GRAPH-INFORMED IMPORTANCE SAMPLING

APPLICATION IN DYNAMIC RARE EVENT SIMULATION

Guillaume Chennetier [1,2] · Hassane Chraibi [1] · Anne Dutfoy [1] · Josselin Garnier [2]

[1] EDF Lab Paris-Saclay, Boulevard Gaspard Monge, 91120 Palaiseau, France.
[2] CMAP, Ecole Polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France.

Monte Carlo method

Quantity of Interest $\quad \mathbf{\Phi} := \mathbb{E}_{\mathbf{X}\sim\mathbf{p}}\left[\phi(\mathbf{X})\right]$

↪ No access to direct observations of $\mathbf{X}$

↪ Nominal distribution $\mathbf{p}$ is numerically samplable

Classical Monte Carlo $\quad$ Generating i.i.d. sample $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim \mathbf{p}$

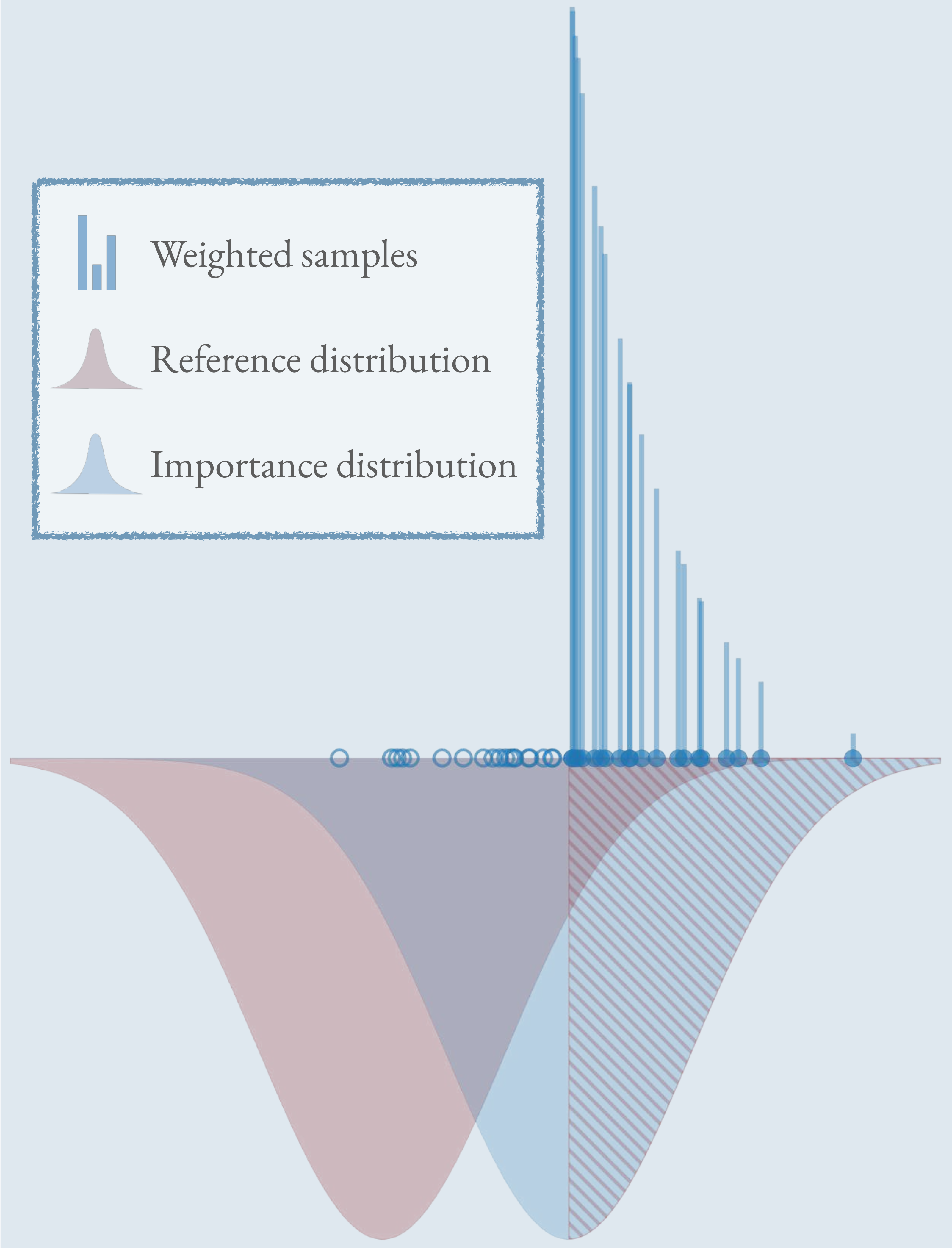
$$\widehat{\mathbf{\Phi}}_n^{\mathsf{CMC}} := \frac{1}{n}\sum_{k=1}^{n}\phi(\mathbf{X}_k)$$

↪ High relative variance when $\mathbf{p}$ puts its mass where $|\phi|$ is small

↪ Rare event case: $\quad \phi(\mathbf{X}) = 1_{\mathbf{X}\in\mathbf{F}}$

PART I

IMPORTANCE SAMPLING



Importance sampling for variance reduction

Importance sampling trick Using an alternative distribution $\mathbf{g}$

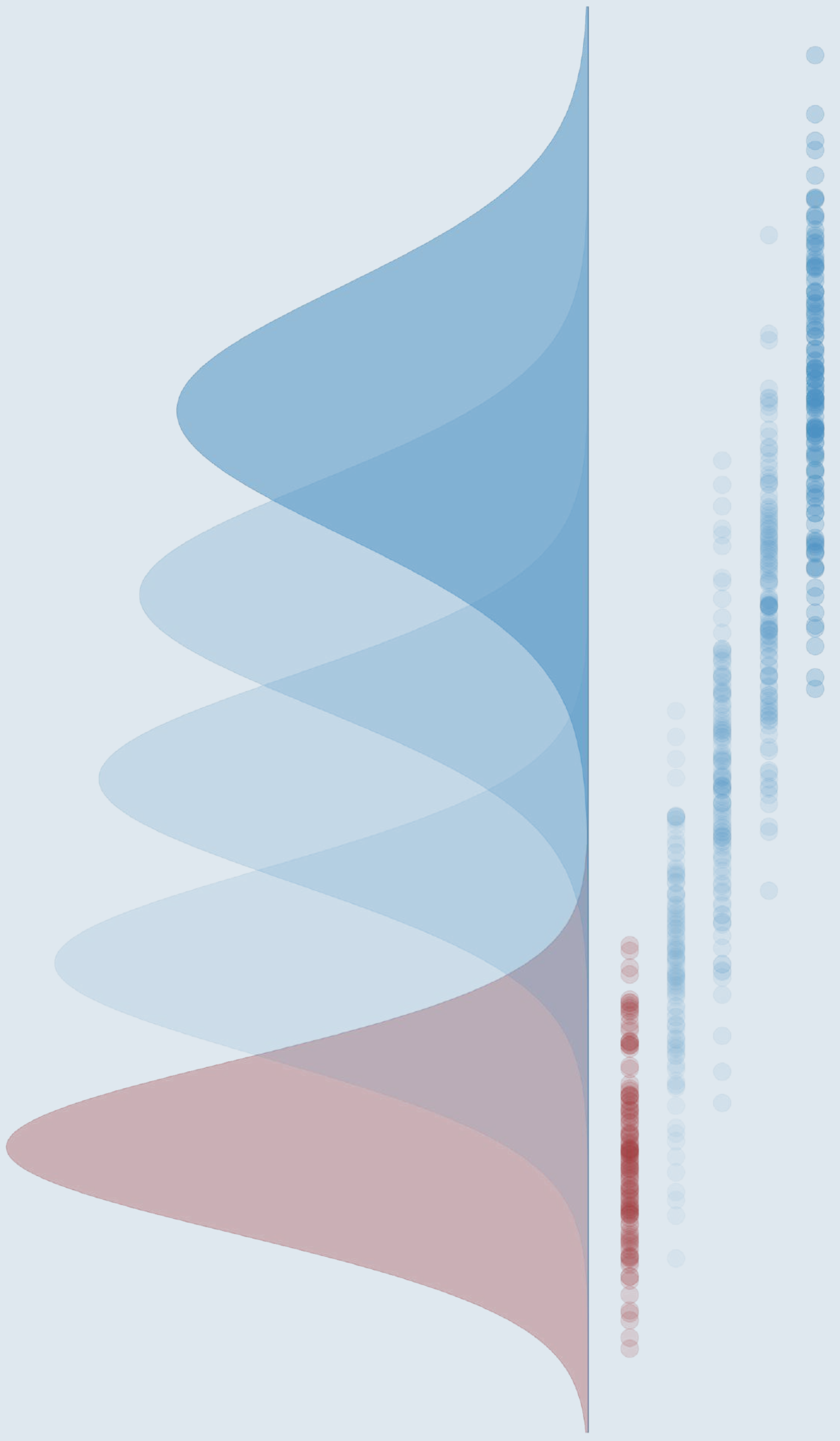
- Let $\mathbf{p}$ and $\mathbf{g}$ be probability densities function with respect to a measure μ on $\mathcal{X}$ such that $\phi(\mathbf{x})\mathbf{p}(\mathbf{x}) \neq 0 \Rightarrow \mathbf{g}(\mathbf{x}) \neq 0$

$$\mathbf{\Phi} = \int_{\mathcal{X}} \phi(\mathbf{x}) \frac{\mathbf{p}(\mathbf{x})}{\mathbf{g}(\mathbf{x})} \mathbf{g}(\mathbf{x}) \mu(\mathrm{d}\mathbf{x}) = \mathbb{E}_{\mathbf{X} \sim \mathbf{g}} \left[\phi(\mathbf{X}) \frac{\mathbf{p}(\mathbf{X})}{\mathbf{g}(\mathbf{X})} \right]$$

IS estimator Generating i.i.d. sample $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim \mathbf{g}$

$$\widehat{\mathbf{\Phi}}_n^{\mathrm{IS}} := \frac{1}{n} \sum_{k=1}^{n} \phi(X_k) \frac{\mathbf{p}(X_k)}{\mathbf{g}(X_k)}$$

- Variance relies on the choice of $\mathbf{g}$
- Optimal but untractable IS p.d.f. $\mathbf{g}^* : \mathbf{x} \propto |\phi(\mathbf{x})| \times \mathbf{p}(\mathbf{x})$



Adaptive importance sampling

Cross entropy procedure Finding the best proposal in a family $(\mathbf{g}_\theta)_{\theta\in\Theta}$

$$\arg\min_{\theta\in\Theta} \mathcal{D}_{\mathsf{KL}}(\mathbf{g}^*\|\mathbf{g}_\theta) = \arg\max_{\theta\in\Theta} \mathbb{E}_{\mathbf{X}\sim\mathbf{p}}\left[|\phi(\mathbf{X})| \log \mathbf{g}_\theta(\mathbf{X})\right]$$

Sequential recycling At iteration $t = 1, \ldots, T$

- Simulation step $\mathbf{X}_{t,1}, \ldots, \mathbf{X}_{t,n_t} \sim \mathbf{g}_{\theta^{(t)}}$
- Optimization step

$$\theta^{(t+1)} \in \arg\max_{\theta\in\Theta} \sum_{\ell=1}^{t} \sum_{k=1}^{n_\ell} \frac{|\phi(\mathbf{X}_{\ell,k})|\, \mathbf{p}(\mathbf{X}_{\ell,k})}{\mathbf{g}_{\theta^{(\ell)}}(\mathbf{X}_{\ell,k})} \log \mathbf{g}_\theta(\mathbf{X}_{\ell,k})$$

Final estimator

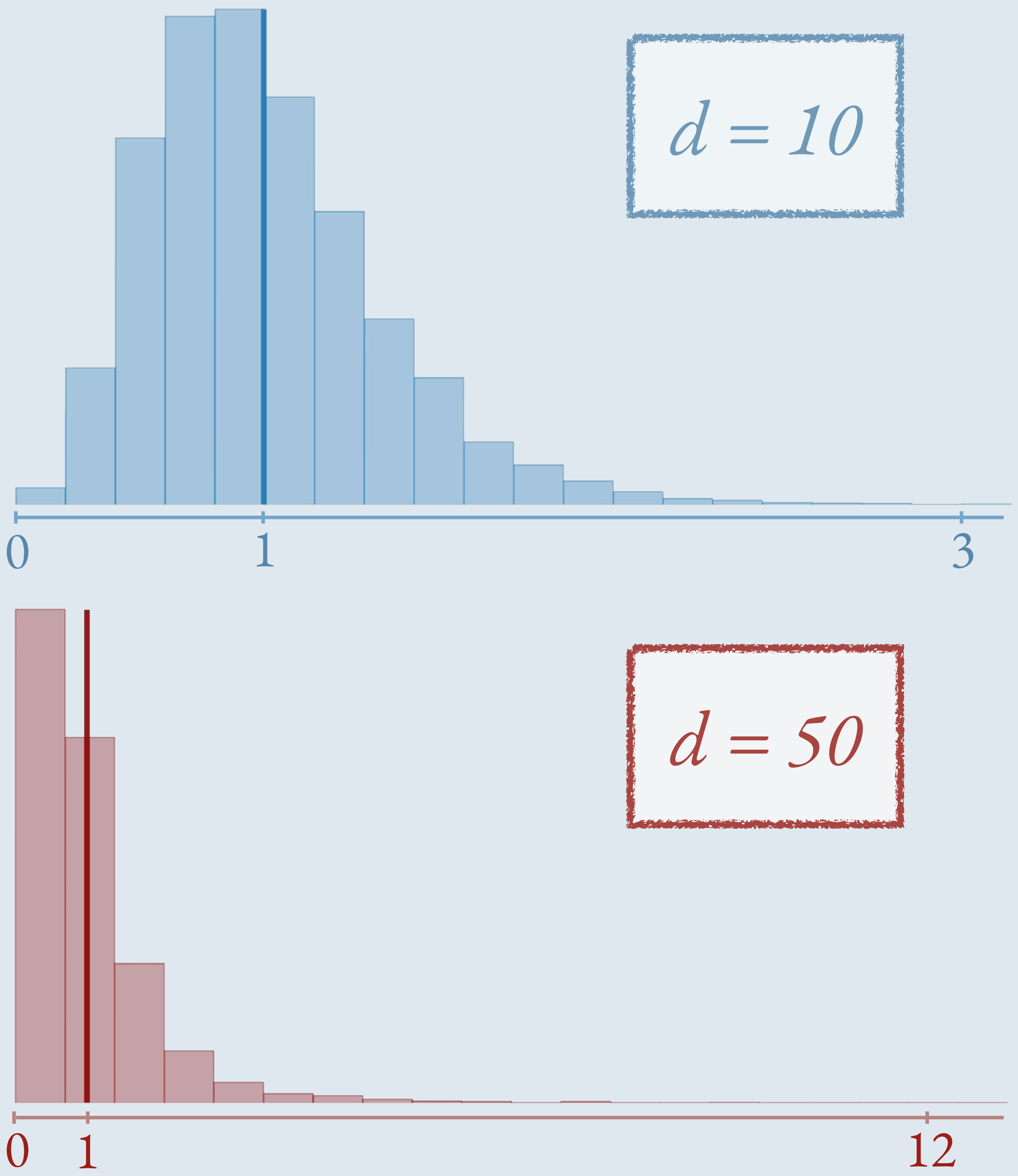
$$\widehat{\mathbf{\Phi}}^{\mathsf{AIS}} = \frac{1}{T}\sum_{\ell=1}^{t} \frac{1}{n_t} \sum_{k=1}^{n_\ell} \frac{\phi(\mathbf{X}_{\ell,k})\, \mathbf{p}(\mathbf{X}_{\ell,k})}{\mathbf{g}_{\theta^{(\ell)}}(\mathbf{X}_{\ell,k})}$$

Importance sampling in high dimension

Histogram of $\frac{\mathbf{p}(\mathbf{X})}{\mathbf{g}(\mathbf{X})}$ with $\mathbf{X} \sim \mathbf{g}$

$$\mathbf{p} \sim \mathcal{N}(0_d, I_d)$$

$$\mathbf{g} \sim \mathcal{N}(0_d + 0.1, I_d)$$



Weights degeneracy

- For any distribution $\mathbf{g}$, we have $\mathbb{E}_{\mathbf{X}\sim\mathbf{g}}\left[\frac{\mathbf{p}(\mathbf{X})}{\mathbf{g}(\mathbf{X})}\right] = 1$
- But the more $\mathbf{p}$ and $\mathbf{g}$ differ, the more often $\frac{\mathbf{p}(\mathbf{X})}{\mathbf{g}(\mathbf{X})}$ is close to 0

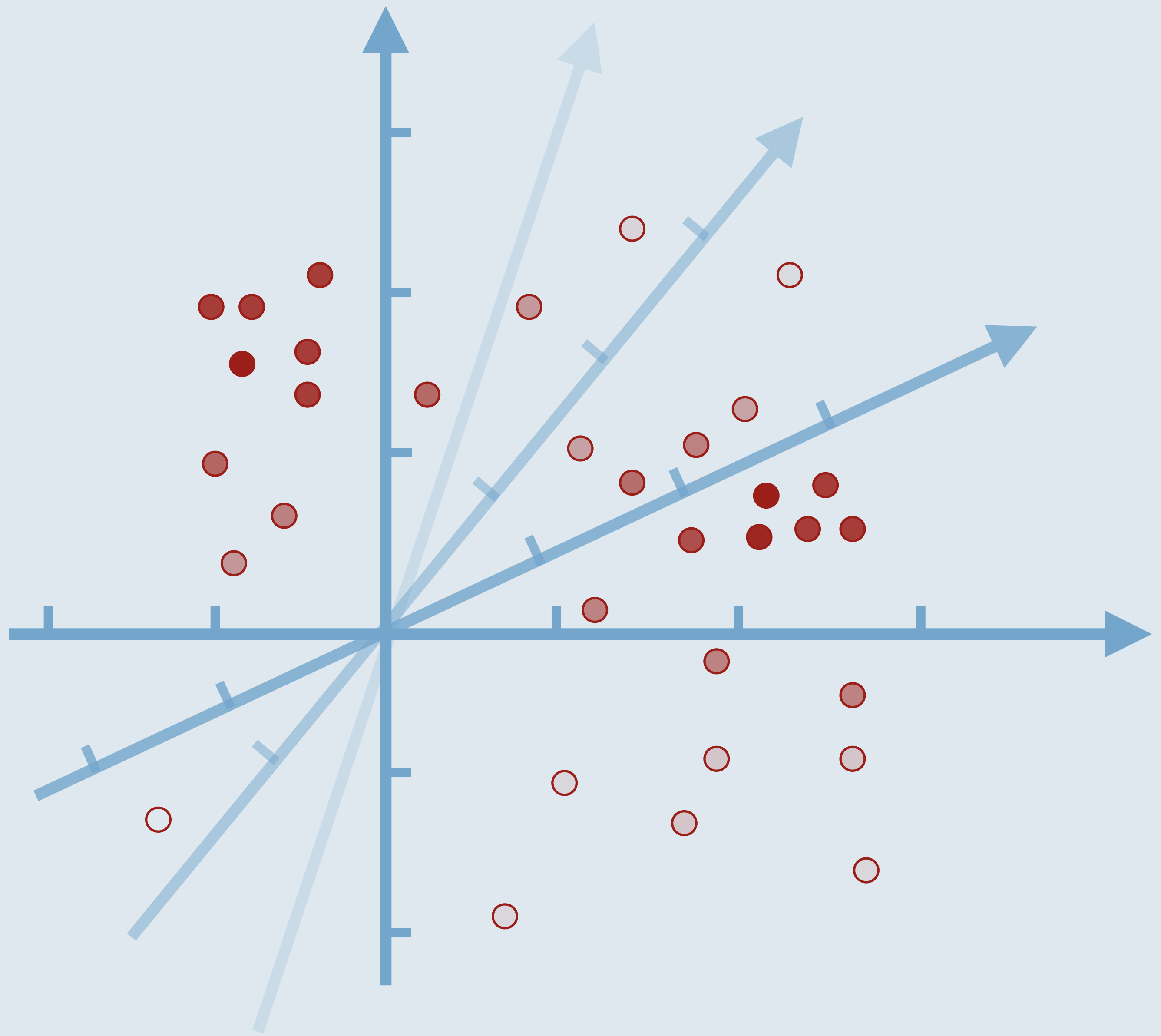
AIS as a density estimation problem

- Slower convergence with complex and large distribution families
- Bad performance with simple and small distribution families

Propositions in the literature

- Dimension reduction with projection in well-chosen subspaces
- Generative models with good properties in high dimension (Julien's talk)

Stochastic process, entropy and dimension



High dimensional spaces can contain

- Vectors with a large number of coordinates $\mathbf{X} \in \mathbb{R}^d$
- But also trajectories of a stochastic process $\mathbf{X} = (X_t)_{t \in [0, s_{\max}]}$

  No obvious way to measure the dimension of a space of trajectories

Large entropy means large space to explore

- $H(\mathbf{X}) = \mathbb{E}_{\mathbf{X}\sim\mathbf{p}}\left[-\log \mathbf{p}(\mathbf{X})\right]$ is large when $\mathbb{E}_{\mathbf{X}\sim\mathbf{p}}\left[\mathbf{p}(\mathbf{X})\right]$ is small
- Since $\int_{\mathcal{X}} \mathbf{p}(\mathbf{x})\mu(d\mathbf{x}) = 1$, it means that we integrate on a large space

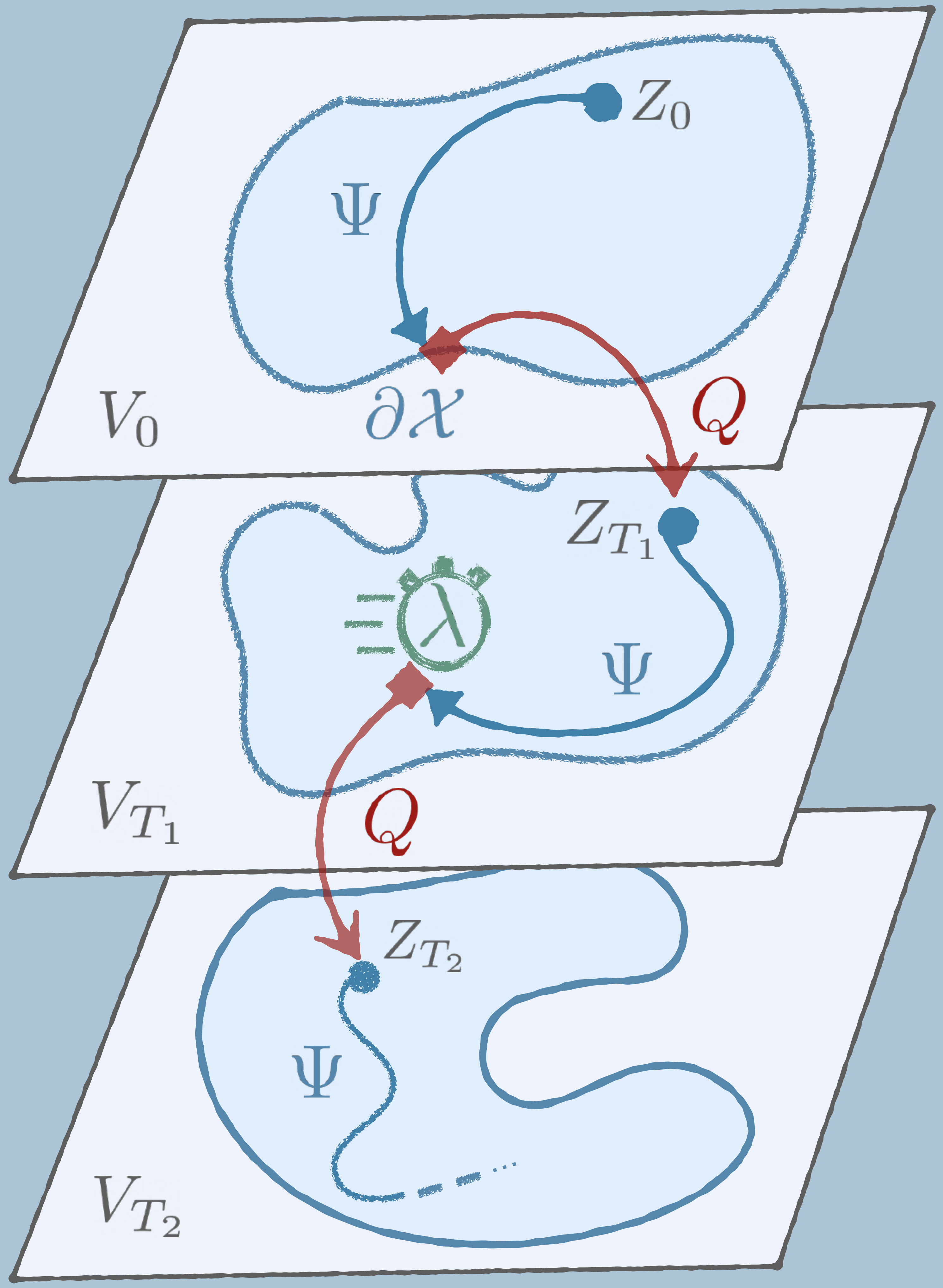
Simple Poisson process example

- Let $\mathbf{X}$ be a trajectory of a simple Poisson process of intensity 1
- Then $H(\mathbf{X}) = H\big(\mathcal{N}(0_d, I_d)\big)$ with $d = \dfrac{2 \times s_{\max}}{1 + \log(2\pi)}$

PART II

PIECEWISE DETERMINISTIC MARKOV PROCESSES

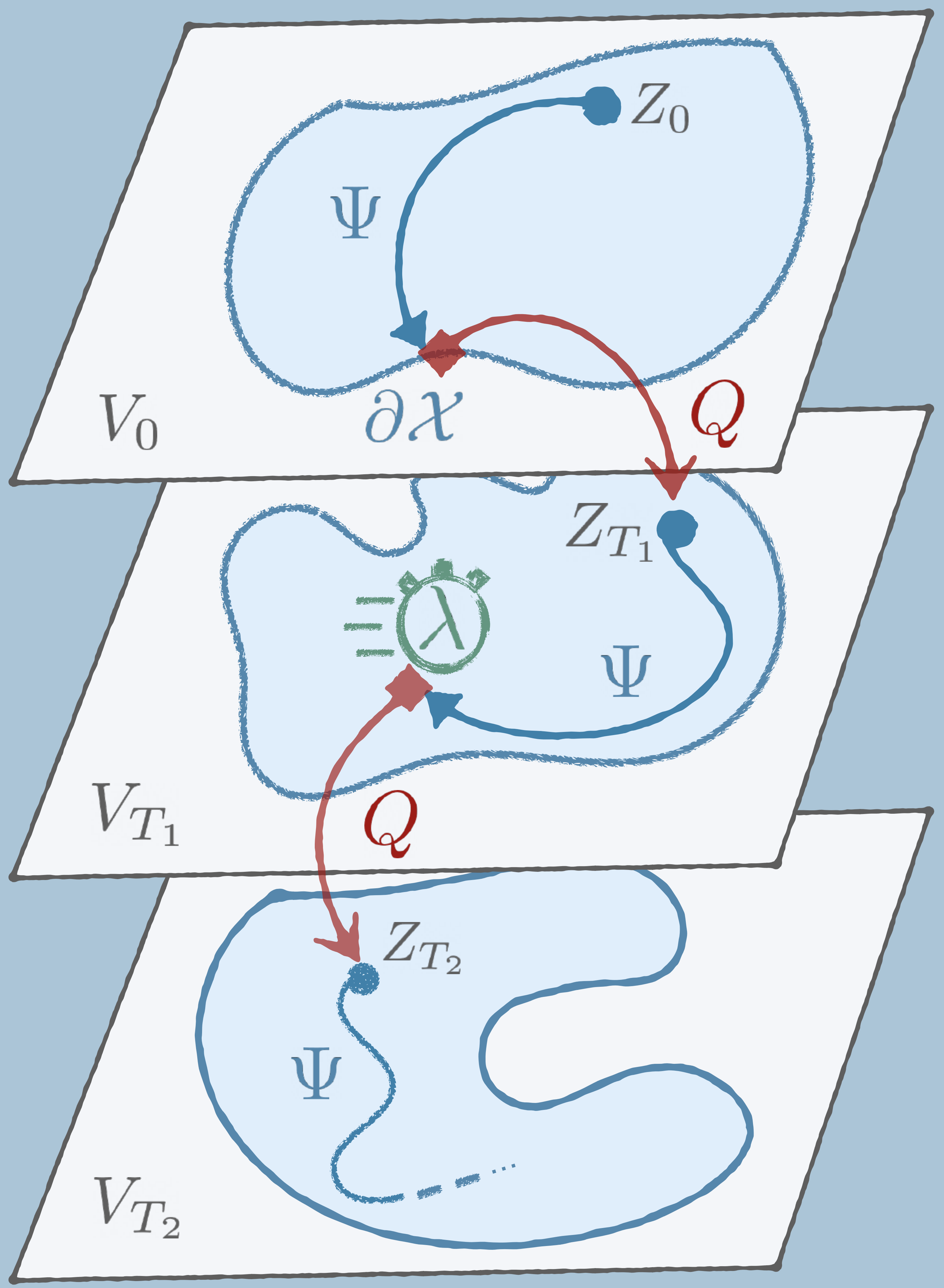
Piecewise Deterministic Markov Processes



PDMP	Class of all non-diffusive Markov processes
	Mark H Davis 1984
Hybrid process	$X_t = (Z_t, V_t) \in \mathcal{Z} \times \mathcal{V} = \mathcal{X}$
	$Z_t \in \mathcal{Z}$ is continuous and called « position »
	$V_t \in \mathcal{V}$ is discrete and called « regime »

Local characteristics of the PDMP

Flow	Ψ	Deterministic dynamics between two jumps
Intensity	λ	Gives the distribution of the random jump times
Kernel	Q	Gives the distribution of the post-jump locations



Modeling dynamic industrial systems

Position	Physical quantities (e.g. temperature, pressure)
Regime	Status (e.g. `ON`, `OFF`) of the system components
Flow	ODEs given by physical laws and parameterized by the status of the components
Boundaries	Physical constraints and control mechanisms when thresholds are reached
Random jumps	Random failures and repairs of the components
Intensity and kernel	Given by components jump rates, which may depend on physical variables

B. de Saporta et al. Numerical methods for simulation and optimization of piecewise deterministic Markov processes: application to reliability

Reliability assessment

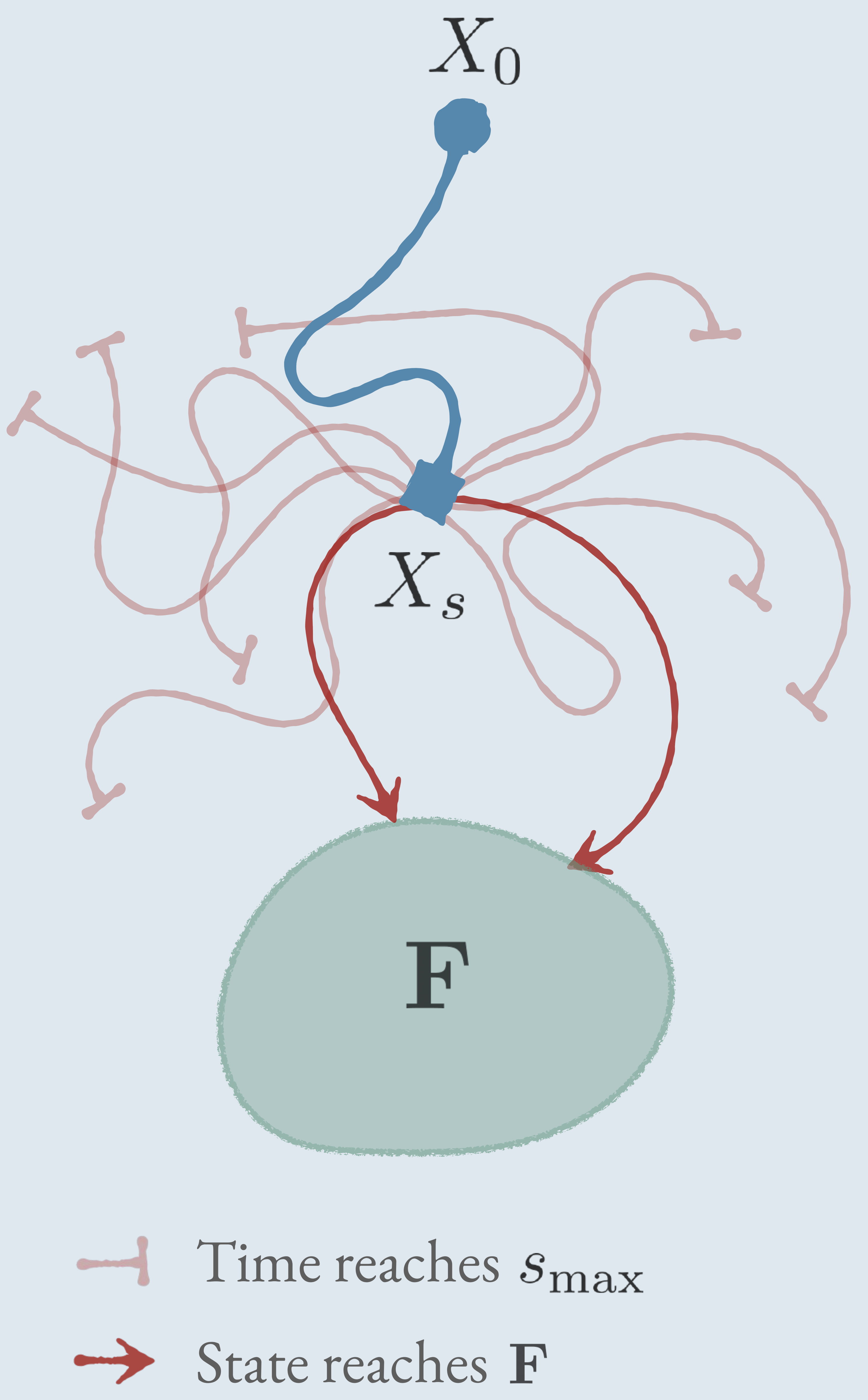
Aim Estimating the probability of critical failure of the system

$$\mathcal{P}_{\mathbf{F}} := \mathbb{P}_{\mathbf{p}}\left(\mathbf{X} \in \mathcal{T}_{\mathbf{F}}\right)$$

Notations

$\mathbf{F} \subset \mathcal{X}$	Critical failure domain
$s_{\max}$	Maximal duration of a PDMP trajectory
$\mathcal{T}_{\mathbf{F}}$	Set of faulty trajectories $\left\{(X_t)_{t\in[0,s_{\max}]} \mid \exists t \ : \ X_t \in \mathbf{F}\right\}$
$\mathbf{X}$	Complete PDMP trajectory $(X_t)_{t\in[0,s_{\max}]}$
$\mathbf{p}$	Reference distribution of the PDMP trajectory
$\mathcal{P}_{\mathbf{F}}$	Probability to reach $\mathbf{F}$ before time $s_{\max}$

Optimal importance sampling of PDMP



A PDMP distribution is characterized by its intensity and kernel.
Optimal choice relies on the knowledge of the **committor function**

Thomas Galtier's Phd Thesis 2021

Committor and edge committor functions

$$\xi^*(x,s) = \mathbb{P}_{\mathbf{p}}\left(\mathbf{X} \in \mathcal{T}_{\mathbf{F}} \mid X_s = x\right)$$

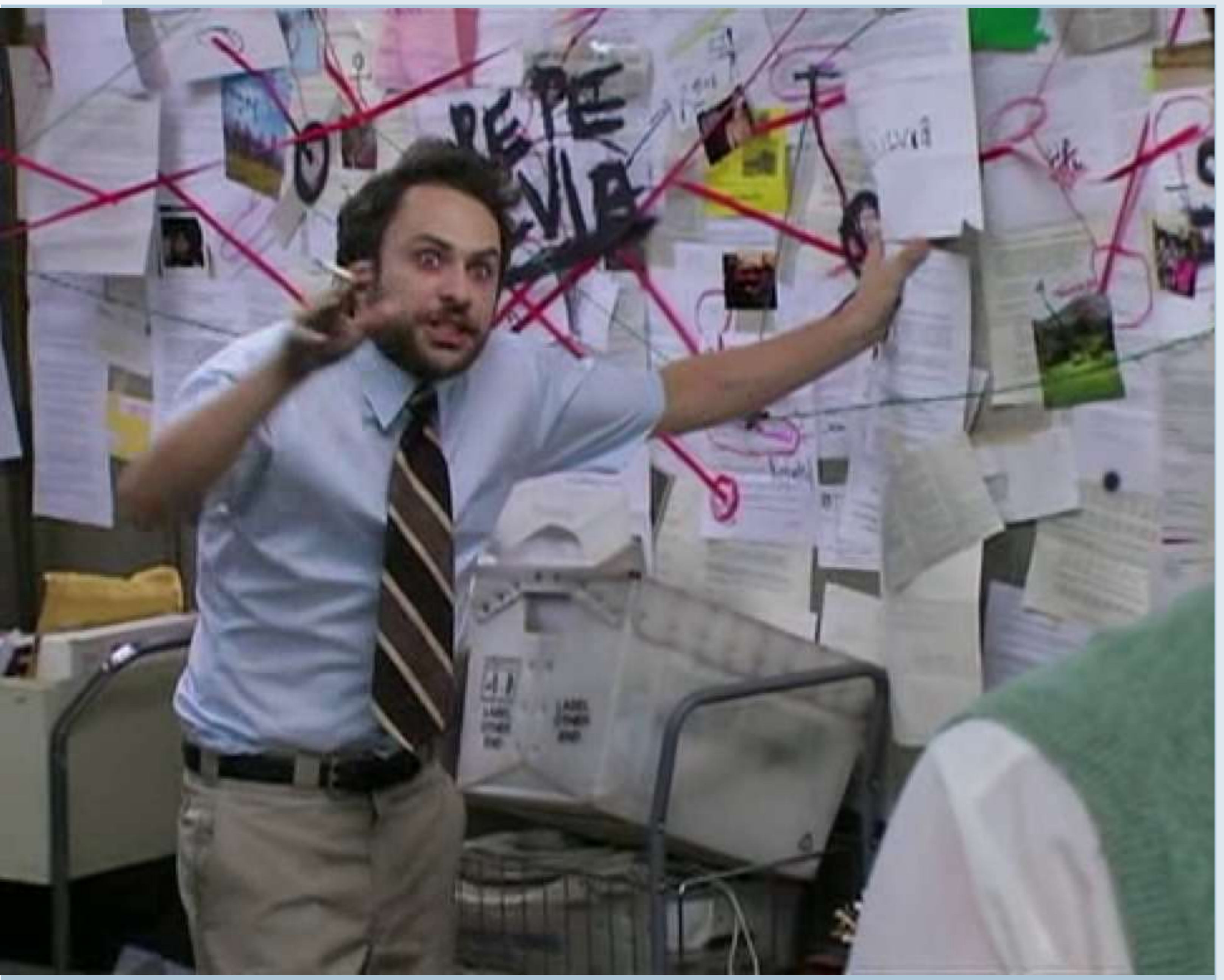
$$\xi^{*-}(x^-,s) = \mathbb{E}_{X_s \sim Q(\cdot|x^-)}\left[\xi^*(X_s,s)\right]$$

Optimal jump intensity and jump kernel

$$\lambda^*(x,s) = \lambda(x) \times \frac{\xi^{*-}(x,s)}{\xi^*(x,s)}$$

$$Q^*(x,s \mid x^-) = Q(x \mid x^-) \times \frac{\xi^*(x,s)}{\xi^{*-}(x^-,s)}$$

Plan of attack



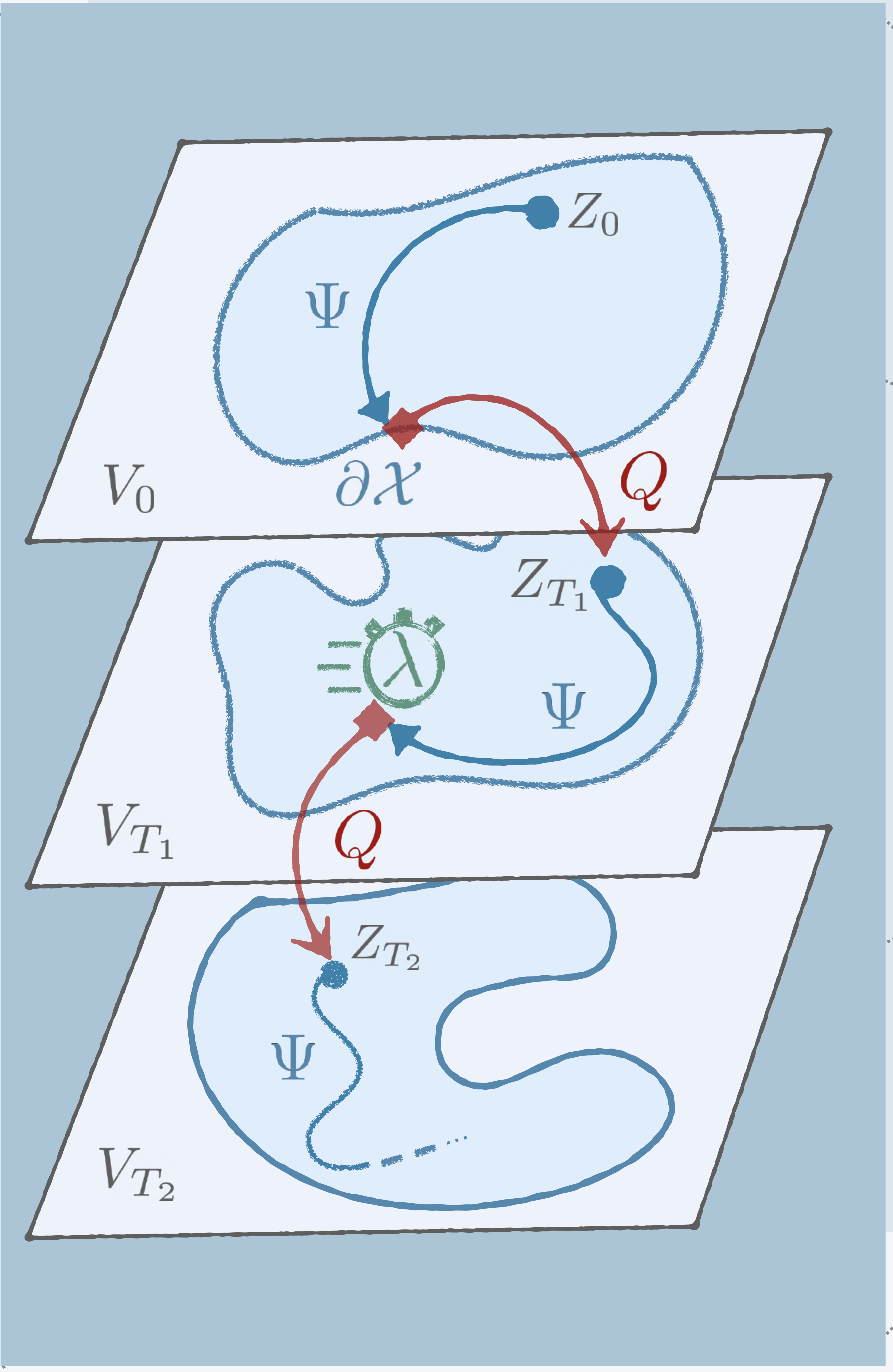
We simply wish to perform informed adaptive importance sampling of piecewise deterministic Markov processes for rare event simulation in reliability assessment.

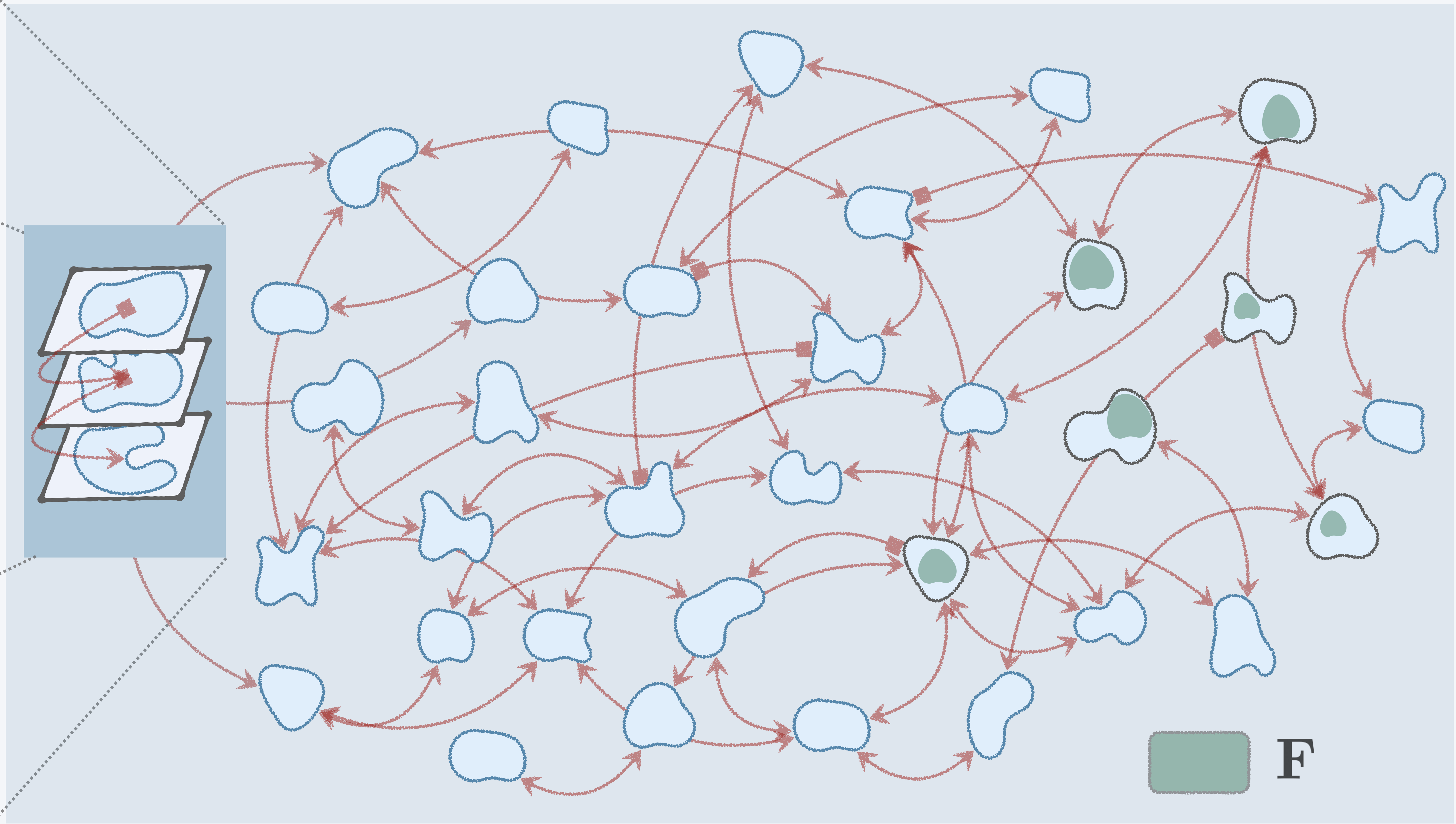
Methodology	Chennetier et al. (2024)
1. Approximating the committor	Choosing a parametric family $(\xi_\theta)_{\theta\in\Theta}$ of approximations of the committor function ξ^*
2. Importance distributions	Determine the family $(\mathbf{g}_\theta)_{\theta\in\Theta}$ by replacing ξ^* by ξ_θ in the previous optimality expressions
3. Cross entropy procedure	Both select a good importance distribution $\mathbf{g}_\theta$ and estimate the probability of failure $\mathcal{P}_\mathbf{F}$
4. Gaussian confidence intervals	We proved convergence and asymptotic normality of the estimator with recycling scheme under simple conditions on the PDMP and on $(\xi_\theta)_{\theta\in\Theta}$

PART III

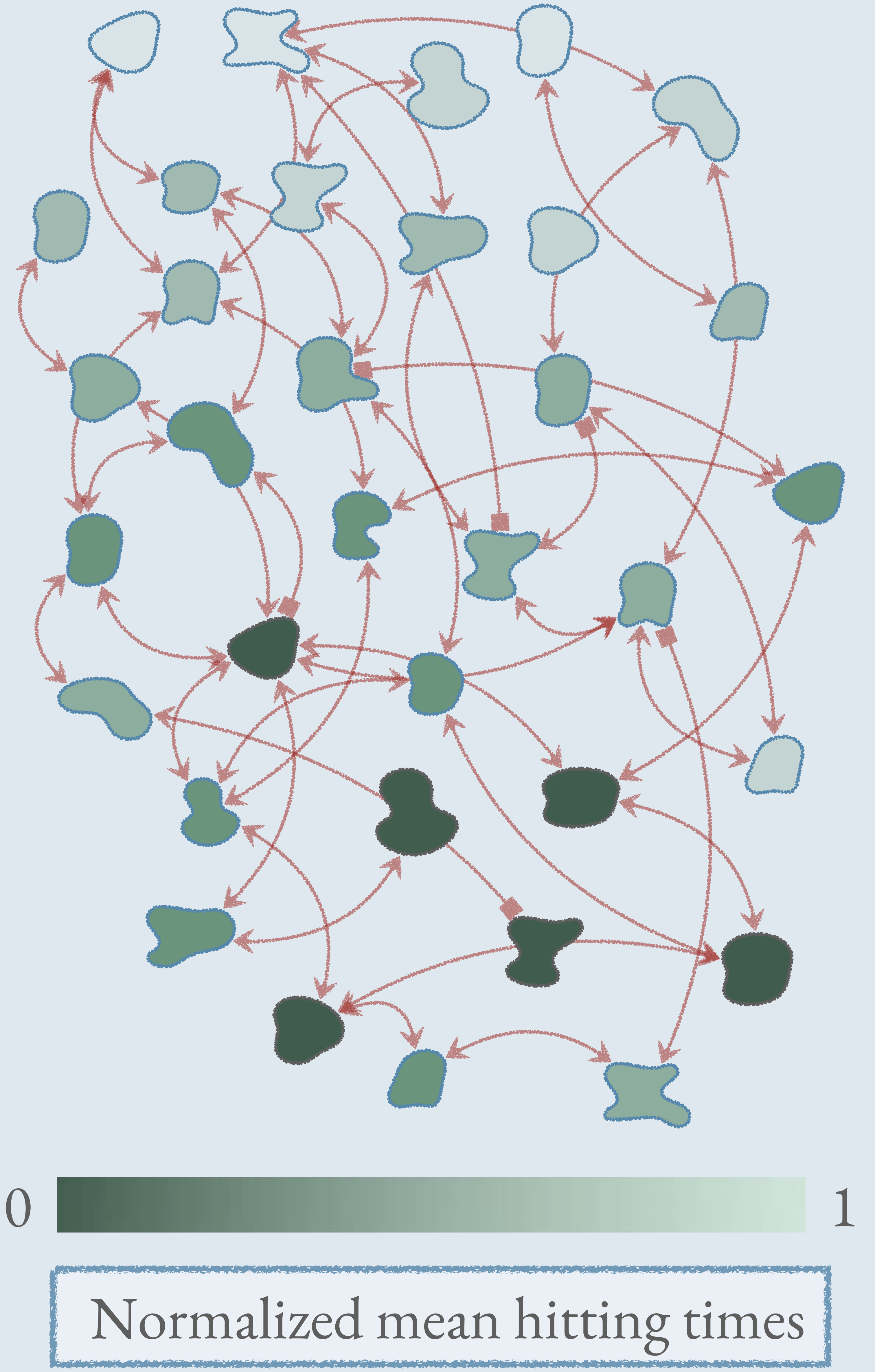
GRAPH-BASED APPROXIMATION

Random walk on a graph





The regime of the PDMP evolves according to a non-Markovian random walk on a graph



Mean hitting times

Failure regimes $\mathcal{V}_{\mathbf{F}} := \{v \in \mathcal{V} \mid \exists z \in \mathcal{Z} \, : \, x = (z,v) \in \mathbf{F}\}$

Idea Computation of mean hitting times of $\mathcal{V}_{\mathbf{F}}$ for a Markovian time-homogeneous random walk $(Y_t)_t$ with generator A on $\mathcal{V}$

Mean hitting time

$$\tau_v = \inf\{t \leq 0 \, : \, Y_t \in \mathcal{V}_{\mathbf{F}} \mid Y_0 = v\}$$

$$h_v = \mathbb{E}_{(Y_t)_t \sim A}\left[\tau_v\right]$$

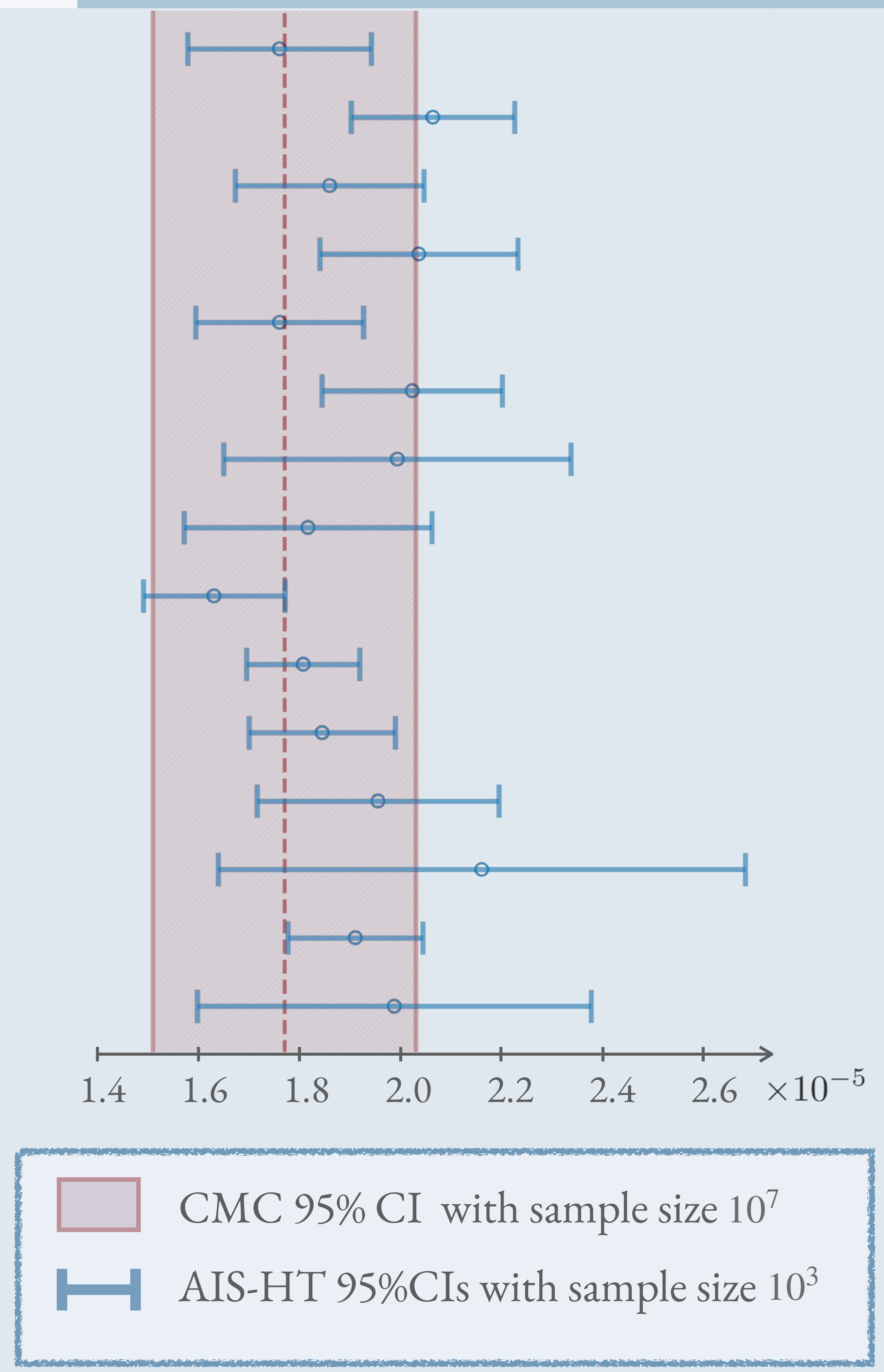
Dynkin's formula consequence

$$\begin{cases} \sum_{v' \in \mathcal{V}} A[v,v'] h_{v'} = -1 & \text{if } v \notin \mathcal{V}_{\mathbf{F}} \\ h_v = 0 & \text{if } v \in \mathcal{V}_{\mathbf{F}} \end{cases}$$

Committor function approximation

$$\xi_\theta(v) = \exp\left[-\sum_{i=1}^{d_\Theta} \theta_i \times (h_v)^i\right]$$

Numerical results

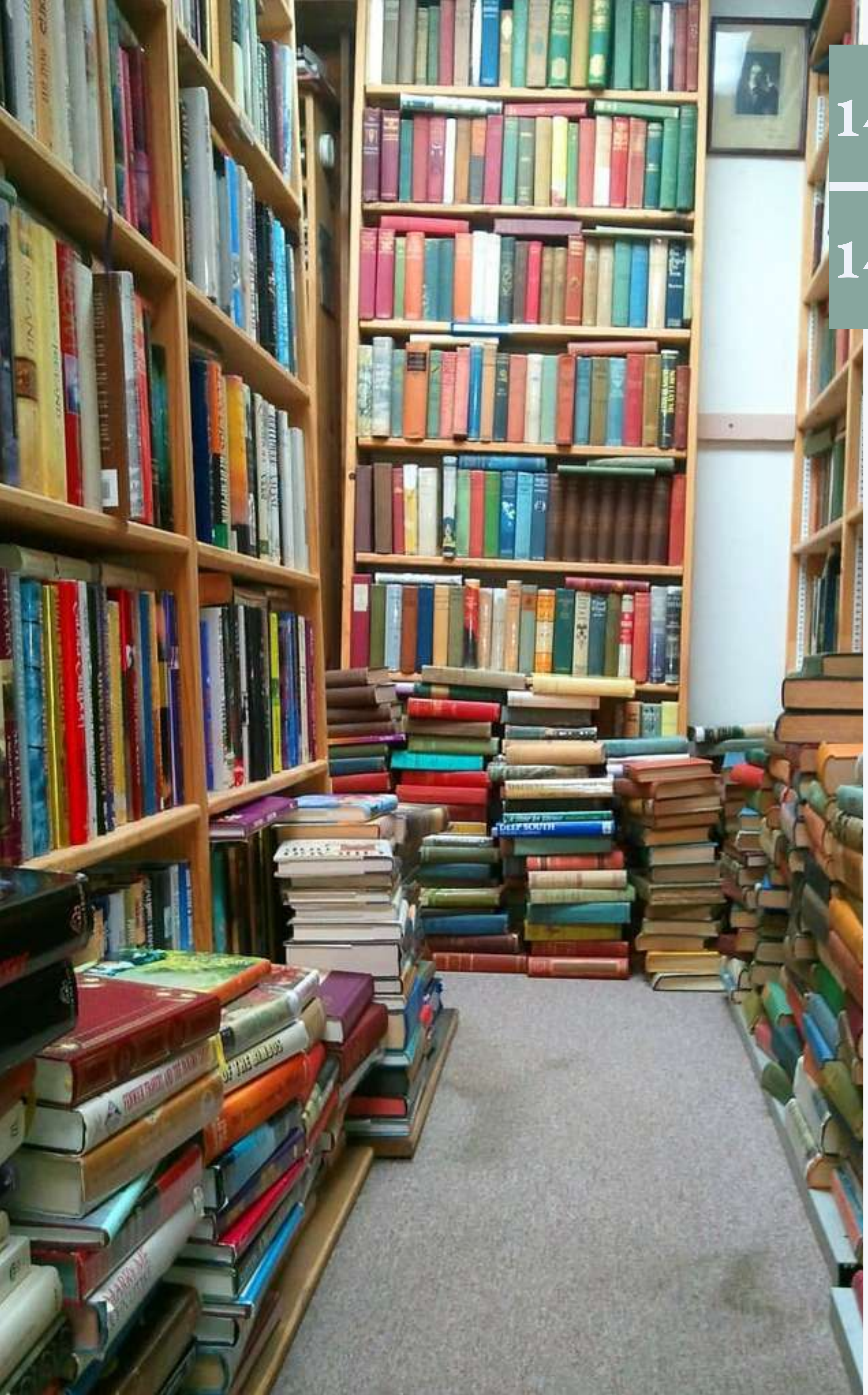


Comparison between classical Monte Carlo on an industrial test case from nuclear industry (spent fuel pool). The corresponding graph has 32,768 vertices.

Method	N	$\widehat{\mathcal{P}}_{\mathbf{F}} \times 10^5$	**C.o.v**	**95% CI** $\times 10^5$
CMC	10^5	2	223.60	[0 ; 4.77]
	10^6	1.3	277.35	[0.59 ; 2.01]
	10^7	1.77	237.68	[1.51 ; 2.03]
AIS-HT	10^3	1.86	1.62	[1.67 ; 2.04]
	10^4	2.01	0.88	[1.98 ; 2.05]

↪ Variance reduction factor about 10,000

THANK YOU



Bibliography

- M. H. A. Davis (1984), Piecewise-Deterministic Markov processes: A general class of non-diffusion stochastic models. *J. R. Stat. Soc. B 46:353-388.*
- B. de Saporta, F. Dufour and H. Zhang (2015), Numerical methods for simulation and optimization of piecewise deterministic Markov processes: application to reliability. *Mathematics and statistics series Wiley-ISTE.*
- T. Galtier (2019), Accelerated Monte-Carlo methods for Piecewise Deterministic Markov Processes for a faster reliability assessment of power generation systems within the PyCATSHOO toolbox. *PhD thesis.*
- G. Chennetier, H. Chraibi, A. Dutfoy and J. Garnier (2024), Adaptive importance sampling based on fault tree analysis for piecewise deterministic Markov process. *SIAM/ASA Journal of Uncertainty Quantification, 12(1), 128-156.*