Yet another Talk on Multilevel Monte Carlo
Different I Promise!

Robert Scheichl

Department of Mathematical Sciences

University of Bath

Joint work over 8 years with a large number of collaborators, including

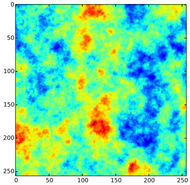
G Detommaso, **I Graham**, **E Müller**, M Parkinson & **T Shardlow** (all Bath);
J Charrier (Marseille); A Cliffe †; T Dodwell (Exeter); M Giles (Oxford);
A Teckentrup (Edinburgh); E Ullmann (TUM)

SAMBa Integrative Think Tank 6
BRLSI, Bath, June 5th 2017

Motivation

INPUT $\longrightarrow$ **MODEL** $\longrightarrow$ **OUTPUT**

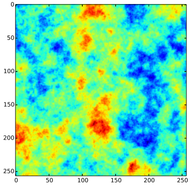
INPUT $\longrightarrow$ **MODEL** $\longrightarrow$ **OUTPUT**



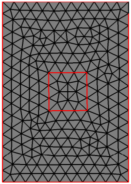
Rock permeability

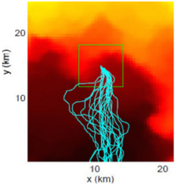
Motivation

INPUT $\longrightarrow$ MODEL $\longrightarrow$ OUTPUT



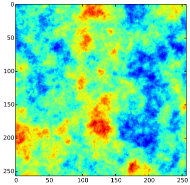
Rock permeability



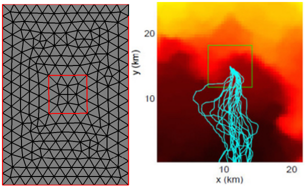


FE analysis of leaking waste

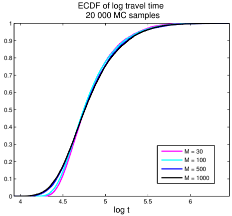
INPUT ⟶ MODEL ⟶ OUTPUT



Rock permeability



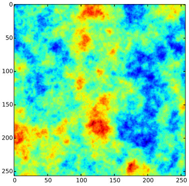
FE analysis of leaking waste



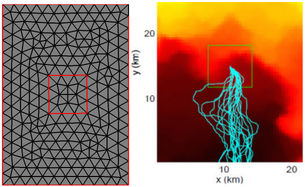
Radionuclides reaching drinkwater?

Motivation

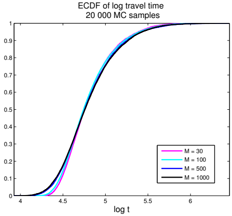
INPUT ⟶ **MODEL** ⟶ **OUTPUT**



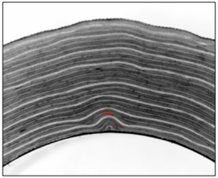
Rock permeability



FE analysis of leaking waste



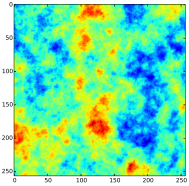
Radionuclides reaching drinkwater?



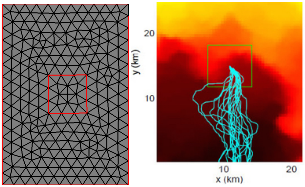
Composite material

Motivation

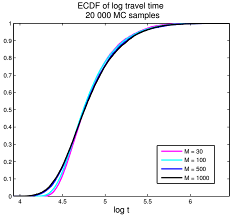
INPUT ⟶ MODEL ⟶ OUTPUT



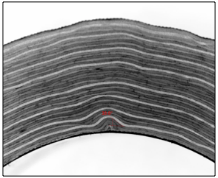
Rock permeability



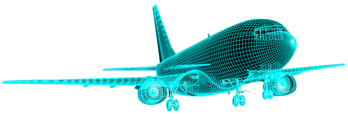
FE analysis of leaking waste



Radionuclides reaching drinkwater?

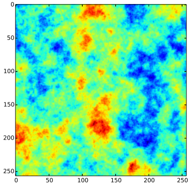


Composite material

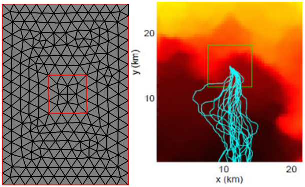


FE analysis of aircraft wing

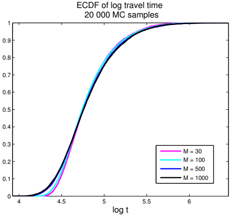
INPUT ⟶ MODEL ⟶ OUTPUT



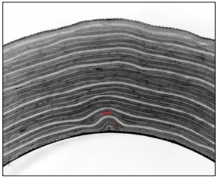
Rock permeability



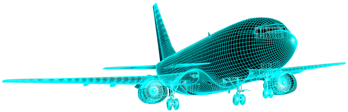
FE analysis of leaking waste



Radionuclides reaching drinkwater?



Composite material



FE analysis of aircraft wing

Wing failing?

Uncertainty Quantification / Stochastic Simulation

$$Z(\omega) \in X \xrightarrow{\text{Model}} u(\omega) \in V \xrightarrow{\text{Output}} Q(\omega) \in \mathbb{R}(\text{or } \mathbb{R}^J) \xrightarrow{\text{Statistics}} \mathbb{E}[Q]$$

noise process | state of process | quantity of interest

Uncertainty Quantification / Stochastic Simulation

$Z(\omega) \in X$	$\xrightarrow{\text{Model}}$	$u(\omega) \in V$	$\xrightarrow{\text{Output}}$	$Q(\omega) \in \mathbb{R}(\text{or } \mathbb{R}^J)$	$\xrightarrow{\text{Statistics}}$	$\mathbb{E}[Q]$
noise process		state of process		quantity of interest		
uncertain input		"latent" field		quantity of interest		

$$Z(\omega) \in X \xrightarrow{\text{Model}} u(\omega) \in V \xrightarrow{\text{Output}} Q(\omega) \in \mathbb{R}(\text{or } \mathbb{R}^J) \xrightarrow{\text{Statistics}} \mathbb{E}[Q]$$

noise process	state of process	quantity of interest
uncertain input	"latent" field	quantity of interest

SDE: $Z = Z_t$, e.g. the driving Brownian (or Levy) process W_t;
$u = u_t$, unknown process (e.g. option price at time t);
Q (non)linear functional of u_t (at end time T or along whole path)

Uncertainty Quantification / Stochastic Simulation

$$Z(\omega) \in X \xrightarrow{\text{Model}} u(\omega) \in V \xrightarrow{\text{Output}} Q(\omega) \in \mathbb{R}(\text{or } \mathbb{R}^J) \xrightarrow{\text{Statistics}} \mathbb{E}[Q]$$

noise process	state of process	quantity of interest
uncertain input	"latent" field	quantity of interest

SDE: $Z = Z_t$, e.g. the driving Brownian (or Levy) process W_t;
$u = u_t$, unknown process (e.g. option price at time t);
Q (non)linear functional of u_t (at end time T or along whole path)

UQ: $Z = Z(x)$ (or $Z(x,t)$), spatial (or spatiotemporal) random field;
$u = u(x)$ (or $u(x,t)$), model state (e.g. PDE solution);
Q (non)linear functional of u (incl. moments or failure probabilities)

$Z(\omega) \in X \xrightarrow{\text{Model}} u(\omega) \in V \xrightarrow{\text{Output}} Q(\omega) \in \mathbb{R}(\text{or } \mathbb{R}^J) \xrightarrow{\text{Statistics}} \mathbb{E}[Q]$

noise process	state of process	quantity of interest
uncertain input	"latent" field	quantity of interest

SDE: $Z = Z_t$, e.g. the driving Brownian (or Levy) process W_t;
$u = u_t$, unknown process (e.g. option price at time t);
Q (non)linear functional of u_t (at end time T or along whole path)

UQ: $Z = Z(x)$ (or $Z(x,t)$), spatial (or spatiotemporal) random field;
$u = u(x)$ (or $u(x,t)$), model state (e.g. PDE solution);
Q (non)linear functional of u (incl. moments or failure probabilities)

Other applications: imaging, biological/chemical reaction networks, interacting particle systems, ...

Uncertainty Quantification / Stochastic Simulation

$$Z(\omega) \in X \xrightarrow{\text{Model}} u(\omega) \in V \xrightarrow{\text{Output}} Q(\omega) \in \mathbb{R}(\text{or } \mathbb{R}^J) \xrightarrow{\text{Statistics}} \mathbb{E}[Q]$$

noise process	state of process	quantity of interest
uncertain input	"latent" field	quantity of interest

SDE: $Z = Z_t$, e.g. the driving Brownian (or Levy) process W_t;
$u = u_t$, unknown process (e.g. option price at time t);
Q (non)linear functional of u_t (at end time T or along whole path)

UQ: $Z = Z(x)$ (or $Z(x,t)$), spatial (or spatiotemporal) random field;
$u = u(x)$ (or $u(x,t)$), model state (e.g. PDE solution);
Q (non)linear functional of u (incl. moments or failure probabilities)

Other applications: imaging, biological/chemical reaction networks, interacting particle systems, ...

Even though $Q(\omega)$ may only be a **single random variable**. Its distribution is often defined only implicitly via the distributions of the **latent** fields $u \in V$ and $Z \in X$ which may be **infinite or high dimensional**!

Practical Implementation: Discretisation / Approximation

$$Z_\ell(\omega) \in X_\ell \xrightarrow{\text{Model}(\ell)} u_\ell(\omega) \in V_\ell \xrightarrow{\text{Output}} Q_\ell(\omega) \in \mathbb{R}(\text{or } \mathbb{R}^J) \xrightarrow{\text{Quadrature}} \widehat{Q}_\ell$$

$$Z_\ell(\omega) \in X_\ell \stackrel{\text{Model}(\ell)}{\longrightarrow} u_\ell(\omega) \in V_\ell \stackrel{\text{Output}}{\longrightarrow} Q_\ell(\omega) \in \mathbb{R} (\text{or } \mathbb{R}^J) \stackrel{\text{Quadrature}}{\longrightarrow} \widehat{Q}_\ell$$

SDE: Discretisation with step size h_ℓ

$Z_\ell = (\Delta W_{\ell,j})_{j=1}^{M_\ell}$ vector of Brownian increments;

$u_\ell = (U_{\ell,j})_{j=1}^{M_\ell}$ vector of states at time $t_j = jh_\ell$.

Practical Implementation: Discretisation / Approximation

$$Z_\ell(\omega) \in X_\ell \stackrel{\text{Model}(\ell)}{\longrightarrow} u_\ell(\omega) \in V_\ell \stackrel{\text{Output}}{\longrightarrow} Q_\ell(\omega) \in \mathbb{R}(\text{or } \mathbb{R}^J) \stackrel{\text{Quadrature}}{\longrightarrow} \widehat{Q}_\ell$$

SDE: Discretisation with step size h_ℓ

$Z_\ell = (\Delta W_{\ell,j})_{j=1}^{M_\ell}$ vector of Brownian increments;

$u_\ell = (U_{\ell,j})_{j=1}^{M_\ell}$ vector of states at time $t_j = jh_\ell$.

UQ: Discretisation with mesh size h_ℓ (here only spatial)

$Z_\ell = (Z_{\ell,j})_{j=1}^{s_\ell}$ coeffs in KL or discrete Fourier expansion of RF Z;

$u_\ell = (U_{\ell,j})_{j=1}^{M_\ell}$ FE coefficient vector associated with FE mesh $\mathcal{T}_\ell$.

Practical Implementation: Discretisation / Approximation

$$Z_\ell(\omega) \in X_\ell \stackrel{\text{Model}(\ell)}{\longrightarrow} u_\ell(\omega) \in V_\ell \stackrel{\text{Output}}{\longrightarrow} Q_\ell(\omega) \in \mathbb{R} (\text{or } \mathbb{R}^J) \stackrel{\text{Quadrature}}{\longrightarrow} \widehat{Q}_\ell$$

SDE: Discretisation with step size h_ℓ

$Z_\ell = (\Delta W_{\ell,j})_{j=1}^{M_\ell}$ vector of Brownian increments;

$u_\ell = (U_{\ell,j})_{j=1}^{M_\ell}$ vector of states at time $t_j = jh_\ell$.

UQ: Discretisation with mesh size h_ℓ (here only spatial)

$Z_\ell = (Z_{\ell,j})_{j=1}^{s_\ell}$ coeffs in KL or discrete Fourier expansion of RF Z;

$u_\ell = (U_{\ell,j})_{j=1}^{M_\ell}$ FE coefficient vector associated with FE mesh $\mathcal{T}_\ell$.

Other approximations: p–**refinement**, **model order reduction**, change of **physical model**, **coarse graining** in particle systems, ...

$$Z_\ell(\omega) \in X_\ell \stackrel{\text{Model}(\ell)}{\longrightarrow} u_\ell(\omega) \in V_\ell \stackrel{\text{Output}}{\longrightarrow} Q_\ell(\omega) \in \mathbb{R}(\text{or } \mathbb{R}^J) \stackrel{\text{Quadrature}}{\longrightarrow} \widehat{Q}_\ell$$

SDE: Discretisation with step size h_ℓ

$Z_\ell = (\Delta W_{\ell,j})_{j=1}^{M_\ell}$ vector of Brownian increments;

$u_\ell = (U_{\ell,j})_{j=1}^{M_\ell}$ vector of states at time $t_j = jh_\ell$.

UQ: Discretisation with mesh size h_ℓ (here only spatial)

$Z_\ell = (Z_{\ell,j})_{j=1}^{s_\ell}$ coeffs in KL or discrete Fourier expansion of RF Z;

$u_\ell = (U_{\ell,j})_{j=1}^{M_\ell}$ FE coefficient vector associated with FE mesh $\mathcal{T}_\ell$.

Other approximations: p–**refinement**, **model order reduction**, change of **physical model**, **coarse graining** in particle systems, ...

This provides us with a natural **model hierarchy** (parametrised by ℓ). Assume, there exist $\alpha > 0$ and $\gamma > 0$ such that

(A1) $|\mathbb{E}[Q - Q_\ell]| = \mathcal{O}(2^{-\alpha\ell})$ and **(A2)** $\mathbb{E}[\text{Cost}_\ell] = \mathcal{O}(2^{\gamma\ell})$

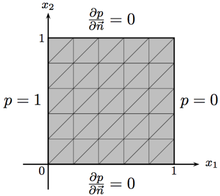
where Cost_ℓ is the cost to compute one realisation of Q_ℓ.

Running Example ("Fruit fly" of UQ)

- Single-phase subsurface flow on unit square D:

$$-\nabla \cdot \left(e^{Z(\omega)} \nabla u(\omega) \right) = 0$$

subject to Neumann BC $\nabla u \cdot \nu = 0$ (top & bottom)
& Dirichlet BC $u = 1$ (left) and $u = 0$ (right)

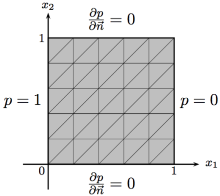


Running Example ("Fruit fly" of UQ)

- Single-phase subsurface flow on unit square D:

$$-\nabla \cdot \left(e^{Z(\omega)} \nabla u(\omega) \right) = 0$$

subject to Neumann BC $\nabla u \cdot \nu = 0$ (top & bottom)
& Dirichlet BC $u = 1$ (left) and $u = 0$ (right)



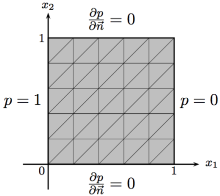
- $u \in V := H^1_\Gamma(D)$ & $Z \in X := L^\infty(D)$ (∞-dim'l function spaces)

Running Example ("Fruit fly" of UQ)

- Single-phase subsurface flow on unit square D:

$$-\nabla \cdot \left(e^{Z(\omega)} \nabla u(\omega) \right) = 0$$

subject to Neumann BC $\nabla u \cdot \nu = 0$ (top & bottom)
& Dirichlet BC $u = 1$ (left) and $u = 0$ (right)



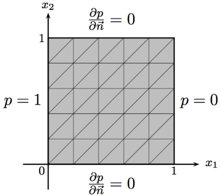
- $u \in V := H^1_\Gamma(D)$ & $Z \in X := L^\infty(D)$ (∞-dim'l function spaces)
- Gaussian $Z(\omega)$ w. **exponential** covariance (i.e. $e^{Z(\omega)}$ lognormal)
- Parametrised by $Z_\ell(\omega) \in X_\ell := \mathbb{R}^{s_\ell}$ with $Z_{\ell,j} \sim \mathcal{N}(0, 1)$ i.i.d.
  (e.g. via truncated KL-expansion of Z or via circulant embedding & FFT)

Running Example ("Fruit fly" of UQ)

- Single-phase subsurface flow on unit square D:

$$-\nabla \cdot \left(e^{Z(\omega)} \nabla u(\omega)\right) = 0$$

subject to Neumann BC $\nabla u \cdot \nu = 0$ (top & bottom)
& Dirichlet BC $u = 1$ (left) and $u = 0$ (right)

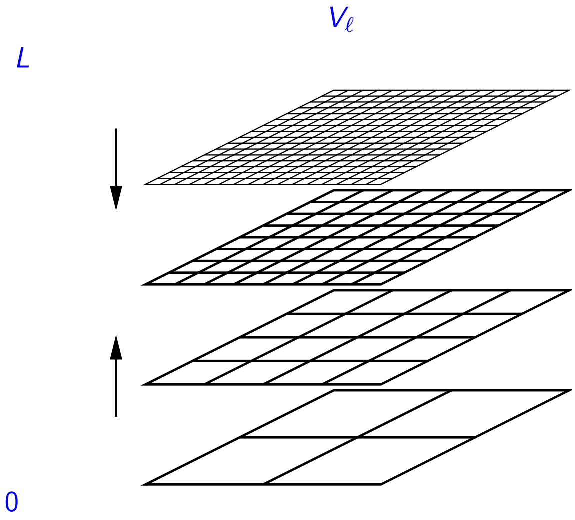


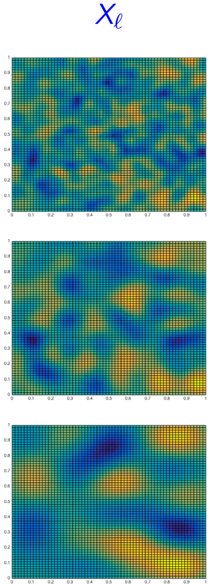
- $u \in V := H^1_\Gamma(D)$ & $Z \in X := L^\infty(D)$ (∞-dim'l function spaces)
- Gaussian $Z(\omega)$ w. **exponential** covariance (i.e. $e^{Z(\omega)}$ lognormal)
- Parametrised by $Z_\ell(\omega) \in X_\ell := \mathbb{R}^{s_\ell}$ with $Z_{\ell,j} \sim \mathcal{N}(0,1)$ i.i.d.
  (e.g. via truncated KL-expansion of Z or via circulant embedding & FFT)
- FE discretisation: $u_\ell \in V_\ell \subset V$
  (e.g. continuous p.w. linears w.r.t. $\mathcal{T}_\ell$ with mesh size $h_\ell = 2^{-\ell+1/2}$):

$$\int_D \nabla v_\ell \cdot \left(e^{Z_\ell(\omega)} \nabla u_\ell(\omega)\right) = 0 \quad \forall v_\ell \in V_\ell \quad \Leftrightarrow \quad A_\ell(\omega) U_\ell(\omega) = b$$

$M_\ell \times M_\ell$ random lin. sys.

Running Example – Model Hierarchy





Here $\alpha \approx 1$ (smooth functionals) and $\gamma \approx 2$ (with AMG)

Monte Carlo

- The **standard Monte Carlo** estimator for $\mathbb{E}[Q]$ is

$$\hat{Q}_L^{\mathrm{MC}} := \frac{1}{N}\sum_{i=1}^{N} Q_L^{(i)}, \quad Q_L^{(i)} \text{ i.i.d. samples with } \mathsf{Model}(L)$$

Monte Carlo

- The **standard Monte Carlo** estimator for $\mathbb{E}[Q]$ is

$$\hat{Q}_L^{\mathrm{MC}} := \frac{1}{N}\sum_{i=1}^{N} Q_L^{(i)}, \quad Q_L^{(i)} \text{ i.i.d. samples with } \mathrm{Model}(L)$$

- Convergence of plain vanilla MC **(mean square error)**:

$$\underbrace{\mathbb{E}\Big[\big(\hat{Q}_L^{\mathrm{MC}} - \mathbb{E}[Q]\big)^2\Big]}_{=:\,\mathrm{MSE}} = \underbrace{\frac{\mathbb{V}[Q_L]}{N}}_{\textbf{sampling error}} + \underbrace{\Big(\mathbb{E}[Q_L - Q]\Big)^2}_{\textbf{model error ("bias")}}$$

Monte Carlo

- The **standard Monte Carlo** estimator for $\mathbb{E}[Q]$ is

$$\hat{Q}_L^{\mathrm{MC}} := \frac{1}{N}\sum_{i=1}^{N} Q_L^{(i)}, \quad Q_L^{(i)} \text{ i.i.d. samples with } \mathrm{Model}(L)$$

- Convergence of plain vanilla MC **(mean square error)**:

$$\underbrace{\mathbb{E}\Big[\big(\hat{Q}_L^{\mathrm{MC}} - \mathbb{E}[Q]\big)^2\Big]}_{=:\ \mathrm{MSE}} = \underbrace{\frac{\mathbb{V}[Q_L]}{N}}_{\textbf{sampling error}} + \underbrace{\Big(\mathbb{E}[Q_L - Q]\Big)^2}_{\textbf{model error (“bias”)}}$$

- Recall $|\mathbb{E}[Q_\ell - Q]| = \mathcal{O}(2^{-\alpha\ell})$ and $\mathbb{E}\left[\mathrm{Cost}_\ell\right] = \mathcal{O}(2^{\gamma\ell})$.
- To get $\mathrm{MSE} = \mathcal{O}(\varepsilon^2)$, we need $L \sim \log_2(\varepsilon^{-1})\alpha^{-1}$ & $N \sim \varepsilon^{-2}$

- The **standard Monte Carlo** estimator for $\mathbb{E}[Q]$ is

$$\hat{Q}_L^{\mathrm{MC}} := \frac{1}{N}\sum_{i=1}^{N} Q_L^{(i)}, \quad Q_L^{(i)} \text{ i.i.d. samples with Model}(L)$$

- Convergence of plain vanilla MC **(mean square error)**:

$$\underbrace{\mathbb{E}\left[\left(\hat{Q}_L^{\mathrm{MC}} - \mathbb{E}[Q]\right)^2\right]}_{=:\ \mathrm{MSE}} = \underbrace{\frac{\mathbb{V}[Q_L]}{N}}_{\textbf{sampling error}} + \underbrace{\left(\mathbb{E}[Q_L - Q]\right)^2}_{\textbf{model error ("bias")}}$$

- Recall $|\mathbb{E}[Q_\ell - Q]| = \mathcal{O}(2^{-\alpha\ell})$ and $\mathbb{E}\left[\mathrm{Cost}_\ell\right] = \mathcal{O}(2^{\gamma\ell})$.
- To get $\mathrm{MSE} = \mathcal{O}(\varepsilon^2)$, we need $L \sim \log_2(\varepsilon^{-1})\alpha^{-1}$ & $N \sim \varepsilon^{-2}$

Complexity Theorem for (plain vanilla) Monte Carlo

$\mathrm{Cost}(\hat{Q}_L^{\mathrm{MC}}) = \mathcal{O}(NM_L) = \mathcal{O}\left(\varepsilon^{-2-\gamma/\alpha}\right)$ to obtain $\mathrm{MSE} = \mathcal{O}(\varepsilon^2)$.

- The **standard Monte Carlo** estimator for $\mathbb{E}[Q]$ is

$$\hat{Q}_L^{\mathrm{MC}} := \frac{1}{N}\sum_{i=1}^{N} Q_L^{(i)}, \quad Q_L^{(i)} \text{ i.i.d. samples with Model}(L)$$

- Convergence of plain vanilla MC **(mean square error)**:

$$\underbrace{\mathbb{E}\left[\left(\hat{Q}_L^{\mathrm{MC}} - \mathbb{E}[Q]\right)^2\right]}_{=: \, \mathrm{MSE}} = \underbrace{\frac{\mathbb{V}[Q_L]}{N}}_{\textbf{sampling error}} + \underbrace{\left(\mathbb{E}[Q_L - Q]\right)^2}_{\textbf{model error ("bias")}}$$

- Recall $|\mathbb{E}[Q_\ell - Q]| = \mathcal{O}(2^{-\alpha\ell})$ and $\mathbb{E}\left[\mathrm{Cost}_\ell\right] = \mathcal{O}(2^{\gamma\ell})$.
- To get $\mathrm{MSE} = \mathcal{O}(\varepsilon^2)$, we need $L \sim \log_2(\varepsilon^{-1})\alpha^{-1}$ & $N \sim \varepsilon^{-2}$

Complexity Theorem for (plain vanilla) Monte Carlo

$\mathrm{Cost}(\hat{Q}_L^{\mathrm{MC}}) = \mathcal{O}(NM_L) = \mathcal{O}\left(\varepsilon^{-2-\gamma/\alpha}\right)$ to obtain $\mathrm{MSE} = \mathcal{O}(\varepsilon^2)$.

For **fruit fly**: $\mathrm{Cost}(\hat{Q}_L^{\mathrm{MC}}) \approx \mathcal{O}(\varepsilon^{-4})$ **Prohibitively expensive!**

Basic Idea: Note that trivially

$$Q_L = Q_0 + \sum_{\ell=1}^{L} Q_\ell - Q_{\ell-1}$$

Basic Idea: Note that trivially (due to linearity of $\mathbb{E}$)

$$\mathbb{E}[Q_L] = \mathbb{E}[Q_0] \; + \; \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell - Q_{\ell-1}]$$

Basic Idea: Note that trivially (due to linearity of $\mathbb{E}$)

$$\mathbb{E}[Q_L] = \mathbb{E}[Q_0] \; + \; \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell - Q_{\ell-1}]$$

Define the following **multilevel MC** estimator for $\mathbb{E}[Q]$:

$$\widehat{Q}_L^{MLMC} := \hat{Q}_0^{\mathrm{MC}} \; + \; \sum_{\ell=1}^{L} \hat{Y}_\ell^{\mathrm{MC}} \quad \text{where} \quad Y_\ell := Q_\ell - Q_{\ell-1}$$

Basic Idea: Note that trivially (due to linearity of $\mathbb{E}$)

$$\mathbb{E}[Q_L] = \mathbb{E}[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell - Q_{\ell-1}]$$

Define the following **multilevel MC** estimator for $\mathbb{E}[Q]$:

$$\widehat{Q}_L^{MLMC} := \hat{Q}_0^{\mathrm{MC}} + \sum_{\ell=1}^{L} \hat{Y}_\ell^{\mathrm{MC}} \quad \text{where} \quad Y_\ell := Q_\ell - Q_{\ell-1}$$

Key Observation: (Variance Reduction! Corrections cheaper!)

Level L: $\mathbb{V}[Q_L - Q_{L-1}] \to 0$ as $L \to \infty$ $\Rightarrow$ $N_L = \mathcal{O}(1)$ (best case)

Basic Idea: Note that trivially (due to linearity of $\mathbb{E}$)

$$\mathbb{E}[Q_L] = \mathbb{E}[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell - Q_{\ell-1}]$$

Define the following **multilevel MC** estimator for $\mathbb{E}[Q]$:

$$\widehat{Q}_L^{MLMC} := \hat{Q}_0^{\mathrm{MC}} + \sum_{\ell=1}^{L} \hat{Y}_\ell^{\mathrm{MC}} \quad \text{where} \quad Y_\ell := Q_\ell - Q_{\ell-1}$$

Key Observation: (Variance Reduction! Corrections cheaper!)

Level L: $\mathbb{V}[Q_L - Q_{L-1}] \to 0$ as $L \to \infty$ $\Rightarrow$ $N_L = \mathcal{O}(1)$ (best case)

Level 0: $N_0 \sim N$ **but** $\mathrm{Cost}_0 = \mathcal{O}(M_0) = \mathcal{O}(1)$

Multilevel Monte Carlo [Heinrich, '98], [Giles, '07]

Basic Idea: Note that trivially (due to linearity of $\mathbb{E}$)

$$\mathbb{E}[Q_L] = \mathbb{E}[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell - Q_{\ell-1}]$$

Define the following **multilevel MC** estimator for $\mathbb{E}[Q]$:

$$\widehat{Q}_L^{MLMC} := \hat{Q}_0^{\mathrm{MC}} + \sum_{\ell=1}^{L} \hat{Y}_\ell^{\mathrm{MC}} \quad \text{where} \quad Y_\ell := Q_\ell - Q_{\ell-1}$$

Key Observation: (Variance Reduction! Corrections cheaper!)

Level L: $\mathbb{V}[Q_L - Q_{L-1}] \to 0$ as $L \to \infty \Rightarrow N_L = \mathcal{O}(1)$ (best case)

$\vdots$

Level ℓ: N_ℓ optimised to "balance" cost with levels 0 and L

$\vdots$

Level 0: $N_0 \sim N$ **but** $\mathrm{Cost}_0 = \mathcal{O}(M_0) = \mathcal{O}(1)$

Complexity Theorem [Giles, '07], [Cliffe, Giles, RS, Teckentrup, '11]

Assume approximation error $\mathcal{O}(2^{-\alpha\ell})$, Cost/sample $\mathcal{O}(2^{\gamma\ell})$ **and**

$$\mathbb{V}[Q_\ell - Q_{\ell-1}] = \mathcal{O}(2^{-\beta\ell}) \qquad \text{(strong error/variance reduction)}$$

Then there exist L, $\{N_\ell\}_{\ell=0}^{L}$ to obtain $\text{MSE} = \mathcal{O}(\varepsilon^2)$ with

$$\text{Cost}(\widehat{Q}_L^{MLMC}) = \mathcal{O}\left(\varepsilon^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)}\right) + \text{possible log-factor}$$

using **dependent** or **independent** estimators $\hat{Q}_0^{\text{MC}}$, and $\left(\hat{Y}_\ell^{\text{MC}}\right)_{\ell=1}^{L}$.

Complexity Theorem [Giles, '07], [Cliffe, Giles, RS, Teckentrup, '11]

Assume approximation error $\mathcal{O}(2^{-\alpha\ell})$, Cost/sample $\mathcal{O}(2^{\gamma\ell})$ **and**

$$\mathbb{V}[Q_\ell - Q_{\ell-1}] = \mathcal{O}(2^{-\beta\ell}) \qquad \text{(strong error/variance reduction)}$$

Then there exist L, $\{N_\ell\}_{\ell=0}^{L}$ to obtain $\mathrm{MSE} = \mathcal{O}(\varepsilon^2)$ with

$$\mathrm{Cost}(\widehat{Q}_L^{MLMC}) = \mathcal{O}\left(\varepsilon^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)}\right) + \text{possible log-factor}$$

using **dependent** or **independent** estimators $\hat{Q}_0^{\mathrm{MC}}$, and $\left(\hat{Y}_\ell^{\mathrm{MC}}\right)_{\ell=1}^{L}$.

Fruit fly (with smooth functionals & AMG): $\alpha \approx 1$, $\beta \approx 2$, $\gamma \approx 2$

$$\mathrm{Cost}(\widehat{Q}_L^{MLMC}) = \mathcal{O}\left(\varepsilon^{-\max\left(2,\frac{\gamma}{\alpha}\right)}\right) = \mathcal{O}\left(\max(N_0, M_L)\right) \approx \mathcal{O}(\varepsilon^{-2})$$

Complexity Theorem [Giles, '07], [Cliffe, Giles, RS, Teckentrup, '11]

Assume approximation error $\mathcal{O}(2^{-\alpha\ell})$, Cost/sample $\mathcal{O}(2^{\gamma\ell})$ **and**

$$\mathbb{V}[Q_\ell - Q_{\ell-1}] = \mathcal{O}(2^{-\beta\ell}) \qquad \text{(strong error/variance reduction)}$$

Then there exist L, $\{N_\ell\}_{\ell=0}^{L}$ to obtain $\mathrm{MSE} = \mathcal{O}(\varepsilon^2)$ with

$$\mathrm{Cost}(\widehat{Q}_L^{MLMC}) = \mathcal{O}\left(\varepsilon^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)}\right) \quad + \text{ possible log-factor}$$

using **dependent** or **independent** estimators $\hat{Q}_0^{\mathrm{MC}}$, and $\left(\hat{Y}_\ell^{\mathrm{MC}}\right)_{\ell=1}^{L}$.

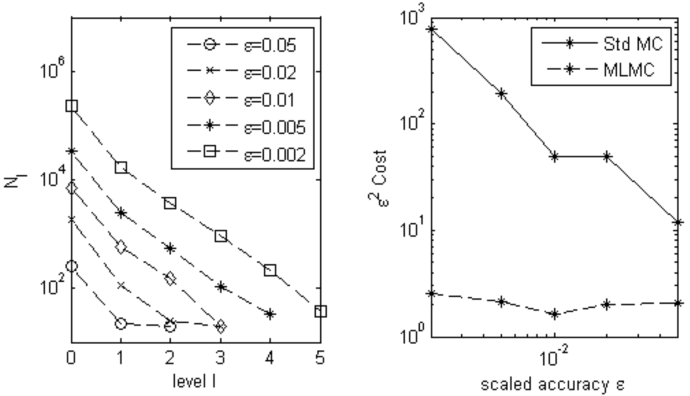
Fruit fly (with smooth functionals & AMG): $\alpha \approx 1$, $\beta \approx 2$, $\gamma \approx 2$

$$\mathrm{Cost}(\widehat{Q}_L^{MLMC}) = \mathcal{O}\left(\varepsilon^{-\max\left(2,\frac{\gamma}{\alpha}\right)}\right) = \mathcal{O}\left(\max(N_0, M_L)\right) \approx \mathcal{O}(\varepsilon^{-2})$$

Optimality: Asymptotic cost of **one** deterministic solve (to tol$=\varepsilon$) **!**

Numerical Example (Multilevel MC)

Fruit fly with $Q = \|u\|_{L_2(D)}$ & circulant embedding with $s_\ell = \mathcal{O}(M_\ell)$



$$\sigma^2 = 1, \quad \lambda = 0.3, \quad h_0 = \frac{1}{8}$$

Observations, Extensions and Applications

- **Gains** even for small number of levels **(see below).**
- Substantial **practical gains** (not only asymptotic as $\varepsilon \to 0$)
- Models do **not** have to be **nested** (could even couple FE & MD)

Observations, Extensions and Applications

- **Gains** even for small number of levels **(see below).**
- Substantial **practical gains** (not only asymptotic as $\varepsilon \to 0$)
- Models do **not** have to be **nested** (could even couple FE & MD)
- **Other approximations:** multiscale methods, model order reduction, smoothing, homogenisation, coarse graining, . . .
- **Different quadrature**: ML Quasi-MC, ML Stoch. Collocation,. . .

Observations, Extensions and Applications

- **Gains** even for small number of levels **(see below).**
- Substantial **practical gains** (not only asymptotic as $\varepsilon \to 0$)
- Models do **not** have to be **nested** (could even couple FE & MD)
- **Other approximations:** multiscale methods, model order reduction, smoothing, homogenisation, coarse graining, ...
- **Different quadrature**: ML Quasi-MC, ML Stoch. Collocation,...
- **Not restricted** to differential equations:
  - continuous time Markov chains, biological/chemical reaction networks, kinetic MC, ...
  - interacting particle syst. (**coarse graining**), nested simulation
  - Boltzmann/neutron transport (**integrodifferential equation**)

Observations, Extensions and Applications

- **Gains** even for small number of levels **(see below).**
- Substantial **practical gains** (not only asymptotic as $\varepsilon \to 0$)
- Models do **not** have to be **nested** (could even couple FE & MD)
- **Other approximations:** multiscale methods, model order reduction, smoothing, homogenisation, coarse graining, ...
- **Different quadrature**: ML Quasi-MC, ML Stoch. Collocation,...
- **Not restricted** to differential equations:
  - continuous time Markov chains, biological/chemical reaction networks, kinetic MC, ...
  - interacting particle syst. (**coarse graining**), nested simulation
  - Boltzmann/neutron transport (**integrodifferential equation**)

Strong (sample-wise) coupling is key: $\mathbb{V}[Q_\ell - Q_{\ell-1}] \ll \mathbb{V}[Q_\ell]$

Not always easy!!

Refs.: https://people.maths.ox.ac.uk/gilesm/mlmc_community.html

Not just theory & Not just for the Fruit Fly
(for simplicity consider only two levels)

Model 1		**Model 0**
$\downarrow$		$\downarrow$
$Q_1^{(i)}$	$\overset{\text{strong coupling}}{\longleftrightarrow}$	$Q_0^{(i)}$

Model 1		Model 0
$\downarrow$		$\downarrow$
$Q_1^{(i)}$	strong coupling $\longleftrightarrow$	$Q_0^{(i)}$

Assume:

1. $\mathbb{E}[\mathrm{Cost}_0] = X\,\mathbb{E}[\mathrm{Cost}_1]$, for some $X < 1$
2. $\mathbb{V}[Q_0] \approx \mathbb{V}[Q_1]$ & $\mathbb{V}[Q_1 - Q_0] = Y^2\,\mathbb{V}[Q_0]$, for some $Y < 1$

Not just theory & Not just for the Fruit Fly
(for simplicity consider only two levels)

$$\boxed{\text{Model 1}} \qquad\qquad\qquad\qquad \boxed{\text{Model 0}}$$

$$\downarrow \qquad\qquad\qquad\qquad\qquad\qquad \downarrow$$

$$Q_1^{(i)} \quad \overset{\text{strong coupling}}{\longleftrightarrow} \quad Q_0^{(i)}$$

Assume:

1. $\mathbb{E}[\mathrm{Cost}_0] = X\,\mathbb{E}[\mathrm{Cost}_1]$, for some $X < 1$
2. $\mathbb{V}[Q_0] \approx \mathbb{V}[Q_1]$ & $\mathbb{V}[Q_1 - Q_0] = Y^2\,\mathbb{V}[Q_0]$, for some $Y < 1$

$$\textbf{Gain} \;=\; \frac{\mathrm{Cost}(\widehat{Q}_1^{\mathrm{MC}})}{\mathrm{Cost}(\widehat{Q}_1^{MLMC})} \;=\; \frac{1}{X + Y^2(1+X)}$$

Not just theory & Not just for the Fruit Fly
(for simplicity consider only two levels)

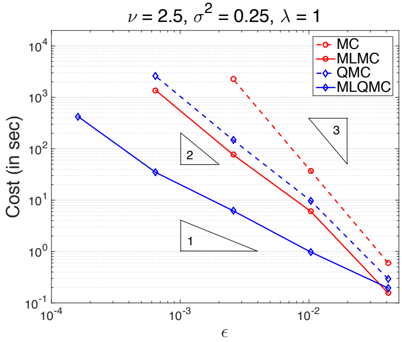
Model 1 $\downarrow$ $Q_1^{(i)}$ $\overset{\text{strong coupling}}{\longleftrightarrow}$ $Q_0^{(i)}$ $\downarrow$ **Model 0**

Assume:

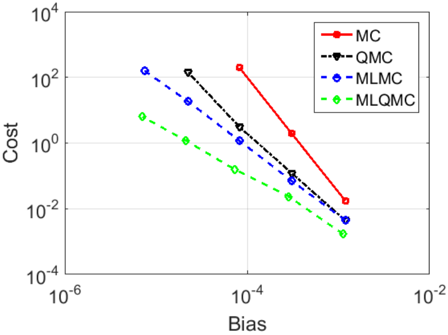
1. $\mathbb{E}[\mathrm{Cost}_0] = X\,\mathbb{E}[\mathrm{Cost}_1]$, for some $X < 1$
2. $\mathbb{V}[Q_0] \approx \mathbb{V}[Q_1]$ & $\mathbb{V}[Q_1 - Q_0] = Y^2\,\mathbb{V}[Q_0]$, for some $Y < 1$

Examples / Gains	X	$Y = 0.5$	$Y = 0.1$	$Y = 0.05$
2D elliptic ($h_0 = 2h_1$)	1/4	1.8	3.8	4.0
3D elliptic ($h_0 = 2h_1$)	1/8	2.4	7.1	8.0
3D parab. ($h_0 = 2h_1$)	1/16	3.0	13.7	15.9
3D elliptic ($h_0 = 4h_1$)	1/64	3.7	38.8	62.4

Even higher gains with multiple levels!



Fruit fly (with Matern covariance)
[Kuo, RS, Schwab, Sloan, Ullmann, '17]



Neutron transport (Boltzmann)
[Graham, Parkinson, RS, '17(pre)]

Extension to Bayesian inference / Data assimilation

$$Z(\omega) \in X \xrightarrow{\text{Model}} u(\omega) \in V \xrightarrow{\text{Output}} Q(\omega) \in \mathbb{R}(\text{or } \mathbb{R}^J) \xrightarrow{\text{Statistics}} \mathbb{E}_\pi[Q]$$

conditioned on data $\mathbf{y}^{\text{obs}}$ — posterior expectation

Extension to Bayesian inference / Data assimilation

$$Z(\omega) \in X \xrightarrow{\text{Model}} u(\omega) \in V \xrightarrow{\text{Output}} Q(\omega) \in \mathbb{R}(\text{or } \mathbb{R}^J) \xrightarrow{\text{Statistics}} \mathbb{E}_\pi[Q]$$

conditioned on data $\mathbf{y}^{\text{obs}}$ posterior expectation

- **Multilevel Markov Chain Monte Carlo**
  [Hoang, Schwab, Stuart 13], **[Dodwell, Ketelsen, RS, Teckentrup, 15]**

Extension to Bayesian inference / Data assimilation

$$Z(\omega) \in X \xrightarrow{\text{Model}} u(\omega) \in V \xrightarrow{\text{Output}} Q(\omega) \in \mathbb{R}(\text{or } \mathbb{R}^J) \xrightarrow{\text{Statistics}} \mathbb{E}_\pi[Q]$$

conditioned on data $\mathbf{y}^{\text{obs}}$ posterior expectation

- **Multilevel Markov Chain Monte Carlo**
  [Hoang, Schwab, Stuart 13], **[Dodwell, Ketelsen, RS, Teckentrup, 15]**
- **Multilevel Sequential Monte Carlo**
  [Beskos, Jasra, Law, Tempone, Zhou, 17], [Del Moral, Jasra, Law, 17]
- **Multilevel Filtering**
  [Jasra, Kamatani, Law, Zhou, 15(pre)], [Gregory, Cotter, Reich, 16], [Gregory, Cotter 16(pre)]
- **Multilevel Ensemble Kalman Filter**
  [Hoel, Law, Tempone, 15], [Chernov, Hoel, Law, Nobile, Temp., 16(pre)]

Extension to Bayesian inference / Data assimilation

$Z(\omega) \in X \xrightarrow{\text{Model}} u(\omega) \in V \xrightarrow{\text{Output}} Q(\omega) \in \mathbb{R}(\text{or } \mathbb{R}^J) \xrightarrow{\text{Statistics}} \mathbb{E}_\pi[Q]$

conditioned on data $\mathbf{y}^{\text{obs}}$ — posterior expectation

- **Multilevel Markov Chain Monte Carlo**
  [Hoang, Schwab, Stuart 13], **[Dodwell, Ketelsen, RS, Teckentrup, 15]**
- **Multilevel Sequential Monte Carlo**
  [Beskos, Jasra, Law, Tempone, Zhou, 17], [Del Moral, Jasra, Law, 17]
- **Multilevel Filtering**
  [Jasra, Kamatani, Law, Zhou, 15(pre)], [Gregory, Cotter, Reich, 16], [Gregory, Cotter 16(pre)]
- **Multilevel Ensemble Kalman Filter**
  [Hoel, Law, Tempone, 15], [Chernov, Hoel, Law, Nobile, Temp., 16(pre)]

Similar gains possible! More difficult to achieve **both** consistency (collapsing sum) + variance reduction (strong coupling).

Multilevel Markov Chain Monte Carlo – Idea
Dodwell, Ketelsen, RS, Teckentrup, JUQ 2015

Posterior distribution (Bayes):

$$\pi^\ell(\mathbf{Z}_\ell|\mathbf{y}^{\rm obs}) \eqsim \exp(-\|\mathbf{y}^{\rm obs} - F_\ell(\mathbf{Z}_\ell)\|^2_{\Sigma_{\mathbf{obs}}})\,\pi_{\rm prior}(\mathbf{Z}_\ell)$$

What were the **key ingredients** of "standard" multilevel Monte Carlo?

Posterior distribution (Bayes):

$$\pi^\ell(\mathbf{Z}_\ell|\mathbf{y}^{\mathrm{obs}}) \eqsim \exp(-\|\mathbf{y}^{\mathrm{obs}} - F_\ell(\mathbf{Z}_\ell)\|^2_{\Sigma_{\mathbf{obs}}})\, \pi_{\mathrm{prior}}(\mathbf{Z}_\ell)$$

What were the **key ingredients** of "standard" multilevel Monte Carlo?

- **Telescoping sum:** $\mathbb{E}[\mathcal{Q}_L] = \mathbb{E}[\mathcal{Q}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\mathcal{Q}_\ell - \mathcal{Q}_{\ell-1}]$
- Models on coarser levels **much cheaper** to solve ($M_0 \ll M_L$).
- $\mathbb{V}[\mathcal{Q}_\ell - \mathcal{Q}_{\ell-1}] \stackrel{\ell\to\infty}{\longrightarrow} 0$ as $\implies$ much **fewer samples** on finer levels.

Multilevel Markov Chain Monte Carlo – Idea
Dodwell, Ketelsen, RS, Teckentrup, JUQ 2015

Posterior distribution (Bayes):

$$\pi^\ell(\mathbf{Z}_\ell|\mathbf{y}^{\text{obs}}) \;\eqsim\; \exp(-\|\mathbf{y}^{\text{obs}} - F_\ell(\mathbf{Z}_\ell)\|^2_{\Sigma_{\text{obs}}})\,\pi_{\text{prior}}(\mathbf{Z}_\ell)$$

What were the **key ingredients** of "standard" multilevel Monte Carlo?

- **Telescoping sum:** $\mathbb{E}[\mathcal{Q}_L] = \mathbb{E}[\mathcal{Q}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\mathcal{Q}_\ell - \mathcal{Q}_{\ell-1}]$
- Models on coarser levels **much cheaper** to solve ($M_0 \ll M_L$).
- $\mathbb{V}[\mathcal{Q}_\ell - \mathcal{Q}_{\ell-1}] \stackrel{\ell\to\infty}{\longrightarrow} 0$ as $\implies$ much **fewer samples** on finer levels.

But Important! In MCMC the target distribution π^ℓ **depends on** ℓ:

$$\mathbb{E}_{\pi^L}[\mathcal{Q}_L] = \mathbb{E}_{\pi^0}[\mathcal{Q}_0] + \sum_\ell \mathbb{E}_{\pi^\ell}[\mathcal{Q}_\ell] - \mathbb{E}_{\pi^{\ell-1}}[\mathcal{Q}_{\ell-1}]$$

Posterior distribution (Bayes):

$$\pi^\ell(\mathbf{Z}_\ell|\mathbf{y}^{\mathrm{obs}}) \eqsim \exp(-\|\mathbf{y}^{\mathrm{obs}} - F_\ell(\mathbf{Z}_\ell)\|^2_{\Sigma_{\mathbf{obs}}})\, \pi_{\mathrm{prior}}(\mathbf{Z}_\ell)$$

What were the **key ingredients** of "standard" multilevel Monte Carlo?

- **Telescoping sum:** $\mathbb{E}[\mathcal{Q}_L] = \mathbb{E}[\mathcal{Q}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\mathcal{Q}_\ell - \mathcal{Q}_{\ell-1}]$
- Models on coarser levels **much cheaper** to solve ($M_0 \ll M_L$).
- $\mathbb{V}[\mathcal{Q}_\ell - \mathcal{Q}_{\ell-1}] \stackrel{\ell\to\infty}{\longrightarrow} 0$ as $\implies$ much **fewer samples** on finer levels.

But Important! In MCMC the target distribution π^ℓ **depends on** ℓ:

$$\mathbb{E}_{\pi^L}[\mathcal{Q}_L] = \underbrace{\mathbb{E}_{\pi^0}[\mathcal{Q}_0]}_{\text{standard MCMC}} + \sum_\ell \underbrace{\mathbb{E}_{\pi^\ell}[\mathcal{Q}_\ell] - \mathbb{E}_{\pi^{\ell-1}}[\mathcal{Q}_{\ell-1}]}_{\text{multilevel MCMC (NEW)}}$$

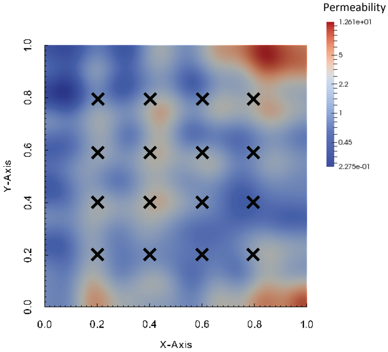
$$\widehat{Q}^{\mathbf{MLMetH}}_{h,s} := \frac{1}{N_0}\sum_{n=1}^{N_0} \mathcal{Q}_0(\mathbf{Z}^n_{0,0}) + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} \left(\mathcal{Q}_\ell(\mathbf{Z}^n_{\ell,\ell}) - \mathcal{Q}_{\ell-1}(\mathbf{Z}^n_{\ell,\ell-1})\right)$$

chains strongly coupled!

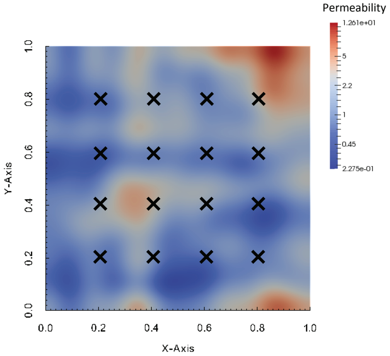
- **Prior:** Separable exponential covariance with $\sigma^2 = 1$, $\lambda = 0.5$.

Fruit fly (i.e. 2D lognormal diffusion on $D = (0,1)^2$ with linear FEs)

- **Prior:** Separable exponential covariance with $\sigma^2 = 1$, $\lambda = 0.5$.
- **"Data"** $\mathbf{y}^{\text{obs}}$: Pressure at 16 points $x_j^* \in D$ and $\Sigma^{\text{obs}} = 10^{-4} I$.



Data

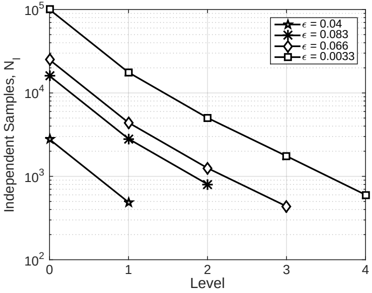


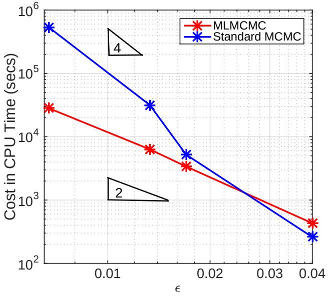
Posterior Sample

MLMCMC – Numerical Example

Fruit fly (i.e. 2D lognormal diffusion on $D = (0,1)^2$ with linear FEs)

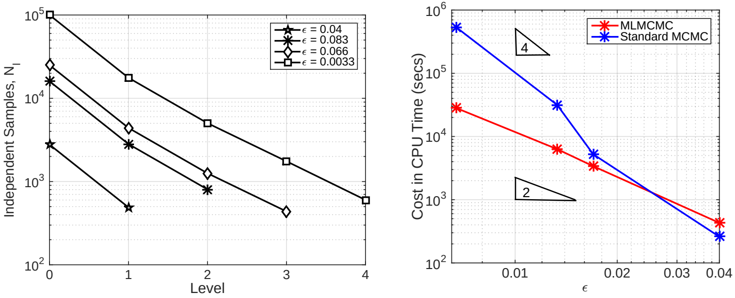
- 5-level method w. #KL modes increasing from $s_0 = 50$ to $s_4 = 150$





Fruit fly (i.e. 2D lognormal diffusion on $D = (0,1)^2$ with linear FEs)

- 5-level method w. #KL modes increasing from $s_0 = 50$ to $s_4 = 150$



- #independent samples $= \dfrac{N_\ell}{t_\ell}$ (w. t_l … integrated autocorrelation time)

Level ℓ	0	1	2	3	4
a.c. time t_ℓ	136.23	3.66	2.93	1.46	1.23

Overall Summary

- **Huge potential** for multilevel Monte Carlo & model hierarchies (in general) in stochastic simulation and in UQ
- A vibrant research area with **many open questions**
- A **"no-brainer"** in practice (if you have a model hierarchy)
- Many **new application areas** await exploration
- I believe, we have only **scratched the surface**, especially in context of Bayesian inference & data assimilation
- Significant further improvements are possible with using adaptive, sample-dependent hierarchies (current work!)

Overall Summary

- **Huge potential** for multilevel Monte Carlo & model hierarchies (in general) in stochastic simulation and in UQ
- A vibrant research area with **many open questions**
- A "**no-brainer**" in practice (if you have a model hierarchy)
- Many **new application areas** await exploration
- I believe, we have only **scratched the surface**, especially in context of Bayesian inference & data assimilation
- Significant further improvements are possible with using adaptive, sample-dependent hierarchies (current work!)

Thank You !