Coarse-graining errors and numerical optimization using a relative entropy framework

Cite as: J. Chem. Phys. **134**, 094112 (2011); https://doi.org/10.1063/1.3557038
Submitted: 22 November 2010 . Accepted: 01 February 2011 . Published Online: 03 March 2011

Aviel Chaimovich and M. Scott Shell

Coarse-graining errors and numerical optimization using a relative entropy framework

Aviel Chaimovich and M. Scott Shell[a)]
Department of Chemical Engineering, University of California Santa Barbara, Santa Barbara, California 93106-5080, USA

(Received 22 November 2010; accepted 1 February 2011; published online 3 March 2011)

The ability to generate accurate coarse-grained models from reference fully atomic (or otherwise "first-principles") ones has become an important component in modeling the behavior of complex molecular systems with large length and time scales. We recently proposed a novel coarse-graining approach based upon variational minimization of a configuration-space functional called the relative entropy, S_{rel}, that measures the information lost upon coarse-graining. Here, we develop a broad theoretical framework for this methodology and numerical strategies for its use in practical coarse-graining settings. In particular, we show that the relative entropy offers tight control over the errors due to coarse-graining in arbitrary microscopic properties, and suggests a systematic approach to reducing them. We also describe fundamental connections between this optimization methodology and other coarse-graining strategies like inverse Monte Carlo, force matching, energy matching, and variational mean-field theory. We suggest several new numerical approaches to its minimization that provide new coarse-graining strategies. Finally, we demonstrate the application of these theoretical considerations and algorithms to a simple, instructive system and characterize convergence and errors within the relative entropy framework. © *2011 American Institute of Physics*. [doi:10.1063/1.3557038]

I. INTRODUCTION

Coarse-grained models have long been used to understand and probe the properties of complex, many-body molecular systems. In recent years, there has been a surge of interest in rigorous approaches to their development.[1–16] Two factors have motivated such efforts: the need for tractable theoretical treatments or numerical simulations, in order to evaluate system behavior, and the natural emergence of effective interactions at larger scales that suggests a cleaner understanding of the system, enabling in some cases closer connections with experiment.

One aspect of this problem has received particular attention: how to optimize the intermolecular interaction potential $U_{\text{CG}}(\mathbf{R})$ of a putative coarse-grained (CG) model with pseudoatom coordinates $\mathbf{R}$ so as to reproduce the behavior of a reference "first-principles" (FP) model with possibly more site coordinates $\mathbf{r}$ and known interactions $U_{\text{FP}}(\mathbf{r})$. Typically, the reference FP system is described by an all-atom classical potential, and the desired CG version is constructed either with fewer interaction sites or a simpler form of the interaction potential (or both). This problem requires a specification of how the CG model is designed; namely, it requires a method for mapping any FP configuration to a corresponding CG one, typically described by a mapping operator as $\mathbf{R} = \mathbf{M}(\mathbf{r})$.[11,12] For example, an equation for the center of mass might be invoked as an operation in mapping a group of FP atoms onto a sole CG molecular site. While not discussed here, it is noteworthy that optimization of this operator constitutes a separate problem.

In a thermodynamic sense, this problem has a well-known solution:[11,17,18] the optimal CG potential is simply the FP potential of mean force (PMF) projected along the coarse-grained degrees of freedom, $W_{\text{FP}}(\mathbf{R})$. In the canonical ensemble,

$$U_{\text{CG}}(\mathbf{R}) = W_{\text{FP}}(\mathbf{R}) = -k_B T \ln \int e^{-\beta U_{\text{FP}}(\mathbf{r})} \delta[\mathbf{M}(\mathbf{r}) - \mathbf{R}]\, d\mathbf{r}, \tag{1}$$

where as usual $\beta = 1/k_B T$, the inverse temperature scaled by the Boltzmann constant. While the integral proceeds over the entire FP configuration space, the Dirac delta function selects only those configurations compatible with a given CG one. Equation (1) gives the ideal interaction because it guarantees that the configuration-space probabilities $\wp_{\text{CG}}(\mathbf{R})$ generated by the CG model, under the action of its potential, are the same as the mapped probabilities $\wp_{\text{FP}}(\mathbf{R})$ of the original model, i.e.,

$$\wp_{\text{CG}}(\mathbf{R}) = \wp_{\text{FP}}(\mathbf{R}) = \frac{e^{-\beta W_{\text{FP}}(\mathbf{R})}}{\int e^{-\beta W_{\text{FP}}(\mathbf{R})} d\mathbf{R}} = \frac{\int e^{-\beta U_{\text{FP}}(\mathbf{r})} \delta[\mathbf{M}(\mathbf{r}) - \mathbf{R}]\, d\mathbf{r}}{\int e^{-\beta U_{\text{FP}}(\mathbf{r})} d\mathbf{r}}. \tag{2}$$

While this solution has been long appreciated, it has not been practically realizable due to the highly multidimensional nature of the PMF, which is not generally well-represented by simple decompositions. Thus, coarse-graining procedures have instead focused on finding a good

a)Author to whom correspondence should be addressed. Electronic mail: shell@engineering.ucsb.edu.

approximation to the PMF. It is typical to assume particular functionalities for U_{CG} (e.g., standard bond, angle, torsional, and nonbonded pair terms), and consequently, coarse-graining proceeds by optimizing any free parameters in this functional space. This basic idea underlies two widely used classes of coarse-graining methodologies in the literature, Boltzmann inversion approaches[1,2,4,16] and force matching.[6,7,11]

In this work, we discuss an alternative approach to this problem based on minimizing a configuration-space functional called the relative entropy.[12,19,20] Importantly, minimization of the relative entropy has a direct relationship with finding a CG potential that reproduces the multidimensional PMF.[12] We show here that this new approach provides a systematic framework for approximating W_{FP} and reducing coarse-graining errors, and we demonstrate connections with these existing coarse-graining algorithms.

Boltzmann inversion methods[1,2,4,16] have been a pioneering coarse-graining approach; rather than confronting the multibody PMF of Eq. (1), these methods typically focus on capturing the two-body one. A foundational principle of many of these approaches is Henderson's uniqueness theorem, which states that for a certain pairwise structure, there is a distinct corresponding pairwise potential.[21] Building on a scheme of McGreevy and Pusztai,[22] Lyubartsev and Laaksonen developed an inverse Monte Carlo method:[1] they used simulations to iteratively construct a finely discretized tabulation of a CG pair interaction so as to replicate the corresponding FP pair distribution. Closely related are the class of iterative Boltzmann inversion methods that perform analogous iterations to generate tabulated potentials.[2,4] Over the past decade, many groups have adopted variations of this methodology in studying a variety of soft matter systems (e.g., polymers[23,24] and membranes[15,25]). While these iterative approaches appear to be the most frequently employed numerical strategy for Boltzmann inversion, alternatives exist. For example, Head-Gordon and Stillinger invoked the Ornstein–Zernike integral equation for the experimental pairwise structure of water to estimate an effective isotropic interaction.[26] Transferability of such Boltzmann-inverted potentials has also received considerable interest. Louis found that the optimal model at one set of state conditions is not strictly optimal at another;[27] for example, isotropic water models were shown to have significant variability in state space.[19,28,29] In fact, the multidimensional PMF of Eq. (1) is clearly a function of temperature and volume, and the state dependence of CG potentials is intrinsic to any such approach. Ongoing efforts are addressing such transferability issues,[13] but we do not discuss them further here.

Coarse-graining by force matching, introduced by Izvekov and Voth,[6,7,11] has recently emerged as a viable and powerful alternative to Boltzmann inversion. The basic idea is to generate a CG potential that recapitulates the same mean forces on CG sites that would be predicted by the reference model, an idea stemming from the earlier work of Ercolessi and Adams.[30] In this approach, a functional measuring the squared error in CG predictions for the average FP forces is minimized variationally with respect to force field parameters, such as knot values in spline potentials and site partial charges. This strategy was shown to provide a rigorous way to reproduce the multidimensional PMF when U_{CG} is not limited to a specific functional space; on the other hand, for the usual approach in which a specific form of is specified, an elegant formalism projects the coarse-graining procedure onto a basis function space.[11] Recently, Mullinax and Noid created an extended ensemble version of force matching that suggests a systematic route to developing a single CG model able to describe several FP models simultaneously, increasing transferability.[13] Using the force matching formalism, they also developed a generalized Yvon–Born–Green theory in which CG interactions can be determined directly, without iteration, using FP configurational correlation functions.[14] Their work, in particular, highlights basic differences of force matching with inverse Boltzmann methods; in coarse-graining to a pair interaction, for example, their method requires both three- and two-body correlations, whereas Boltzmann inversion requires only the customary two-body one. On a practical level, Andrienko and co-workers have developed a computational toolkit which is able to evaluate the benefits and drawbacks of these different optimization methods.[31]

An alternative variational coarse-graining approach that we recently proposed is based upon minimizing a different objective function, called the relative entropy. A discrete-state version of this quantity is given by[12]

$$S_{\mathrm{rel}} = \sum_i \wp_{\mathrm{FP}}(i) \ln\left(\frac{\wp_{\mathrm{FP}}(i)}{\wp_{\mathrm{CG}}(M(i))}\right) + \langle S_{\mathrm{map}}\rangle_{\mathrm{FP}}. \tag{3}$$

Here, i is an index that proceeds over all configurational microstates of the FP system, $M(i)$ describes the operation of the mapping function on FP configuration i to generate a corresponding CG configuration $I = M(i)$, and the microstate probabilities $\wp$ follow from the action of the respective interaction potentials on the FP and CG systems. For example, as usual in the canonical ensemble, $\wp_{\mathrm{FP}}(i) = Z_{\mathrm{FP}}^{-1} e^{-\beta U_{\mathrm{FP}}(i)}$ and $\wp_{\mathrm{CG}}(I) = Z_{\mathrm{CG}}^{-1} e^{-\beta U_{\mathrm{CG}}(I)}$, where Z is the configurational partition function. The mapping entropy in Eq. (3), $\langle S_{\mathrm{map}}\rangle_{\mathrm{FP}}$, stems from the average degeneracy of FP configurations mapping to the same CG one,[12]

$$S_{\mathrm{map}}(I) = \ln \sum_i \delta_{I,M(i)}, \tag{4}$$

where δ gives the Kronecker delta function. It is particularly important to note that the mapping entropy has no dependence on the CG interaction potential, but rather is a unique function of the mapping operator and the FP configurational weights. Analogous expressions for the relative entropy exist for systems with continuous degrees of freedom,

$$S_{\mathrm{rel}} = \int \wp_{\mathrm{FP}}(\mathbf{r}) \ln\left(\frac{\wp_{\mathrm{FP}}(\mathbf{r})}{\wp_{\mathrm{CG}}(\mathbf{M}(\mathbf{r}))}\right) d\mathbf{r} + \langle S_{\mathrm{map}}\rangle_{\mathrm{FP}}, \tag{5}$$

$$S_{\mathrm{map}}(\mathbf{R}) = \ln \int \delta[\mathbf{M}(\mathbf{r}) - \mathbf{R}]\, d\mathbf{r}. \tag{6}$$

In the continuous case, the mapping entropy measures a hyperarea in the FP configuration space for states that map to

the same CG configuration. For example, consider the coarse-graining of an anisotropic (rigid) molecule onto an isotropic point; the degeneracy for each molecule can be readily determined by invoking Euler angles, and the corresponding mapping entropy evaluates at $\ln[8\pi^2]$.

In crude terms, the relative entropy measures the information lost when moving from the FP to the CG configurational distribution.[12] In probability theory, it is also known as the Kullback–Leibler divergence,[32] which has a wide role as an asymmetric distance metric between probability distributions. In the multiscale context here, S_{rel} is based on the likelihood that random sampling of the CG ensemble can correctly reconstruct the FP microstate distribution.[12] At a more physically intuitive level, we showed that S_{rel} is a measure of the fluctuations in the potential energy difference between the two systems across their configurational landscapes (described in greater detail below).[20] Furthermore, we found that S_{rel} is inherently tied to and may predict many different coarse-graining errors.[20] By the above equation, it is clear that S_{rel} is zero only if $\wp_{\text{CG}}(\mathbf{M}(\mathbf{r})) \propto \wp_{\text{FP}}(\mathbf{r})$, and in such a scenario, the relative entropy vanishes together with many coarse-graining errors. As the probability distributions deviate, the value of the relative entropy increases, as do the errors. As such, the relative entropy provides a convenient metric for the relevance of a given CG model to a reference FP one. We, therefore, proposed a general coarse-graining strategy in which optimal CG models are those that minimize this quantity.[12,19,20]

In this work, we show that the relative entropy approach bears much in common with other coarse-graining techniques, such as force matching and iterative Boltzmann inversion techniques, and under some conditions, recovers the same CG potentials as these. These findings extend previous work by other groups that have noted the link between the relative entropy and Boltzmann inversion. Silloco developed a procedure based on minimization of a pairwise-version of S_{rel} to extract pair potentials that reproduce target pair distributions;[33] a subsequent effort developed a Monte Carlo scheme which consistently refines the sought for pair potential.[34] Ming and Wall used the relative entropy to optimize elastic network models of protein conformational fluctuations.[35] Furthermore, Murtola *et al.* demonstrated connections between the relative entropy and Henderson's uniqueness theorem in the context of the inverse Monte Carlo method.[15] In addition, we previously parameterized a spherically symmetric water model using the relative entropy,[19] attaining analogous state behavior as observed by Head-Gordon and co-workers via the inverse Boltzmann approach.[28,29]

Despite these connections to other methods, the relative entropy approach admits features that may offer new coarse-graining routes. Foremost, this framework places few restrictions on the type of the CG model: it is equally applicable to systems with discrete or continuous degrees of freedom [Eqs. (3) and (5)], capable of addressing models with discontinuous or ill-defined forces. Most importantly, its close connections with coarse-graining errors[20] suggest a powerful two-part coarse-graining platform, as we discuss below: first, it provides a precise manner by which to reduce coarse-graining errors in arbitrary microscopic quantities, providing an informed strategy for selecting basis functions in the CG potential, and second, the value of the relative entropy itself provides a potentially potent metric for the overall quality of a given CG model. Importantly, this framework not only permits direct optimization of specific CG models, but can also inform the selection and design of the models in detail.

The goal of this paper is to develop a comprehensive and general coarse-graining framework based on the relative entropy, moving beyond our earlier simple tests.[12,19,20] In Sec. II, we derive a set of basic relations for S_{rel}-optimized models that, in particular, demonstrate systematic approaches to reducing coarse-graining errors. We discuss the relationship of these approaches to other coarse-graining methods in Sec. III. In Sec. IV, we describe a number of numerical strategies for the relative entropy coarse-graining and computational implementation of the ideas in Sec. II. Finally, in Sec. V, we illustrate several newly developed aspects of the relative entropy approach with an instructive test involving the lattice gas.

II. RELATIVE ENTROPY COARSE-GRAINING FRAMEWORK

Upon substitution of the appropriate configurational probabilities in Eqs. (3) or (5), we obtain the following canonical expression for the relative entropy for either the discrete or continuous cases:

$$S_{\text{rel}} = \beta \langle U_{\text{CG}} - U_{\text{FP}} \rangle_{\text{FP}} - \beta (A_{\text{CG}} - A_{\text{FP}}) + \langle S_{\text{map}} \rangle_{\text{FP}} \qquad (7)$$

with $A = -k_B T \ln Z$. Here, all averaging is performed in the FP ensemble, i.e., with weights according to the populations of the FP configurations. The mapping entropy remains rigorously independent of the CG interaction, and thus, can be treated as constant during S_{rel} minimization. This equation forms the practical starting point for our coarse-graining framework.

In what follows, we describe a procedure for extracting effective CG force fields based on minimizing Eq. (7) above. While much of this discussion centers on the values of the *parameters* at the relative entropy minimum, we stress that the actual *value* of S_{rel} may also provide a broader metric for "scoring" a given CG model. Evidence to this lies in the fact that the relative entropy consistently signals various coarse-graining errors, as we have shown previously.[20] Thus, an additional capability of this multiscale framework may be the CG model design. That is, when faced with a variety of CG models, the relative entropy may be a promising route to scoring their efficacy without comprehensive evaluation of properties.

A. Basic optimization principles using the relative entropy

To parameterize a CG model, optimization in the relative entropy framework corresponds to minimizing its value as a function of adjustable features in the CG model.[12] In the most basic coarse-graining scenario, a parameter λ in the CG potential is to be optimized; this amounts to finding the global minimum in S_{rel} as a function of λ. Provided that λ is

continuous, the following relation holds at optimality by Eq. (7):

$$0 = \frac{\partial S_{\rm rel}}{\partial \lambda} = \beta \left\langle \frac{\partial U_{\rm CG}}{\partial \lambda} \right\rangle_{\rm FP} - \beta \left\langle \frac{\partial U_{\rm CG}}{\partial \lambda} \right\rangle_{\rm CG}. \quad (8)$$

According to Eq. (8), optimization of the force field finds the point at which the derivative of the trial CG energy function averages to the same value in the CG and FP ensembles. While this condition is true for any extremum, a minimum can be verified by the positivity of the second derivative,

$$\frac{\partial^2 S_{\rm rel}}{\partial \lambda^2} = \beta \left\langle \frac{\partial^2 U_{\rm CG}}{\partial \lambda^2} \right\rangle_{\rm FP} - \beta \left\langle \frac{\partial^2 U_{\rm CG}}{\partial \lambda^2} \right\rangle_{\rm CG} + \beta^2 \left\langle \frac{\partial U_{\rm CG}}{\partial \lambda}^2 \right\rangle_{\rm CG} - \beta^2 \left\langle \frac{\partial U_{\rm CG}}{\partial \lambda} \right\rangle_{\rm CG}^2. \quad (9)$$

Though Eq. (9) is not rigorously positive for arbitrary λ, careful numerical minimization strategies can ensure convergence to a local minimum, and the global minimum might also be found by judicious choice of an initial value for λ. We describe several such approaches in Sec. IV.

This coarse-graining approach simplifies for parameters that are linear coefficients in the potential, $U_{\rm CG}(\mathbf{R}) = \ldots + \lambda f(\mathbf{R})$. Such parameters appear frequently in typical force field components, such as in Lennard-Jones coefficients, spring constants, spline knots, and values in tabulated potentials. They give rise to the following optimality condition:

$$0 = \frac{\partial S_{\rm rel}}{\partial \lambda} = \beta \langle f \rangle_{\rm FP} - \beta \langle f \rangle_{\rm CG}. \quad (10)$$

In this case, minimization guarantees that the average value of f is the same in both ensembles. Furthermore, the second derivative simplifies to

$$\frac{\partial^2 S_{\rm rel}}{\partial \lambda^2} = \beta^2 \left\langle f^2 \right\rangle_{\rm CG} - \beta^2 \langle f \rangle_{\rm CG}^2. \quad (11)$$

The implied curvature is simply proportional to a variance in the CG ensemble, which is never negative, suggesting that the relative entropy possesses a single global minimum for such linear parameters.

Alternatively, one might optimize not a parameter but an entire function in the CG force field. Practically, such a function may be represented by a finely discretized grid, a spline, or a combination of basis functions; here, we consider the formal case that serves as the limit of these approximations for an infinite number of points or basis functions. Let us specify the function for optimization as $f(y)$, where y is an argument that depends on the CG coordinate set, i.e., $y = y(\mathbf{R})$. For example, if f corresponds to a pairwise potential then y gives the distance between two sites. Regardless, functional minimization of the relative entropy with respect to f gives

$$0 = \frac{\delta S_{\rm rel}}{\delta f(y)} = \beta \left\langle \frac{\partial U_{\rm CG}}{\partial f} \delta \left[y - y(\mathbf{M}(\mathbf{r})) \right] \right\rangle_{\rm FP} - \beta \left\langle \frac{\partial U_{\rm CG}}{\partial f} \delta \left[y - y(\mathbf{R}) \right] \right\rangle_{\rm CG}, \quad (12)$$

where we have left a possible explicit dependence of the force field on f. If the potential is linear in f, evaluation of the delta functions leads to

$$\wp_{\rm CG}(y) = \wp_{\rm FP}(y), \quad (13)$$

where $\wp(y)$ is the probability distribution for y. Thus, functional minimization obliges that the probability distribution for the argument of f is the same for both ensembles.

B. Ramifications of optimization on coarse-graining errors

These optimality conditions for CG potentials parameterized by relative entropy minimization lead to a number of simple relations for minimizing coarse-graining errors. Consider a configurational observable X. Suppose that it is known that X plays an important role in the FP phenomena, and it is thus desirable to correctly describe fluctuations or correlations in X upon coarse-graining. Examples include simple metrics such as distances or angles among particular sites, or more complex ones such as the radius of gyration or solvent-accessible surface area. We first require that the observable is measurable in both systems: as $X(\mathbf{M}(\mathbf{r}))$ in the original FP system and as $X(\mathbf{R})$ in the CG system. This, of course, also affects the choice of the mapping function and hence, the type of the CG model.

As a basic step, one might design a CG potential to correctly reproduce the average, $\langle X \rangle_{\rm CG} = \langle X \rangle_{\rm FP}$. In the relative entropy formalism, a straightforward approach is to include a term in $U_{\rm CG}$ that is proportional to X,

$$U_{\rm CG}(\mathbf{R}) = \ldots + aX(\mathbf{R}), \quad (14)$$

where a is a parameter to be tuned. By the optimality condition in Eq. (10), relative entropy minimization chooses a, such that

$$\langle X \rangle_{\rm CG} = \langle X \rangle_{\rm FP}, \quad (15)$$

which exactly accomplishes the task at hand. Here, the parameter a appears implicitly in the configurational weights in the CG average on the left hand side. Beyond a simple agreement of averages, a similar approach can be used to capture the variance in X using a two-term addition to the potential,

$$U_{\rm CG}(\mathbf{R}) = \ldots + aX(\mathbf{R}) + bX(\mathbf{R})^2. \quad (16)$$

The two optimality conditions for choosing a and b then lead directly to

$$\langle X \rangle_{\rm CG} = \langle X \rangle_{\rm FP}, \ \langle X^2 \rangle_{\rm CG} = \langle X^2 \rangle_{\rm FP}. \quad (17)$$

Together these give a condition for the observable's variance, $\mathrm{var}_{\rm CG}(X) = \mathrm{var}_{\rm FP}(X)$. Thus, simultaneous relative entropy optimization for a and b explicitly demands equality of the average and variance of X between the two ensembles. Notice that the value of a is most likely different between the CG models of Eqs. (14) and (16), as the two models invoke different optimization equations.

By extension, it is easy to see that the addition of higher-order terms in X with corresponding $S_{\rm rel}$ minimization leads to CG models that increasingly replicate higher moments of the FP distribution of X (only dependent on the choice of how many terms to include). As the number of terms approaches

infinity, a complete set of basis functions are formed that implies a Taylor expansion for some function of X. Indeed, one may proceed directly to include such a (yet unknown) function directly in the force field,

$$U_{\mathrm{CG}}(\mathbf{R}) = \ldots + f(X(\mathbf{R})). \tag{18}$$

By the expression in Eq. (13), functional minimization with respect to f gives

$$\wp_{\mathrm{CG}}(X) = \wp_{\mathrm{FP}}(X). \tag{19}$$

This type of relative entropy minimization thus enables one to match the entire fluctuation spectrum of X, which approximates the case when f is suitably flexible, for example, composed of splines or finely discretized tabulations.

These ideas readily extend to cases when more than one observable is of interest. Consider now two configurational observables X_1 and X_2. One might add terms to the CG interaction potential in the following manner:

$$U_{\mathrm{CG}}(\mathbf{R}) = \ldots + a_1 X_1(\mathbf{R}) + a_2 X_2(\mathbf{R}) + b_{1,2} X_1(\mathbf{R}) X_2(\mathbf{R}). \tag{20}$$

Upon relative entropy minimization with respect to the force field parameters a_1, a_2, and $b_{1,2}$, the following equalities will then hold:

$$\langle X_1\rangle_{\mathrm{CG}} = \langle X_1\rangle_{\mathrm{FP}},\ \langle X_2\rangle_{\mathrm{CG}} = \langle X_2\rangle_{\mathrm{FP}},\ \langle X_1 X_2\rangle_{\mathrm{CG}} = \langle X_1 X_2\rangle_{\mathrm{FP}}. \tag{21}$$

These are easily combined to show that the covariance between the two observables is the same in either ensemble, $\mathrm{cov}_{\mathrm{CG}}(X_1, X_2) = \mathrm{cov}_{\mathrm{FP}}(X_1, X_2)$. By extension, the complete range of coupled fluctuations of X_1 and X_2 can be captured by the CG model if a two-dimensional function of them is included in the force field,

$$U_{\mathrm{CG}}(\mathbf{R}) = \ldots + f(X_1(\mathbf{R}), X_2(\mathbf{R})). \tag{22}$$

By Eq. (13) for the two-vector (X_1, X_2), the functional minimum of the relative entropy gives

$$\wp_{\mathrm{CG}}(X_1(\mathbf{R}), X_2(\mathbf{R})) = \wp_{\mathrm{FP}}(X_1(\mathbf{R}), X_2(\mathbf{R})). \tag{23}$$

This result shows that the complete *joint* distribution of fluctuations of the two observables is reproduced by this kind of CG model. Note that this expression is distinct from the one in which separate functions for each observable are added to the potential, in which case the individual distributions will be equal, $\wp_{\mathrm{CG}}(X_1) = \wp_{\mathrm{FP}}(X_1)$ and $\wp_{\mathrm{CG}}(X_2) = \wp_{\mathrm{FP}}(X_2)$, but the joint one may not be equal.

These findings have several implications for relative entropy optimization. Most importantly, they demonstrate a systematic strategy for improving CG models by reducing coarse-graining errors in arbitrary configurational properties. One can control the degree of replication of a particular property (e.g., variance versus complete distribution) through the addition of appropriate terms in the CG force field. A second implication is the effect of S_{rel} minimization on typical force field components in the CG model. By Eq. (15), optimized harmonic bond and angle terms will reproduce the average and variance of the corresponding lengths and angles as the FP model. Lennard-Jones interactions will reproduce specific pair distance correlations ($\langle r_{ij}^{-12}\rangle$ and $\langle r_{ij}^{-6}\rangle$) among the sites to which they apply. Moreover, by Eq. (19), optimized splines or finely discretized tabulations applied to pairwise and torsional interactions will produce CG models that replicate the entire respective distance and orientation distributions. To summarize, consider the argument of a force field component (e.g., distance and orientation); while the full distribution of the argument is likely to be captured by flexible potentials, other potentials may only reproduce particular moments.

Further consideration suggests that the kind of correlations correctly modeled by the CG system is directly tied to the type of the potential used. This feature can be particularly illustrated for n-body interactions: a spatially dependent mean-field interaction can generate correct spatial distributions, two-body terms can generate correct pair correlations, three-body terms for triplet correlations, and so on. That is not to say that any S_{rel}-optimized two-body CG potential will not correctly capture higher correlations; while this may be possible, the relative entropy formalism only links the exact reproduction of n-body statistics with an n-body potential. It is important to emphasize that these kinds of relationships fall naturally out of S_{rel} minimization, but they are not explicitly included in the original formulation of the relative entropy, which makes no reference to specific n-body interactions. This contrasts with several inverse Boltzmann coarse-graining methodologies that explicitly seek to reproduce particular correlation functions.

To formally show the correspondence between different interaction modes and configurational correlations, we write the CG energy function in a generalized form with different component functions:

$$U_{\mathrm{CG}}(\mathbf{R}) = \sum_k \sum_{\vec{\imath}\in\mathbf{N}^k} u_k\left(\mathbf{X}_k\left(\mathbf{R}_{\vec{\imath}}\right)\right). \tag{24}$$

Here, the summation over k proceeds over all distinct interactions u_k in the force field. This index tabulates interactions of different order (i.e., the number of sites involved, n_k) as well as those which apply to different sets of sites (e.g., bonded interactions occur only between neighboring sites). The second summation involves a vector index $\vec{\imath}$ of length n_k, and it proceeds over the set $\mathbf{N}^k$, which contains all unique combinations of n_k sites to which potential k applies. For example, for the pair-interacting Lennard-Jones system, the second sum is simply the pairwise loop, with $\mathbf{N}^k$ being the set of sites $(i, j < i)$. In any case, each instance of the potential u_k for a subset of sites depends on a particular subset of $\mathbf{R}$, labeled $\mathbf{R}_{\vec{\imath}}$. Finally, the argument to the potential itself may not be all of these coordinates, but some reduced set of geometric descriptors $\mathbf{X}_k$, a vector with length m_k. Conventional examples for $\mathbf{X}_k$ include the separation distance for a pair potential or the dihedral rotation for a torsional potential. While for both of these, $m_k = 1$, it can be greater in general, as in the three pair distances required, in principle, for a three-body potential.

Relative entropy optimization in the space of functions defined by the expression above corresponds to its functional

minimization with respect to each u_k,

$$0 = \frac{\delta S_{\text{rel}}}{\delta u_k(\mathbf{X}_k)} = \beta \left\langle \sum_{\vec{\imath} \in \mathbf{N}^k} \delta\left[\mathbf{X}_k - \mathbf{X}_k\left(\mathbf{M}_{\vec{\imath}}(\mathbf{r})\right)\right] \right\rangle_{\text{FP}} - \beta \left\langle \sum_{\vec{\imath} \in \mathbf{N}^k} \delta\left[\mathbf{X}_k - \mathbf{X}_k\left(\mathbf{R}_{\vec{\imath}}\right)\right] \right\rangle_{\text{CG}}. \quad (25)$$

Normalizing this expression by the number of terms in the sums and rearranging, we obtain

$$\wp_{\text{CG}}(\mathbf{X}_k) = \wp_{\text{FP}}(\mathbf{X}_k). \quad (26)$$

The interpretation of this expression is that relative entropy minimization achieves a CG model that reproduces the joint distribution of all of the arguments to the component interaction u_k, i.e., the m_k-dimensional multivariate distribution of this potential's arguments. This result only applies in the functional limit, i.e., when a perfectly flexible representation of the interaction u_k is used (e.g., an m_k-dimensional spline). As $m_k = 1$ for many conventional interactions, simple spline functionalities should provide numerically robust ways to reproduce such probability distributions upon S_{rel} optimization.

III. CONNECTIONS WITH OTHER COARSE-GRAINING METHODOLOGIES

We now proceed to show similarities, where possible, between the relative entropy and other coarse-graining approaches. None of these approaches are generally equivalent with S_{rel} coarse-graining, but we demonstrate that under particular conditions, approximations, or types of CG models, the relative entropy framework will lead to identical optimized models. The starting point for this discussion is the most general case in which the CG potential is unrestricted, i.e., it can take on any function in the space of N-body interactions, where N is the number of CG degrees of freedom. Note that this function space contains not only every pairwise decomposition of the energy, but also *all* possible multibody interactions in the system. We find the optimal potential by functional minimization of the relative entropy as outlined in Eq. (12),

$$0 = \frac{\delta S_{\text{rel}}}{\delta U_{\text{CG}}(\mathbf{R}')} = \beta \left\langle \delta\left[\mathbf{R}' - \mathbf{M}(\mathbf{r})\right] \right\rangle_{\text{FP}} - \beta \left\langle \delta\left[\mathbf{R}' - \mathbf{R}\right] \right\rangle_{\text{CG}}. \quad (27)$$

Here, the terms inside the averages indicate a product of Dirac delta functions that proceeds over all components of $\mathbf{R}'$. These averages can immediately be evaluated,

$$0 = Z_{\text{FP}}^{-1} e^{-\beta W_{\text{FP}}(\mathbf{R}')} - Z_{\text{CG}}^{-1} e^{-\beta U_{\text{CG}}(\mathbf{R}')}. \quad (28)$$

Here, Eq. (1) is used to insert the FP multidimensional PMF in place of the FP average. Rearrangement and switching $\mathbf{R}'$ for $\mathbf{R}$ finally gives

$$U_{\text{CG}}(\mathbf{R}) = W_{\text{FP}}(\mathbf{R}) + \text{cnst}. \quad (29)$$

The additive constant simply serves to shift the CG force field globally in energy, with no effect on configurational probabilities, and it can be taken to be zero without loss of generality. Thus, the so-called "unrestricted" relative entropy optimization, in which the CG potential can have any functional form, rigorously returns the thermodynamically optimal potential identified in the Introduction. In general, of course, numerical limitations prevent the determination of such multibody functions; instead, pair or perhaps three-body decompositions are typically used. Still, this analytical result shows the limiting tendency of S_{rel} minimization to equate the CG potential with the corresponding FP free energy surface.

A. Force matching

The result above permits a formal connection with the force matching technique.[6,7,11] In brief, the configurational derivative of Eq. (29) gives a relationship between the instantaneous CG and average FP interatomic forces. Consider a particular scalar CG coordinate $R = \mathbf{R}_I$ that is the Ith entry in $\mathbf{R}$ (e.g., the x position of a particular site). Then, $\nabla_R U_{\text{CG}}(\mathbf{R}) = \nabla_R W_{\text{FP}}(\mathbf{R})$,or using the definition of W_{FP},

$$F_{\text{CG}}(\mathbf{R}) = -k_B T e^{\beta W_{\text{FP}}(\mathbf{R})} \int e^{-\beta U_{\text{FP}}(\mathbf{r})} \nabla_R \delta\left[\mathbf{M}(\mathbf{r}) - \mathbf{R}\right] d\mathbf{r}, \quad (30)$$

where $F_{\text{CG}} = \mathbf{F}_{I,\text{CG}}$ denotes the force acting on coordinate R in the CG system. If the CG coordinates are defined by nonoverlapping combinations of the center of masses of groups of FP sites, as is the usual case, the integral can be evaluated as[11] [see also supplementary information (SI)][36]

$$F_{\text{CG}}(\mathbf{R}) = \int F(\mathbf{r})\, e^{-\beta U_{\text{FP}}(\mathbf{r})} \delta\left[\mathbf{M}(\mathbf{r}) - \mathbf{R}\right] d\mathbf{r} = \langle F \rangle_{\text{FP}}(\mathbf{R}), \quad (31)$$

where $F(\mathbf{r})$ is the total force due to all FP coordinates that define R. By dependence on the CG coordinates $\mathbf{R}$, the integral gives a conditional average of the net force F over all FP configurations mapping to the same CG one. Equation (31) easily generalizes to the complete force vector $\mathbf{F}$ operating on all CG coordinates. This basic relationship was derived rigorously by Noid *et al.*[11] and provides the theoretical basis for the force matching approach; an alternative derivation using a delta function identity derived by Darve[37] is given in the SI.[36] Force matching proceeds by minimizing force residuals between a trial CG force field and average forces from a reference FP ensemble of configurations. The implication of this result is that, in principle, unrestricted optimization by force matching should equate U_{CG} with the reference free energy surface W_{FP}, up to an unimportant additive constant. In this limit, force matching and S_{rel} minimization techniques, therefore, yield identical CG potentials when applied to a given system. However, in practical settings where potentials are constrained to be pair additive and cannot directly capture high-order correlations, the force matching and relative entropy approaches are not likely to return identical results. In particular, the Yvon–Born–Green formulation of Mullinax and Noid shows that force matching requires higher-order correlation functions to parameterize an n-body potential,[14] whereas S_{rel} minimization requires only the n-body ones per the considerations in Sec. II.

B. Iterative Boltzmann inversion

A connection also exists between S_{rel} optimization and the inverse Monte Carlo method or the iterative Boltzmann inversion approach.[1,2,4,16] Murtola *et al.* first noted this close correspondence;[15] here we elaborate in the context of a more specific, practical implementation. We consider a typical case in which one constructs a pair potential $u(r)$ that recapitulates a measured or computed reference radial-distribution function, both represented on a finely discretized grid in the space of pairwise distances. Let this grid have bin spacings denoted by Δr, where a specific bin i contains distances in the range $i\Delta r \leq r < (i+1)\Delta r$. For the sake of illustration, we consider the distance distribution corresponding to two specific CG sites. We formally relate this discretized two-site distance distribution function to the CG system by $\wp_{i,\text{CG}} = \langle \theta_i(\mathbf{R})\rangle_{\text{CG}}$, where $\theta_i(\mathbf{R})$ is an indicator function that takes a value of 1 when the specific pair distance lies within bin i and zero when outside of it. Thus, $\wp_{i,\text{CG}}$ gives probabilities on the discrete distance space indexed by i, which is formally related to a discretized and suitably normalized radial-distribution function. An analogous expression can be written for the reference FP distribution to which we will match, $\wp_{i,\text{FP}} = \langle \theta_i(\mathbf{M}(\mathbf{r}))\rangle_{\text{FP}}$.

We construct a tabulated CG pair potential with yet unknown energy values u_i,

$$U_{\text{CG}}(\mathbf{R}) = \sum_i u_i \theta_i(\mathbf{R}). \tag{32}$$

Here, we have included only terms in the potential related to the specific pair interaction, as additional terms have no effect on the results. In this Boltzmann inversion method, the potential is refined iteratively using simulations of the trial CG potential to compute corresponding distance distributions. Such iterations typically take the following form:

$$u_i^{k+1} = u_i^k + k_B T \ln\left(\frac{\wp_{i,\text{CG}}^k}{\wp_{i,\text{FP}}}\right), \quad \text{for all } i. \tag{33}$$

Here, k is an iteration index and $\wp_{i,\text{CG}}^k$ gives the distance distribution evaluated from a simulation using CG potential k. The reference quantity $\wp_{i,\text{FP}}$ is the matching distribution, evaluated once prior to the iterations. Ultimately, this kind of iteration converges to a unique potential when $\wp_{i,\text{CG}}^k = \wp_{i,\text{FP}}$ as $k \to \infty$.

The relative entropy approach instead finds an optimal set of values u_i by variational minimization. Note that the CG potential is linear in the u_i parameters, which according to Eq. (11) implies the existence of a single relative entropy minimum. By Eq. (10), this minimum then satisfies $\langle \theta_i\rangle_{\text{FP}} = \langle \theta_i\rangle_{\text{CG}}$ for all i, and thus, $\wp_{i,\text{FP}} = \wp_{i,\text{CG}}$ for all i. This implies that the discretized distance distribution generated under the action of the S_{rel}-optimized CG potential reproduces exactly its counterpart in the reference FP ensemble. By the uniqueness principle, relative entropy minimization then locates the same tabulated pair potential that would be found with iterative Boltzmann inversion, although the algorithm used to locate the optimum may be different. These ideas also directly reflect the concepts put forth in Sec. II that describe the connection between S_{rel}-optimized pair potentials and pairwise correlations. One important difference, however, is that relative entropy minimization also provides a quantitative, whole-ensemble scoring metric for the quality of the CG model.

C. Energy matching

Another coarse-graining strategy matches the potential energy landscape of a reference ensemble to a corresponding CG one. While this approach is not frequently used in the development of pseudoatom models, it is reminiscent of techniques for parameterizing all-atom classical potentials from quantum calculations. Let us introduce the following functional to measure the discrepancy in the two potentials:

$$\tilde{S}_{\text{rel}} = \frac{\beta^2}{2}\int \left[U_{\text{FP}}(\mathbf{r}) - U_{\text{CG}}(\mathbf{M}(\mathbf{r})) - C\right]^2 \wp_{\text{FP}}(\mathbf{r})\, d\mathbf{r}. \tag{34}$$

This quantity gives the average squared error in the energy prediction using FP ensemble weights, as the corresponding configurational probabilities underlie the model of interest to match. Since configurational probabilities do not depend on absolute energies, we also shift the energy difference by some constant C chosen to minimize the squared error; it evaluates to $C = \langle U_{\text{FP}} - U_{\text{CG}}\rangle_{\text{FP}}$. The functional is nondimensionalized by the thermal energy; the prefactor of $1/2$ does not affect the minimization and is chosen for convenience. Consequently,

$$\begin{aligned}\tilde{S}_{\text{rel}} &= \frac{\beta^2}{2}\int \left[U_{\text{FP}}(\mathbf{r}) - U_{\text{CG}}(\mathbf{M}(\mathbf{r})) - \langle U_{\text{FP}} - U_{\text{CG}}\rangle_{\text{FP}}\right]^2 \wp_{\text{FP}}(\mathbf{r})\, d\mathbf{r} \\ &= \frac{\beta^2}{2}\text{var}_{\text{FP}}(U_{\text{FP}} - U_{\text{CG}}).\end{aligned} \tag{35}$$

Energy matching proceeds by minimizing this functional with respect to any adjustable features of the CG model, such as a parameter λ in U_{CG}. In this case, the optimal λ satisfies the condition,

$$\langle U_{\text{FP}} - U_{\text{CG}}\rangle_{\text{FP}} \left\langle \frac{\partial U_{\text{CG}}}{\partial \lambda}\right\rangle_{\text{FP}} = \left\langle (U_{\text{FP}} - U_{\text{CG}})\frac{\partial U_{\text{CG}}}{\partial \lambda}\right\rangle_{\text{FP}}. \tag{36}$$

Apparently, coarse-graining via energy matching for a particular parameter in the CG potential guarantees that the corresponding derivative becomes uncorrelated with the energy difference between the two systems. Note that the correlation vanishes only in the FP ensemble.

It is no coincidence that this functional is notated similarly to the relative entropy. This quantity $\tilde{S}_{\text{rel}}$ becomes equal to S_{rel} when the latter is expanded to a second order perturbation (part of the so-called "Gaussian assumption" that we later describe in Sec. IV). At the moment, it suffices to state that the assumption holds when $U_{\text{FP}}(\mathbf{r}) \approx U_{\text{CG}}(\mathbf{M}(\mathbf{r}))$ and when $\beta \to 0$. This implies that optimization using S_{rel} approach naturally performs a kind of an energy matching procedure at elevated thermal energies, as well as for cases in which the CG force field is a near replica of the FP one. This of course is not the case in general, likely violated at moderate temperatures and when CG potentials are unable to capture the complexity of FP energy landscapes. Still, Eq. (36) may be useful in quickly locating estimates for CG parameters: all averages are in the FP ensemble, and thus, evaluation of the CG

configurational distribution is not required. For full S_{rel} minimization, this can provide an initial guess that accelerates and stabilizes the numerical minimization procedure.

D. Classical variational (mean-field) theory

Common practice in the analytical evaluation of statistical-mechanical models is to minimize the venerable Gibbs–Bogoliubov–Feynman bound to obtain a mean-field closure.[38] We previously demonstrated a connection of S_{rel} to this classic approach;[12] here we elaborate on this relationship. We restrict our discussion to cases in which the mapping entropy is zero (the mapping function is the identity operator). This choice is motivated by the fact that often the interactions rather than the degrees of freedom are coarse-grained in these classical scenarios, in the sense that multibody terms are replaced with single-body ones. Nevertheless, the formalism outlined below does not restrict the CG force field (it is not necessarily a mean field).

We introduce a functional that we will show to underly the classical variational approach,

$$S_{\text{var}} = \int \wp_{\text{CG}}(\mathbf{r}) \ln\left(\frac{\wp_{\text{CG}}(\mathbf{r})}{\wp_{\text{FP}}(\mathbf{r})}\right) d\mathbf{r}. \tag{37}$$

This quantity bears stark resemblance to the relative entropy, except that the roles of the FP and CG systems have been switched; it is, essentially, the relative entropy for the reverse multiscale procedure, involving the "discovery" of the FP system most appropriate to a CG model. In the canonical ensemble,

$$S_{\text{var}} = \beta \langle U_{\text{FP}} - U_{\text{CG}} \rangle_{\text{CG}} - \beta (A_{\text{FP}} - A_{\text{CG}}). \tag{38}$$

By the Gibbs–Bogoliubov–Feynman bound, this is always non-negative and the classical variational approach corresponds to minimizing this functional, in effect. Any arbitrary parameter λ in U_{CG} can be evaluated at the minimum, which yields the following optimality (or closure) condition:

$$\langle U_{\text{FP}} - U_{\text{CG}} \rangle_{\text{CG}} \left\langle \frac{\partial U_{\text{CG}}}{\partial \lambda} \right\rangle_{\text{CG}} = \left\langle (U_{\text{FP}} - U_{\text{CG}}) \frac{\partial U_{\text{CG}}}{\partial \lambda} \right\rangle_{\text{CG}}. \tag{39}$$

The striking similarity of this expression with the optimality condition for energy matching is particularly interesting, with the sole difference being the ensemble in which the covariance vanishes. As the averaging here is performed in the CG ensemble, this method is potentially more convenient than energy matching because it does not require sampling of the more expensive FP ensemble, traditionally allowing for analytical closures. In this sense, Eq. (39) also provides a means to estimate the minimum in S_{rel}, but without the evaluation of the FP state probabilities.

Comparison of Eq. (38) with the relative entropy reveals the following relationship:

$$S_{\text{var}} = -S_{\text{rel}} - \beta \langle U_{\text{FP}} - U_{\text{CG}} \rangle_{\text{FP}} + \beta \langle U_{\text{FP}} - U_{\text{CG}} \rangle_{\text{CG}}. \tag{40}$$

Therefore, minimization of S_{var} is distinct from that of S_{rel}. However, this expression can be simplified using the same perturbation expansion for the relative entropy mentioned in the discussion on energy matching (the "Gaussian assumption," also described in detail below). Such an approximation gives [see SI (Ref. 36)]

$$\beta \langle U_{\text{FP}} - U_{\text{CG}} \rangle_{\text{FP}} - \beta \langle U_{\text{FP}} - U_{\text{CG}} \rangle_{\text{CG}} \approx -2S_{\text{rel}}. \tag{41}$$

Upon substitution into Eq. (40), one has that

$$S_{\text{var}} \approx S_{\text{rel}}. \tag{42}$$

Thus, to the extent that Eq. (41) holds, the variational approach and relative entropy minimization become identical coarse-graining strategies. The two methods are equivalent so long as the Gaussian assumption holds, i.e., at high temperature or when the CG energy landscape is able to closely replicate the FP one. As energy matching also approaches relative entropy optimization under these conditions, minimization of S_{rel} suggests that in *both* ensembles one finds a vanishing correlation between $\partial U_{\text{CG}}/\partial\lambda$ and $(U_{\text{FP}} - U_{\text{CG}})$ and an equivalence of the average values of $\partial U_{\text{CG}}/\partial\lambda$.

IV. NUMERICAL STRATEGIES FOR RELATIVE ENTROPY OPTIMIZATION

Here, we discuss practical optimization of CG force fields by minimization of the relative entropy. In addition, as the value of the relative entropy itself signals the behavior of coarse-graining errors and serves as a model quality metric,[20] we also discuss numerical aspects involved in computing it.

A. Calculating the relative entropy

The relative entropy can be computed by the canonical expression of Eq. (7), involving energetic differences between the CG and FP systems and the respective mapping entropy. The mapping entropy might be found analytically (particularly if it is constant in phase space), and the potential energy averages can be readily determined in molecular simulation. Conversely, the free energies are more numerically challenging, but may by computed using established advanced algorithms.[39]

An alternative approach exists, however, for computing S_{rel}. One can reformulate the free energy difference in Eq. (7) via standard free energy perturbation,[20]

$$S_{\text{rel}} = \ln\langle e^{\Delta - \langle\Delta\rangle_{\text{FP}}} \rangle_{\text{FP}}, \tag{43}$$

where $\Delta(\mathbf{r}) = \beta[U_{\text{FP}}(\mathbf{r}) - U_{\text{CG}}(\mathbf{M}(\mathbf{r}))]$. The function Δ is simply a temperature-normalized dimensionless measure of the potential energy difference between a given FP configuration and a corresponding CG configuration. The effect of the $\langle\Delta\rangle_{\text{FP}}$ term in the exponential serves to remove arbitrary global shifts in the corresponding potentials that do not affect configurational probabilities. This expression gives a physical interpretation for the relative entropy: it measures differences in the potential energy landscapes of the FP and CG ensembles.[20] For perturbation purposes, Δ at most can have singularities for a measure-zero region of phase space, and S_{map} must be a configurational constant; we assume as such in what follows.

On first sight, Eq. (43) seems particularly attractive for the calculation of S_{rel} as it lacks free energies, requires no special simulation techniques beyond averaging, and eliminates

the need to construct the CG ensemble. However, the average over the exponential leads to an increase in statistical errors that are typical of free energy perturbation calculations[39] and related expressions such as the Jarzynski equality:[40] large contributions to the average stem from infrequently sampled states with large values of Δ. For n uncorrelated measurements of Δ, we estimate the uncertainty in S_{rel} via Eq. (43) to be the following [see SI (Ref. 36)],

$$\varepsilon_{S_{\text{rel}}} \approx \frac{1}{\sqrt{n}} e^{S_{\text{rel}}}. \tag{44}$$

To leading order, the statistical error in the relative entropy increases exponentially with its value. One implication of this result is that poorly constructed CG models with large S_{rel} values will pose difficult numerical challenges for Eq. (43). In such cases, biased sampling techniques may permit the use of Eq. (43) by explicitly probing a broader distribution over Δ. Nevertheless, in the simulation case study below (Sec. V), we find that this error is a gross overestimation of the actual statistical error, suggesting that Eq. (43) is perhaps more promising than expected.

While Eq. (43) presents a method for computing the relative entropy exactly, it also provides a starting point for systematic approximations. Let us suppose that the distribution of Δ in the FP ensemble is Gaussian, with standard deviation $\sigma = \sqrt{\text{var}_{\text{FP}}(\Delta)}$ and mean $\Delta_0 = \langle\Delta\rangle_{\text{FP}}$. Then, this Gaussian assumption gives [see Ref. 20 and SI (Ref. 36)]

$$S_{\text{rel}} \approx \tilde{S}_{\text{rel}} = \tfrac{1}{2}\text{var}_{\text{FP}}(\Delta). \tag{45}$$

In other words, the relative entropy in this case is proportional to the variance of fluctuations in the dimensionless energy difference in the FP ensemble. This result is particularly significant because the approximate quantity $\tilde{S}_{\text{rel}}$ takes a form identical to the energy-based variational objective function discussed in Sec. III. From now on, we refer to $\tilde{S}_{\text{rel}}$ as the Gaussian approximate for the relative entropy.

The Gaussian assumption provides a simpler, albeit approximate route to computing the relative entropy, without averages over exponentials. With n uncorrelated measurements of Δ, we estimate the uncertainty in computing $\tilde{S}_{\text{rel}}$ via Eq. (45) to be [see SI (Ref. 36)]

$$\varepsilon_{\tilde{S}_{\text{rel}}} \approx \sqrt{\frac{2}{n}}\tilde{S}_{\text{rel}}. \tag{46}$$

The statistical error from the Gaussian approximate, therefore, increases linearly with its value, much better behaved than the exact relative entropy. Nevertheless, $\tilde{S}_{\text{rel}}$ has an inherent systematic error in estimating the true value of the relative entropy. We only expect $S_{\text{rel}} = \tilde{S}_{\text{rel}}$ if the Gaussian distribution holds, particularly in its tails.

The validity of the Gaussian assumption can be addressed through a cumulant expansion of Eq. (43),

$$\begin{aligned} S_{\text{rel}} &= \ln\left(1 + \langle\Delta - \langle\Delta\rangle_{\text{FP}}\rangle_{\text{FP}} + \tfrac{1}{2}\left\langle(\Delta - \langle\Delta\rangle_{\text{FP}})^2\right\rangle_{\text{FP}} + \ldots\right) \\ &= \tfrac{1}{2}\left(\left\langle\Delta^2\right\rangle_{\text{FP}} - \langle\Delta\rangle_{\text{FP}}^2\right) + \ldots, \end{aligned} \tag{47}$$

which shows that $S_{\text{rel}} \approx \tilde{S}_{\text{rel}}$ (i.e., the approximation is valid) for small relative variations in Δ and in turn, small values of S_{rel}. This occurs for two cases: at high temperatures as $\beta \to 0$ and for $U_{\text{FP}}(\mathbf{r}) \approx U_{\text{CG}}(\mathbf{M}(\mathbf{r}))$ when CG models are close-replications of FP models. Consider the former limit as the temperature approaches infinity,

$$\lim_{\beta\to 0} S_{\text{rel}} = \frac{\beta^2}{2}\text{var}_{\mathbf{r}}(U_{\text{FP}} - U_{\text{CG}}), \tag{48}$$

where $\text{var}_{\mathbf{r}}$ indicates an unweighted variance over the entire FP configurational space. While it appears that the relative entropy scales universally with β^2 at extremely high temperatures (irrespective of other state conditions),[20] the prefactor of the power law is distinct for each case and is identical with the corresponding β-curvature of the relative entropy. Thus, computation of $\text{var}_{\mathbf{r}}(U_{\text{FP}} - U_{\text{CG}})$ might serve as a crude estimate of the first-order behavior of S_{rel} in state space.

The Gaussian assumption also permits another approach to estimating the relative entropy on the basis of different ensemble averages of Δ, as we have shown in previous work[20] [and in the SI (Ref. 36)],

$$S_{\text{rel}} \approx -\tfrac{1}{2}\left(\langle\Delta\rangle_{\text{FP}} - \langle\Delta\rangle_{\text{CG}}\right). \tag{49}$$

While not equivalent to $\tilde{S}_{\text{rel}}$ in Eq. (45), this expression is particularly convenient for interactions that are coarse-grained into a mean field. Namely, one can show that in such cases the relative entropy estimate depends solely on correlations between the degrees of freedom in the FP ensemble, independent of the particular CG force field [see SI (Ref. 36)].

B. Minimization of the relative entropy

Here, we suggest numerical coarse-graining algorithms for minimizing S_{rel}. Previously, we proposed to use independent Newton–Raphson minimization equations for each free parameter.[19] Here, we propose a more robust strategy suitable to multiparameter models using a single, coupled Newton–Raphson equation. Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots)$ denote the parameters. Then

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \chi\mathbf{H}^{-1}\cdot\nabla_\lambda S_{\text{rel}}, \tag{50}$$

where the first derivatives of S_{rel} are given by

$$\nabla_\lambda S_{\text{rel}} = \beta\left\langle\frac{\partial U_{\text{CG}}}{\partial\boldsymbol{\lambda}}\right\rangle_{\text{FP}} - \beta\left\langle\frac{\partial U_{\text{CG}}}{\partial\boldsymbol{\lambda}}\right\rangle_{\text{CG}}, \tag{51}$$

and $\mathbf{H}$ gives the Hessian matrix of S_{rel} second derivatives,

$$\begin{aligned} \mathbf{H}_{ij} = {} & \beta\left\langle\frac{\partial^2 U_{\text{CG}}}{\partial\lambda_i\partial\lambda_j}\right\rangle_{\text{FP}} - \beta\left\langle\frac{\partial^2 U_{\text{CG}}}{\partial\lambda_i\partial\lambda_j}\right\rangle_{\text{CG}} + \beta^2\left\langle\frac{\partial U_{\text{CG}}}{\partial\lambda_i}\frac{\partial U_{\text{CG}}}{\partial\lambda_j}\right\rangle_{\text{CG}} \\ & - \beta^2\left\langle\frac{\partial U_{\text{CG}}}{\partial\lambda_i}\right\rangle_{\text{CG}}\left\langle\frac{\partial U_{\text{CG}}}{\partial\lambda_j}\right\rangle_{\text{CG}}. \end{aligned} \tag{52}$$

In the Newton–Raphson update of Eq. (50), k gives an iteration index along which $\boldsymbol{\lambda}$ converges to its value at the relative entropy minimum and χ is a mixing coefficient that can be adjusted to ensure stability of the algorithm ($\chi \to 1$ for usual updates close to the minimum). The two kinds of averages appearing here correspond to CG energy derivatives in different ensembles. The FP averages can be computed by first performing a single, well-converged simulation of the fully detailed reference system; for any given CG model, these averages are then calculated by either reprocessing U_{CG} for

CG configurations of the saved FP trajectory or by using the interatomic distance or angle distributions on which the $\boldsymbol{\lambda}$-components of U_{CG} are dependent. The CG averages, on the other hand, require complete evaluation at each iteration since the CG ensemble changes with $\boldsymbol{\lambda}$, and can be computed with a short CG simulation. It is important to note that the derivatives are inside the averages, and thus, simply evaluated from the analytical form of the CG potential alongside the energy evaluation itself.

The Newton–Raphson approach requires that the initial parameter values be in the vicinity of the minimum and the denominator (second derivative) be positive. For linear parameters, the second derivative is always positive; for nonlinear ones, a good initial guess is required. Alternatively, a steepest descent iteration based on the first derivative alone may be used to place the parameter near the minimum, using a suitable step size γ,

$$\lambda^{k+1} = \lambda^k - \gamma \nabla_\lambda S_{\mathrm{rel}}. \tag{53}$$

The CG averages appearing in these iteration equations can be evaluated by short equilibrium simulations using the CG system and the potential, similar to the approach of the inverse Monte Carlo method. Such a procedure introduces a stochastic element in the minimization; as such, the optimal parameters λ may need to be found by evaluating their mean value over a number of iterations at the final stages of the procedure. Alternatively, to make the CG averages deterministic, one can generate an initial reference CG trajectory using a known parameter set $\boldsymbol{\lambda}^0$ and then reweight the saved configurations for every new parameter set $\boldsymbol{\lambda}^k$ generated during the iterations,

$$\left\langle \frac{\partial U_{\mathrm{CG}}\left(\boldsymbol{\lambda}^k\right)}{\partial \boldsymbol{\lambda}} \right\rangle_{\mathrm{CG},\boldsymbol{\lambda}^k} = \left\langle \frac{\partial U_{\mathrm{CG}}\left(\boldsymbol{\lambda}^k\right)}{\partial \boldsymbol{\lambda}} e^{\beta \Delta U_{\mathrm{CG}}} \right\rangle_{\mathrm{CG},\boldsymbol{\lambda}^0} \left\langle e^{\beta \Delta U_{\mathrm{CG}}} \right\rangle_{\mathrm{CG},\boldsymbol{\lambda}^0}^{-1}, \tag{54}$$

where $\Delta U_{\mathrm{CG}}(\mathbf{R}) = U_{\mathrm{CG}}\left(\mathbf{R};\boldsymbol{\lambda}^0\right) - U_{\mathrm{CG}}\left(\mathbf{R};\boldsymbol{\lambda}^k\right)$. This approach would make the entire iteration procedure deterministic, but its success requires the initial parameters to lie sufficiently close to their ultimate optimal values for accurate reweighting. A related approach would be to reweight the reference FP trajectory directly,

$$\left\langle \frac{\partial U_{\mathrm{CG}}\left(\boldsymbol{\lambda}^k\right)}{\partial \boldsymbol{\lambda}} \right\rangle_{\mathrm{CG},\boldsymbol{\lambda}^k} = \left\langle \frac{\partial U_{\mathrm{CG}}\left(\boldsymbol{\lambda}^k\right)}{\partial \boldsymbol{\lambda}} e^{\Delta} \right\rangle_{\mathrm{FP}} \left\langle e^{\Delta} \right\rangle_{\mathrm{FP}}^{-1}. \tag{55}$$

These averages give the same expressions resulting from differentiation of the exponential average expression for S_{rel} in Eq. (43). While attractive for not requiring any CG simulations, one must again be mindful of associated statistical and systematic errors as described above.

Initial guesses for parameter values may be facilitated by using the high-temperature, energy-matching approximation to the relative entropy $\tilde{S}_{\mathrm{rel}}$ as in Eq. (45). A simple steepest-descent protocol for minimizing this quantity may be sufficient to locate a region of parameter space suitable to further refinement using the full expression for the relative entropy, of the form,

$$\lambda^{k+1} = \lambda^k - \gamma \left(\frac{\partial \tilde{S}_{\mathrm{rel}}}{\partial \lambda} \right) = \lambda^k - \gamma\beta^2 \left[\left\langle U_{\mathrm{FP}} - U_{\mathrm{CG}} \right\rangle_{\mathrm{FP}} \left\langle \frac{\partial U_{\mathrm{CG}}}{\partial \lambda} \right\rangle_{\mathrm{FP}} - \left\langle \left(U_{\mathrm{FP}} - U_{\mathrm{CG}}\right) \frac{\partial U_{\mathrm{CG}}}{\partial \lambda} \right\rangle_{\mathrm{FP}} \right]. \tag{56}$$

All of the averages in this expression stem from the FP ensemble alone, and thus, can simply be computed upon each iteration by reprocessing the reference trajectory. An especially simple scheme follows for linear parameters of the form $U_{\mathrm{CG}}(\mathbf{R}) = \lambda f(\mathbf{R})$. In this case, a closed expression for the optimal parameter value exists that requires only a reference FP simulation,

$$\lambda^0 = \left[\left\langle U_{\mathrm{FP}} f \right\rangle_{\mathrm{FP}} - \left\langle U_{\mathrm{FP}} \right\rangle_{\mathrm{FP}} \left\langle f \right\rangle_{\mathrm{FP}} \right] \left[\left\langle f^2 \right\rangle_{\mathrm{FP}} - \left\langle f \right\rangle_{\mathrm{FP}}^2 \right]^{-1}. \tag{57}$$

This approach is readily extended to the case in which multiple parameters are optimized at once.

V. CASE STUDY: LATTICE GAS

To demonstrate these conceptual and numerical strategies, we address a classic problem, that of generating a mean-field approximation to the 2D lattice gas. This system enables extensive sampling to characterize algorithmic convergence and statistical errors in computing the relative entropy. We focus largely on these issues here, while a previous study using it addressed the ability of S_{rel} to predict coarse-graining errors due to the use of a mean field.[20] This model also demonstrates the applicability of the relative entropy approach using a system for which many conventional coarse-graining strategies cannot be directly applied.

In this model, the interactions rather than the degrees of freedom are coarse-grained. The lattice gas studied has $M = 16^2$ lattice sites $\vec{\imath}$, each with $n_{\vec{\imath}} \in \{0, 1\}$ molecules such that their total number is $N = \sum_{\vec{\imath}} n_{\vec{\imath}}$. The FP model interacts through nearest-neighbors, $U_{\mathrm{FP}}(i) = -\sum_{\vec{\imath},1} n_{\vec{\imath}} n_{\vec{\imath}+\vec{1}}$, while the CG one interacts through a mean field, $U_{\mathrm{CG}}(I) = -\lambda \sum_{\vec{\imath}} n_{\vec{\imath}}$. As no degrees of freedom are removed, the FP configuration i is identical with its CG configuration I, and the mapping entropy vanishes. The energy scale is set by the FP interaction, and λ of the CG potential is the mean-field parameter for optimization. In both cases, the gas is maintained at a constant chemical potential μ. By expressing the interactions in terms of a grand energy, $[U - \mu N]$, one can make use of the above canonical framework.

In previous work, we compared the S_{rel}-optimized mean-field model with that of the traditional mean-field approach.[20] We found an explicit, closed-form solution for the location of the minimum in S_{rel} via Eq. (10) with respect to the mean field λ,

$$\lambda = -k_B T \ln\left[\frac{M}{\langle N \rangle_{\mathrm{FP}}} - 1 \right] - \mu, \tag{58}$$

which requires the average occupancy in the FP ensemble. This is in contrast to the traditional transcendental mean-field solution which entails the average occupancy in the CG ensemble (obtained by minimizing S_{var}, as discussed in Sec. III). Importantly, the CG system utilizing the

S_{rel}-optimized λ replicated properties of the lattice gas (e.g., $\langle N\rangle$ and $\langle N^2\rangle$) better than the one using the S_{var}-optimized parameter. In particular, as the mean-field parameter is linear in particle number, the coarse-graining error considerations in Sec. II show that S_{rel} minimization explicitly reproduces the average occupancy, $\langle N\rangle_{\text{CG}} = \langle N\rangle_{\text{FP}}$. Most importantly, we found that the relative entropy was a robust indicator of the failure of the mean-field description. Near the critical point, its value increased together with various coarse-graining errors, especially those of particle correlations. We were also able to develop several scaling laws for the relationship of these errors to S_{rel} that agreed well with numerical results.

We here consider the same coarse-graining scenario, but rather than appealing to the analytical result in Eq. (58), we characterize the numerical strategies outlined in Sec. IV in finding the optimal mean field λ and in estimating the magnitude of S_{rel}. This, thus, serves as a basic test of these numerical procedures that can be compared to a known exact result. We focus on the perturbation reformulation of the relative entropy of Eq. (43), involving an exponential average, and the associated Gaussian approximate shown in Eq. (45); we consider the corresponding iterative scheme for numerical minimization. The practical advantage of these equations is that all involve only averages in the FP ensemble. Thus for clarity in the remainder of this section, ensemble references in averages are omitted but always implicitly pertain to the FP system.

Each coarse-graining procedure requires a reference trajectory of the FP system. We perform Monte Carlo simulations for 10^6 steps using single-particle occupancy moves; every 20 steps, the instantaneous particle number and the potential energy are recorded. This is performed for a grid of state points in β and μ. Importantly, the lattice gas behaves symmetrically about the critical value of the chemical potential, $\mu_C = -2$, and the critical temperature is at $\beta_C \approx 1.6$. For each state point, 10 replicates are generated. In the following, average results over each of these 10 are shown and when relevant, the respective errors are also provided.

A. Numerical model optimization and estimates of the relative entropy

Relative entropy determination of the mean-field parameter λ starts with the energy-matching estimate of Eq. (57), $\lambda^0 = -\text{cov}(U, N)/\text{var}(N)$, and repeatedly continues by Newton–Raphson iterations as per Eq. (55) until there is a negligible change in its value,

$$\lambda^{k+1} = \lambda^k + \chi\left[\frac{\langle N\rangle - \langle Nw\rangle}{\beta\langle N^2 w\rangle - \beta\langle Nw\rangle^2}\right], \quad w = \frac{e^{\beta(U+\lambda N)}}{\langle e^{\beta(U+\lambda N)}\rangle}. \tag{59}$$

For this simple test system, thousands of samples are averaged per iteration and parameters appear to converge quickly to an optimum (typically five iterations to converge to within two decimal points).

The final value of this iteration is representative of the S_{rel}-optimal parameter, and it is referred throughout as the exact parameter. Its value in state space is shown in Fig. 1

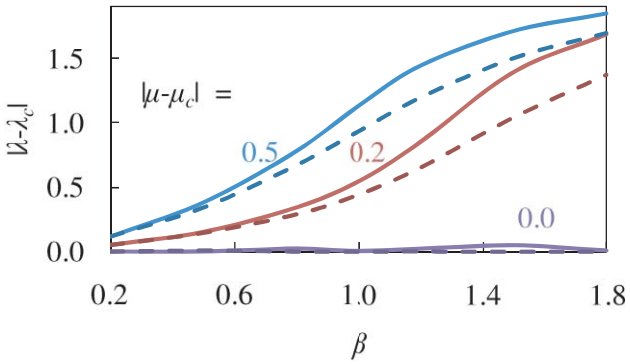

FIG. 1. Optimized mean-field parameters in state space. Two optimizations are shown: the solid curves are for the values attained by minimization of S_{rel} given by Eq. (43), while the dashed curves from minimization of the approximate $\tilde{S}_{\text{rel}}$ of Eq. (45). Because of symmetry, the parameter is shifted by its critical value $\lambda_C = 2$; values are shown in terms of the inverse temperature, and each pair of isolines corresponds to a different chemical potential. The same notation for determination of the mean-field parameter (solid vs dotted lines) persists in subsequent figures.

(solid curves). Regardless of temperature, all optimization at the critical value of the chemical potential yields roughly the same mean-field parameter, $\lambda_C = 2$, which was observed in our previous work, even for the traditional mean-field approach.[20] Nevertheless, there is much statistical uncertainty for this isoline, especially near the critical point (where it slightly surpasses 2% error); this stems from large fluctuations in this region, which in turn give rise to larger numerical errors for the exponentials appearing in Eq. (59). The magnitude of these errors is discussed in greater depth below. For other chemical potentials, the mean-field parameter exhibits a notable inflection with respect to β; at high temperatures, however, it approaches its critical value. Compared to our previous work in which the optimal parameter is determined by the analytical Eq. (58), this iterative approach attains nearly the same optimal values; they are only noticeably misestimated (with the error in λ being above 0.02, but still below 0.1) for the few state points within the critical radius, where the relative entropy is fairly high.

The initial value of the parameter before iteration is a minimizer of the approximate $\tilde{S}_{\text{rel}}$, and it is referred to throughout as the approximate parameter. Its value in state space is also shown in Fig. 1 (dashed curves). Surprisingly, this value is identical to the "full" S_{rel} minimum along the μ_C isoline, and it does so with much better statistics (e.g., its statistical uncertainty is never above 0.1%). It is not likely that the Gaussian assumption for the relative entropy is valid here, as the actual relative entropy value is overestimated; therefore, the correspondence of the exact and approximate parameters may simply result from the lattice gas' symmetry about the critical chemical potential. For other chemical potentials, the two parameters also agree at high temperatures as a result of the Gaussian assumption's increasing accuracy there. In regions of moderate to low temperature and away from the critical chemical potential, however, deviations between the approximate and exact parameters become significant. We examine these two parameterizations throughout the rest of the case study. The notation in all figures remains consistent with

that of Fig. 1: solid curves and filled symbols are for the model using the exact parameter and dashed curves with empty symbols for the one using the approximate parameter.

We now proceed to examine the value of the relative entropy for these parameterizations, keeping with the notations as previously defined. Eq. (43) for $S_{\rm rel}$ gives in this case,

$$S_{\rm rel}(\lambda) = \ln\left\langle e^{\beta(U-\langle U\rangle)+\beta\lambda(N-\langle N\rangle)}\right\rangle. \quad (60)$$

We first evaluate it for models utilizing the exact mean-field parameter. This is representative of a *twofold-exact* coarse-graining scheme, in which the Gaussian assumption is never invoked, neither in the optimization procedure nor in the computation of the relative entropy. The top panel of Fig. 2 shows the behavior of $S_{\rm rel}$ in state space (solid curves). In tandem with our previous work, the critical μ-isoline sharply increases toward the critical temperature, while other μ-isolines exhibit a maximum. Thus, the growth of the relative entropy in the vicinity of the critical point naturally signals the inadequacy of the mean-field description there.[20]

At the same time, $\tilde{S}_{\rm rel}$ can be evaluated by Eq. (45),

$$\tilde{S}_{\rm rel}(\lambda) = \tfrac{1}{2}\beta^2\left[{\rm var}(U) + 2\lambda{\rm cov}(U,N) + \lambda^2{\rm var}(N)\right]. \quad (61)$$

In computing it for the models utilizing the approximate mean-field parameter, this is representative of a *twofold-approximate* coarse-graining scheme, in which the Gaussian assumption is invoked in both the optimization procedure and the computation of the relative entropy. The bottom panel of Fig. 2 shows the behavior of this value of the relative entropy in state space (dashed curves). Importantly, the Gaussian assumption attains qualitatively similar behavior as that for the twofold-exact version, but with significant quantitative differences (being significantly greater). Thus, the Gaussian assumption appears to provide a reasonable strategy if one is interested only in the comparative behavior of the relative entropy.

Let us now extend our study beyond these "twofold" scenarios and examine "cross" effects in using the Gaussian assumption in only one of the optimization or $S_{\rm rel}$ estimation approaches. Using Eqs. (60) and (61), we compute the exact $S_{\rm rel}$ for the approximate value of the mean-field parameter given by minimization of $\tilde{S}_{\rm rel}$ (top of Fig. 2, dashed curves), and the approximate $\tilde{S}_{\rm rel}$ for the exact value of the mean-field parameter given by minimization of $S_{\rm rel}$ (bottom of Fig. 2, solid curves). Unsurprisingly, $S_{\rm rel}$ values in the top panel for the exact mean field are always lower than those for the approximate one, as guaranteed by explicit minimization of $S_{\rm rel}$ for the former; for similar reasons, the opposite is true for $\tilde{S}_{\rm rel}$ values in the bottom panel. Regardless, the full and approximate relative entropy estimates are fairly insensitive to the method by which the optimal mean field is located, though their values relative to each other differ more substantially. The comparison of the two estimates is also much better than what would be predicted if λ were extracted using the traditional mean-field closure (as discussed in our previous study). This suggests that invoking the Gaussian assumption for parameterization (i.e., energy matching) provides a reasonable first estimate of the optimal parameters.

It is noteworthy to examine the relationship between the exact and approximate relative entropy estimates. For each optimized mean field, we compare its $\tilde{S}_{\rm rel}$ and $S_{\rm rel}$ values in the bottom panel of Fig. 3. All points are above the unity-slope line, which suggests systematic overestimation of the relative entropy by the Gaussian assumption, at least for this case study. For low values of the relative entropy, the Gaussian approximate predicts the correct value within 5% relative error, but for high values, the error soars above 100%. Thus, careful consideration of the magnitude of the relative entropy might inform the accuracy of the Gaussian assumption. These findings reinforce our results above: namely, the Gaussian approximate might provide a qualitative description of the

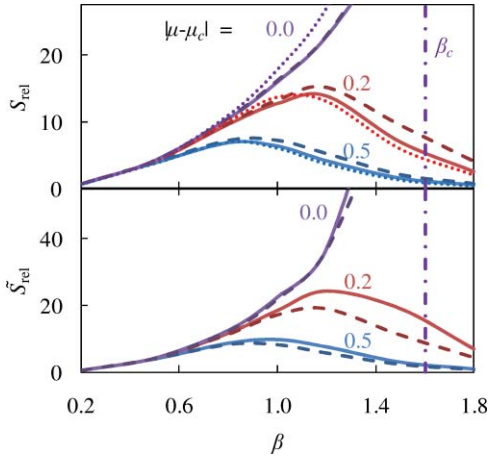

FIG. 2. Relative entropy in state space. Values are shown in terms of the inverse temperature, and each pair of isolines corresponds to a different chemical potential. The top panel is for the exact relative entropy while the bottom panel is for the approximate one. Note that the scale of the bottom panel is twice as much as the top one. Both parameterizations for the mean field appear in each panel. In the top panel, the dotted curve stems from the approximation to the relative entropy of Eq. (49) that is independent of the mean-field value (i.e., that does not require CG optimization).

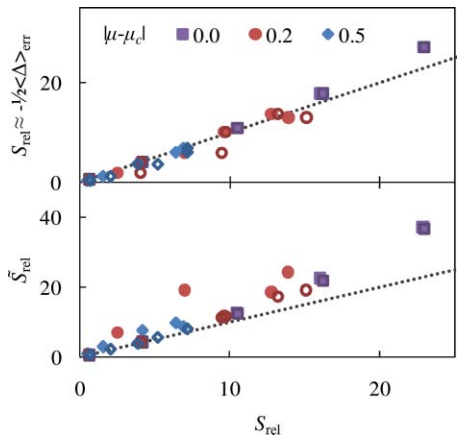

FIG. 3. Analysis of estimates for the relative entropy. The top panel is for the estimate of Eq. (62) (i.e., independent of CG optimization), with $\langle\Delta\rangle_{\rm err} = \langle\Delta\rangle_{\rm FP} - \langle\Delta\rangle_{\rm CG}$, while the bottom panel is for the direct application of the Gaussian approximate of Eq. (61); both of these are shown in terms of the full value of the relative entropy by Eq. (60). In both, a line of unity slope is given (the ideal scenario in which each estimate is perfect). For each panel, two parameterizations are shown: filled symbols for the values attained by $S_{\rm rel}$ minimization, and empty symbols for values attained by $\tilde{S}_{\rm rel}$ minimization.

relative entropy's behavior, but has systematic quantitative inconsistencies.

Equation (49), which is a further manipulation of the Gaussian assumption, provides an estimate of the relative entropy without requiring a value of λ. In this case, one finds, [see SI (Ref. 36)]

$$S_{\text{rel}} \approx M\beta(\langle n_{\vec{0}} n_{\vec{1}}\rangle - \langle n\rangle^2), \tag{62}$$

where the term $\langle n_{\vec{0}} n_{\vec{1}}\rangle$ indicates the correlation of nearest-neighbor occupancies. On an intuitive basis, one expects the relative entropy to relate to such correlations in the reference system as they are ultimately tied to the failure of the mean-field closure in describing critical phenomena. For other coarse-graining scenarios, similar simplifications to the relative entropy on the basis of the kinds of approximations made here might also hold promising for pointing toward physical attributes and correlations that make a particular CG model deficient. In addition, we note that, unlike the previous expressions of Eqs. (60) and (61), neither the mean field nor any other information from the CG model enters this equation beyond the mere mapping to a one-body, mean-field interaction. Equation (62) gives, in effect, a broad estimate for the entire function space of mean-field models that does not require optimization. Figure 2 (top panel, dotted curves) and Fig. 3 (top panel) show that this CG-free approximation follows the exact S_{rel}, with surprising accuracy.

In the infinite temperature limit, the relative entropy becomes inversely proportional with the square of the temperature and the prefactor is the variance of the configuration-space averaged potential energy difference, as by Eq. (48),

$$\lim_{\beta\to 0} S_{\text{rel}} = \frac{M}{16}\beta^2, \tag{63}$$

where we use the finding that the mean-field parameter for any chemical potential approaches $-\mu_C$ at high temperature. In Fig. 2, all curves at high temperature do, in fact, collapse onto this particular power law. Remarkably, we observe that the relative entropy along the critical value of the chemical potential (at least in the supercritical region) also appears to follow this scaling law, within uncertainty. This is another aspect of the case study for which there is a similarity between $\beta \to 0$ and μ_C (the other resemblance being that optimization via any method yields λ_C for both of these state conditions). This power law, which starts at an infinite temperature for all isolines, continues toward the critical temperature only for the critical isoline.[20] By further examination of Fig. 2, it appears that this power law creates an upper bound for the relative entropy at a given temperature (for optimal models).

B. Statistical and systematic errors

Of additional interest for this case study are statistical errors within the perturbation-based relative entropy coarse-graining strategy. A central characterization is the number of (independent) reference FP ensemble samples required to converge the coarse-graining procedure and to estimate the relative entropy. To do so, we examine fluctuations in quantities optimized or computed from the 10 simulation replicates at each state point. We vary the number of samples taken from each replicate to assess the dependence of computed quantities on simulation length. We verify that these samples are always statistically independent by ensuring that they are taken at times spaced farther apart than the correlation times in the energy and particle number. Below, we present a representative set of analyses for a small selection of state points.

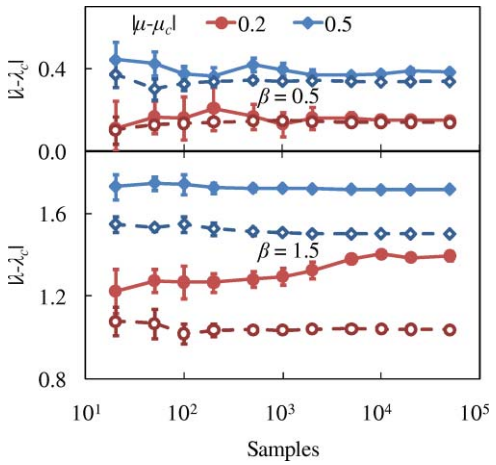

FIG. 4. Convergence of the optimized mean-field parameters. The optimal parameters are shown as a function of the number of independent samples used in computing them. The panels differ in temperature, with $\beta = 0.5$ for the top and $\beta = 1.5$ for the bottom. The y-axis scales in the panels are identical.

Figure 4 shows the convergence behavior of the mean-field parameter for both the exact and approximate parameterization methods. Both approaches converge at long times. The approximate parameter converges with only $\sim 10^2$ samples for all of these state points, while the exact parameter shows significant variability in the rate of convergence: at $\beta = 1.5$, $|\mu - \mu_C| = 0.5$ has very fast convergence ($\sim 10^2$ samples) while $|\mu - \mu_C| = 0.2$ has very slow convergence ($\sim 10^4$ samples). Clearly, the approximate method leads to more efficient convergence, as it consistently reaches stable values of λ with less samples than the exact one. Still, in some cases, notably those near the critical locus, the approximate parameter converges to a significantly different value from that of the exact method.

Figure 5 shows that the corresponding convergence behavior of the relative entropy (exact and approximate) is reminiscent of that of the mean-field parameters. $\tilde{S}_{\text{rel}}$ converges with only $\sim 10^2$ samples, while S_{rel} shows significant variability, with the same extremes as discussed before. One trend present here but absent in the convergence of the mean-field parameter is the consistent growth of the relative entropy with the number of samples. This is likely due to the effects of rare fluctuations in Δ that have a dramatic effect on the value of the relative entropy.

Why do models based on $\tilde{S}_{\text{rel}}$ consistently converge with $\sim 10^2$ samples, while models based on S_{rel} require as many as $\sim 10^4$ samples to converge? These findings can be interpreted using the statistical error estimates for the relative entropy described in Sec. IV. These earlier considerations show that the value of the relative entropy should signal its own uncertainty. To test these relationships, we compute the statistical errors in the relative entropy using the complete trajectories from the

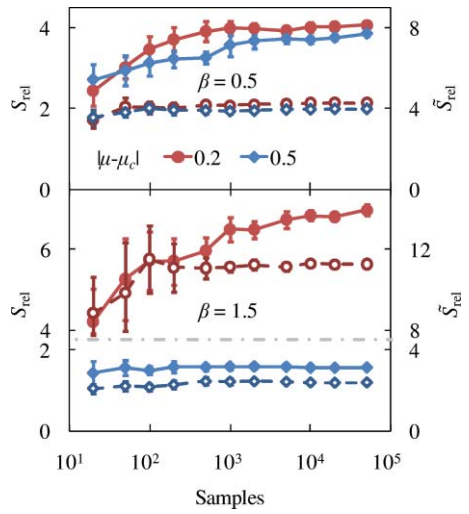

FIG. 5. Convergence of the relative entropy. The left axis is for the full relative entropy, while the right axis is for the approximate one. The panels differ in temperature as per Fig. 4, but the scales of their y-axes are the same, except for a break between 2 and 4 in the bottom panel.

10 replicates. Figure 6 shows the error in S_{rel} (top panel) as well as the corresponding analytical relations expected for the errors of both S_{rel} and $\tilde{S}_{rel}$ [Eqs. (44) and (46)]. For all values of the relative entropy, the prediction of the error by the exponential relation greatly overestimates the actual error. This encouraging result suggests that calculation of the relative entropy via the perturbation expression (exponential average) is more promising than initially expected. Interestingly, while the error seems to be bounded above by the exponential prediction, it is also bounded below by the linear relation derived for the corresponding error of $\tilde{S}_{rel}$. Consequently, the relative

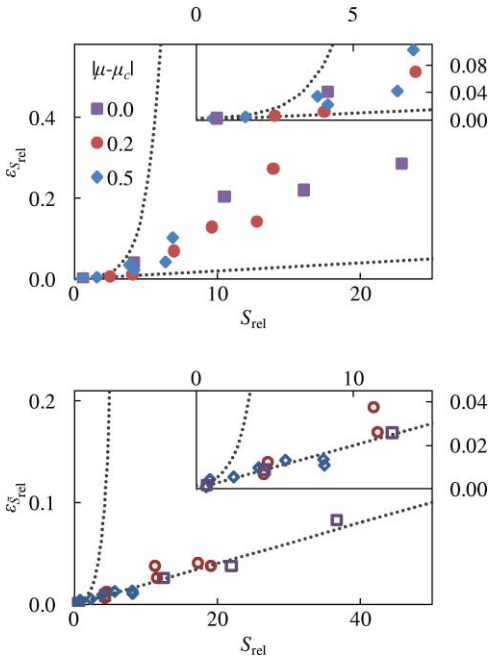

FIG. 6. Statistical error in the relative entropy. The top portion is for the exact relative entropy, while the bottom for the approximate one. In both panels, analytical estimates for the error are shown (dotted curves). The insets magnify the comparison with these relations closer to the origin.

uncertainty grows for higher values of S_{rel}. This explains why there is fast (slow) convergence for models with low (high) S_{rel}. The much lower error in $\tilde{S}_{rel}$, shown in the bottom of Fig. 6, follows the analytical prediction remarkably well, and exhibits a strongly linear relation with $\tilde{S}_{rel}$. The near constancy of the relative error in computing $\tilde{S}_{rel}$ might explain the near constant convergence rates for it.

VI. CONCLUSIONS

In this work, we presented a broad framework for coarse-graining using the relative entropy. This quantity measures the deficiency of a particular CG model relative to a reference FP one,[12] and it is a natural indicator of coarse-graining errors (i.e., discrepancies between properties of the two models).[20] A central concept in this approach is that optimal CG models have minimal values of the relative entropy. On one level, the relative entropy potentially provides a scoring metric for discriminating the quality of CG models differing in design, i.e., with different numbers of sites. On another, the interaction potential for a given CG model can be optimized by minimizing the relative entropy with respect to its free parameters. In particular, we showed that the form of the terms in the CG potential directly implies which coarse-graining errors will be eliminated by relative entropy minimization. As such, we showed that this framework provides a strategy for rationally designing (selecting basis functions for) the CG potential so as to eliminate errors in the average values, correlations, or distributions of arbitrary microscopic observables.

We also demonstrated connections between relative entropy minimization and other major coarse-graining methodologies. We discussed its equivalence with the force matching technique[6,7,11] in the limit of an unconstrained CG energy function (i.e., when the CG model achieves the ideal multidimensional PMF), and with iterative Boltzmann inversion[1,2,4,16] in the case of finely tabulated (typically pair additive) potentials. We also show that the energy matching and the traditional mean-field variational closure bear stark similarities but are distinct from the relative entropy approach; however, all three methods become identical under special conditions (e.g., infinite temperature).

We outlined several numerical approaches to computing the relative entropy and to locating its minimum in the CG model parameter space. We developed an exact expression for the relative entropy involving an ensemble average and an approximate one subject to less statistical error. Using these formulations with Newton–Raphson descent, we presented several strategies for minimizing the relative entropy. To test the new formulations, we examined numerical convergence and statistical errors in developing a mean-field approximation to the classic lattice gas, a problem that has for comparison a known analytical expression via the relative entropy. Our results found good convergence behavior and excellent agreement of some of the approximate relative entropy expressions with their true values.

In total, we believe that these results suggest that the relative entropy is a powerful and practical new coarse-graining strategy, applicable to a wide variety of systems as well as providing unique ways to assess and eliminate

coarse-graining errors. Future work will focus on applying these new numerical and theoretical techniques to more complex and realistic systems.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the Camille & Henry Dreyfus Foundation (New Faculty Award) and the National Science Foundation (Award No. CBET-0845074).

[1] A. P. Lyubartsev and A. Laaksonen, Phys. Rev. E **52**, 3730 (1995).
[2] A. K. Soper, Chem. Phys. **202**, 295 (1996).
[3] J. Baschnagel, K. Binder, P. Doruker, A. A. Gusev, O. Hahn, K. Kremer, W. L. Mattice, F. Müller-Plathe, M. Murat, and W. Paul, Adv. Polym. Sci. **152**, 41 (2000).
[4] F. Muller-Plathe, Chemphyschem **3**, 754 (2002).
[5] J. J. de Pablo, AIChE J. **51**, 2372 (2005).
[6] S. Izvekov and G. A. Voth, J. Phys. Chem. B **109**, 2469 (2005).
[7] S. Izvekov and G. A. Voth, J. Chem. Phys. **123**, 134105 (2005).
[8] G. S. Ayton, W. G. Noid, and G. A. Voth, Curr. Opin. Struct. Biol. **17**, 192 (2007).
[9] P. Sherwood, B. R. Brooks, and M. S. P. Sansom, Curr. Opin. Struct. Biol. **18**, 630 (2008).
[10] M. Praprotnik, L. D. Site, and K. Kremer, Annu. Rev. Phys. Chem. **59**, 545 (2008).
[11] W. G. Noid, J. W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, J. Chem. Phys. **128**, 244114 (2008).
[12] M. S. Shell, J. Chem. Phys. **129**, 144108 (2008).
[13] J. W. Mullinax and W. G. Noid, J. Chem. Phys. **131**, 104110 (2009).
[14] J. W. Mullinax and W. G. Noid, J. Phys. Chem. C **114**, 5661 (2009).
[15] T. Murtola, M. Karttunen, and I. Vattulainen, J. Chem. Phys. **131**, 055101 (2009).
[16] A. Lyubartsev, A. Mirzoev, L. J. Chen, and A. Laaksonen, Faraday Discuss. **144**, 43 (2010).
[17] J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).
[18] T. L. Hill, *An Introduction to Statistical Thermodynamics* (Addison-Wesley, Reading, MA, 1960).
[19] A. Chaimovich and M. S. Shell, Phys. Chem. Chem. Phys. **11**, 1901 (2009).
[20] A. Chaimovich and M. S. Shell, Phys. Rev. E **81**, 060104 (2010).
[21] R. L. Henderson, Phys. Lett. **49**, 197 (1974).
[22] R. L. McGreevy and L. Pusztai, Mol. Simul. **1**, 359 (1988).
[23] R. L. C. Akkermans and W. J. Briels, J. Chem. Phys. **114**, 1020 (2001).
[24] D. Reith, M. Puetz, and F. Mueller-Plathe, J. Comput. Chem. **24**, 1624 (2003).
[25] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein, J. Phys. Chem. B **105**, 4464 (2001).
[26] T. Head-Gordon and F. H. Stillinger, J. Chem. Phys. **98**, 3313 (1993).
[27] A. A. Louis, J. Phys.: Condens. Matter **14**, 9187 (2002).
[28] M. E. Johnson, T. Head-Gordon, and A. A. Louis, J. Chem. Phys. **126**, 144509 (2007).
[29] T. Head-Gordon and M. E. Johnson, J. Chem. Phys. **130**, 214510 (2009).
[30] F. Ercolessi and J. B. Adams, Europhys. Lett. **26**, 583 (1994).
[31] V. Ruhle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, J. Chem. Theory Comput. **5**, 3211 (2009).
[32] S. Kullback and R. A. Leibler, Ann. Math. Stat. **22**, 79 (1951).
[33] F. Cilloco, J. Mol. Struct. **296**, 253 (1993).
[34] M. D'Alessandro and F. Cilloco, Phys. Rev. E **82**, 021128 (2010).
[35] D. Ming and M. E. Wall, Phys. Rev. Lett. **95**, 198103 (2005).
[36] See supplementary material at http://dx.doi.org/10.1063/1.3557038 for a detailed derivation of some of the expressions presented.
[37] E. Darve, *New Algorithms for Macromolecular Simulation* (Springer, Berlin, 2006), vol. 49, pg. 213.
[38] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987).
[39] C. Chipot and A. Pohorille, *Free Energy Calculations: Theory and Applications in Chemistry and Biology* (Springer, Berlin, 2007).
[40] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).