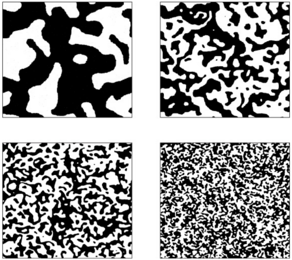
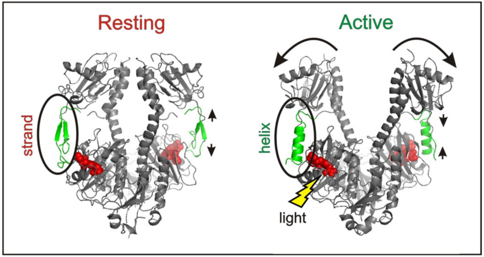
Boltzmann Generators

F. Noé[1]

December 8, 2018

Limitations of Monte Carlo Sampling



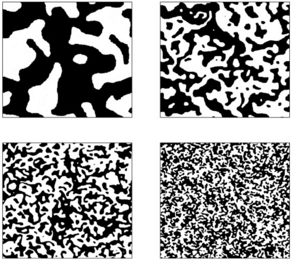


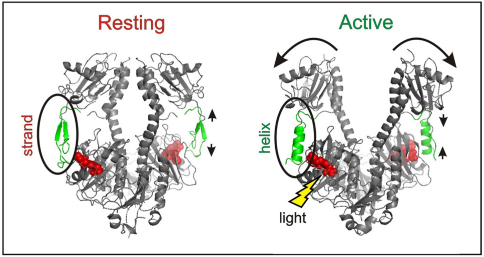
- **Input**: Reduced Potential Energy $u(\mathbf{x})$ in coordinates $\mathbf{x} \in \mathbb{R}^n$, e.g. $u(\mathbf{x}) = U(\mathbf{x})/k_B T$ (canonical ensemble).
- **Aim**: Generate *independent* Samples from Equilibrium Distribution.

$$\mu(\mathbf{x}) \propto e^{-u(\mathbf{x})}$$

- **Problem**: For large n, subvolume of low-energy configurations is vanishingly small compared to $\mathbb{R}^n$ and has a complex shape.
  - No protocol to place all n degrees of freedom is known such that direct MC sampling (proposal + rejection or reweighting) works.
  - Standard approach: MD/MCMC with local moves $\rightarrow$ sampling problem.

Limitations of Monte Carlo Sampling

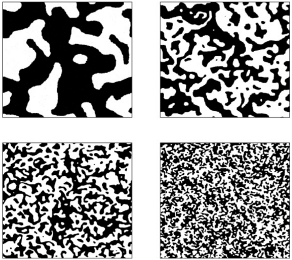


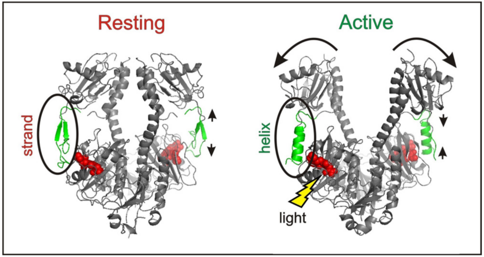


- **Input**: Reduced Potential Energy $u(\mathbf{x})$ in coordinates $\mathbf{x} \in \mathbb{R}^n$, e.g. $u(\mathbf{x}) = U(\mathbf{x})/k_B T$ (canonical ensemble).
- **Aim**: Generate *independent* Samples from Equilibrium Distribution.

$$\mu(\mathbf{x}) \propto \mathrm{e}^{-u(\mathbf{x})}$$

- **Problem**: For large n, subvolume of low-energy configurations is vanishingly small compared to $\mathbb{R}^n$ and has a complex shape.
  - No protocol to place all n degrees of freedom is known such that direct MC sampling (proposal + rejection or reweighting) works.
  - Standard approach: MD/MCMC with local moves $\rightarrow$ sampling problem.



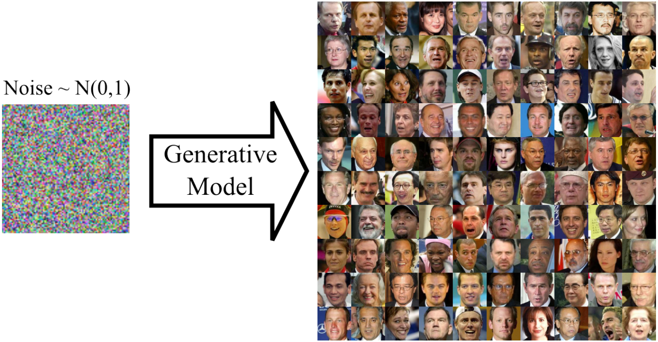


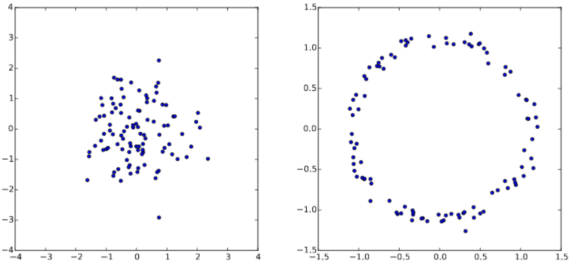
- **Input**: Reduced Potential Energy $u(\mathbf{x})$ in coordinates $\mathbf{x} \in \mathbb{R}^n$, e.g. $u(\mathbf{x}) = U(\mathbf{x})/k_B T$ (canonical ensemble).
- **Aim**: Generate *independent* Samples from Equilibrium Distribution.

$$\mu(\mathbf{x}) \propto \mathrm{e}^{-u(\mathbf{x})}$$

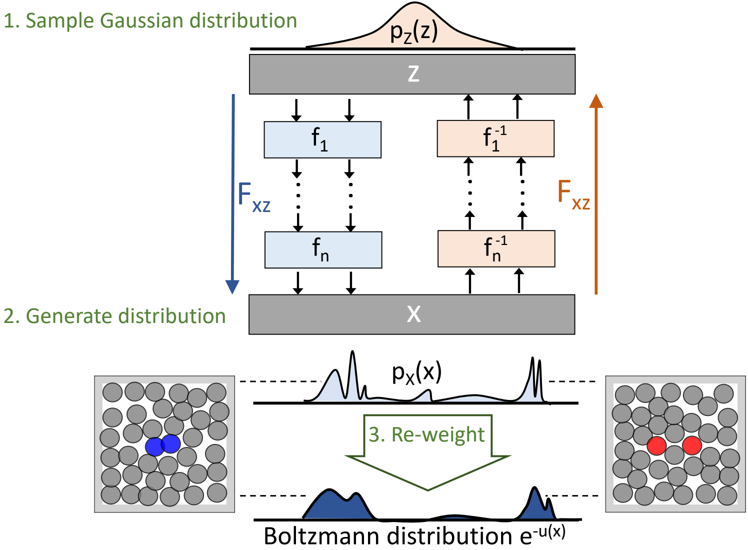
- **Problem**: For large n, subvolume of low-energy configurations is vanishingly small compared to $\mathbb{R}^n$ and has a complex shape.
  - No protocol to place all n degrees of freedom is known such that direct MC sampling (proposal + rejection or reweighting) works.
  - Standard approach: MD/MCMC with local moves $\rightarrow$ sampling problem.

Idea: Use Generative Neural Networks





Boltzmann Generators[1]



[1] https://arxiv.org/abs/1812.01729

- Invertible transformation:

$$\mathbf{z} = F_{xz}(\mathbf{x};\theta)$$
$$\mathbf{x} = F_{zx}(\mathbf{z};\theta).$$

- Jacobians:

$$\mathbf{J}_{zx}(\mathbf{z};\theta) = \left[\frac{\partial F_{zx}(\mathbf{z};\theta)}{\partial z_1}, \ldots, \frac{\partial F_{zx}(\mathbf{z};\theta)}{\partial z_n}\right]$$
$$\mathbf{J}_{xz}(\mathbf{x};\theta) = \left[\frac{dF_{xz}(\mathbf{x};\theta)}{dx_1}, \ldots, \frac{dF_{xz}(\mathbf{x};\theta)}{dx_n}\right]$$

- Transformation of random variables:

$$\begin{aligned} p_X(\mathbf{x}) &= p_Z(\mathbf{z}) \left|\det \mathbf{J}_{zx}(\mathbf{z})\right|^{-1} \\ &= p_Z(T_{xz}(\mathbf{x})) \left|\det \mathbf{J}_{xz}(\mathbf{x})\right| \end{aligned} \tag{1}$$

$$\begin{aligned} p_Z(\mathbf{z}) &= p_X(\mathbf{x}) \left|\det \mathbf{J}_{xz}(\mathbf{x})\right|^{-1} \\ &= p_X(T_{zx}(\mathbf{z})) \left|\det \mathbf{J}_{zx}(\mathbf{z})\right| \end{aligned} \tag{2}$$

[1] https://arxiv.org/abs/1812.01729

Transformation of random variables

- Invertible transformation:

$$\mathbf{z} = F_{xz}(\mathbf{x}; \theta)$$
$$\mathbf{x} = F_{zx}(\mathbf{z}; \theta).$$

- Jacobians:

$$\mathbf{J}_{zx}(\mathbf{z}; \theta) = \left[\frac{\partial F_{zx}(\mathbf{z}; \theta)}{\partial z_1}, \ldots, \frac{\partial F_{zx}(\mathbf{z}; \theta)}{\partial z_n} \right]$$
$$\mathbf{J}_{xz}(\mathbf{x}; \theta) = \left[\frac{dF_{xz}(\mathbf{x}; \theta)}{dx_1}, \ldots, \frac{dF_{xz}(\mathbf{x}; \theta)}{dx_n} \right]$$

- Transformation of random variables:

$$\begin{aligned} p_X(\mathbf{x}) &= p_Z(\mathbf{z}) \left|\det \mathbf{J}_{zx}(\mathbf{z})\right|^{-1} \\ &= p_Z(T_{xz}(\mathbf{x})) \left|\det \mathbf{J}_{xz}(\mathbf{x})\right| \end{aligned} \tag{1}$$

$$\begin{aligned} p_Z(\mathbf{z}) &= p_X(\mathbf{x}) \left|\det \mathbf{J}_{xz}(\mathbf{x})\right|^{-1} \\ &= p_X(T_{zx}(\mathbf{z})) \left|\det \mathbf{J}_{zx}(\mathbf{z})\right| \end{aligned} \tag{2}$$

[1] https://arxiv.org/abs/1812.01729

Transformation of random variables

- Invertible transformation:

$$\mathbf{z} = F_{xz}(\mathbf{x};\theta)$$
$$\mathbf{x} = F_{zx}(\mathbf{z};\theta).$$

- Jacobians:

$$\mathbf{J}_{zx}(\mathbf{z};\theta) = \left[\frac{\partial F_{zx}(\mathbf{z};\theta)}{\partial z_1}, \ldots, \frac{\partial F_{zx}(\mathbf{z};\theta)}{\partial z_n}\right]$$
$$\mathbf{J}_{xz}(\mathbf{x};\theta) = \left[\frac{dF_{xz}(\mathbf{x};\theta)}{dx_1}, \ldots, \frac{dF_{xz}(\mathbf{x};\theta)}{dx_n}\right]$$

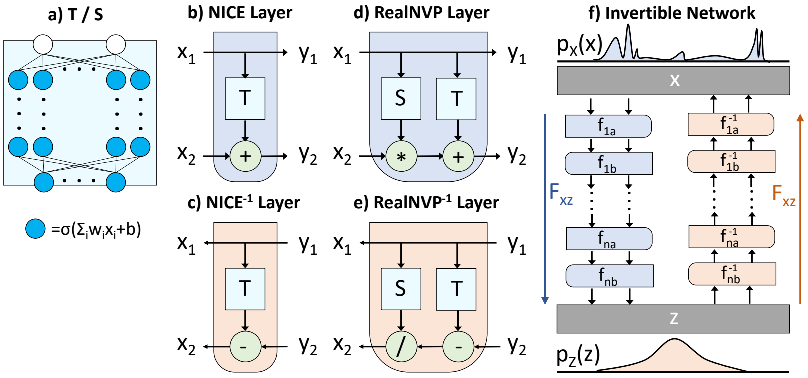
- Transformation of random variables:

$$\begin{aligned} p_X(\mathbf{x}) &= p_Z(\mathbf{z}) \left|\det \mathbf{J}_{zx}(\mathbf{z})\right|^{-1} \\ &= p_Z(T_{xz}(\mathbf{x})) \left|\det \mathbf{J}_{xz}(\mathbf{x})\right| \end{aligned} \tag{1}$$

$$\begin{aligned} p_Z(\mathbf{z}) &= p_X(\mathbf{x}) \left|\det \mathbf{J}_{xz}(\mathbf{x})\right|^{-1} \\ &= p_X(T_{zx}(\mathbf{z})) \left|\det \mathbf{J}_{zx}(\mathbf{z})\right| \end{aligned} \tag{2}$$

[1] https://arxiv.org/abs/1812.01729

Boltzmann Generators[1]



- **NICE**: Dinh, Krueger, Y. Bengio, ICLR 2015
- **RealNVP**: Dinh, Sohl-Dickstein, S. Bengio, ICLR 2017
- More advanced flow operations (e.g., normalizing flows)

[1] https://arxiv.org/abs/1812.01729

Invertible network components

Layer	f_{xz}	$\lvert\det \mathbf{J}_{xz}\rvert$	f_{zx}	$\lvert\det \mathbf{J}_{zx}\rvert$
NICE	$\begin{aligned}\mathbf{z}_1 &= \mathbf{x}_1\\ \mathbf{z}_2 &= \mathbf{x}_2 + T(\mathbf{x}_1;\theta)\end{aligned}$	1	$\begin{aligned}\mathbf{x}_1 &= \mathbf{z}_1\\ \mathbf{x}_2 &= \mathbf{z}_2 - T(\mathbf{y}_1;\theta)\end{aligned}$	1
Scaling, Exp	$\mathbf{z} = \mathrm{e}^{\mathbf{k}} \circ \mathbf{x}$	$\mathrm{e}^{\sum_i k_i}$	$\mathbf{x} = \mathrm{e}^{-\mathbf{k}} \circ \mathbf{z}$	$\mathrm{e}^{-\sum_i k_i}$
RealNVP	$\begin{aligned}\mathbf{z}_1 &= \mathbf{x}_1\\ \mathbf{z}_2 &= \mathbf{x}_2 \odot \exp(S(\mathbf{x}_1;\theta))\\ &\quad + T(\mathbf{x}_1;\theta)\end{aligned}$	$\mathrm{e}^{\sum_i S_i(\mathbf{x}_1;\theta)}$	$\begin{aligned}\mathbf{x}_1 &= \mathbf{z}_1\\ \mathbf{x}_2 &= (\mathbf{z}_2 - T(\mathbf{x}_1;\theta))\\ &\quad \odot \exp(-S(\mathbf{z}_1;\theta))\end{aligned}$	$\mathrm{e}^{-\sum_i S_i(\mathbf{z}_1;\theta)}$

[1] https://arxiv.org/abs/1812.01729

Learning Problem

- **Distributions**:

$$\begin{array}{rcccl} \text{Prior} & \mu_Z(\mathbf{z}) & \xrightarrow{F_{zx}} & p_X(\mathbf{x}) & \text{generated} \\ \text{Boltzmann} & \mu_X(\mathbf{x}) & \xrightarrow{F_{xz}} & p_Z(\mathbf{z}) & \text{generated} \end{array}$$

- **Aim**: sample configurations $\mathbf{x}$ from **Boltzmann distribution**

$$\mu_X(\mathbf{x}) = Z_X^{-1} e^{-u(\mathbf{x})} \tag{3}$$

- **Prior distribution**: Sample input in $\mathbf{z}$ from isotropic Gaussian:

$$q_Z(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}) = Z_Z^{-1} e^{-\frac{1}{2}\|\mathbf{z}\|^2/\sigma^2}, \tag{4}$$

Prior energy:

$$u_Z(\mathbf{z}) = -\log q_Z(\mathbf{z}) = \frac{1}{2\sigma^2}\|\mathbf{z}\|^2 + \text{const}$$

[1] https://arxiv.org/abs/1812.01729

Learning Problem

- **Distributions**:

$$\begin{array}{rcccl} \text{Prior} & \mu_Z(\mathbf{z}) & \xrightarrow{F_{zx}} & p_X(\mathbf{x}) & \text{generated} \\ \text{Boltzmann} & \mu_X(\mathbf{x}) & \xrightarrow{F_{xz}} & p_Z(\mathbf{z}) & \text{generated} \end{array}$$

- **Aim**: sample configurations $\mathbf{x}$ from **Boltzmann distribution**

$$\mu_X(\mathbf{x}) = Z_X^{-1} \mathrm{e}^{-u(\mathbf{x})} \tag{3}$$

- **Prior distribution**: Sample input in $\mathbf{z}$ from isotropic Gaussian:

$$q_Z(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}) = Z_Z^{-1} \mathrm{e}^{-\frac{1}{2}\|\mathbf{z}\|^2/\sigma^2}, \tag{4}$$

Prior energy:

$$u_Z(\mathbf{z}) = -\log q_Z(\mathbf{z}) = \frac{1}{2\sigma^2} \|\mathbf{z}\|^2 + \text{const}$$

[1] https://arxiv.org/abs/1812.01729

- **Distributions**:

$$\begin{array}{rccccl} \text{Prior} & \mu_Z(\mathbf{z}) & \xrightarrow{F_{zx}} & p_X(\mathbf{x}) & \text{generated} \\ \text{Boltzmann} & \mu_X(\mathbf{x}) & \xrightarrow{F_{xz}} & p_Z(\mathbf{z}) & \text{generated} \end{array}$$

- **Aim**: sample configurations $\mathbf{x}$ from **Boltzmann distribution**

$$\mu_X(\mathbf{x}) = Z_X^{-1} \mathrm{e}^{-u(\mathbf{x})} \tag{3}$$

- **Prior distribution**: Sample input in $\mathbf{z}$ from isotropic Gaussian:

$$q_Z(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}) = Z_Z^{-1} \mathrm{e}^{-\frac{1}{2}\|\mathbf{z}\|^2/\sigma^2}, \tag{4}$$

Prior energy:

$$u_Z(\mathbf{z}) = -\log q_Z(\mathbf{z}) = \frac{1}{2\sigma^2}\|\mathbf{z}\|^2 + \text{const}$$

[1] https://arxiv.org/abs/1812.01729

Latent KL divergence

- Loss function:

$$J = \underbrace{w_{ML} J_{ML}}_{\text{max likelihood}} + \underbrace{w_{KL} J_{KL}}_{\text{Kullback–Leibler}} + \underbrace{w_{RC} J_{RC}}_{\text{reaction coordinate}}.$$

- KL divergence between generated $p_X(\mathbf{x})$ and Boltzmann distribution:

$$\begin{aligned} \mathrm{KL}_\theta [\mu_Z \parallel p_Z] &= \int \mu_Z(\mathbf{z}) [\log \mu_Z(\mathbf{z}) - \log p_Z(\mathbf{z};\theta)] \, \mathrm{d}\mathbf{z}, \\ &= -H_Z - \int \mu_Z(\mathbf{z}) \log p_Z(\mathbf{z};\theta) \mathrm{d}\mathbf{z}, \\ &= -H_Z - \int \mu_Z(\mathbf{z}) [\log \mu_X(F_{zx}(\mathbf{z};\theta)) + \log |\mathbf{J}_{zx}(\mathbf{z};\theta)|] \, \mathrm{d}\mathbf{z}, \\ &= \underbrace{-H_Z + \log Z_X}_{\text{const}} + \underbrace{\mathbb{E}_{\mathbf{z}\sim\mu_Z(\mathbf{z})} [u(F_{zx}(\mathbf{z};\theta)) - \log |\mathbf{J}_{zx}(\mathbf{z};\theta)|]}_{\text{free energy}} \end{aligned}$$

- KL loss:

$$J_{KL} = \mathbb{E}_{\mathbf{z}\sim\mu_Z(\mathbf{z})} [u(F_{zx}(\mathbf{z};\theta)) - \log |\mathbf{J}_{zx}(\mathbf{z};\theta)|] . \qquad (5)$$

- See also:
  - $\mathrm{KL}_\theta [q_Z \parallel p_Z]$ equals *probability distillation loss* (van den Oord et al, *PMLR 2018*).
  - Neural Renormalization group (Li and Wang, *arXiv:1802.02840*)

[1] https://arxiv.org/abs/1812.01729

Latent KL divergence

- Loss function:

$$J = \underbrace{w_{ML} J_{ML}}_{\text{max likelihood}} + \underbrace{w_{KL} J_{KL}}_{\text{Kullback−Leibler}} + \underbrace{w_{RC} J_{RC}}_{\text{reaction coordinate}}.$$

- KL divergence between generated $p_X(\mathbf{x})$ and Boltzmann distribution:

$$\begin{aligned}
\mathrm{KL}_\theta\left[\mu_Z \parallel p_Z\right] &= \int \mu_Z(\mathbf{z}) \left[\log \mu_Z(\mathbf{z}) - \log p_Z(\mathbf{z};\theta)\right] \mathrm{d}\mathbf{z}, \\
&= -H_Z - \int \mu_Z(\mathbf{z}) \log p_Z(\mathbf{z};\theta) \mathrm{d}\mathbf{z}, \\
&= -H_Z - \int \mu_Z(\mathbf{z}) \left[\log \mu_X(F_{zx}(\mathbf{z};\theta)) + \log\left|\mathbf{J}_{zx}(\mathbf{z};\theta)\right|\right] \mathrm{d}\mathbf{z}, \\
&= \underbrace{-H_Z + \log Z_X}_{\text{const}} + \underbrace{\mathbb{E}_{\mathbf{z}\sim\mu_Z(\mathbf{z})}\left[u(F_{zx}(\mathbf{z};\theta)) - \log\left|\mathbf{J}_{zx}(\mathbf{z};\theta)\right|\right]}_{\text{free energy}}
\end{aligned}$$

- KL loss:

$$J_{KL} = \mathbb{E}_{\mathbf{z}\sim\mu_Z(\mathbf{z})}\left[u(F_{zx}(\mathbf{z};\theta)) - \log\left|\mathbf{J}_{zx}(\mathbf{z};\theta)\right|\right]. \tag{5}$$

- See also:
  - $\mathrm{KL}_\theta\left[q_Z \parallel p_Z\right]$ equals *probability distillation loss* (van den Oord et al, *PMLR 2018*).
  - Neural Renormalization group (Li and Wang, *arXiv:1802.02840*)

[1] https://arxiv.org/abs/1812.01729

Latent KL divergence

- Loss function:

$$J = \underbrace{w_{ML} J_{ML}}_{\text{max likelihood}} + \underbrace{w_{KL} J_{KL}}_{\text{Kullback–Leibler}} + \underbrace{w_{RC} J_{RC}}_{\text{reaction coordinate}}.$$

- KL divergence between generated $p_X(\mathbf{x})$ and Boltzmann distribution:

$$\begin{aligned}
\mathrm{KL}_\theta\left[\mu_Z \parallel p_Z\right] &= \int \mu_Z(\mathbf{z}) \left[\log \mu_Z(\mathbf{z}) - \log p_Z(\mathbf{z};\theta)\right] \mathrm{d}\mathbf{z}, \\
&= -H_Z - \int \mu_Z(\mathbf{z}) \log p_Z(\mathbf{z};\theta) \mathrm{d}\mathbf{z}, \\
&= -H_Z - \int \mu_Z(\mathbf{z}) \left[\log \mu_X(F_{zx}(\mathbf{z};\theta)) + \log \left|\mathbf{J}_{zx}(\mathbf{z};\theta)\right|\right] \mathrm{d}\mathbf{z}, \\
&= \underbrace{-H_Z + \log Z_X}_{\text{const}} + \underbrace{\mathbb{E}_{\mathbf{z}\sim\mu_Z(\mathbf{z})}\left[u(F_{zx}(\mathbf{z};\theta)) - \log \left|\mathbf{J}_{zx}(\mathbf{z};\theta)\right|\right]}_{\text{free energy}}
\end{aligned}$$

- KL loss:

$$J_{KL} = \mathbb{E}_{\mathbf{z}\sim\mu_Z(\mathbf{z})}\left[u(F_{zx}(\mathbf{z};\theta)) - \log \left|\mathbf{J}_{zx}(\mathbf{z};\theta)\right|\right]. \qquad (5)$$

- See also:
  - $\mathrm{KL}_\theta\left[q_Z \parallel p_Z\right]$ equals *probability distillation loss* (van den Oord et al, *PMLR 2018*).
  - Neural Renormalization group (Li and Wang, *arXiv:1802.02840*)

[1] https://arxiv.org/abs/1812.01729

Latent KL divergence

- Loss function:

$$J = \underbrace{w_{ML} J_{ML}}_{\text{max likelihood}} + \underbrace{w_{KL} J_{KL}}_{\text{Kullback−Leibler}} + \underbrace{w_{RC} J_{RC}}_{\text{reaction coordinate}}.$$

- KL divergence between generated $p_X(\mathbf{x})$ and Boltzmann distribution:

$$\begin{aligned}
\mathrm{KL}_\theta\left[\mu_Z \parallel p_Z\right] &= \int \mu_Z(\mathbf{z}) \left[\log \mu_Z(\mathbf{z}) - \log p_Z(\mathbf{z};\theta)\right] \mathrm{d}\mathbf{z}, \\
&= -H_Z - \int \mu_Z(\mathbf{z}) \log p_Z(\mathbf{z};\theta) \mathrm{d}\mathbf{z}, \\
&= -H_Z - \int \mu_Z(\mathbf{z}) \left[\log \mu_X(F_{zx}(\mathbf{z};\theta)) + \log \left|\mathbf{J}_{zx}(\mathbf{z};\theta)\right|\right] \mathrm{d}\mathbf{z}, \\
&= \underbrace{-H_Z + \log Z_X}_{\text{const}} + \underbrace{\mathbb{E}_{\mathbf{z}\sim\mu_Z(\mathbf{z})}\left[u(F_{zx}(\mathbf{z};\theta)) - \log\left|\mathbf{J}_{zx}(\mathbf{z};\theta)\right|\right]}_{\text{free energy}}
\end{aligned}$$

- KL loss:

$$J_{KL} = \mathbb{E}_{\mathbf{z}\sim\mu_Z(\mathbf{z})}\left[u(F_{zx}(\mathbf{z};\theta)) - \log\left|\mathbf{J}_{zx}(\mathbf{z};\theta)\right|\right]. \tag{5}$$

- See also:
  - $\mathrm{KL}_\theta\left[q_Z \parallel p_Z\right]$ equals *probability distillation loss* (van den Oord et al, *PMLR 2018*).
  - Neural Renormalization group (Li and Wang, *arXiv:1802.02840*)

[1] https://arxiv.org/abs/1812.01729

Configuration KL divergence

- Loss function:

$$J = \underbrace{w_{ML} J_{ML}}_{\text{max likelihood}} + \underbrace{w_{KL} J_{KL}}_{\text{Kullback-Leibler}} + \underbrace{w_{RC} J_{RC}}_{\text{reaction coordinate}}.$$

- KL divergence between generated distribution $q_Z(\mathbf{z})$ and Gaussian:

$$\begin{aligned} \mathrm{KL}_\theta(\mu_X \parallel p_X) &= -H_X - \int \mu_X(\mathbf{x}) \log p_X(\mathbf{x};\theta) \mathrm{d}\mathbf{x} \\ &= -H_X - \int \mu_X(\mathbf{x}) \left[\log \mu_Z(F_{xz}(\mathbf{x};\theta)) + \log |\mathbf{J}_{xz}(\mathbf{z};\theta)|\right] \mathrm{d}\mathbf{x}. \\ &= \underbrace{-H_X + \log Z_Z}_{\text{const}} + \mathbb{E}_{\mathbf{x}\sim\mu(\mathbf{x})}\left[\frac{1}{\sigma^2}\|F_{xz}(\mathbf{x};\theta)\|^2 - \log |\mathbf{J}_{xz}(\mathbf{x};\theta)|\right] \end{aligned}$$

Problem: sampling $\mathbf{x} \sim \mu(\mathbf{x})$ is difficult and our goal!

- **Maximum Likelihood** loss for initial data distribution $\rho(\mathbf{x})$:

$$\begin{aligned} J_{ML} &= -\mathbb{E}_{\mathbf{x}\sim\rho(\mathbf{x})}\left[\log p_X(\mathbf{x};\theta)\right] \\ &= \mathbb{E}_{\mathbf{x}\sim\rho(\mathbf{x})}\left[\frac{1}{\sigma^2}\|F_{xz}(\mathbf{x};\theta)\|^2 - \log |\mathbf{J}_{xz}(\mathbf{x};\theta)|\right] \end{aligned}$$

[1] https://arxiv.org/abs/1812.01729

Configuration KL divergence

- Loss function:

$$J = \underbrace{w_{ML} J_{ML}}_{\text{max likelihood}} + \underbrace{w_{KL} J_{KL}}_{\text{Kullback−Leibler}} + \underbrace{w_{RC} J_{RC}}_{\text{reaction coordinate}}.$$

- KL divergence between generated distribution $q_Z(\mathbf{z})$ and Gaussian:

$$\begin{aligned}\mathrm{KL}_\theta(\mu_X \parallel p_X) &= -H_X - \int \mu_X(\mathbf{x}) \log p_X(\mathbf{x};\theta)\mathrm{d}\mathbf{x} \\ &= -H_X - \int \mu_X(\mathbf{x}) \left[\log \mu_Z(F_{xz}(\mathbf{x};\theta)) + \log |\mathbf{J}_{xz}(\mathbf{z};\theta)|\right] \mathrm{d}\mathbf{x}. \\ &= \underbrace{-H_X + \log Z_Z}_{\text{const}} + \mathbb{E}_{\mathbf{x}\sim\mu(\mathbf{x})}\left[\frac{1}{\sigma^2}\left\|F_{xz}(\mathbf{x};\theta)\right\|^2 - \log|\mathbf{J}_{xz}(\mathbf{x};\theta)|\right]\end{aligned}$$

Problem: sampling $\mathbf{x} \sim \mu(\mathbf{x})$ is difficult and our goal!

- **Maximum Likelihood** loss for initial data distribution $\rho(\mathbf{x})$:

$$\begin{aligned}J_{ML} &= -\mathbb{E}_{\mathbf{x}\sim\rho(\mathbf{x})}\left[\log p_X(\mathbf{x};\theta)\right] \\ &= \mathbb{E}_{\mathbf{x}\sim\rho(\mathbf{x})}\left[\frac{1}{\sigma^2}\left\|F_{xz}(\mathbf{x};\theta)\right\|^2 - \log|\mathbf{J}_{xz}(\mathbf{x};\theta)|\right]\end{aligned}$$

[1] https://arxiv.org/abs/1812.01729

Configuration KL divergence

- Loss function:

$$J = \underbrace{w_{ML} J_{ML}}_{\text{max likelihood}} + \underbrace{w_{KL} J_{KL}}_{\text{Kullback-Leibler}} + \underbrace{w_{RC} J_{RC}.}_{\text{reaction coordinate}}$$

- KL divergence between generated distribution $q_Z(\mathbf{z})$ and Gaussian:

$$\begin{aligned}
\mathrm{KL}_\theta(\mu_X \parallel p_X) &= -H_X - \int \mu_X(\mathbf{x}) \log p_X(\mathbf{x};\theta) \mathrm{d}\mathbf{x} \\
&= -H_X - \int \mu_X(\mathbf{x}) \left[\log \mu_Z(F_{xz}(\mathbf{x};\theta)) + \log |\mathbf{J}_{xz}(\mathbf{z};\theta)|\right] \mathrm{d}\mathbf{x}. \\
&= \underbrace{-H_X + \log Z_Z}_{\text{const}} + \mathbb{E}_{\mathbf{x}\sim\mu(\mathbf{x})} \left[\frac{1}{\sigma^2} \|F_{xz}(\mathbf{x};\theta)\|^2 - \log |\mathbf{J}_{xz}(\mathbf{x};\theta)| \right]
\end{aligned}$$

Problem: sampling $\mathbf{x} \sim \mu(\mathbf{x})$ is difficult and our goal!

- **Maximum Likelihood** loss for initial data distribution $\rho(\mathbf{x})$:

$$\begin{aligned}
J_{ML} &= -\mathbb{E}_{\mathbf{x}\sim\rho(\mathbf{x})} \left[\log p_X(\mathbf{x};\theta)\right] \\
&= \mathbb{E}_{\mathbf{x}\sim\rho(\mathbf{x})} \left[\frac{1}{\sigma^2} \|F_{xz}(\mathbf{x};\theta)\|^2 - \log |\mathbf{J}_{xz}(\mathbf{x};\theta)| \right]
\end{aligned}$$

[1] https://arxiv.org/abs/1812.01729

Reaction coordinate loss

- Loss function:

$$J = \underbrace{w_{ML} J_{ML}}_{\text{max likelihood}} + \underbrace{w_{KL} J_{KL}}_{\text{Kullback}-\text{Leibler}} + \underbrace{w_{RC} J_{RC}.}_{\text{reaction coordinate}}$$

- Reaction coordinate loss:

$$J_{RC} = \int p(R(\mathbf{x})) \log p(R(\mathbf{x}))\, \mathrm{d}R(\mathbf{x})$$
$$= \mathbb{E}_{\mathbf{x} \sim p_X(\mathbf{x})} \log p(R(\mathbf{x})).$$

- Implementation:
  - Reaction coordinate function R is user input
  - Min and max bounds are given
  - $p(R(\mathbf{x}))$ is computed as a batchwise kernel density estimate.

[1] https://arxiv.org/abs/1812.01729

Reaction coordinate loss

- Loss function:

$$J = \underbrace{w_{ML} J_{ML}}_{\text{max likelihood}} + \underbrace{w_{KL} J_{KL}}_{\text{Kullback}-\text{Leibler}} + \underbrace{w_{RC} J_{RC}.}_{\text{reaction coordinate}}$$

- Reaction coordinate loss:

$$\begin{aligned} J_{RC} &= \int p\left(R(\mathbf{x})\right) \log p\left(R(\mathbf{x})\right) \,\mathrm{d}R(\mathbf{x}) \\ &= \mathbb{E}_{\mathbf{x} \sim p_X(\mathbf{x})} \log p\left(R(\mathbf{x})\right). \end{aligned}$$

- Implementation:
  - Reaction coordinate function R is user input
  - Min and max bounds are given
  - $p\left(R(\mathbf{x})\right)$ is computed as a batchwise kernel density estimate.

[1] https://arxiv.org/abs/1812.01729

Reaction coordinate loss

- Loss function:

$$J = \underbrace{w_{ML} J_{ML}}_{\text{max likelihood}} + \underbrace{w_{KL} J_{KL}}_{\text{Kullback}-\text{Leibler}} + \underbrace{w_{RC} J_{RC}.}_{\text{reaction coordinate}}$$

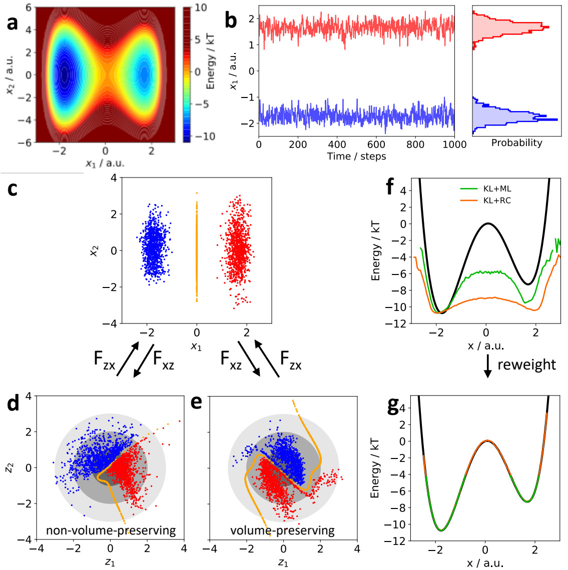
- Reaction coordinate loss:

$$\begin{aligned} J_{RC} &= \int p\left(R(\mathbf{x})\right) \log p\left(R(\mathbf{x})\right) \mathrm{d}R(\mathbf{x}) \\ &= \mathbb{E}_{\mathbf{x} \sim p_X(\mathbf{x})} \log p\left(R(\mathbf{x})\right). \end{aligned}$$

- Implementation:
  - Reaction coordinate function R is user input
  - Min and max bounds are given
  - $p\left(R(\mathbf{x})\right)$ is computed as a batchwise kernel density estimate.

[1]https://arxiv.org/abs/1812.01729

Boltzmann Generator - 1D Example



Reweighting

- Probability reweighting: assign to each generated configuration $\mathbf{x}$ the statistical weight:

$$w_X(\mathbf{x} \mid \mathbf{z}) = \frac{\mu_X(\mathbf{x})}{p_X(\mathbf{x})} = \frac{p_Z(\mathbf{z})}{\mu_Z(\mathbf{z})}. \tag{6}$$
$$\propto \mathrm{e}^{-u_X(T_{zx}(\mathbf{z}))+u_Z(\mathbf{z})+\log|\det(\mathbf{J}_{zx}(\mathbf{z}))|}$$

- Weighted expectation values

$$\mathbb{E}[O] \approx \frac{\sum_{i=1}^{N} w_X(\mathbf{x}) O(\mathbf{x})}{\sum_{i=1}^{N} w_X(\mathbf{x})}.$$

[1]https://arxiv.org/abs/1812.01729

Reweighting

- Probability reweighting: assign to each generated configuration $\mathbf{x}$ the statistical weight:

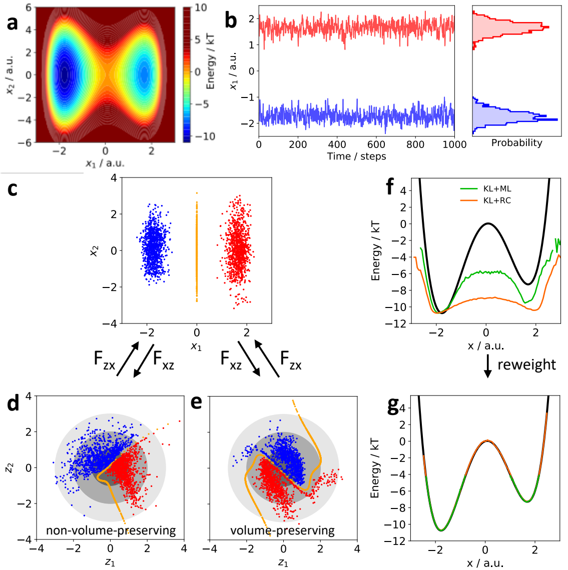
$$w_X(\mathbf{x} \mid \mathbf{z}) = \frac{\mu_X(\mathbf{x})}{p_X(\mathbf{x})} = \frac{p_Z(\mathbf{z})}{\mu_Z(\mathbf{z})}. \tag{6}$$
$$\propto e^{-u_X(T_{zx}(\mathbf{z}))+u_Z(\mathbf{z})+\log|\det(\mathbf{J}_{zx}(\mathbf{z}))|}$$

- Weighted expectation values

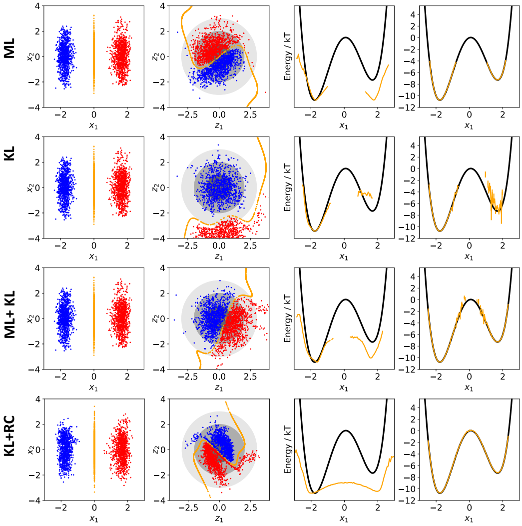
$$\mathbb{E}[O] \approx \frac{\sum_{i=1}^{N} w_X(\mathbf{x}) O(\mathbf{x})}{\sum_{i=1}^{N} w_X(\mathbf{x})}.$$

[1] https://arxiv.org/abs/1812.01729

Boltzmann Generator - 1D Example



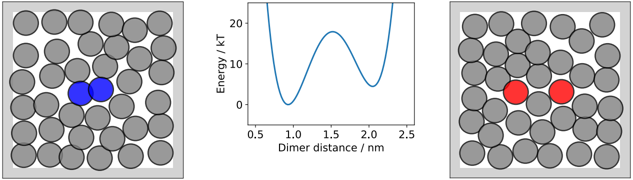
Different training methods



Architecture $\mathrm{N_{10}S}$, $nl_{layers} = 2$, $nl_{hidden} = 100$, $\sigma = \tanh$

[1]https://arxiv.org/abs/1812.01729

Boltzmann Generator - Particle dimer

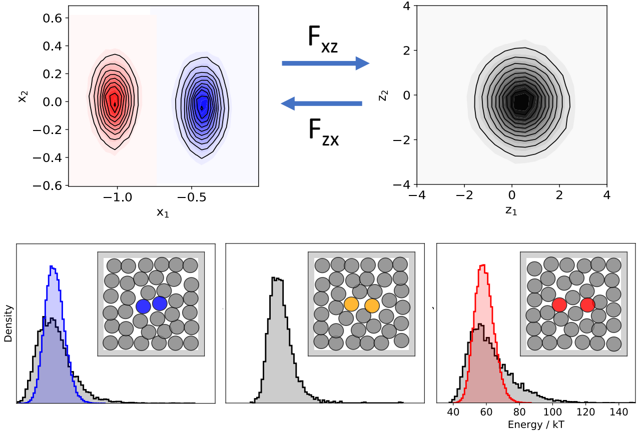


- Configuration: $\mathbf{x} = \left[\mathbf{x}_{1x}, \mathbf{x}_{1y}, \mathbf{x}_{2x}, \mathbf{x}_{2y}, \ldots, \mathbf{x}_{(n_s+2)x}, \mathbf{x}_{(n_s+2)y}\right]$
- Dimer distance $d = \|\mathbf{x}_1 - \mathbf{x}_2\|$, **potential energy**:

$$
\begin{aligned}
U(\mathbf{x}) = \frac{1}{4}a(d-d_0)^4 - \frac{1}{2}b(d-d_0)^2 + c(d-d_0)^4 \\
+\varepsilon \sum_{i=1}^{n+1} \sum_{j=i+1, j\neq 2}^{n+2} \left(\frac{\sigma}{\|\mathbf{x}_i - \mathbf{x}_j\|}\right)^{12} \\
+\,\text{box potential}
\end{aligned}
$$

[1] https://arxiv.org/abs/1812.01729

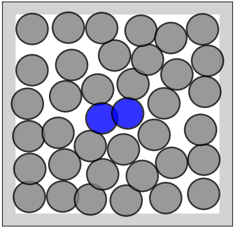
Boltzmann Generator - Particle dimer

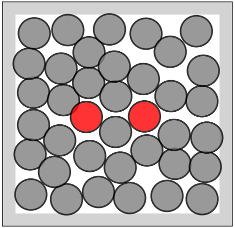


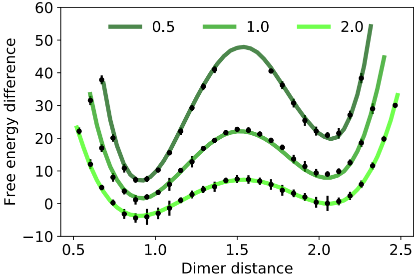
Architecture R_8, $nl_{layers} = 3$, $nl_{hidden} = 200$, $\sigma = \tanh$, $w_{KL} = 1$, $w_{ML} = 0.01$, $w_{RC} = 10$

[1] https://arxiv.org/abs/1812.01729

Boltzmann Generator - Particle dimer



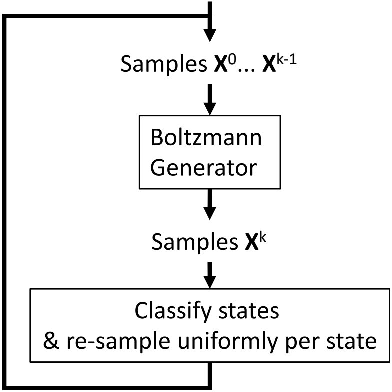




7 orders of magnitude speedup compared to MD

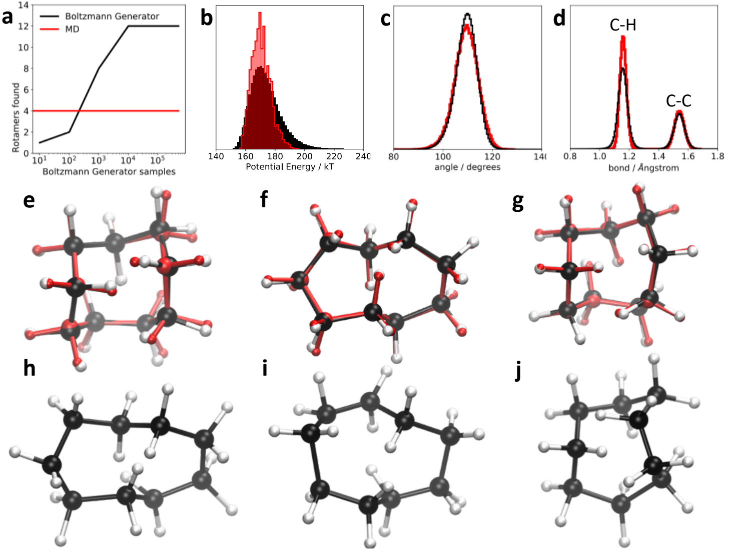
[1]https://arxiv.org/abs/1812.01729

Iterative exploration



[1] https://arxiv.org/abs/1812.01729

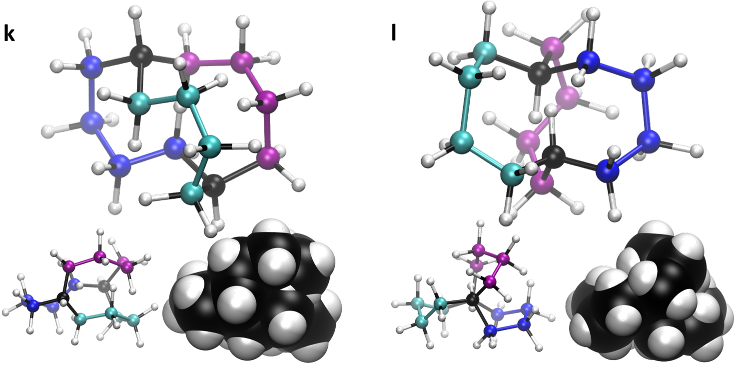
Boltzmann Generator - Hydrocarbons



$R_4,\ nl_{layers} = 2,\ nl_{hidden} = 100,\ \sigma = \tanh,\ w_{KL} = 0.01,\ w_{ML} = 1$

[1]https://arxiv.org/abs/1812.01729

Boltzmann Generator - Hydrocarbons



R_4, $nl_{layers} = 2$, $nl_{hidden} = 100$, $\sigma = \tanh$, $w_{KL} = 0.01$, $w_{ML} = 1$

[1] https://arxiv.org/abs/1812.01729

Acknowledgements

Collaborations

Cecilia Clementi (Rice University)
Christof Schütte (FU Berlin)
Eric Vanden-Eijnden (Courant Institut NY)
Thomas Weikl (MPI Potsdam)
Edina Rosta (King's College London)

Vijay Pande (Stanford)
Volker Haucke (FMP Berlin)
Stephan Sigrist (FU Berlin)
Katja, Faelber, Oliver Daumke (MDC)
John Chodera (MSKCC NY)
Gianni de Fabritiis (Barcelona)

Funding

erc
EINSTEIN Foundation.de
DFG Deutsche Forschungsgemeinschaft
Freie Universität Berlin

Positions available!