


Fantastic Path RND  
and Where to Find Them

Jiajun He, Francisco Vargas

# “Density Ratio” and Radon-Nikodym Derivative

 Don't freak out about the name Radon-Nikodym Derivative  
--- it's just the “*density ratio*”

 Very informally, let  $\mathbf{P}$  and  $\mathbf{Q}$  be two measures with density  $p$  and  $q$ , their density ratio is the **Radon-Nikodym Derivative (RND)**, denoted as

$$\frac{p(x)}{q(x)} = \frac{d\mathbf{P}}{d\mathbf{Q}}(x)$$

 The density is essentially the RND w.r.t to Lebesgue measure

$$p(x) = \frac{d\mathbf{P}}{d\mu}(x), q(x) = \frac{d\mathbf{Q}}{d\mu}(x)$$

 RND is helpful for spaces without Lebesgue measure

# Stochastic Differential Equations

▶▶ Forward SDE

$$dX_t = f(X_t, t)dt + \sigma_t dW_t$$

◀◀ Backward SDE

$$dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t$$

💡 Intuitive understanding by **Eular-Maruyama Discretisation**:

$$X_{n+1} - X_n = f(X_n, t_n)\Delta t + \sigma_n \sqrt{\Delta t} \epsilon$$

$$X_{n+1} - X_n = g(X_{n+1}, t_{n+1})\Delta t + \sigma_{n+1} \sqrt{\Delta t} \epsilon'$$

# From Gaussian Density Ratio to Path RND

$$X_{n+1} - X_n = f(X_n, t_n)\Delta t + \sigma_n\sqrt{\Delta t}\epsilon$$

? for a **discretised** path sample  $\{X_1, X_2, \dots, X_N\}$ , what is its density?

✓ **Transition density:**  $p(X_{n+1}|X_n) = N(X_{n+1}|X_n + f(X_n, t_n)\Delta t, \sigma_n^2\Delta t)$

✓ **Full path density:**  $p(X_1, X_2, \dots, X_N) = p(X_1)\prod p(X_{n+1}|X_n)$

# From Gaussian Density Ratio to Path RND

Now take a closer look at

$$N(X_{n+1} | X_n + f(X_n, t_n)\Delta t, \sigma_n^2 \Delta t)$$

$$\log p = \frac{-(\sigma_n \sqrt{\Delta t} \epsilon)^2}{2\sigma_n^2 \Delta t} - \log \sigma_n - \frac{1}{2} \log \Delta t + C$$

 density diverge when  $\Delta t \rightarrow 0$

# From Gaussian Density Ratio to Path RND

But what if we have another SDE:

$$p_1 = N(X_{n+1} | X_n + f(X_n, t_n)\Delta t, \sigma_n^2 \Delta t)$$

$$p_2 = N(X_{n+1} | X_n + h(X_n, t_n)\Delta t, \sigma_n^2 \Delta t)$$

$$\log p_1 - \log p_2 = \frac{(2X_{n+1} - 2X_n - h\Delta t - f\Delta t)(h\Delta t - f\Delta t)}{2\sigma_n^2 \Delta t}$$

 density ratio did NOT diverge when  $\Delta t \rightarrow 0$

# From Gaussian Density Ratio to Path RND

For solution  $X$  to one SDE:  $dX_t = f(X_t, t)dt + \sigma_t dW_t$ ,

we cannot define its density  $p(X_0) \prod p(X_{t+dt} | X_t)$

But with another SDE:  $dX_t = h(X_t, t)dt + \sigma_t dW_t$ ,

we can define density ratio (**Radon-Nikodym Derivative**) as a whole:

$$\frac{d\mathbf{P}}{d\mathbf{Q}}(X)$$

# Forward-forward RND and Girsanov

$$\begin{aligned}\mathbf{P} : dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \mathbf{Q} : dX_t &= h(X_t, t)dt + \sigma_t dW_t, X_0 \sim q_0\end{aligned}$$

$$\frac{d\mathbf{P}}{d\mathbf{Q}}(X) \approx \frac{p(X_0) \prod N_1(X_{n+1}|X_n)}{q(X_0) \prod N_2(X_{n+1}|X_n)}$$



# Forward-forward RND and Girsanov

$$\begin{aligned} \mathbf{P} : dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \mathbf{Q} : dX_t &= h(X_t, t)dt + \sigma_t dW_t, X_0 \sim q_0 \end{aligned}$$

$$\frac{d\mathbf{P}}{d\mathbf{Q}}(X) = \underbrace{\frac{p(X_0)}{q(X_0)}}_{\text{Initial density ratio}} \exp \left( \underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t - \frac{f_t^2(X_t)}{2\sigma_t^2} dt}_{\text{Forward Ito Integral}} - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot dX_t + \frac{g_t^2(X_t)}{2\sigma_t^2} dt \right)$$

Forward Ito Integral  $\int a_t(X_t) \cdot dX_t = \lim \sum a_n(X_n) \cdot (X_{n+1} - X_n)$

# Forward-backward RND

$$\begin{aligned} \mathbf{P} : dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \overleftarrow{\mathbf{Q}} : dX_t &= g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim q_1 \end{aligned}$$

$$\frac{d\mathbf{P}}{d\overleftarrow{\mathbf{Q}}}(X) \approx \frac{p_0(X_0) \prod N_1(X_{n+1}|X_n)}{q_1(X_1) \prod N_2(X_n|X_{n+1})}$$

# Forward-backward RND

$$\begin{aligned} \mathbf{P} : dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \overleftarrow{\mathbf{Q}} : dX_t &= g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim q_1 \end{aligned}$$

$$\frac{d\mathbf{P}}{d\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp \left( \int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t - \frac{f_t^2(X_t)}{2\sigma_t^2} dt - \underbrace{\int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t + \frac{g_t^2(X_t)}{2\sigma_t^2} dt}_{\text{Backward Ito Integral}} \right)$$

$$\int a_t(X_t) \cdot \overleftarrow{dX}_t = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

# Forward-backward RND

$$\begin{aligned} \mathbf{P} : dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \overleftarrow{\mathbf{Q}} : dX_t &= g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim q_1 \end{aligned}$$

$$\frac{d\mathbf{P}}{d\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp \left( \underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t - \frac{f_t^2(X_t)}{2\sigma_t^2} dt - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t + \frac{g_t^2(X_t)}{2\sigma_t^2} dt}_{\lim \frac{\prod N_1(X_{n+1}|X_n)}{\prod N_2(X_n|X_{n+1})}} \right)$$

# A Side Note on Stochastic Integrals

Ito forward integral

$$\int a_t(X_t) \cdot dX_t = \lim \sum a_n(X_n) \cdot (X_{n+1} - X_n)$$

Ito backward integral

$$\int a_t(X_t) \cdot \overleftarrow{dX}_t = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

Stratonovich integral

$$\int a_t(X_t) \circ dX_t = \lim \sum \frac{a_n(X_n) + a_{n+1}(X_{n+1})}{2} \cdot (X_{n+1} - X_n)$$

# A Side Note on Stochastic Integrals

Ito forward integral

$$\int a_t(X_t) \cdot dX_t = \lim \sum a_n(X_n) \cdot (X_{n+1} - X_n)$$

Ito backward integral

$$\int a_t(X_t) \cdot \overleftarrow{dX}_t = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

Conversion rule:

$$\int a_t(X_t) \cdot dX_t - \int a_t(X_t) \cdot \overleftarrow{dX}_t = - \int \sigma_t^2 \nabla \cdot a_t dt$$

# Time-reversal and Nelson's relation

$$\begin{aligned} \mathbf{P} : dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \overleftarrow{\mathbf{Q}} : dX_t &= g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim p_1 \end{aligned}$$

$$\overleftarrow{\mathbf{Q}} = \mathbf{P}, \text{ i. e. }, \frac{\overleftarrow{d\mathbf{Q}}}{d\mathbf{P}} = 1$$

iff

$$g(\cdot, t) = f(\cdot, t) - \sigma_t^2 \nabla \log p_t(\cdot)$$

# Time-reversal and Nelson's relation

$$\begin{aligned} \mathbf{P} : dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \overleftarrow{\mathbf{Q}} : dX_t &= g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim p_1 \end{aligned}$$

$$\overleftarrow{\mathbf{Q}} = \mathbf{P}, \text{ i. e., } \frac{\overleftarrow{d\mathbf{Q}}}{d\mathbf{P}} = 1$$

$$g(\cdot, t) = f(\cdot, t) \overset{\text{iff}}{-} \sigma_t^2 \nabla \log p_t(\cdot)$$

e.g., 0 in VE process      score



# Where to Find Path RND?

Path RND is no different than other Density Ratios

👉 You can do **Importance Sampling**

👉 You can do **Variational Inference**

lead to a variety of methods!

# Importance Sampling with Path RND

**Importance Sampling:**

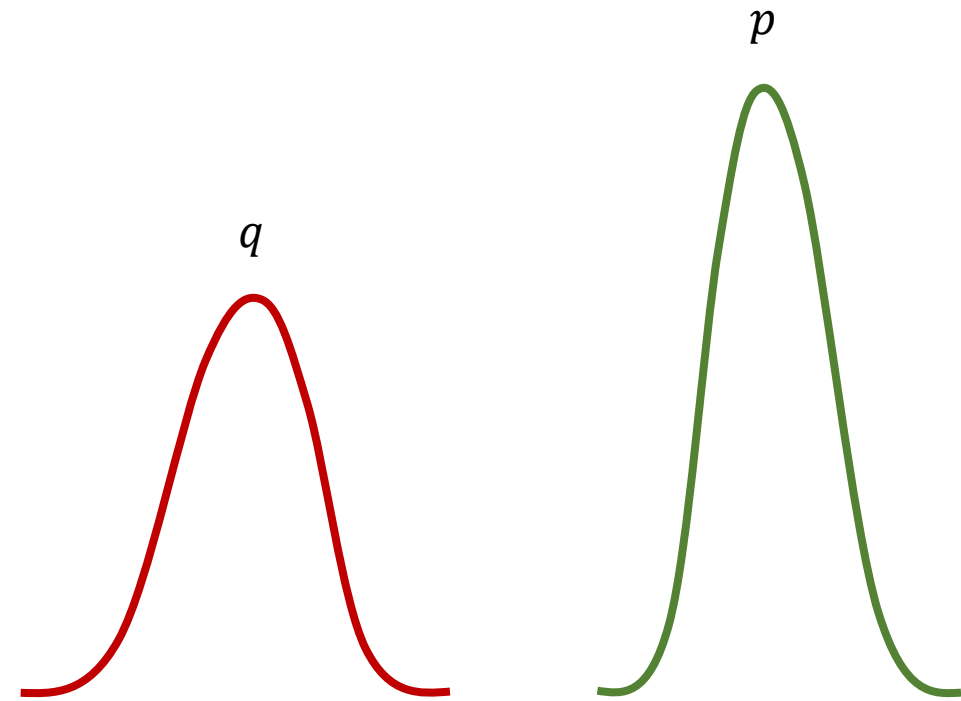
$$\mathbf{E}_{x \sim p}[f(x)] = \mathbf{E}_{x \sim q} \left[ \frac{p(x)}{q(x)} f(x) \right]$$

**Importance Sampling in Path space:**

$$\mathbf{E}_{X \sim \mathbf{P}}[f(X)] = \mathbf{E}_{X \sim \mathbf{Q}} \left[ \frac{d\mathbf{P}}{d\mathbf{Q}}(X) f(X) \right]$$

# Importance Sampling with Path RND: AIS

Estimate:  $\mathbf{E}_{x \sim p}[f(x)]$  with proposal  $q$



# Importance Sampling with Path RND: AIS

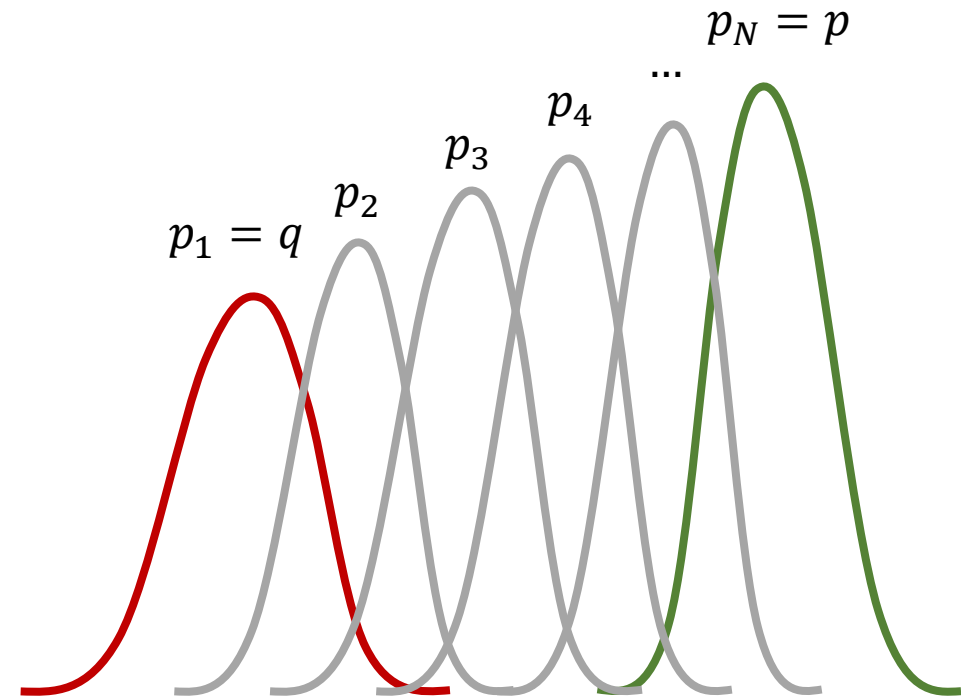
Estimate:  $\mathbf{E}_{x \sim p}[f(x)]$  with proposal  $q$

$X_1 \sim p_1$      $X_t \sim \text{MCMC}_{p_t}(X_{t-1})$     **“Forward Process”**

$X_N \sim p_N$      $X_{t-1} \sim \text{MCMC}_{p_t}(X_t)$     **“Backward Process”**

$$\begin{aligned} \mathbf{E}_{x \sim p}[f(x)] &= \int f(x_N) p(x_N) \prod p(x_{n-1} | x_n) dx \\ &= \int \frac{f(x_N) p(x_N) \prod p(x_{n-1} | x_n)}{q(x_1) \prod p(x_n | x_{n-1})} q(x_1) \prod p(x_n | x_{n-1}) dx \end{aligned}$$

**Proposal**



# Importance Sampling with Path RND: AIS

Estimate:  $\mathbf{E}_{x \sim p}[f(x)]$  with proposal  $q$

$$X_1 \sim p_1$$

$$X_t \sim \text{ULA}(X_{t-1})$$

“Forward Process”

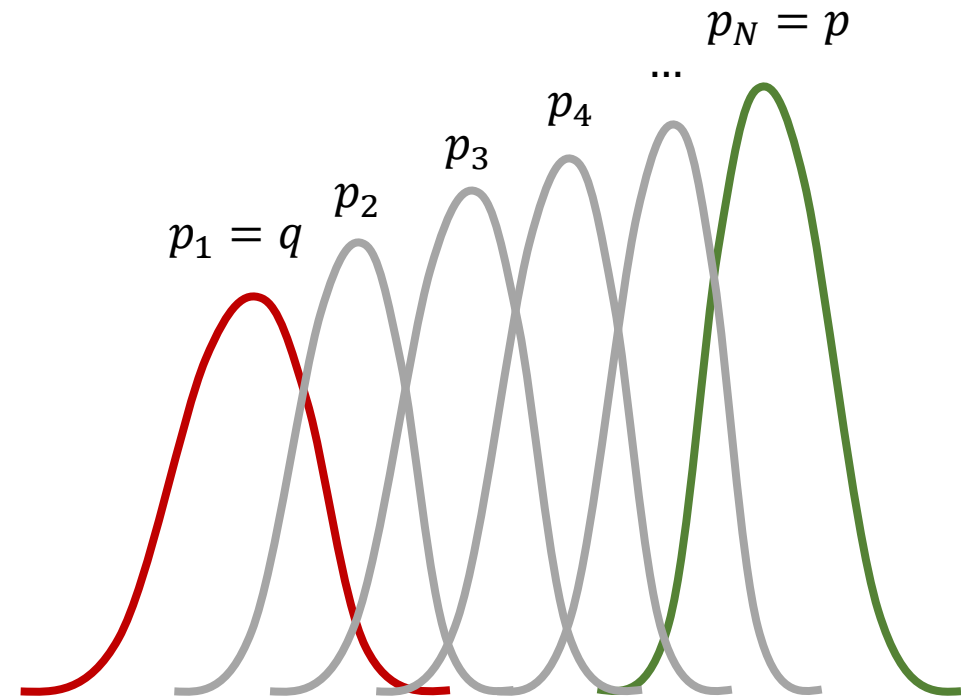
$$X_N \sim p_N$$

$$X_{t-1} \sim \text{ULA}(X_t)$$

“Backward Process”

$$\begin{aligned} \mathbf{E}_{x \sim p}[f(x)] &= \int f(x_N) p(x_N) \prod N(x_{n-1} | x_n) dx \\ &= \int \frac{f(x_N) p(x_N) \prod N(x_{n-1} | x_n)}{q(x_1) \prod N(x_n | x_{n-1})} q(x_1) \prod N(x_n | x_{n-1}) dx \end{aligned}$$

Proposal



# Importance Sampling with Path RND: AIS

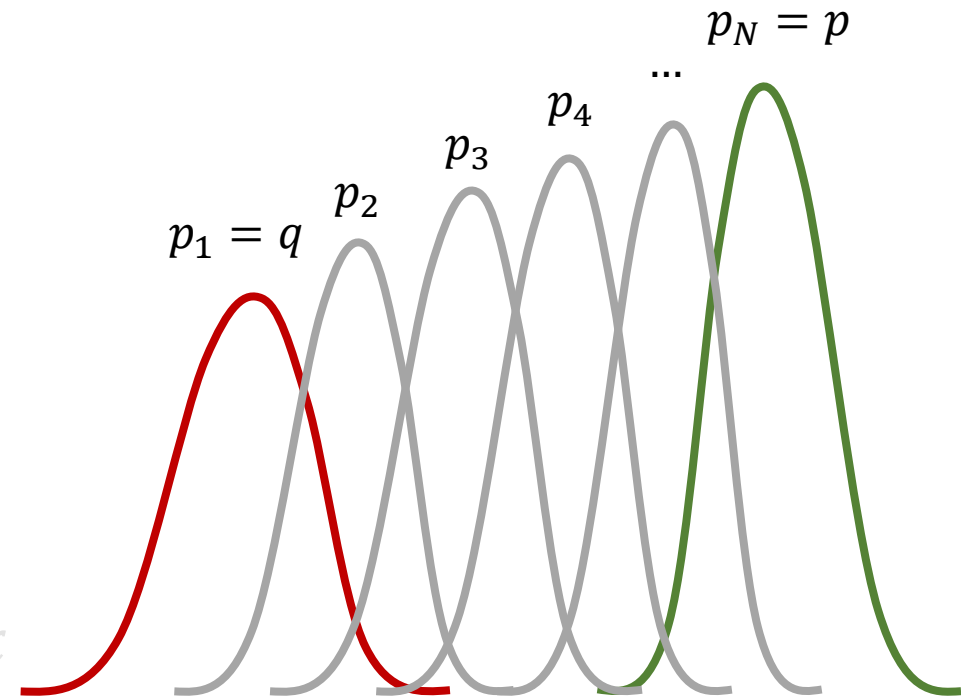
Estimate:  $\mathbf{E}_{x \sim p}[f(x)]$  with proposal  $q$

$X_1 \sim p_1$        $X_t \sim \text{ULA}(X_{t-1})$       “Forward Process”

$X_N \sim p_N$        $X_{t-1} \sim \text{ULA}(X_t)$       “Backward Process”

Taking the limit... ( $\infty$  intermediate distributions)

$$\begin{aligned} \mathbf{E}_{x \sim p}[f(x)] &= \int f(x_N) p(x_N) \prod N(x_{n-1} | x_n) dx \\ &= \int f(x_N) \frac{p(x_N) \prod N(x_{n-1} | x_n)}{q(x_1) \prod N(x_n | x_{n-1})} q(x_1) \prod N(x_n | x_{n-1}) dx \end{aligned}$$



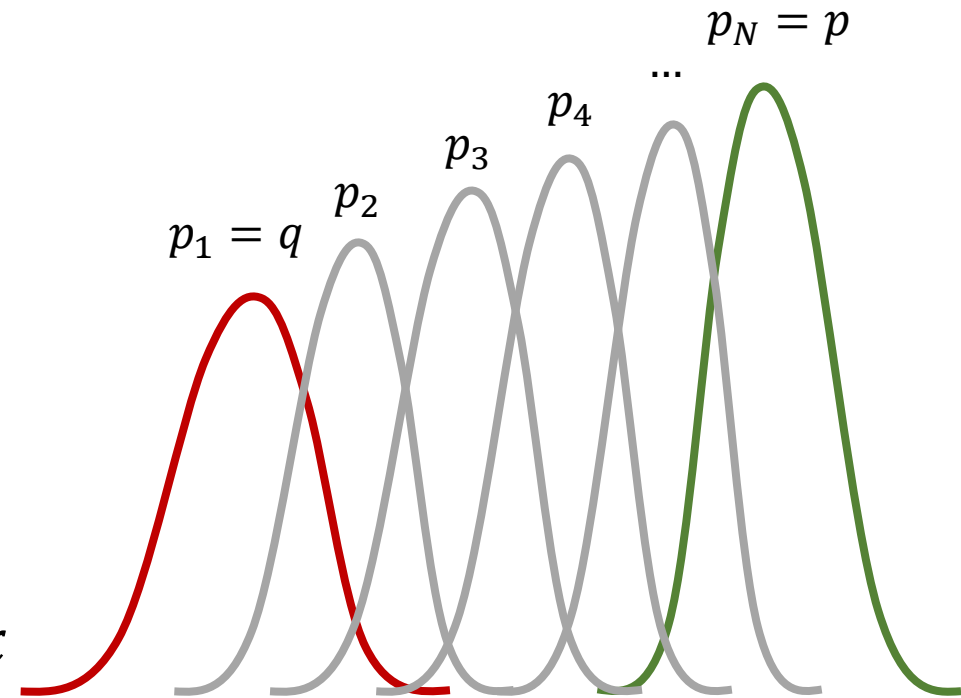
# Importance Sampling with Path RND: AIS

Estimate:  $\mathbf{E}_{x \sim p}[f(x)]$  with proposal  $q$

$X_1 \sim q = p_1$      $X_t \sim \text{ULA}(X_{t-1})$     **“Forward Process”**

$X_N \sim p = p_N$      $X_{t-1} \sim \text{ULA}(X_t)$     **“Backward Process”**

$$\begin{aligned} \mathbf{E}_{x \sim p}[f(x)] &= \int f(x_N) p(x_N) \prod N(x_{n-1} | x_n) dx \\ &= \int f(x_N) \frac{p(x_N) \prod N(x_{n-1} | x_n)}{q(x_1) \prod N(x_n | x_{n-1})} q(x_1) \prod N(x_n | x_{n-1}) dx \end{aligned}$$



# Importance Sampling with Path RND: AIS

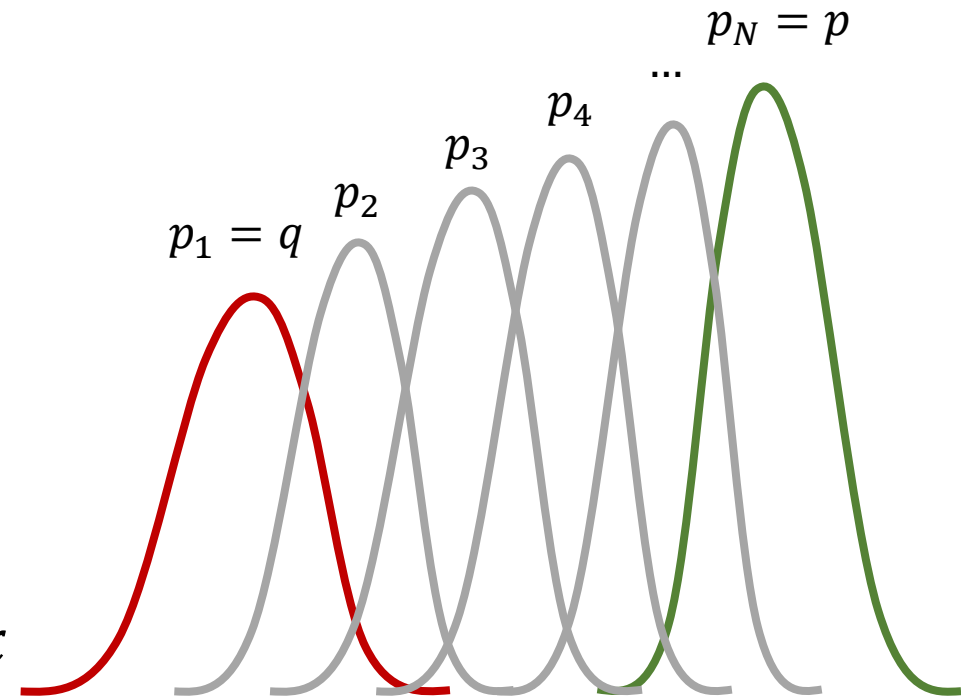
Estimate:  $\mathbf{E}_{x \sim p}[f(x)]$  with proposal  $q$

$$X_1 \sim q = p_1 \quad X_t \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t, \text{ "Euler-Arnold Process"}$$

$$X_N \sim p = p_N \quad X_{t-1} \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t, \text{ "Euler-Arnold Process"}$$

$$\mathbf{E}_{x \sim p}[f(x)] = \int f(x_N) p(x_N) \prod N(x_{n-1} | x_n) dx$$

$$= \int f(x_N) \frac{p(x_N) \prod N(x_{n-1} | x_n)}{q(x_1) \prod N(x_n | x_{n-1})} q(x_1) \prod N(x_n | x_{n-1}) dx$$





# Importance Sampling with Path RND: AIS

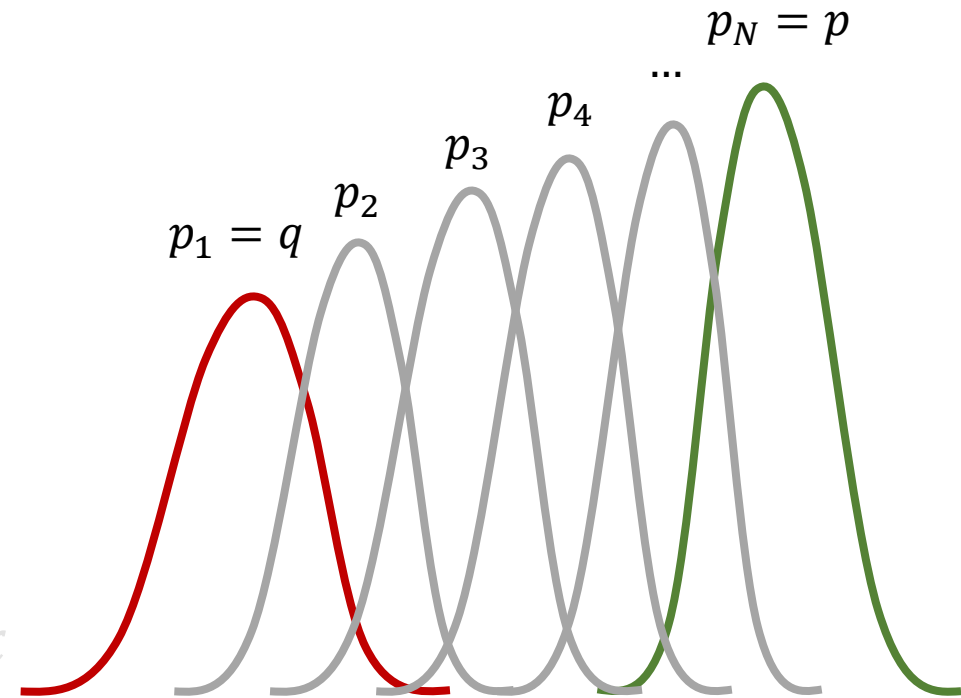
Estimate:  $\mathbf{E}_{x \sim p}[f(x)]$  with proposal  $q$

$$X_1 \sim q = p_1 \quad X_t \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t, \text{ "Euler-Maruyama Process"}$$

$$X_N \sim p = p_N \quad X_{t-1} \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t, \text{ "Euler-Maruyama Process"}$$

$$\mathbf{E}_{x \sim p}[f(x)] = \int f(x_N) p(x_N) \prod N(x_{n-1} | x_n) dx$$

$$= \int f(x_N) \frac{p(x_N) \prod \overleftarrow{dP}(x_{n-1} | x_n)}{q(x_1) \prod \overrightarrow{dQ}(x_n | x_{n-1})} q(x_1) \prod N(dQ | x_{n-1}) dx$$



$$\mathbf{E}_{X \sim P}[f(X)] = \mathbf{E}_{X \sim Q} \left[ \frac{dP}{dQ}(X) f(X) \right]$$

# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{\overrightarrow{dQ}} =$$

# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{p(X_1)}{q(X_0)} \exp \left( \int \frac{\nabla U_t}{2} \cdot dX_t + \frac{\sigma_t^2}{4} |\nabla U_t|^2 dt + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{dX}_t - \frac{\sigma_t^2}{4} |\nabla U_t|^2 dt \right)$$

# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\begin{aligned} \frac{\overleftarrow{dP}}{dQ} &= \frac{p(X_1)}{q(X_0)} \exp \left( \int \frac{\nabla U_t}{2} \cdot dX_t + \frac{\sigma_t^2}{4} |\nabla U_t|^2 dt + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{dX}_t - \frac{\sigma_t^2}{4} |\nabla U_t|^2 dt \right) \\ &= \frac{p(X_1)}{q(X_0)} \exp \left( \int \frac{\nabla U_t}{2} \cdot dX_t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{dX}_t \right) \end{aligned}$$

# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\begin{aligned} \frac{\overleftarrow{dP}}{dQ} &= \frac{p(X_1)}{q(X_0)} \exp \left( \int \frac{\nabla U_t}{2} \cdot dX_t + \frac{\sigma_t^2}{4} |\nabla U_t|^2 dt + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{dX}_t - \frac{\sigma_t^2}{4} |\nabla U_t|^2 dt \right) \\ &= \frac{p(X_1)}{q(X_0)} \exp \left( \int \frac{\nabla U_t}{2} \cdot dX_t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{dX}_t \right) \quad \text{👉 conversion rule} \end{aligned}$$

# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{p(X_1)}{q(X_0)} \exp \left( \int \frac{\nabla U_t}{2} \cdot dX_t + \frac{\sigma_t^2}{4} |\nabla U_t|^2 dt + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{dX}_t - \frac{\sigma_t^2}{4} |\nabla U_t|^2 dt \right)$$

$$= \frac{p(X_1)}{q(X_0)} \exp \left( \int \frac{\nabla U_t}{2} \cdot dX_t + \int \frac{\nabla U_t}{2} \cdot \overleftarrow{dX}_t \right)$$

 conversion rule

$$= \frac{p(X_1)}{q(X_0)} \exp \left( \int \nabla U_t \cdot dX_t + \int \sigma_t^2 \Delta U_t dt \right)$$

# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{p(X_1)}{q(X_0)} \exp \left( \int \nabla U_t \cdot dX_t + \int \sigma_t^2 \Delta U_t dt \right)$$

# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{p(X_1)}{q(X_0)} \exp\left(\int \nabla U_t \cdot dX_t + \int \sigma_t^2 \Delta U_t dt\right) \quad \text{👉 Ito's lemma}$$
$$df_t(X_t) = (\partial_t f(X_t) + \sigma_t^2 \Delta f) dt + \nabla U_t \cdot dX_t$$



# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\begin{aligned} \frac{\overleftarrow{dP}}{dQ} &= \frac{p(X_1)}{q(X_0)} \exp \left( \int \nabla U_t \cdot dX_t + \int \sigma_t^2 \Delta U_t dt \right) \\ &= \frac{p(X_1)}{q(X_0)} \exp \left( \int dU_t(X_t) - \partial_t U_t(X_t) dt \right) \end{aligned}$$

# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\begin{aligned} \frac{\overleftarrow{dP}}{dQ} &= \frac{p(X_1)}{q(X_0)} \exp \left( \int \nabla U_t \cdot dX_t + \int \sigma_t^2 \Delta U_t dt \right) \\ &= \frac{p(X_1)}{q(X_0)} \exp \left( \int dU_t(X_t) - \partial_t U_t(X_t) dt \right) \\ &= \frac{p(X_1)}{q(X_0)} \exp \left( U_1(X_1) - U_0(X_0) + \int -\partial_t U_t(X_t) dt \right) \end{aligned}$$

# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\begin{aligned} \frac{\overleftarrow{dP}}{dQ} &= \frac{p(X_1)}{q(X_0)} \exp \left( \int \nabla U_t \cdot dX_t + \int \sigma_t^2 \Delta U_t dt \right) \\ &= \frac{p(X_1)}{q(X_0)} \exp \left( \int dU_t(X_t) - \partial_t U_t(X_t) dt \right) \\ &= \frac{Z_0 \exp(-U_1(X_1))}{Z_1 \exp(-U_0(X_0))} \exp \left( U_1(X_1) - U_0(X_0) + \int -\partial_t U_t(X_t) dt \right) \end{aligned}$$

# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\begin{aligned} \frac{\overleftarrow{dP}}{dQ} &= \frac{p(X_1)}{q(X_0)} \exp \left( \int \nabla U_t \cdot dX_t + \int \sigma_t^2 \Delta U_t dt \right) \\ &= \frac{p(X_1)}{q(X_0)} \exp \left( \int dU_t(X_t) - \partial_t U_t(X_t) dt \right) \\ &= \frac{Z_0 \exp(\cancel{U_1(X_1)})}{Z_1 \exp(\cancel{-U_0(X_0)})} \exp \left( \cancel{U_1(X_1)} - \cancel{U_0(X_0)} + \int -\partial_t U_t(X_t) dt \right) \end{aligned}$$

# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\begin{aligned} \frac{\overleftarrow{dP}}{dQ} &= \frac{p(X_1)}{q(X_0)} \exp \left( \int \nabla U_t \cdot dX_t + \int \sigma_t^2 \Delta U_t dt \right) \\ &= \frac{p(X_1)}{q(X_0)} \exp \left( \int dU_t(X_t) - \partial_t U_t(X_t) dt \right) \\ &= \frac{Z_0}{Z_1} \exp \left( \int -\partial_t U_t(X_t) dt \right) \end{aligned}$$

# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{\overrightarrow{dQ}} = \frac{Z_0}{Z_1} \exp\left(\int -\partial_t U_t(X_t) dt\right)$$

 Crooks Fluctuation Theorem

# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{Z_0}{Z_1} \exp\left(\int -\partial_t U_t(X_t) dt\right) \quad \text{💡 Crooks Fluctuation Theorem}$$

$$\mathbf{E}_Q \left[ \frac{\overleftarrow{dP}}{dQ} \right] = \mathbf{E}_Q \left[ \frac{Z_0}{Z_1} \exp\left(\int -\partial_t U_t(X_t) dt\right) \right] = 1$$

# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{Z_0}{Z_1} \exp\left(\int -\partial_t U_t(X_t) dt\right)$$

 Crooks Fluctuation Theorem

$$\mathbf{E}_Q \left[ \exp\left(\int -\partial_t U_t(X_t) dt\right) \right] = \frac{Z_1}{Z_0}$$



# From AIS to Jarzynski Equation & Crooks Theorem

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{Z_0}{Z_1} \exp\left(\int -\partial_t U_t(X_t) dt\right)$$

💡 Crooks Fluctuation Theorem

$$\mathbf{E}_Q \left[ \exp\left(\int -\partial_t U_t(X_t) dt\right) \right] = \frac{Z_1}{Z_0}$$

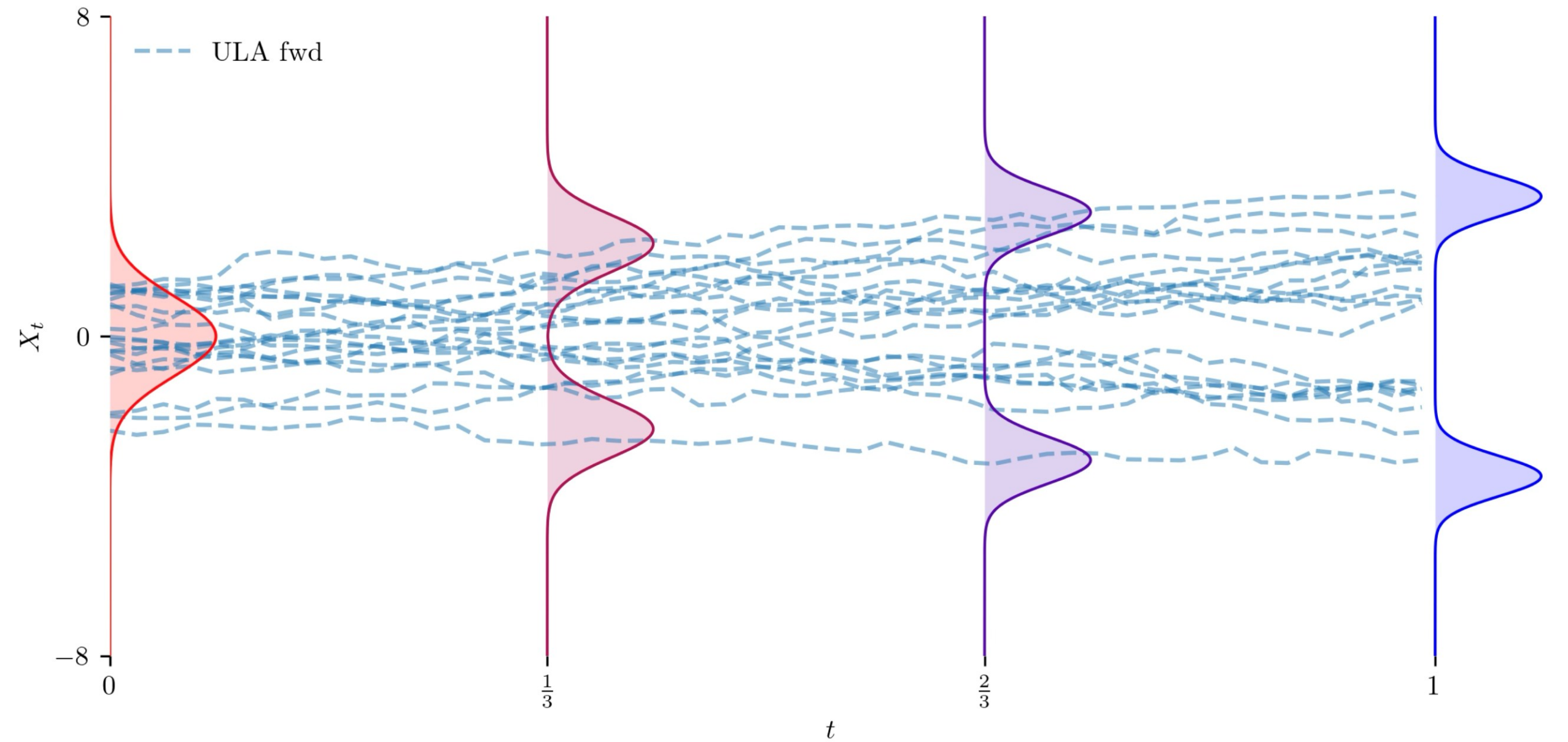
💡 Jarzynski Equation

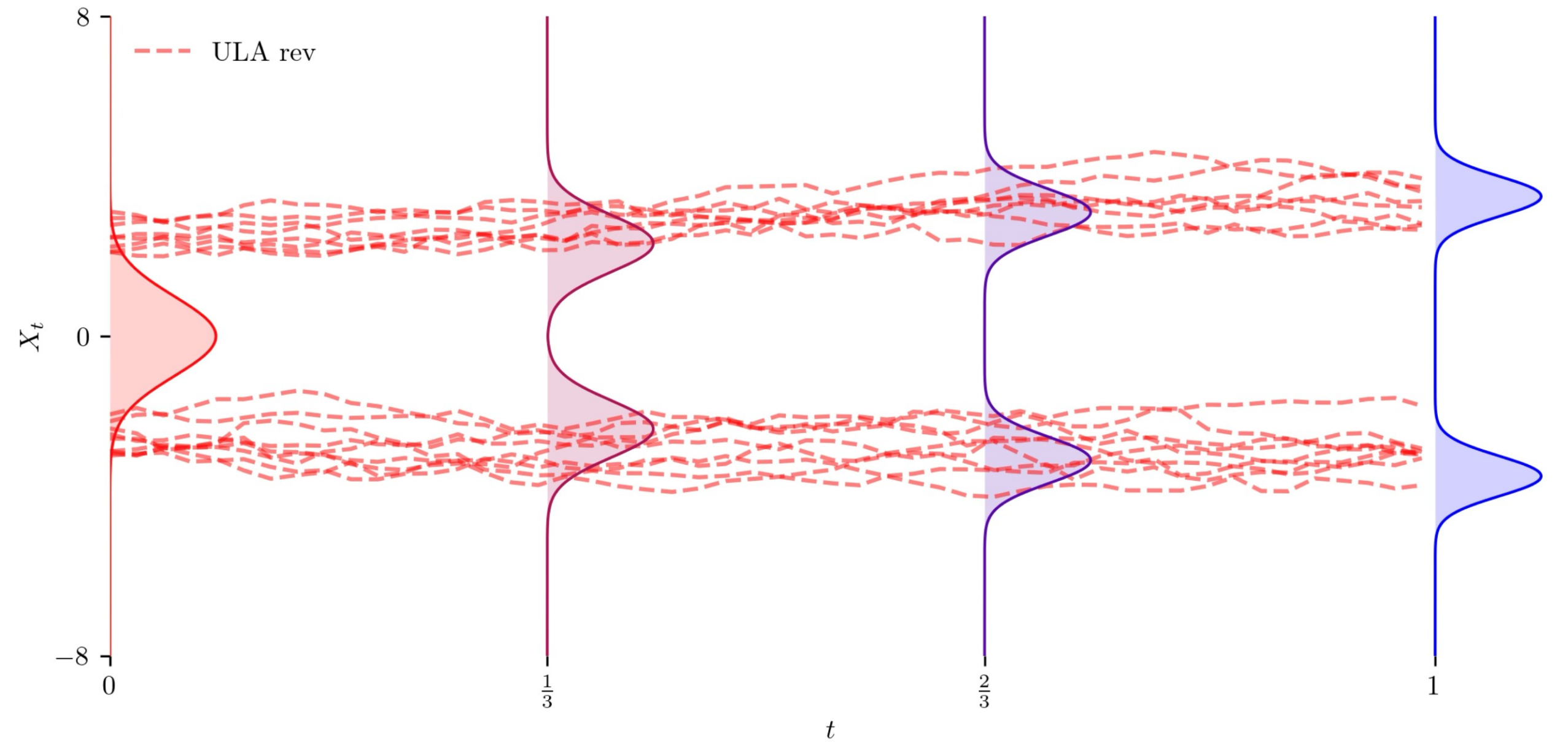
# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = -\sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = \sigma^2 \nabla U_t(X_t) dt + \sigma \sqrt{2} \overleftarrow{dW}_t,$$

The evolving of samples is slower than that of energy



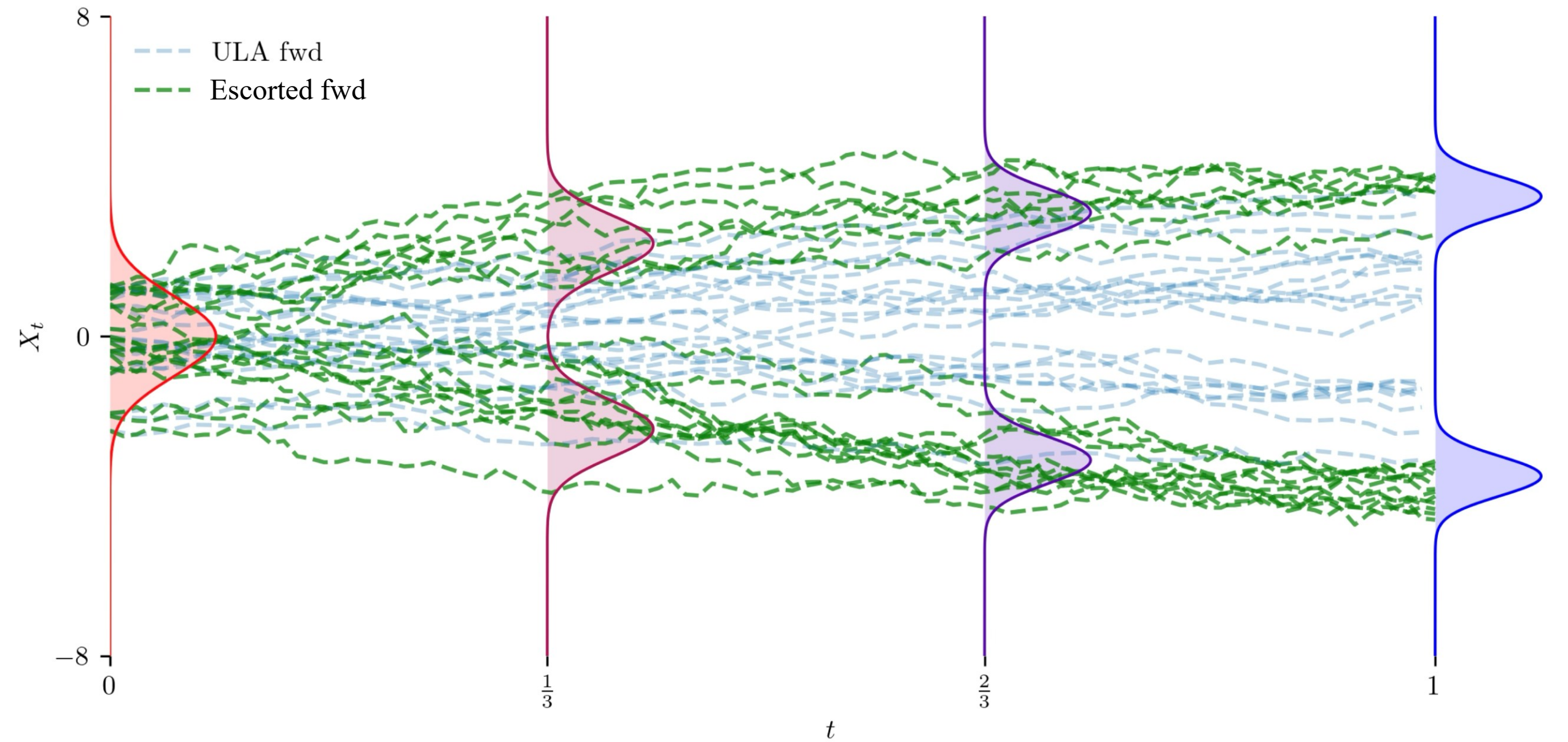


# From Jarzynski to Escorted Jarzynski

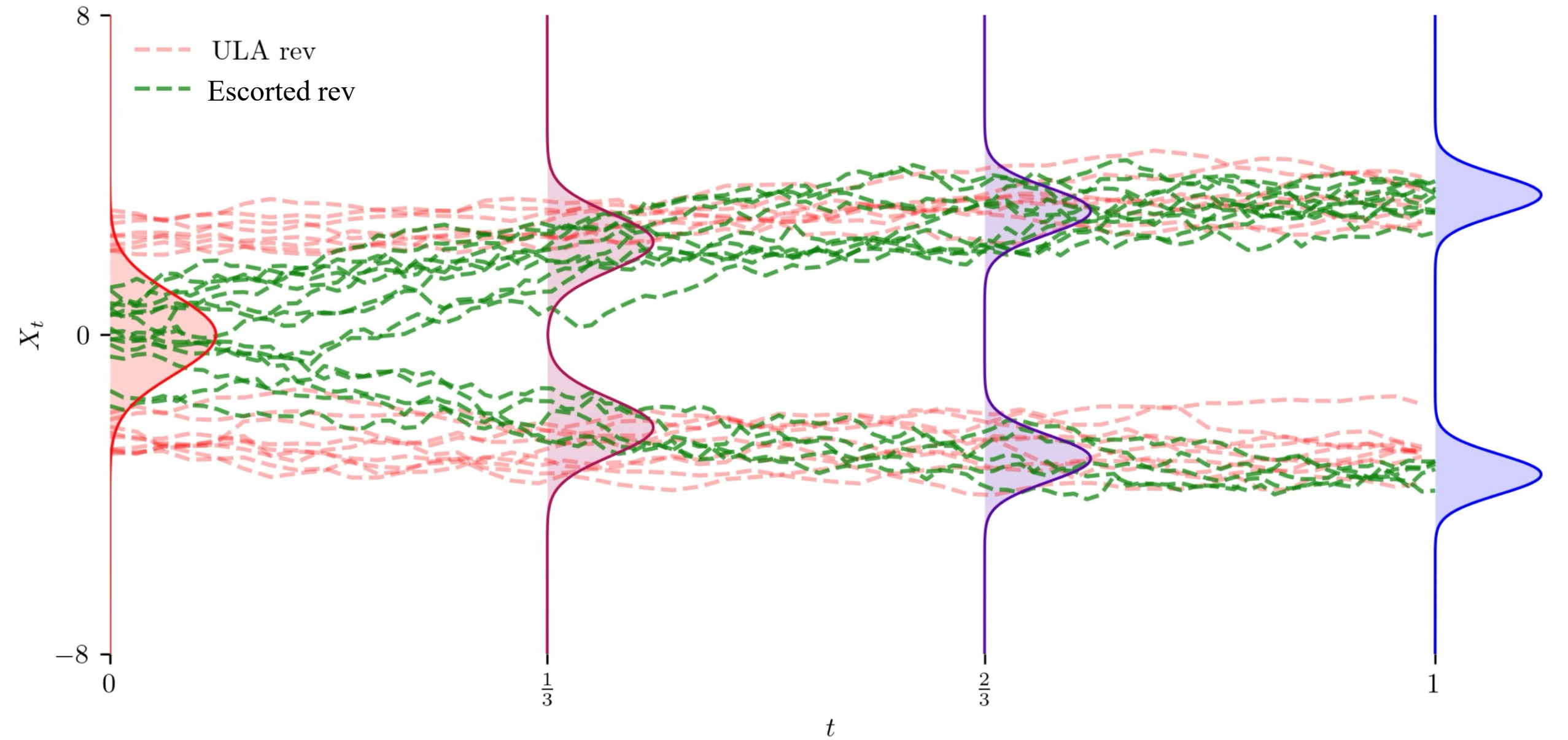
$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

The evolving of samples is closer to than that of energy







# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} =$$



# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{p(X_1)}{q(X_0)} \exp \left( \int \frac{\sigma^2 \nabla U_t - u_t}{2\sigma^2} \cdot dX_t + \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t - u_t|^2 dt + \int \frac{\sigma^2 \nabla U_t + u_t}{2\sigma^2} \cdot \overleftarrow{dX}_t - \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t + u_t|^2 dt \right)$$

# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} d\overrightarrow{W}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} d\overleftarrow{W}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{p(X_1)}{q(X_0)} \exp \left( \int \frac{\sigma^2 \nabla U_t - u_t}{2\sigma^2} \cdot dX_t + \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t - u_t|^2 dt + \int \frac{\sigma^2 \nabla U_t + u_t}{2\sigma^2} \cdot d\overleftarrow{X}_t - \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t + u_t|^2 dt \right)$$

...conversion rule...

# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{p(X_1)}{q(X_0)} \exp \left( \int \frac{\sigma^2 \nabla U_t - u_t}{2\sigma^2} \cdot dX_t + \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t - u_t|^2 dt + \int \frac{\sigma^2 \nabla U_t + u_t}{2\sigma^2} \cdot \overleftarrow{dX}_t - \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t + u_t|^2 dt \right)$$

...conversion rule...

...Ito's Lemma...

# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{p(X_1)}{q(X_0)} \exp \left( \int \frac{\sigma^2 \nabla U_t - u_t}{2\sigma^2} \cdot dX_t + \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t - u_t|^2 dt + \int \frac{\sigma^2 \nabla U_t + u_t}{2\sigma^2} \cdot \overleftarrow{dX}_t - \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t + u_t|^2 dt \right)$$

...conversion rule...

...Ito's Lemma...

...cancel  $U_1$  and  $U_0$ ...

# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} d\overrightarrow{W}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} d\overleftarrow{W}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{p(X_1)}{q(X_0)} \exp \left( \int \frac{\sigma^2 \nabla U_t - u_t}{2\sigma^2} \cdot dX_t + \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t - u_t|^2 dt + \int \frac{\sigma^2 \nabla U_t + u_t}{2\sigma^2} \cdot d\overleftarrow{X}_t - \frac{1}{4\sigma_t^2} |\sigma_t^2 \nabla U_t + u_t|^2 dt \right)$$

...conversion rule...

...Ito's Lemma...

...cancel  $U_1$  and  $U_0$ ...

$$\frac{\overleftarrow{dP}}{dQ} = \frac{Z_0}{Z_1} \exp \left( \int -\partial_t U_t(X_t) dt - \nabla U_t \cdot u_t dt + \nabla \cdot u_t dt \right)$$

# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} d\overrightarrow{W}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{d}W_t,$$

$$\frac{\overleftarrow{d}\mathbf{P}}{d\mathbf{Q}} = \frac{Z_0}{Z_1} \exp\left(\int -\partial_t U_t(X_t)dt - \nabla U_t \cdot u_t dt + \nabla \cdot u_t dt\right)$$

# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} d\overrightarrow{W}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} d\overleftarrow{W}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{Z_0}{Z_1} \exp \left( \int -\partial_t U_t(X_t) dt - \nabla U_t \cdot u_t dt + \nabla \cdot u_t dt \right)$$

 Controlled Crooks Fluctuation Theorem

# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{Z_0}{Z_1} \exp\left(\int -\partial_t U_t(X_t)dt - \nabla U_t \cdot u_t dt + \nabla \cdot u_t dt\right)$$

💡 Controlled Crooks Fluctuation Theorem

$$\mathbf{E}_Q \left[ \exp\left(\int -\partial_t U_t(X_t)dt - \nabla U_t \cdot u_t dt + \nabla \cdot u_t dt\right) \right] = \frac{Z_1}{Z_0}$$

💡 Escorted Jarzynski Equation



# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

$$\frac{\overleftarrow{dP}}{dQ} = \frac{Z_0}{Z_1} \exp\left(\int -\partial_t U_t(X_t)dt - \nabla U_t \cdot u_t dt + \nabla \cdot u_t dt\right)$$

💡 Controlled Crooks Fluctuation Theorem

$$\mathbf{E}_Q \left[ \exp\left(\int -\partial_t U_t(X_t)dt - \nabla U_t \cdot u_t dt + \nabla \cdot u_t dt\right) \right] = \frac{Z_1}{Z_0}$$

💡 Escorted Jarzynski Equation

**Can also be derived via PDEs [1] / Feynman-Kac formula [2]:**

[1] Albergo, M. S., & Vanden-Eijnden, E (2025). NETS: A Non-equilibrium Transport Sampler. *ICML 2025*.

[2] Skreta, M., Akhound-Sadegh, T., Ohanesian, V., Bondesan, R., Aspuru-Guzik, A., Doucet, A., ... & Neklyudov, K. (2025).

Feynman-kac correctors in diffusion: Annealing, guidance, and product of experts. *ICML 2025*.

# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

**Question: How to find  $u_t$ ?**

# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

**Question: How to find  $u_t$ ?**

👉 Setting 1: access sample for  $q$  + density of  $q$  + energy of  $p$  (neural samplers)

# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

**Question: How to find  $u_t$ ?**

👉 Setting 1: access sample for  $q$  + density of  $q$  + energy of  $p$  (neural samplers)

💡 match the forward & backward process

# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

**Question: How to find  $u_t$ ?**

👉 Setting 1: access sample for  $q$  + density of  $q$  + energy of  $p$  (neural samplers)

💡 match the forward & backward process

💡 match the marginal  $p_t$  of sampling process to  $U_t$

# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

**Question: How to find  $u_t$ ?**

👉 Setting 1: access sample for  $q$  + density of  $q$  + energy of  $p$  (neural samplers)

💡 match the forward & backward process

👉 CMCD [1]

💡 match the marginal  $p_t$  of sampling process to  $U_t$

👉 NETS [2]

[1] Vargas, F., Padhy, S., Blessing, D., & Nüsken, N. (2024). Transport meets variational inference: Controlled monte carlo diffusions. *ICLR 2024*.

[2] Albergo, M. S., & Vanden-Eijnden, E. (2025). Nets: A non-equilibrium transport sampler. *ICML 2025*.

# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

**Question: How to find  $u_t$ ?**

👉 Setting 1: access sample for  $q$  + density of  $q$  + energy of  $p$  (neural samplers)

💡 match the forward & backward process

👉 CMCD [1]

💡 match the marginal  $p_t$  of sampling process to  $U_t$

👉 NETS [2]

👉 Setting 2: access sample and energy for  $q$  and  $p$  (e.g., aim to estimate  $Z_0/Z_1$ )

[1] Vargas, F., Padhy, S., Blessing, D., & Nüsken, N. (2024). Transport meets variational inference: Controlled monte carlo diffusions. *ICLR 2024*.

[2] Albergo, M. S., & Vanden-Eijnden, E. (2025). Nets: A non-equilibrium transport sampler. *ICML 2025*.

# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

**Question: How to find  $u_t$ ?**

👉 Setting 1: access sample for  $q$  + density of  $q$  + energy of  $p$  (neural samplers)

💡 match the forward & backward process

👉 CMCD [1]

💡 match the marginal  $p_t$  of sampling process to  $U_t$

👉 NETS [2]

👉 Setting 2: access sample and energy for  $q$  and  $p$  (e.g., aim to estimate  $Z_0/Z_1$ )

💡 learn  $\nabla U_t$  and  $u_t$  at the same time (e.g., by stochastic interpolant)

[1] Vargas, F., Padhy, S., Blessing, D., & Nüsken, N. (2024). Transport meets variational inference: Controlled monte carlo diffusions. *ICLR 2024*.

[2] Albergo, M. S., & Vanden-Eijnden, E. (2025). Nets: A non-equilibrium transport sampler. *ICML 2025*.



# From Jarzynski to Escorted Jarzynski

$$X_0 \sim q \quad dX_t = [-\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overrightarrow{dW}_t,$$

$$X_1 \sim p \quad dX_t = [\sigma^2 \nabla U_t(X_t) + u_t(X_t)]dt + \sigma\sqrt{2} \overleftarrow{dW}_t,$$

**Question: How to find  $u_t$ ?**

👉 Setting 1: access sample for  $q$  + density of  $q$  + energy of  $p$  (neural samplers)

💡 match the forward & backward process

👉 CMCD [1]

💡 match the marginal  $p_t$  of sampling process to  $U_t$

👉 NETS [2]

👉 Setting 2: access sample and energy for  $q$  and  $p$  (e.g., aim to estimate  $Z_0/Z_1$ )

💡 learn  $\nabla U_t$  and  $u_t$  at the same time (e.g., by stochastic interpolant) 👉 FEAT [3]

[1] Vargas, F., Padhy, S., Blessing, D., & Nüsken, N. (2024). Transport meets variational inference: Controlled monte carlo diffusions. *ICLR 2024*.

[2] Albergo, M. S., & Vanden-Eijnden, E. (2025). Nets: A non-equilibrium transport sampler. *ICML 2025*.

[3] He, J., Du, Y., Vargas, F., Wang, Y., Gomes, C. P., Hernández-Lobato, J. M., & Vanden-Eijnden, E. (2025). FEAT: Free energy Estimators with Adaptive Transport. *arXiv*.

# Importance Sampling with Path RND

AIS  $\mathbf{E}_{X \sim \mathbf{P}}[f(X)] = \mathbf{E}_{X \sim \mathbf{Q}} \left[ \frac{d\mathbf{P}}{d\mathbf{Q}}(X) f(X) \right]$

Jarzynski  $\mathbf{E}_{\mathbf{Q}} \left[ \exp \left( \int -\partial_t U_t(X_t) dt \right) \right] = \frac{Z_1}{Z_0}$

Escorted Jarzynski  $\mathbf{E}_{\mathbf{Q}} \left[ \exp \left( \int -\partial_t U_t(X_t) dt - \nabla U_t \cdot u_t dt + \nabla \cdot u_t dt \right) \right] = \frac{Z_1}{Z_0}$

# Importance Sampling with Path RND

AIS  $\mathbf{E}_{X \sim P}[f(X)] = \mathbf{E}_{X \sim Q} \left[ \frac{dP}{dQ}(X) f(X) \right]$

Jarzynski  $\mathbf{E}_Q \left[ \exp \left( \int -\partial_t U_t(X_t) dt \right) \right] = \frac{Z_1}{Z_0}$

Escorted Jarzynski  $\mathbf{E}_Q \left[ \exp \left( \int -\partial_t U_t(X_t) dt - \nabla U_t \cdot u_t dt + \nabla \cdot u_t dt \right) \right] = \frac{Z_1}{Z_0}$

$$Z = \int \exp(-U(X)) dX$$

# Importance Sampling with Path RND

AIS  $\mathbf{E}_{X \sim P}[f(X)] = \mathbf{E}_{X \sim Q} \left[ \frac{dP}{dQ}(X) f(X) \right]$

Jarzynski  $\mathbf{E}_Q \left[ \exp \left( \int -\partial_t U_t(X_t) dt \right) \right] = \frac{Z_1}{Z_0}$

Escorted Jarzynski  $\mathbf{E}_Q \left[ \exp \left( \int -\partial_t U_t(X_t) dt - \nabla U_t \cdot u_t dt + \nabla \cdot u_t dt \right) \right] = \frac{Z_1}{Z_0}$

$$Z = \int \exp(-U(X)) dX$$

$$p(X) = \int p(Z, X) dZ \quad \text{💡 } p(X) \text{ is the normalization factor for } p(Z|X) \propto p(Z, X)$$

# Importance Sampling with Path RND

$$dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N(0, \nu)$$

? What is  $p_0(x)$  at  $t = 0$ ?

# Importance Sampling with Path RND

$$dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N(0, \nu)$$

? What is  $p_0(x)$  at  $t = 0$ ?

$$x \sim N(X|Z), p(Z) \sim N(0, \nu)$$

? What is  $p(x)$ ?

# Importance Sampling with Path RND

Marginal density relation with importance sampling:

$$p(x) = \mathbf{E}_{z \sim p(z|x)} [1] p(x)$$

# Importance Sampling with Path RND

Marginal density relation with importance sampling:

$$\begin{aligned} p(x) &= \mathbf{E}_{z \sim p(z|x)} [1] p(x) \\ &= \mathbf{E}_{z \sim p(z|x)} [p(x)] \end{aligned}$$



# Importance Sampling with Path RND

Marginal density relation with importance sampling:

$$\begin{aligned} p(x) &= \mathbf{E}_{z \sim p(z|x)} [1] p(x) \\ &= \mathbf{E}_{z \sim p(z|x)} [p(x)] \\ &= \mathbf{E}_{z \sim q(z|x)} \left[ \frac{p(x)p(z|x)}{q(z|x)} \right] \end{aligned}$$

# Importance Sampling with Path RND

Marginal density relation with importance sampling:

$$\begin{aligned} p(x) &= \mathbf{E}_{z \sim p(z|x)} [1] p(x) \\ &= \mathbf{E}_{z \sim p(z|x)} [p(x)] \\ &= \mathbf{E}_{z \sim q(z|x)} \left[ \frac{p(x)p(z|x)}{q(z|x)} \right] \\ &= \mathbf{E}_{z \sim q(z|x)} \left[ \frac{p(z)p(x|z)}{q(z|x)} \right] \end{aligned}$$

# Importance Sampling with Path RND

Marginal density relation with importance sampling:

 **Let's rewrite it in a path!**

$$\begin{aligned} p(x) &= \mathbf{E}_{z \sim p(z|x)} [1] p(x) & p_0(x) &= \\ &= \mathbf{E}_{z \sim p(z|x)} [p(x)] & &= \\ &= \mathbf{E}_{z \sim q(z|x)} \left[ \frac{p(x)p(z|x)}{q(z|x)} \right] & &= \\ &= \mathbf{E}_{z \sim q(z|x)} \left[ \frac{p(z)p(x|z)}{q(z|x)} \right] & &= \end{aligned}$$

# Importance Sampling with Path RND

Marginal density relation with importance sampling:

 **Let's rewrite it in a path!**

$$\begin{aligned} p(x) &= \mathbf{E}_{z \sim p(z|x)} [1] p(x) & p_0(x) &= \mathbf{E}_{X_{1:N} \sim p(X_{1:N}|X_0=x)} [1] p_0(x) \\ &= \mathbf{E}_{z \sim p(z|x)} [p(x)] & &= \\ &= \mathbf{E}_{z \sim q(z|x)} \left[ \frac{p(x)p(z|x)}{q(z|x)} \right] & &= \\ &= \mathbf{E}_{z \sim q(z|x)} \left[ \frac{p(z)p(x|z)}{q(z|x)} \right] & &= \end{aligned}$$

# Importance Sampling with Path RND

Marginal density relation with importance sampling:

 **Let's rewrite it in a path!**

$$\begin{aligned} p(x) &= \mathbf{E}_{z \sim p(z|x)} [1] p(x) & p_0(x) &= \mathbf{E}_{X_{1:N} \sim p(X_{1:N}|X_0=x)} [1] p_0(x) \\ &= \mathbf{E}_{z \sim p(z|x)} [p(x)] & &= \mathbf{E}_{X_{1:N} \sim p(X_{1:N}|X_0=x)} [p_0(x)] \\ &= \mathbf{E}_{z \sim q(z|x)} \left[ \frac{p(x)p(z|x)}{q(z|x)} \right] & &= \\ &= \mathbf{E}_{z \sim q(z|x)} \left[ \frac{p(z)p(x|z)}{q(z|x)} \right] & &= \end{aligned}$$

# Importance Sampling with Path RND

Marginal density relation with importance sampling:

 **Let's rewrite it in a path!**

$$\begin{aligned} p(x) &= \mathbf{E}_{z \sim p(z|x)} [1] p(x) & p_0(x) &= \mathbf{E}_{X_{1:N} \sim p(X_{1:N}|X_0=x)} [1] p_0(x) \\ &= \mathbf{E}_{z \sim p(z|x)} [p(x)] & &= \mathbf{E}_{X_{1:N} \sim p(X_{1:N}|X_0=x)} [p_0(x)] \\ &= \mathbf{E}_{z \sim q(z|x)} \left[ \frac{p(x)p(z|x)}{q(z|x)} \right] & &= \mathbf{E}_{X_{1:N} \sim q(X_{1:N}|X_0=x)} \left[ \frac{p_0(x)p(X_{1:N}|X_0=x)}{q(X_{1:N}|X_0=x)} \right] \\ &= \mathbf{E}_{z \sim q(z|x)} \left[ \frac{p(z)p(x|z)}{q(z|x)} \right] & &= \end{aligned}$$

# Importance Sampling with Path RND

Marginal density relation with importance sampling:

 **Let's rewrite it in a path!**

$$\begin{aligned} p(x) &= \mathbf{E}_{z \sim p(z|x)} [1] p(x) & p_0(x) &= \mathbf{E}_{X_{1:N} \sim p(X_{1:N}|X_0=x)} [1] p_0(x) \\ &= \mathbf{E}_{z \sim p(z|x)} [p(x)] & &= \mathbf{E}_{X_{1:N} \sim p(X_{1:N}|X_0=x)} [p_0(x)] \\ &= \mathbf{E}_{z \sim q(z|x)} \left[ \frac{p(x)p(z|x)}{q(z|x)} \right] & &= \mathbf{E}_{X_{1:N} \sim q(X_{1:N}|X_0=x)} \left[ \frac{p_0(x)p(X_{1:N}|X_0=x)}{q(X_{1:N}|X_0=x)} \right] \\ &= \mathbf{E}_{z \sim q(z|x)} \left[ \frac{p(z)p(x|z)}{q(z|x)} \right] & &= \mathbf{E}_{X_{1:N} \sim q(X_{1:N}|X_0=x)} \left[ \frac{p(X_{1:N}, X_0=x)}{q(X_{1:N}|X_0=x)} \right] \end{aligned}$$

# Importance Sampling with Path RND

$$p(x) = \mathbf{E}_{X_{1:N} \sim q(X_{1:N} | X_0 = x)} \left[ \frac{p(X_{1:N}, X_0 = x)}{q(X_{1:N} | X_0 = x)} \right]$$



# Importance Sampling with Path RND

$$p(x) = \mathbf{E}_{X_{1:N} \sim q(X_{1:N}|X_0=x)} \left[ \frac{p(X_{1:N}, X_0 = x)}{q(X_{1:N}|X_0 = x)} \right]$$

$$p: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_N \sim N$$

$$q: dX_t = h(X_t, t)dt + \sigma_t dW_t$$

# Importance Sampling with Path RND

$$p(x) = \mathbf{E}_{X_{1:N} \sim q(X_{1:N}|X_0=x)} \left[ \frac{p(X_{1:N}, X_0 = x)}{q(X_{1:N}|X_0 = x)} \right]$$

$$p: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_N \sim N$$

$$q: dX_t = h(X_t, t)dt + \sigma_t dW_t$$

$$p(x) = \mathbf{E}_{X_{1:N} \sim q(X_{1:N}|X_0=x)} \left[ \frac{\prod N_p(X_n|X_{n+1})}{\prod N_q(X_{n+1}|X_n)} N(X_N) \right]$$

# Importance Sampling with Path RND

$$p: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_N \sim N$$

$$q: dX_t = h(X_t, t)dt + \sigma_t dW_t$$

$$p(x) = \mathbf{E}_{X_{1:N} \sim q(X_{1:N} | X_0 = x)} \left[ \frac{\prod N_p(X_n | X_{n+1})}{\prod N_q(X_{n+1} | X_n)} N(X_N) \right]$$

# Importance Sampling with Path RND

$$p: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_N \sim N$$

$$q: dX_t = h(X_t, t)dt + \sigma_t dW_t$$

$$p(x) = \mathbf{E}_{X_{1:N} \sim q(X_{1:N} | X_0 = x)} \left[ \frac{\prod N_p(X_n | X_{n+1})}{\prod N_q(X_{n+1} | X_n)} N(X_N) \right]$$

## Forward-backward RND

$$\begin{aligned} \mathbf{P} : dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \mathbf{Q} : dX_t &= g(X_t, t)dt + \sigma_t dW_t, X_1 \sim q_1 \end{aligned}$$

$$\frac{d\mathbf{P}}{d\mathbf{Q}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp \left( \underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t - \frac{f_t^2(X_t)}{2\sigma_t^2} dt - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t + \frac{g_t^2(X_t)}{2\sigma_t^2} dt}_{\text{Backward Ito Integral}} \right)$$

$$\int a_t(X_t) \cdot \overleftarrow{dX}_t = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

# RECALL: Forward-backward RND

$$\begin{aligned} \mathbf{P} : dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \overleftarrow{\mathbf{Q}} : dX_t &= g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim q_1 \end{aligned}$$

$$\frac{d\mathbf{P}}{d\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp \left( \int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t - \frac{f_t^2(X_t)}{2\sigma_t^2} dt - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t + \frac{g_t^2(X_t)}{2\sigma_t^2} dt \right)$$

Initial densities

$$\lim \frac{\prod N_1(X_{n+1}|X_n)}{\prod N_2(X_n|X_{n+1})}$$

# Importance Sampling with Path RND

$$p: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_N \sim N$$

$$q: dX_t = h(X_t, t)dt + \sigma_t dW_t$$

$$p(x) = \mathbf{E}_{X_{1:N} \sim q(X_{1:N} | X_0 = x)} \left[ \frac{\prod N_g(X_n | X_{n+1})}{\prod N_h(X_{n+1} | X_n)} N(X_N) \right]$$

## Forward-backward RND

$$\begin{aligned} \mathbf{P} : dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \mathbf{Q} : dX_t &= g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim q_1 \end{aligned}$$

$$\frac{d\mathbf{P}}{d\mathbf{Q}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp \left( \underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t - \frac{f_t^2(X_t)}{2\sigma_t^2} dt - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t + \frac{g_t^2(X_t)}{2\sigma_t^2} dt}_{\text{Backward Ito Integral}} \right)$$

$$\int a_t(X_t) \cdot \overleftarrow{dX}_t = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

# Importance Sampling with Path RND

$$\bar{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N$$

$$\mathbf{Q}: dX_t = f(X_t, t)dt + \sigma_t dW_t, X_1 \sim q(X_0)$$

$$p(x) = \mathbf{E}_{\mathbf{Q}(\cdot|X_0)} \left[ N(X_1) \exp \left( - \int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t + \frac{f_t^2(X_t)}{2\sigma_t^2} dt + \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t - \frac{g_t^2(X_t)}{2\sigma_t^2} dt \right) \right]$$

## Forward-backward RND

$$\begin{aligned} \mathbf{P}: dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \bar{\mathbf{Q}}: dX_t &= g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim q_1 \end{aligned}$$

$$\frac{d\mathbf{P}}{d\bar{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp \left( \underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t - \frac{f_t^2(X_t)}{2\sigma_t^2} dt - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t + \frac{g_t^2(X_t)}{2\sigma_t^2} dt}_{\text{Backward Ito Integral}} \right)$$

$$\int a_t(X_t) \cdot \overleftarrow{dX}_t = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

# Importance Sampling with Path RND

$$\bar{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N$$

$$\mathbf{Q}: dX_t = f(X_t, t)dt + \sigma_t dW_t, X_1 \sim q(X_0)$$

$$p(x) = \mathbf{E}_{\mathbf{Q}(\cdot|X_0)} \left[ N(X_1) \exp \left( - \int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t + \frac{f_t^2(X_t)}{2\sigma_t^2} dt + \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t - \frac{g_t^2(X_t)}{2\sigma_t^2} dt \right) \right]$$

## Forward-backward RND

$$\mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0$$

$$\bar{\mathbf{Q}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim q_1$$

$$\frac{d\mathbf{P}}{d\mathbf{Q}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp \left( \underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t - \frac{f_t^2(X_t)}{2\sigma_t^2} dt - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t + \frac{g_t^2(X_t)}{2\sigma_t^2} dt}_{\text{Backward Ito Integral}} \right)$$

$$\int a_t(X_t) \cdot \overleftarrow{dX}_t = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

For simplicity, we hereafter call

$$R_f^g(X) = \exp \left( - \int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t + \frac{f_t^2(X_t)}{2\sigma_t^2} dt + \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t - \frac{g_t^2(X_t)}{2\sigma_t^2} dt \right)$$



# Importance Sampling with Path RND

$$\bar{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N$$

$$\mathbf{Q}: dX_t = f(X_t, t)dt + \sigma_t dW_t, X_1 \sim q(X_0)$$

$$p(x) = \mathbf{E}_{\mathbf{Q}(\cdot|X_0)} \left[ N(X_1) \exp \left( - \int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t + \frac{f_t^2(X_t)}{2\sigma_t^2} dt + \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t - \frac{g_t^2(X_t)}{2\sigma_t^2} dt \right) \right]$$

## Forward-backward RND

$$\mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0$$

$$\bar{\mathbf{Q}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim q_1$$

$$\frac{d\mathbf{P}}{d\mathbf{Q}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp \left( \underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t - \frac{f_t^2(X_t)}{2\sigma_t^2} dt - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t + \frac{g_t^2(X_t)}{2\sigma_t^2} dt}_{\text{Backward Ito Integral}} \right)$$

$$\int a_t(X_t) \cdot \overleftarrow{dX}_t = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

For simplicity, we hereafter call

$$R_f^g(X) = \lim \frac{\prod N_g(X_n | X_{n+1})}{\prod N_h(X_{n+1} | X_n)}$$

# Importance Sampling with Path RND

$$\tilde{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N$$

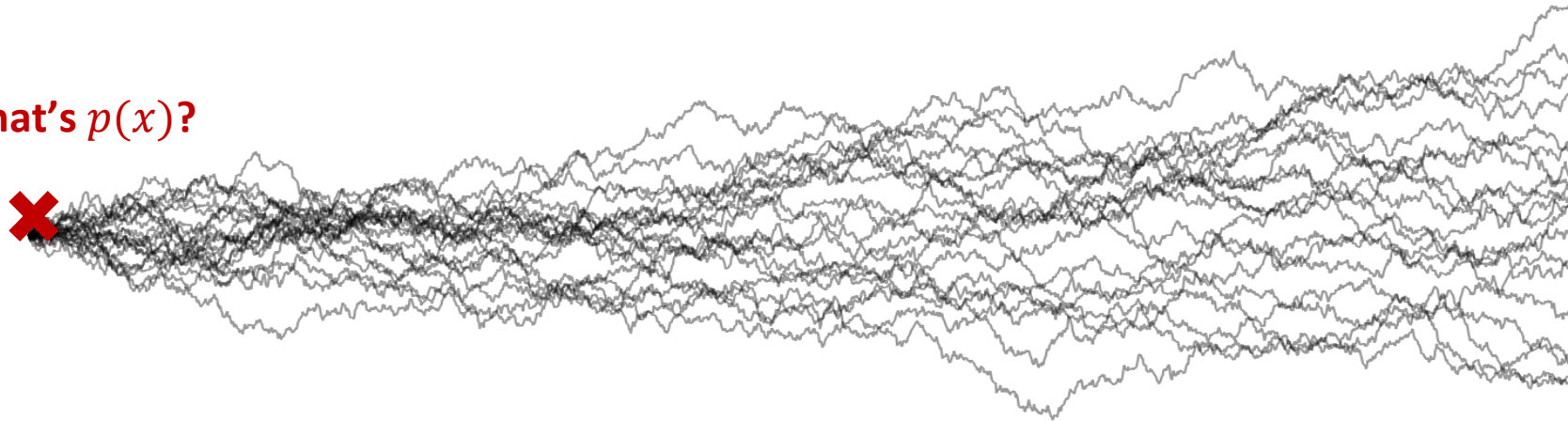
$$\mathbf{Q}: dX_t = f(X_t, t)dt + \sigma_t dW_t, X_1 \sim q(X_0)$$

$$p(x) = \mathbf{E}_{\mathbf{Q}(\cdot|X_0)} \left[ N(X_1) R_f^g(X_{[0,1]}) \right]$$

# Importance Sampling with Path RND

$$\overleftarrow{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N$$

What's  $p(x)$ ?

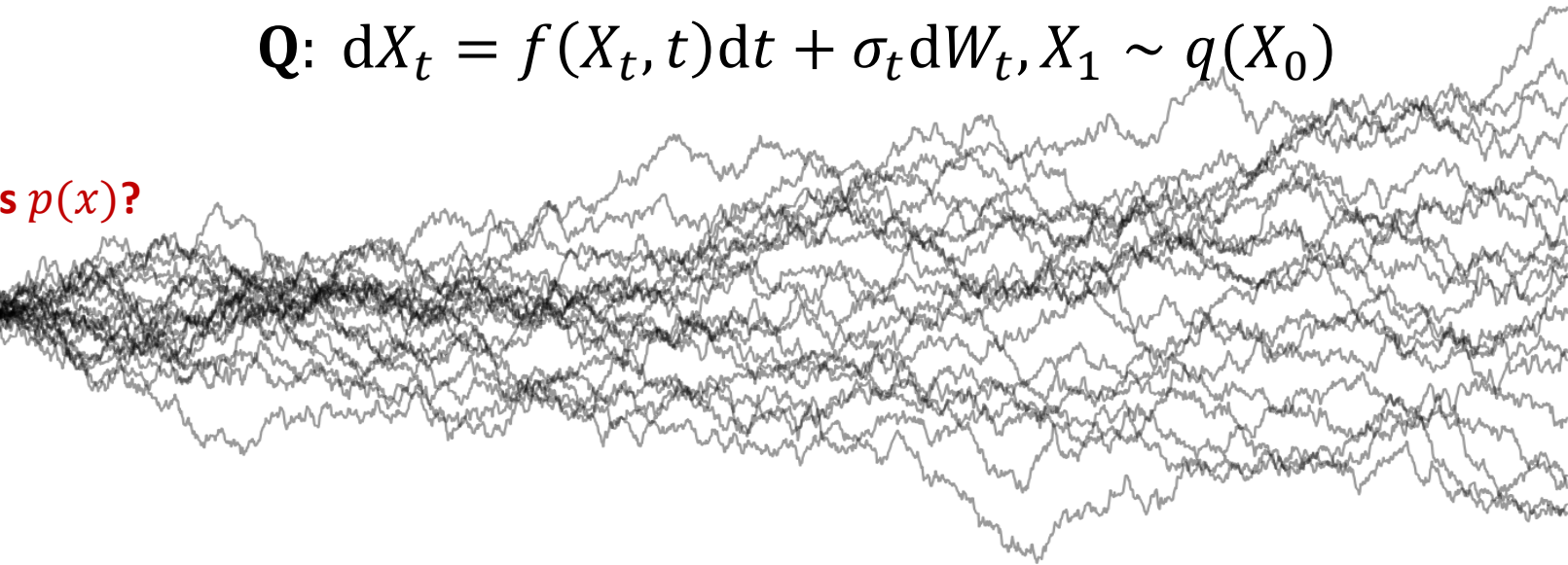


# Importance Sampling with Path RND

$$\overleftarrow{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N$$

$$\mathbf{Q}: dX_t = f(X_t, t)dt + \sigma_t dW_t, X_1 \sim q(X_0)$$

What's  $p(x)$ ?



# Importance Sampling with Path RND

$$\overleftarrow{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N$$

$$\mathbf{Q}: dX_t = f(X_t, t)dt + \sigma_t dW_t, X_1 \sim q(X_0)$$

What's  $p(x)$ ?



$$p(x) = \mathbf{E}_{\mathbf{Q}(\cdot|X_0)} \left[ N(X_1) R_f^g(X_{[0,1]}) \right]$$

# Importance Sampling with Path RND

$$\overleftarrow{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N$$

$$\mathbf{Q}: dX_t = f(X_t, t)dt + \sigma_t dW_t, X_1 \sim q(X_0)$$

What's  $p(x)$ ?



$$p(x) = \mathbf{E}_{\mathbf{Q}(\cdot|X_0)} \left[ N(X_1) R_f^g(X_{[0,1]}) \right]$$

Can also be derived from Feynman-Kac formulation as shown in [1] and [2]:

[1] Huang, C. W., Lim, J. H., & Courville, A. C. (2021). A variational perspective on diffusion-based generative models and score matching. *NeurIPS 2021*.

[2] Premkumar, A. (2024). Diffusion density estimators. *arXiv*.

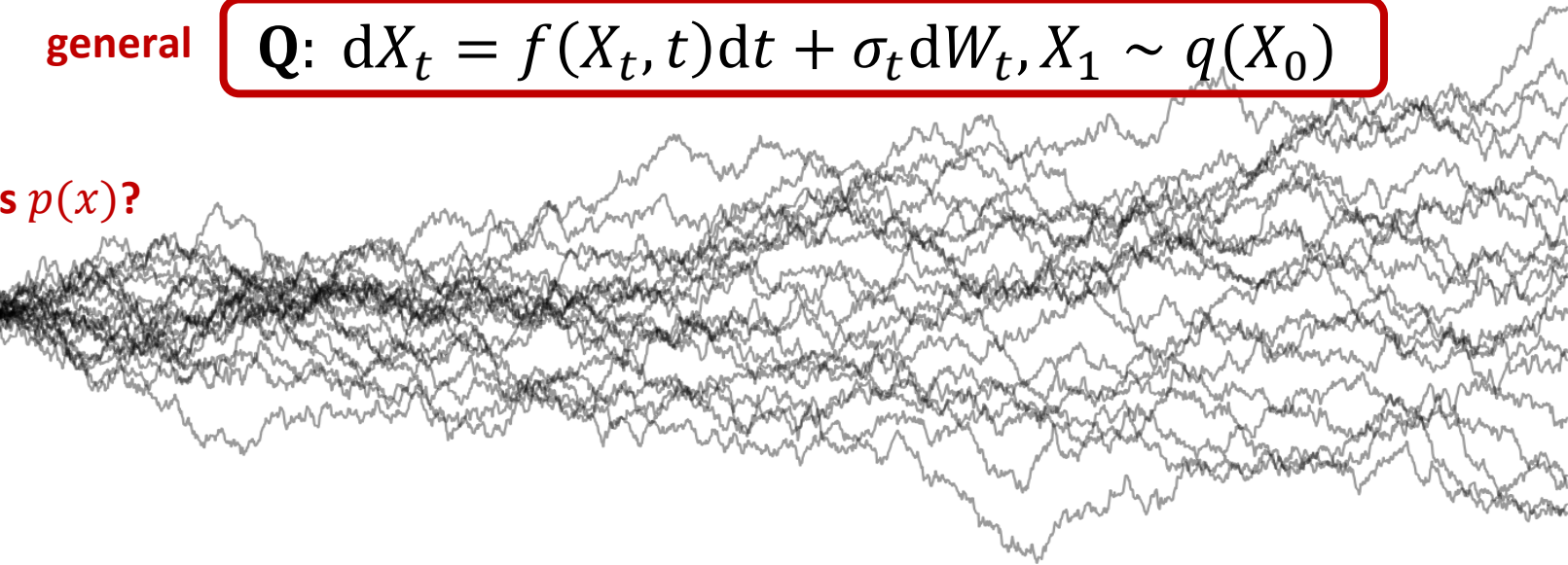
# Importance Sampling with Path RND

$$\overleftarrow{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N$$

general

$$\mathbf{Q}: dX_t = f(X_t, t)dt + \sigma_t dW_t, X_1 \sim q(X_0)$$

What's  $p(x)$ ?



# Importance Sampling with Path RND

Special Case: **time-reversal proposal**



# Importance Sampling with Path RND

Special Case: **time-reversal proposal**

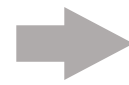
$$\overleftarrow{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N$$

$$\mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0$$

# Importance Sampling with Path RND

Special Case: **time-reversal proposal**

$$\begin{aligned} \overleftarrow{\mathbf{P}}: dX_t &= g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N \\ \mathbf{P}: dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \end{aligned}$$



$$\frac{\overleftarrow{d\mathbf{P}}}{d\mathbf{P}}(X_0 = x, X_{(0,1]}) = \mathbf{1}$$

## Forward-backward RND

$$\begin{aligned} \mathbf{P}: dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \overleftarrow{\mathbf{Q}}: dX_t &= g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim q_1 \end{aligned}$$

$$\frac{d\mathbf{P}}{d\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp \left( \underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t - \frac{f_t^2(X_t)}{2\sigma_t^2} dt - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t + \frac{g_t^2(X_t)}{2\sigma_t^2} dt}_{\text{Backward Ito Integral}} \right)$$

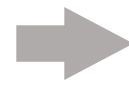
$$\int a_t(X_t) \cdot \overleftarrow{dX}_t = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

# Importance Sampling with Path RND

Special Case: **time-reversal proposal**

$$\overleftarrow{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N$$

$$\mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0$$



$$\frac{\overleftarrow{d\mathbf{P}}}{d\mathbf{P}}(X_0 = x, X_{(0,1]}) = \mathbf{1}$$

$$\frac{N(X_1)}{p_0(X_0 = x)} R_f^g(X_{[0,1]}) = 1$$

## Forward-backward RND

$$\begin{aligned} \mathbf{P}: dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \overleftarrow{\mathbf{Q}}: dX_t &= g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim q_1 \end{aligned}$$

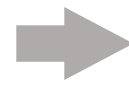
$$\frac{d\mathbf{P}}{d\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp \left( \underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t - \frac{f_t^2(X_t)}{2\sigma_t^2} dt - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t + \frac{g_t^2(X_t)}{2\sigma_t^2} dt}_{\text{Backward Ito Integral}} \right)$$

$$\int a_t(X_t) \cdot \overleftarrow{dX}_t = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

# Importance Sampling with Path RND

Special Case: **time-reversal proposal**

$$\begin{aligned} \overleftarrow{\mathbf{P}}: dX_t &= g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N \\ \mathbf{P}: dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \end{aligned}$$



$$\frac{\overleftarrow{d\mathbf{P}}}{d\mathbf{P}}(X_0 = x, X_{(0,1]}) = \mathbf{1}$$

$$N(X_1) R_f^g(X_{[0,1]}) = p_0(X_0 = x)$$

## Forward-backward RND

$$\begin{aligned} \mathbf{P}: dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \overleftarrow{\mathbf{Q}}: dX_t &= g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim q_1 \end{aligned}$$

$$\frac{d\mathbf{P}}{d\overleftarrow{\mathbf{Q}}}(X) = \underbrace{\frac{p_0(X_0)}{q_1(X_1)}}_{\text{Initial densities}} \exp \left( \underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t - \frac{f_t^2(X_t)}{2\sigma_t^2} dt - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t + \frac{g_t^2(X_t)}{2\sigma_t^2} dt}_{\text{Backward Ito Integral}} \right)$$

$$\int a_t(X_t) \cdot \overleftarrow{dX}_t = \lim \sum a_{n+1}(X_{n+1}) \cdot (X_{n+1} - X_n)$$

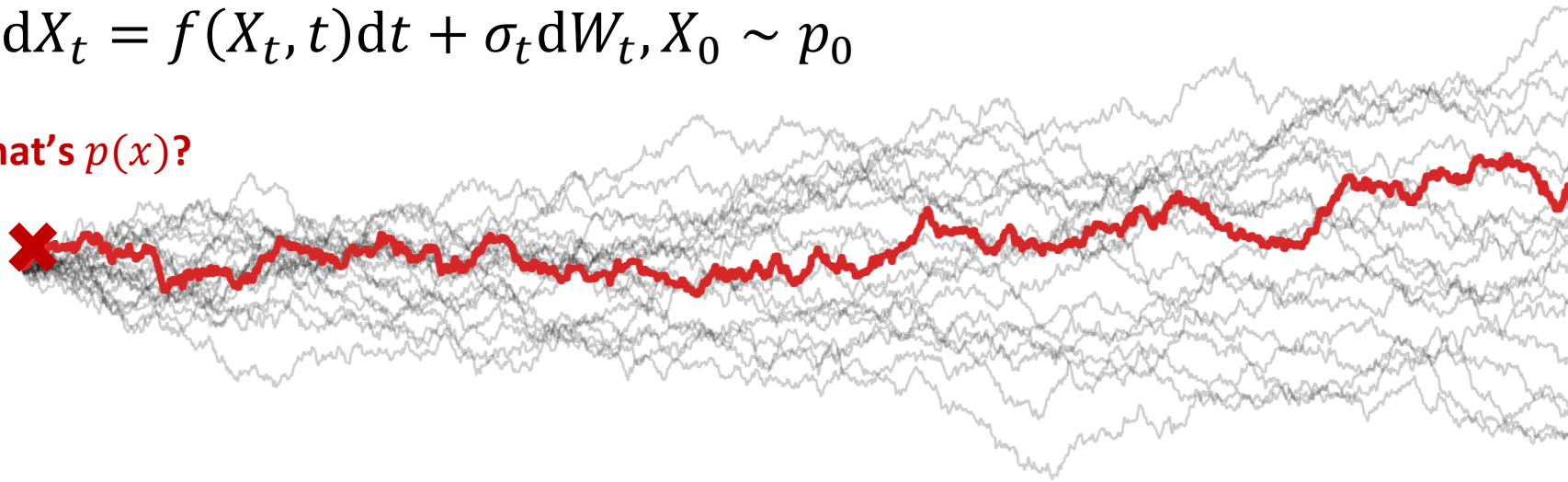
# Importance Sampling with Path RND

Special Case: **time-reversal proposal**

$$\overleftarrow{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, X_1 \sim N$$

$$\mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0$$

What's  $p(x)$ ?



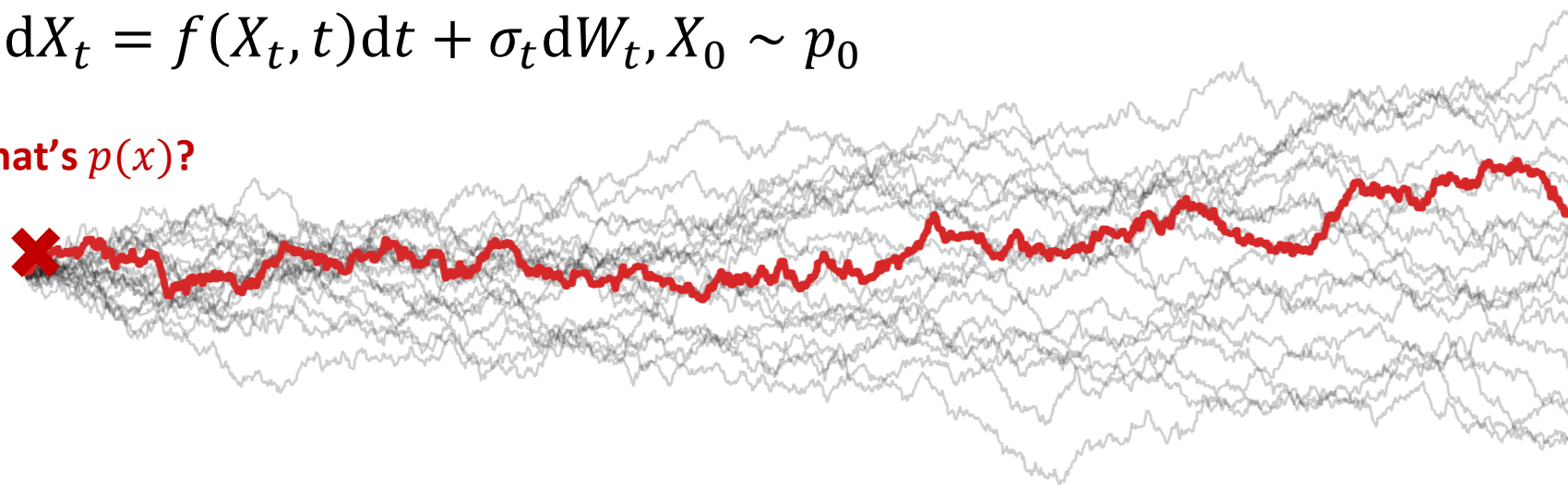
# Importance Sampling with Path RND

Special Case: **time-reversal proposal**

$$\overleftarrow{\mathbf{P}}: dX_t = g(X_t, t)dt + \overleftarrow{\sigma}_t d\overleftarrow{W}_t, X_1 \sim N$$

$$\mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0$$

What's  $p(x)$ ?



**Can be derived by PDEs as shown in [1] and [2]:**

[1] Karczewski, R., Heinonen, M., & Garg, V. (2025). Diffusion Models as Cartoonists: The Curious Case of High Density Regions. ICLR 2025.

[2] Skreta, M., Atanackovic, L., Bose, A. J., Tong, A., & Neklyudov, K. (2025). The Superposition of Diffusion Models Using the Ito Density Estimator. ICLR 2025

# Importance Sampling with Path RND

Special Case: **time-reversal proposal**

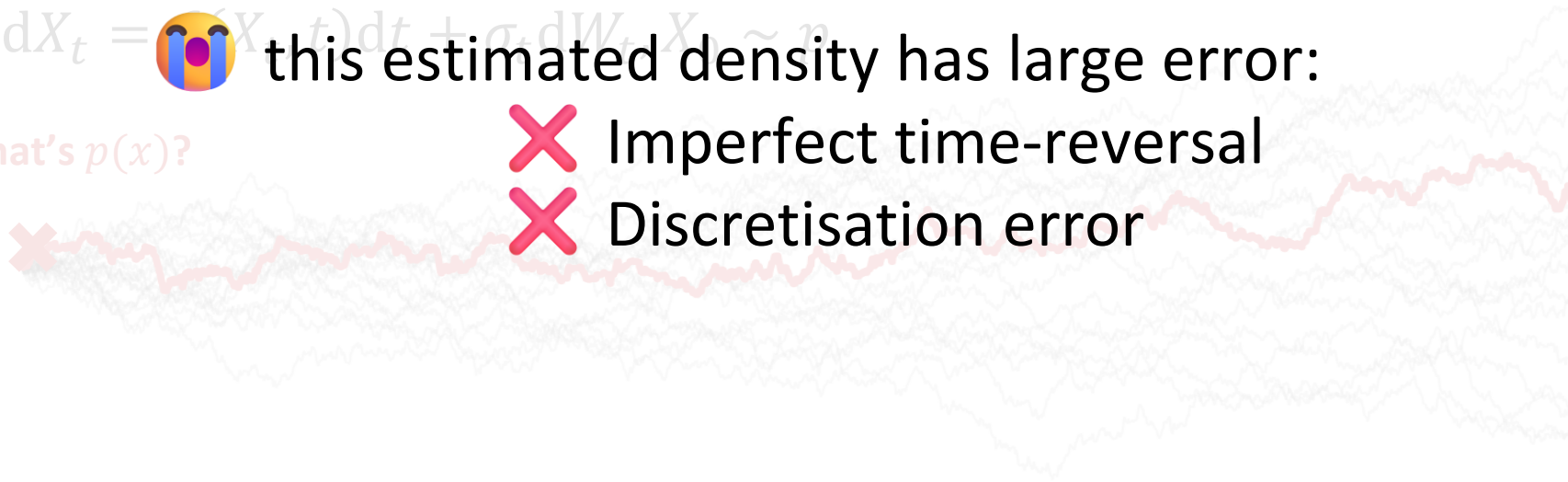
$$\overleftarrow{\mathbb{P}}: dX_t = g(X_t, t)dt + \overleftarrow{\sigma}_t d\overleftarrow{W}_t, X_1 \sim N$$

$\mathbb{P}: dX_t = \text{😭} (X_t, t)dt + \sigma_t dW_t, X_0 \sim p$  this estimated density has large error:

What's  $p(x)$ ?

✗ Imperfect time-reversal

✗ Discretisation error



**Can be derived by PDEs as shown in [1] and [2]:**

[1] Karczewski, R., Heinonen, M., & Garg, V. (2025). Diffusion Models as Cartoonists: The Curious Case of High Density Regions. ICLR 2025.

[2] Skreta, M., Atanackovic, L., Bose, A. J., Tong, A., & Neklyudov, K. (2025). The Superposition of Diffusion Models Using the Ito Density Estimator. ICLR 2025

# Importance Sampling with Path RND

Special Case: **time-reversal proposal**

$$\overleftarrow{\mathbb{P}}: dX_t = g(X_t, t)dt + \overleftarrow{\sigma}_t d\overleftarrow{W}_t, X_1 \sim N$$

$\mathbb{P}: dX_t = g(X_t, t)dt + \sigma_t dW_t, X_0 \sim p$  this estimated density has large error:

What's  $p(x)$ ?

✗ Imperfect time-reversal

✗ Discretisation error

✓ Only look at small intervals

**Can be derived by PDEs as shown in [1] and [2]:**

[1] Karczewski, R., Heinonen, M., & Garg, V. (2025). Diffusion Models as Cartoonists: The Curious Case of High Density Regions. ICLR 2025.

[2] Skreta, M., Atanackovic, L., Bose, A. J., Tong, A., & Neklyudov, K. (2025). The Superposition of Diffusion Models Using the Ito Density Estimator. ICLR 2025



# Importance Sampling with Path RND

- ✓ Only look at small intervals

$$\begin{array}{l} \bar{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t \quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau'] \\ \mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t \quad X_{\tau} \sim p_{\tau} \quad t \in [\tau, \tau'] \end{array} \quad \longrightarrow \quad \overleftarrow{\frac{d\mathbf{P}}{d\mathbf{P}}}(X_{[\tau, \tau']}) = 1$$

$$\frac{p_{\tau'}(X_{\tau'})}{p_{\tau}(X_{\tau})} R_f^g(X_{[\tau, \tau']}) = 1$$

# Importance Sampling with Path RND

✓ Only look at small intervals

$$\begin{array}{l} \overleftarrow{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t \quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau'] \\ \mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t \quad X_{\tau} \sim p_{\tau} \quad t \in [\tau, \tau'] \end{array} \quad \longrightarrow \quad \frac{\overleftarrow{d\mathbf{P}}}{d\mathbf{P}}(X_{[\tau, \tau']}) = 1$$

👉 Equal density sampling:

for two diffusion models, find sample has **same density under both models**

👉 Calculate SMC weights:

for one (or several) DMs, **steer the distribution** by tilting/annealing/composition

👉 Energy regularisation:

Train an **energy-based diffusion** model by ensuring this relation for all time intervals

# Example: Diffusion Inference-time Steering with Path RND

 Problem Setup:

Given a pretrained model for  $p_0$ , generate samples  $\sim p_0(x)\exp(r(x))$

# Example: Diffusion Inference-time Steering with Path RND

 Problem Setup:

Given a pretrained model for  $p_0$ , generate samples  $\sim p_0(x)\exp(r(x))$

 Strategy:

# Example: Diffusion Inference-time Steering with Path RND

 Problem Setup:

Given a pretrained model for  $p_0$ , generate samples  $\sim p_0(x)\exp(r(x))$

 Strategy:

- Choose a heuristic guidance process;

# Example: Diffusion Inference-time Steering with Path RND

 Problem Setup:

Given a pretrained model for  $p_0$ , generate samples  $\sim p_0(x)\exp(r(x))$

 Strategy:

- Choose a heuristic guidance process;
- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;

# Example: Diffusion Inference-time Steering with Path RND



Problem Setup:

Given a pretrained model for  $p_0$ , generate samples  $\sim p_0(x)\exp(r(x))$



Strategy:

- Choose a heuristic guidance process;
- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do importance-resampling to move samples at  $q_{t'}$  to  $q_t$  ( $t < t'$ )

# Example: Diffusion Inference-time Steering with Path RND

 Problem Setup:

Given a pretrained model for  $p_0$ , generate samples  $\sim p_0(x)\exp(r(x))$

 Strategy:

- Choose a heuristic guidance process;
- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do importance-resampling to move samples at  $q_{t'}$  to  $q_t$  ( $t < t'$ )



# Example: Diffusion Inference-time Steering with Path RND

## Problem Setup:

Given a pretrained model for  $p_0$ , generate samples  $\sim p_0(x)\exp(r(x))$

## Strategy:

- Choose a heuristic guidance process;
- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do importance-resampling to move samples at  $q_{t'}$  to  $q_t$  ( $t < t'$ )

We already learned about this pipeline from Raghav (Feynman-Kac Steering); Marta (Feynman-Kac Corrector); Luhuan (RDSMC) during the talks

## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;
- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do **importance-resampling**

## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;

$$dX_t = (\text{score} + \text{guidance}) dt + \sigma_t \overleftarrow{dW}_t, \quad X_{\tau'} \sim q_{\tau'}$$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t) \exp(r_t(x_t))$ ;
- Do **importance-resampling**

## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;

“proposal” 
$$dX_t = a(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, \quad X_{\tau'} \sim q_{\tau'}$$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do **importance-resampling**

## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;

“proposal”  $dX_t = a(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, \quad X_{\tau'} \sim q_{\tau'}$

“target”?  $X_{\tau} \sim q_{\tau}$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do **importance-resampling**

## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;

“proposal”  $dX_t = a(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, \quad X_{\tau'} \sim q_{\tau'}$

“target”?  $dX_t = b(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau} \sim q_{\tau}$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do **importance-resampling**

## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;

“proposal”  $dX_t = a(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, \quad X_{\tau'} \sim q_{\tau'}$

“target”?  $dX_t = b(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau} \sim q_{\tau}$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do **importance-resampling**

$$w \propto \frac{\text{target}}{\text{proposal}}$$



## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;

“proposal”  $dX_t = a(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, \quad X_{\tau'} \sim q_{\tau'}$

“target”?  $dX_t = b(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau} \sim q_{\tau}$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do **importance-resampling**

$$w \propto \frac{\text{target}}{q_{\tau'}(X_{\tau'}) \prod N_a(X_n | X_{n+1})}$$

## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;

“proposal”  $dX_t = a(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, \quad X_{\tau'} \sim q_{\tau'}$

“target”?  $dX_t = b(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau} \sim q_{\tau}$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do **importance-resampling**

$$w \propto \frac{q_{\tau}(X_{\tau}) \prod N_b(X_{n+1}|X_n)}{q_{\tau'}(X_{\tau'}) \prod N_a(X_n|X_{n+1})}$$

## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;

“proposal”  $dX_t = a(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, \quad X_{\tau'} \sim q_{\tau'}$

“target”?  $dX_t = b(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau} \sim q_{\tau}$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do **importance-resampling**

$$w \propto \frac{q_{\tau}(X_{\tau}) \prod N_b(X_{n+1}|X_n)}{q_{\tau'}(X_{\tau'}) \prod N_a(X_n|X_{n+1})}$$

## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;

“proposal”  $dX_t = a(X_t, t)dt + \sigma_t \overleftarrow{dW}_t,$

“target”?  $dX_t = b(X_t, t)dt + \sigma_t dW_t$

- Define a sequence of intermediate target
- Do **importance-resampling**

$$R_f^g(X) = \lim \frac{\prod N_g(X_n | X_{n+1})}{\prod N_h(X_{n+1} | X_n)} \exp(r_t(x_t));$$

$$w \propto \frac{q_{\tau}(X_{\tau}) \prod N_b(X_{n+1} | X_n)}{q_{\tau'}(X_{\tau'}) \prod N_a(X_n | X_{n+1})}$$

## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;

“proposal”  $dX_t = a(X_t, t)dt + \sigma_t \overleftarrow{dW}_t,$

“target”?  $dX_t = b(X_t, t)dt + \sigma_t dW_t$

- Define a sequence of intermediate target  $(x_t) \exp(r_t(x_t));$
- Do **importance-resampling**

$1/R_b^a(X_{[\tau, \tau']})$

$$w \propto \frac{q_{\tau}(X_{\tau}) \prod N_b(X_{n+1} | X_n)}{q_{\tau'}(X_{\tau'}) \prod N_a(X_n | X_{n+1})}$$

## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;

“proposal”  $dX_t = a(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, \quad X_{\tau'} \sim q_{\tau'}$

“target”?  $dX_t = b(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau} \sim q_{\tau}$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do **importance-resampling**

$$w \propto \frac{q_{\tau}(X_{\tau})}{q_{\tau'}(X_{\tau'})} \mathbf{1}/R_b^a(X_{[\tau, \tau']})$$

## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;

“proposal”  $dX_t = a(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, \quad X_{\tau'} \sim q_{\tau'}$

“target”?  $dX_t = b(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau} \sim q_{\tau}$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do **importance-resampling**

$$w \propto \frac{p_{\tau}(x_{\tau})\exp(r_{\tau}(x_{\tau}))}{p_{\tau'}(x_{\tau'})\exp(r_{\tau'}(x_{\tau'}))} \mathbf{1}/R_b^a(X_{[\tau,\tau']})$$

## Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

- Choose a heuristic guidance process;

“proposal”  $dX_t = a(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, \quad X_{\tau'} \sim q_{\tau'}$

“target”?  $dX_t = b(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau} \sim q_{\tau}$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do **importance-resampling**

$$w \propto \frac{p_{\tau}(x_{\tau})\exp(r_{\tau}(x_{\tau}))}{p_{\tau'}(x_{\tau'})\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau,\tau']})$$

?



# Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

$\overleftarrow{\mathbf{P}}$ :  $dX_t = \text{diffusion denoising } dt + \sigma_t dW_t$   $X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$

- Choose a heuristic guidance process;

“proposal”  $dX_t = a(X_t, t)dt + \sigma_t dW_t$ ,  $X_{\tau'} \sim q_{\tau'}$

“target”?  $dX_t = b(X_t, t) \frac{p_{\tau}(X_{\tau})}{p_{\tau'}(X_{\tau'})} dt + \sigma_t dW_t$ ,  $X_{\tau} \sim q_{\tau}$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t) \exp(r_t(x_t))$ ;
- Do importance-resampling

$$w \propto \frac{p_{\tau}(x_{\tau}) \exp(r_{\tau}(x_{\tau}))}{p_{\tau'}(x_{\tau'}) \exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau, \tau']})$$

?

# Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

$\bar{\mathbf{P}}$ :  $dX_t = \text{diffusion denoising } dt + \sigma_t dW_t$   $X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$

- Choose a heuristic guidance process;

$\mathbf{P}$ :  $dX_t = \text{diffusion noising } dt + \sigma_t dW_t$   $X_{\tau} \sim p_{\tau} \quad t \in [\tau, \tau']$

“proposal”

$dX_t = a(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau'} \sim q_{\tau'}$

“target”?

$dX_t = b(X_t, t) \frac{p_{\tau}(X_{\tau})}{p_{\tau'}(X_{\tau'})} dt + \sigma_t dW_t, \quad X_{\tau} \sim q_{\tau}$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t) \exp(r_t(x_t))$ ;
- Do importance-resampling

$$w \propto \frac{p_{\tau}(x_{\tau}) \exp(r_{\tau}(x_{\tau}))}{p_{\tau'}(x_{\tau'}) \exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau, \tau']})$$

?

# Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

$$\overleftarrow{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t dW_t \quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$$

- Choose a heuristic guidance process;

$$\mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t \quad X_{\tau} \sim p_{\tau} \quad t \in [\tau, \tau']$$

“proposal”

$$dX_t = a(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau'} \sim q_{\tau'}$$

“target”?

$$dX_t = b(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau} \sim q_{\tau}$$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do importance-resampling

$$w \propto \frac{p_{\tau}(x_{\tau})\exp(r_{\tau}(x_{\tau}))}{p_{\tau'}(x_{\tau'})\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau, \tau']})$$

?

# Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

$$\begin{aligned} \overleftarrow{\mathbf{P}}: dX_t &= g(X_t, t)dt + \sigma_t dW_t & X_{\tau'} &\sim p_{\tau'} & t \in [\tau, \tau'] \\ \mathbf{P}: dX_t &= f(X_t, t)dt + \sigma_t dW_t & X_{\tau} &\sim p_{\tau} & t \in [\tau, \tau'] \end{aligned} \quad \Rightarrow \quad \frac{d\mathbf{P}}{d\overleftarrow{\mathbf{P}}}(X_{[\tau, \tau']}) = 1$$

- Choose a heuristic guidance process;

“proposal”

$$dX_t = a(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau'} \sim q_{\tau'}$$

“target”?

$$dX_t = b(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau} \sim q_{\tau}$$

$$\frac{p_{\tau}(X_{\tau})}{p_{\tau'}(X_{\tau'})} = R_f^g(X_{[\tau, \tau']})$$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do importance-resampling

$$w \propto \frac{p_{\tau}(x_{\tau})\exp(r_{\tau}(x_{\tau}))}{p_{\tau'}(x_{\tau'})\exp(r_{\tau'}(x_{\tau'}))} \frac{1}{R_b^a(X_{[\tau, \tau']})}$$

?

# Example: Diffusion Inference-time Steering with Path RND

have  $\{x\} \sim q_{\tau'}$ , how to obtain exact sample  $\{x\} \sim q_{\tau}$

$$\begin{array}{l} \overleftarrow{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t dW_t \quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau'] \\ \mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t \quad X_{\tau} \sim p_{\tau} \quad t \in [\tau, \tau'] \end{array} \Rightarrow \frac{d\overleftarrow{\mathbf{P}}}{d\mathbf{P}}(X_{[\tau, \tau']}) = 1$$

- Choose a heuristic guidance process;

“proposal”

$$dX_t = a(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau'} \sim q_{\tau'}$$

“target”?

$$dX_t = b(X_t, t)dt + \sigma_t dW_t, \quad X_{\tau} \sim q_{\tau}$$

$$\frac{p_{\tau}(X_{\tau})}{p_{\tau'}(X_{\tau'})} = R_f^g(X_{[\tau, \tau']})$$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do importance-resampling

$$w \propto \boxed{R_f^g(X_{[\tau, \tau']})} \frac{\exp(r_{\tau}(x_{\tau}))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau, \tau']})$$

## Example: Diffusion Inference-time Steering with Path RND

- Choose a heuristic guidance process;

“proposal”  $dX_t = a(X_t, t)dt + \sigma_t \overleftarrow{dW}_t, \quad X_{\tau'} \sim q_{\tau'}$

“target”  $dX_t = b(X_t, t)dt + \sigma_t dW_t, \quad X_\tau \sim q_\tau$

- Define a sequence of intermediate target densities  $q_t \propto p_t(x_t)\exp(r_t(x_t))$ ;
- Do **importance-resampling**

$$w \propto R_f^g(X_{[\tau, \tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau, \tau']})$$

## Example: Diffusion Inference-time Steering with Path RND

$$w \propto R_f^g(X_{[\tau, \tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau, \tau']})$$

### Summary:

- 👉 Define proposal and target process
- 👉 Define intermediate densities  $q_t$  (by steering diffusion's  $p_t$ )
- 👉 Replace ratio between  $p_t$  by forward-backward kernel ratio  $R$

## Example: Diffusion Inference-time Steering with Path RND

$$w \propto R_f^g(X_{[\tau, \tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau, \tau']})$$

### Summary:

- 👉 Define proposal and target process
- 👉 Define intermediate densities  $q_t$  (by steering diffusion's  $p_t$ )
- 👉 Replace ratio between  $p_t$  by forward-backward kernel ratio  $R$



## Example: Diffusion Inference-time Steering with Path RND

$$w \propto R_f^g(X_{[\tau, \tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau, \tau']})$$

### Summary:

- 👉 Define proposal and target process
- 👉 Define intermediate densities  $q_t$  (by steering diffusion's  $p_t$ )
- 👉 Replace ratio between  $p_t$  by forward-backward kernel ratio  $R$

## Example: Diffusion Inference-time Steering with Path RND

$$w \propto R_f^g(X_{[\tau, \tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau, \tau']})$$

### Summary:

- 👉 Define proposal and target process
- 👉 Define intermediate densities  $q_t$  (by steering diffusion's  $p_t$ )
- 👉 Replace ratio between  $p_t$  by forward-backward kernel ratio  $R$

## Example: Diffusion Inference-time Steering with Path RND

$$w \propto R_f^g(X_{[\tau, \tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau, \tau']})$$

🌟 Anneal target  $p_t^\beta$

🌟 Composition/CFG between 2 diffusions  $\left(p_t^{(1)}\right)^\beta \left(p_t^{(2)}\right)^\alpha$

## Example: Diffusion Inference-time Steering with Path RND

$$w \propto R_f^g(X_{[\tau, \tau']}) \frac{\exp(r_\tau(x_\tau))}{\exp(r_{\tau'}(x_{\tau'}))} 1/R_b^a(X_{[\tau, \tau']})$$

🌟 Anneal target  $p_t^\beta$

$$w \propto \left[ R_f^g(X_{[\tau, \tau']}) \right]^\beta 1/R_b^a(X_{[\tau, \tau']})$$

🌟 Composition/CFG between 2 diffusions  $\left( p_t^{(1)} \right)^\beta \left( p_t^{(2)} \right)^\alpha$

$$w \propto \left[ R_{f_1}^{g_1}(X_{[\tau, \tau']}) \right]^\beta \left[ R_{f_2}^{g_2}(X_{[\tau, \tau']}) \right]^\alpha 1/R_b^a(X_{[\tau, \tau']})$$

## Example: Diffusion Inference-time Steering with Path RND

$$R_f^g(X) = \exp\left(-\int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t + \frac{f_t^2(X_t)}{2\sigma_t^2} dt + \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t - \frac{g_t^2(X_t)}{2\sigma_t^2} dt\right)$$

## Example: Diffusion Inference-time Steering with Path RND

$$R_f^g(X) = \exp\left(-\int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t + \frac{f_t^2(X_t)}{2\sigma_t^2} dt + \int \frac{g_t(X_t)}{\sigma_t^2} \cdot \overleftarrow{dX}_t - \frac{g_t^2(X_t)}{2\sigma_t^2} dt\right)$$

If we (1) plug-in this  $R$ ,  
and (2) choose target process  $f$  to be simple noising process:



**We recover a path RND perspective for FKC [1]:**

[1] Skreta, M., Akhound-Sadegh, T., Ohanesian, V., Bondesan, R., Aspuru-Guzik, A., Doucet, A., ... & Neklyudov, K. (2025). Feynman-kac correctors in diffusion: Annealing, guidance, and product of experts. *ICML 2025*.

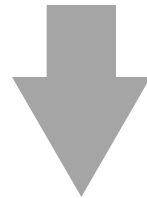
## Example: Diffusion Inference-time Steering with Path RND

$$R_f^g(X) = \lim \frac{\prod N_g(X_n | X_{n+1})}{\prod N_h(X_{n+1} | X_n)}$$

## Example: Diffusion Inference-time Steering with Path RND

$$R_f^g(X) = \lim \frac{\prod N_g(X_n | X_{n+1})}{\prod N_h(X_{n+1} | X_n)}$$

If we (1) plug-in this  $R$ ,  
and (2) choose target process  $f$  to be simple noising process,  
and (3) focus on reward-tilting



**We recover a path RND perspective for TDS [1] / FKS[2]:**

[1] Wu, L., Trippe, B., Naesseth, C., Blei, D., & Cunningham, J. P. (2023). Practical and asymptotically exact conditional sampling in diffusion models. *NeurIPS 2023*.

[2] Singhal, R., Horvitz, Z., Teehan, R., Ren, M., Yu, Z., McKeown, K., & Ranganath, R. (2025). A general framework for inference-time scaling and steering of diffusion models. *ICML 2025*.



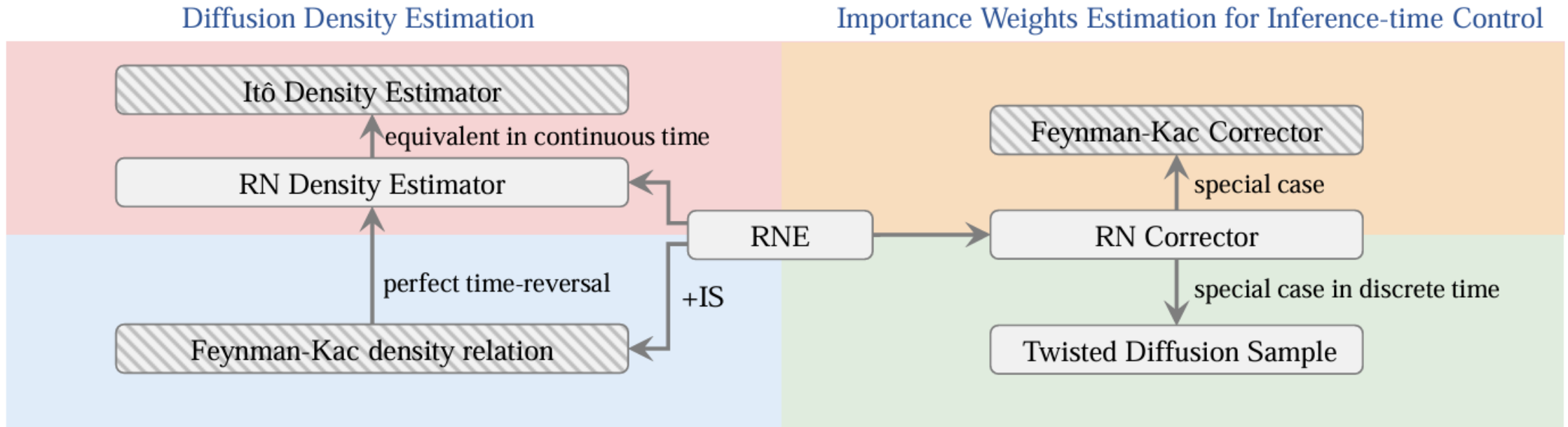
# Example: Diffusion Inference-time Steering with Path RND

## Inference-time control with path RND (RN Estimator) [1]:

- 👉 flexible! (any target process, any proposal process)
- 👉 plug-and-play! (as long as there is some  $p_t/p_{t'}$ )
- 👉 Unifying! (Ito density estimator/FKC/TDS/FKS...)

[1] He, J., Hernández-Lobato, J. M., Du, Y., & Vargas, F. (2025). RNE: a plug-and-play framework for diffusion density estimation and inference-time control. *arXiv*.

# Example: Diffusion Inference-time Steering with Path RND



# Example: Diffusion Inference-time Steering with Path RND

Inference-time Annealing: (Alanine Dipeptide 800K to 300K)

Metric	Energy TV( $\downarrow$ )	Distance TV( $\downarrow$ )	Sample $W_2$ ( $\downarrow$ )
Anneal score (wo SMC)	0.794	0.023	<b>0.173*</b>
FKC	0.338	0.022	0.289
In theory = FKC 👉 RNC ( $c_a = 1, c_b = 0$ )	0.386	0.017	0.282
Flexible choices 👉 RNC ( $c_a = 0.6, c_b = 0.4$ )	<b>0.034</b>	<b>0.011</b>	0.253

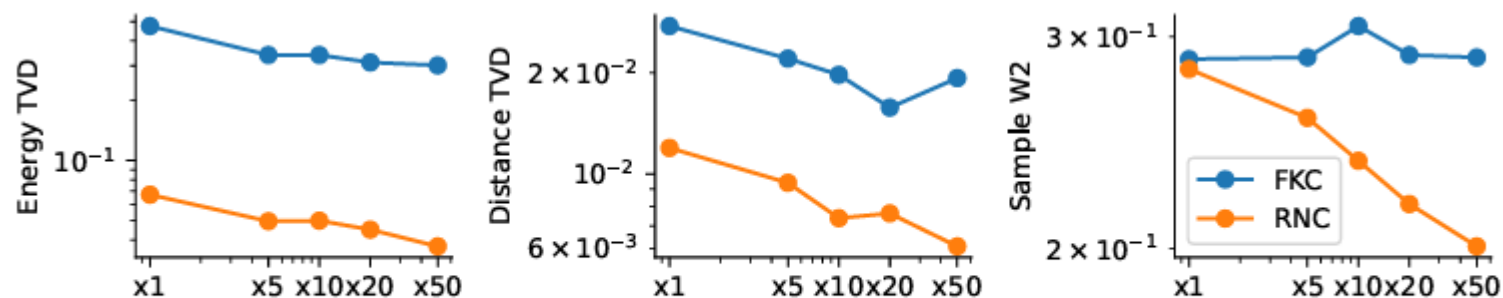


Figure 6: Performance scaling with particle numbers in SMC.

# Takeaways So Far...

Path RND connects **transition kernels** with **marginal densities**  
(known) (unknown)

- 👉 Density estimation
- 👉 SMC corrector

# Takeaways So Far...

Path RND connects **transition kernels** with **marginal densities**  
(known) (unknown)

- 👉 Density estimation
- 👉 SMC corrector
- ❓ What else? Where do we also need density?

# Example: Energy-parameterised Diffusion Training

 Problem Setup:

train a diffusion model, parameterised by energy instead of score

# Example: Energy-parameterised Diffusion Training



Problem Setup:

train a diffusion model, parameterised by energy instead of score


$$\begin{array}{l} \tilde{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t \quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau'] \\ \mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t \quad X_{\tau} \sim p_{\tau} \quad t \in [\tau, \tau'] \end{array} \quad \rightarrow \quad \frac{\overleftarrow{d\mathbf{P}}}{d\mathbf{P}}(X_{[\tau, \tau']}) = 1$$

$$\frac{p_{\tau'}(X_{\tau'})}{p_{\tau}(X_{\tau})} R_f^g(X_{[\tau, \tau']}) = 1$$

# Example: Energy-parameterised Diffusion Training

 Problem Setup:

train a diffusion model, parameterised by energy instead of score

Learned network   $\tilde{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t \quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$   $\rightarrow \frac{d\tilde{\mathbf{P}}}{d\mathbf{P}}(X_{[\tau, \tau']}) = 1$

$\mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t \quad X_\tau \sim p_\tau \quad t \in [\tau, \tau']$


$$\frac{p_{\tau'}(X_{\tau'})}{p_\tau(X_\tau)} R_f^g(X_{[\tau, \tau']}) = 1$$




# Example: Energy-parameterised Diffusion Training

 Problem Setup:

train a diffusion model, parameterised by energy instead of score

Learned network   $\tilde{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t \quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$

Noising process   $\mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t \quad X_{\tau} \sim p_{\tau} \quad t \in [\tau, \tau']$


$\Rightarrow \frac{d\mathbf{P}}{d\tilde{\mathbf{P}}}(X_{[\tau, \tau']}) = 1$


$$\frac{p_{\tau'}(X_{\tau'})}{p_{\tau}(X_{\tau})} R_f^g(X_{[\tau, \tau']}) = 1$$

# Example: Energy-parameterised Diffusion Training


 Problem Setup:


train a diffusion model, parameterised by energy instead of score

Learned network   $\bar{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t \quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$

Noising process   $\mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t \quad X_\tau \sim p_\tau \quad t \in [\tau, \tau']$

$\Rightarrow \frac{d\bar{\mathbf{P}}}{d\mathbf{P}}(X_{[\tau, \tau']}) = 1$


Learned network   $\frac{p_{\tau'}(X_{\tau'})}{p_\tau(X_\tau)} R_f^g(X_{[\tau, \tau']}) = 1$


Learned network   $\frac{p_{\tau'}(X_{\tau'})}{p_\tau(X_\tau)} R_f^g(X_{[\tau, \tau']}) = 1$

# Example: Energy-parameterised Diffusion Training


 Problem Setup:


train a diffusion model, parameterised by energy instead of score

Learned network   $\bar{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t \quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$

Noising process   $\mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t \quad X_\tau \sim p_\tau \quad t \in [\tau, \tau']$

$\Rightarrow \frac{\overleftarrow{d}\mathbf{P}}{d\mathbf{P}}(X_{[\tau, \tau']}) = 1$

Learned network 


Learned network 


$$\left| \left| \frac{p_{\tau'}(X_{\tau'})}{p_\tau(X_\tau)} R_f^g(X_{[\tau, \tau']}) - 1 \right| \right|^2$$

# Example: Energy-parameterised Diffusion Training


 Problem Setup:


train a diffusion model, parameterised by energy instead of score

Learned network   $\bar{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t \quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$

Noising process   $\mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t \quad X_\tau \sim p_\tau \quad t \in [\tau, \tau']$

$\Rightarrow \frac{\overleftarrow{d}\mathbf{P}}{d\mathbf{P}}(X_{[\tau, \tau']}) = 1$

Learned network 

Learned network 


$$\left| \left| \frac{p_{\tau'}(X_{\tau'})}{p_\tau(X_\tau)} R_f^g(X_{[\tau, \tau']}) - 1 \right| \right|^2$$


$\tau' = \tau + \Delta t$

# Example: Energy-parameterised Diffusion Training


 Problem Setup:


train a diffusion model, parameterised by energy instead of score

Learned network   $\bar{\mathbf{P}}: dX_t = g(X_t, t)dt + \sigma_t \overleftarrow{dW}_t \quad X_{\tau'} \sim p_{\tau'} \quad t \in [\tau, \tau']$

Noising process   $\mathbf{P}: dX_t = f(X_t, t)dt + \sigma_t dW_t \quad X_\tau \sim p_\tau \quad t \in [\tau, \tau']$

$\Rightarrow \frac{\overleftarrow{d}\mathbf{P}}{d\mathbf{P}}(X_{[\tau, \tau']}) = 1$

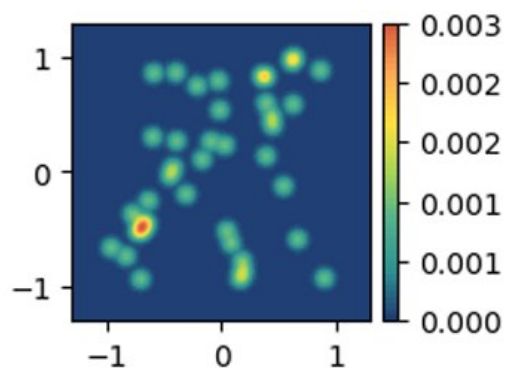
Learned network 

Learned network 

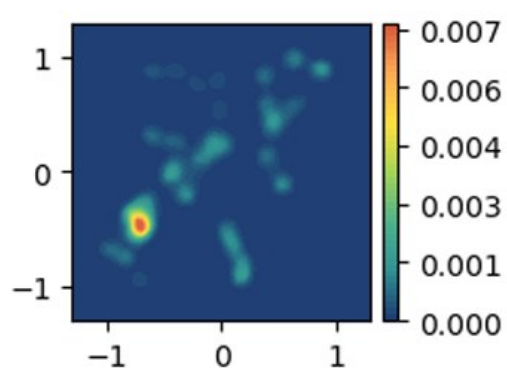
$$\left| \left| \frac{p_{\tau'}(X_{\tau'})}{p_\tau(X_\tau)} R_f^g(X_{[\tau, \tau']}) - 1 \right| \right|^2 + \text{DSM}$$

$\tau' = \tau + \Delta t$

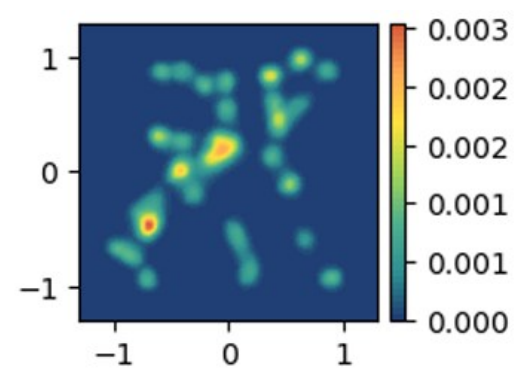
# Example: Energy-parameterised Diffusion Training



(a) Ground truth



(b) DSM

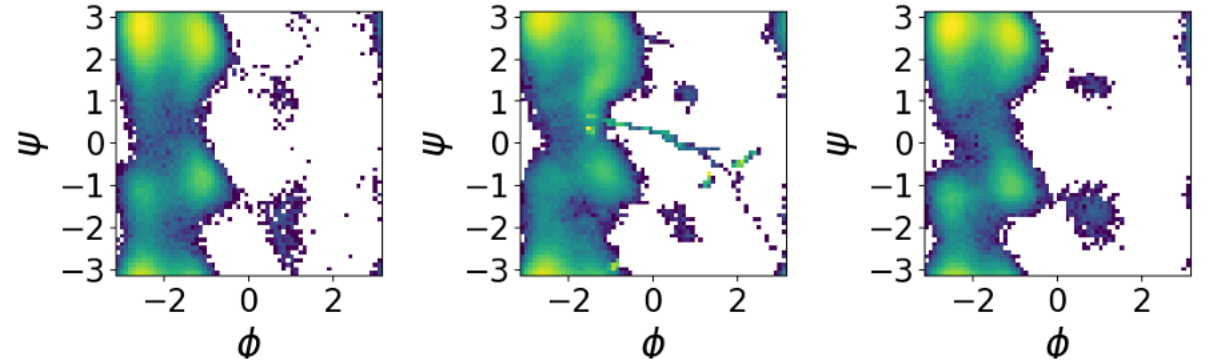


(c) RNE regularise

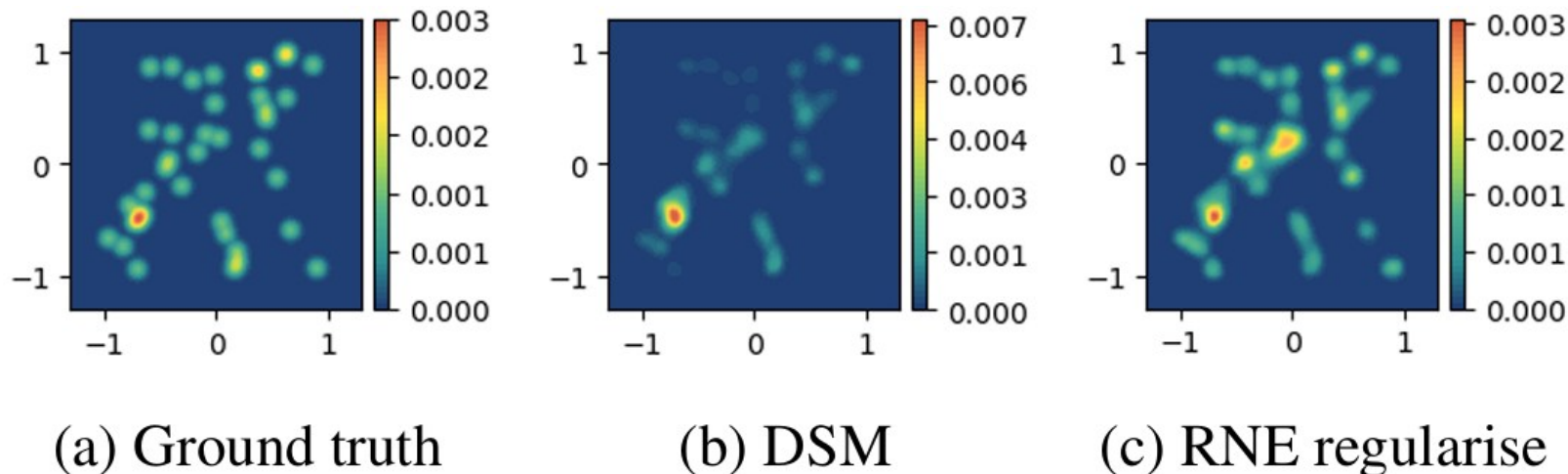
# Example: Energy-parameterised Diffusion Training

Table 4: Quality of samples obtained by running denoising process (denoted as DM) and running MCMC on learned energy at  $t = 0$ .

Training method	Sample Method	Sample $W_2$
DSM	DM	0.1811
	MCMC	0.9472
RNE Reg	DM	0.1809
	MCMC	0.1836



(a) Ground truth (b) DSM (c) RNE regularise  
Figure 8: Ramachandran plot of samples by MCMC (with Metropolis–Hastings) on learned energy.



(a) Ground truth

(b) DSM

(c) RNE regularise

# Importance Sampling with Path RND

**Importance Sampling:**

$$\mathbf{E}_{x \sim p}[f(x)] = \mathbf{E}_{x \sim q} \left[ \frac{p(x)}{q(x)} f(x) \right]$$

**Importance Sampling in Path space:**

$$\mathbf{E}_{X \sim \mathbf{P}}[f(X)] = \mathbf{E}_{X \sim \mathbf{Q}} \left[ \frac{d\mathbf{P}}{d\mathbf{Q}}(X) f(X) \right]$$



# Importance Sampling with Path RND

**Importance Sampling:**

$$\mathbf{E}_{x \sim p}[f(x)] = \mathbf{E}_{x \sim q} \left[ \frac{p(x)}{q(x)} f(x) \right]$$

? What is the optimal proposal for  $f > 0$ ?

**Importance Sampling in Path space:**

$$\mathbf{E}_{X \sim \mathbf{P}}[f(X)] = \mathbf{E}_{X \sim \mathbf{Q}} \left[ \frac{d\mathbf{P}}{d\mathbf{Q}}(X) f(X) \right]$$

# Importance Sampling with Path RND

**Importance Sampling:**

$$\mathbf{E}_{x \sim p}[f(x)] = \mathbf{E}_{x \sim q} \left[ \frac{p(x)}{q(x)} f(x) \right]$$

? What is the optimal proposal for  $f > 0$ ?

$$q \propto pf$$

**Importance Sampling in Path space:**

$$\mathbf{E}_{X \sim \mathbf{P}}[f(X)] = \mathbf{E}_{X \sim \mathbf{Q}} \left[ \frac{d\mathbf{P}}{d\mathbf{Q}}(X) f(X) \right]$$

$$\frac{d\mathbf{Q}}{d\mathbf{P}} \propto f$$

# Importance Sampling with Path RND

Likelihood

**Importance Sampling:**

? What is the optimal proposal for  $L > 0$ ?

$$\mathbf{E}_{x \sim p}[f(x)] = \mathbf{E}_{x \sim q} \left[ \frac{p(x)}{q(x)} f(x) \right]$$

$$q \propto pL$$

**Importance Sampling in Path space:**

$$\mathbf{E}_{X \sim \mathbf{P}}[f(X)] = \mathbf{E}_{X \sim \mathbf{Q}} \left[ \frac{d\mathbf{P}}{d\mathbf{Q}}(X) f(X) \right]$$

$$\frac{d\mathbf{Q}}{d\mathbf{P}} \propto L$$

# Variational Inference with Path RND

**Importance Sampling:**

$$\mathbf{E}_{x \sim p}[f(x)] = \mathbf{E}_{x \sim q} \left[ \frac{p(x)}{q(x)} f(x) \right]$$

**Importance Sampling in Path space:**

$$\mathbf{E}_{X \sim \mathbf{P}}[f(X)] = \mathbf{E}_{X \sim \mathbf{Q}} \left[ \frac{d\mathbf{P}}{d\mathbf{Q}}(X) f(X) \right]$$

**Variational Inference:**

$$q^* \propto pL$$
$$q = \min D[q, q^*]$$

**Variational Inference in Path space:**

$$\frac{d\mathbf{Q}^*}{d\mathbf{P}} \propto L$$
$$\mathbf{Q} = \min D[\mathbf{Q}, \mathbf{Q}^*]$$

# Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{dQ^*}{dP} \propto L$$

$$Q = \min D[Q, Q^*]$$

# Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{d\mathbf{Q}^*}{d\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min D[\mathbf{Q}, \mathbf{Q}^*]$$

$$\mathbf{P}: dX_t = b(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

$$\mathbf{Q}: dX_t = (b + \sigma_t u)(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

# Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{d\mathbf{Q}^*}{d\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \text{KL}[\mathbf{Q}||\mathbf{Q}^*]$$

$$\mathbf{P}: dX_t = b(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

$$\mathbf{Q}: dX_t = (b + \sigma_t u)(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

# Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{d\mathbf{Q}^*}{d\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathbf{E}_{\mathbf{Q}} \left[ \log \frac{d\mathbf{Q}}{d\mathbf{Q}^*} \right]$$

$$\mathbf{P}: dX_t = b(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

$$\mathbf{Q}: dX_t = (b + \sigma_t u)(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$



# Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{d\mathbf{Q}^*}{d\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathbf{E}_{\mathbf{Q}} \left[ \log \frac{d\mathbf{Q}}{d\mathbf{Q}^*} \right] = \min \mathbf{E}_{\mathbf{Q}} \left[ \log \frac{d\mathbf{Q}}{d\mathbf{P}} + \log \frac{d\mathbf{P}}{d\mathbf{Q}^*} \right]$$

$$\mathbf{P}: dX_t = b(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

$$\mathbf{Q}: dX_t = (b + \sigma_t u)(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

# Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{d\mathbf{Q}^*}{d\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathbf{E}_{\mathbf{Q}} \left[ \log \frac{d\mathbf{Q}}{d\mathbf{Q}^*} \right] = \min \mathbf{E}_{\mathbf{Q}} \left[ \log \frac{d\mathbf{Q}}{d\mathbf{P}} - \log L \right]$$

$$\mathbf{P}: dX_t = b(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

$$\mathbf{Q}: dX_t = (b + \sigma_t u)(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

# Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{d\mathbf{Q}^*}{d\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathbf{E}_{\mathbf{Q}} \left[ \log \frac{d\mathbf{Q}}{d\mathbf{Q}^*} \right] = \min \mathbf{E}_{\mathbf{Q}} \left[ \log \frac{d\mathbf{Q}}{d\mathbf{P}} - \log L \right]$$

$$\mathbf{P}: dX_t = b(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

$$\mathbf{Q}: dX_t = (b + \sigma_t u)(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

Forward-forward RND and Girsanov

$$\begin{aligned} \mathbf{P}: dX_t &= f(X_t, t)dt + \sigma_t dW_t, X_0 \sim p_0 \\ \mathbf{Q}: dX_t &= h(X_t, t)dt + \sigma_t dW_t, X_0 \sim q_0 \end{aligned}$$

$$\frac{d\mathbf{P}}{d\mathbf{Q}}(X) = \frac{p(X_0)}{q(X_0)} \exp \left( \underbrace{\int \frac{f_t(X_t)}{\sigma_t^2} \cdot dX_t - \frac{f_t^2(X_t)}{2\sigma_t^2} dt}_{\text{Forward Ito Integral}} - \int \frac{g_t(X_t)}{\sigma_t^2} \cdot dX_t + \frac{g_t^2(X_t)}{2\sigma_t^2} dt \right)$$

Initial density ratio      Forward Ito Integral  $\int a_t(X_t) \cdot dX_t = \lim \sum a_n(X_n) \cdot (X_{n+1} - X_n)$

# Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{d\mathbf{Q}^*}{d\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathbf{E}_{\mathbf{Q}} \left[ \log \frac{d\mathbf{Q}}{d\mathbf{Q}^*} \right] = \min \mathbf{E}_{\mathbf{Q}} \left[ \int \frac{1}{2} \|u_t\|^2 dt - \log L \right]$$

$$\mathbf{P}: dX_t = b(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

$$\mathbf{Q}: dX_t = (b + \sigma_t u)(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

# Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{d\mathbf{Q}^*}{d\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathbf{E}_{\mathbf{Q}} \left[ \log \frac{d\mathbf{Q}}{d\mathbf{Q}^*} \right] = \min \mathbf{E}_{\mathbf{Q}} \left[ \int \frac{1}{2} \|u_t\|^2 dt - \log L \right]$$

$$\mathbf{P}: dX_t = b(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

$$\mathbf{Q}: dX_t = (b + \sigma_t u)(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

**stochastic optimal control**

# Variational Inference with Path RND

Variational Inference in Path space:

$$\frac{d\mathbf{Q}^*}{d\mathbf{P}} \propto L$$

$$\mathbf{Q} = \min \mathbf{E}_{\mathbf{Q}} \left[ \log \frac{d\mathbf{Q}}{d\mathbf{Q}^*} \right] = \min \mathbf{E}_{\mathbf{Q}} \left[ \int \frac{1}{2} \|u_t\|^2 dt - \log L \right]$$

$$\mathbf{P}: dX_t = b(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

$$\mathbf{Q}: dX_t = (b + \sigma_t u)(X_t, t)dt + \sigma_t dW_t \quad X_0 \sim p_0$$

**stochastic optimal control**

👉 **fine-tune diffusion models [1]**

👉 **neural samplers [2,3,4...]**

[1] Domingo-Enrich, C., Drozdal, M., Karrer, B., & Chen, R. T. (2024). Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. ICLR 2025

[2] Havens, A., Miller, B. K., Yan, B., Domingo-Enrich, C., Sriram, A., Wood, B., ... & Chen, R. T. (2025). Adjoint sampling: Highly scalable diffusion samplers via adjoint matching. ICML 2025.

[3] Liu, G. H., Choi, J., Chen, Y., Miller, B. K., & Chen, R. T. (2025). Adjoint Schrodinger Bridge Sampler. arXiv.

[4] Zhu, Y., Guo, W., Choi, J., Liu, G. H., Chen, Y., & Tao, M. (2025). MDNS: Masked Diffusion Neural Sampler via Stochastic Optimal Control. arXiv.

# Collaborators on FEAT




# Collaborators on RNE





# Fantastic Path RNDs

 *intuitive understand from sequence of Gaussian kernels*

 *Forward-forward RND with Girsanov theorem*

 *Forward-backward RND*



# Where to Find Them?

## Importance Sampling with Path RND

*AIS, Jarzynski and Crook's Fluctuation Theorem*

*Free-energy estimation*

*Density estimation*

*Generation control*

*Energy regularisation*

## Variational Inference with Path RND

*Neural samplers*

*Diffusion model fine-tuning*



Thank You!

Jiajun He (jh2383@cam.ac.uk)

Francisco Vargas (vargfran@gmail.com)