Testing the mutual information expansion of entropy with multivariate Gaussian distributions

Martin Goethe, Ignacio Fita, and J. Miguel Rubi

Articles you may be interested in

Principal component analysis on a torus: Theory and application to protein dynamics
The Journal of Chemical Physics **147**, 244101 (2017); 10.1063/1.4998259

Perspective: Maximum caliber is a general variational principle for dynamical systems
The Journal of Chemical Physics **148**, 010901 (2018); 10.1063/1.5012990

Another resolution of the configurational entropy paradox as applied to hard spheres
The Journal of Chemical Physics **147**, 224503 (2017); 10.1063/1.4999483

On-the-fly analysis of molecular dynamics simulation trajectories of proteins using the Bayesian inference method
The Journal of Chemical Physics **147**, 124108 (2017); 10.1063/1.4997099

Transition path time distributions
The Journal of Chemical Physics **147**, 214103 (2017); 10.1063/1.5000423

General application of Tolman's concept of activation energy
The Journal of Chemical Physics **147**, 224111 (2017); 10.1063/1.5009751

Testing the mutual information expansion of entropy with multivariate Gaussian distributions

Martin Goethe,[1,a)] Ignacio Fita,[2] and J. Miguel Rubi[1]

[1]*Department of Condensed Matter Physics, University of Barcelona, Carrer Martí i Franqués 1, 08028 Barcelona, Spain*
[2]*Molecular Biology Institute of Barcelona (IBMB-CSIC, Maria de Maeztu Unit of Excellence), Carrer Baldiri Reixac 4-8, 08028 Barcelona, Spain*

The mutual information expansion (MIE) represents an approximation of the configurational entropy in terms of low-dimensional integrals. It is frequently employed to compute entropies from simulation data of large systems, such as macromolecules, for which brute-force evaluation of the full configurational integral is intractable. Here, we test the validity of MIE for systems consisting of more than $m = 100$ degrees of freedom (dofs). The dofs are distributed according to multivariate Gaussian distributions which were generated from protein structures using a variant of the anisotropic network model. For the Gaussian distributions, we have semi-analytical access to the configurational entropy as well as to all contributions of MIE. This allows us to accurately assess the validity of MIE for different situations. We find that MIE diverges for systems containing long-range correlations which means that the error of consecutive MIE approximations grows with the truncation order n for all tractable $n \ll m$. This fact implies severe limitations on the applicability of MIE, which are discussed in the article. For systems with correlations that decay exponentially with distance, MIE represents an asymptotic expansion of entropy, where the first successive MIE approximations approach the exact entropy, while MIE also diverges for larger orders. In this case, MIE serves as a useful entropy expansion when truncated up to a specific truncation order which depends on the correlation length of the system. *Published by AIP Publishing.* https://doi.org/10.1063/1.4996847

I. INTRODUCTION

Computer simulations have become a major tool for unraveling the microscopic mechanisms behind biological processes. Trajectories of molecular dynamics simulations give full access to the time evolution of all simulated atoms, which, in principle, allows to watch biology at play on the computer screen. Snapshots of Monte Carlo simulations, albeit not in chronological order, still permit to measure the exact spatial distribution of the atoms at equilibrium. It is for these powerful capabilities that both techniques have been applied extremely successfully to a variety of biological systems during the last 40 years.[1]

Many biological processes are equilibrium processes which occur spontaneously with no energy consumption required. Those processes are often governed by a delicate balance between the (time-averaged) energy and entropy.[2–6] For a theoretical understanding of the process, it is essential to obtain good estimates for both terms from the simulation. Taking only energy into account might be insufficient since entropy has strong impact on important processes such as protein folding and ligand binding.[7–16] The behavior of more flexible proteins, such as intrinsically disordered proteins, is even strongly controlled by entropy.[17–19]

While acceptable energy estimates can directly be obtained from the simulations, entropy is less accessible since entropy is an ensemble measure. Precisely, the entropy S is defined as the integral

$$S \equiv -k_{\mathrm{B}} \int dx^m \, \rho(\boldsymbol{x}) \log \rho(\boldsymbol{x}) \tag{1}$$

performed over all m degrees of freedom (dofs) of the system, $x_1, x_2, \ldots, x_m$. Here $\rho(\boldsymbol{x})$ denotes the joint probability density function (pdf) of the dofs and k_{B} denotes the Boltzmann constant. Computing the integral from the simulation snapshots is hindered by the "curse of dimensionality"[20] which implies that measuring S accurately requires a number of snapshots M_{data} which scales exponentially with m. This renders brute-force evaluation of Eq. (1) impossible for large m.

To obtain entropy estimates also for macromolecules, approximations of the integral (1) need to be considered. Various approaches have been proposed for this purpose which can be classified into parametric and non-parametric methods.[21] Parametric methods are based on the assumption that the dofs are distributed according to a particular pdf for which Eq. (1) can be integrated analytically. The data are fitted to this pdf, and the associated entropy of the fitted pdf is used as an estimate for S. Most approaches are based on the multivariate Gaussian distribution[22–25] but also other distributions[26,27] have been employed. Parametric methods have the technical advantage that no numerical integration is involved. Their accuracy is determined by the quality of the involved fit, and hence,

a)Electronic mail: martingoethe@ub.edu

depends crucially on the used coordinate representation.[16,28] While the parametric assumption may hold sufficiently well in a particular coordinate system, it may fail entirely in another.

Non-parametric methods rely on an alternative approach unrelated to any particular pdf. Equation (1) is approximated in terms of low-dimensional integrals which represent the marginal entropies of subsets containing up to n dofs. The truncation order n is chosen sufficiently small such that all integrals can be computed reliably from the snapshots via binning or more elaborated integration techniques.[29,30] In real-world applications, n equals a small integer, typically less than six, since the curse of dimensionality dictates $n = \mathcal{O}(\log(M_{\text{data}}))$.

Two unrelated non-parametric approaches have been proposed which are feasible for large systems: the maximum information spanning tree expansion (MIST)[31,32] and the mutual information expansion (MIE).[33–35] MIST is based on an expansion of Eq. (1) which is truncated at order n. This expansion is mathematically rigorous as guaranteed by the exact inequalities $S_{\text{MIST}}^{(1)} \geq S_{\text{MIST}}^{(2)} \geq \cdots \geq S_{\text{MIST}}^{(m-1)} \geq S_{\text{MIST}}^{(m)} = S$, for consecutive MIST approximations $S_{\text{MIST}}^{(n)}$.[31] Hence, MIST truncated at successive orders $n = 1, 2, \ldots, m$ approximate S with increasing accuracy.

MIE provides an expression for S which can be derived from Eq. (1) by iterative application of the generalized Kirkwood superposition approximation.[34] Also the untruncated MIE expression equals S exactly (see Sec. II). However, unlike MIST, successive truncations of MIE are not guaranteed to yield entropy approximations of increasing quality. Meaning, it has not been proven that the error $|S - S_{\text{MIE}}^{(n)}|$ of the nth MIE approximation $S_{\text{MIE}}^{(n)}$ decreases for consecutive truncation orders n. Moreover, MIE has never been tested for high-dimensional systems containing more than $m \approx 20$ dofs. Hence, it is unknown whether MIE truncated at low order $n \ll m$ approximates S acceptably well for large systems. This is of particular importance since MIE has been applied to a multitude of systems.[11,19,36–41]

In this work, we test MIE for large systems of up to $m = 150$ dofs and various truncation orders $n = 1, 2, \ldots, 6$. This represents the natural situation for which MIE has been proposed, namely, when brute-force estimation of S is intractable. We use multivariate Gaussian distribution for which S as well as all MIE contributions can be calculated semi-analytically to arbitrary precision. Our test highlights severe limitations for the applicability of MIE.

II. METHODS

A. The mutual information expansion

MIE[33–35] expresses the high-dimensional integral of Eq. (1) in terms of specific low-dimensional ones which can reliably be computed from the simulation snapshots. We consider a system of m dofs described by continuous coordinates $x_1, x_2, \ldots, x_m$. The set of all coordinates is denoted by $\mathcal{A}_m \equiv \{x_1, x_2, \ldots, x_m\}$ while subsets containing exactly k coordinates are denoted by $\mathcal{U}_k$. As short hand for a coordinate set $\{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$, we use the set of indices $\{i_1, i_2, \ldots, i_k\}$. The marginal entropy of a set of dofs $\mathcal{U}_k = \{i_1, i_2, \ldots, i_k\}$ is

the k dimensional integral

$$S_k(\mathcal{U}_k) \equiv -k_{\text{B}} \int dx_{i_1} \, dx_{i_2} \ldots dx_{i_k} \, \rho(\mathcal{U}_k) \log \rho(\mathcal{U}_k), \quad (2)$$

where $\rho(\mathcal{U}_k)$ denotes the joint probability distribution of $\mathcal{U}_k$. The lower index k of S_k reflects the dimensionality of the integral. The mutual information (MI) of two dofs $\mathcal{U}_2 = \{1, 2\}$ is defined as

$$I_2(\mathcal{U}_2) \equiv S_1(1) + S_1(2) - S_2(1, 2). \quad (3)$$

MI has been generalized to larger sets of dofs $\mathcal{U}_k$ with $k > 2$. It is defined as

$$I_k(\mathcal{U}_k) \equiv \sum_{l=1}^{k} (-1)^{l+1} \sum_{\mathcal{X}_l \subset \mathcal{U}_k} S_l(\mathcal{X}_l), \quad (4)$$

where $\mathcal{X}_l$ denotes a subset of $\mathcal{U}_k$ containing exactly l coordinates and where the second sum runs over all $\binom{k}{l}$ possible subsets $\mathcal{X}_l$. For example, the third order MI for $\mathcal{U}_3 = \{1, 2, 3\}$ is given by

$$I_3(\mathcal{U}_3) = S_1(1) + S_1(2) + S_1(3) - S_2(1, 2)$$
$$- S_2(1, 3) - S_2(2, 3) + S_3(1, 2, 3). \quad (5)$$

Using these definitions, one can write the total entropy S as

$$S = S_m(\mathcal{A}_m) = \sum_{k=1}^{m} (-1)^{k+1} T_k, \quad (6)$$

where

$$T_k \equiv \sum_{\mathcal{U}_k \subset \mathcal{A}_m} I_k(\mathcal{U}_k) \quad (7)$$

is the sum of the mutual information for all possible $\binom{m}{k}$ sets $\mathcal{U}_k \subset \mathcal{A}_m$. Note that the lower index k of I_k and T_k reflects the highest dimension of all involved integrals.

The mutual information expansion of order n represents a truncation of Eq. (6) after n terms,[33–35] i.e.,

$$S_{\text{MIE}}^{(n)} \equiv \sum_{k=1}^{n} (-1)^{k+1} T_k \quad (8)$$

$$= \sum_{k=1}^{n} (-1)^{n-k} \binom{m-k-1}{n-k} \sum_{\mathcal{U}_k \subset \mathcal{A}_m} S_k(\mathcal{U}_k). \quad (9)$$

Equation (9) is obtained by rearrangement of all occurring marginal entropy terms (see the Appendix). This "unrolled" representation is useful as it allows to compute $S_{\text{MIE}}^{(n)}$ in an iterative way where each marginal entropy contribution $S_k(\mathcal{U}_k)$ occurs only once throughout the calculation. For $n = m$, Eq. (9) further shows that the equality of Eq. (6) holds trivially just because all S_k with $k < m$ cancel, i.e., Eq. (6) indeed reads $S_m(\mathcal{A}_m) = S_m(\mathcal{A}_m)$ after cancellation.

B. MIE for a multivariate Gaussian distribution

We limit our analysis to a particular case for which we have semi-analytical access to all involved integrals. This allows us to accurately estimate the error due to MIE since the integrals can easily be computed to arbitrary precision. We assume that the m dofs are distributed according to a multivariate Gaussian distribution with zero mean and positive-definite covariance

matrix Σ. The joint probability density function (pdf) is given by

$$\rho(\mathcal{A}_m) \equiv \frac{1}{(2\pi)^{m/2} [\det \Sigma]^{1/2}} \exp\left(-\frac{1}{2} x^t \Sigma^{-1} x\right), \qquad (10)$$

where the superscript $(.)^t$ denotes transposition and det denotes the determinant. The associated entropy of the distribution reads

$$S = S_m(\mathcal{A}_m) = \frac{k_B}{2} \log \det \Sigma + m\eta \qquad (11)$$

with $\eta \equiv k_B \log (2\pi e)/2 \approx 1.419 k_B$.[42] The multivariate Gaussian distribution has the desirable property that the marginal pdf of any subset $\mathcal{U}_k$ of dofs is also a multivariate Gaussian with $(k \times k)$-dimensional covariance matrix Σ_{U_k}. Σ_{U_k} is readily generated from Σ by deleting all rows and columns which correspond to dofs not included in $\mathcal{U}_k$.[43] Consequently, the associated marginal entropy $S_k(\mathcal{U}_k)$ for $\mathcal{U}_k \subset \mathcal{A}_m$ simply reads

$$S_k(\mathcal{U}_k) = \frac{k_B}{2} \log \det \Sigma_{\mathcal{U}_k} + k\eta. \qquad (12)$$

Hence, for a multivariate Gaussian distribution, computing all terms of Eq. (9) reduces to the calculation of determinants of sub-matrices of Σ. This was done efficiently and accurately by using Cholesky decomposition and arbitrary-precision arithmetics. Since the number of summands contributing to $S_{MIE}^{(n)}$ grows very rapidly with m and n (where $n \ll m$), we additionally rearranged all sums as sums of adequate partitions of the data to prevent precision loss. We further made use of the fact that all contributions proportional to η cancel in I_k with $k > 1$.

It is advantageous to express the covariance matrix $\Sigma = D R D$ in terms of the correlation matrix R and the fluctuation matrix $D \equiv \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_m)$ which is a diagonal matrix with the standard-deviations $\sigma_i \equiv \text{std-dev}(x_i)$, $(i = 1, \ldots, m)$ on the main diagonal. The total entropy S then decomposes into $S = S_{fluct} + S_{corr}$ where $S_{fluct} \equiv \sum_{i=1}^m S_1(i)$ is the contribution to the entropy produced by the fluctuations of the dofs if they were all independent, and $S_{corr} \equiv \frac{k_B}{2} \log \det R \leq 0$ is a negative contribution to the entropy entirely caused by the correlations of the dofs. The first MIE contribution T_1 equals S_{fluct}, i.e., $S_{MIE}^{(1)}$ accounts only for the fluctuation part of the entropy, neglecting all correlations. This equality is exact and does not need to be tested. Unclear is, whether $S_{MIE}^{(n)} - T_1$ truncated to higher order $2 \leq n \ll m$ approaches S_{corr}. Trivial equality $S_{MIE}^{(m)} - T_1 = S_{corr}$ is found for $n = m$, as discussed above. In this work, we check whether the error $|S - S_{MIE}^{(n)}|$ decreases with the truncation order n for small values of $n \ll m$. Precisely, we generate covariance matrices Σ (see below), calculate R, and compare $S_{MIE}^{(n)} - T_1$ with S_{corr}. This tests the situation relevant for application: for a system of many dofs m, one calculates $S_{MIE}^{(n)}$ for small truncation order n, say $n = 2$ to 5, and expects that $S_{MIE}^{(n)}$ is a better approximation to S than simply S_{fluct}.

C. Anisotropic network model

We employ the anisotropic network model (ANM)[44,45] to generate covariance matrices Σ from protein structures. The model stems from the potential energy

$$\hat{V}(x) = \frac{\gamma}{2} \sum_{\substack{i,j; \\ i<j}}^N (r_{ij} - d_{ij})^2 \, \theta(r_c - d_{ij}) \qquad (13)$$

representing N beads in 3 dimensions coupled by harmonic springs if their equilibrium distance is smaller than a cutoff distance r_c. Here, γ is the coupling constant setting the energy scale, $d_{ij} = \text{dist}(O_i, O_j)$, O_i is the equilibrium positions of bead i, $\text{dist}(A, B)$ is the Euclidean distance between A and B, $r_{ij} = \text{dist}(O_i + \xi_i - O_j - \xi_j)$, ξ_i is the displacement vectors of bead i from its equilibrium positions O_i, θ is the Heaviside step function, and $x = [\xi_1^t, \xi_2^t, \ldots, \xi_N^t]^t$ denotes the $3N$ dimensional vector composed of all displacement vectors $\{\xi_i\}$. By expanding $\hat{V}$ around $x = 0$ to second order in x, one obtains the potential energy

$$V_0 = x^t \mathcal{H} x \qquad (14)$$

of the ANM in terms of the $(3N \times 3N)$-dimensional Hessian matrix

$$\mathcal{H} = \begin{pmatrix} H_{11} & H_{12} & \cdots & H_{1N} \\ H_{21} & H_{22} & \cdots & H_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{N1} & H_{N2} & \cdots & H_{NN} \end{pmatrix} \qquad (15)$$

which is composed of (3×3)-dimensional sub-matrices H_{ij}. To construct $\mathcal{H}$ from a set of given equilibrium positions $\{O_i\}$, we first initialize all sub-matrices equal to zero. For all pairs (O_i, O_j) with $i < j$ and $\text{dist}(O_i, O_j) < r_c$, we then update $H_{ii} \leftarrow H_{ii} + W$, $H_{jj} \leftarrow H_{jj} + W$, $H_{ij} \leftarrow H_{ij} - W$, and $H_{ji} \leftarrow H_{ji} - W$ where $W = (dd^t)/(d^t d)$ and $d = O_i - O_j$. The Hessian $\mathcal{H}$ possesses at least 6 eigenvalues equal to zero because Eq. (13) is invariant under translation and rotation. Additional zero eigenvalues occur if a bead is coupled to less than three other beads and/or if the positions $\{O_i\}$ show additional symmetries.

We also study a variant of the ANM referred to as the modified anisotropic network model (mANM) which is found to exhibit only short-range correlations. It is defined by $V_\varepsilon = x^t \mathcal{H}_\varepsilon x$ with

$$\mathcal{H}_\varepsilon = \mathcal{H} + \varepsilon \mathbb{1}, \qquad (16)$$

where $\mathbb{1}$ denotes the unitary matrix and $\varepsilon > 0$ is a parameter. The ANM is recovered for $\varepsilon = 0$.

Since V_ε is a quadratic form in x, the displacement vector x is multivariate Gaussian distributed with zero mean and covariance matrix,

$$\Sigma = \frac{k_B T}{\gamma} \mathcal{H}_\varepsilon^{-1}, \qquad (17)$$

where $\mathcal{H}_\varepsilon^{-1}$ denotes the (pseudo-)inverse of $\mathcal{H}_\varepsilon$ and $k_B T$ denotes the product of the Boltzmann constant and temperature. For simplicity, we set $\gamma = k_B T$. The (3×3)-dimensional sub-matrix of Σ for the coordinates of the beads i, j is denoted by Σ_{ij}.

III. RESULTS

Before we consider more realistic models of macromolecules, we first check the validity of MIE for the ANM in one dimension. To this end, we place N beads at the positions $O_i = (i, 0, 0)^t$ for $i = 1, \ldots, N$ and couple only nearest neighbors by setting $r_c = 1.5$. The Hessian has the following tridiagonal form:

$$\mathcal{H}(\alpha_1, \alpha_2) = \begin{pmatrix} 2 - \alpha_1 & -1 & 0 & \dots & & \alpha_2 \\ -1 & 2 & -1 & \ddots & & \vdots \\ 0 & -1 & 2 & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & & -1 \\ \alpha_2 & \dots & 0 & -1 & 2 - \alpha_1 \end{pmatrix} \quad (18)$$

after reduction to the rows and columns which correspond to the x-coordinates of the beads. The parameters α_1 and α_2 select for the applied boundary conditions. Open boundary conditions are established for $(\alpha_1, \alpha_2) = (1, 0)$ while periodic boundary conditions for $(\alpha_1, \alpha_2) = (0, -1)$. In both cases, $\mathcal{H}$ possesses one zero eigenvalue (due to the symmetry of V under parallel translation of all beads) which renders the MIE test impossible. Therefore, we introduce a third type of boundary conditions by setting $\alpha_1 = \alpha_2 = 0$ which breaks the symmetry and ensures $\det \mathcal{H}(0, 0) > 0$. Anyway, except for the zero eigenvalue, all three models have the same asymptotic behavior since independently of the boundary condition, $\log \det^+ \mathcal{H} = \mathcal{O}(\log(N))$ with $\det^+$ being the pseudo-determinant.

For the linear-chain model given by $\mathcal{H}(0, 0)$, Fig. 1(a) shows the MIE contributions T_2, T_3, T_4, and T_5 of Eq. (7) as a function of the number of dofs m (where m equals the number of beads in 1D). To leading order, all terms grow with a different power of m, i.e.,

$$T_k = \mathcal{O}(m^k) \quad (19)$$

for $m \to \infty$ and k fixed. Consequently, also $S_{\text{MIE}}^{(n)}$ is of order $\mathcal{O}(m^n)$ while the true entropy S is of order $\mathcal{O}(m)$. Hence, for this model, consecutive MIE approximations do not approach

S. Instead, MIE diverges since the error $|S - S_{\text{MIE}}^{(n)}| = \mathcal{O}(m^n)$ increases dramatically with n for $1 < n \ll m$. To be precise, we use the term "divergence" to describe that the error of MIE grows with n for all tractable values of $n \ll m$. We remind the reader that MIE is exact for $n = m$ (see Sec. II) which shows that the error of MIE will occasionally decrease with n for intractable large $n \sim m$. However, this crossover behavior does not matter here since n is limited to small values in practice.

The divergence of MIE for the linear-chain model can be traced back to the strong correlations between the dofs of the model. Figure 1(b) shows the correlation matrix $R = D^{-1} \mathcal{H}(0, 0)^{-1} D^{-1}$ of the model. While $\mathcal{H}(0, 0)$ is sparse since only nearest-neighbor beads are coupled, R takes non-zero values everywhere. Also distant beads are strongly correlated since correlations are "transmitted" along the chain of interacting beads. The associated correlation function $\mathcal{C}(d)$ decays asymptotically as

$$\mathcal{C}(d)/\text{const.} \sim 1 - \text{const.}\frac{d}{m} \quad (20)$$

with distance $d = |i - j|$ (see the Appendix) which only vanishes for $d = \mathcal{O}(m)$ such that each dof is correlated to $\mathcal{O}(m)$ others. This fact also emerges in the sub-linear behavior $S - \eta m = -\frac{k_B}{2} \log (m + 1)$ instead of the usual linear behavior for systems with short-range correlations.

These strong correlations give rise to the scaling behavior of Eq. (19) which can readily be illustrated for $T_2 = \sum_{\mathcal{U}_2 \subset \mathcal{A}_m} I_2(\mathcal{U}_2)$. There are $\binom{m}{2} = \mathcal{O}(m^2)$ many summands $I_2 \geq 0$. Therefore, T_2 can only be $\mathcal{O}(m)$ if each bead is independent of all but $\mathcal{O}(1)$ many beads. In the linear-chain model,
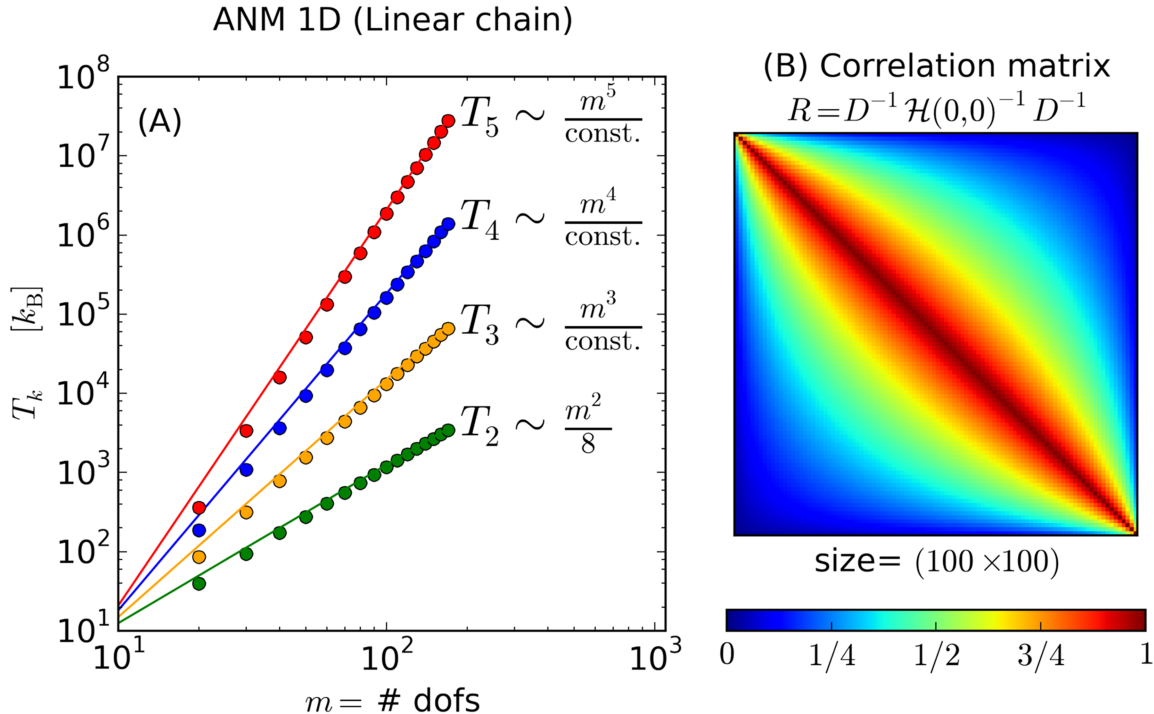

FIG. 1. (a) MIE contributions T_k ($k = 2, 3, 4, 5$) as a function of m for the ANM in 1D, i.e., the linear-chain model defined by $\mathcal{H}(0, 0)$ of Eq. (18). Asymptotically, $T_k = \mathcal{O}(m^k)$ which implies that MIE strongly diverges for this model since $|S - S_{\text{MIE}}^{(n)}| = \mathcal{O}(m^n)$. The first asymptotic, $T_2 \sim m^2/8$, is derived in the Appendix. (b) Illustration of the correlation matrix $R = D^{-1} \mathcal{H}(0, 0)^{-1} D^{-1}$ for the ANM in 1D and $m = 100$. The ANM exhibits strong correlations. All dofs are mutually correlated.

all beads are mutually correlated which leads to $T_2/k_B \sim m^2/8 + o(m^2)$ (see the Appendix). Also the higher order contributions T_k with $2 < k \leq n \ll m$ will in general only be of order $\mathcal{O}(m)$ for systems with no long-range dependencies. A proof for a general model cannot be given here because I_k is not of defined sign for $k > 2$. However, exact cancellation is unlikely to occur in general.

Summarizing, MIE truncated at order n with $1 < n \ll m$ does not serve as a useful entropy expansion for systems which exhibit long-range correlations. This represents a severe limitation for the applicability of MIE. Long-range correlations can occur for various reasons, e.g., due to thermal fluctuations close to a critical point[46] or due to insufficiently screened Coulomb interactions.[47] Moreover, long-range correlations can also be generated artificially if the system is represented inappropriately. This can happen if m dofs are represented by $M > m$ coordinates which fulfill $M - m$ constraints. The M coordinates (or subsets of them) become mutually dependent as they are coupled through the constraints. This can cause the divergence of MIE if sufficiently many coordinates are affected. Arguably the most prominent example for this situation is the use of Cartesian coordinates to describe the atom positions of macromolecules. The stiffness of the bond lengths causes dependencies between the Cartesian coordinates leading to the divergence of MIE. Furthermore, collective motion of the dofs will also cause MIE to diverge if the collective motion is inherent also in the coordinate representation. Assume an ideal gas in a container which performs oscillatory movement. The gas atoms are independent when represented in a coordinate system which is linked to the container while they are strongly correlated through the collective motion when expressed in an external coordinate system. Various types of collective motion have been observed in proteins[48–50] such as hinge bending motion where distinct subregions perform periodic relative motion or breathing modes where the total volume of the protein oscillates periodically. In the literature, however, MIE was applied to internal coordinates[51] when used to estimate molecular entropies. For such coordinates, long-range correlations caused by collective motion are vastly reduced, even below the resolution threshold for μs simulations.[52]

In the remaining, we test MIE for systems that only exhibit short-range correlations. To this end, we study the mANM of Eq. (16) which represents the ANM extended by an additional force driving all beads independently to their equilibrium positions such that long-range correlations are suppressed.

The Hessian of the mANM in 1D reads $\mathcal{H}_\varepsilon = \mathcal{H}(0, -1) + \varepsilon \mathbb{1}$ with $\varepsilon > 0$ when periodic boundary conditions are applied which reduces finite-size effects but does not alter the asymptotic behavior of the model. The associated correlation function

$$\mathcal{C}(d)/\text{const.} \sim \exp(-d/\xi) \tag{21}$$

decays exponentially at large distances with the correlation length

$$\xi = \left[\text{arcosh} \left(\frac{2+\varepsilon}{2} \right) \right]^{-1} \tag{22}$$

(see the Appendix). Hence, each bead is now correlated to only $\mathcal{O}(1)$ many beads as shown in Fig. 2(a). This meets the necessary condition for MIE established previously since for a multivariate Gaussian distribution, the absence of correlations implies independence. It further leads to the usual $\mathcal{O}(m)$ behavior $S - \eta m \sim -\frac{k_B}{2\xi} m$ (see the Appendix).

For the mANM in 1D, all MIE terms T_k are asymptotically proportional to m as shown in Fig. 2(b) where
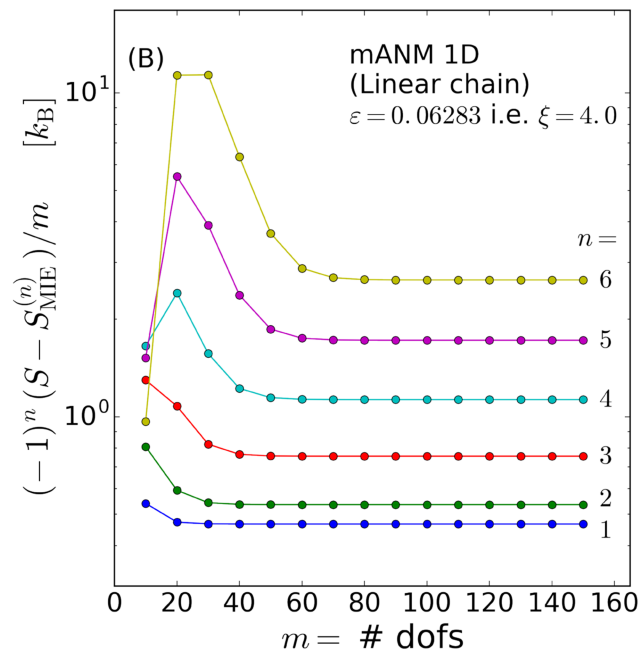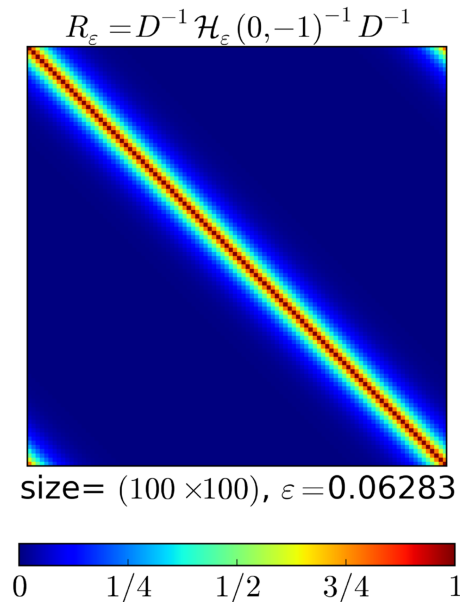
(A) Corr. matrix for mANM 1D
$R_\varepsilon = D^{-1} \mathcal{H}_\varepsilon(0,-1)^{-1} D^{-1}$

size= (100×100), $\varepsilon = 0.06283$

FIG. 2. (a) Correlation matrix R_ε for the mANM in 1D with periodic boundary conditions, defined by $\mathcal{H}_\varepsilon = \mathcal{H}(0,-1) + \varepsilon \mathbb{1}$. We choose $\varepsilon = 0.062\,83$ which corresponds to a correlation length $\xi \approx 4.0$. Each dof of the mANM is correlated to only $\mathcal{O}(m)$ many dofs which meets the necessary condition for MIE established previously. (b) Unsigned error of MIE per dof, $(-1)^n (S - S_{\text{MIE}}^{(n)})/m$, as a function of m for the same model as in (a). For the mANM, $T_k = \mathcal{O}(m)$ since the curves flatten out for $m \to \infty$. MIE also diverges for $\xi \approx 4.0$ since the error grows with n. Note the additional oscillations absorbed into the factor $(-1)^n$.
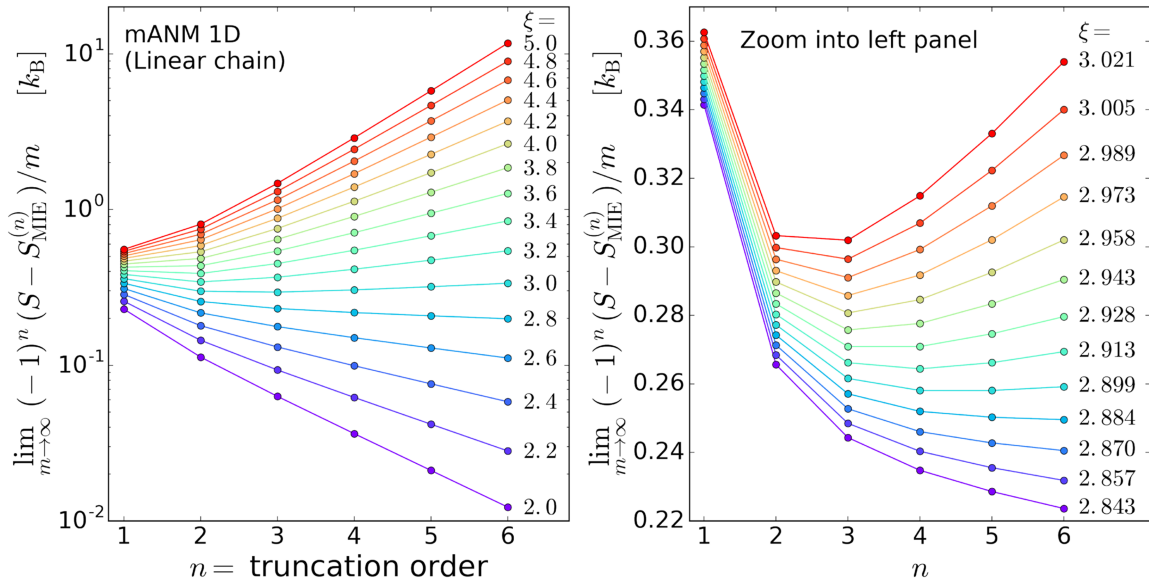
FIG. 3. The asymptotic value $\lim_{m \to \infty}(-1)^n(S - S_{\mathrm{MIE}}^{(n)})/m$ as a function of the truncation order n for various values of $\xi \in [2.0, 5.0]$. The data suggest that MIE represents an asymptotic expansion of entropy. Consecutive MIE approximations $S_{\mathrm{MIE}}^{(n)}$ approach S for truncation orders smaller than some critical value $n_0(\xi)$ while MIE diverges for larger values of n (i.e., for $n_0 < n \ll m$). $n_0(\xi) = 1$ for large correlation lengths $\xi \geq 3.6$, i.e., MIE is unable to improve the naive entropy estimate $S_{\mathrm{fluct}} = S_{\mathrm{MIE}}^{(1)} = \sum_{i=1}^{m} S_1(i)$ obtained by neglecting all correlations. For intermediate $\xi \in [2.899, 3.4]$, we find $1 < n_0 < 6$, i.e., MIE improves S_{fluct} when truncated at n with $1 < n \leq n_0$. For small $\xi \leq 2.884$, we conjecture that $n_0 \geq 6$.

$(-1)^n(S - S_{\mathrm{MIE}}^{(n)})/m$ is plotted as a function of m for $\varepsilon = 0.062\,83$ corresponding to $\xi \approx 4.0$. The different curves for $S_{\mathrm{MIE}}^{(1)}$, $S_{\mathrm{MIE}}^{(2)}$, ..., $S_{\mathrm{MIE}}^{(6)}$ flatten out for $m \gtrsim 80$ confirming $T_k = \mathcal{O}(m)$. However, reading off the asymptotic values $k_B^{-1} \lim_{m \to \infty}(S - S_{\mathrm{MIE}}^{(n)})/m$ yields -0.466, 0.535, -0.755, 1.130, -1.723, 2.645 for $m = 1, 2, 3, 4, 5, 6$, respectively. Hence, MIE still diverges since the absolute error increases with the truncation order n.

To assess the dependence of ξ on the behavior of MIE, we show in Fig. 3 the asymptotic value $\lim_{m \to \infty}(-1)^n(S - S_{\mathrm{MIE}}^{(n)})/m$ as a function of the truncation order n for various values of $\xi \in [2.0, 5.0]$. MIE diverges for large values of $\xi \geq 3.6$ while $S_{\mathrm{MIE}}^{(1)}$, $S_{\mathrm{MIE}}^{(2)}$, ..., $S_{\mathrm{MIE}}^{(6)}$ approach S for small $\xi \leq 2.884$. For intermediate values $\xi \in [2.899, 3.4]$, a crossover behavior is visible. While $S_{\mathrm{MIE}}^{(n)}$ approaches S for the first truncation orders n, it eventually diverges for larger n. These data suggest that MIE represents an asymptotic expansion of entropy for systems with correlations that decay exponentially with distance. Consecutive MIE approximations $S_{\mathrm{MIE}}^{(n)}$ approach S for $n = 1, 2, \ldots, n_0(\xi)$ up to a critical value $n_0(\xi)$ which depends on the correlation length. Beyond $n_0(\xi)$, MIE diverges instead, i.e., $|S - S_{\mathrm{MIE}}^{(n)}|$ grows with n for truncation orders n with $n_0 < n \ll m$. Our analysis for $n \leq 6$ reveals this behavior for $\xi \geq 2.899$ where $n_0(\xi) < 6$. We conjecture that MIE represents an asymptotic expansion also for $\xi < 2.899$ with a larger value $n_0 \geq 6$. Summarizing, MIE represents a useful entropy expansion for small values of ξ where $S_{\mathrm{MIE}}^{(1)}, S_{\mathrm{MIE}}^{(2)}, \ldots, S_{\mathrm{MIE}}^{(n_0)}$ approach S (up to oscillations).

Finally, we show that the previous results obtained in one dimension also hold in three dimensions. For this purpose, we constructed mANMs in 3D using 85 protein structures to define the equilibrium positions $\{O_i\}$. The selected proteins contained 50 to 53 amino acids and showed at most 30% sequence similarity. O_i's were placed at the alpha-carbon

positions of the inner 50 residues. Pairs of beads were coupled up to a cut-off distance of $r_c = 15$ Å.

Correlations decay exponentially with distance also for the mANM in 3D. Figure 4 shows the correlation function

$$\mathcal{C}_{3D}(d) = \frac{d^t \, \Sigma_{ij} \, d}{\left[(d^t \, \Sigma_{ii} \, d) \cdot (d^t \, \Sigma_{jj} \, d)\right]^{1/2}}, \tag{23}$$

where Σ_{ij} is the sub-matrix of Σ for beads i, j and $d = O_i - O_j$. No analytical expression is available for the dependence
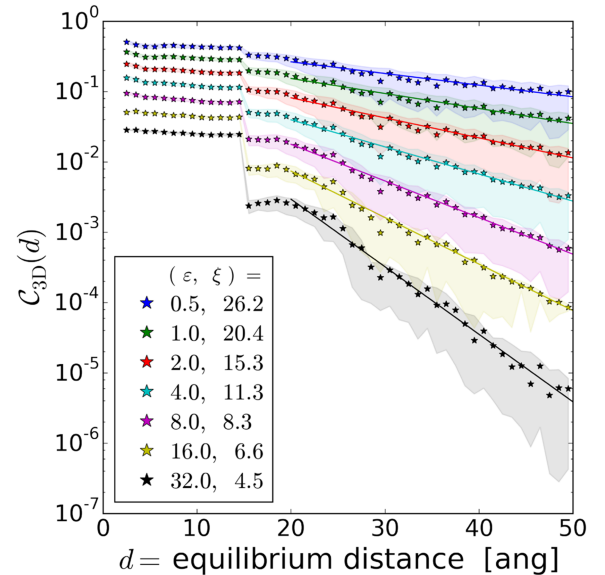

FIG. 4. The correlation function of the mANM in 3D [Eq. (23)] for various values of $\varepsilon \in [0.5, 32]$ measured from 229 proteins (of maximal 30% sequence similarity and 50 to 59 residues). Data points represent the median of all $\mathcal{C}_{3D}$ values for pairs of beads with distance d. Shaded areas show the first and third quartiles. Data points for $d > 20$ Å were fitted to an exponential decay resulting in the correlation lengths ξ given in the legend.
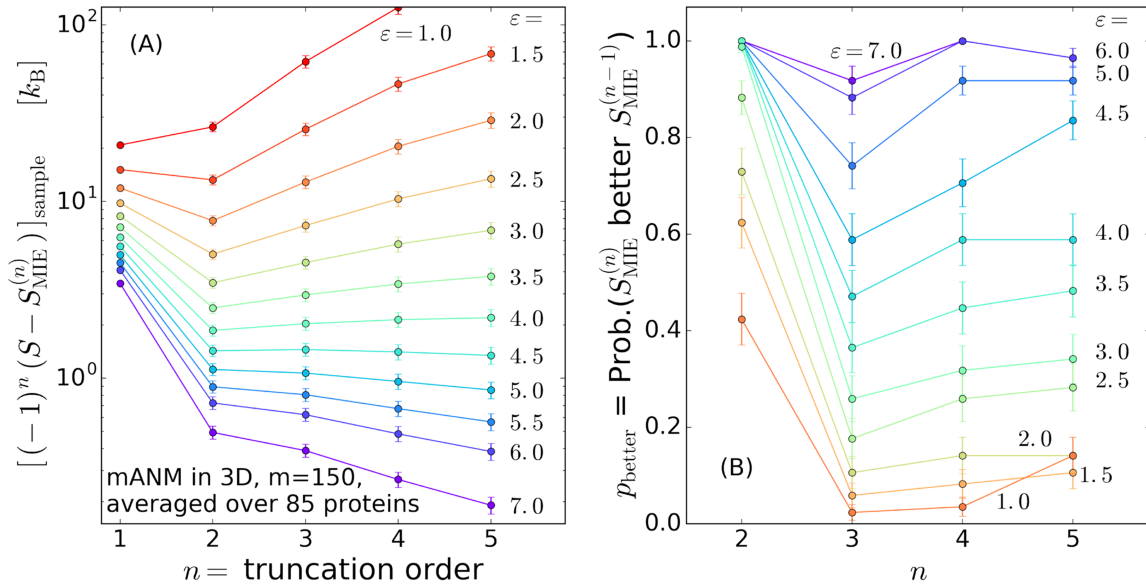
FIG. 5. (a) Performance of MIE for the mANM in 3D ($m = 150$, constructed from 85 proteins). The sample average $[(-1)^n(S - S_{\mathrm{MIE}}^{(n)})]_{\mathrm{sample}}$ shows the same qualitative behavior found for the mANM in 1D (see Fig. 3). Error bars indicate one-sigma confidence intervals in both panels. (b) Probability p_{better} that $|S - S_{\mathrm{MIE}}^{(n)}| < |S - S_{\mathrm{MIE}}^{(n-1)}|$ vs. n. Consecutive MIE approximations are likely more accurate for large values of ε and $n = 1, \ldots, 5$, while MIE should be truncated at $n = 1$ for small values of ε.

of the correlation length ξ on ε; however, $\xi(\varepsilon)$ can readily be obtained from the data. This confirmed that $\xi \to 0$ for $\varepsilon \to \infty$ as in 1D (see Fig. 4). Equation (23) represents a possible definition of the correlation function in 3D which measures the amount of correlation along $\boldsymbol{d}$. Alternative definitions have been proposed in the literature[53] and the precise $\xi(\varepsilon)$ dependence varies between definitions. Qualitatively, we obtained the same behavior for all possible definitions (not shown): the correlations decay exponentially at large distance with $\xi(\varepsilon) \to 0$ for $\varepsilon \to \infty$. Hence, the necessary condition for MIE is met since each dof depends only on $\mathcal{O}(1)$ many other dofs.

We compute MIE up to $n = 5$ for the mANM's of all proteins and various values of ε. Figure 5(a) shows the average $[(-1)^n(S - S_{\mathrm{MIE}}^{(n)})]_{\mathrm{sample}}$ vs. n where $[\ldots]_{\mathrm{sample}}$ denotes averaging over proteins. These averages show the same qualitative behavior as described previously for 1D. For large values of ε thus small correlation lengths, $S_{\mathrm{MIE}}^{(n)}$ approaches S for $n = 1$, $2, \ldots, 5$. For intermediate values, consecutive MIE approximations first approach S but finally diverge for larger orders n. For small ε or large ξ, MIE purely diverges such that $S_{\mathrm{MIE}}^{(1)}$ represents the best approximation. The sign of $S - S_{\mathrm{MIE}}^{(n)}$ does not alternate strictly with n for all proteins. However, most proteins do follow the pattern such that $[|S - S_{\mathrm{MIE}}^{(n)}|]_{\mathrm{sample}}$ behaves quantitatively very similarly to $[(-1)^n(S - S_{\mathrm{MIE}}^{(n)})]_{\mathrm{sample}}$ (not shown).

Another representation of the data is given in Fig. 5(b) where we show the probability p_{better} that $|S - S_{\mathrm{MIE}}^{(n)}| < |S - S_{\mathrm{MIE}}^{(n-1)}|$ estimated via counting statistics. $S_{\mathrm{MIE}}^{(n)}$ is likely to be a better approximation than $S_{\mathrm{MIE}}^{(n-1)}$ for small values of ξ and $n < n_0$, while MIE should be truncated at $n = 1$ for large ξ.

IV. DISCUSSION

We have tested the performance of MIE for multivariate Gaussian distributions for which the exact entropy and all MIE

contributions can be computed to arbitrary precision. We studied systems containing a large number m of dofs and truncated MIE at low orders $n = 1, 2, \ldots, 6 \ll m$. This case represents the typical situation for which MIE was proposed, namely, when m is too large for brute-force evaluation of Eq. (1) while low-dimensional integrals up to n dimensions can be computed from the available data.

We found that MIE diverges for systems which exhibit long-range correlations. This severely limits the applicability of MIE to systems in which each dof depends only on $\mathcal{O}(1)$ many other dofs. MIE can neither be applied to systems held in the vicinity of a critical point nor to systems containing unscreened long-range interactions. Moreover, MIE may also diverge for purely short-range systems when the dofs are inappropriately described in a coordinate representation which artificially generates long-range correlations. This can happen when m dofs are described by $M > m$ coordinates that satisfy $M - m$ additional constraints. In general, the M coordinates will be mutually dependent through the constraints which cause MIE to diverge. This result questions the application of MIE to macromolecules described in Cartesian coordinates (even after subtraction of roto-translation).

If the correlation function decays exponentially at large distances, our results are strong evidence that MIE represents an asymptotic expansion of entropy. $S_{\mathrm{MIE}}^{(n)}$ occasionally diverges with n for sufficiently large n, namely, for $n_0(\xi) < n \ll m$ where $n_0(\xi)$ depends on the correlation length ξ of the system. We found that $n_0 = 1$ for large ξ such that MIE diverges for all n. For small ξ, $n_0 > 1$ such that the first consecutive MIE approximations approach the correct result. For those cases, MIE represents a useful entropy expansion when truncated at order $n \leq n_0$.

Arguably the most important application of MIE is for macromolecules represented with internal coordinates. Our approach cannot be generalized to cover this case since internal

coordinates do not follow a multivariate Gaussian distribution. To our knowledge, it is unknown how the correlations of these coordinates decay with distance; it is even not obvious which distance definition should be used. If the associated correlation function decays sub-exponentially at a large distance then our results do not carry over to internal coordinates. However, sub-exponential decay is extremely unusual. Realistic systems generally exhibit correlations which asymptotically decay exponentially with distance. Assuming that such behavior is also found for internal coordinates then it is natural to believe that MIE represents an asymptotic expansion of entropy also for internal coordinates. The value of the associated correlation length needs further investigation, but it is probably rather small.[52] For these reasons, we expect that MIE serves as a useful entropy expansion for macromolecules at physiological conditions when described in internal coordinates.

APPENDIX: SUPPORTING CALCULATIONS

Equation (9) follows from the equality

$$S_{MIE}^{(n)} = \sum_{k=1}^{n} \lambda(m,n,k) \sum_{\mathcal{U}_k \subset \mathcal{A}_m} S_k(\mathcal{U}_k) \quad (A1)$$

with

$$\lambda(m,n,k) \equiv \sum_{j=0}^{n-k} (-1)^j \binom{m-k}{j} \quad (A2)$$

established in Ref. 37 since

$$\lambda(m,n,k) = (-1)^{n-k} \binom{m-k-1}{n-k}, \quad (A3)$$

where $\binom{\cdot}{\cdot}$ denotes the binomial coefficients generalized to negative arguments.[54] In particular, $\lambda(m,m,k) = 0$ for $0 < k < m$ and $\lambda(m,m,m) = 1$ for $m > 0$. It follows that Eq. (6) holds trivially because all marginal entropies S_k with $k < m$ cancel on the right hand side of Eq. (6).

The remaining appendix outlines various exact results for the (modified) linear-chain model defined by the $(m \times m)$-dimensional Hessian $\mathcal{H}_\varepsilon(m) = \mathcal{H}(0,0) + \varepsilon \mathbb{1}$ of tridiagonal form with $2 + \varepsilon$ on the main diagonal and -1 on both next-to-main diagonals.

The recursion relation

$$\det \mathcal{H}_\varepsilon(m) = (2+\varepsilon) \cdot \det \mathcal{H}_\varepsilon(m-1) - \det \mathcal{H}_\varepsilon(m-2) \quad (A4)$$

for the determinant of $\mathcal{H}_\varepsilon$ is readily obtained via Laplace expansion. The recursion is solved using linear algebra yielding $\log \det \Sigma = \log \det \mathcal{H}_\varepsilon^{-1} = -\log \det \mathcal{H}_\varepsilon$ with

$$\log \det \mathcal{H}_\varepsilon(m) = \begin{cases} \log(m+1) & \text{for } \varepsilon = 0 \\ m/\xi + o(m) & \text{for } \varepsilon > 0 \end{cases} \quad (A5)$$

with ξ defined in Eq. (22). This illustrates the fundamental difference of the ANM and the mANM. An analytic expression for the inverse matrix $\Sigma = \mathcal{H}_\varepsilon^{-1}$ is available,[55]

$$\Sigma_{i,j} = \begin{cases} \dfrac{(i+j-|i-j|)(2m+2-|i-j|-i-j)}{4(m+1)} & \text{for } \varepsilon = 0 \\ \dfrac{\cosh[(m+1-|j-i|)/\xi] - \cosh[(m+1-i-j)/\xi]}{2\sinh(1/\xi) \cdot \sinh[(m+1)/\xi]} & \text{for } \varepsilon > 0 \end{cases} \quad (A6)$$

from which the asymptotic behavior of the correlation function [Eqs. (20) and (21)] can be read off. Here, $\Sigma_{i,j}$ denotes the element (i,j) of Σ. This should not be confused with Σ_{ij} of the main text which denotes a (3×3)-dimensional sub-matrix for the 3D case. By using Eq. (A6) and dropping all sub-dominant terms, one can express the asymptotic behavior of

$$T_2 = \sum_{i<j} I_2(i,j) = -\frac{k_B}{2} \sum_{i<j} \log\left(1 - \frac{\Sigma_{i,j}^2}{\Sigma_{i,i}\Sigma_{j,j}}\right) \quad (A7)$$

for $\varepsilon = 0$ in terms of the Riemann integral

$$T_2 = -\frac{k_B}{2}m^2 \int_0^1 dx \int_x^1 dy \log\left(\frac{x-y}{y(x-1)}\right) + o(m^2) \quad (A8)$$

over a right triangle of length unity which can be solved elementarily.[56] In this way, we obtained $T_2 = \frac{k_B}{8}m^2 + o(m^2)$ shown in Fig. 1. To prove $T_k = \mathcal{O}(m^k)$ [i.e., Eq. (19)] for $k > 2$, one can proceed in an analogous way. With Eq. (A6), a Riemann integral representation for T_k is found which is multiplied by a factor m^k. Showing that the integral is finite can be

done (without explicitly solving it) by bounding the integral appropriately.[57]

[1] Nobel Media AB, "The Nobel Prize in Chemistry 2013," www.nobelprize.org (2013), Awarded jointly to Martin Karplus, Michael Levitt, and Arieh Warshel.

[2] C. N. Pace, "Contribution of the hydrophobic effect to globular protein stability," J. Mol. Biol. **226**, 29–35 (1992).

[3] L. Sawle and K. Ghosh, "How do thermophilic proteins and proteomes withstand high temperature?," Biophys. J. **101**, 217–227 (2011).

[4] J. D. Chodera and D. L. Mobley, "Entropy-enthalpy compensation: Role and ramifications in biomolecular ligand recognition and design," Annu. Rev. Biophys. **42**, 121–142 (2013).

[5] J.-L. Popot and D. M. Engelman, "Membrane protein folding and oligomerization: The two-stage model," Biochemistry **29**, 4031–4037 (1990).

[6] A. T. D. Poian, A. C. Oliveira, and J. L. Silva, "Cold denaturation of an icosahedral virus. The role of entropy in virus assembly," Biochemistry **34**, 2672–2677 (1995).

[7] B. Ma, C.-J. Tsai, and R. Nussinov, "A systematic study of the vibrational free energies of polypeptides in folded and random states," Biophys. J. **79**, 2739–2753 (2000).

[8] I. N. Berezovsky, W. W. Chen, P. J. Choi, and E. I. Shakhnovich, "Entropic stabilization of proteins and its proteomic consequences," PLoS Comput. Biol. **1**, 322–332 (2005).

[9] J. Zhang and J. S. Liu, "On side-chain conformational entropy of proteins," PLoS Comput. Biol. **2**, 1586–1591 (2006).

[10]C.-E. Chang, W. Chen, and M. K. Gilson, "Ligand configurational entropy and protein binding," Proc. Natl. Acad. Sci. U. S. A. **104**, 1534–1539 (2007).

[11]B. J. Killian, J. Y. Kravitz, S. Somani, P. Dasgupta, Y.-P. Pang, and M. K. Gilson, "Configurational entropy in protein–peptide binding: Computational study of Tsg101 ubiquitin E2 variant domain with an HIV-derived PTAP nonapeptide ," J. Mol. Biol. **389**, 315–335 (2009).

[12]D. Sciretti, P. Bruscolini, A. Pelizzola, M. Pretti, and A. Jaramillo, "Computational protein design with side-chain conformational entropy," Proteins: Struct., Funct., Bioinf. **74**, 176–191 (2009).

[13]E. Balog, D. Perahia, J. C. Smith, and F. Merzel, "Vibrational softening of a protein on ligand binding," J. Phys. Chem. B **115**, 6811–6817 (2011).

[14]A. J. Wand, "The dark energy of proteins comes to light: Conformational entropy and its role in protein function revealed by NMR relaxation," Curr. Opin. Struct. Biol. **23**, 75–81 (2013).

[15]M. Rossi, M. Scheffler, and V. Blum, "Impact of vibrational entropy on the stability of unsolvated peptide helices with increasing length," J. Phys. Chem. B **117**, 5574–5584 (2013).

[16]M. Goethe, I. Fita, and J. M. Rubi, "Vibrational entropy of a protein: Large differences between distinct conformations," J. Chem. Theory Comput. **11**, 351–359 (2015).

[17]P. Tompa and A. Fersht, *Structure and Function of Intrinsically Disordered Proteins* (CRC Press, 2009).

[18]R. I. Cukier, "Dihedral angle entropy measures for intrinsically disordered proteins," J. Phys. Chem. B **119**, 3621–3634 (2015).

[19]A. Baruah, P. Rani, and P. Biswas, "Conformational entropy of intrinsically disordered proteins from amino acid triads," Sci. Rep. **5**, 11740 (2015).

[20]R. Bellman, *Adaptive Control Processes: A Guided Tour* (Princeton University Press, 1961), available at http://www.jstor.org/stable/j.ctt183ph6v.

[21]D. Suárez and N. Díaz, "Direct methods for computing single-molecule entropies from molecular simulations," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **5**, 1 (2015).

[22]M. Karplus and J. N. Kushick, "Method for estimating the configurational entropy of macromolecules," Macromolecules **14**, 325–332 (1981).

[23]O. Edholm and H. J. C. Berendsen, "Entropy estimation from simulations of non-diffusive systems," Mol. Phys. **51**, 1011–1028 (1984).

[24]J. Schlitter, "Estimation of absolute and relative entropies of macromolecules using the covariance matrix," Chem. Phys. Lett. **215**, 617 (1993).

[25]I. Andricioaei and M. Karplus, "On the calculation of entropy from covariance matrices of the atomic fluctuations," J. Chem. Phys. **115**, 6289–6292 (2001).

[26]D.-W. Li and R. Brüschweiler, "*In silico* relationship between configurational entropy and soft degrees of freedom in proteins and peptides," Phys. Rev. Lett. **102**, 118108 (2009).

[27]G. Gyimesi, P. Závodszky, and A. Szilágyi, "Calculation of configurational entropy differences from conformational ensembles using Gaussian mixtures," J. Chem. Theory Comput. **13**, 29–41 (2017).

[28]C.-E. Chang, W. Chen, and M. K. Gilson, "Evaluating the accuracy of the quasiharmonic approximation," J. Chem. Theory Comput. **1**, 1017–1028 (2005).

[29]O. F. Lange and H. Grubmüller, "Full correlation analysis of conformational protein dynamics," Proteins: Struct., Funct., Bioinf. **70**, 1294–1312 (2008).

[30]V. Hnizdo, E. Darian, A. Fedorowicz, E. Demchuk, S. Li, and H. Singh, "Nearest-neighbor nonparametric method for estimating the configurational entropy of complex molecules," J. Comput. Chem. **28**, 655–668 (2007).

[31]B. M. King and B. Tidor, "MIST: Maximum information spanning trees for dimension reduction of biological data sets," Bioinformatics **25**, 1165–1172 (2009).

[32]B. M. King, N. W. Silver, and B. Tidor, "Efficient calculation of molecular configurational entropies using an information theoretic approximation," J. Phys. Chem. B **116**, 2891–2904 (2012).

[33]H. Matsuda, "Physical nature of higher-order mutual information: Intrinsic correlations and frustration," Phys. Rev. E **62**, 3096–3102 (2000).

[34]B. J. Killian, J. Y. Kravitz, and M. K. Gilson, "Extraction of configurational entropy from molecular simulations via an expansion approximation," J. Chem. Phys. **127**, 024107 (2007).

[35]V. Hnizdo, J. Tan, B. J. Killian, and M. K. Gilson, "Efficient calculation of configurational entropy from molecular simulations by combining the mutual-information expansion and nearest-neighbor methods," J. Comput. Chem. **29**, 1605–1614 (2008).

[36]U. Hensen, O. F. Lange, and H. Grubmüller, "Estimating absolute configurational entropies of macromolecules: The minimally coupled subspace approach," PLoS One **5**, e9179 (2010).

[37]E. Suárez, N. Díaz, and D. D. Suárez, "Entropy calculations of single molecules by combining the rigid-rotor and harmonic-oscillator approximations with conformational entropy estimations from molecular dynamics simulations," J. Chem. Theory Comput. **7**, 2638–2653 (2011).

[38]A. T. Fenley, H. S. Muddana, and M. K. Gilson, "Entropy–enthalpy transduction caused by conformational shifts can obscure the forces driving protein–ligand binding," Proc. Natl. Acad. Sci. U. S. A. **109**, 20006–20011 (2012).

[39]E. Suárez, N. Díaz, J. Méndez, and D. Suárez, "CENCALC: A computational tool for conformational entropy calculations from molecular simulations," J. Comput. Chem. **34**, 2041–2054 (2013).

[40]A. T. Fenley, B. J. Killian, V. Hnizdo, A. Fedorowicz, D. S. Sharp, and M. K. Gilson, "Correlation as a determinant of configurational entropy in supramolecular and protein systems," J. Phys. Chem. B **118**, 6447–6455 (2014).

[41]M. Fleck, A. A. Polyansky, and B. Zagrovic, "PARENT: A parallel software suite for the calculation of configurational entropy in biomolecular systems," J. Chem. Theory Comput. **12**, 2055–2065 (2016).

[42]T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley, 2006).

[43]Y. L. Tong, *The Multivariate Normal Distribution* (Springer-Verlag, Berlin, 1990).

[44]A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, "Anisotropy of fluctuation dynamics of proteins with an elastic network model," Biophys. J. **80**, 505–515 (2001).

[45]H. Lei, X. Yang, B. Zheng, G. Lin, and N. A. Baker, "Constructing surrogate models of complex systems with enhanced sparsity: Quantifying the influence of conformational uncertainty in biomolecular solvation," Multiscale Model. Simul. **13**, 1327–1353 (2015).

[46]N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group* (Addison-Wesley Publishing Company, Boston, 2002).

[47]A. Hunt, "Approximate thermodynamical treatment of the Coulomb gap," Philos. Mag. Lett. **62**, 371 (1990).

[48]J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids* (Cambridge University Press, Cambridge, 1987).

[49]T. Ichiye and M. Karplus, "Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations," Proteins: Struct., Funct., Bioinf. **11**, 205–217 (1991).

[50]K. Henzler-Wildman and D. Kern, "Dynamic personalities of proteins," Nature **450**, 964–972 (2007).

[51]V. Hnizdo and M. K. Gilson, "Thermodynamic and differential entropy under a change of variables," Entropy **12**, 578–590 (2010).

[52]D.-W. Li, D. Meng, and R. Brüschweiler, "Short–range coherence of internal protein dynamics revealed by high-precision *in silico* study," J. Am. Chem. Soc. **131**, 14610–14611 (2009).

[53]L. Meinhold and J. C. Smith, "Fluctuations and correlations in crystalline protein dynamics: A simulation analysis of staphylococcal nuclease," Biophys. J. **88**, 2554 (2005).

[54]M. J. Kronenburg, "The binomial coefficient for negative arguments," e-print arXiv:1105.3689 (2011).

[55]G. Y. Hu and R. F. O'Connell, "Analytical inversion of symmetric tridiagonal matrices," J. Phys. A: Math. Gen. **29**, 1511 (1996).

[56]We use $\int dy \log\left(\frac{x-y}{y(x-1)}\right) = x\log(x/y) + (y - x)\log\left(\frac{x-y}{y(x-1)}\right)$ and $\int dx\, x \log(x) = x^2(2\log(x) - 1)/4$.

[57]For T_2, one may use $\int_0^1 dx \int_x^1 dy \log\left(\frac{x-y}{y(x-1)}\right) \leq \int_0^1 dx \int_x^1 dy \left(\frac{x-y}{y(x-1)} - 1\right) = \frac{3}{2} - \frac{\pi^2}{6} \approx -0.145$.