A generalized free energy perturbation theory accounting for end states with differing configuration space volume – Supporting Information –

R. Thomas Ullmann^{*} and G. Matthias Ullmann[†] Structural Biology/Bioinformatics, University of Bayreuth, Universitätsstrasse 30, BGI, 95447 Bayreuth, Germany (Dated: December 16, 2010)

^{*}Author to whom correspondence should be addressed. Electronic mail: thomas.ullmann@uni-bayreuth.de †E-mail: matthias.ullmann@uni-bayreuth.de; Website: www.bisb.uni-bayreuth.de

CONTENTS

A.	The classical particle in the box with a single-well or a double-well harmonic potential	3
	1. Some useful mathematical relations	4
	2. Description of the system	4
	3. Partition functions and probability distributions	6
	a. Zustandssumme and configuration integral in classical statistical thermodynamics	6
	b. The particle in the box with a single-well harmonic potential	9
	c. The particle in the box with a double-well harmonic potential	10
	4. Effective phase space and configuration space volume	13
	a. The particle in the box with a single-well harmonic potential	13
	b. The particle in the box with a double-well harmonic potential	16
	5. Extension of the formalism to multiple non-interacting particles	18
	6. Pressure	19
	7. Correction terms for the free energy perturbation schemes	20
	a. Definition of an external particle reservoir	22
	b. Model transformations	23
B.	Efficiency of FEP simulations - bias and statistical uncertainty	30
	1. Staging	31
	2. The Bennett acceptance ratio method	32
C.	Relation between Boltzmann entropy and Süßmann entropy	33
	References	36

quantity	harmonic potential type		
	single-well	double-well	
configurational partition function (configuration integral)	$Q^{\rm sw} = \frac{1}{N!} \left(\sqrt{\frac{2\pi}{\beta m \omega^2}} \operatorname{Erf}\left[\sqrt{\frac{\beta m \omega^2}{2}} \frac{L}{2} \right] \right)^N$	$Q^{\mathrm{dw}} = \frac{1}{N!} \left(\sqrt{\frac{8\pi}{\beta m \omega^2}} \operatorname{Erf}\left[\sqrt{\frac{\beta m \omega^2}{2}} \frac{L}{4} \right] \right)^N$	
	in SI units of $[m^{1,1}]$		
Zustandssumme (phase integral, partition function)	$Z^{\rm sw} = \frac{1}{N!} \left(\frac{2\pi}{\beta\omega} \operatorname{Erf}\left[\sqrt{\frac{\beta m \omega^2}{2}} \frac{L}{2} \right] \right)^N$	$Z^{\mathrm{dw}} = \frac{1}{N!} \left(\frac{4\pi}{\beta\omega} \mathrm{Erf}\left[\sqrt{\frac{\beta m \omega^2}{2}} \frac{L}{4} \right] \right)^N$	
	in SI units of $[(Js)^N]$		
effective configuration space volume	$\tilde{\Omega}^{\mathrm{sw},\mathbf{q}} = \frac{1}{N!} \left(\sqrt{\frac{4\pi}{\beta m \omega^2}} \frac{\left(\mathrm{Erf}\left[\sqrt{\frac{\beta m \omega^2}{2}} \frac{L}{2} \right] \right)^2}{\mathrm{Erf}\left[\sqrt{\beta m \omega^2} \frac{L}{2} \right]} \right)^N$	$\tilde{\Omega}^{\mathrm{dw},\mathbf{q}} = \frac{1}{N!} \left(\sqrt{\frac{16\pi}{\beta m \omega^2}} \frac{\left(\mathrm{Erf}\left[\sqrt{\frac{\beta m \omega^2}{2}} \frac{L}{4} \right] \right)^2}{\mathrm{Erf}\left[\sqrt{\beta m \omega^2} \frac{L}{4} \right]} \right)^N$	
	in SI units of $[\mathbf{m}^N]$		
effective phase space volume	$\tilde{\Omega}^{\mathrm{sw},\mathbf{p},\mathbf{q}} = \frac{1}{N!} \left(\frac{4\pi}{\beta\omega} \frac{\left(\mathrm{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}} \frac{L}{2} \right] \right)^2}{\mathrm{Erf}\left[\sqrt{\beta m\omega^2} \frac{L}{2} \right]} \right)^N$	$\tilde{\Omega}^{\mathrm{dw},\mathbf{p},\mathbf{q}} = \frac{1}{N!} \left(\frac{8\pi}{\beta\omega} \frac{\left(\mathrm{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}} \frac{L}{4} \right] \right)^2}{\mathrm{Erf}\left[\sqrt{\beta m\omega^2} \frac{L}{4} \right]} \right)^N$	
	in SI units of $\left[(J s)^N \right]$		

Appendix A: The classical particle in the box with a single-well or a double-well harmonic potential

TABLE I. Compilation of formulas for the description of the particle in the box models presented in the example section of the main text. The particles are identical, non-interacting (ideal gas) and indistinguishable. N is the number of particles, ω is the angular frequency, m is the particle mass, L is the box length, $\beta^{-1} = k_B T$ and Erf is the error function (see Eq. (A.8)).

1. Some useful mathematical relations

The listed relations can be found in standard math tables.

L

x-

$$\int_{-\infty}^{\infty} \exp\left[-ax^2\right] \, \mathrm{d}x = \sqrt{\frac{\pi}{a}} \tag{A.1}$$

$$\int_{-\frac{L}{2}}^{\frac{\pi}{2}} \exp\left[-ax^{2}\right] \, \mathrm{d}x = \sqrt{\frac{\pi}{a}} \mathrm{Erf}\left[\sqrt{a}\frac{L}{2}\right] \tag{A.2}$$

$$\int_{-\frac{L}{4}}^{\frac{T}{4}} \exp\left[-ax^{2}\right] \, \mathrm{d}x = \sqrt{\frac{\pi}{a}} \mathrm{Erf}\left[\sqrt{a}\frac{L}{4}\right] \tag{A.3}$$

$$\lim_{x \to \infty} \operatorname{Erf}(x) = 1 \tag{A.4}$$

$$\lim_{x \to 0} \operatorname{Erf}(x) = \frac{2x}{\sqrt{\pi}}$$
(A.5)

$$\frac{\mathrm{d}\ln\left[ax\right]}{\mathrm{d}x} = \frac{1}{x} \tag{A.6}$$

$$\frac{\mathrm{d}\ln\left[a\mathrm{Erf}\left[bx\right]\right]}{\mathrm{d}x} = \frac{2b\exp\left[-b^2x^2\right]}{\sqrt{\pi}\mathrm{Erf}\left[bx\right]} \tag{A.7}$$

where

$$\operatorname{Erf}\left(x\right) = \frac{2}{\pi} \int_{0}^{x} \exp\left[-t^{2}\right] \mathrm{d}t \tag{A.8}$$

is the error function.

2. Description of the system

This model example demonstrates the application of our formalism to a dynamic system formulated in non-dimensionless, continuous coordinates. The model system consists of a particle of mass m in a one-dimensional box of length L. The spatial coordinate q is defined on the interval $-\frac{L}{2} \leq q \leq \frac{L}{2}$. The momentum p must not be confused with the occupation probability p. The potential energy is given by a single-well or double-well harmonic potential denoted by sw or dw, respectively. The Hamiltonian of the system is given by

$$H^{\rm sw} = \frac{\mathbf{p}^2}{2m} + \frac{m\omega^2 \mathbf{q}^2}{2} \tag{A.9}$$

$$H^{\rm dw} = \frac{\mathbf{p}^2}{2m} + \begin{cases} \frac{m\omega^2}{2} \left(\mathbf{q} + \frac{L}{4}\right)^2 \text{ if } \mathbf{q} \le 0\\ \frac{m\omega^2}{2} \left(\mathbf{q} - \frac{L}{4}\right)^2 \text{ if } \mathbf{q} > 0 \end{cases}$$
(A.10)



FIG. 1. Potential energy and probability distribution of the particle in a box models with a single-well or a double-well harmonic potential. The particle mass is set to $m = 2 \times 10^{-26}$ [kg]. The box length of the end state is $L = 1.6 \times 10^{-9}$ [m] and the angular frequency is $\omega = 1.6 \times 10^{12}$ [s⁻¹]. (a) single-well (black curve) or double-well (red curve) harmonic potential as function of the configuration **q** (b) configuration space probability distribution for the position of a single particle in a box with a single-well (black curve) or double-well (red curve) harmonic potential (c) phase space probability distribution for the position of a single particle in a box with a single-well harmonic potential (d) phase space probability distribution for the position of a single particle in a box with a double-well harmonic potential The regions of phase space with a probability density higher than an 31.35 [(Js)⁻¹] are projected into the phase space plane and contoured bold black to schematically show the effectively populated regions of phase space whose volume is given by $\tilde{\Omega}$. Please note that this cutoff value for the probability density is arbitrary and that the effective phase space volume, defined by Eq. (38), does actually not correspond to a region of phase space with sharp boundaries since the probability density is non-zero throughout the whole phase space region on which the model is defined.

where the kinetic energy is given by the first terms on the right hand sides, whereas the potential energy is given by the second terms on the right hand sides. In a classical formulation as adopted here p and q are independent. The strength of the harmonic potential is determined by the angular frequency ω of the oscillator. The potential energy for the single-well and double-well models are plotted in Fig. 1(a) as function of the spatial coordinate q. In the following sections, we will derive the necessary formalism to treat the particle in the box models. The derivations are given in full detail to demonstrate the applicability of our FEP formalism to models in continuous, non-dimensionless coordinates. The most important equations used for the particle in the box example of the main text are compiled in Table I for quick reference.

In case of the single-well harmonic potential and an infinitely large box, the model is equivalent to the well known classical formulation of the one-dimensional harmonic oscillator. The models can readily be extended to the case of multiple non-interacting (ideal gas) particles. In this case the model with a single-well potential and an infinitely large box can also be interpreted as a collection of independent, classical harmonic oscillators.

3. Partition functions and probability distributions

We begin this section, with a short introduction of partition functions in general and their usefulness in (classical) statistical thermodynamics. In the main part of this section, we derive analytical expressions for the classical partition functions and probability density distributions of the harmonic oscillator models introduced in the previous section.

a. Zustandssumme and configuration integral in classical statistical thermodynamics

The Zustandssumme Z is given by the integral of the Boltzmann factor $\exp[-\beta H]$ over the whole phase space known as phase integral

$$Z = \int \exp\left[-\beta H\left(\mathbf{p}, \mathbf{q}\right)\right] \,\mathrm{d}\mathbf{p} \,\mathrm{d}\mathbf{q} \tag{A.11}$$

where the microstate of the system is given by a particular set of values for \mathbf{p} and \mathbf{q} . For classical systems, H is separable in a kinetic energy term E^{kin} depending only on \mathbf{p} and a potential energy term E^{pot} depending only on \mathbf{q} , and thus the phase integral can be expressed as product of two

independent integrals termed momentum integral and configuration integral

$$Z = \int \exp\left[-\beta E^{kin}\left(\mathbf{p}\right)\right] \, \mathrm{d}\mathbf{p} \int \exp\left[-\beta E^{pot}\left(\mathbf{q}\right)\right] \, \mathrm{d}\mathbf{q}$$
(A.12)

Typically, the range of momentum values is not restricted. Thus, the momentum integral is typically carried out over the interval $-\infty \leq \mathbf{p} \leq \infty$ yielding a constant factor $\sqrt{2\pi m\beta^{-1}}^{DN}$ where N is the number of particles in the system and D is the dimensionality of the system, that would be D = 1 for our one-dimensional box. Since we formulated our model in continuous coordinates, the ratio of the Boltzmann factor and the Zustandssumme denotes a probability density

$$\rho^{\mathbf{p},\mathbf{q}} = \frac{\exp\left[-\beta H\right]}{Z} \tag{A.13}$$

which upon integration over a certain region of phase space yields the equilibrium probability, for the occupation of this phase space region by the system

$$p = \int_{\substack{\mathbf{p}, \mathbf{q} \\ \mathbf{q}_{start} \\ \mathbf{q}_{start}}}^{\mathbf{p}_{end}} \rho^{\mathbf{p}, \mathbf{q}} \, \mathrm{d}\mathbf{p} \, \mathrm{d}\mathbf{q}$$
(A.14)

Often, we are not interested in the full phase space probability distribution of the system but only in the probability distribution of the system configuration, which can be obtained by integrating the probability density at each configuration over the entire range of momentum values

$$\rho^{\mathbf{q}} = \frac{1}{Z} \int_{\mathbf{p}=-\infty}^{\mathbf{p}=\infty} \exp\left[-\beta H\right] d\mathbf{p}$$

$$= \frac{\int_{\mathbf{p}=-\infty}^{\mathbf{p}=\infty} \exp\left[-\beta E^{kin}\left(\mathbf{p}\right)\right] d\mathbf{p} \quad \exp\left[-\beta E^{pot}\left(\mathbf{q}\right)\right]}{\int_{\mathbf{p}=-\infty}^{\mathbf{p}=-\infty} \exp\left[-\beta E^{kin}\left(\mathbf{p}\right)\right] d\mathbf{p} \int \exp\left[-\beta E^{pot}\left(\mathbf{q}\right)\right] d\mathbf{q}}$$
(A.15)

where we see that the momentum integral cancels because it occurs in numerator and denominator. Thus, the probability distribution of the configuration is given by

$$\rho^{\mathbf{q}} = \frac{\exp\left[-\beta E^{pot}\right]}{Q} \tag{A.16}$$

where the configuration integral is given by the integral

$$Q = \int \exp\left[-\beta E^{pot}\right] \,\mathrm{d}\mathbf{q} \tag{A.17}$$

taken over the whole configuration space. Due to its role in determining ρ^{q} , Q is also called configurational partition function. As for the phase space probability density, a probability to find the system in a given region of the configuration space can be obtained by integrating ρ^{q} over this region

$$p = \int_{\mathbf{q}_{start}}^{\mathbf{q}_{end}} \rho^{\mathbf{q}} \, \mathrm{d}\mathbf{q} \tag{A.18}$$

Often, the configurational partition function is partitioned further in a product of partition functions each accounting for different energy terms and/or degrees of freedom, where the configuration **q** can be expressed in terms of different coordinate systems. The separability of the configuration integral, e.g., into translational, rotational and vibrational partition functions does involve the approximation that these motions are uncorrelated [1, 2]. Derived coordinate systems, such as internal coordinates, are not necessarily mutually independent (or orthogonal) and do not preserve the units of the initial, e.g., cartesian coordinates. Jacobians must be used for the back-transformation into the desired coordinate system [3–7]. For some example applications of Jacobians to molecular simulations in dihedral angle space see Refs. [8–10].

In the literature, often a factor $\mathcal{N} = \frac{1}{h^{DN}}$ is added to Eq. (A.11) to match the high temperature limit of the corresponding quantum mechanical solution. The factor is omitted here, because it is of no relevance to our purely classical treatment and cancels from all derived expressions like free energy differences [1, 2]. Likewise, it would also be necessary to add the factor \mathcal{N} to the Boltzmann factor in the numerator of Eq. A.13 to yield the normalization of the probability density to the prescribed configuration space volume (characterized by finite potential energy) as expected from a classical system. For further information on the meaning of \mathcal{N} and possible interpretations, the reader is referred to standard statistical mechanics textbooks [1, 11]. Also Refs. [2, 12, 13] might be helpful.

If the system can adopt several indistinguishable configurations, an additional factor needs to be added to the partition functions Eqs. (A.11) and (A.17) and likewise to the Boltzmann factor of such degenerate microstates in Eqs. (A.13), (A.15) and (A.16) to correct for this reduction of the entropy. If for example, the system consists of N, indistinguishable particles, a prefactor of $\frac{1}{N!}$ would have to be added to both partition functions and the Boltzmann factors of all microstates and configurations in Eqs. (A.13), (A.15) and (A.16).

Besides determining the microstate probability distributions, the partition functions or their derivatives with respect to some parameter permit the calculation of all thermodynamic state func-

tions of the system, for instance free energy, internal energy, enthalpy, entropy, chemical potential and pressure. This is exemplified at the calculation of the pressure in Sec. A 6.

b. The particle in the box with a single-well harmonic potential

The Hamiltonian is given by Eq. A.9. Consequently, the partition function is given by

$$Z^{\rm sw} = \int_{-\infty}^{\infty} \exp\left[-\frac{\beta \mathbf{p}^2}{2m}\right] \,\mathrm{d}\mathbf{p} \int_{-\frac{L}{2}}^{\frac{\pi}{2}} \exp\left[-\frac{\beta m\omega^2 \mathbf{q}^2}{2}\right] \,\mathrm{d}\mathbf{q} \tag{A.19}$$

The configuration integral can be solved using Eq. (A.2) giving the configurational partition function

$$Q^{\rm sw} = \sqrt{\frac{2\pi}{\beta m \omega^2}} \operatorname{Erf}\left[\sqrt{\frac{\beta m \omega^2}{2}} \frac{L}{2}\right] \qquad \text{in SI units of } [m] \tag{A.20}$$

The momentum integral can be solved using Eq. (A.1) giving $\sqrt{2\pi m\beta^{-1}}$ in $\left[\frac{\text{kgm}}{\text{s}}\right]$. The product of momentum and configuration integral gives the total partition function or Zustandssumme

$$Z^{\rm sw} = \frac{2\pi}{\beta\omega} \operatorname{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}} \frac{L}{2}\right] \qquad \text{in SI units of}\left[\frac{\operatorname{kg} \operatorname{m}^2}{s}\right] = [\operatorname{Js}] \qquad (A.21)$$

where Q possesses units of length [m] and Z possesses units of action [J s]. In the limit of an infinitely large box length, the well known expression for the partition function of the classical one-dimensional harmonic oscillator is recovered using Eq. (A.4)

$$\lim_{L \to \infty} Z^{\rm sw} = \frac{2\pi}{\beta\omega} \qquad \text{in SI units of } [\rm J\,s] \tag{A.22}$$

With the derived partition functions Eqs. (A.20) and (A.21), and Eqs. (A.13) and (A.16) the probability density distributions in phase space and configuration space can be written as

$$\rho^{\mathrm{sw},\mathbf{q}} = \frac{\exp\left[-\frac{\beta m\omega^{2}\mathbf{q}^{2}}{2}\right]}{\sqrt{\frac{2\pi}{\beta m\omega^{2}}} \mathrm{Erf}\left[\sqrt{\frac{\beta m\omega^{2}}{2}\frac{L}{2}}\right]} \qquad \text{in SI units of } [\mathrm{m}^{-1}] \tag{A.23}$$

$$\rho^{\mathrm{sw},\mathbf{p},\mathbf{q}} = \frac{\exp\left[-\frac{\beta\mathbf{p}^2}{2m} - \frac{\beta m\omega^2 \mathbf{q}^2}{2}\right]}{\frac{2\pi}{\beta\omega} \mathrm{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}\frac{L}{2}}\right]} \quad \text{in SI units of } \left[(\mathrm{J\,s})^{-1}\right]$$
(A.24)

In the last part of this section we will check our results for consistency with physical expectations. The simplest check concerns the proper normalization of the derived probability distributions. The integral of the probability distributions Eq. (A.23) and (A.24) over the whole configuration or phase space should give a probability of unity to find the particle inside the integrated region. This can be seen to be the case, because the integrals of the variable exponential terms in the numerator of Eqs. (A.23) and (A.24) yield again the partition functions, which also occur as constants in the denominator of these equations. Finally, we will consider the limiting case of a vanishing harmonic potential, *i.e*, $\omega \rightarrow 0$. In this case, one would expect a uniform probability distribution throughout the entire box. The corresponding limit of the configuration space probability density is found using Eq. (A.5)

$$\lim_{\omega \to 0} \rho^{\mathrm{sw},\mathbf{q}} = \frac{\sqrt{\frac{\beta m \omega^2}{2\pi}} \lim_{\omega \to 0} \exp\left[-\frac{\beta m \omega^2 \mathbf{q}^2}{2}\right]}{\lim_{\omega \to 0} \operatorname{Erf}\left[\sqrt{\frac{\beta m \omega^2}{2}}\frac{L}{2}\right]}$$

$$= \frac{\sqrt{\frac{\beta m \omega^2}{2\pi}}1}{\frac{2}{\sqrt{\pi}}\sqrt{\frac{\beta m \omega^2}{2}}\frac{L}{2}}$$

$$= \frac{\sqrt{\frac{\beta m \omega^2}{2\pi}}1}{\sqrt{\frac{\beta m \omega^2}{2\pi}}L}$$

$$= \frac{1}{L} \quad \text{in SI units of } [\mathrm{m}^{-1}]$$
(A.25)

and confirms our expectation.

The configuration and phase space probability density distributions are plotted for concrete example parameters in Fig. 1(b) and (c)

c. The particle in the box with a double-well harmonic potential

The Hamiltonian is given by Eq. A.10. Consequently, the partition function is given by

$$Z^{\mathrm{dw}} = \int_{-\infty}^{\infty} \exp\left[-\frac{\beta \mathbf{p}^2}{2m}\right] \,\mathrm{d}\mathbf{p} \left(\int_{-\frac{L}{2}}^{0} \exp\left[-\frac{\beta m\omega^2 \left(\mathbf{q} + \frac{L}{4}\right)^2}{2}\right] \,\mathrm{d}\mathbf{q} + \int_{0}^{\frac{L}{2}} \exp\left[-\frac{\beta m\omega^2 \left(\mathbf{q} - \frac{L}{4}\right)^2}{2}\right] \,\mathrm{d}\mathbf{q}\right)$$
(A.26)

where we need to perform a piecewise integration to obtain the configuration integral, because of the piecewise definition of the potential energy. The configuration integral can be solved using Eq. (A.3) with a variable substitution setting $\mathbf{q}' = \mathbf{q} \pm \frac{L}{4}$. After re-substitution we obtain

$$Q^{\rm dw} = \sqrt{\frac{8\pi}{\beta m \omega^2}} \operatorname{Erf}\left[\sqrt{\frac{\beta m \omega^2}{2}} \frac{L}{4}\right] \qquad \text{in SI units of } [m] \tag{A.27}$$

The momentum integral can be solved using Eq. (A.1) giving $\sqrt{2\pi m\beta^{-1}}$ in $\left[\frac{\text{kgm}}{\text{s}}\right]$. The product of momentum and configuration integral gives the total partition function or Zustandssumme

$$Z^{\rm dw} = \frac{4\pi}{\beta\omega} \operatorname{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}} \frac{L}{4}\right] \qquad \text{in SI units of}\left[\frac{\operatorname{kg} \operatorname{m}^2}{s}\right] = [\operatorname{Js}] \qquad (A.28)$$

where Q possesses units of length [m] and Z possesses units of action [J s]. In the limit of an infinitely large box length, twice the well known expression for the partition function of the classical one-dimensional harmonic oscillator is recovered using Eq. (A.4)

$$\lim_{L \to \infty} Z^{\rm dw} = \frac{4\pi}{\beta\omega} \qquad \text{in SI units of } [\rm J\,s] \tag{A.29}$$

In the limit of $L \to \infty$, the potential wells are separated infinitely far from each other, thus each potential well should be equivalent to a system with a single-well harmonic potential of equal strength. Consequently, the partition function of the system in this limit should be twice that of the equivalent system with the single well harmonic potential, which is confirmed by comparison of Eq. (A.22)) with Eq. (A.29. With the derived partition functions Eqs. ((A.27) and (A.28), and Eqs. (A.13) and ((A.16) the probability density distributions in phase space and configuration space can be written as

$$\rho^{\mathrm{dw},\mathbf{q}} = \frac{\exp\left[-\frac{\beta m\omega^2 \mathbf{q}^2}{2}\right]}{\sqrt{\frac{8\pi}{\beta m\omega^2}} \mathrm{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}\frac{L}{4}}\right]} \qquad \text{in SI units of } [\mathrm{m}^{-1}] \tag{A.30}$$

$$\rho^{\mathrm{dw,\mathbf{p},\mathbf{q}}} = \frac{\exp\left[-\frac{\beta\mathbf{p}^2}{2m} - \frac{\beta m\omega^2 \mathbf{q}^2}{2}\right]}{\frac{4\pi}{\beta\omega} \mathrm{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}\frac{L}{4}}\right]} \quad \text{in SI units of } \left[(\mathrm{J\,s})^{-1}\right]$$
(A.31)

In the last part of this section we will check our results for consistency with physical expectations. The simplest check concerns the proper normalization of the derived probability distributions. The integral of the probability distributions Eqs. (A.30) and (A.31) over the whole configuration or phase space should give a probability of unity to find the particle inside the integrated region. This can be seen to be the case, because the integrals of the variable exponential terms in the numerator of Eqs. (A.30) and (A.31) yield again the partition functions, which also occur as constants in the denominator of these equations. Next, we will consider the limiting case of a vanishing harmonic potential, *i.e*, $\omega \rightarrow 0$. In this case, one would expect a uniform probability distribution throughout the entire box. The corresponding limit of the configuration space probability density is found using Eq. (A.5) and exploiting the symmetry of the harmonic potential



FIG. 2. Effective configuration space volume and effective phase space volume of the particle in a box models with a single-well or a double-well harmonic potential. The effective phase space volume of a given model differs from its the effective configuration space volume only by a constant factor $\sqrt{\frac{4\pi m}{\beta}}$. The particle mass is set to $m = 2 \times 10^{-26}$ [kg]. The angular frequency for (a) and (b) is set to $\omega = 10^{-26}$ [s⁻¹]. The box length for (c) and (d) is set to $L = 1.6 \times 10^{-9}$ [m]. (a) configuration space volume of the single-well (black curve) or double-well model (red curve) as function of the box length L (b) phase space volume of the single-well (black curve) or double-well model (red curve) as function of the box length L (c) configuration space volume of the single-well (black curve) or double-well model (red curve) as function of the box length L (c) configuration space volume of the angular frequency ω (c) phase space volume of the single-well (black curve) or double-well model (red curve) as function of the angular frequency ω (c) phase space volume of the single-well (black curve) or double-well model (red curve) as function of the angular frequency ω

about $\mathbf{q} = 0$

$$\lim_{\omega \to 0} \rho^{\mathrm{sw},\mathbf{q}} = \frac{\sqrt{\frac{\beta m \omega^2}{8\pi}} \lim_{\omega \to 0} \exp\left[-\frac{\beta m \omega^2 (|\mathbf{q}| - \frac{L}{4})^2}{2}\right]}{\lim_{\omega \to 0} \operatorname{Erf}\left[\sqrt{\frac{\beta m \omega^2}{2}}\frac{L}{4}\right]}$$

$$= \frac{\sqrt{\frac{\beta m \omega^2}{8\pi}}1}{\frac{2}{\sqrt{\pi}}\sqrt{\frac{\beta m \omega^2}{2}}\frac{L}{4}}$$

$$= \frac{\sqrt{\frac{\beta m \omega^2}{8\pi}}1}{\sqrt{\frac{\beta m \omega^2}{8\pi}}L}$$

$$= \frac{1}{L} \quad \text{in SI units of } [\mathrm{m}^{-1}]$$
(A.32)

and confirms our expectation.

The configuration and phase space probability density distributions are plotted for concrete example parameters in Fig. 1(b) and (d)

4. Effective phase space and configuration space volume

The effective configuration (phase) space volume was introduced in the main text as measure for the volume of the effectively populated regions of configuration (phase) space. In the context of our model formulated in a continuous coordinates, the effective configuration and phase space volumes are obtained from

$$\tilde{\Omega}^{\mathbf{q}} = \frac{1}{\int \left(\rho^{\mathbf{q}}\right)^2 \,\mathrm{d}\mathbf{q}} \tag{A.33}$$

$$\tilde{\Omega}^{\mathbf{p},\mathbf{q}} = \frac{1}{\int \left(\rho^{\mathbf{p},\mathbf{q}}\right)^2 \,\mathrm{d}\mathbf{p} \,\mathrm{d}\mathbf{q}} \tag{A.34}$$

In the following two sections, we will derive and examine concrete expressions for the phase and configuration space volume of our particle in the box models with a single-well and double-well harmonic potential. The effective configuration space and phase space volumes for these models are plotted in Fig. 2 as function of the box length L and as function of the angular frequency ω .

a. The particle in the box with a single-well harmonic potential

The effective configuration space volume is found from Eqs. (A.33) and (A.23) with the aid of Eq. (A.2)

$$\frac{1}{\tilde{\Omega}^{\mathrm{sw},\mathbf{q}}} = \int_{-\frac{L}{2}}^{\frac{L}{2}} (\rho^{\mathrm{sw},\mathbf{q}})^2 \,\mathrm{d}\mathbf{q}$$

$$= \frac{\frac{\beta m\omega^2}{2\pi}}{\left(\operatorname{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}}\frac{L}{2}\right]\right)^2 - \frac{L}{2}} \exp\left[-\beta m\omega^2 \mathbf{q}^2\right] \,\mathrm{d}\mathbf{q}$$

$$= \frac{\frac{\beta m\omega^2}{2\pi}}{\left(\operatorname{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}}\frac{L}{2}\right]\right)^2} \sqrt{\frac{\pi}{\beta m\omega^2}} \operatorname{Erf}\left[\sqrt{\beta m\omega^2}\frac{L}{2}\right]$$

$$= \sqrt{\frac{\beta m\omega^2}{4\pi}} \frac{\operatorname{Erf}\left[\sqrt{\beta m\omega^2}\frac{L}{2}\right]}{\left(\operatorname{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}}\frac{L}{2}\right]\right)^2}$$
(A.35)

from which we obtain

$$\tilde{\Omega}^{\mathrm{sw},\mathbf{q}} = \sqrt{\frac{4\pi}{\beta m \omega^2}} \frac{\left(\mathrm{Erf}\left[\sqrt{\frac{\beta m \omega^2}{2}} \frac{L}{2}\right]\right)^2}{\mathrm{Erf}\left[\sqrt{\beta m \omega^2} \frac{L}{2}\right]} \qquad \text{in SI units of } [\mathrm{m}] \qquad (A.36)$$

The limit of the effective configuration space volume as L approaches infinity is found by making use of Eq. A.4

$$\lim_{L \to \infty} \tilde{\Omega}^{\mathrm{sw}, \mathbf{q}} = \sqrt{\frac{4\pi}{\beta m \omega^2}} \tag{A.37}$$

In the limit of a vanishing harmonic potential, the effective configuration space volume approaches the box length L

$$\lim_{\omega \to 0} \tilde{\Omega}^{\mathrm{sw},\mathbf{q}} = \sqrt{\frac{4\pi}{\beta m \omega^2}} \frac{\left(\frac{2}{\sqrt{\pi}}\sqrt{\frac{\beta m \omega^2}{2}}\frac{L}{2}\right)^2}{\frac{2}{\sqrt{\pi}}\sqrt{\beta m \omega^2}\frac{L}{2}}$$

$$= \sqrt{\frac{4\pi}{\beta m \omega^2}} \frac{\frac{\beta m \omega^2}{2\pi}L^2}{\sqrt{\frac{\beta m \omega^2}{\pi}L}}$$

$$= L$$
(A.38)

where we made use of Eq. A.5. This confirms what one would expect on physical grounds, because in absence of the harmonic potential the particle should be free to move anywhere inside the box.

The effective phase space volume is found from Eqs. (A.34) and (A.24) with the aid of Eqs. (A.1) and (A.2)

$$\frac{1}{\tilde{\Omega}^{\mathrm{sw,p,q}}} = \int_{-\frac{L}{2}}^{\frac{L}{2}} (\rho^{\mathrm{sw,p,q}})^2 \, \mathrm{d}\mathbf{p} \, \mathrm{d}\mathbf{q}$$

$$= \frac{\int_{-\infty}^{\infty} \exp\left[-\frac{\beta\mathbf{p}^2}{m}\right] \, \mathrm{d}\mathbf{p} \int_{-\frac{L}{2}}^{\frac{L}{2}} \exp\left[-\beta m\omega^2 \mathbf{q}^2\right] \, \mathrm{d}\mathbf{q}}{\frac{4\pi^2}{\beta^2\omega^2} \left(\operatorname{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}}\frac{L}{2}\right]\right)^2} \quad (A.39)$$

$$= \frac{\frac{\pi}{\beta\omega} \left(\operatorname{Erf}\left[\sqrt{\beta m\omega^2}\frac{L}{2}\right]\right)}{\frac{4\pi^2}{\beta^2\omega^2} \left(\operatorname{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}}\frac{L}{2}\right]\right)^2} \\
= \frac{\beta\omega}{4\pi} \frac{\operatorname{Erf}\left[\sqrt{\beta m\omega^2}\frac{L}{2}\right]}{\left(\operatorname{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}}\frac{L}{2}\right]\right)^2}$$

from which we obtain

$$\tilde{\Omega}^{\mathrm{sw, \mathbf{p}, \mathbf{q}}} = \frac{4\pi}{\beta\omega} \frac{\left(\operatorname{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}} \frac{L}{2}\right]\right)^2}{\operatorname{Erf}\left[\sqrt{\beta m\omega^2} \frac{L}{2}\right]} \quad \text{in SI units of } [\mathrm{J\,s}] \quad (A.40)$$

The limit of the effective phase space volume as L approaches infinity is found by making use of Eq. A.4

$$\lim_{L \to \infty} \tilde{\Omega}^{\mathrm{sw}, \mathbf{p}, \mathbf{q}} = \frac{4\pi}{\beta\omega}$$
(A.41)

In the limit of a vanishing harmonic potential, the effective phase space volume approaches the box length *L* times the constant factor $\sqrt{\frac{4\pi m}{\beta}}$ originating from the additional integration over the momentum space in Eqs. (A.21) and (A.35) in comparison to Eqs. (A.20) and (A.35) which are the analogous equations formulated in configuration space

$$\lim_{\omega \to 0} \tilde{\Omega}^{\mathrm{sw},\mathbf{p},\mathbf{q}} = \frac{4\pi}{\beta\omega} \frac{\left(\frac{2}{\sqrt{\pi}}\sqrt{\frac{\beta m \omega^2}{2}}\frac{L}{2}\right)^2}{\frac{2}{\sqrt{\pi}}\sqrt{\beta m \omega^2}\frac{L}{2}}$$

$$= \frac{4\pi}{\beta\omega} \frac{\frac{\beta m \omega^2}{2\pi}L^2}{\sqrt{\frac{\beta m \omega^2}{\pi}L}}$$

$$= \sqrt{\frac{4\pi m}{\beta}}L$$
(A.42)

where we made use of Eq. A.5.

A last interesting comparison of $\lim_{L\to\infty} \tilde{\Omega}^{\mathbf{q}}$ can be made to the amplitude A of an equivalent harmonic oscillator moving under deterministic Newtonian mechanics without the stochastic fluctuations induced by the heat bath. The total energy of the oscillator is set equal to the mean total energy $\langle H \rangle = \beta^{-1}$ of the stochastic oscillator. The amplitude is found by setting this energy equal to the maximum potential energy (see Eq. (A.9)) of the oscillator at an extension $\mathbf{q} = A$

$$E_{max}^{pot} = \frac{1}{2}m\omega^2 \mathbf{q}_{max}^2$$

$$\beta^{-1} = \frac{1}{2}m\omega^2 A^2$$
(A.43)

from which we obtain

$$A = \sqrt{\frac{2}{\beta m \omega^2}} \tag{A.44}$$

Comparison to Eq. (A.41) shows

$$\frac{\lim_{L \to \infty} \tilde{\Omega}^{\mathbf{q}}}{A} = \sqrt{2\pi} \tag{A.45}$$

that is, the effective configuration space volume is a factor $\sqrt{2\pi}$ larger than the corresponding deterministic amplitude.

b. The particle in the box with a double-well harmonic potential

The effective configuration space volume is found from Eqs. (A.33) and (A.30) exploiting the symmetry of the harmonic potential about $\mathbf{q} = 0$ using Eq. (A.1) for the momentum dependent integral and Eq. (A.3) for the configuration dependent integral using the temporary variable substitution $\mathbf{q}' = \mathbf{q} + \frac{L}{4}$ for the integration

$$\frac{1}{\tilde{\Omega}^{\mathrm{dw},\mathbf{q}}} = \int_{-\frac{L}{2}}^{\frac{\pi}{2}} \left(\rho^{\mathrm{dw},\mathbf{q}}\right)^{2} d\mathbf{q}$$

$$= \frac{\frac{\beta m \omega^{2}}{2\pi}}{\left(\operatorname{Erf}\left[\sqrt{\frac{\beta m \omega^{2}}{2}}\frac{L}{2}\right]\right)^{2}} 2\int_{-\frac{L}{2}}^{0} \exp\left[-\beta m \omega^{2} \left(\mathbf{q} + \frac{L}{4}\right)^{2}\right] d\mathbf{q}$$

$$= \frac{\frac{\beta m \omega^{2}}{\pi}}{\left(\operatorname{Erf}\left[\sqrt{\frac{\beta m \omega^{2}}{2}}\frac{L}{4}\right]\right)^{2}} \sqrt{\frac{\pi}{16\beta m \omega^{2}}} \operatorname{Erf}\left[\sqrt{\beta m \omega^{2}}\frac{L}{4}\right]$$

$$= \sqrt{\frac{\beta m \omega^{2}}{16\pi}} \frac{\operatorname{Erf}\left[\sqrt{\beta m \omega^{2}}\frac{L}{4}\right]}{\left(\operatorname{Erf}\left[\sqrt{\frac{\beta m \omega^{2}}{2}}\frac{L}{4}\right]\right)^{2}}$$
(A.46)

from which we obtain

$$\tilde{\Omega}^{\mathrm{dw},\mathbf{q}} = \sqrt{\frac{16\pi}{\beta m \omega^2}} \frac{\left(\mathrm{Erf}\left[\sqrt{\frac{\beta m \omega^2}{2}} \frac{L}{4}\right]\right)^2}{\mathrm{Erf}\left[\sqrt{\beta m \omega^2} \frac{L}{4}\right]} \qquad \text{in SI units of } [\mathrm{m}] \tag{A.47}$$

The limit of the effective configuration space volume as L approaches infinity is found by making use of Eq. A.4

$$\tilde{\Omega}^{\mathrm{dw},\mathbf{q}} \lim_{L \to \infty} = \sqrt{\frac{16\pi}{\beta m \omega^2}}$$
(A.48)

Eq. (A.48) is twice the corresponding limit of the configuration space volume for the single-well model Eq. (A.37), The factor 2 can be rationalized by recognizing that the double-well model is in this limit equivalent to two single-well models separated by an infinitely large distance. In the limit of a vanishing harmonic potential, the effective configuration space volume approaches the

$$\tilde{\Omega}^{\mathrm{dw,q}} \lim_{\omega \to 0} = \sqrt{\frac{16\pi}{\beta m \omega^2}} \frac{\left(\frac{2}{\sqrt{\pi}} \sqrt{\frac{\beta m \omega^2}{2}} \frac{L}{4}\right)^2}{\frac{2}{\sqrt{\pi}} \sqrt{\beta m \omega^2} \frac{L}{4}}$$

$$= \sqrt{\frac{16\pi}{\beta m \omega^2}} \frac{\frac{\beta m \omega^2}{\pi} \frac{L^2}{8}}{\sqrt{\frac{\beta m \omega^2}{\pi} \frac{L}{2}}}$$

$$= L$$
(A.49)

where we made use of Eq. A.5. This confirms what one would expect on physical grounds, because in absence of the harmonic potential the particle should be free to move anywhere inside the box.

The effective phase space volume is found from Eqs. (A.34) and (A.31) with the aid of Eq. (A.1) for the integration of the momentum dependent integral. During the integration of the configuration dependent integral, the symmetry of the harmonic potential about $\mathbf{q} = 0$ is exploited and a temporary variable substitution $\mathbf{q}' = \mathbf{q} + \frac{L}{4}$ is used to be able to employ Eq. (A.3)

$$\frac{1}{\tilde{\Omega}^{\mathrm{dw},\mathbf{p},\mathbf{q}}} = \int_{-\frac{L}{2}}^{\frac{L}{2}} \left(\rho^{\mathrm{dw},\mathbf{p},\mathbf{q}}\right)^{2} d\mathbf{p} d\mathbf{q} \\
= \frac{\int_{-\infty}^{\infty} \exp\left[-\frac{\beta\mathbf{p}^{2}}{m}\right] d\mathbf{p} \ 2 \int_{-\frac{L}{2}}^{0} \exp\left[-\beta m\omega^{2}\left(\mathbf{q}+\frac{L}{4}\right)^{2}\right] d\mathbf{q}}{\frac{16\pi^{2}}{\beta^{2}\omega^{2}} \left(\operatorname{Erf}\left[\sqrt{\frac{\beta m\omega^{2}}{2}}\frac{L}{4}\right]\right)^{2}} \qquad (A.50)$$

$$= \frac{\frac{2\pi}{\beta\omega} \left(\operatorname{Erf}\left[\sqrt{\beta m\omega^{2}}\frac{L}{4}\right]\right)}{\frac{16\pi^{2}}{\beta^{2}\omega^{2}} \left(\operatorname{Erf}\left[\sqrt{\frac{\beta m\omega^{2}}{2}}\frac{L}{4}\right]\right)^{2}} \\
= \frac{\beta\omega}{8\pi} \frac{\operatorname{Erf}\left[\sqrt{\beta m\omega^{2}}\frac{L}{4}\right]}{\left(\operatorname{Erf}\left[\sqrt{\frac{\beta m\omega^{2}}{2}}\frac{L}{4}\right]\right)^{2}}$$

from which we obtain

$$\tilde{\Omega}^{\mathrm{dw, \mathbf{p}, \mathbf{q}}} = \frac{8\pi}{\beta\omega} \frac{\left(\mathrm{Erf}\left[\sqrt{\frac{\beta m\omega^2}{2}}\frac{L}{4}\right]\right)^2}{\mathrm{Erf}\left[\sqrt{\beta m\omega^2}\frac{L}{4}\right]} \qquad \text{in SI units of } [\mathrm{J\,s}]$$
(A.51)

The limit of the effective phase space volume as L approaches infinity is found by making use of Eq. A.4

$$\tilde{\Omega}^{\mathrm{dw, \mathbf{p}, \mathbf{q}}} \lim_{L \to \infty} = \frac{8\pi}{\beta\omega}$$
(A.52)

which is twice the corresponding phase space volume of the corresponding single-well model Eq. (A.52), as can be understood by the same arguments employed in the interpretation of the configuration space volume Eq. (A.48) in this limit. In the limit of a vanishing harmonic potential, the effective phase space volume approaches the box length L times the constant factor $\sqrt{\frac{4\pi m}{\beta}}$ originating from the additional integration over the momentum space in Eqs. (A.28) and (A.46) in comparison to Eqs. (A.27) and (A.46) which are the analogous equations formulated in configuration space

$$\tilde{\Omega}^{\mathrm{dw, \mathbf{p}, \mathbf{q}}} \lim_{\omega \to 0} = \frac{8\pi}{\beta\omega} \frac{\left(\frac{2}{\sqrt{\pi}}\sqrt{\frac{\beta m \omega^2}{2}}\frac{L}{4}\right)^2}{\frac{2}{\sqrt{\pi}}\sqrt{\beta m \omega^2}\frac{L}{4}}$$

$$= \frac{8\pi}{\beta\omega} \frac{\frac{\beta m \omega^2}{2\pi}\frac{L^2}{4}}{\sqrt{\frac{\beta m \omega^2}{\pi}\frac{L}{2}}}$$

$$= \sqrt{\frac{4\pi m}{\beta}}L$$
(A.53)

where we made use of Eq. A.5.

5. Extension of the formalism to multiple non-interacting particles

The particle in the box models presented in the previous sections are readily extended to the case of multiple, non-interacting (ideal gas) particles residing in the box. Since the particles do not interact, the integration of the partition function can be done for each particle separately and the total partition functions are given by the product of the individual one-particle partition functions.

$$Z = \prod_{i=1}^{N} Z_i = Z_{\text{one-particle}}^{N}$$
(A.54)

$$Q = \prod_{i=1}^{N} Q_i = Q_{\text{one-particle}}^N$$
(A.55)

The same applies to the total configuration space volume Ψ , the effective configuration space volume $\tilde{\Omega}^{\mathbf{q}}$ and the effective phase space volume $\tilde{\Omega}^{\mathbf{p},\mathbf{q}}$

$$\Psi = \prod_{i=1}^{N} \Psi_i = \Psi_{\text{one-particle}}^N$$
(A.56)

$$\tilde{\Omega}^{\mathbf{p}} = \prod_{i=1}^{N} \tilde{\Omega}_{i}^{\mathbf{p}} = \left(\tilde{\Omega}_{\text{one-particle}}^{\mathbf{p}}\right)^{N}$$
(A.57)

$$\tilde{\Omega}^{\mathbf{p},\mathbf{q}} = \prod_{i=1}^{N} \tilde{\Omega}_{i}^{\mathbf{p},\mathbf{q}} = \left(\tilde{\Omega}_{\text{one-particle}}^{\mathbf{p},\mathbf{q}}\right)^{N}$$
(A.58)

If the particles are indistinguishable, the number of actual microstates of the system is reduced by a factor $\frac{1}{N!}$, which is given by the number of possible permutations of the particle identities denoted in Eqs. (A.54), (A.55), (A.56), (A.57) and (A.58) by the index variable *i*. Consequently, the factor $\frac{1}{N!}$ has to be added to Eqs. (A.54), (A.55), (A.56), (A.56), (A.57) and (A.58) if the particles are indistinguishable.

6. Pressure

In this section we will derive expressions for the pressure of our particle in the box models from the partition functions obtained in Sec. A 3 as example for the derivation of thermodynamic state functions from the partition function. The resulting expressions for the pressure in certain limiting cases will be verified by comparison to well known results of elementary statistical thermodynamics. The text-book expression for the external pressure P^{ext} exerted against the walls of the box is given by the negative derivative of the free energy with respect to the box length

$$P^{\text{ext}} = \frac{\mathrm{d}F}{\mathrm{d}L}$$

$$\beta P^{\text{ext}} = \frac{-\mathrm{d}\ln Z}{\mathrm{d}L}$$
(A.59)

where we note that the momentum integral and other possibly added constant factors independent on L, like the factor $\frac{1}{N!}$ introduced in the previous section to account for the indistinguishability of the N particles can be separated off the configurational integral and vanish upon differentiation. For simplicity we compile all factors independent on L in the constant c, and express the configurational partition function in terms of the one-particle partition function, assuming again N identical, non-interacting particles

$$\beta P^{\text{ext}} = \frac{-\mathrm{d}\ln Q_{\text{one-particle}}^N}{\mathrm{d}L} + \frac{\mathrm{d}\ln c}{\mathrm{d}L}$$

$$\frac{\beta P^{\text{ext}}}{N} = \frac{-\mathrm{d}\ln Q_{\text{one-particle}}}{\mathrm{d}L}$$
(A.60)

The external pressure of the single-well and double well harmonic potentials can now be derived by substituting $Q_{\text{one-particle}}$ with the corresponding configurational partition function given by Eq. (A.20) or Eq. (A.27), respectively. The differentiation is carried out by making use of Eq. (A.7) from which we obtain

$$\frac{\beta P^{\text{ext,sw}}}{N} = \frac{\sqrt{\frac{\beta m \omega^2}{2\pi}} \exp\left[-\frac{\beta m \omega^2}{8}L^2\right]}{\operatorname{Erf}\left[\sqrt{\frac{\beta m \omega^2}{2}}\frac{L}{2}\right]}$$
(A.61)

$$\frac{\beta P^{\text{ext,dw}}}{N} = \frac{\sqrt{\frac{\beta m \omega^2}{2\pi}} \exp\left[-\frac{\beta m \omega^2}{32}L^2\right]}{2\text{Erf}\left[\sqrt{\frac{\beta m \omega^2}{2}}\frac{L}{4}\right]}$$
(A.62)

To confirm result we consider the limit of Eqs. (A.61) and (A.62) as ω approaches zero, that is in the absence of the harmonic potentials. Making use of Eq. (A.5), we obtain for the single-well model

$$\lim_{L \to 0} \frac{\beta P^{\text{ext,sw}}}{N} = \frac{\sqrt{\frac{\beta m \omega^2}{2\pi} 1}}{\frac{2}{\sqrt{\pi}} \sqrt{\frac{\beta m \omega^2}{2} \frac{L}{2}}}$$

$$= \frac{1}{L}$$
(A.63)

Making again use of Eq. (A.5), we obtain for the double-well model

$$\lim_{L \to 0} \frac{\beta P^{\text{ext,dw}}}{N} = \frac{\sqrt{\frac{\beta m \omega^2}{2\pi}} 1}{2\frac{2}{\sqrt{\pi}}\sqrt{\frac{\beta m \omega^2}{2}} \frac{L}{4}}$$

$$= \frac{1}{L}$$
(A.64)

The limiting cases Eqs. (A.63) and (A.64) match the ideal gas law, confirming our result. Another interesting limit of the external pressure is obtained considering the limit $L \to \infty$, which is equal to the limit for $\omega \to \infty$

$$\lim_{L \to \infty} \frac{\beta P^{\text{ext,sw}}}{N} = \lim_{\omega \to \infty} \frac{\beta P^{\text{ext,sw}}}{N} = 0$$
(A.65)

In both cases, the particle does essentially never reach the walls of the box, explaining why the external pressure approaches zero in these limits.

7. Correction terms for the free energy perturbation schemes

In this section, we derive explicit expressions for the FEP correction terms introduced in Sec. 2 of the main article as applied to the particle in the box models. We will demonstrate the significance of these correction terms at our example systems with special emphasize on the units acquired by the correction terms and the single move FEP integrals themselves. Furthermore, we

will extend the example transformations of the particle in the box models presented in Sec. 3.4 of the main paper with two more model transformations in which the particle number in the system changes during the transformation.

The correction terms for the fast and slow single move FEP schemes derived in Sec. 2.3 are given by (from Eqs. (46) and (45))

$$\Delta\Delta F_{0\to1}^{\text{rand}} = -\beta^{-1} \ln \frac{\Psi_1}{\Psi_0} \tag{A.66}$$

$$\Delta \Delta F_{0 \to 1}^{\text{equi}} = -\beta^{-1} \ln \frac{\tilde{\Omega}_1}{\Psi_0}$$
(A.67)

From Eqs. (23) and (31) we know that

$$\Delta \tilde{F}_{0 \to 1}^{\text{rand}} = \Delta F_{0 \to 1} - \Delta \Delta F_{0 \to 1}^{\text{rand}}$$
(A.68)

$$\Delta \tilde{F}_{0 \to 1}^{\text{equi}} = \Delta F_{0 \to 1} - \Delta \Delta F_{0 \to 1}^{\text{equi}}$$
(A.69)

Substituting

$$\Delta F_{0\to 1} = -\beta^{-1} \ln \frac{Q_1}{Q_0},\tag{A.70}$$

 $\Delta\Delta F_{0\to 1}^{\text{rand}}$ with Eq. (A.66) and $\Delta\Delta F_{0\to 1}^{\text{equi}}$ with Eq. (A.67) we obtain

$$\Delta \tilde{F}_{0 \to 1}^{\text{rand}} = -\beta^{-1} \ln \frac{Q_1 \Psi_0}{Q_0 \Psi_1}$$
(A.71)

$$\Delta \tilde{F}_{0\to 1}^{\text{equi}} = -\beta^{-1} \ln \frac{Q_1 \Psi_0}{Q_0 \tilde{\Omega}_1} \tag{A.72}$$

From Secs. A 5 and A 5 we know that the configurational partition function Q as well as the total configuration space volume Ψ and the effective configuration space volume $\tilde{\Omega}$ have units of $[m^{DN}]$, where D is the dimensionality of the system (D = 1 in our case) and N is the number of particles in the system. From the expressions for the free energy difference Eq. (A.70), for the free energy estimates from the single move FEP schemes Eqs. (A.71) and (A.72) and their corresponding correction terms Eqs. (A.66) and (A.67), it can be seen that the dimensionality and the total number of particles is required to be constant during the transformation for the units to cancel. Now one might ask how the formalism can be applied then to transformations involving a change in the number of particles residing in the system. In principle there can be two reasons for a change in the particle number. The first possibility is a redistribution of mass among the particles (for example by aggregation or disaggregation of particles). In this case the number of actual particles (for example atoms) would in reality still be the same, but their spatial position might be highly correlated if the interaction of the particles is very strong. Nevertheless, the particles will still retain some freedom of relative movement, because the entropic cost of fixing their relative position would be infinitely high. If the aggregates have a very well-defined, almost rigid structure, the entropic contribution of their relative movement to the free energy can be made implicit assuming a constant energy term describing the formation of the aggregate, and the particle aggregate can then be treated as one particle during the simulation. The second possibility is the exchange of particles with an external reservoir. In this case the formulation of the problem must include the particles of the reservoir. In the calculation of binding free energies, the unbound particles in the reservoir are often treated implicitly by inclusion of a chemical potential term accounting for the free energy cost of removing them from the reservoir (see for example Ref. [14]). An explicit inclusion of the reservoir particles in the present formalism transforms Eqs. (A.66) and (A.67) into

$$\Delta \Delta F_{0 \to 1}^{\text{rand}} = -\beta^{-1} \ln \frac{\Psi_1^{\text{main}} \Psi_1^{\text{res}}}{\Psi_0^{\text{main}} \Psi_0^{\text{res}}}$$
(A.73)

$$\Delta \Delta F_{0 \to 1}^{\text{equi}} = -\beta^{-1} \ln \frac{\tilde{\Omega}_1^{\mathbf{q}, \text{main}} \tilde{\Omega}_1^{\mathbf{q}, \text{res}}}{\Psi_0^{\text{main}} \Psi_0^{\text{res}}}$$
(A.74)

where the superscripts *main* and *res* indicate the main model and the particle reservoir, respectively. Similarly, the expression for the transformation free energy becomes

$$\Delta F_{0\to 1} = -\beta^{-1} \ln \frac{Q_1^{\text{main}} Q_1^{\text{res}}}{Q_0^{\text{main}} Q_0^{\text{res}}}$$
(A.75)

a. Definition of an external particle reservoir

As mentioned above, the treatment of a transformation involving the addition of particles from an external reservoir to our main system requires the inclusion of the particles within the reservoir into the description of the initial state. In order to include the reservoir in the formalism we need a knowledge about the nature of the reservoir and its energetics. For our example, we will use the simplest possible definition of a particle reservoir depicted in Fig 3. The reservoir consists of an (infinitely) large number X of one-dimensional boxes of length L_0 (the same length as our particle in the box model in the initial state 0). Each individual box contains exactly one particle identical to the particles in the main system. The potential energy within the reservoir boxes is zero. The reservoir has the same temperature as the main system. During the transformation, ΔN particles will be removed from the reservoir and inserted into the main system. The particles are not exchangeable between the individual boxes and are thus treated as distinguishable. With this model of the particle reservoir and the number of removed particles, we can write the total



FIG. 3. An infinite particle reservoir formed by one-dimensional boxes of length L_0 containing one particle each. The particles can not be exchanged among the boxes and are consequently treated as distinguishable.

configuration space volume of the particle reservoir in the initial state 0 and the final state 1 as

$$\Psi_0^{\text{res}} = \prod_{i=1}^X L_0 \tag{A.76}$$

$$\Psi_1^{\text{res}} = \prod_{i=1}^{X - \Delta N} L_0 \tag{A.77}$$

$$\frac{\Psi_1^{\text{res}}}{\Psi_0^{\text{res}}} = L_0^{-\Delta N} \tag{A.78}$$

One can see, that, with this simple definition of the particle reservoir, we do not need knowledge about the total number of particles within the reservoir but only about the change in the reservoir's particle number accompanying the transformation to be able to specify $\frac{\Psi_1^{\text{res}}}{\Psi_0^{\text{res}}}$ in Eqs. (A.73) and (A.74). Note that, since the potential energy in the reservoir boxes is zero, the effective configuration space volume is equal to the total configuration space volume and also to the configurational partition function and thus

$$\frac{\tilde{\Omega}_1^{\mathbf{q},\mathrm{res}}}{\Psi_0^{\mathrm{res}}} = L_0^{-\Delta N} \tag{A.79}$$

For the same reason, also the configurational partition function is equal to the total configuration space volume. Like the total configuration space volume, the total configurational partition function is given by the product of partition functions for the individual reservoir boxes and thus

$$\frac{Q_1^{\text{res}}}{Q_0^{\text{res}}} = L_0^{-\Delta N} \tag{A.80}$$

b. Model transformations

In this section, we will state explicit expressions for the correction terms in case of several example transformations of our particle in the box models including the example transformation presented in the main text and examine their magnitude at concrete numeric examples. The transformations can involve a change in box length, angular frequency, potential type (single-well or double-well) and number of (ideal gas) particles:

- 1: scaling of the angular frequency by a factor a and scaling of the box length by a factor b
- 2: scaling of the angular frequency by a factor a and adding of ΔN particles
- 3: adding of ΔN particles and scaling of the box length by a factor b

The model transformations 2 and 3 can also be interpreted as a simplified model of binding, where the number of particles in the main system would represent the number of bound particles, the magnitude of the angular frequency would determine the strength of the interaction of ligand and receptor, and where a change in the box length during the transformation would correspond to a change in the total available configuration space volume upon binding. For a real binding reaction, the available configuration space volume will mostly decrease upon binding because the ligand and receptor need to come into close proximity and need to adopt configurations (overall rotation, internal conformation) compatible with the bound state.

In the following, concrete expressions are stated for the corresponding correction terms required for the application of the random single-move (fast) and for the single-move FEP scheme according to the equilibrium distribution of the target end state (slow). For the slow single-move FEP scheme, the correction terms are given for a transition from a single-well particle in the box model as end state 0 and a single-well or double-well model as end state 1, respectively. In the fast single-move FEP scheme, the correction term does not depend on the effective configuration space volume of the target end state, but on the total configuration space volume only. Thus, the correction terms are identical for the single-well or double-well particle in the box models as target end state for the fast single-move FEP scheme. Consequently, three variants are given for each of the three transformation types:

- sw → sw: transition from a single-well to a single-well model in the fast single-move FEP scheme
- sw \rightarrow sw: transition from a single-well to a single-well model in the slow single-move FEP scheme
- sw \rightarrow dw: transition from a single-well to a double-well model in the slow single-move FEP scheme

The particle mass is set to m = 2.0E-26 [kg]. For all three transformation types, the box length of the initial state 0 is $L_0 = 1.6\text{E}-09$ [m], its angular frequency is $\omega_0 = 1.0\text{E}+12$ [s⁻¹] and its particle number is $N_0 = 1$. Furthermore, the initial state possesses always a single-well harmonic potential. The length of the individual boxes of the particle reservoir is L_0 . See Sec. A 7 a for a detailed description of the particle reservoir. In the following the transformation type will be indicated by the superscripts attached to the symbols of the free energy changes and correction terms.

In the first transformation type, the frequency and the box length are scaled by a constant factor so that $\omega_1 = a\omega_0$ and $L_1 = bL_0$ where $\{a, b\} > 0$ are constant factors. The free energy differences for the transition from a single-well model to a single-well model or a double-well model are plotted in Fig. 4(a) and (b). The correction terms corresponding to the three variants are plotted in Fig. 5(a)-(c). Expressions for the transformation free energies are formulated according to Eq. (A.75) using the configurational partition functions of the particle in the box models Eqs. (A.20) and Eqs. (A.27). The ratio of the reservoir configurational partition functions is substituted with Eq. (A.80).

$$\Delta F_{\rm sw \to sw}^{1} = -\beta^{-1} \ln \frac{\operatorname{Erf}\left[\sqrt{\frac{\beta m (a\omega_{0})^{2}}{2}} \frac{bL_{0}}{2}\right]}{a \operatorname{Erf}\left[\sqrt{\frac{\beta m \omega_{0}^{2}}{2}} \frac{L_{0}}{2}\right]}$$
(A.81)
$$\Delta F_{\rm sw \to dw}^{1} = -\beta^{-1} \ln \frac{\operatorname{Erf}\left[\sqrt{\frac{\beta m (a\omega_{0})^{2}}{2}} \frac{bL_{0}}{4}\right]}{2a \operatorname{Erf}\left[\sqrt{\frac{\beta m \omega_{0}^{2}}{2}} \frac{L_{0}}{2}\right]}$$
(A.82)

Expressions for the correction terms can be obtained by inserting the total and effective configuration space volumes into Eqs. (A.73) and Eqs. (A.74). The effective configuration space volumes are taken from Eqs. (A.57) and (A.36) or (A.47) for the single-well and double-well models, respectively. With this, we obtain

$$\Delta \Delta F_{\text{xw} \to \text{xw}}^{\text{fast},1} = -\beta^{-1} \ln b \tag{A.83}$$

$$\Delta\Delta F_{\rm sw\to sw}^{\rm slow,1} = -\beta^{-1} \ln \frac{\sqrt{\frac{4\pi}{\beta m (a\omega_0)^2}} \left(\operatorname{Erf} \left[\sqrt{\frac{\beta m (a\omega_0)^2}{2}} \frac{bL_0}{2} \right] \right)}{L_0 \operatorname{Erf} \left[\sqrt{\beta m (a\omega_0)^2} \frac{bL_0}{2} \right]}$$
(A.84)

$$\Delta\Delta F_{\rm sw\to dw}^{\rm slow,1} = -\beta^{-1} \ln \frac{\sqrt{\frac{16\pi}{\beta m(a\omega_0)^2}} \left(\operatorname{Erf}\left[\sqrt{\frac{\beta m(a\omega_0)^2}{2}} \frac{bL_0}{4}\right] \right)^2}{L_0 \operatorname{Erf}\left[\sqrt{\beta m \left(a\omega_0\right)^2} \frac{bL_0}{4}\right]}$$
(A.85)

In the second transformation type, the frequency is scaled by a constant factor a > 0 so that $\omega_1 = a\omega_0$ and ΔN particles are taken from an external reservoir described in Sec. A 7 a and added to the system. The correction terms corresponding to the three variants are plotted in Fig. 5(d)-(f). Expressions for the transformation free energies are formulated according to Eq. (A.75) using the configurational partition functions of the particle in the box models Eqs. (A.20) and Eqs. (A.27) and the rules to calculate the N particle partition functions for indistinguishable particles from them (see Sec. A 5). The ratio of the reservoir configurational partition functions is substituted with Eq. (A.80).

$$\Delta F_{\rm sw\to sw}^2 = -\beta^{-1} \ln \frac{\left(\sqrt{\frac{2\pi}{\beta m (a\omega_0)^2}} \operatorname{Erf}\left[\sqrt{\frac{\beta m (a\omega_0)^2}{2}} \frac{L_0}{2}\right]\right)^{\Delta N+1}}{(\Delta N+1)! \sqrt{\frac{2\pi}{\beta m\omega_0^2}} \operatorname{Erf}\left[\sqrt{\frac{\beta m \omega_0^2}{2}} \frac{L_0}{2}\right] L_0^{\Delta N}}$$
(A.86)
$$\Delta F_{\rm sw\to dw}^2 = -\beta^{-1} \ln \frac{\left(\sqrt{\frac{8\pi}{\beta m (a\omega_0)^2}} \operatorname{Erf}\left[\sqrt{\frac{\beta m (a\omega_0)^2}{2}} \frac{L_0}{4}\right]\right)^{\Delta N+1}}{(\Delta N+1)! \sqrt{\frac{2\pi}{\beta m\omega_0^2}} \operatorname{Erf}\left[\sqrt{\frac{\beta m \omega_0^2}{2}} \frac{L_0}{2}\right] L_0^{\Delta N}}$$
(A.87)

Expressions for the correction terms can be obtained by inserting the total and effective configuration space volumes into Eqs. (A.73) and Eqs. (A.74). The effective configuration space volumes are taken from Eqs. (A.57) and (A.36) or (A.47) for the single-well and double-well models, respectively. Eq. (A.78) or Eq. (A.79) is inserted for the relative configuration space volume of the reservoir at the two end states. With this, we obtain

$$\Delta\Delta F_{\rm xw\to xw}^{\rm fast,2} = -\beta^{-1} \ln \frac{1}{(\Delta N+1)!} \tag{A.88}$$

$$\Delta\Delta F_{\rm sw\to sw}^{\rm slow,2} = -\beta^{-1} \ln \frac{\left(\sqrt{\frac{4\pi}{\beta m(a\omega_0)^2}} \left(\operatorname{Erf}\left[\sqrt{\frac{\beta m(a\omega_0)^2}{2}}\frac{L_0}{2}\right]\right)^2\right)}{\left(\Delta N+1\right)! \left(L_0 \operatorname{Erf}\left[\sqrt{\beta m\left(a\omega_0\right)^2}\frac{L_0}{2}\right]\right)^{\Delta N+1}}$$
(A.89)

$$\Delta\Delta F_{\rm sw\to dw}^{\rm slow,2} = -\beta^{-1} \ln \frac{\left(\sqrt{\frac{16\pi}{\beta m(a\omega_0)^2}} \left(\operatorname{Erf}\left[\sqrt{\frac{\beta m(a\omega_0)^2}{2}}\frac{L_0}{4}\right]\right)^2\right)^{\Delta N+1}}{\left(\Delta N+1\right)! \left(L_0 \operatorname{Erf}\left[\sqrt{\beta m\left(a\omega_0\right)^2}\frac{L_0}{4}\right]\right)^{\Delta N+1}}$$
(A.90)

where the factor $(\Delta N + 1)!$ accounts for the indistinguishability of the particles within the main system (see Sec. A 5).

In the third transformation type, the length is scaled by a constant factor b > 0 so that $L_1 = bL_0$ and ΔN particles are taken from an external reservoir described in Sec. A7 a and added to

the system. The correction terms corresponding to the three variants are plotted in Fig. 5(g)-(i). Expressions for the transformation free energies are formulated according to Eq. (A.75) using the configurational partition functions of the particle in the box models Eqs. (A.20) and Eqs. (A.27) and the rules to calculate the N particle partition functions for indistinguishable particles from them (see Sec. A 5). The ratio of the reservoir configurational partition functions is substituted with Eq. (A.80).

$$\Delta F_{\rm sw\to sw}^{3} = -\beta^{-1} \ln \frac{\left(\sqrt{\frac{2\pi}{\beta m\omega_{0}^{2}}} \operatorname{Erf}\left[\sqrt{\frac{\beta m\omega_{0}^{2}}{2}} \frac{bL_{0}}{2}\right]\right)^{\Delta N+1}}{(\Delta N+1)! \sqrt{\frac{2\pi}{\beta m\omega_{0}^{2}}} \operatorname{Erf}\left[\sqrt{\frac{\beta m\omega_{0}^{2}}{2}} \frac{L_{0}}{2}\right] L_{0}^{\Delta N}}$$

$$\Delta F_{\rm sw\to dw}^{3} = -\beta^{-1} \ln \frac{\left(\sqrt{\frac{8\pi}{\beta m\omega_{0}^{2}}} \operatorname{Erf}\left[\sqrt{\frac{\beta m\omega_{0}^{2}}{2}} \frac{bL_{0}}{4}\right]\right)^{\Delta N+1}}{(\Delta N+1)! \sqrt{\frac{2\pi}{\beta m\omega_{0}^{2}}} \operatorname{Erf}\left[\sqrt{\frac{\beta m\omega_{0}^{2}}{2}} \frac{L_{0}}{2}\right] L_{0}^{\Delta N}}$$
(A.91)
(A.92)

Expressions for the correction terms can be obtained by inserting the total and effective configuration space volumes into Eqs. (A.73) and Eqs. (A.74). The effective configuration space volumes are taken from Eqs. (A.57) and (A.36) or (A.47) for the single-well and double-well models, respectively. Eq. (A.78) or Eq. (A.79) is inserted for the relative configuration space volume of the reservoir at the two end states. With this, we obtain

$$\Delta\Delta F_{\text{xw}\to\text{xw}}^{\text{fast},3} = -\beta^{-1} \ln \frac{b^{\Delta N+1}}{(\Delta N+1)!}$$
(A.93)
$$\left(\sum_{k=0}^{\infty} \left(\sum_{k=0}^{\infty} \left($$

$$\Delta\Delta F_{\rm sw\to sw}^{\rm slow,3} = -\beta^{-1} \ln \frac{\left(\sqrt{\frac{4\pi}{\beta m\omega_0^2}} \left(\operatorname{Erf}\left[\sqrt{\frac{\beta m\omega_0^2}{2}}\frac{bL_0}{2}\right]\right)\right)}{(\Delta N+1)! \left(L_0 \operatorname{Erf}\left[\sqrt{\beta m\omega_0^2}\frac{bL_0}{2}\right]\right)^{\Delta N+1}}$$
(A.94)

$$\Delta\Delta F_{\rm sw\to dw}^{\rm slow,3} = -\beta^{-1} \ln \frac{\left(\sqrt{\frac{16\pi}{\beta m\omega_0^2}} \left(\operatorname{Erf}\left[\sqrt{\frac{\beta m\omega_0^2}{2}} \frac{bL_0}{4}\right]\right)^2\right)^{\Delta N+1}}{(\Delta N+1)! \left(L_0 \operatorname{Erf}\left[\sqrt{\beta m\omega_0^2} \frac{bL_0}{4}\right]\right)^{\Delta N+1}}$$
(A.95)

where the factor $(\Delta N + 1)!$ accounts for the indistinguishability of the particles within the main system (see Sec. A 5).

From comparison of Figs. 4 and 5, it can be seen that the correction terms are, especially for the slow single move scheme, often comparable in magnitude to the transformation free energies themselves.



FIG. 4. The plots show the free energy difference for the three different transformation types of the particle in the box models. For each of the transformation types the free energy is plotted for the transition of a single-well model to a single-well model (left column) and for the transition of a single-well model to a double-well model (right column) Positive and negative values are indicated by blue and red color, respectively. Contour values are given in kcal/mol. Parameter values are given in the text (same as in Fig. 6 of the main article). (a)-(b) The transformation involves a scaling of the frequency by a factor *a* and a scaling of the box length by a factor *b*. The plots are identical to the ones presented in Fig. Figure 5 of the main article and are reproduced here to facilitate comparison with the additional plots. (c)-(d) The transformation consists of a scaling of the frequency by a factor *a* and adding ΔN particles from an external reservoir. (e)-(f) The transformation consists of adding ΔN particles from an external reservoir and scaling of the box length by a factor *b*. (a), (c), (e) $\Delta F_{0\rightarrow 1}$ for transformations with a transition from a single-well to another single-well harmonic potential. (b), (d), (f) $\Delta F_{0\rightarrow 1}$ for transformations with a transition from a single-well to a double-well harmonic potential.



FIG. 5. The plots show the correction terms of the single move FEP schemes applied to three three different transformations of the harmonic oscillator models. Positive and negative values are indicated by blue and red color, respectively. Contour values are given in kcal/mol. Parameter values are given in the text (same as in Fig. 6 of the main article). (a)-(c) The transformation involves a scaling of the frequency by a factor *a* and a scaling of the box length by a factor *b*. The plots are identical to the ones presented in Fig. Figure 5 of the main article and are reproduced here to facilitate comparison with the additional plots. (d)-(f) The transformation consists of a scaling of the frequency by a factor *a* and adding ΔN particles from an external reservoir. (g)-(i) The transformation consists of adding ΔN particles from an external reservoir and scaling of the box length by a factor *b*. (a), (d), (g) $\Delta\Delta F_{0\rightarrow 1}^{\text{rand}}$ Given otherwise equal parameters, the correction term is identical for transformations leading from single-well to single-well or from single-well to double-well harmonic potential. (c), (f), (i) $\Delta\Delta F_{0\rightarrow 1}^{\text{equi}}$ for transformations with a transition from a single-well to a double-well harmonic potential.

Appendix B: Efficiency of FEP simulations - bias and statistical uncertainty

To avoid confusion, we would like to begin this section by stressing that the correction terms for the single-move FEP schemes derived in Sec. 2.3 of the main article are unrelated to bias or statistical error and the connected issues of imperfect sampling. All the example problems presented in the main article and in the previous section are solved analytically, that is with perfect sampling, giving exact results. Actual problems of interest to the researcher however are mostly not amenable to analytical solutions. Consequently, an estimate of the free energy difference has to be obtained from numerical simulation data which suffers from statistical uncertainty and bias caused by the finite sample size [15–17]. Thus, methods to minimize the inaccuracy of the free energy estimates are of great interest. To pursue this aim, two principle routes are available.

The magnitude of the statistical uncertainty in the standard Zwanzig formalism is mainly determined by the overlap of the occupation probability distributions of the end states [16], and ultimately by the width and overlap of the distributions of energy difference values in forward and backward direction [16, 18]. In our theoretical framework, the overlap measures how likely it is to encounter the environmental subsystem in the same configuration when drawing samples from the equilibrium distribution of the two end states. The first route aims at minimizing the statistical uncertainty of the free energy estimate by optimizing the way the simulation is carried out, that is by maximizing the overlap of the probability distributions. A common method to pursue this route in FEP simulations is staging[19]. Staging breaks up the simulation in multiple steps by introducing fictitious intermediate states - so called stages to increase the overlap between consecutive stages, i.e., their similarity. The theoretical basis of staging is briefly introduced in the next section.

The second route aims at minimizing the statistical uncertainty of a free energy estimate calculated from a given body of simulation data. An example for methods following this route is the Bennett acceptance ratio method [16] sometimes also called Bennett-Pande Method to recognize its rediscovery and popularization by Vijay Pande and coworkers [20]. The Bennett acceptance ratio method can be readily extended to comprise the FEP simulation schemes presented in this work, as will be sketched in Section B 2.

From a theoretical perspective, the distinction between the two routes is not always sharp. The Bennett acceptance ratio method can for example also be interpreted theoretically as introduction of an intermediate state in a position which maximizes overlap between the intermediate and both end states [16, 21–23]. In addition, numerous other, sometimes related methods to improve sam-

pling or minimize the statistical error exist which cannot be reviewed here. The interested reader is referred to the works of Shirts et al. [20, 24, 25], Kofke et al. [22, 25], Götte and Grubmüller [26], Straatsma and McCammon [27] and that of Christ and van Gunsteren [23] for a first approach to this topic.

1. Staging

Instead of directly carrying the system from one end-state to the other the pathway is broken down into $N^{int} + 2$ stages and $M = N^{int} + 1$ steps by introducing N^{int} fictitious intermediates along a transformation coordinate $0 \le \lambda \le 1$ between the end states. Each stage is treated in the same way as the direct perturbation between the end states within the FEP formalism. The total free energy difference is then found by summing up the free energy differences of all the steps

$$\Delta F_{0\to 1} = \sum_{i=1}^{M} \Delta F_i \tag{B.1}$$

In the simplest case, the state energy of an intermediate is linearly interpolated between the end states 0 and 1

$$E_{\lambda} = (1 - \lambda) E_0 + \lambda E_1 \tag{B.2}$$

but other, more complicated definitions can be used as appropriate since the total free energy difference is pathway independent. Likewise, the placement of the intermediates along the transformation coordinate can be chosen freely to optimize the efficiency of the calculation. The configuration space volume of the intermediate states is given by

$$\Psi_{\lambda} = \begin{cases} \Psi^{e}\Psi_{0}^{r} & \text{if } \lambda = 0 \\ \Psi^{e}\Psi_{1}^{r} & \text{if } \lambda = 1 \\ \Psi^{e}\Psi_{0}^{r}\Psi_{1}^{r} & \text{if } 0 < \lambda < 1 \end{cases}$$
(B.3)

The alert reader may have noted, that the correction terms for the single move FEP schemes Eqs. (22) and (30) are not always dimensionless for perturbation steps leading from a physical end state to one of the alchemical intermediate states or vice versa, and that these correction terms cancel in the summation of the perturbation steps in Eq. (B.1). This is as it should be, since the alchemical states possess an artificially increased dimensionality. Only the total free energy difference between the end states is physically meaningful and independent on the chosen definition

for the alchemical energy function, while the free energy profile along the coupling parameter λ as transformation coordinate depends on this definition and thus is not physically meaningful.

2. The Bennett acceptance ratio method

For FEP between states with one reactive subsystem configuration each, Bennett showed [16] that the free energy estimate minimizing the statistical error can be obtained by iteratively solving

$$\sum_{n_{0\to 1}} \frac{1}{1 + \exp\left[\beta \left(E_1 - E_0\right) - C\right]\right)} = \sum_{n_{1\to 0}} \frac{1}{1 + \exp\left[\beta \left(E_1 - E_0\right) + C\right]\right)}$$
(B.4)

for the constant C where the sums run over all $n_{0\to 1}, n_{1\to 0}$ being the number of perturbation samples collected in forward and reverse direction, respectively. From $C, \Delta F_{0\to 1}$ can be calculated via

$$\Delta F_{0\to 1} = \beta^{-1} \left(\ln C - \ln \frac{n_{1\to 0}}{n_{0\to 1}} \right)$$
(B.5)

The standard deviation of the free energy estimate is given by

$$\sigma^{2} = \frac{\overline{f\left[\beta\left(E_{1}-E_{0}-C\right)\right]^{2}}_{0\to1} - \overline{f\left[\beta\left(E_{1}-E_{0}-C\right)\right]^{2}}_{0\to1}}{n_{0\to1}\overline{f\left[\beta\left(E_{1}-E_{0}-C\right)\right]^{2}}_{0\to1}} + \frac{\overline{f\left[\beta\left(E_{0}-E_{1}+C\right)\right]^{2}}_{1\to0} - \overline{f\left[\beta\left(E_{0}-E_{1}+C\right)\right]^{2}}_{1\to0}}{n_{1\to0}\overline{f\left[\beta\left(E_{0}-E_{1}+C\right)\right]^{2}}_{1\to0}}$$
(B.6)

where the overbars denote the arithmetic average, the subscripts of the averages indicate the samples collected in forward and reverse direction as for the numbers of samples n and f is the Fermi function

$$f[x] = \frac{1}{1 + \exp[-\beta x]}$$
 (B.7)

The formalism is extended to comprise the single move FEP schemes defined by Eqs. (23) and (31) by correcting for the relative volume of the configuration spaces that contribute the energy differences. For the random single-move scheme, this relative configuration space volume is $\ln \frac{\Psi_1}{\Psi_0}$. The corresponding relative configuration space volume for the equilibrated single-move scheme is $\ln \frac{\Omega_1}{\Omega_0}$. Accordingly, Eq. (B.5) is modified for the random-single move scheme

$$\Delta F_{0\to 1} = \beta^{-1} \left(\ln C - \ln \frac{n_{1\to 0}}{n_{0\to 1}} - \ln \frac{\Psi_1}{\Psi_0} \right)$$
(B.8)

This expression can be checked for consistency by inserting $\frac{\Psi_1}{\Psi_0}$ times as much samples in reverse direction as in the forward direction into Eq. (B.5).

Inserting Eq. (B.8) into Eq. (B.4) gives after a small rearrangement the iteration formula corresponding to the multi-move FEP scheme defined by Eq. (13) as

$$\sum_{n_{0\to 1}} \frac{1}{1 + \frac{1}{\Psi_{0}} \int \exp\left[\beta \left(E_{1} - E_{0}\right) - C\right] \, \mathrm{d}\mathbf{q}_{1}^{\mathrm{r}})} = \sum_{n_{1\to 0}} \frac{1}{1 + \frac{1}{\Psi_{1}} \int \exp\left[\beta \left(E_{1} - E_{0}\right) + C\right] \, \mathrm{d}\mathbf{q}_{0}^{\mathrm{r}})}$$
(B.9)

The free energy difference is then obtained again from Eq. (B.5) and the corresponding standard deviation of the free energy estimate is given by

$$\sigma^{2} = \frac{\overline{f\left[\beta\left(E_{1}-E_{0}-C\right)\right]^{2}}_{0\to1} - \overline{f\left[\beta\left(E_{1}-E_{0}-C\right)\right]^{2}}_{0\to1}}{n_{0\to1}\Psi_{0}\overline{f\left[\beta\left(E_{1}-E_{0}-C\right)\right]^{2}}_{0\to1}} + \frac{\overline{f\left[\beta\left(E_{0}-E_{1}+C\right)\right]^{2}}_{1\to0} - \overline{f\left[\beta\left(E_{0}-E_{1}+C\right)\right]^{2}}_{1\to0}}{n_{1\to0}\Psi_{1}\overline{f\left[\beta\left(E_{0}-E_{1}+C\right)\right]^{2}}_{1\to0}}$$
(B.10)

where the averages are taken over the entire space $n_{1\to 0}\Psi_1$ of collected state energy difference samples.

Recently, extended versions of the Bennett acceptance ratio method were proposed by two groups. The method of Maragakis et al. [28] includes collected data for multiple transformations sharing a common reference end state in a simultaneous maximum-entropy estimation of all free energy differences. The complementary method of Shirts and coworkers includes all the intermediate steps of a FEP simulation with staging simultaneously to obtain the maximum-likelihood estimate of the free energy difference [25]. Both extensions offer improved results relative to the original method. It seems likely that these extended methods can also be applied to the FEP simulation schemes presented in the present paper with minor modifications.

Appendix C: Relation between Boltzmann entropy and Süßmann entropy

In this section, we want to carry on the discussion of the relation between the canonical Boltzmann entropy and the Süßmann entropy begun in Sec. 2.4 of the main article. By examining Figure 2(b), we can observe that Ω and $\tilde{\Omega}$ adopt equal values in the extreme limits of $K \to 0$ and $K \to \infty$ and that they differ in the intermediate range in which the state energy E of the occupied microstate can differ. From comparison of Eqs. (40) and (41), it can be seen that $S = \tilde{S}$ if the



FIG. 6. Possible interpretation of the difference between Boltzmann and Süßmann entropy as measure for the range of thermally accessible microstate energies. See text for detailed description of the system. **left:** The difference $TS - T\tilde{S}$ is plotted for three different arrangements of the microstates along the coordinate as a function of the harmonic potential strenghth *K*. **right:** The three different arrangements of the microstates along the coordinate **q**. Each circle represents a microstate. The colors of the circles are the same as for the corresponding graphs on the left. In the bottom arrangement (black), the microstates are evenly distributed along the coordinate. In the middle arrangement (red), the microstates are all found in the center of the coordinate range. In the arrangement shown at the top (green), one microstate is found at the center of the coordinate range while all other microstates are found at the minimum or maximum of the coordinate range.

exponential average of the state energy equals the arithmetic average. This condition is only met in the microcanonical ensemble or a degenerate system in which the state energy is also constant, since generally $-\beta^{-1} \ln \langle \exp[-\beta E] \rangle \leq \langle E \rangle$. The difference between Boltzmann entropy Eq. (40) and Süßmann entropy Eq. (41) is given by

$$T\left(S-\tilde{S}\right) = \langle E \rangle + \beta^{-1} \ln \int \rho \exp\left[-\beta E\right] d\mathbf{q}$$
 (C.1)

where we wrote the exponential average of Eq. (41) explicitly as integral. Next, we write the arithmetic average in the form $\langle E \rangle = -\beta^{-1} \ln \exp \left[-\beta \langle E \rangle\right]$

$$T\left(S-\tilde{S}\right) = -\beta^{-1}\ln\exp\left[-\beta\left\langle E\right\rangle\right] + \beta^{-1}\ln\int\rho\exp\left[-\beta E\right]\mathrm{d}\mathbf{q}$$
(C.2)

By exploiting the relation $\ln x - \ln y = \ln \frac{x}{y}$, we can write

$$T\left(S - \tilde{S}\right) = \beta^{-1} \ln \frac{\int \rho \exp\left[-\beta E\right] d\mathbf{q}}{\exp\left[-\beta \left\langle E\right\rangle\right]}$$
(C.3)

Finally, we can write Eq.(C.3) as

$$T\left(S - \tilde{S}\right) = \beta^{-1} \ln \int \rho \exp\left[-\beta \left(E - \langle E \rangle\right)\right] d\mathbf{q}$$

= $\beta^{-1} \ln \left\langle \exp\left[-\beta \left(E - \langle E \rangle\right)\right] \right\rangle$ (C.4)

From Eq. (C.4), a possible interpretation of the difference between the Boltzmann entropy and the Süßmann entropy could be proposed as a measure of the interval width over which the state energy can vary between samples drawn from the equilibrium distribution, i.e., a measure of the thermal fluctuations of the state energy. Thus, $T(S - \tilde{S})$ can be expected to be on the order of k_BT . Alternatively, we could interpret this difference as measure of the uncertainty about the present value of the state energy as opposed to the interpretation of the Süßmann entropy as uncertainty in the position of the system in configuration space. Accordingly, the Boltzmann entropy would measure the uncertainty in all microscopic system parameters at the same time. Here, uncertainty is used in the same meaning as that used by Jaynes in the sense of incomplete knowledge about the current state of the system [29]. That is we know the possible microstates, their respective energies and the resulting occupation probability distribution, but we can not know in before the exact identity of a microstate sampled from the equilibrium distribution and thus also not the state energy of the drawn sample.

In the remaining part of this section, a variant of the example system from Sec. 3.1 of the main article is used to obtain a qualitative picture of how the difference between Boltzmann and Süßmann entropy in depends on the energy landscape of a system. The coordinate q might be conceived as a one-dimensional projection of a more complex configuration space. Consequently the microstates might be arranged in a way that differs from an even distribution along q. Fig. 6 shows the difference between Boltzmann and Süßmann entropies for three different arrangements of the 21 microstates along q. For the arrangement shown on the middle right scheme, there is no difference between the entropies, because all microstates are equivalent. Consequently the exponential and arithmetic averages of the microstate energy E are equal for any potential strength and the range of possible energy values is infinitely narrow. The linear arrangement in the bottom right scheme is the same as in Sec. 3.1. The difference between the entropies is small for low and

high potential strengths but significantly positive at intermediate values of K. This behavior can be explained as follows. At low potential strength, the microstates have nearly the same energy so that the range of possible values of the microstate energy is very narrow. This results in a small difference between the entropy measures. At high potential strength, essentially only the central microstate with the lowest energy is thermally accessible so that the interval width of accessible microstate energies is again effectively equal to zero. At intermediate potential strength, several microstates of different energy can be populated and the interval width of thermally accessible microstate energies is non-zero. A more extreme variant of the linear arrangement is used in the arrangement shown in the top-right scheme (green). The corresponding entropy difference shows a sharper extremum at intermediate values of the potential strength. For lower potential strengths, the microstates have all very similar energies resulting in a narrow interval of thermally accessible microstate energies and a small difference between the entropies. At higher potential strengths, the energies of the outer microstates is so high in relation to the energy of the central microstate that the outer microstates are not thermally accessible. Consequently the central microstate is the only significantly populated microstate resulting in an interval of thermally accessible energy interval of effectively zero width. The maximum difference is higher than for the linear arrangement because all states except the central one are concentrated at the extremes of the coordinate range. A higher number of microstates at a certain energy level entropically favors its population. This favoring can compensate for unfavorably high energies and widen the range of microstate energies effectively accessible to a system. This mechanism can lead to much higher differences between Boltzmann and Süßmann entropies than for the example system shown here. In larger systems, the difference can reach several $k_B T$.

In conclusion, the qualitative picture of the dependence of the difference between Boltzmann and Süßmann entropy on the energy landscape of a system supports our proposed interpretation of this difference as measure of the interval width over which the microstate energy can vary in an ensemble.

- [2] H.-X. Zhou and M. K. Gilson, Chem. Rev. 109, 4092 (2009).
- [3] K. S. Pitzer, J. Chem. Phys. 10, 428 (1942); 14, 239 (1946).

^[1] T. L. Hill, An Introduction to Statistical Thermodynamics (Dover Publ. Inc., New York, 1960).

- [4] D. R. Herschbach, H. S. Johnston, and D. Rapp, J. Chem. Phys. 31, 1652 (1959).
- [5] N. Gō and H. A. Scheraga, J. Chem. Phys. 31, 1652 (1976).
- [6] S. Boresch and M. Karplus, J. Chem. Phys. 105, 5145 (1996).
- [7] M. J. Potter and M. K. Gilson, J. Phys. Chem. A 106, 563 (2002).
- [8] L. R. Dodd, T. D. Boone, and D. N. Theodorou, Molec. Phys. 78, 961 (1993).
- [9] D. Hoffmann and E. W. Knapp, Phys. Rev. E 53, 4221 (1996); Eur. Biophys. J. 24, 387 (1996).
- [10] J. P. Ulmschneider and W. L. Jorgensen, J. Chem. Phys. 118, 4261 (2003).
- [11] G. Wedler, Lehrbuch der Physikalischen Chemie, 4th ed. (Wiley-VCH, 1997).
- [12] E. T. Jaynes, "The gibbs paradox," in *Maximum Entropy and Bayesian Methods, Seattle, 1991* (Kluwer Academic Publishers, 1992) Chap. 1, pp. 1–22, http://bayes.wustl.edu/etj/etj.html.
- [13] G. Süßmann, Zeitschrift für Naturforschung A 52a, 49 (1997).
- [14] M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, Biophys. J. 72, 1047 (1997).
- [15] R. H. Wood, W. C. F. Mühlbauer, and P. T. Thompson, J. Chem. Phys. 95, 6670 (1991).
- [16] C. H. Bennett, J. Comp. Phys. 22, 245 (1976).
- [17] A. Pohorille, C. Jarzynski, and C. Chipot, J. Phys. Chem. B 114, 10235 (2010).
- [18] D. M. Ferguson, J. Chem. Phys. 99, 10086 (1993).
- [19] J. P. Valleau and D. N. Card, J. Chem. Phys. 57, 5457 (1972).
- [20] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, Phys. Rev. Lett 91, 140601 (2003).
- [21] K.-K. Han, Phys. Lett. A 165, 28 (1992).
- [22] N. Lu, J. K. Singh, and D. A. Kofke, J. Chem. Phys. 118, 2977 (2003).
- [23] C. D. Christ and W. van Gunsteren, J. Chem. Phys. 126, 184110 (2007); 128, 174112 (2008).
- [24] M. R. Shirts and V. Pande, J. Chem. Phys. 122, 134508 (2005).
- [25] M. R. Shirts and J. D. Chodera, J. Chem. Phys. 129, 124105 (2008); N. Lu, D. A. Kofke, and T. B. Woolf, J. Comp. Chem. 25, 28 (2003); D. Wu and D. A. Kofke, J. Chem. Phys. 123, 054103 (2005); 123, 084109 (2005).
- [26] M. Götte and H. Grubmüller, J. Comp. Chem. 30, 447 (2001).
- [27] T. P. Straatsma and J. A. McCammon, J. Chem. Phys. 90, 3300 (1989).
- [28] P. Maragakis, M. Spichty, and M. Karplus, Phys. Rev. Lett 96, 100602 (2006).
- [29] E. T. Jaynes, Phys. Rev. 106, 620 (1957); 108, 171 (1957); "Information Theory and Statistical Mechanics," in *Brandeis University Summer Institute Lectures in Theoretical Physics*, Vol. 3 (W. A. Benjamin, New York, 1963) pp. 181–218.