

## S1 Derivation of ensemble ratios

Following Ref. [1] the analytical ratio between observable distributions at two different state points can be used to verify the sampled ensemble. Specifically, the phase-space probability densities of the canonical (NVT), the isothermal-isobaric (NPT), and the grand-canonical ( $\mu$ VT) have similar mathematical forms,

$$P(\mathbf{r}, \mathbf{p} \mid \beta) \sim e^{-\beta \mathcal{H}(\mathbf{r}, \mathbf{p})} \quad \text{NVT} \quad (1)$$

$$P(\mathbf{r}, \mathbf{p}, V \mid \beta, P) \sim e^{-\beta \mathcal{H}(\mathbf{r}, \mathbf{p}) - PV} \quad \text{NPT} \quad (2)$$

$$P(\mathbf{r}, \mathbf{p}, \mathbf{N} \mid \beta, \boldsymbol{\mu}) \sim e^{-\beta \mathcal{H}(\mathbf{r}, \mathbf{p}) + \sum_i \beta \mu_i N_i} \quad \mu\text{VT}, \quad (3)$$

where  $P(a \mid b)$  denotes the probability of microstate  $a$  in the ensemble defined by parameters  $b$ . The microstate is defined by the positions  $\mathbf{r}$  and momenta  $\mathbf{p}$  of the atoms in the system, and potentially the volume  $V$  (for NPT) and the number of particles  $N_i$  for each species  $i$  (for  $\mu$ VT). The ensemble of the physical system is defined by the temperature  $T$ , the pressure  $P$ , the chemical potentials  $\mu_i$  and the functional form of the Hamiltonian  $\mathcal{H}(\mathbf{r}, \mathbf{p})$ .

The probability of observing a specific total energy  $E = \mathcal{H}(\mathbf{r}, \mathbf{p})$  (in NVT), a specific total energy  $E$  along with a specific volume  $V$  (in NPT) or a specific total energy  $E$  along with a specific species composition  $\mathbf{N}$  (in  $\mu$ VT) is then given by

$$P(E \mid \beta) \sim \Omega(E) e^{-\beta E} \quad \text{NVT} \quad (4)$$

$$P(E, V \mid \beta, P) \sim \Omega(E, V) e^{-\beta E - PV} \quad \text{NPT} \quad (5)$$

$$P(E, \mathbf{N} \mid \beta, \boldsymbol{\mu}) \sim \Omega(E, \mathbf{N}) e^{-\beta E + \sum_i \beta \mu_i N_i} \quad \mu\text{VT}, \quad (6)$$

where  $\Omega$  denotes the density of states, which is generally a non-trivial, system-dependent quantity. The presence of  $\Omega$  explains why no direct statistical tests for the probability distribution of configurational quantities exist.

$\Omega$  is, however, independent from the state parameters of the ensemble, as can be verified by looking at at Eqs. 4 – 6. The ratios of the probability densities of observables of the identical system at different state points are hence independent of  $\Omega$ ,

$$\frac{P(E \mid \beta_2)}{P(E \mid \beta_1)} \sim e^{-(\beta_2 - \beta_1)E} \quad \text{NVT} \quad (7)$$

$$\frac{P(E, V \mid \beta_2, P_2)}{P(E, V \mid \beta_1, P_1)} \sim e^{-(\beta_2 - \beta_1)E - (\beta_2 P_2 - \beta_1 P_1)V} \quad \text{NPT} \quad (8)$$

$$\frac{P(E, \mathbf{N} \mid \beta_2, \boldsymbol{\mu}_2)}{P(E, \mathbf{N} \mid \beta_1, \boldsymbol{\mu}_1)} \sim e^{-(\beta_2 - \beta_1)E + \sum_i (\beta_2 \mu_{2,i} - \beta_1 \mu_{1,i}) N_i} \quad \mu\text{VT}. \quad (9)$$

Taking the logarithm of these ratios, it becomes obvious that in all cases, a linear dependence of the observables exist:

$$\log \frac{P(E \mid \beta_2)}{P(E \mid \beta_1)} \sim -(\beta_2 - \beta_1)E \quad \text{NVT} \quad (10)$$

$$\log \frac{P(E, V \mid \beta_2, P_2)}{P(E, V \mid \beta_1, P_1)} \sim -(\beta_2 - \beta_1)E - (\beta_2 P_2 - \beta_1 P_1)V \quad \text{NPT} \quad (11)$$

$$\log \frac{P(E, \mathbf{N} \mid \beta_2, \boldsymbol{\mu}_2)}{P(E, \mathbf{N} \mid \beta_1, \boldsymbol{\mu}_1)} \sim -(\beta_2 - \beta_1)E + \sum_i (\beta_2 \mu_{2,i} - \beta_1 \mu_{1,i}) N_i \quad \mu\text{VT}. \quad (12)$$

## References

1. Shirts MR. Simple Quantitative Tests to Validate Sampling from Thermodynamic Ensembles. Journal of Chemical Theory and Computation. 2013;9(2):909–926. doi:10.1021/ct300688p.