Phys.* **55**, 580 (1971).
- [31] J.J. Burton, *J. Chem. Phys.* **56**, 3133 (1972).
- [32] M.R. Hoare, *Adv. Chem. Phys.* **40**, 49 (1979).
- [33] F.H. Stillinger and T.A. Weber, *Phys. Rev. A* **25**, 978 (1982).
- [34] F.H. Stillinger and T.A. Weber, *Phys. Rev. A* **28**, 2408 (1983).
- [35] P.G. Mezey, *Potential Energy Hypersurfaces*, Studies in Physical and Theoretical Chemistry Vol. 53 (Elsevier, New York, 1987).
- [36] G. Franke, E.R. Hilf, and P. Borrmann, *J. Chem. Phys.* **98**, 3496 (1993).
- [37] D.J. Wales, *Mol. Phys.* **78**, 151 (1993).
- [38] J.P.K. Doye and D.J. Wales, *J. Chem. Phys.* **102**, 9659 (1995).
- [39] M.B. Boisen, Jr. and G.V. Gibbs, *Mathematical Crystallography*, 2nd ed., Reviews in Mineralogy Vol. 15 (Mineralogical Soc. of America, Washington, D.C., 1990).
- [40] A. Rahman and F.H. Stillinger, *J. Chem. Phys.* **57**, 4009 (1972).
- [41] W.H. Bragg, *Proc. Phys. Soc. London* **34**, 98 (1922).
- [42] W.H. Barnes, *Proc. R. Soc. London, Ser. A* **324**, 127 (1929).
- [43] To provide examples of the meaning of cis and trans H bonds in ice-Ih, we note that all of the H bonds in the configuration of Fig. 4 are cis. In Fig. 6, two bonds along the *c* axis are shown in both structures. (Another two *c* axis bonds are part of the unit cell, but are not shown because they connect to atoms in adjoining cells.) The left of the two *c* axis bonds in both structures of Fig. 6 is trans, while the right one is cis. In the left-hand structure the bonds in the lower *a-b* plane are all trans, while two bonds in the upper *a-b* plane are cis and four are trans. In the right-hand structure the configurations of the *a-b* planes are reversed. Figure 9 provides two examples with very large fraction of trans H bonds.
- [44] J.A. Hayward and J.R. Reimers, *J. Chem. Phys.* **106**, 1518 (1997).
- [45] E. Cota and W.G. Hoover, *J. Chem. Phys.* **67**, 3839 (1977).