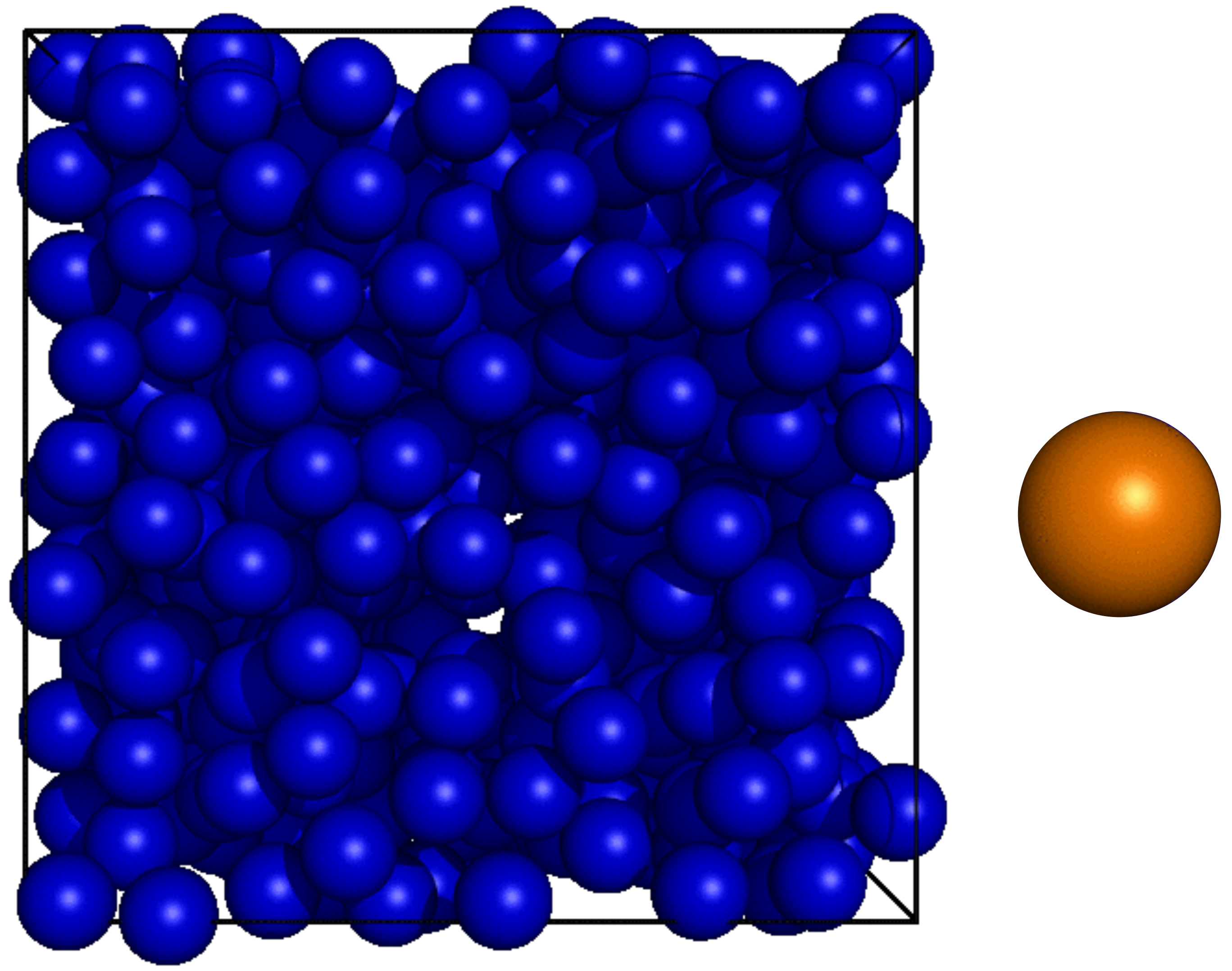
SOLVATION FREE ENERGIES WITH NEURAL TI

Bálint Máté, François Fleuret, Tristan Bereau

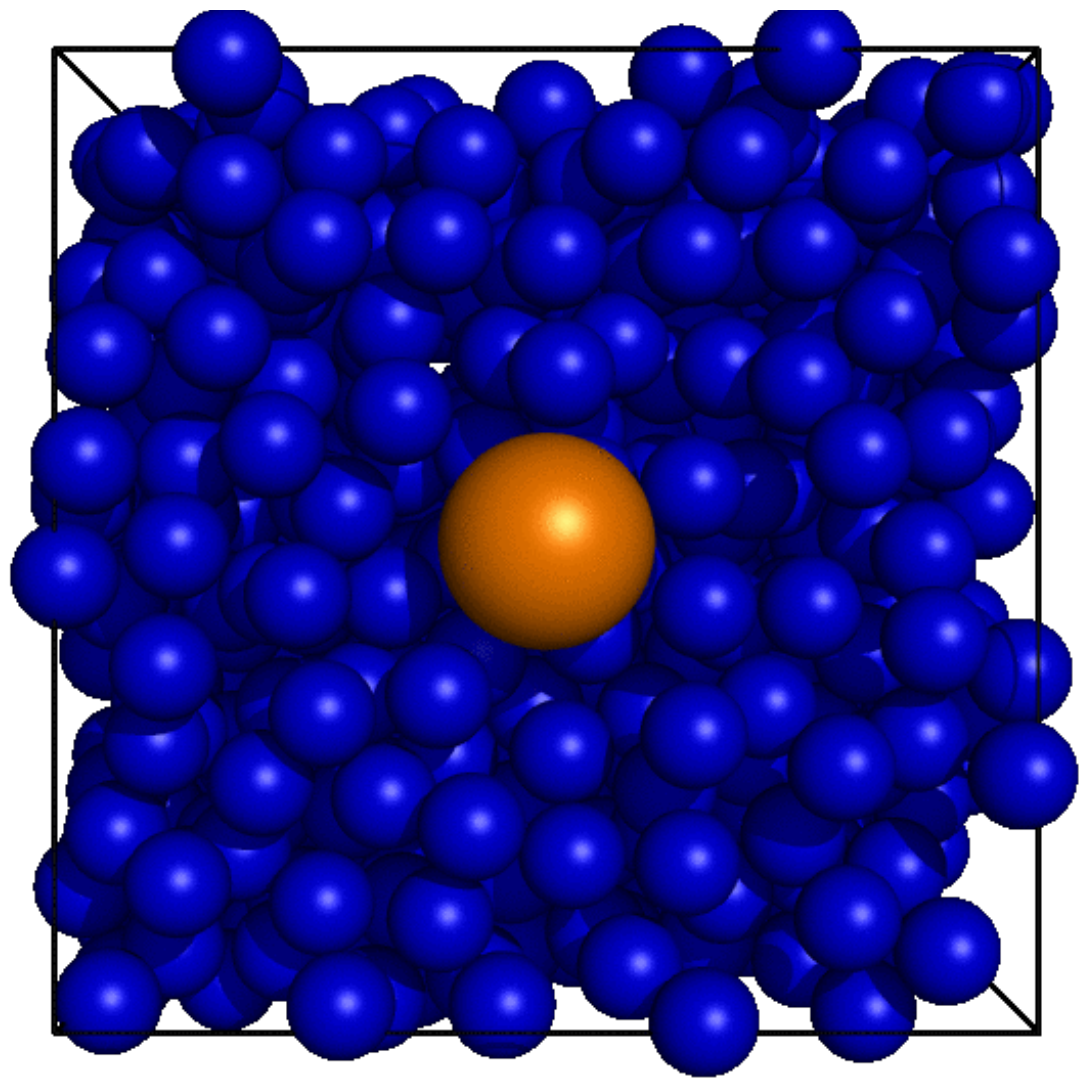
23/10/2024

arXiv:2410.15815

MOTIVATION



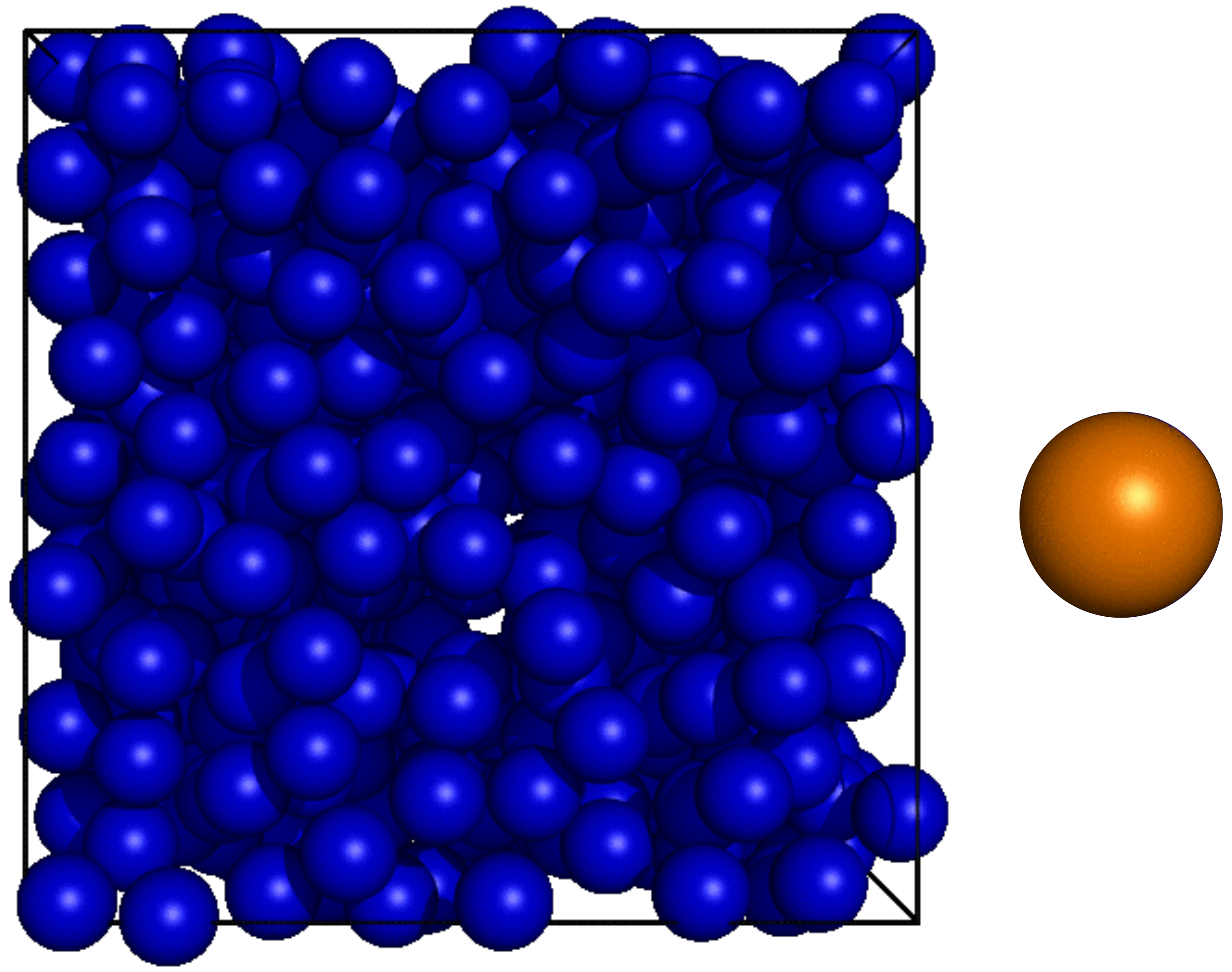
$$U_{\text{solvent}} + U_{\text{solute}}$$



$$U_{\text{solvent}} + U_{\text{solute}} + U_{\text{solute-solvent}}$$

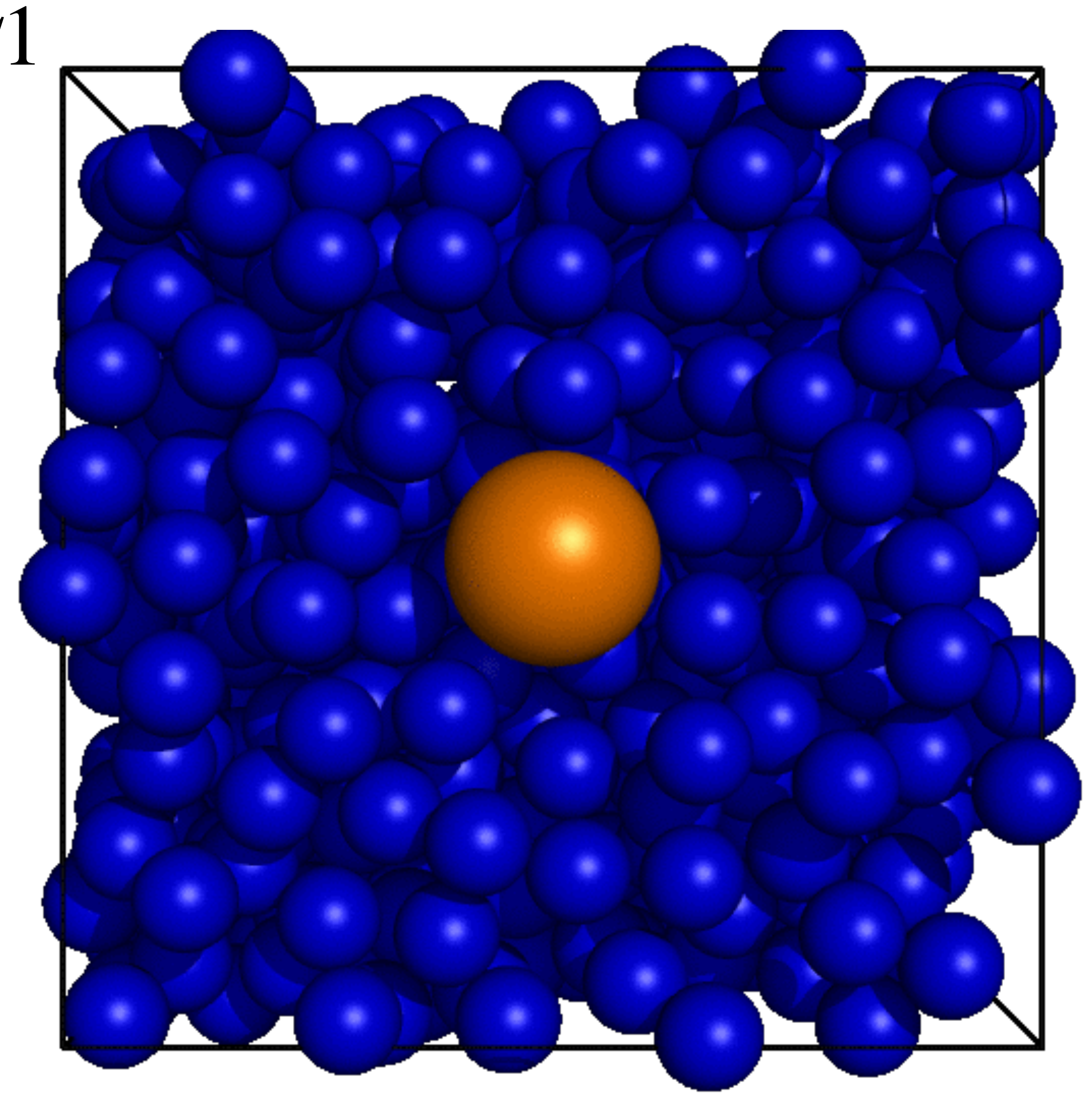
MOTIVATION

$$\frac{1}{Z_0} e^{-\beta(U_{\text{solvent}} + U_{\text{solute}})}$$



$$U_{\text{solvent}} + U_{\text{solute}}$$

$$\frac{1}{Z_1} e^{-\beta(U_{\text{solvent}} + U_{\text{solute}} + U_{\text{solute-solvent}})}$$

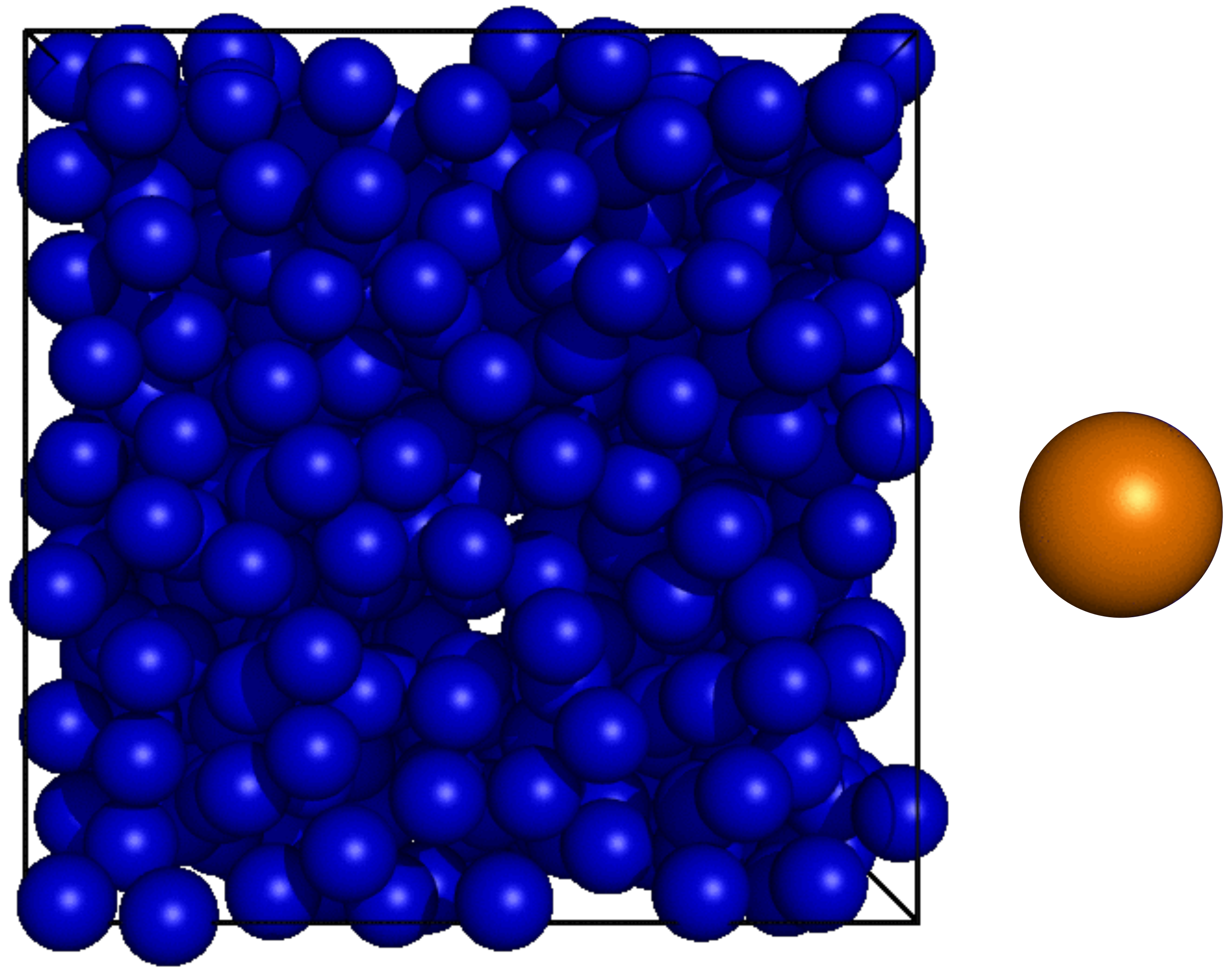


$$U_{\text{solvent}} + U_{\text{solute}} + U_{\text{solute-solvent}}$$

MOTIVATION

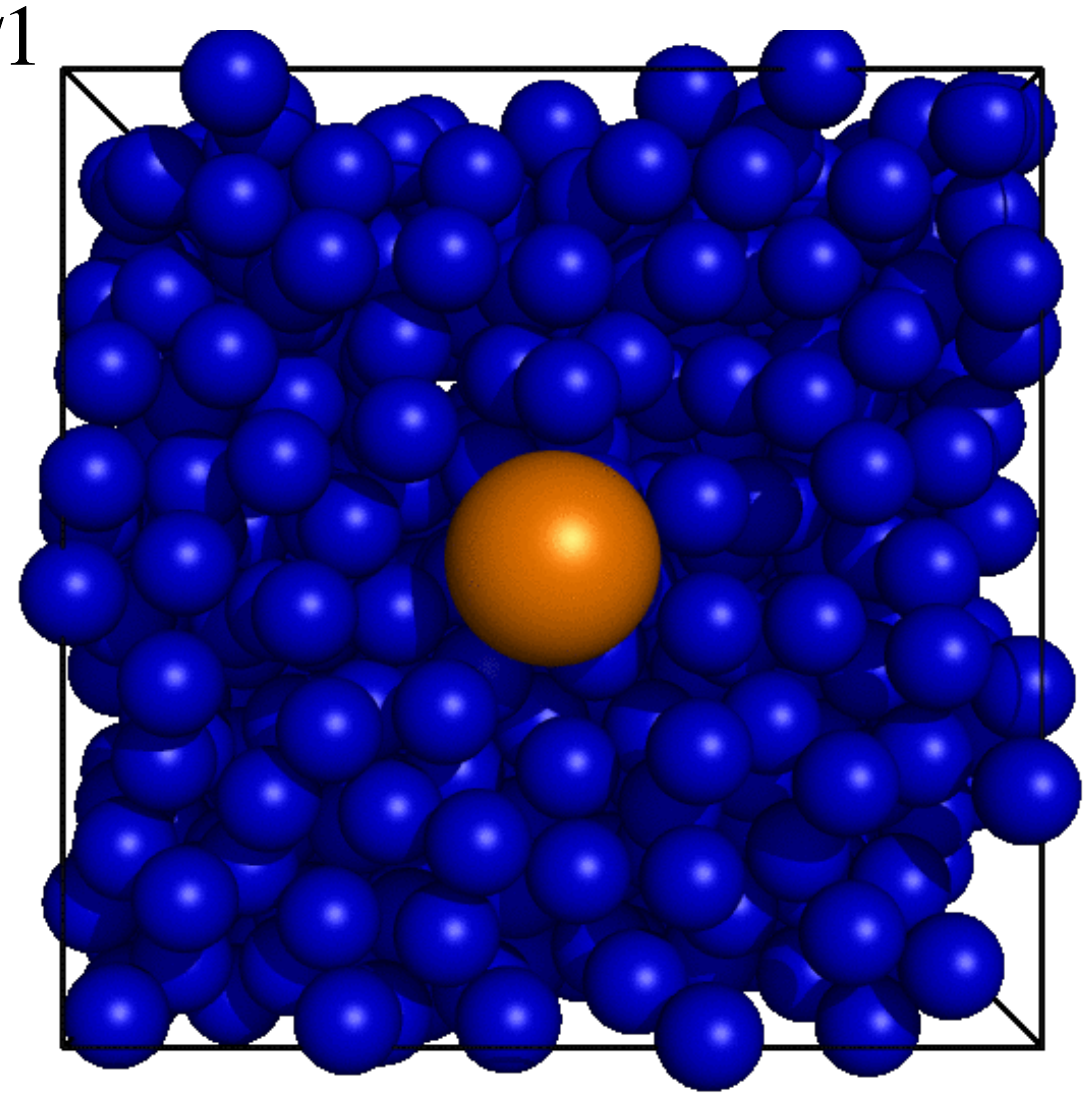
$$\Delta F_{0\to 1} = \beta^{-1}(\log Z_0 - \log Z_1) = ?$$

$$\frac{1}{Z_0} e^{-\beta(U_{\text{solvent}} + U_{\text{solute}})}$$



$$U_{\text{solvent}} + U_{\text{solute}}$$

$$\frac{1}{Z_1} e^{-\beta(U_{\text{solvent}} + U_{\text{solute}} + U_{\text{solute-solvent}})}$$



$$U_{\text{solvent}} + U_{\text{solute}} + U_{\text{solute-solvent}}$$

THERMODYNAMIC INTEGRATION

$$\Delta F_{0\to 1} = \int_0^1 d\lambda \langle \partial_\lambda U \rangle_\lambda$$

THERMODYNAMIC INTEGRATION

$$\Delta F_{0\to 1} = \int_0^1 d\lambda \langle \partial_\lambda U \rangle_\lambda$$

- Two components:

THERMODYNAMIC INTEGRATION

$$\Delta F_{0\to 1} = \int_0^1 d\lambda \langle \partial_\lambda U \rangle_\lambda$$

- Two components:
    - interpolating family of potentials

THERMODYNAMIC INTEGRATION

$$\Delta F_{0 \to 1} = \int_0^1 d\lambda \langle \partial_\lambda U \rangle_\lambda$$

- Two components:
  - interpolating family of potentials
  - samples from the equilibrium distributions of the interpolating potentials

THERMODYNAMIC INTEGRATION

$$\Delta F_{0\to 1} = \int_0^1 d\lambda \langle \partial_\lambda U \rangle_\lambda$$

- Two components:
    - interpolating family of potentials
    - samples from the equilibrium distributions of the interpolating potentials
- standard TI [Kirkwood, 1935]:

THERMODYNAMIC INTEGRATION

$$\Delta F_{0\to 1} = \int_0^1 d\lambda \langle \partial_\lambda U \rangle_\lambda$$

- Two components:
  - interpolating family of potentials
  - samples from the equilibrium distributions of the interpolating potentials
- standard TI [Kirkwood, 1935]:
  - start with an interpolation of potentials

THERMODYNAMIC INTEGRATION

$$\Delta F_{0\to 1} = \int_0^1 d\lambda \langle \partial_\lambda U \rangle_\lambda$$

- Two components:
    - interpolating family of potentials
    - samples from the equilibrium distributions of the interpolating potentials
- standard TI [Kirkwood, 1935]:
    - start with an interpolation of potentials
    - sample from the equilibrium distribution of those potentials

THERMODYNAMIC INTEGRATION

$$\Delta F_{0\to 1} = \int_0^1 d\lambda \langle \partial_\lambda U \rangle_\lambda$$

- Two components:
  - interpolating family of potentials
  - samples from the equilibrium distributions of the interpolating potentials
- standard TI [Kirkwood, 1935]:
  - start with an interpolation of potentials
  - sample from the equilibrium distribution of those potentials
- Neural TI [Máté, Fleuret, Bereau, 2024]:

THERMODYNAMIC INTEGRATION

$$\Delta F_{0\to 1} = \int_0^1 d\lambda \langle \partial_\lambda U \rangle_\lambda$$

- Two components:
    - interpolating family of potentials
    - samples from the equilibrium distributions of the interpolating potentials
- standard TI [Kirkwood, 1935]:
    - start with an interpolation of potentials
    - sample from the equilibrium distribution of those potentials
- Neural TI [Máté, Fleuret, Bereau, 2024]:
    - start with a way of generating intermediate samples

THERMODYNAMIC INTEGRATION

$$\Delta F_{0\to 1} = \int_0^1 d\lambda \langle \partial_\lambda U \rangle_\lambda$$

- Two components:
  - interpolating family of potentials
  - samples from the equilibrium distributions of the interpolating potentials
- standard TI [Kirkwood, 1935]:
  - start with an interpolation of potentials
  - sample from the equilibrium distribution of those potentials
- Neural TI [Máté, Fleuret, Bereau, 2024]:
  - start with a way of generating intermediate samples
  - learn the corresponding equilibrium potentials

HOW DO WE GENERATE INTERMEDIATE SAMPLES?

Assuming that we have access to samples from the endpoint distributions, $x_0 \sim \rho_0, x_1 \sim \rho_1$.

HOW DO WE GENERATE INTERMEDIATE SAMPLES?

Assuming that we have access to samples from the endpoint distributions, $x_0 \sim \rho_0, x_1 \sim \rho_1$.

- Stochastic interpolants [Albergo et al.]

HOW DO WE GENERATE INTERMEDIATE SAMPLES?

Assuming that we have access to samples from the endpoint distributions, $x_0 \sim \rho_0, x_1 \sim \rho_1$.

- Stochastic interpolants [Albergo et al.]
  - Example to keep in mind:

HOW DO WE GENERATE INTERMEDIATE SAMPLES?

Assuming that we have access to samples from the endpoint distributions, $x_0 \sim \rho_0, x_1 \sim \rho_1$.

- Stochastic interpolants [Albergo et al.]
    - Example to keep in mind:
        - $x_t = (1 - t)x_0 + tx_1 + \sigma(t)z$

HOW DO WE GENERATE INTERMEDIATE SAMPLES?

Assuming that we have access to samples from the endpoint distributions, $x_0 \sim \rho_0, x_1 \sim \rho_1$.

- Stochastic interpolants [Albergo et al.]
  - Example to keep in mind:
    - $x_t = (1 - t)x_0 + tx_1 + \sigma(t)z$
    - By definition, $x_t \sim \rho_t(x)$ when $x_0 \sim \rho_0, x_1 \sim \rho_1, z \sim \mathcal{N}(0,1)$

HOW DO WE GENERATE INTERMEDIATE SAMPLES?

Assuming that we have access to samples from the endpoint distributions, $x_0 \sim \rho_0, x_1 \sim \rho_1$.

- Stochastic interpolants [Albergo et al.]
  - Example to keep in mind:
    - $x_t = (1-t)x_0 + tx_1 + \sigma(t)z$
    - By definition, $x_t \sim \rho_t(x)$ when $x_0 \sim \rho_0, x_1 \sim \rho_1, z \sim \mathcal{N}(0,1)$
  - We can learn $\nabla \log \rho_t$ with Denoising Score Matching (DSM) if $\sigma(t) > 0$

HOW DO WE GENERATE INTERMEDIATE SAMPLES?

Assuming that we have access to samples from the endpoint distributions, $x_0 \sim \rho_0, x_1 \sim \rho_1$.

- Stochastic interpolants [Albergo et al.]
  - Example to keep in mind:
    - $x_t = (1-t)x_0 + tx_1 + \sigma(t)z$
    - By definition, $x_t \sim \rho_t(x)$ when $x_0 \sim \rho_0, x_1 \sim \rho_1, z \sim \mathcal{N}(0,1)$
  - We can learn $\nabla \log \rho_t$ with Denoising Score Matching (DSM) if $\sigma(t) > 0$
    - To do TI, we will need an energy-based model, $-\nabla U_t^\theta \approx \nabla \log \rho_t$

MAKING USE OF THE TEMPORAL DERIVATIVE

MAKING USE OF THE TEMPORAL DERIVATIVE

- TI relies on the temporal derivative: $\Delta F_{0\rightarrow 1} = \int_0^1 dt \langle \partial_t U(t, x) \rangle_t$

MAKING USE OF THE TEMPORAL DERIVATIVE

- TI relies on the temporal derivative: $\Delta F_{0\to 1} = \int_0^1 dt \langle \partial_t U(t,x) \rangle_t$
- During training we only need the spatial gradient $\nabla_x U(t,x)$

MAKING USE OF THE TEMPORAL DERIVATIVE

- TI relies on the temporal derivative: $\Delta F_{0\rightarrow 1} = \int_0^1 dt \langle \partial_t U(t,x) \rangle_t$
- During training we only need the spatial gradient $\nabla_x U(t,x)$
- 2 roles for $\partial_t U(t,x)$:

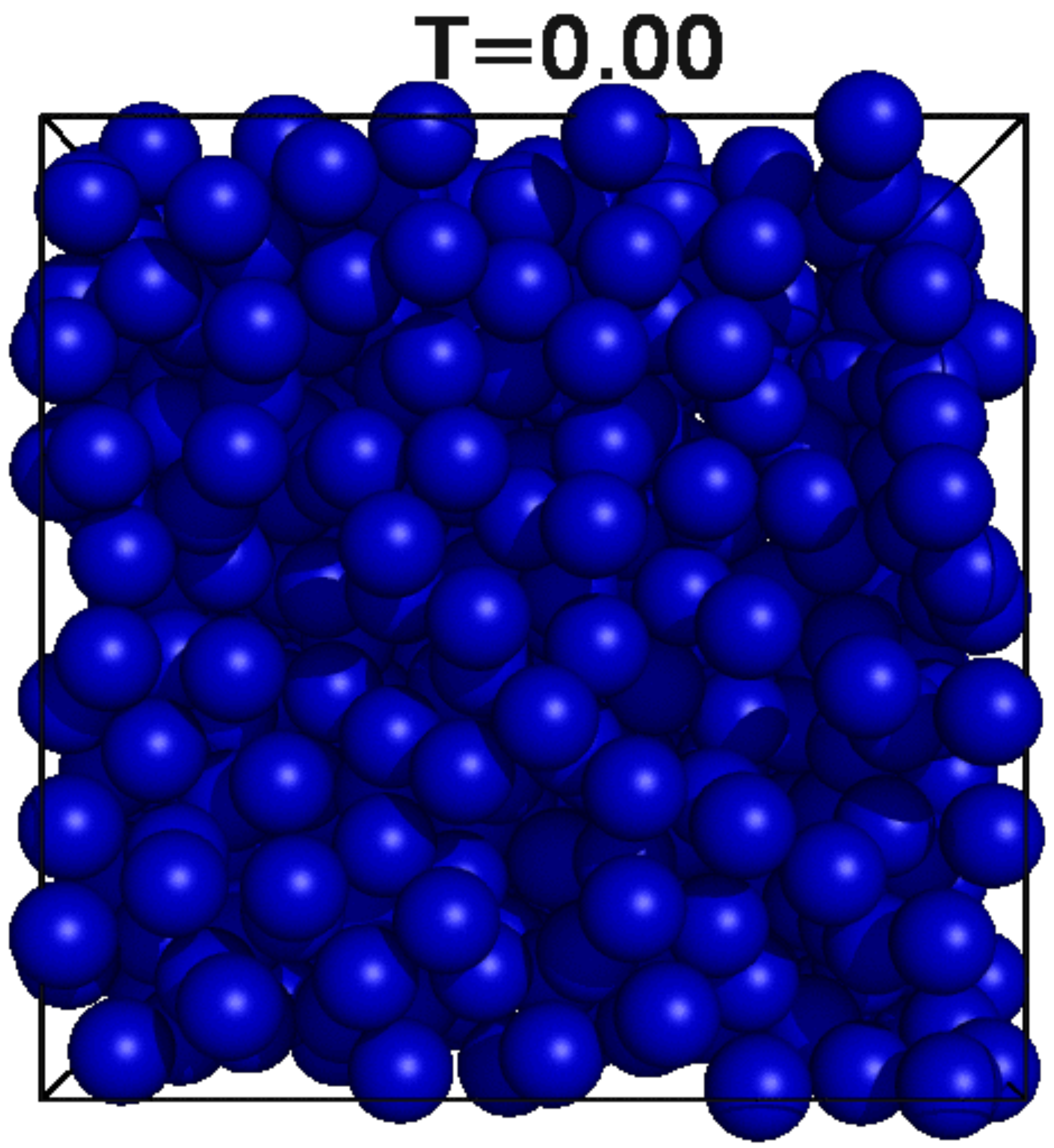
MAKING USE OF THE TEMPORAL DERIVATIVE

- TI relies on the temporal derivative: $\Delta F_{0\to1} = \int_0^1 dt \langle \partial_t U(t,x) \rangle_t$
- During training we only need the spatial gradient $\nabla_x U(t,x)$
- 2 roles for $\partial_t U(t,x)$:
    - rolling estimate of the free energy difference $\Delta \hat{F}_{0\to1} = \mathbb{E}_{t\sim U([0,1]), x\sim\rho_t}\left[\partial_t U(t,x)\right]$

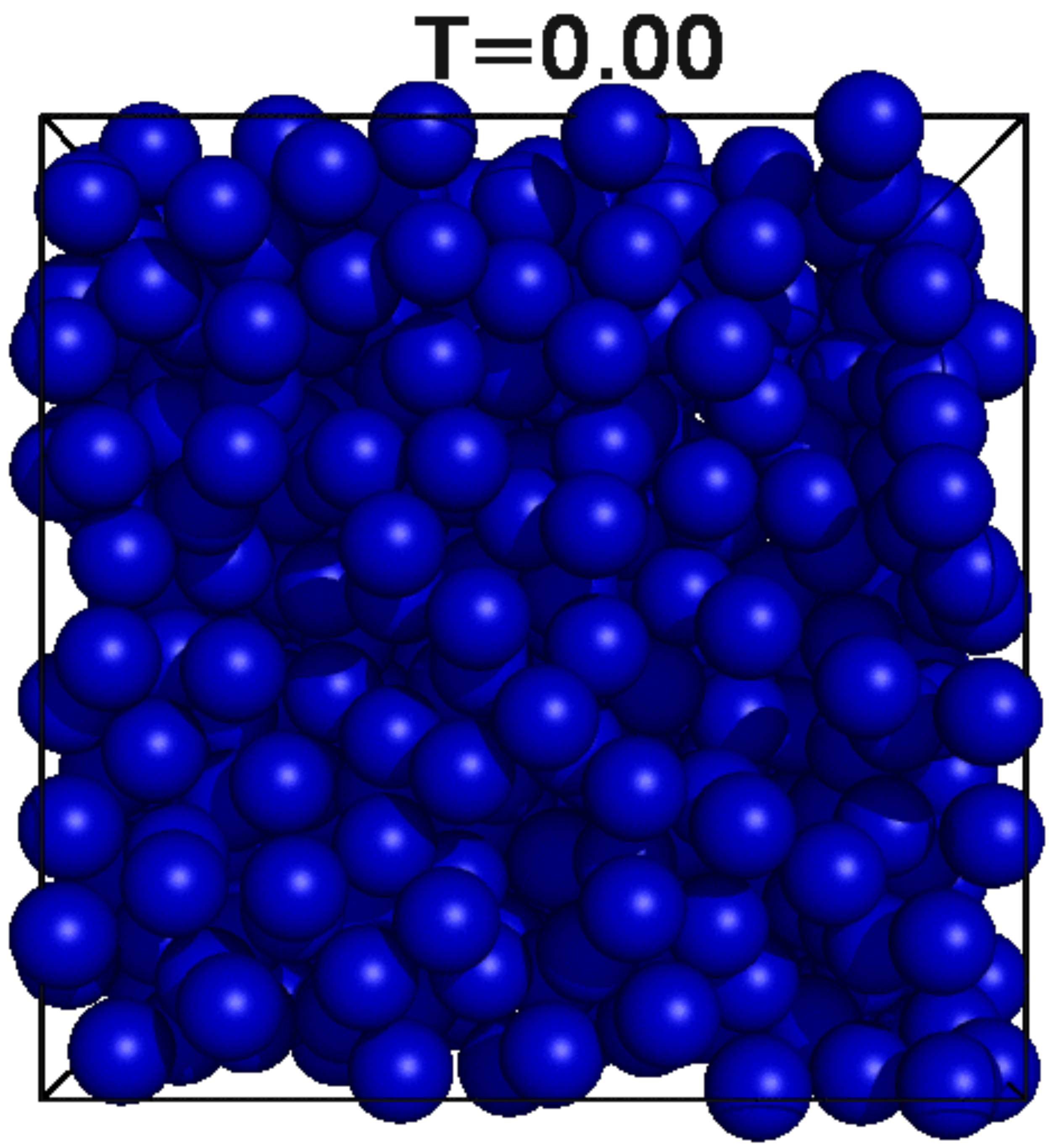
MAKING USE OF THE TEMPORAL DERIVATIVE

- TI relies on the temporal derivative: $\Delta F_{0\to 1} = \int_0^1 dt \langle \partial_t U(t,x) \rangle_t$
- During training we only need the spatial gradient $\nabla_x U(t,x)$
- 2 roles for $\partial_t U(t,x)$:
  - rolling estimate of the free energy difference $\Delta \hat{F}_{0\to 1} = \mathbb{E}_{t\sim U([0,1]), x\sim\rho_t}\left[\partial_t U(t,x)\right]$
  - Carefully regularizing $\partial_t U(t,x)^2$ to prevent it from exploding

EXPERIMENT 1

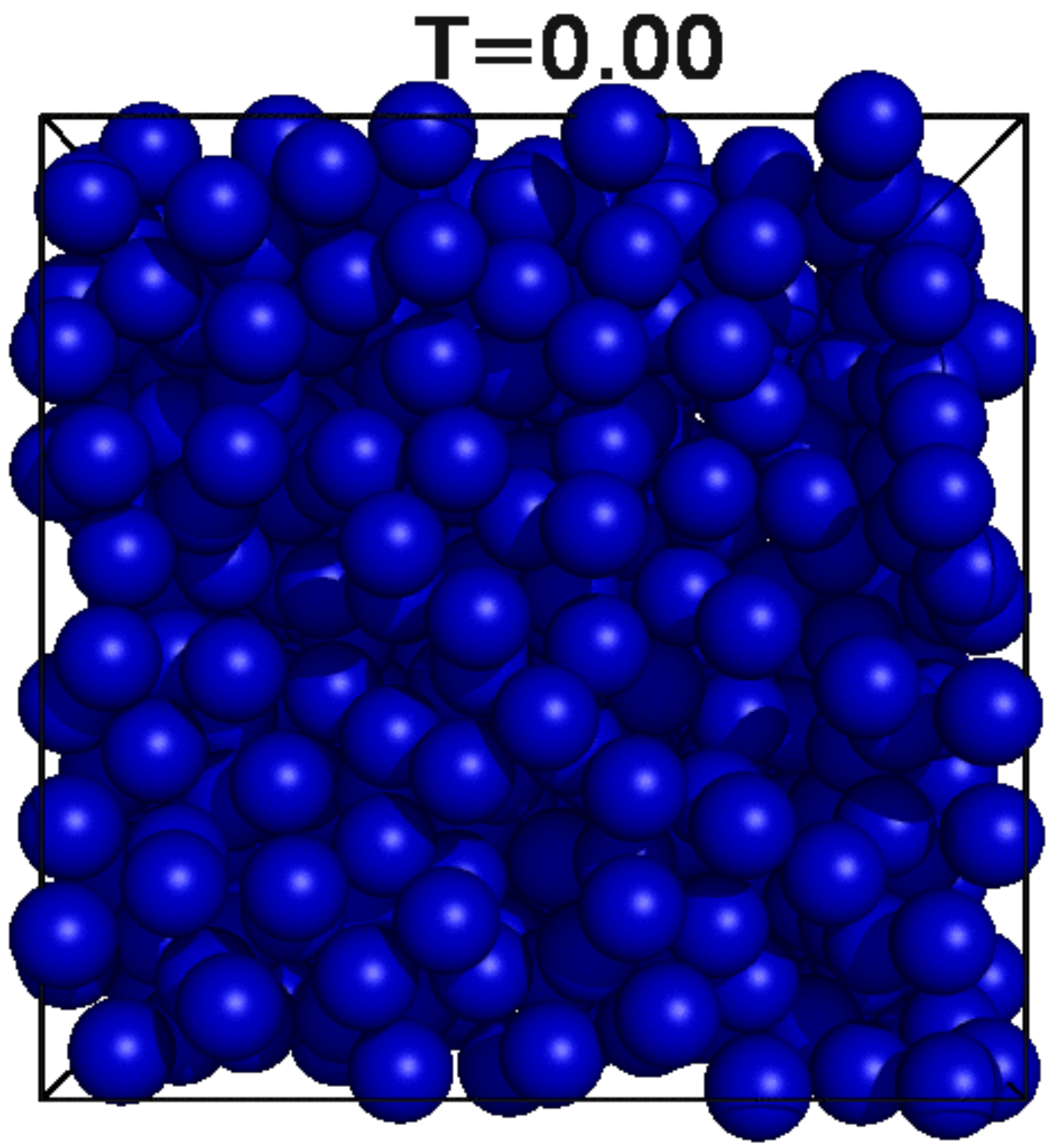


EXPERIMENT 1



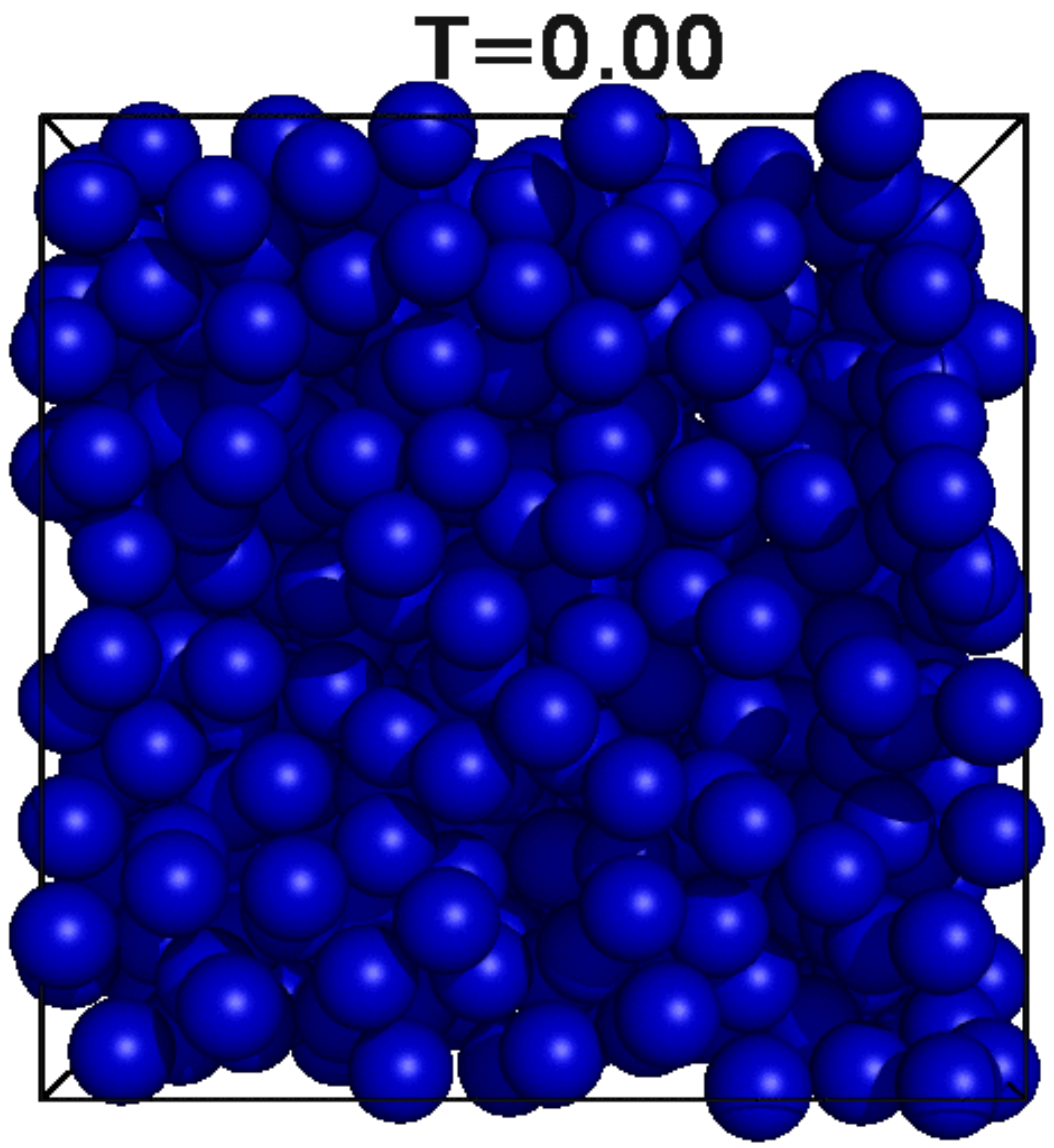
EXPERIMENT 1

- 512 LJ solvent particles(σ_A, ϵ_A) in a box of size $9\sigma_A \times 9\sigma_A \times 9\sigma_A$



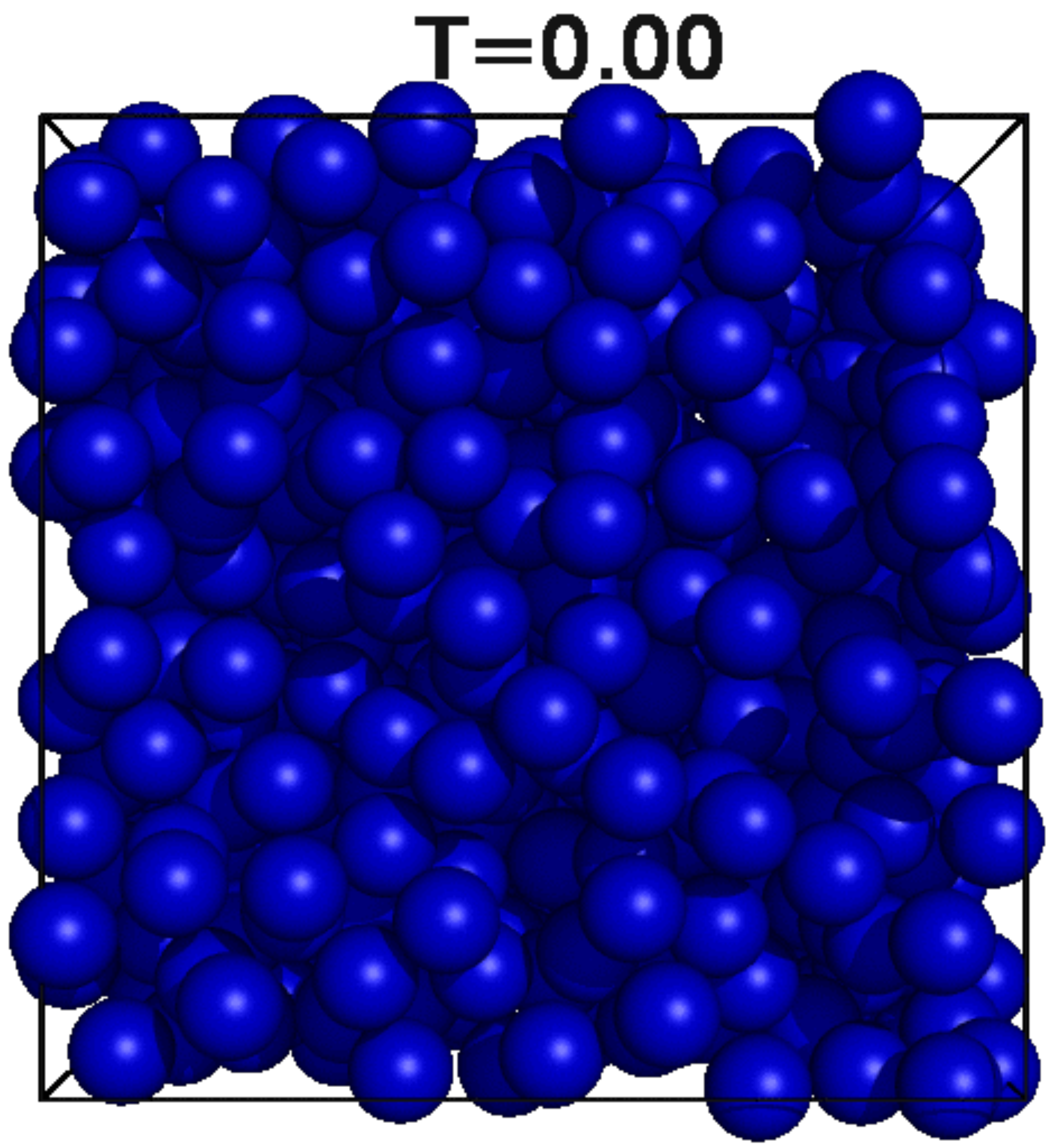
EXPERIMENT 1

- 512 LJ solvent particles(σ_A, ϵ_A) in a box of size $9\sigma_A \times 9\sigma_A \times 9\sigma_A$
    - + 1solute LJ-particle with ($\sigma_B = 2\sigma_A, \epsilon_B = 2\epsilon_A$)



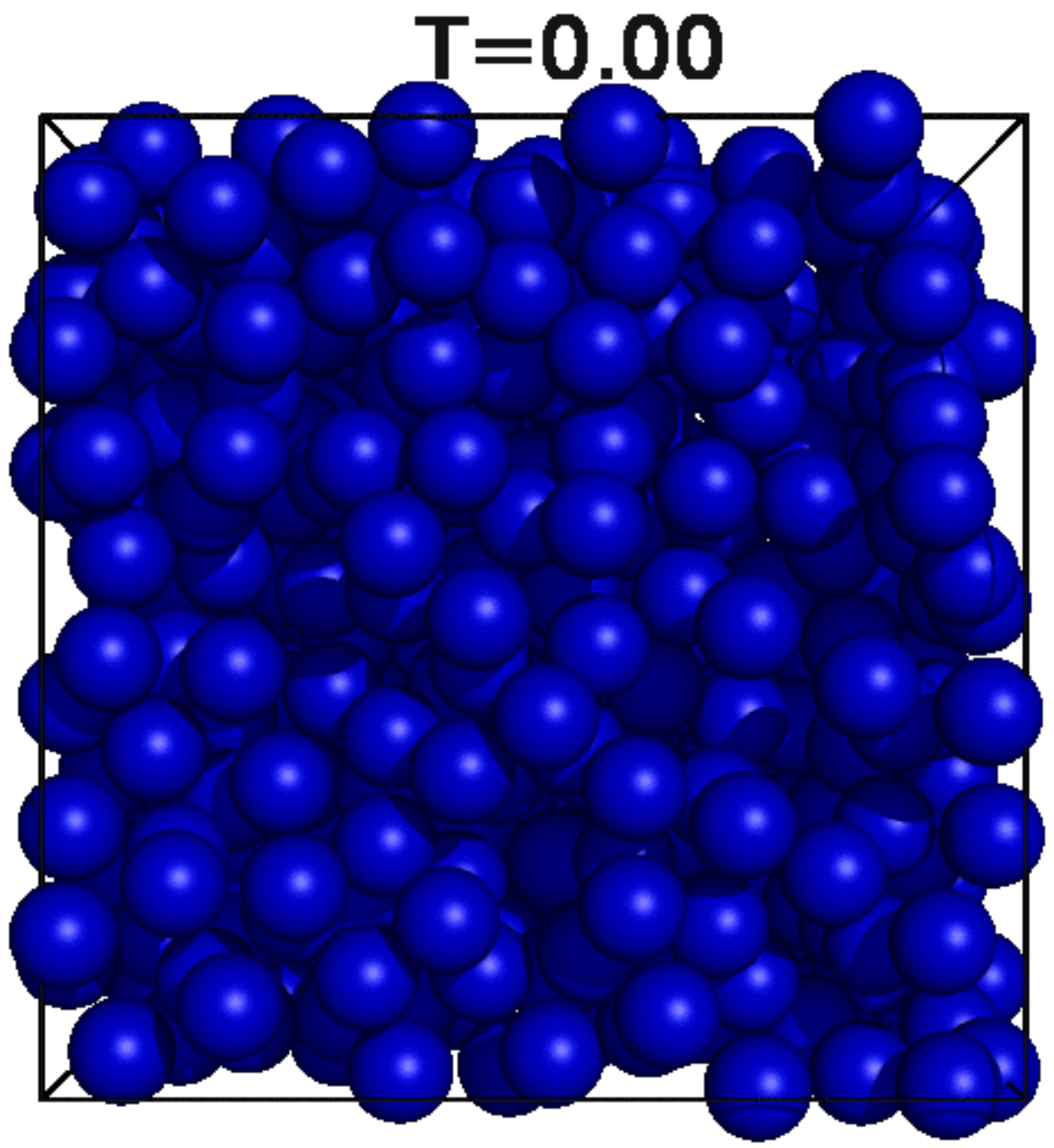
EXPERIMENT I

- 512 LJ solvent particles(σ_A, ϵ_A) in a box of size $9\sigma_A \times 9\sigma_A \times 9\sigma_A$
  - + 1solute LJ-particle with $(\sigma_B = 2\sigma_A, \epsilon_B = 2\epsilon_A)$
  - solute-solvent interactions: $\sigma_{AB} = \frac{\sigma_A + \sigma_B}{2}, \epsilon_{AB} = \sqrt{\epsilon_A \epsilon_B}$



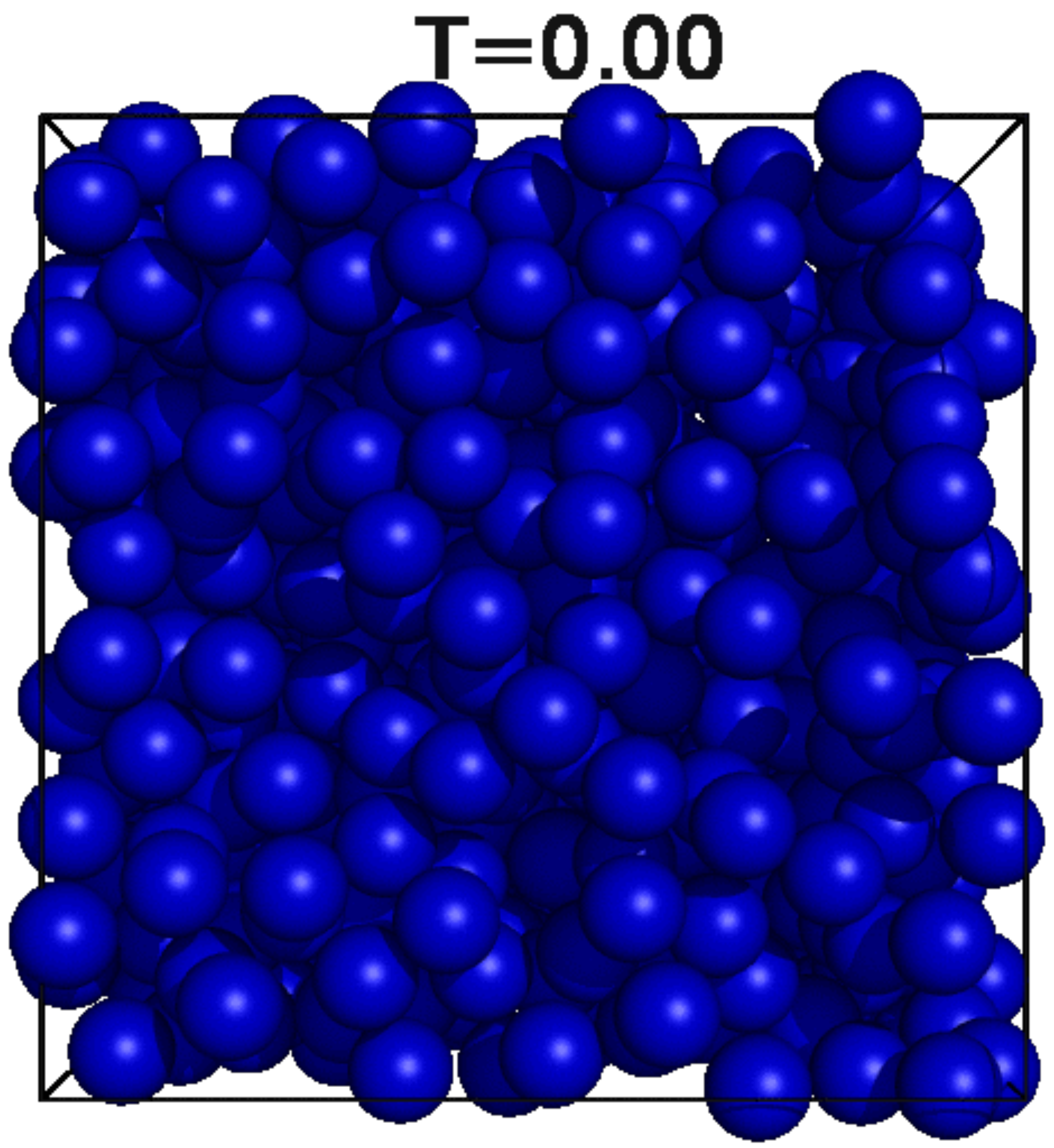
EXPERIMENT I

- 512 LJ solvent particles(σ_A, ϵ_A) in a box of size $9\sigma_A \times 9\sigma_A \times 9\sigma_A$
    - + 1solute LJ-particle with $(\sigma_B = 2\sigma_A, \epsilon_B = 2\epsilon_A)$
    - solute-solvent interactions: $\sigma_{AB} = \frac{\sigma_A + \sigma_B}{2}, \epsilon_{AB} = \sqrt{\epsilon_A \epsilon_B}$
- TI between two potentials:



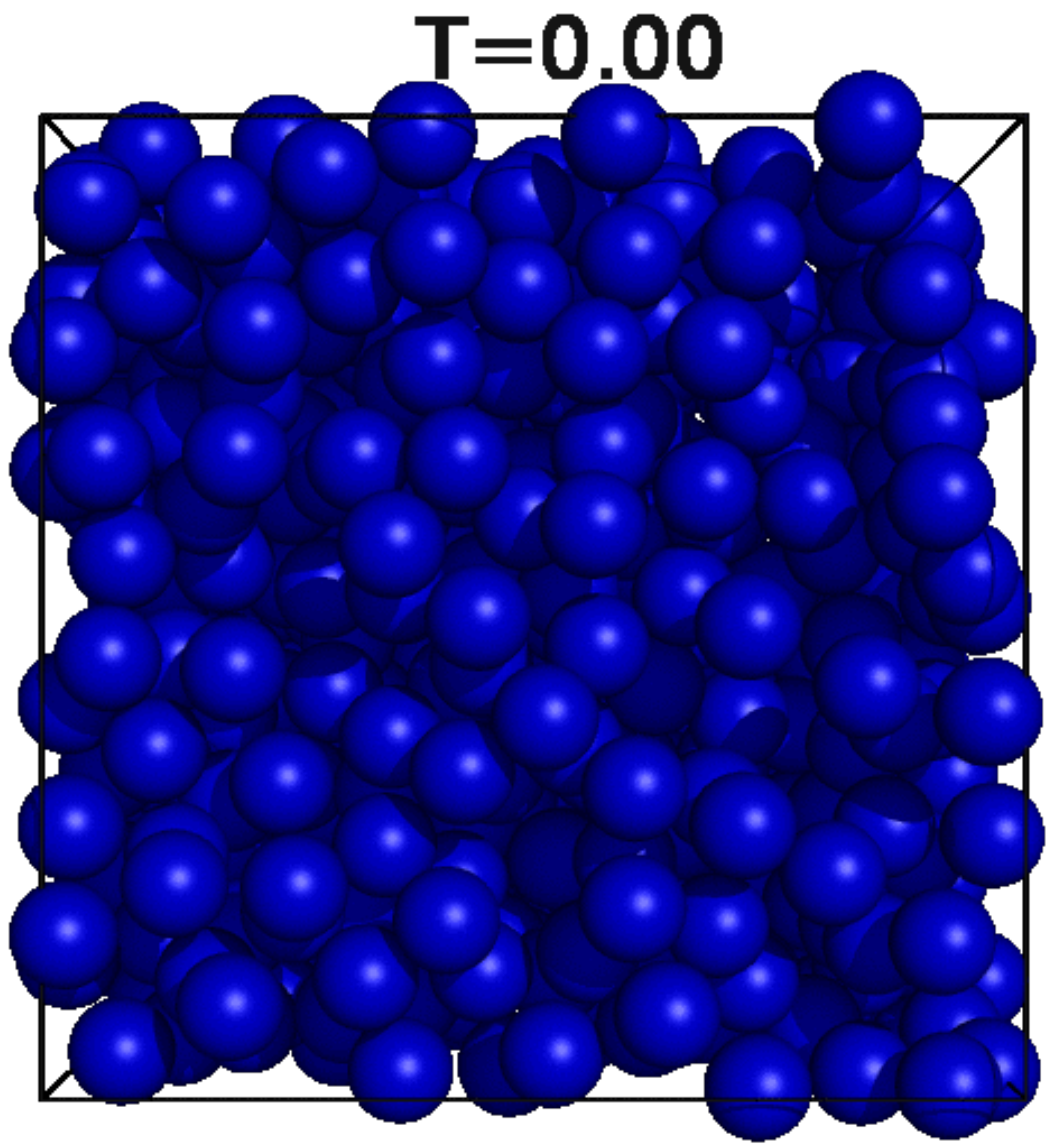
EXPERIMENT I

- 512 LJ solvent particles(σ_A, ϵ_A) in a box of size $9\sigma_A \times 9\sigma_A \times 9\sigma_A$
    - + 1solute LJ-particle with ($\sigma_B = 2\sigma_A, \epsilon_B = 2\epsilon_A$)
    - solute-solvent interactions: $\sigma_{AB} = \frac{\sigma_A + \sigma_B}{2}, \epsilon_{AB} = \sqrt{\epsilon_A \epsilon_B}$
- TI between two potentials:
    - $U_0 = U_{\text{solvent}}\ (+U_{\text{solute}})$



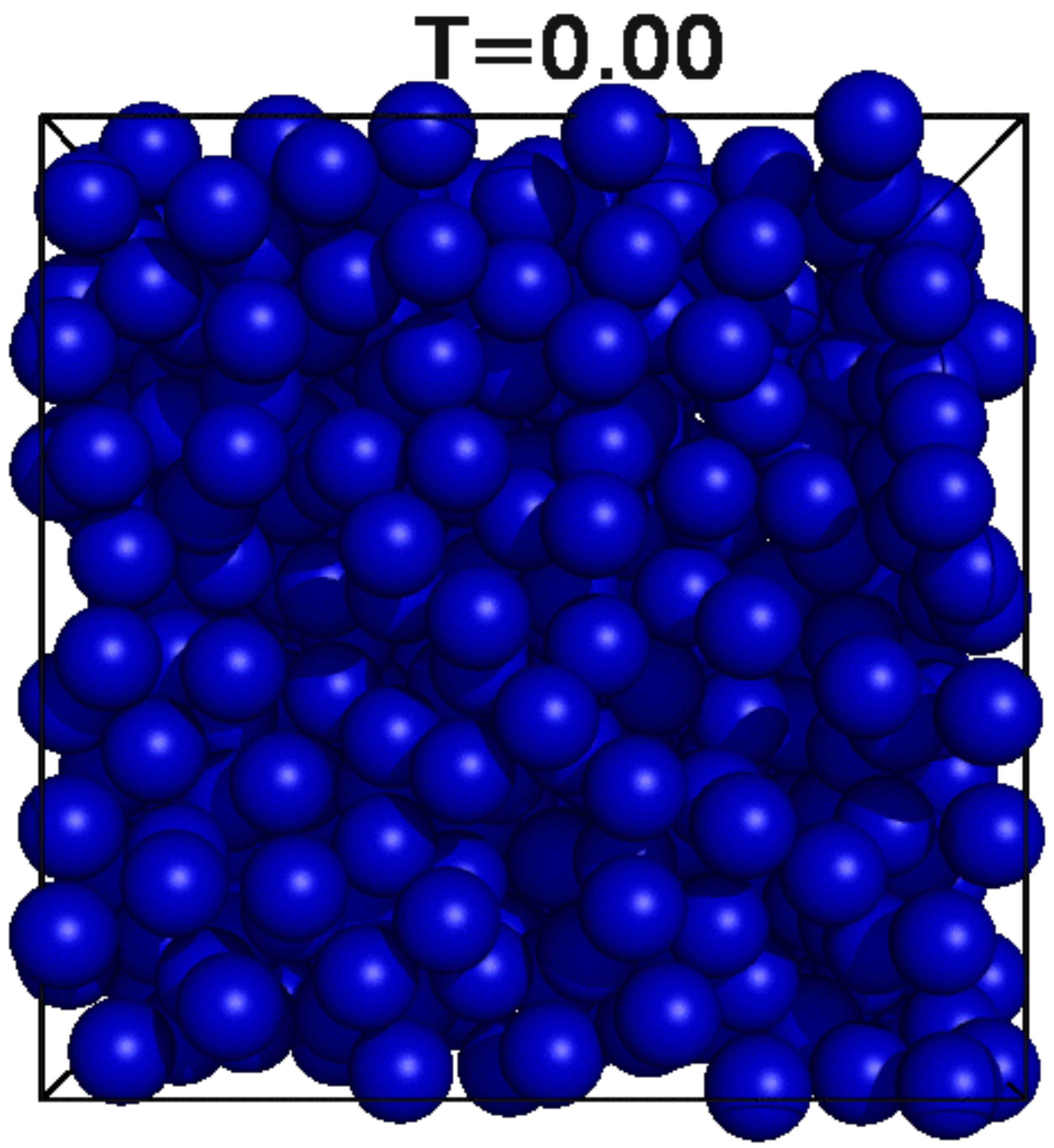
EXPERIMENT I

- 512 LJ solvent particles(σ_A, ϵ_A) in a box of size $9\sigma_A \times 9\sigma_A \times 9\sigma_A$
  - + 1solute LJ-particle with ($\sigma_B = 2\sigma_A, \epsilon_B = 2\epsilon_A$)
  - solute-solvent interactions: $\sigma_{AB} = \frac{\sigma_A + \sigma_B}{2}, \epsilon_{AB} = \sqrt{\epsilon_A \epsilon_B}$
- TI between two potentials:
  - $U_0 = U_{\text{solvent}}\ (+U_{\text{solute}})$
  - $U_1 = U_0 + U_{\text{solute-solvent}}$

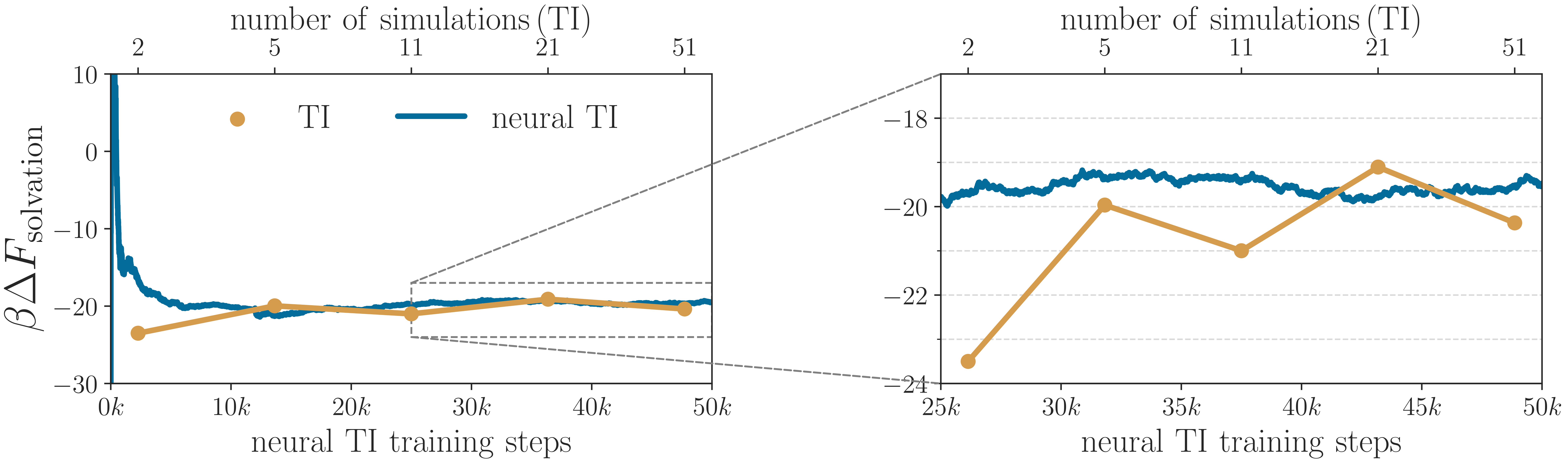


EXPERIMENT I

- 512 LJ solvent particles(σ_A, ϵ_A) in a box of size $9\sigma_A \times 9\sigma_A \times 9\sigma_A$
  - $+$ 1solute LJ-particle with $(\sigma_B = 2\sigma_A, \epsilon_B = 2\epsilon_A)$
  - solute-solvent interactions: $\sigma_{AB} = \frac{\sigma_A + \sigma_B}{2}, \epsilon_{AB} = \sqrt{\epsilon_A \epsilon_B}$
- TI between two potentials:
  - $U_0 = U_{\text{solvent}}\ (+U_{\text{solute}})$
  - $U_1 = U_0 + U_{\text{solute-solvent}}$
  - $U_t(x) = b_t^{\text{w}} U_{\text{solvent}}(x, a_t^{\text{w}}) + b_t^{\text{ws}} U_{\text{solvent-solute}}(x, a_t^{\text{ws}}) + b_t^{\text{V}} V_t^{\theta}(x)$



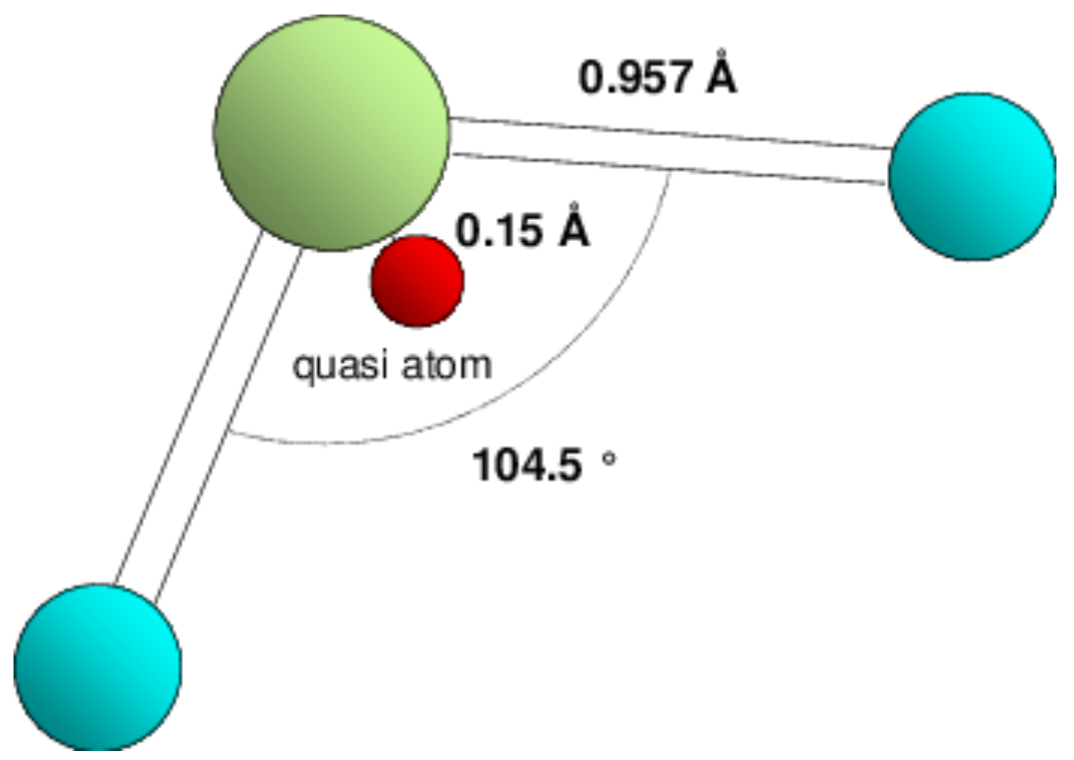
EXPERIMENT I



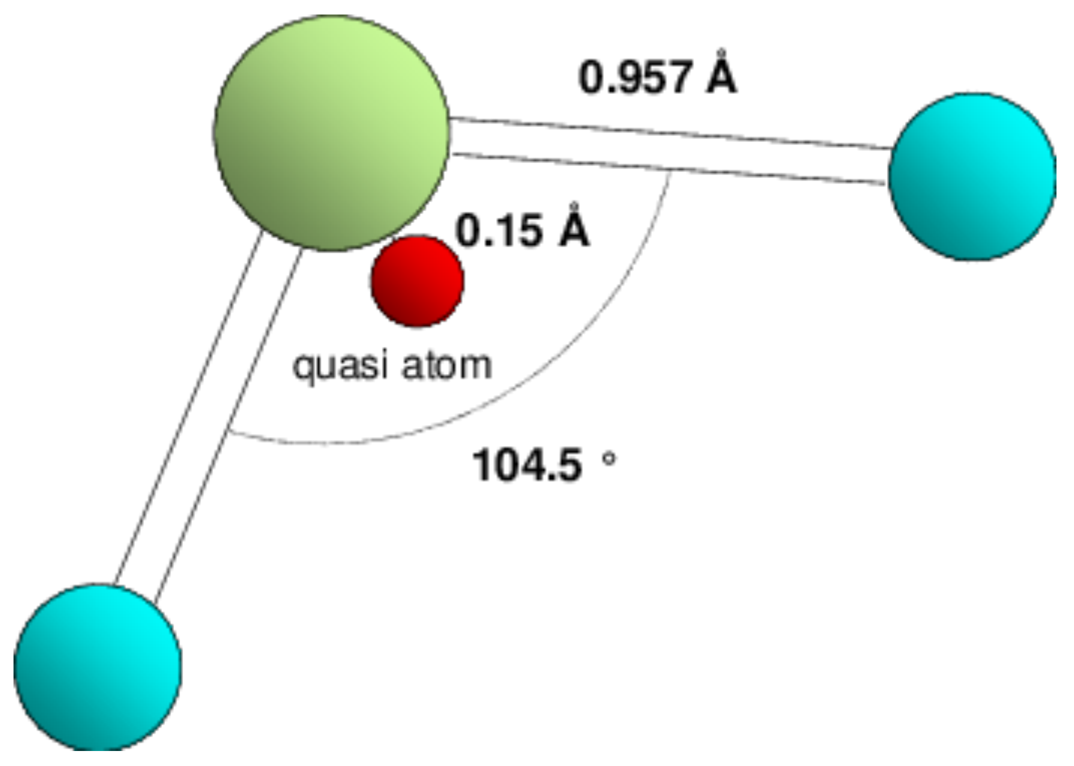
EXPERIMENT 2

EXPERIMENT 2

- 216 water molecules (TIP4P)

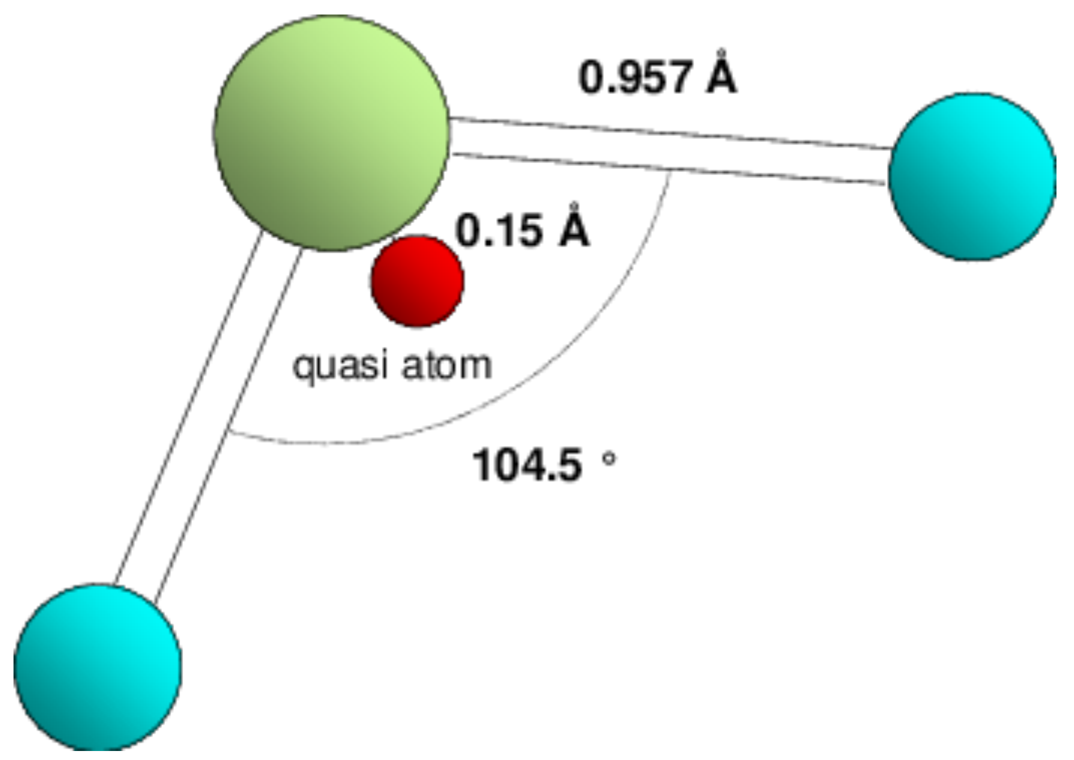


EXPERIMENT 2



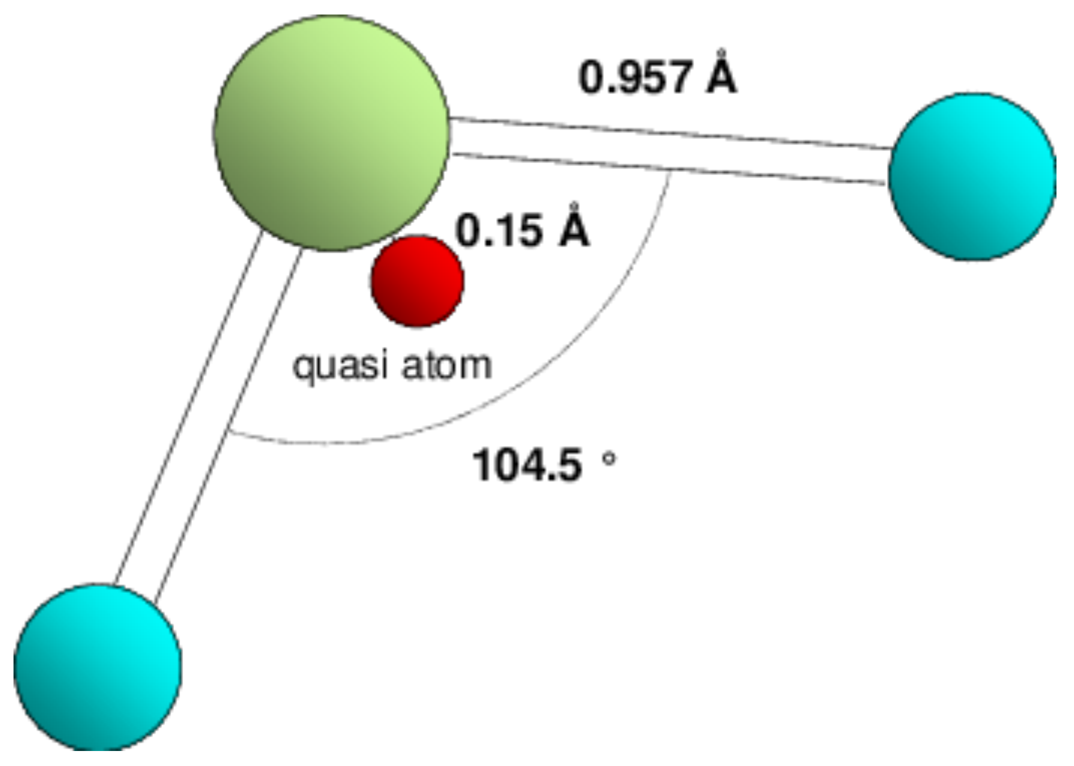
- 216 water molecules (TIP4P)
    - + solute molecule (water/methane)

EXPERIMENT 2



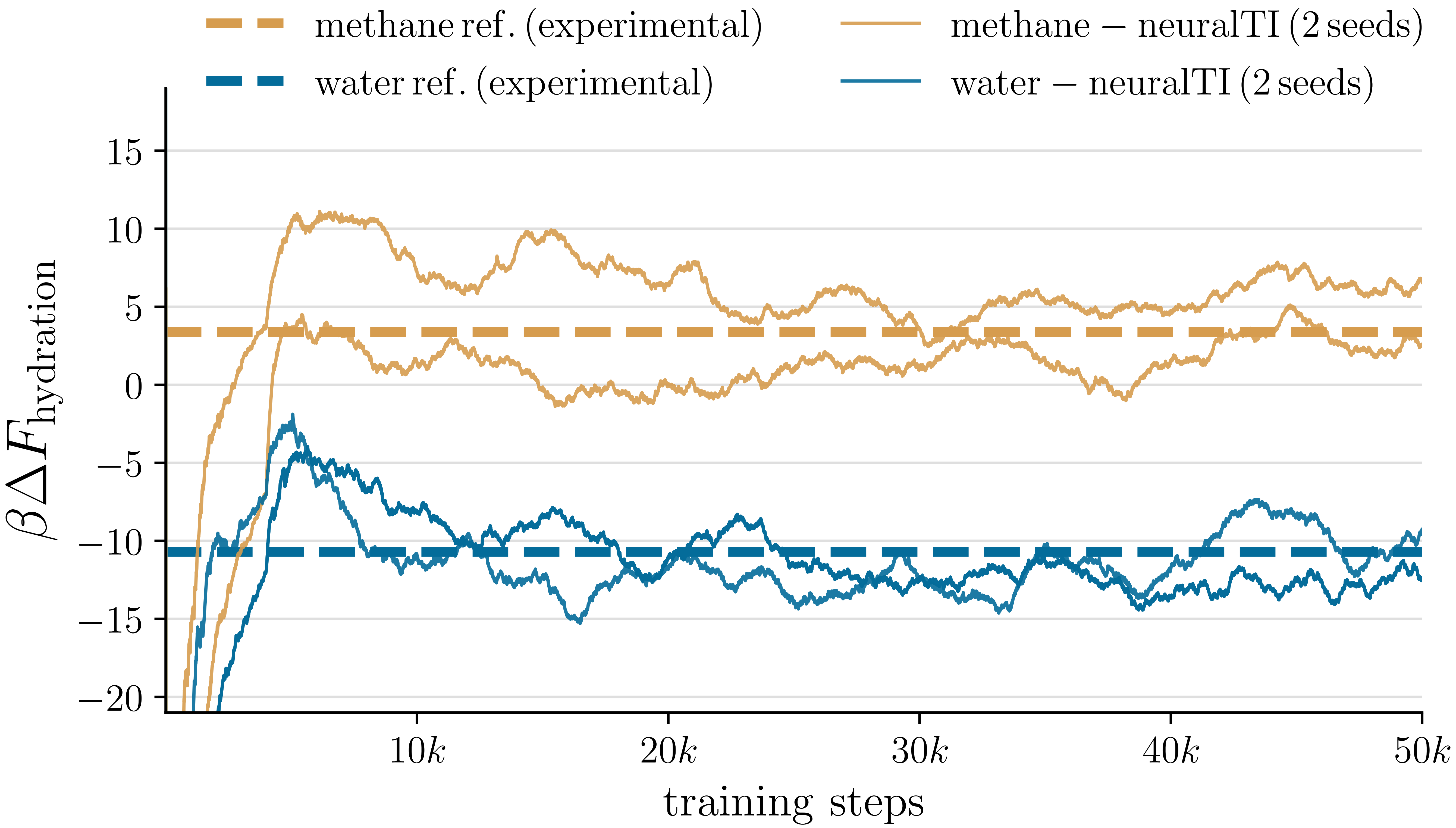
- 216 water molecules (TIP4P)
  - + solute molecule (water/methane)
  - rigid molecules, the configuration space is $[\mathbb{T}^3 \times SO(3)]^N$

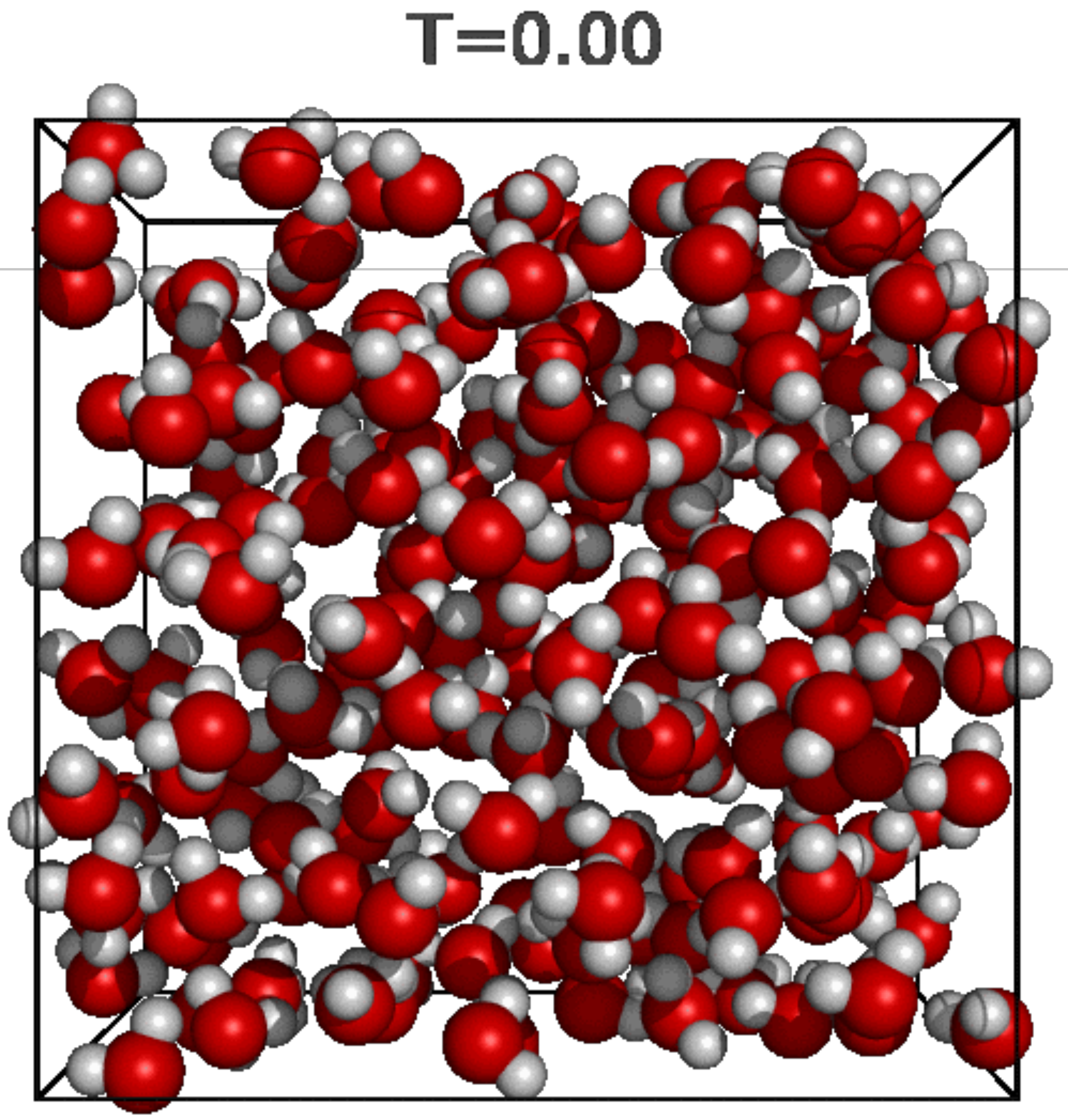
EXPERIMENT 2

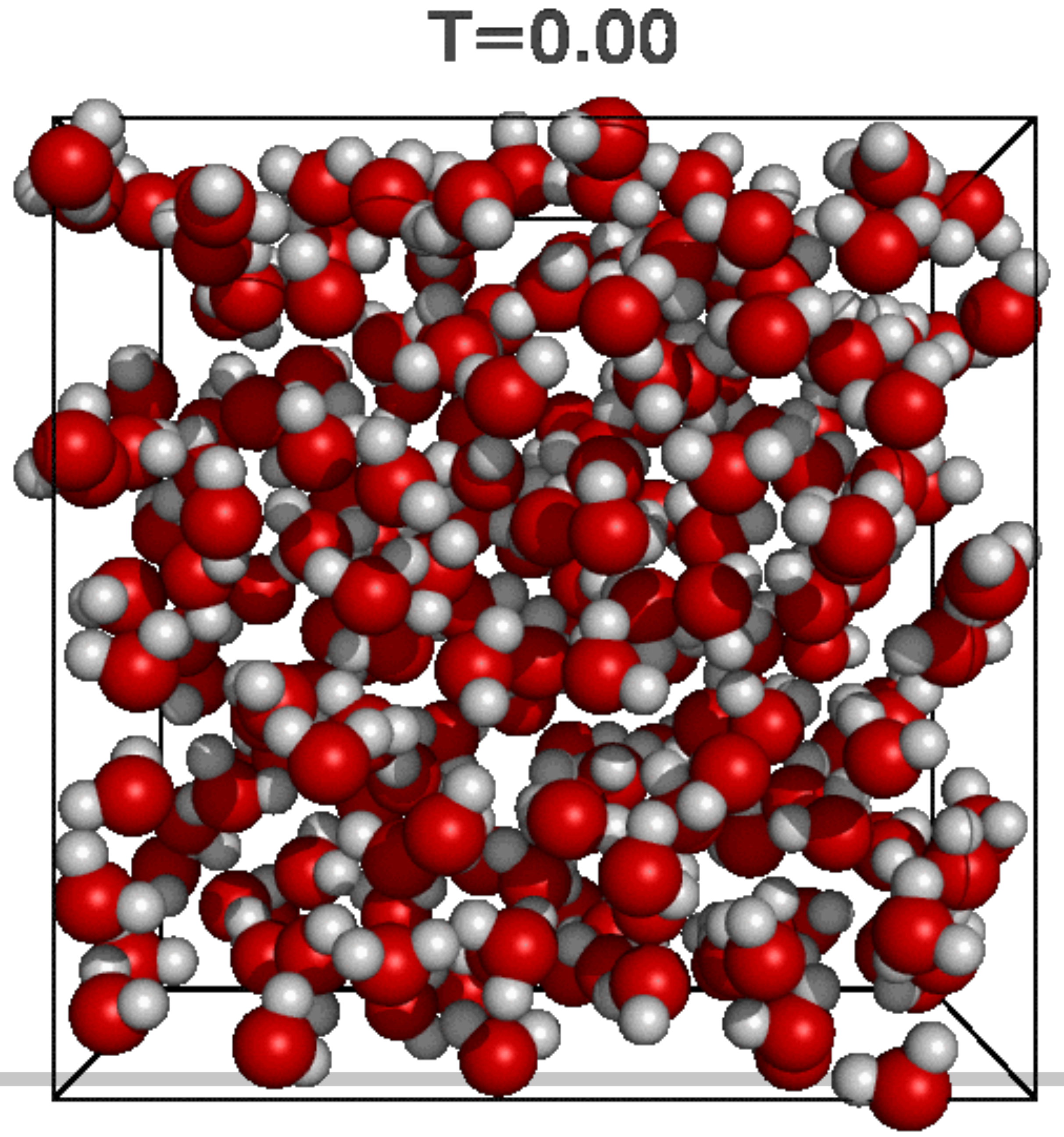


- 216 water molecules (TIP4P)
    - + solute molecule (water/methane)
    - rigid molecules, the configuration space is $[\mathbb{T}^3 \times SO(3)]^N$
    - LJ + Coulomb interactions

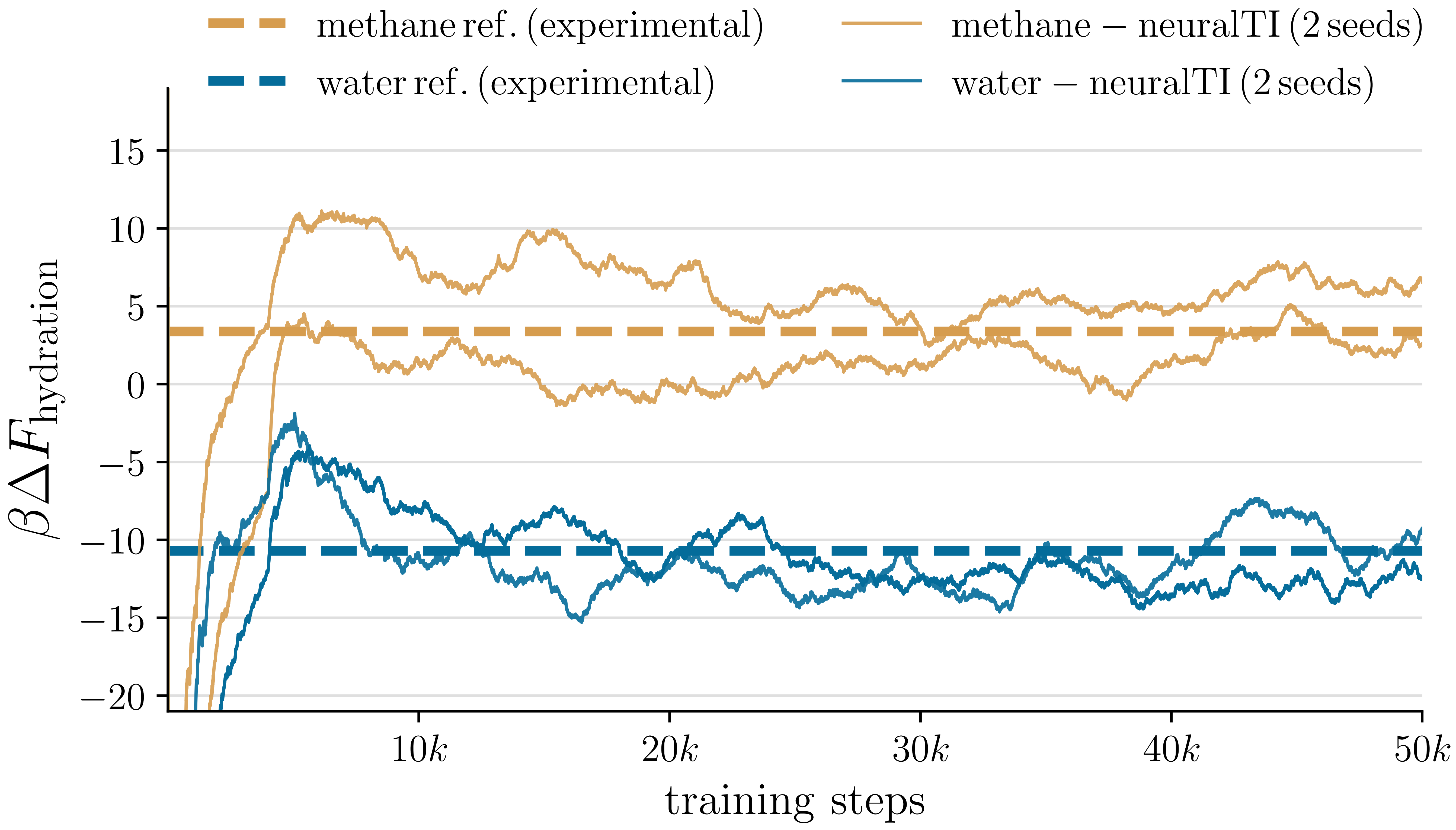
EXPERIMENT 2

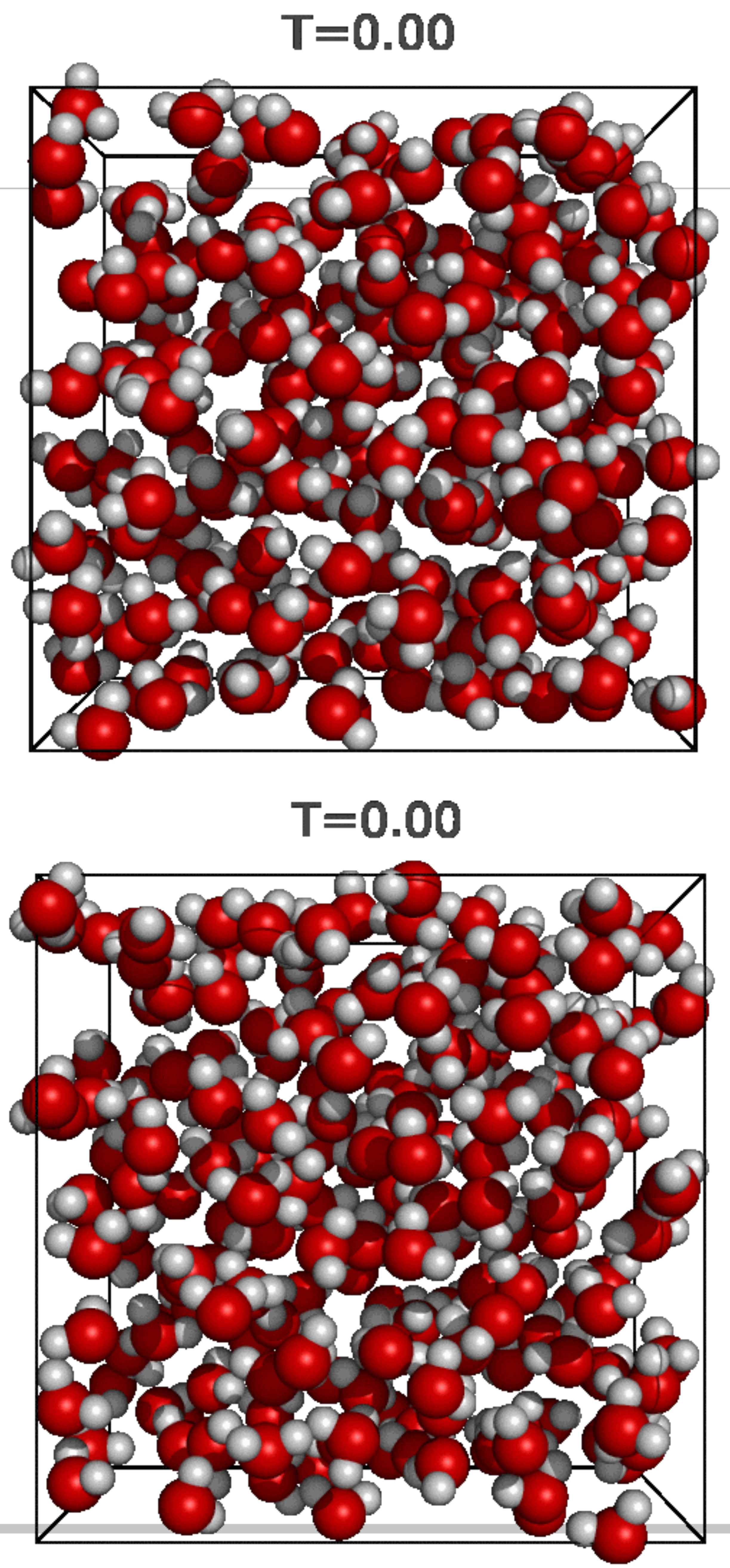






EXPERIMENT 2





CAN WE LEARN U_t BY OPTIMIZING ∇U_t?

CAN WE LEARN U_t BY OPTIMIZING ∇U_t?

- If $\nabla U = \nabla \hat{U}$ everywhere, then $U = \hat{U}$ up to an additive constant,

CAN WE LEARN U_t BY OPTIMIZING ∇U_t?

- If $\nabla U = \nabla \hat{U}$ everywhere, then $U = \hat{U}$ up to an additive constant,
  - * if the domain is connected

CAN WE LEARN U_t BY OPTIMIZING ∇U_t?

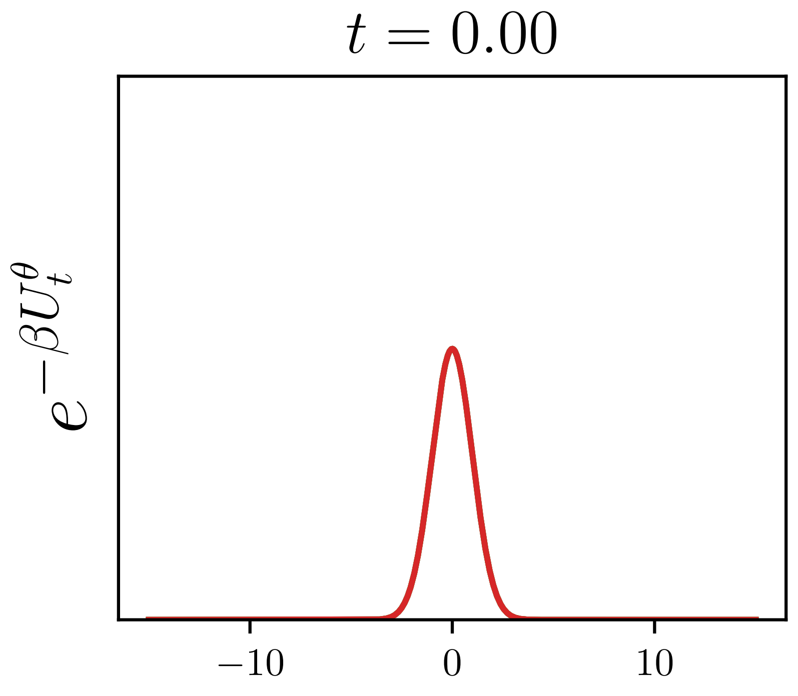
- If $\nabla U = \nabla \hat{U}$ everywhere, then $U = \hat{U}$ up to an additive constant,
  - * if the domain is connected
  - $e^{-\beta U}$ is multimodal ≈ disconnected domain

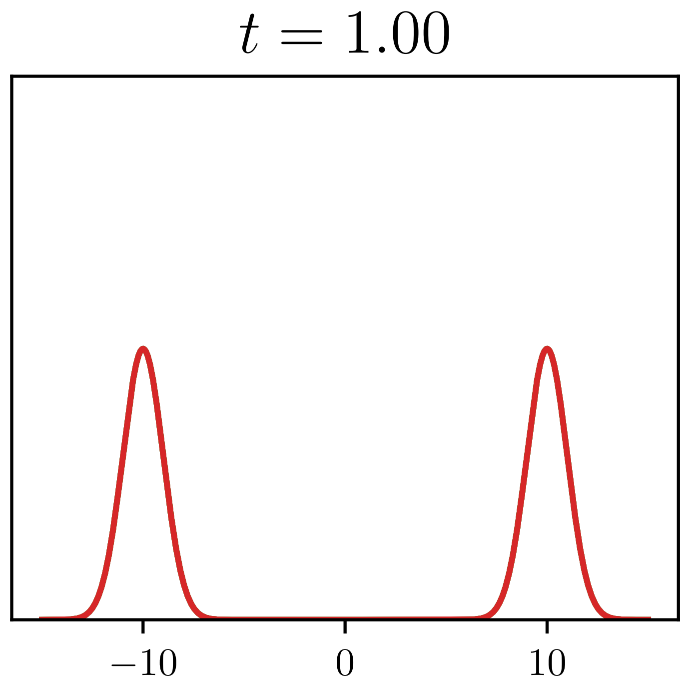
CAN WE LEARN U_t BY OPTIMIZING ∇U_t?

- If $\nabla U = \nabla \hat{U}$ everywhere, then $U = \hat{U}$ up to an additive constant,
  - * if the domain is connected
  - $e^{-\beta U}$ is multimodal $\approx$ disconnected domain
  - $e^{-\beta U_t^\theta} \propto \rho_t$ only locally

CAN WE LEARN U_t BY OPTIMIZING ∇U_t?

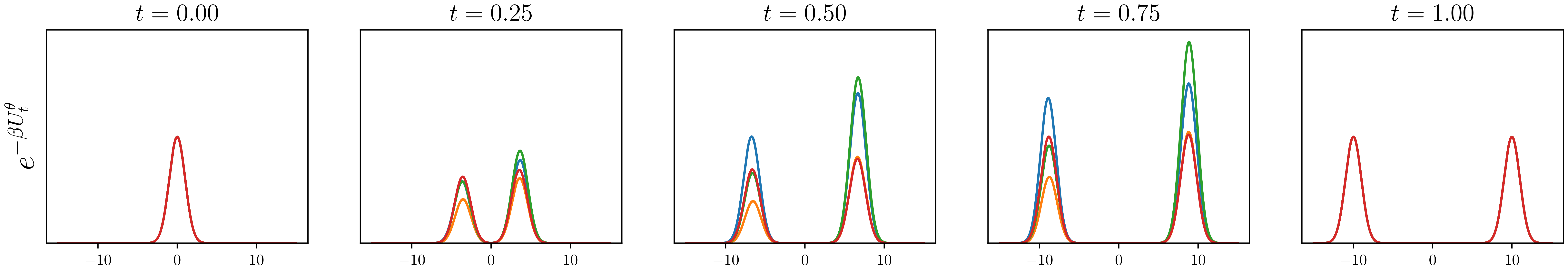
- If $\nabla U = \nabla \hat{U}$ everywhere, then $U = \hat{U}$ up to an additive constant,
  - * if the domain is connected
  - $e^{-\beta U}$ is multimodal $\approx$ disconnected domain
  - $e^{-\beta U_t^\theta} \propto \rho_t$ only locally





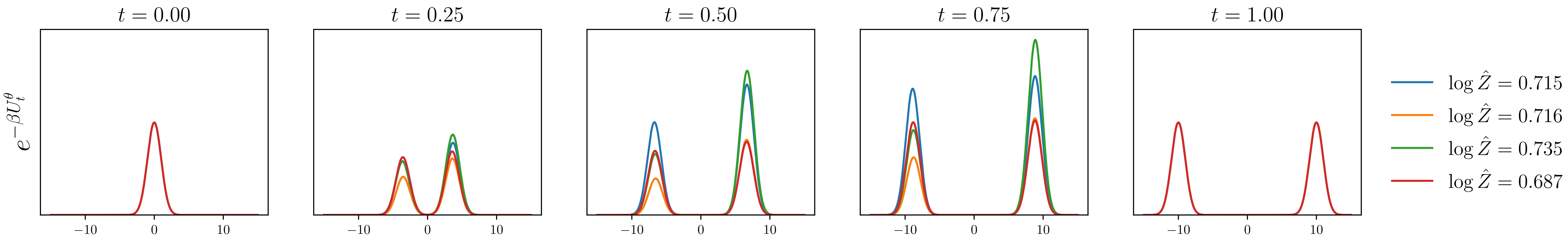
CAN WE LEARN U_t BY OPTIMIZING ∇U_t?

- If $\nabla U = \nabla \hat{U}$ everywhere, then $U = \hat{U}$ up to an additive constant,
  - * if the domain is connected
  - $e^{-\beta U}$ is multimodal $\approx$ disconnected domain
  - $e^{-\beta U_t^\theta} \propto \rho_t$ only locally



CAN WE LEARN U_t BY OPTIMIZING ∇U_t?

- If $\nabla U = \nabla \hat{U}$ everywhere, then $U = \hat{U}$ up to an additive constant,
  - * if the domain is connected
  - $e^{-\beta U}$ is multimodal $\approx$ disconnected domain
  - $e^{-\beta U_t^\theta} \propto \rho_t$ only locally



QUESTIONS?