

# Multilevel Monte Carlo for Stochastic PDEs



Inti-Raymi Carhuanchu Mantripp  
Mansfield College  
University of Oxford

A dissertation submitted for the degree of  
*MSc in Mathematical Modelling & Scientific Computing*  
Trinity 2025

# Contents

# List of Figures

# Chapter 1

## Introduction

The purpose of this dissertation is to investigate the Multilevel Monte Carlo (MLMC) method as a technique for reducing the computational cost in estimating expectations of random variables for two classes of parabolic stochastic partial differential equations (SPDEs).

This is motivated by the slow convergence of the Monte Carlo (MC) estimator. Its root mean square error decays  $O(N^{-1/2})$  where  $N$  is the number of independent samples obtained, and therefore achieving a high accuracy estimate requires a high number of samples. As most SPDEs do not admit analytic solutions, the main approach available for estimating expectations of functionals of SPDEs is Monte Carlo methods [?]. In the context of SPDEs, obtaining samples typically requires discretising the spatial domain via finite difference or finite element methods and then evolving the system to obtain the desired quantity. Accurate estimates require fine discretisations, and as such, the cost of obtaining a sample can become expensive. This, combined with the slow convergence of the standard MC estimator, can make the standard Monte Carlo method prohibitively expensive.

A number of techniques exist that aim to improve on the convergence of the MC estimator. Their goal is to achieve an equivalent mean squared error in their estimate at a reduced computational cost. The MLMC method is one such method. At its heart is an approach of computing a blend of cheaper and more expensive samples, unlike the MC estimator which uses only samples of the same cost. Many inexpensive samples provide the bulk of the statistical signal, while comparatively few more costly samples then correct any residual bias. In the context of SPDEs, this means obtaining samples using meshes at varying *levels* of discretisation. By taking many samples at cheaper, coarser levels and fewer at finer, more expensive levels, the MLMC estimator aims to achieve an estimate of equivalent accuracy at reduced cost compared to the MC estimator.

The practical efficiency of the MLMC method, however, relies on correlating samples obtained at different levels, in a manner analogous to that of the control variates method ([?], Section 1.2). This is because the cost savings achievable are dependent on the extent to which samples across two adjacent levels of accuracy covary. Practically, this entails aligning the randomness used to generate pairs of coarse and fine samples, a process referred to as *coupling*. To provide intuition as to what this means, Figure ?? illustrates what the finite difference grid used by two adjacent estimators might look like. At each node in these grids, a random noise is required to be generated. This raises a question: what methods are best suited to aligning the randomness across samples? Further, how large an impact can different coupling methods have?

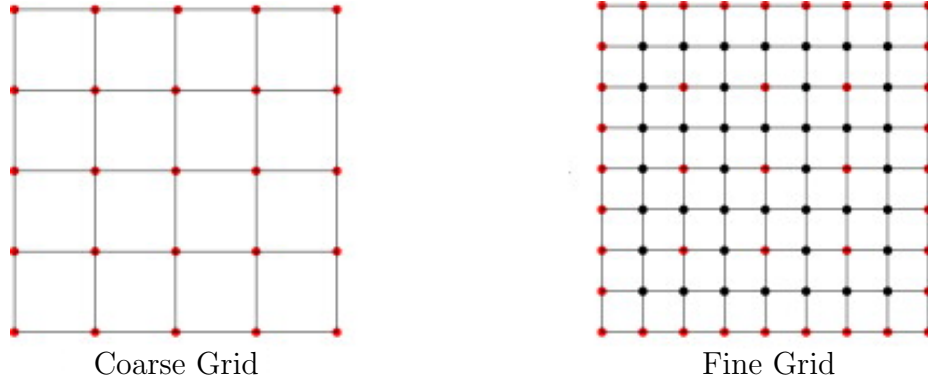


Figure 1.1: An illustrative coarse and fine finite difference grid, representing two adjacent levels of discretisation. The MLMC method computes samples on both grids to estimate a quantity of interest. Red dots indicate the vertices of the coarse grid, which are also present on the fine grid. This overlap allows for the correlation of the random noise generated at these points, a process known as coupling. Black dots indicate vertices that exist only on the more accurate fine grid.

There are three main questions this dissertation aims to answer therefore.

1. **What cost savings can MLMC achieve over standard MC in the context of two parabolic SPDEs?**
2. **What are the effects of different coupling mechanisms?**
3. **Monte Carlo methods are embarrassingly parallelisable. What further cost reductions are therefore achievable in high performance implementations of these methods if we take advantage of the highly parallel nature of MC methods?**

To answer these questions, this dissertation is structured as follows.

Chapter ?? provides the necessary theoretical background and a review of the relevant literature. We formally introduce the Monte Carlo and Multilevel Monte Carlo methods, provide a concise overview of the two parabolic SPDEs that serve as our case studies, and conclude with a literature review to situate this investigation within the current research landscape.

Chapter ?? details the numerical methodology and its validation. We present the MLMC Algorithm and derive the finite difference schemes for both the Stochastic Heat Equation and Dean-Kawasaki equation. We describe for each of these implementations the noise coupling strategies to be investigated, and for the Stochastic Heat Equation, derive theoretical targets and convergence rates with which we validate our implementation. We further validate our MLMC implementation of the Dean-Kawasaki equation, replicating the results of another paper [?].

Chapter SECTION HERE presents a comprehensive performance analysis. We directly address our first two research questions by comparing the computational cost of our MLMC implementation against a standard MC estimator for a range of target accuracies. Furthermore, we analyse the practical impact of different coupling mechanisms on overall performance.

Chapter SECTION HERE explores the potential for further cost reductions through high-performance computing. We discuss the design of a parallelised MLMC algorithm that leverages modern multi-core architectures and present performance metrics, such as computational speedup, to quantify the benefits of this approach.

Finally, Chapter SECTION HERE concludes the dissertation. We summarise our key findings, revisit the initial research questions, and discuss potential avenues for future investigation.

# Chapter 2

## Preliminary Concepts and Methods

This chapter presents the theoretical groundwork upon which the investigations of this dissertation is built. We begin by developing the theory of the Monte Carlo (MC) and Multilevel Monte Carlo (MLMC) methods, analysing their respective convergence and computational complexities.

Following this, we provide a concise introduction to the stochastic partial differential equations (SPDEs) we will focus on. We focus on two parabolic equations which serve as our case studies: The Stochastic Heat Equation and the Dean-Kawasaki equation. We outline their mathematical formulations and briefly discuss the challenges associated with their numerical solution, which are particularly pertinent to the Dean-Kawasaki equation.

The chapter concludes with a literature review that surveys the development and application of the MLMC method within the context of SPDEs, situating the research questions of this dissertation within the broader academic landscape.

### 2.1 Monte Carlo and Multilevel Monte Carlo

The primary objective in many applications involving SPDEs is not to find a single, particular solution, but rather to compute the expected value of a quantity of interest that depends on the solution. For example, we may want to find the average temperature at a specific point in a domain governed by the stochastic heat equation, or the average particle density fluctuation in a system described by the Dean-Kawasaki equation.

Let  $P$  represent such a quantity of interest, which we assume to be a real-valued random variable. Our goal throughout this section is to develop an efficient numerical

method for estimating its expectation,  $\mathbb{E}[P]$ . Since analytical expressions for  $\mathbb{E}[P]$  are rarely available in this context, we must turn to computational methods. The most fundamental of these is defined below.

**Definition 1 Monte Carlo Estimator.**

Let  $\{P^{(n)}\}$  for  $n = 1, \dots, N$  be a set of  $N$  independent and identically distributed (i.i.d) samples of a random variable  $P$ . The standard Monte Carlo estimator,  $\hat{P}_{MC}$ , of the expectation  $\mathbb{E}[P]$  is the sample mean:

$$\hat{P}_{MC} = \frac{1}{N} \sum_{n=1}^N P^{(n)}$$

By linearity of the expectation operator, the MC estimator in Definition ?? is unbiased. The accuracy of an estimator is typically estimated via its Mean Squared Error (MSE) and Root Mean Square Error (RMSE), which we define now.

**Definition 2 Mean Squared Error and Root Mean Squared Error.**

Let  $P$  be a fixed unknown quantity and let  $\hat{P}$  be an estimator for  $P$ . The **Mean Square Error (MSE)** of the estimator is the expected value of the squared error:

$$MSE(\hat{P}) = \mathbb{E}[(\hat{P} - P)^2]$$

This can be decomposed into terms representing the estimator's variance and squared bias:

$$MSE(\hat{P}) = \underbrace{\mathbb{V}[\hat{P}]}_{\text{Variance of estimator}} + \underbrace{(\mathbb{E}[\hat{P}] - \mathbb{E}[P])^2}_{\text{Bias}^2}$$

The **Root Mean Square Error (RMSE)** is the square root of the MSE.

Since the standard MC estimator is unbiased, its bias term is zero. Its MSE is therefore equal to its variance:

$$MSE(\hat{P}_{MC}) = \mathbb{V}[\hat{P}_{MC}] = \frac{1}{N} \mathbb{V}[P]$$

The framework above assumes we can generate perfect samples of the random variable  $P$ . In practice, for complex problems such as SPDEs this is typically impossible. Instead, we must compute numerical approximations which we will denote by  $P_L$ , where  $L$  represents a level of discretisation. For example, and as will be used later in a finite difference scheme,  $L$  could correspond to a mesh with spatial grid spacing  $h_L = 2^{-(L+1)}$ . A higher  $L$  means a finer mesh and a more accurate - but also more computationally expensive - approximation.



This introduces a second source of error. The total error of our estimate is now a combination of the statistical error from the Monte Carlo sampling and the systematic bias from numerical discretisation. Consequently, the MSE of the standard MC estimator using  $N$  samples of the numerical approximation  $P_L$ , denoted  $\hat{P}_{MC,L}$  is:

$$\text{MSE}(\hat{P}_{MC,L}) = \underbrace{\frac{1}{N}\mathbb{V}[P_L]}_{\text{Statistical Error (Variance)}} + \underbrace{(\mathbb{E}[P_L] - \mathbb{E}[P])^2}_{\text{Discretisation Error (Bias}^2)} \quad (2.1)$$

(??) presents the clear dilemma of the standard MC method. To achieve an overall MSE less than a tolerance  $\varepsilon^2$ , both terms must be sufficiently small. Reducing the statistical error requires a large number of samples  $N$ , while reducing the discretisation error requires a fine discretisation level  $L$ . Since the computational cost per sample  $C_L$ , increases sharply with  $L$ , the total cost,  $N \times C_L$ , often becomes prohibitively large.

This is the fundamental challenge that the Multilevel Monte Carlo method is designed to overcome. Instead of estimating the expensive quantity  $\mathbb{E}[P_L]$  directly, MLMC reformulates it using a telescoping sum:

$$\mathbb{E}[P_L] = \sum_{\ell=0}^L \mathbb{E}[Y_\ell], \quad \text{where} \quad Y_\ell := P_\ell - P_{\ell-1} \quad \text{and} \quad P_{-1} := 0.$$

Each correction term  $\mathbb{E}[Y_\ell]$  is then estimated independently with a standard Monte Carlo estimator.

**Definition 3 Multilevel Monte Carlo Estimator.**

Let  $Y_\ell = P_\ell - P_{\ell-1}$  be the correction at level  $\ell$ . The **Multilevel Monte Carlo (MLMC) estimator**,  $\hat{P}_{\text{MLMC}}$ , for the expectation  $\mathbb{E}[P_L]$  is:

$$\hat{P}_{\text{MLMC}} = \sum_{\ell=0}^L \hat{Y}_\ell, \quad \text{where} \quad \hat{Y}_\ell = \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} Y_\ell^{(n)}.$$

By the linearity of expectation,  $\hat{P}_{\text{MLMC}}$  is an unbiased estimator for  $\mathbb{E}[P_L]$ . Since the estimates at each level are independent, its variance is the sum of the individual variances. We define the cost and variance of a single sample at level  $\ell$  as  $C_\ell$  and  $V_\ell$  respectively:

$$\begin{aligned} C_\ell &:= \text{Cost}(Y_\ell) \\ V_\ell &:= \mathbb{V}[Y_\ell] = \mathbb{V}[P_\ell - P_{\ell-1}] \\ &= \mathbb{V}[P_\ell] + \mathbb{V}[P_{\ell-1}] - 2\text{Cov}(P_\ell, P_{\ell-1}). \end{aligned}$$

The total cost and variance of the MLMC estimator are therefore:

$$C_{\text{MLMC}} = \sum_{\ell=0}^L N_{\ell} C_{\ell}, \quad \mathbb{V}[\hat{P}_{\text{MLMC}}] = \sum_{\ell=0}^L \frac{V_{\ell}}{N_{\ell}}.$$

The success of the MLMC method hinges on the behaviour of the level variances,  $V_{\ell}$ . The key is to ensure that  $V_{\ell}$  decreases rapidly as the level  $\ell$  (and therefore cost  $C_{\ell}$ ) increases. This is achieved by using the same underlying source of randomness to generate pairs of samples  $(P_{\ell}, P_{\ell-1})$ . This technique, known as *coupling*, ensures the samples are strongly correlated. Because  $P_{\ell}$  and  $P_{\ell-1}$  are approximations of the same underlying quantity, their difference is small, and consequently the variance of this difference,  $V_{\ell}$ , is much smaller than the variance of either term individually. As the level  $\ell$  increases,  $P_{\ell}$  converges to  $P_{\ell-1}$ , and so we expect that  $V_{\ell} \rightarrow 0$ .

This decay of the level variance allows for a crucial trade-off: we use a small number of samples  $N_{\ell}$  for the expensive, high-level correction terms (where  $V_{\ell}$  is small) and compensate by using a large number of samples for the cheap, low-level terms where the variance is high. The optimal allocation of samples across levels can be determined by solving a constrained optimisation problem. For a fixed variance,  $\varepsilon^2$ , choosing the optimal  $\{N_{\ell}\}_{\ell=0}^L$  that minimises the total cost  $C_{\text{MLMC}} = \sum N_{\ell} C_{\ell}$  is solveable with Lagrange multipliers. It yields the optimal number of samples [?]:

$$N_{\ell} = \left\lceil \frac{1}{\varepsilon_{\text{var}}^2} \sqrt{\frac{V_{\ell}}{C_{\ell}}} \sum_{k=0}^L \sqrt{V_k C_k} \right\rceil. \quad (2.2)$$

The ceiling function  $\lceil \cdot \rceil$  ensures the number of samples is an integer. Equation (??) states that we should take more samples when the variance per unit cost of a level is high, and less when it is low.

We also have that the bias  $(\mathbb{E}[P_{\ell}] - \mathbb{E}[P]) \rightarrow 0$  as  $\ell \rightarrow \infty$ . To ensure that MSE is less than  $\varepsilon^2$ , by Definition ?? we can impose that  $(\mathbb{E}[P_L - P])^2 < \frac{\varepsilon^2}{2}$  and  $\mathbb{V}[\hat{P}_{\text{MLMC}}] < \frac{\varepsilon^2}{2}$ .

This leads to the following theorem [?] which makes precise the cost scaling of the MLMC method:

**Theorem 1 MLMC Complexity Theorem.**

*Let  $P$  denote a random variable, and let  $P_{\ell}$  denote the corresponding level  $\ell$  numerical approximation.*

*If there exists independent estimators  $Y_{\ell}$  based on  $N_{\ell}$  Monte Carlo samples, each with expected cost  $C_{\ell}$  and variance  $V_{\ell}$ , and positive constants  $\alpha, \beta, \gamma, c_1, c_2, c_3$  such that  $\alpha \geq \frac{1}{2} \min(\beta, \gamma)$  and we have*

1. **Weak Error (Bias) Decay:**  $|\mathbb{E}[P_\ell - P]| \leq c_1 2^{-\alpha\ell},$
2. **Unbiased Estimators:**  $\mathbb{E}[Y_\ell] = \begin{cases} E[P_0], & l = 0 \\ E[P_l - P_{l-1}], & l > 0, \end{cases}$
3. **Variance Decay:**  $V_\ell \leq c_2 2^{-\beta\ell},$
4. **Cost Growth:**  $C_\ell \leq c_3 2^{\gamma\ell},$

then there exists a positive constant  $c_4$  such that for any  $\varepsilon < e^{-1}$  there are values  $L$  and  $N_L$  for which the multilevel estimator

$$Y = \sum_{\ell=0}^L Y_\ell,$$

has an MSE with bound

$$MSE \equiv \mathbb{E}[(Y - \mathbb{E}[P])^2] < \varepsilon^2 \quad (2.3)$$

with a computational complexity  $C$  with bound

$$\mathbb{E}[C] \leq \begin{cases} c_4 \varepsilon^{-2}, & \beta > \gamma, \\ c_4 \varepsilon^{-2} (\log \varepsilon)^2, & \beta = \gamma, \\ c_4 \varepsilon^{-2 - (\gamma - \beta)/\alpha}, & \beta < \gamma \end{cases} \quad (2.4)$$

To appreciate the significance of the MLMC Complexity Theorem, we first establish the cost of the standard MC method in the above context. To achieve an MSE of  $\mathcal{O}(\varepsilon^2)$ , both the statistical and discretisation errors must be controlled. Controlling the bias to  $O(\varepsilon)$  requires using a fine grid with a step size  $h_L \propto \varepsilon^{1/\alpha}$ . Independently, controlling the statistical error to  $O(\varepsilon^2)$  requires  $N \propto \varepsilon^{-2}$  samples. The total cost is the product of the number of samples and the cost per sample, where  $C_L \propto h_L^{-\gamma}$ . Combining these requirements gives the overall complexity:

$$C_{\text{MC}} \propto N \times C_L \propto \varepsilon^{-2} \times (\varepsilon^{1/\alpha})^{-\gamma} = \varepsilon^{-2 - \gamma/\alpha}.$$

In contrast, in the MLMC case  $\beta > \gamma$ , the dominant computational cost is on coarsest level where the cost per sample  $C_\ell$  is  $O(1)$  [?]. Requiring  $N = O(\varepsilon^{-2})$  samples provides the dominant cost. This is the optimal case.

When  $\beta = \gamma$ , the cost contribution from each level,  $N_\ell C_\ell \propto \sqrt{V_\ell C_\ell}$ , is approximately constant across all levels [?]. The total cost is therefore proportional to the number of levels,  $L$ , which must increase as  $\mathcal{O}(\log \varepsilon)$  to meet the bias requirement.

This results in the total complexity of  $\mathcal{O}(\varepsilon^{-2}(\log \varepsilon)^2)$ , which remains a considerable improvement over the MC method.

In the case  $\beta < \gamma$ , the cost per level grows with  $\ell$ , meaning the total cost is dominated by the work on the finest level,  $L$ . Even in this worst case scenario though, we still arrive at a smaller scaling of cost than that of the MC estimator.

We conclude this section by highlighting the importance of the relationship between the variance decay rate  $\beta$ , and the cost growth rate  $\gamma$ . As shown in the Complexity Theorem, the magnitude of the computational savings offered by an MLMC implementation depends critically on whether the variance decreases faster than the cost increases (i.e. if  $\beta \geq \gamma$ ). Determining this relationship for specific SPDE applications and coupling strategies is a primary goal of this dissertation.

## 2.2 Stochastic Partial Differential Equations

An SPDE is a partial differential equation where at least one of the following is random: coefficients, initial boundary conditions, the domain, and the forcing term [?]. The SPDEs examined in this dissertation are specifically those with a random forcing term.

This section introduces the two parabolic SPDEs that serve as our primary case studies. For each, we will present the formal problem specification and define relevant quantities. These specifications form the basis for the derivation of their respective finite difference schemes and for the analytical calculation of the quantities of interest that are used for validation in Chapter ??.

### 2.2.1 The Stochastic Heat Equation

The stochastic heat equation (SHE) is a canonical example of a parabolic SPDE. Formally, it is the standard heat equation perturbed by a stochastic forcing term introducing spatially and temporally uncorrelated fluctuations. Heat diffusing through a metal bar that is also experiencing heat emitting chemical reactions would be the archetypal scenario the SHE describes.

The equation is fully specified by the SPDE itself, an initial condition and a set of boundary conditions [?, ?].

$$\begin{aligned}
\frac{\partial u(t, \mathbf{x})}{\partial t} &= \Delta u(t, \mathbf{x}) + \xi(t, \mathbf{x}), \quad \text{for } (t, \mathbf{x}) \in (0, T] \times \Omega & (\text{SHE}) \\
u(0, \mathbf{x}) &= u_0(\mathbf{x}), \quad \text{for } \mathbf{x} \in \Omega \\
\mathcal{B}u(t, \mathbf{x}) &= g(t, \mathbf{x}), \quad \text{for } (t, \mathbf{x}) \in (0, T] \times \partial\Omega
\end{aligned}$$

where  $u(t, \mathbf{x})$  is a real valued function at time  $t \in [0, T]$  and spatial position  $\mathbf{x}$  within a domain  $\Omega \subset \mathbb{R}^d$ . The function  $u_0(\mathbf{x})$  is the initial state of the field, the operator  $\mathcal{B}$  represents the boundary condition on the boundary of domain  $\partial\Omega$ , and  $\Delta$  is the standard Laplacian operator.

The defining component is the stochastic forcing term  $\xi(t, \mathbf{x})$ , which denotes space-time white noise. We formally define this as follows ([?], Chapter 1).

**Definition 4 Space-Time White Noise.**

*A space-time white noise  $\xi(t, \mathbf{x})$  on  $[0, T] \times D$  is a centred Gaussian process defined by a collection of random variables  $\{W(\phi)\}$  indexed by test functions  $\phi \in L^2([0, T] \times D)$ , with a covariance structure given by:*

$$\mathbb{E}[W(\phi)W(\psi)] = \langle \phi, \psi \rangle_{L^2}$$

This formal definition has several important consequences. Firstly, it gives rise to the more intuitive heuristic covariance expression  $\mathbb{E}[\xi(t, \mathbf{x})\xi(s, \mathbf{y})] = \delta(t - s)\delta(\mathbf{x} - \mathbf{y})$ , implying the noise is perfectly uncorrelated at every point. Secondly, the abstract process can be understood as the distributional derivative of a more tangible (though still highly irregular) object known as a Brownian sheet (a multidimensional Brownian motion).

Most critically for our purposes, Definition ?? directly informs how we can discretise the noise term in a numerical scheme. The integral of noise over a discrete space-time grid cell,  $C_j^n = [t_n, t_{n+1}] \times [x_j - \frac{\Delta x}{2}, x_j + \frac{\Delta x}{2}]$ , is found by choosing the test function  $\phi$  to be the indicator function of that cell,  $\phi_j^n(t, \mathbf{x}) = \mathbf{1}_{C_j^n}(t, \mathbf{x})$ .

The random variable representing the integrated noise over this cell is therefore  $W(\phi_j^n)$ . From Definition ??, we know this is a centred Gaussian random variable whose variance is given by:

$$\begin{aligned}
\text{Var}(W(\phi_j^n)) &= \mathbb{E}[W(\phi_j^n)^2] = \langle \phi_j^n, \phi_j^n \rangle_{L^2} \\
&= \int_0^T \int_D (\mathbf{1}_{C_j^n}(t, \mathbf{x}))^2 d\mathbf{x} dt \\
&= \int_{t_n}^{t_{n+1}} \int_{x_j - \frac{\Delta x}{2}}^{x_j + \frac{\Delta x}{2}} 1 d\mathbf{x} dt \\
&= \text{Area}(C_j^n) = \Delta t \Delta x
\end{aligned}$$

Since any centred Gaussian random variable with variance  $\sigma^2$  can be written as  $\sigma Z$  where  $Z \sim \mathcal{N}(0, 1)$ , this derivation leads to a crucial result for our finite difference implementations:

$$\int_{t_n}^{t_{n+1}} \int_{x_j - \frac{\Delta x}{2}}^{x_j + \frac{\Delta x}{2}} \xi(t, \mathbf{x}) d\mathbf{x} dt = \sqrt{\Delta t \Delta x} Z_j^n \quad (2.6)$$

where  $Z_j^n$  are independent and identically distributed standard normal random variables.

### 2.2.2 The Dean-Kawasaki Equation

The Dean-Kawasaki (DK) equation, having emerged from the field of fluctuating hydrodynamics, is used to describe the evolution of the density  $\rho(\mathbf{x}, t)$  of a system of  $N \gg 1$  weakly interacting particles. For the non-interacting case investigated in this dissertation, the equation is given by [?]:

$$\begin{aligned}
\frac{\partial \rho(t, \mathbf{x})}{\partial t} &= \frac{1}{2} \Delta \rho(t, \mathbf{x}) + N^{-1/2} \nabla \cdot (\sqrt{\rho(t, \mathbf{x})} \xi(t, \mathbf{x})), & \text{for } (t, \mathbf{x}) \in (0, T] \times \Omega & \quad (\text{DK}) \\
\rho(0, \mathbf{x}) &= \rho_0(\mathbf{x}), & \text{for } \mathbf{x} \in \Omega \\
\rho(t, \mathbf{x}) &\text{ satisfies periodic b.c. on } \partial\Omega, & \text{for } t \in (0, T]
\end{aligned}$$

Here,  $\rho(t, \mathbf{x})$  is the particle density at time  $t$  in a domain  $\Omega = \mathbb{T}^d$ , the  $d$ -dimensional torus. The term  $\frac{1}{2} \Delta \rho$  describes standard particle diffusion. The stochastic forcing term  $N^{-1/2} \nabla \cdot (\sqrt{\rho} \xi)$  models the density fluctuations, where  $\xi$  is space-time white noise, scaled by the local density via the  $\sqrt{\rho}$  factor.

The equation is highly singular, and “strong” or “pathwise” solutions i.e. solutions that exist as a function of every given realisation of the noise, do not exist. Instead, the only existing solutions are martingale solutions, where  $\rho$  corresponds to empirical measures of the underlying particle system:

$$\rho(\mathbf{x}, t) \equiv \mu_t^N(\mathbf{x}) := N^{-1} \sum_{i=1}^N \delta(x - X_i(t)).$$

The solution therefore is not a smooth field but a sum of Dirac-delta distributions.

Despite the singular nature of the solution, Cornalba and Fischer [?] demonstrate that statistical properties of fluctuations around the mean-field limit  $\bar{\rho}$  can be simulated. The quantities of interest,  $P$ , generally take the form of functionals applied to a weighted average of the density fluctuations:

$$P = \psi(N^{1/2} \int (\mu_N^T - \bar{\rho}^T)(\mathbf{x}) \phi(\mathbf{x}) d\mathbf{x}). \quad (2.7)$$

These can be computed for sufficiently regular test functions  $\psi, \phi$ , provided  $Nh^d \gg 1$  (i.e. as long as on average there is more than one particle per grid cell). For example, choosing  $\psi(z) = z^2$  allows for the computation of the variance of the fluctuations.

## 2.3 Literature Review

The MLMC method was formally introduced and popularised by Giles for SDE path simulation in 2008 [?], building on earlier foundational work on multilevel integration by Heinrich starting in 1998 [?]. Early works focussed predominantly on application to SDEs, particularly in computational finance. Other research extended the method's application to a wider range of topics, including various classes of SDEs [?, ?], Lévy processes [?], Numerical Linear Algebra [?], and Reliability theory [?].

MLMC for SPDEs is a more recent and active area of research. A search on Scopus [?] of "Multilevel Monte Carlo Stochastic Partial Differential Equations" at the time of writing returns 88 documents. Changing this to "Multilevel Monte Carlo Parabolic Stochastic Partial Differential Equations" refines this down to 18. Of the work examining MLMC for SPDEs, much of the focus has been on elliptic SPDEs (for example [?, ?, ?]).

This dissertation is concerned with parabolic SPDEs, for which the literature provides a smaller but highly relevant set of foundational papers. This review will focus on three key works that inform the central questions of this research: establishing the theoretical basis for MLMC's efficiency, demonstrating its practical validation, and exploring its application to highly singular equations with different coupling strategies.

Barth, Lang, and Schwab [?] provide a foundational analysis of the convergence and complexity of the MLMC method for a general class of parabolic SPDEs. Using a

Galerkin method in space and a Euler-Maruyama scheme in time, they prove that the MLMC estimator significantly reduces the computational work required to achieve a given accuracy compared to a standard single-level method. Their key result shows that the computational complexity can be reduced from  $O(h_L^{-(d+4)})$  for a standard MC method to nearly  $O(h_L^{-(d+2)})$  for MLMC, where  $d$  is the spatial dimension and  $h_L$  is the finest mesh width. Their work provides theoretical underpinning for the cost savings this dissertation seeks to quantify for the stochastic heat and Dean-Kawasaki equations. However, their analysis does not explore the practical performance improvements of different noise coupling strategies.

In [?], Giles and Reisinger provide a practical demonstration of MLMC's performance improvements for a class of parabolic SPDEs arising in financial modelling. The authors develop and analyse a Milstein finite difference scheme, proving it converges with first-order accuracy in time and second-order in space. They demonstrate a concrete reduction in computational complexity from  $O(\varepsilon^{-7/2})$  for a standard MC approach to the optimal  $O(\varepsilon^{-2})$  for their MLMC implementation, validating this gain through numerical experiments. This work serves as a methodological benchmark, demonstrating how to empirically confirm the theoretical performance gains of the MLMC method.

A very recent and highly relevant contribution is the 2024 paper by Cornalba and Fischer, which develops and analyses an MLMC method specifically for the Dean-Kawasaki equation, one of the two case studies in this dissertation. Their work tackles a highly singular SPDE for which standard convergence proofs fail. By formulating their analysis in terms of the convergence of probability distributions, they prove that MLMC provides a significant computational improvement over standard MC, provided average particle density is sufficiently large. Crucially, they propose and analyse two distinct noise coupling strategies: a "Fourier coupling" and a "Right-Most Nearest Neighbours (NN) coupling". We build directly on this work in this dissertation by also investigating the Dean-Kawasaki equation and proposing an alternative noise coupling strategy.

NEED TO ADD REGARDING HPC MLMC IMPLEMENTATIONS AND SPDES  
(believe none done)



# Chapter 3

## Methodologies and Validation

This chapter details the practical methodologies and numerical validation that form the foundation for the investigations in this dissertation. Having presented the theoretical groundwork in the preceding chapter, we now turn to the numerical implementations.

This chapter is organised into 3 main parts. First, we present the MLMC algorithm that is implemented throughout this work.

Second, we focus on the Stochastic Heat Equation (SHE). We begin by defining the specific SHE problem to be analysed and deriving the explicit finite difference scheme used for its numerical solution. We then detail the MLMC implementation, including the three distinct noise coupling strategies to be investigated. To validate this implementation, we apply the method to two chosen quantities of interest: the squared amplitude of Fourier modes and the system's total energy. We first derive theoretical convergence values and rates for these quantities, and then demonstrate that our implementation aligns with these.

Third, we address the Dean-Kawasaki equation. Following the approach of Cornalba and Fischer [?], we present the finite difference scheme, MLMC implementation and target quantity of interest. We then again validate our implementation by demonstrating its convergence and alignment with the results of Cornalba and Fischer.

By establishing and then rigorously verifying our numerical methods, this chapter provides the necessary groundwork for the performance analysis that follows.

### 3.1 MLMC Algorithm

Having outlined the MLMC theory in Section ??, we now present the MLMC algorithm that is implemented in this dissertation. This is based on the MLMC implementation provided by Giles [?]. The practical implementation of the MLMC method

follows a two phase approach. The first phase is dedicated to empirically estimating the parameters of the MLMC Complexity Theorem ??, while the second phase uses these parameters in an adaptive algorithm to achieve a final result.

Before running the main adaptive algorithm, A fixed, large number of  $N$  samples is simulated on a specified range of discretisation levels. From these simulations we gather empirical estimates of  $|\mathbb{E}[Y_\ell]|, V_\ell, C_\ell$ . Performing linear regression on the logarithms on these quantities against the index level we obtain estimates for the weak error decay rate, variance decay rate, and cost growth rate,  $\alpha, \beta$  and  $\gamma$  respectively.

Those estimates can then be passed to the MLMC algorithm presented here.

## 3.2 Stochastic Heat Equation Implementation

### 3.2.1 Problem Specification and Finite Difference Scheme

We consider the one-dimensional SHE on a unit interval  $[0, 1]$  with homogenous Dirichlet boundary conditions and a zero initial condition. This represents the simplest setting for a parabolic SPDE. The problem is formally defined as finding the real-valued function  $u(t, x)$  that satisfies:

$$\frac{\partial u(t, x)}{\partial t} - \frac{\partial^2 u(t, x)}{\partial x^2} = \xi(x, t), \quad \text{for } (t, x) \in (0, T] \times (0, 1) \quad (3.1a)$$

$$u(0, x) = 0, \quad \text{for } x \in [0, 1] \quad (3.1b)$$

$$u(t, 0) = 0, \quad u(t, 1) = 0, \quad \text{for } t \in (0, T]. \quad (3.1c)$$

To solve (??) numerically, we employ an explicit finite difference scheme obtained using a finite volume approach [?] where (??) is integrated over small, discrete space-time control volumes and the resulting integral terms then approximated. This scheme is:

$$U_j^{n+1} = U_j^n + \frac{\Delta t}{(\Delta x)^2} (U_{j+1}^n - 2U_j^n + U_{j-1}^n) + \Delta W_j^n, \quad (3.2)$$

$$\text{where } \Delta W_j^n = \sqrt{\frac{\Delta t}{\Delta x}} Z_j^n \quad \text{and} \quad Z_j^n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$$

The derivation of this scheme now follows.

First, we define a uniform grid. The spatial domain  $[0, 1]$  is discretised into  $J$  intervals of width  $\Delta x = 1/J$ , with grid points  $x_j = j\Delta x$  for  $j = 0, 1, \dots, J$ . Similarly, let the time interval  $[0, T]$  be discretised into  $N$  steps of size  $\Delta t = T/N$ , with the time

---

**Algorithm 1** Adaptive Multilevel Monte Carlo Algorithm

---

**Input:** Target RMSE  $\varepsilon > 0$ , initial number of samples  $N_0$ ,  $\alpha$ ,  $\beta$  and  $\gamma$  estimates.

**Output:** MLMC estimate  $\hat{P}_{\text{MLMC}}$ , sample counts  $N_\ell$ , variances  $V_\ell$ , and costs  $C_\ell$  for each level.

```
1: Initialize:
2: Set finest level  $L \leftarrow 2$ .
3: Set initial required samples  $\Delta N_\ell \leftarrow N_0$  for  $\ell = 0, \dots, L$ .
4: Set current sample counts  $N_\ell \leftarrow 0$  for  $\ell = 0, \dots, L$ .
5: Initialize statistical accumulators for sums of  $Y_\ell$  and  $Y_\ell^2$ .
6:
7: while  $\sum_{\ell=0}^L \Delta N_\ell > 0$  do                                 $\triangleright$  Loop until all levels have sufficient samples
8:                                      $\triangleright$  1: Generate Samples
9:   for all  $\ell = 0, \dots, L$  do
10:    if  $\Delta N_\ell > 0$  then
11:      Simulate  $\Delta N_\ell$  new samples of the correction term  $Y_\ell = P_\ell - P_{\ell-1}$ .
12:      Update the statistical accumulators (sums of  $Y_\ell$  and  $Y_\ell^2$ ).
13:      Update the current sample count:  $N_\ell \leftarrow N_\ell + \Delta N_\ell$ .
14:    end if
15:  end for
16:
17:                                      $\triangleright$  2: Update Estimates and Optimal Allocation
18:  for all  $\ell = 0, \dots, L$  do
19:    Compute the current estimate of the variance  $V_\ell \leftarrow \mathbb{V}[Y_\ell]$  from the accumulators.
20:  end for
21:  Calculate the optimal number of samples  $N_\ell^{\text{opt}}$  for all levels using the current variance estimates  $V_\ell$  and the cost per sample  $C_\ell$  in equation (??) with a target statistical variance of  $\varepsilon^2/2$ .
22:  Compute the required additional samples:  $\Delta N_\ell \leftarrow \max(0, N_\ell^{\text{opt}} - N_\ell)$ .
23:
24:                                      $\triangleright$  3: Check for Bias Convergence
25:  if  $\sum \Delta N_\ell < 0.01 \sum N_\ell$  then                                 $\triangleright$  Check only if sampling has stabilized
26:    Compute the mean of the finest correction,  $|\hat{Y}_L| = \left| \frac{1}{N_L} \sum_{n=1}^{N_L} Y_L^{(n)} \right|$ .
27:    if  $|\hat{Y}_L| \geq \varepsilon/\sqrt{2}$  then                                 $\triangleright$  If bias is still too large
28:       $L \leftarrow L + 1$ .                                 $\triangleright$  Add a new, finer level.
29:      Initialize accumulators,  $N_L$ , and  $\Delta N_L$  for the new level.
30:    end if
31:  end if
32: end while
33:
34: Compute Final Estimate:  $\hat{P}_{\text{MLMC}} \leftarrow \sum_{\ell=0}^L \left( \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} Y_\ell^{(n)} \right)$ .
```

---

points  $t_n = n\Delta t$  for  $n = 0, 1, \dots, N$ . Our discrete approximations of  $u$  are denoted  $U_j^n \approx u(t_n, x_j)$  with  $U_j^0 = 0$ ,  $U_0^n = U_J^n = 0$  capturing our initial and boundary conditions respectively.

We integrate (??) over a control volume  $C_j^n = [x_j - \frac{\Delta x}{2}, x_j + \frac{\Delta x}{2}] \times [t_n, t_n + 1]$ :

$$\int_{t_n}^{t_{n+1}} \int_{x_j - \frac{\Delta x}{2}}^{x_j + \frac{\Delta x}{2}} \frac{\partial u}{\partial t} dx dt - \int_{t_n}^{t_{n+1}} \int_{x_j - \frac{\Delta x}{2}}^{x_j + \frac{\Delta x}{2}} \frac{\partial^2 u}{\partial x^2} dx dt = \int_{t_n}^{t_{n+1}} \int_{x_j - \frac{\Delta x}{2}}^{x_j + \frac{\Delta x}{2}} \xi(t, x) dx dt \quad (3.3)$$

We then approximate each term in this equation. Focussing first on the time derivative term on the LHS:

$$\begin{aligned} \int_{t_n}^{t_{n+1}} \int_{x_j - \frac{\Delta x}{2}}^{x_j + \frac{\Delta x}{2}} \frac{\partial u}{\partial t} dx dt &= \int_{x_j - \frac{\Delta x}{2}}^{x_j + \frac{\Delta x}{2}} [u(t_{n+1}, x) - u(t_n, x)] dx && \begin{array}{l} \text{(Fundamental} \\ \text{Theorem of Calculus} \\ \text{(FTOC))} \end{array} \\ &\approx \Delta x (U_j^{n+1} - U_j^n) && \text{(Midpoint Rule)} \end{aligned}$$

Similarly, for the spatial derivative term:

$$\begin{aligned} \int_{t_n}^{t_{n+1}} \int_{x_j - \frac{\Delta x}{2}}^{x_j + \frac{\Delta x}{2}} \frac{\partial^2 u}{\partial x^2} dx dt &= \int_{t_n}^{t_{n+1}} \left[ \frac{\partial u}{\partial x} \left( t, x_j + \frac{\Delta x}{2} \right) - \frac{\partial u}{\partial x} \left( t, x_j - \frac{\Delta x}{2} \right) \right] dt && \text{(FTOC)} \\ &\approx \frac{\Delta t}{\Delta x} [U_{j+1}^n - 2U_j^n + U_{j-1}^n] && \begin{array}{l} \text{(Midpoint Rule and} \\ \text{Central Differences)} \end{array} \end{aligned}$$

Finally, the forcing term on the RHS, via equation (??), is equal to

$$\int_{t_n}^{t_{n+1}} \int_{x_j - \frac{\Delta x}{2}}^{x_j + \frac{\Delta x}{2}} \xi(t, x) dx dt = \sqrt{\Delta x \Delta t} Z_j^n$$

where  $Z_j^n$  are independent and identically distributed standard normal random variables. Finally, we substitute the discrete approximations for each of the three terms back into the integral equation (??). This yields the following relation:

$$\Delta x (U_j^{n+1} - U_j^n) = \frac{\Delta t}{(\Delta x)^2} \Delta x (U_{j+1}^n - 2U_j^n + U_{j-1}^n) + \sqrt{\Delta t \Delta x} Z_j^n$$

Rearranging yields the scheme shown in equation (??). This scheme is used to propagate the solution forward in time. It is known to be conditionally stable, requiring the Courant-Friedrichs-Lewy (CFL) condition,  $\frac{\Delta t}{(\Delta x)^2} \leq \frac{1}{2}$ , to be satisfied for convergence, similar to the corresponding deterministic heat equation [?]. A proof of this is given in the Appendix (CITE APPENDIX HERE).

### 3.2.2 Stochastic Heat Equation - MLMC Implementation

We now describe how scheme (??) is implemented for our MLMC algorithm ???. This applies for a generic quantity of interest (QoI),  $P$ . The QoIs tested for this investigation are outlined in section ??.

For each level  $\ell$ , we divide the spatial domain into  $n_\ell = 2^{l+1}$  subdivisions resulting in a spatial step size  $\Delta x_\ell = 1/n_\ell$ . We fix CFL number  $\lambda = \frac{\Delta t}{(\Delta x)^2} = 0.25$ . This yields a timestep of  $\Delta t_\ell = \lambda(\Delta x_\ell)^2$ . This means that for any two consecutive levels the refinement ratios are related via

$$\Delta x_{\ell-1} = 2\Delta x_\ell, \quad \Delta t_{\ell-1} = 4\Delta t_\ell. \quad (3.6)$$

For each level  $\ell$ , to obtain a sample at that level we evolve scheme (??) with discretisations  $\Delta x_\ell$  and  $\Delta t_\ell$  the desired number of time steps, and then compute our discrete estimate of  $P$ ,  $P_\ell$ .

We highlight that, given these refinement ratios, an additional level of refinement results in 8 times as many cells in the grid. Therefore, we anticipate the cost to scale as  $C_\ell \propto 2^{3\ell}$ , implying a cost growth rate of  $\gamma = 3$ .

For noise coupling, we investigate 3 strategies: Right-Most Nearest Neighbours (NN) used in [?], Central Coupling (CC) and a Finite Element (FE) based coupling. We outline each of these now. We follow the convention used in [?] for coupling, describing the  $\ell$ -th level as the *fine* level and the  $\ell - 1$ -th level as the *coarse* level.

#### Right-Most Nearest Neighbours Coupling

For the fine level's noise generation, at each interior fine grid point,  $j$ , and each fine time step,  $n$ , an independent noise increment is generated:

$$\Delta W_{j,f}^n = \sqrt{\frac{\Delta t_f}{\Delta x_f}} Z_j^n, \quad \text{where } Z_j^n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$$

The coarse level's noise at a grid point,  $k$ , corresponding to fine grid point  $2j$ , over a coarse time step,  $m$ , is constructed by aggregating the underlying fine noise. We sum the noise from the corresponding fine grid point and its immediate right-hand Neighbour  $2j + 1$ . This sum is accumulated over the 4 underlying fine timesteps and then rescaled.

$$\Delta W_{k,c}^m = \frac{1}{2} \sum_{n=4m}^{4m+3} (\Delta W_{2j,f}^n + \Delta W_{2j+1,f}^n)$$

The rescaling factor  $\frac{1}{2}$  is essential to ensure that the coarse noise  $\Delta W_{k,c}^m$  has the correct statistical variance of  $\frac{\Delta t_c}{\Delta x_c}$  required for coarse grid simulation.

We highlight that this method discards the final interior noise increment from the fine grid point. We anticipated this having some detrimental effect compared to other coupling strategies, as this clearly leads to imperfect correlation between fine and coarse samples. This motivated the following two strategies.

### Central Coupling

As an alternative to the asymmetric NN method, we propose a centred coupling strategy. This aims to create a more symmetric correlation between the fine and coarse grids by defining the fundamental source of randomness on a "half-cell" refinement of the spatial grid. From this common source of randomness, the noise increments for both the fine and coarse grids are constructed.

We divide each internal fine grid cell  $[x_j - \frac{\Delta x_f}{2}, x_j + \frac{\Delta x_f}{2}] \times [t_n, t_n + \Delta t_f]$  into two half-cells of width  $\frac{\Delta x_f}{2}$ . For each of these half-cells and for each fine time step  $n$ , we generate an independent, fundamental noise increment  $\zeta$ . Let  $\zeta_{j,L}^n$  and  $\zeta_{j,R}^n$  be the half-cell noises on the left and right halves of the  $j$ -th fine grid point during the  $n$ -th time step. These are i.i.d Gaussian random variables with variance equal to the area of the half-cell:

$$\zeta_{j,L}^n, \zeta_{j,R}^n \stackrel{\text{i.i.d}}{\sim} \mathcal{N}(0, \frac{\Delta x_f \Delta t_f}{2})$$

The noises for the fine finite difference scheme are then constructed by aggregating these fundamental half-cell noises.

$$\Delta W_{j,f}^n = \frac{1}{\Delta x_f} (\zeta_{j,L}^n + \zeta_{j,R}^n)$$

where the  $\frac{1}{\Delta x_f}$  coefficient ensures  $\Delta W_{j,f}^n$  has the correct variance.

Similarly to the NN coupling, the coarse noise is constructed as an aggregation of the underlying fine noises accumulated over the four fine timesteps that constitute a single coarse time step, equivalent to using the exact underlying 16 half-cell noises.

$$\Delta W_{k,c}^n = \frac{1}{\Delta x_c} \sum_{n=4m}^{4m+3} (\Delta W_{2j,f}^n + \Delta W_{2j+1}^n) = \frac{1}{\Delta x_c} \sum_{n=4m}^{4m+3} (\zeta_{2j,L}^n + \zeta_{2j,R}^n + \zeta_{2j+1,L}^n + \zeta_{2j+1,R}^n)$$

Again, the scaling factor  $\frac{1}{\Delta x_c}$  ensures the resulting noises have the correct variances required on the coarse grid. By further discretising the spatial grid, this method

aimed to better align the noises used for each grid point in both the coarse and fine grids, unlike the NN method, ensuring no underlying fine noise was discarded in constructing coarse noise.

### Finite Element Coupling Method

The third coupling strategy is derived from a Galerkin Finite Element Method (FEM) [?] spatial discretisation of the SHE. We transform the infinite-dimensional SPDE into a finite-dimensional system of SDEs, which in turn defines a structure of the discrete noise and coupling between levels.

We begin by formulating a weak form of the SHE. We multiply (??) by a sufficiently smooth test function  $\phi(x)$  and integrate over the spatial domain  $D = [0, 1]$ . Applying integration by parts yields:

$$(du, \phi) + (u_x, \phi_x)dt = (dW(t), \phi) \quad (3.7)$$

where  $(\cdot, \cdot)$  denotes the  $L^2$  inner product and  $dW(t) = \xi(t, x)dt$  represents Brownian motion.

The Galerkin method seeks an approximate solution,  $U(t, x)$  within a finite-dimensional subspace spanned by a set of basis functions [?]. For this problem, we use the standard piecewise linear “hat” basis functions  $\phi_j(x)$ , defined on a uniform grid with spacing  $h$ :

$$\phi_j(x) = \max(0, 1 - |x - x_j|/h).$$

The approximate solution is written as

$$U(t, x) = \sum_{j=1}^{J-1} U_j(t) \phi_j(x).$$

Here, the  $U_j(t)$  are the time-dependent coefficients represents the solution’s value at the spatial nodes  $x_j$ . By requiring the weak form to hold for every basis function in this space, we obtain a system of SDEs for the vector of coefficients  $\mathbf{U}(t)$ .

The system of SDEs can be written in matrix form as:

$$M d\mathbf{U}(t) + K\mathbf{U}(t)dt = dW(t) \quad (3.8)$$

where  $M$  is a tri-diagonal matrix with elements  $M_{ij} = (\phi_i, \phi_j)$ . For the hat basis, this gives

$$M_{i,i} = \frac{2}{3}h, \quad M_{i,i\pm 1} = \frac{1}{6}h.$$

$K$  is a tridiagonal matrix with entries  $K_{ij} = (\phi'_i, \phi'_j)$ . This gives

$$K_{i,i} = \frac{2}{h}, \quad K_{i,i\pm 1} = -\frac{1}{h}$$

$dW(t)$  is a vector of normal increments. To determine their variance and covariance, we can derive the following covariance function:

$$\mathbb{E}[(f, dW)(g, dW)] = (f, g)dt$$

for arbitrary spatial functions  $f$  and  $g$ . This follows from Definition ?? by considering test functions  $\phi_f$  and  $\phi_g$ :

$$\begin{aligned} \phi_f(s, x) &= f(x)\mathbf{1}_{[t, t+dt]}(s) \\ \phi_g(s, x) &= g(x)\mathbf{1}_{[t, t+dt]}(s) \end{aligned}$$

where  $\mathbf{1}_{[t, t+dt]}(s)$  is an indicator function which is 1 for  $s \in [t, t+dt]$ , 0 otherwise. The abstract random variable  $W(\phi_f)$  from Definition ?? now represents the noise tested against the spatial function  $f(x)$

$$\begin{aligned} \mathbb{E}[(f, dW)(g, dW)] &= \mathbb{E}[W(\phi_f)W(\phi_g)] \\ &= (\phi_f, \phi_g) \\ &= \int_0^T \int_D \phi_f(s, \mathbf{x})\phi_g(s, \mathbf{x}) d\mathbf{x}ds \\ &= \int_0^T \int_D (f(\mathbf{x})\mathbf{1}_{[t, t+dt]}(s)) (g(\mathbf{x})\mathbf{1}_{[t, t+dt]}(s)) d\mathbf{x}ds \\ &= \left( \int_D f(\mathbf{x})g(\mathbf{x}) d\mathbf{x} \right) \times \left( \int_0^T (\mathbf{1}_{[t, t+dt]}(s))^2 ds \right) \\ &= (f, g) dt \end{aligned}$$

The vector of normal increments have the following expectations:

$$\begin{aligned} \mathbb{E}[dW_i dW_j] &= \mathbb{E}[(dW, \phi_i)(dW, \phi_j)] = (\phi_i, \phi_j)dt \\ \mathbb{E}[dW_i^2] &= \frac{2}{3}hdt, \quad \mathbb{E}[dW_i dW_{i\pm 1}] = \frac{1}{6}hdt \end{aligned}$$



A common simplification known as mass lumping is used, where matrix  $M$  is replaced by the diagonal matrix  $hI$ . Applying this to (??) yields the scheme:

$$h\mathbf{U}^{n+1} = h\mathbf{U}^n - K\mathbf{U}^n\Delta t + \Delta W^n$$

where the discrete noise vector  $\Delta W^n$  has the covariance structure  $M\Delta t$ .

The coupling between a fine level  $\ell$  and a coarse level  $\ell - 1$  is derived from the relationship between the FEM basis functions. A coarse grid basis function,  $\phi_k^c$ , can be expressed as a linear combination of the fine grid basis functions:

$$\phi_k^c(x) = \frac{1}{2}\phi_{2k-1}^f(x) + \phi_{2k}^f(x) + \frac{1}{2}\phi_{2k+1}^f(x).$$

This provide a natural way to construct the coarse noise from the fine noise. The coarse noise increment at a coarse node  $k$ ,  $\Delta W_{k,c}^m$ , is constructed by applying the same linear weighting to the fine noise increments over the corresponding time steps:

$$\Delta W_{k,c}^m = \frac{1}{2} \sum_{n=4m}^{4m+3} \left( \frac{1}{2}\Delta W_{2k-1,f}^n + \Delta W_{2k,f}^n + \frac{1}{2}\Delta W_{2k+1,f}^n \right). \quad (3.9)$$

### 3.2.3 Quantities of Interest for the Stochastic Heat Equation

To validate our numerical schemes and analyse the performance of different MLMC coupling strategies, we compute expectations of two distinct quantities of interest (QoI) derived from the solution of the SHE problem (??). The first is the squared amplitude of the SHE's Fourier Modes, the second is a measure of the total energy of the system. We outline these here, derive the true values that our MLMC version wishes to converge to, and then also determine what the weak error and variance decay values,  $\alpha$  and  $\beta$  respectively, we expect to observe. In the next section we validate that our implementation converges to these results.

#### 3.2.3.1 Squared Amplitude of Fourier Modes

For the SHE problem (??), the solution can be expanded in a Fourier sine series. This decomposition is justified as the sine functions are the eigenfunctions of the Laplacian operator with the given boundary conditions, and they also form a basis for decomposing the stochastic noise term via the Karhunen-Lo  ve theorem ([?], Section 4.1).

$$u(t, x) = \sum_{n=1}^{\infty} u_n(t) \phi_n(x) \quad \phi_n(x) = \sqrt{2} \sin(n\pi x)$$

where  $u_n(t)$  is the  $n$ -th Fourier mode at time  $t$  and  $\phi_n$  our orthonormal sine basis functions.  $u_n(t)$  is given by the  $L^2$  inner product of  $u$  with  $\phi_n$ :

$$u_n(t) = \int_0^1 u(t, x) \phi_n(x) dx$$

thus our QoI, the expected squared Fourier Amplitude is given by:

$$P = \mathbb{E} [u_n(t)^2] = \mathbb{E} \left[ \left( \int_0^1 u(t, x) \phi_n(x) dx \right)^2 \right]$$

Our sample estimates of Fourier modes are obtained via the Discrete Inner Product, as defined in [?].

**Definition 5 Discrete Inner Product.**

For two functions  $V$  and  $W$  defined at interior mesh-points  $x_i$  for  $i = 1, \dots, N$  with spatial step size  $h$ ,

$$(V, W)_h := \sum_{i=1}^N h V_i W_i$$

An  $\ell$ -th level estimate of the squared amplitude of the  $n$ -th Fourier mode at timestep  $j$ ,  $(\hat{u}_n^{(\ell, j)})^2$ , we compute as:

$$(\hat{u}_n^{(\ell, j)})^2 = (U^j, \phi_n)_{\Delta x_\ell}^2$$

We derive first an analytic expression for  $u_n$ , and then the first and second moments of  $u_n$  for validating our MLMC implementation is correct. We will also derive the analytic error we expect to observe and the expected variance decay.

**Analytic Moments of Fourier Modes**

To obtain analytic solutions to the Fourier modes, we first express each term in the SHE governing equation as a Fourier sine series (an approach justified in [?], Chapter 5, and [?], Section 4.1).

$$\begin{aligned} u_t(t, x) &= \sum_{n=1}^{\infty} \frac{du_n(t)}{dt} \phi_n(x) \\ u_{xx}(t, x) &= \sum_{n=1}^{\infty} -(n\pi)^2 u_n(t) \phi_n(x) \\ \xi(x, t) &= \sum_{n=1}^{\infty} \dot{B}_n(t) \phi_n(x) \end{aligned}$$

where  $\dot{B}_n(t)$  represents one-dimensional white noise for each mode. Substituting these expansions into the SHE yields:

$$\sum_{n=1}^{\infty} \left( \frac{du_n(t)}{dt} - (-(n\pi)^2 u_n(t)) - \dot{B}_n(t) \right) \phi_n(x) = 0$$

$$\sum_{n=1}^{\infty} \left( \frac{du_n(t)}{dt} + (n\pi)^2 u_n(t) - \dot{B}_n(t) \right) \phi_n(x) = 0$$

For this infinite sum to be zero for all  $x$ , the coefficient of each basis function must be zero independently. This decouples the partial differential equation into an infinite system of SDEs, one for each mode  $u_n(t)$ :

$$\frac{du_n(t)}{dt} + (n\pi)^2 u_n(t) = \dot{B}_n(t)$$

Writing this in differential form, where  $dB_n(t) = \dot{B}_n(t)dt$  is the increment of a standard one-dimensional Brownian motion, gives:

$$du_n(t) = -(k\pi)^2 u_n(t)dt + dB_n(t)$$

This is the equation for an *Ornstein-Uhlenbeck process*. With a zero initial condition,  $u_n(0) = 0$ , the solution to this SDE is a Gaussian process ([?], Chapter 5.1). Its first two moments are well-known:

$$\mathbb{E}[u_n(t)] = 0 \tag{3.10}$$

$$\mathbb{V}[u_n(t)] = \mathbb{E}[u_n(t)^2] = \frac{1 - e^{-2(n\pi)^2 t}}{2(n\pi)^2} \tag{3.11}$$

Thus,

$$u_n(t) \sim \mathcal{N}\left(0, \frac{1 - e^{-2\lambda_n^2 t}}{2\lambda_n^2}\right), \quad \lambda_n = \pi n$$

$$\mathbb{E}[u_n(t)^2] = \frac{1 - e^{-2\lambda_n^2 t}}{2\lambda_n^2} \tag{3.12}$$

### Weak Error and Variance Decay of Squared Amplitude of Fourier Modes

We now establish the decay rates for our MLMC estimators. We must analyse two key properties: the weak error of the finite difference scheme and the rate of decay of the MLMC variance,  $V_\ell$ . We present two propositions and their accompanying proofs to formally derive these rates for our chosen QoI, the squared amplitude (variance) of a single Fourier mode.

**Proposition 1 Weak Error for the Squared Amplitude of a Fourier Mode.**

*The explicit finite difference scheme for the Stochastic Heat Equation governing the  $n$ -th Fourier mode, with time step  $\Delta t$ , has a weak error in its stationary variance of  $\mathcal{O}(\Delta t)$ . Consequently, for a finite difference implementation of the SHE with a fixed CFL number such that  $\Delta t \propto (\Delta x)^2$ , the error in the squared amplitude of the mode is of order  $\mathcal{O}((\Delta x)^2)$ .*

Before presenting our proof, we conclude from Proposition ?? that, recalling Theorem ??, we have a decay in the weak error of  $\alpha = 2$ . This follows from Proposition ?? as:

$$\begin{aligned}\mathbb{E}[P_\ell - P] &= \mathcal{O}((\Delta x_\ell)^2) \\ &= \mathcal{O}((2^{-(\ell+1)})^2) \quad (\Delta x_\ell = 2^{-(\ell+1)} \text{ for MLMC Implementation}) \\ &= \mathcal{O}(2^{-2\ell})\end{aligned}$$

*Proof.* The proof compares the exact variance of the continuous Ornstein-Uhlenbeck (OU) process with the variance of its finite difference approximation. Following from (??), we can obtain that the stationary variance of the  $n$ -th Fourier mode is the following:

$$\lim_{t \rightarrow \infty} \mathbb{V}[u_n] := \mathbb{V}[u_n^\infty] = \frac{1}{2\lambda_n^2} \quad (3.13)$$

Recalling our finite difference scheme in (??), we express it instead in matrix-vector form, with  $N$  internal points:

$$\mathbf{U}^j := \begin{bmatrix} U_1^j \\ \vdots \\ U_N^j \end{bmatrix} \in \mathbb{R}^N, \quad \mathbf{Z}^j := \begin{bmatrix} Z_1^j \\ \vdots \\ Z_N^j \end{bmatrix} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathbf{I}_N), \quad (3.14)$$

$$\mathbf{U}^{j+1} = \mathbf{U}^j + \Delta t A \mathbf{U}^j + \sqrt{\frac{\Delta t}{\Delta x}} \mathbf{Z}^j, \quad (3.15)$$

$$\text{where } A = \frac{1}{(\Delta x)^2} \begin{bmatrix} -2 & 1 & & \\ 1 & -2 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & -2 \end{bmatrix}. \quad (3.16)$$

We can equivalently express our finite difference estimates  $\mathbf{U}^j$  as a finite series of Fourier modes along with our basis functions  $\phi$ :

$$\mathbf{U}^j = \sum_{n=1}^N \hat{u}_n^j \phi_n \quad \text{where } \phi_n := \sqrt{2} \begin{bmatrix} \sin\left(\frac{n\pi}{N+1} \cdot 1\right) \\ \vdots \\ \sin\left(\frac{n\pi}{N+1} \cdot N\right) \end{bmatrix} \in \mathbb{R}^N$$

where  $\hat{u}_n^j$  represents our estimate of the  $n$ -th Fourier mode at time step  $j$ .

Our discrete basis functions are orthonormal under Definition ?? Discrete Inner product ([?], Chapter 2), i.e.

$$(\phi_n, \phi_m)_{\Delta x} = \sum_{i=1}^N \Delta x \phi_{n,i} \phi_{m,i} = \delta_{n,m}$$

where  $\delta_{n,m}$  is the Kronecker-delta function. Applying the discrete inner product to our scheme (??) and examining each term:

$$\begin{aligned} (\mathbf{U}^{j+1}, \phi_n)_{\Delta x} &= (\mathbf{U}^j, \phi_n)_{\Delta x} + \Delta t (A\mathbf{U}^j, \phi_n)_{\Delta x} \\ &\quad + \sqrt{\frac{\Delta t}{\Delta x}} (\mathbf{Z}^j, \phi_n)_{\Delta x}, \end{aligned} \tag{3.17}$$

$$(\mathbf{U}^{j+1}, \phi_n)_{\Delta x} = \left( \sum_{k=1}^N \hat{u}_k^{j+1} \phi_k, \phi_n \right)_{\Delta x} = \hat{u}_n^{j+1},$$

$$(\mathbf{U}^j, \phi_n)_{\Delta x} = \hat{u}_n^j,$$

$$\Delta t (A\mathbf{U}^j, \phi_n)_{\Delta x} = \Delta t \Delta x (A\mathbf{U}^j)^T \phi_n = \Delta t \Delta x (\mathbf{U}^j)^T A \phi_n \tag{3.18}$$

$$\begin{aligned} &= \Delta t \Delta x \lambda_n^{FD} (\mathbf{U}^j)^T \phi_n \\ &= \Delta t \lambda_n^{FD} (\mathbf{U}^j, \phi_n)_{\Delta x} \\ &= \Delta t \lambda_n^{FD} \hat{u}_n^j \end{aligned} \tag{3.19}$$

$$\text{where } \lambda_n^{FD} = \frac{4}{(\Delta x)^2} \sin^2\left(\frac{\pi n}{2(N+1)}\right),$$

$$\begin{aligned} \sqrt{\frac{\Delta t}{\Delta x}} (\mathbf{Z}^j, \phi_n)_{\Delta x} &= \sqrt{\frac{\Delta t}{\Delta x}} \sum_{k=1}^N \Delta x \mathbf{Z}_k^j (\phi_n)_k \\ &= \sqrt{\Delta t \Delta x} \sum_{k=1}^N \mathbf{Z}_k^j (\phi_n)_k \\ &:= \nu_n^j \sim \mathcal{N}(0, \sigma_n^2), \end{aligned} \tag{3.20}$$

where in going from (??) to (??) we have substituted the eigenvalues,  $\lambda_n^{FD}$ , of the matrix  $A$  ([?], Chapter 2) and applied the orthogonality of the basis functions, and in (??) we have used that the sum of normal random variables is normally distributed.

We obtain that (??) is equal to:

$$\hat{u}_n^{j+1} = (1 - \Delta t \lambda_n^{FD}) \hat{u}_n^j + \nu_n^j \tag{3.21}$$

We now determine what the variance of the random variable  $\nu_n^j$  is:

$$\begin{aligned}
\mathbb{V}[\nu_n^j] &= \mathbb{V} \left[ \sqrt{\Delta t \Delta x} \sum_{k=1}^N \mathbf{Z}_k^j (\phi_n)^k \right] \\
&= \Delta t \Delta x \sum_{k=1}^N \mathbb{V}[\mathbf{Z}_k^j] ((\phi_n)_k)^2 \\
&= \Delta t \sum_{k=1}^N \Delta x ((\phi_n)_k)^2 \\
&= \Delta t
\end{aligned}$$

where we have used the orthonormality of the basis functions,  $(\phi_n, \phi_m)_{\Delta x} = \delta_{n,m}$ . Taking the variance of equation (??) therefore becomes:

$$\mathbb{V}[\hat{u}_n^{j+1}] = (1 - \Delta t \lambda_n^{FD})^2 \mathbb{V}[\hat{u}_n^j] + \Delta t \quad (3.22)$$

For the stationary variance, we have that  $\mathbb{V}[\hat{u}_n^{j+1}] = \mathbb{V}[\hat{u}_n^j] = \mathbb{V}[\hat{u}_j^\infty]$ . Solving for the stationary variance gives:

$$\mathbb{V}[\hat{u}_j^\infty] = \frac{\Delta t}{1 - (1 - \Delta t \lambda_n^{FD})^2} = \frac{1}{2\lambda_n^{FD} - (\lambda_n^{FD})^2 \Delta t} \quad (3.23)$$

$$= \frac{1}{2\lambda_n^{FD} (1 - \frac{\Delta t \lambda_n^{FD}}{2})} \quad (3.24)$$

$$= \frac{1}{2\lambda_n^{FD}} (1 + \frac{\Delta t \lambda_n^{FD}}{2} + \dots) \quad (\text{By Taylor Series}) \quad (3.25)$$

Recalling that  $\lambda_n^{FD} = \frac{4}{(\Delta x)^2} \sin^2 \left( \frac{\pi n}{2(N+1)} \right) = \frac{4}{(\Delta x)^2} \sin^2 \left( \frac{n\pi \Delta x}{2} \right)$ , as  $\Delta x = \frac{1}{N+1}$ , and expanding via another Taylor series yields:

$$\begin{aligned}
\lambda_n^{FD} &= \frac{4}{(\Delta x)^2} \left( \frac{n^2 \pi^2 (\Delta x)^2}{4} + \mathcal{O}((\Delta x)^2) \right) = n^2 \pi^2 + \mathcal{O}((\Delta x)^2) \\
&= \lambda_n^2 + \mathcal{O}((\Delta x)^2)
\end{aligned}$$

Substituting this into (??) yields:

$$\begin{aligned}
\mathbb{V}[\hat{u}_j^\infty] &= \frac{1}{2\lambda_n^{FD}} \left( 1 + \frac{\Delta t \lambda_n^{FD}}{2} + \dots \right) \\
&= \frac{1}{2(\lambda_n^2 + \mathcal{O}(\Delta x^2))} \left( 1 + \frac{\Delta t \lambda_n^{FD}}{2} + \dots \right) \\
&= \frac{1}{2\lambda_n^2 (1 + \mathcal{O}(\Delta x^2))} \left( 1 + \frac{\Delta t \lambda_n^{FD}}{2} + \dots \right) \\
&= \frac{1}{2\lambda_n^2} (1 + \mathcal{O}(\Delta x^2)) \left( 1 + \frac{\Delta t \lambda_n^{FD}}{2} + \dots \right) \\
&= \frac{1}{2\lambda_n^2} (1 + \mathcal{O}(\Delta x^2))
\end{aligned}$$

Finally therefore we obtain our result that the weak error of our finite difference scheme estimate is :

$$|\mathbb{V}[u_n^\infty] - \mathbb{V}[\hat{u}_n^\infty]| = \left| \frac{1}{2\lambda_n^2} - \frac{1}{2\lambda_n^2} (1 + \mathcal{O}(\Delta x^2)) \right| = \mathcal{O}(\Delta x^2)$$

□

Having established the weak error, we now turn to the MLMC variance decay. The following proposition will be used to validate the decay rate achievable in an MLMC implementation.

**Proposition 2 Magnitude of MLMC Variance Decay for the Squared Amplitude of Fourier Mode.**

Let  $P_\ell = \frac{1}{M} \sum_{m=1}^M (\hat{u}_n^{(\ell,m)})^2$  be the MC estimator for the squared amplitude of the the  $n$ -th Fourier mode on level  $\ell$ . The magnitude of the MLMC variance,  $V_\ell = \mathbb{V}[P_\ell - P_{\ell-1}]$ , is determined by the correlation,  $\rho$ , between the underlying discrete finite difference samples,  $\hat{u}_n^{(\ell,n)}$  and  $\hat{u}_n^{(\ell-1,n)}$ .

- If the processes are perfectly correlated ( $\rho^2 = 1$ ),  $V_\ell = \mathcal{O}((\Delta x_\ell)^4)$ .
- If the correlation is imperfect ( $\rho^2 < 1$ ),  $V_\ell = \mathcal{O}((\Delta x_\ell)^2)$ .

Following Proposition ??, we conclude for the MLMC implementation:

$$V_\ell = \begin{cases} \mathcal{O}(2^{-4\ell}), & \text{thus } \beta = 4 & \text{if } \rho = \pm 1 \\ \mathcal{O}(2^{-2\ell}), & \text{thus } \beta = 2 & \text{otherwise} \end{cases}$$

*Proof.* To prove Proposition ?? we will expand the terms in  $V_\ell$  and then solve for each of them.

$$\begin{aligned} V_\ell &= \mathbb{V}[P_\ell - P^{\ell-1}] \\ &= \mathbb{V}[P_\ell] + \mathbb{V}[P_{\ell-1}] - 2\text{Cov}((P_\ell, P_{\ell-1})) \end{aligned}$$

We begin by expanding  $\mathbb{V}[P_\ell]$ :

$$\begin{aligned} \mathbb{V}[P_\ell] &= \mathbb{V}\left[\frac{1}{M} \sum_{m=1}^M \left(\hat{u}_k^{(\ell,m)}\right)^2\right] \\ &= \frac{1}{M} \mathbb{V}[(\hat{u}_k^\ell)^2] \\ &= \frac{1}{M} \left( \mathbb{E}[(\hat{u}_k^\ell)^4] - (\mathbb{E}[(\hat{u}_k^\ell)^2])^2 \right) \end{aligned}$$

This gives us two expectations to determine. Starting with the latter of these, we recall that Fourier modes are normally distributed with zero mean. We therefore have the following:

$$\begin{aligned} \mathbb{V}[(\hat{u}_n^\ell)^2] &= \mathbb{E}[(\hat{u}_n^\ell)^2] - \underbrace{\mathbb{E}[(\hat{u}_n^\ell)^2]^2}_{=0} \\ \mathbb{V}[\hat{u}_n^\ell] &= \mathbb{E}[(\hat{u}_n^\ell)^2] \end{aligned}$$

Referring to the earlier derivation of the variance of the Fourier mode (??):

$$\mathbb{V}[\hat{u}_n^\ell] = \frac{1}{2\lambda_n^{FD} - \Delta t_\ell (\lambda_n^{FD})^2}. \quad (3.26)$$

We denote by  $\Delta x_\ell$  and  $\Delta t_\ell$  the spatial and temporal step sizes in our finite difference scheme at level  $\ell$ . These are related, for our explicit finite difference scheme, via  $\Delta t = \mu(\Delta x)^2$ , where  $\mu$  is the CFL number. Recalling that  $\lambda_n^{FD} = \frac{4}{(\Delta x)^4} \sin^2\left(\frac{n\pi\Delta x}{2}\right) = \lambda_n - \frac{n^4\pi^4}{12}(\Delta x)^2 + \mathcal{O}((\Delta x)^4)$  where  $\lambda_n = n^2\pi^2$ , we have:

$$\begin{aligned} (\lambda_n^{FD})^2 &= \underbrace{\left(\lambda_n - \frac{\pi^4 n^4}{12}(\Delta x)^2 + \mathcal{O}((\Delta x)^4)\right)^2}_{=a} \\ &= \lambda_n^2 - 2a\lambda_n(\Delta x)^2 + \mathcal{O}((\Delta x)^4) \end{aligned}$$

Returning to (??):



$$\begin{aligned}
\mathbb{V}[\hat{u}_n^\ell] &= \frac{1}{2\lambda_n^{FD} - \Delta t_\ell (\lambda_n^{FD})^2} = \frac{1}{2\lambda_n \left(1 - \frac{2a + \mu\lambda_n^2}{2\lambda_n} (\Delta x_\ell)^2 + \mathcal{O}((\Delta x_\ell)^4)\right)} \\
&= \frac{1}{2\lambda_n} \left(1 + \frac{2a + \mu\lambda_n^2}{2\lambda_n} (\Delta x_\ell)^2 + \mathcal{O}((\Delta x_\ell)^4)\right) \quad (\text{Taylor Series}) \\
&= \frac{1}{2\lambda_n} + \underbrace{\frac{2a + \mu\lambda_n^2}{4\lambda_n^2}}_{=b} (\Delta x_\ell)^2 + \mathcal{O}((\Delta x_\ell)^4) \\
&= \frac{1}{2\lambda_n} + b(\Delta x_\ell)^2 + \mathcal{O}((\Delta x_\ell)^4) := \sigma_\ell^2
\end{aligned}$$

We next express  $\mathbb{V}[\hat{u}_n^{\ell-1}]$  in terms of  $\sigma_\ell^2$ . This is straightforward via our geometric relationships between spatial steps at adjacent levels, (??).

$$\sigma_{\ell-1}^2 = \frac{1}{2\lambda_n} + 4b(\Delta x)^2 + \mathcal{O}((\Delta x)^4) \quad (3.27)$$

We now determine what  $\mathbb{E}[(\hat{u}_n^\ell)^4]$  is. Again, as  $\hat{u}_n^\ell$  is normally distributed with zero mean, we use the standard identity that  $\mathbb{E}[X^4] = 3\sigma^4$  for a normally distributed random variable  $X$  with zero mean. This yields:

$$\begin{aligned}
\mathbb{V}[(\hat{u}_n^\ell)^2] &= \mathbb{E}[(\hat{u}_n^\ell)^4] + (\mathbb{E}[(\hat{u}_n^\ell)^2])^2 \\
&= 3\sigma_l^4 - \sigma_l^4 = 2\sigma_l^4
\end{aligned}$$

Finally, we obtain that:

$$\mathbb{V}[P_\ell] = \frac{2\sigma_l^4}{M}, \quad \mathbb{V}[P_{\ell-1}] = \frac{2\sigma_{\ell-1}^4}{M}. \quad (3.28)$$

We now examine  $\text{Cov}(P_\ell, P_{\ell-1})$ .

$$\text{Cov}(P_\ell, P_{\ell-1}) = \frac{1}{M} \text{Cov}(\hat{u}_l^2, \hat{u}_{l-1}^2) = \frac{1}{M} \mathbb{E}[(\hat{u}_n^\ell \hat{u}_n^{\ell-1})^2] - \mathbb{E}[(\hat{u}_n^\ell)^2] \mathbb{E}[(\hat{u}_n^{\ell-1})^2]$$

We already know  $\mathbb{E}[(\hat{u}_n^\ell)^2] = \sigma_\ell^2$ . To solve for the first term, we use Isserlis's Theorem [?], another formula relating properties of normally distributed random variables.

This gives us:

$$\begin{aligned}
\mathbb{E}[(\hat{u}_n^\ell)^2 (\hat{u}_n^{\ell-1})^2] &= \mathbb{E}[(\hat{u}_n^\ell)^2] \mathbb{E}[(\hat{u}_n^{\ell-1})^2] + 2\mathbb{E}[\hat{u}_n^\ell \hat{u}_n^{\ell-1}]^2 \\
&= \sigma_\ell^2 \sigma_{\ell-1}^2 + 2\mathbb{E}[\hat{u}_n^\ell \hat{u}_n^{\ell-1}]^2
\end{aligned}$$

We know that

$$\text{Cov}(\hat{u}_n^\ell, \hat{u}_n^{\ell-1}) = \mathbb{E}[\hat{u}_n^\ell \hat{u}_n^{\ell-1}] - \underbrace{\mathbb{E}[\hat{u}_n^\ell] \mathbb{E}[\hat{u}_n^{\ell-1}]}_{=0} = \rho \sigma_\ell \sigma_{\ell-1}$$

and so we now obtain that:

$$\text{Cov}(P_\ell, P_{\ell-1}) = \frac{1}{M} (\sigma_\ell^2 \sigma_{\ell-1}^2 + 2\rho^2 \sigma_\ell \sigma_{\ell-1} - \sigma_\ell^2 \sigma_{\ell-1}^2) = \frac{2\rho^2 \sigma_\ell^2 \sigma_{\ell-1}^2}{M} \quad (3.29)$$

Putting together (??) and (??) into  $V_\ell$ :

$$V_\ell = \mathbb{V}[P_\ell] + \mathbb{V}[P_{\ell-1}] - 2\text{Cov}(P_\ell, P_{\ell-1}) = \frac{2}{M} (\sigma_\ell^4 + \sigma_{\ell-1}^4 - 2\rho^2 \sigma_\ell^2 \sigma_{\ell-1}^2) \quad (3.30)$$

We expand each of these terms, also setting  $\frac{1}{2\lambda_n} = \sigma^2$ , as it simplifies the expansion and also corresponds to the stationary variance of the continuous Fourier mode (??):

$$\begin{aligned} \sigma_\ell^4 &= (\sigma^2 + b(\Delta x_\ell)^2 + \mathcal{O}((\Delta x_\ell)^4))^2 = \sigma^4 + 2\sigma^2 b(\Delta x_\ell)^2 + \mathcal{O}((\Delta x_\ell)^4) \\ \sigma_{\ell-1}^4 &= (\sigma^2 + 4b(\Delta x_\ell)^2 + \mathcal{O}((\Delta x_\ell)^4))^2 = \sigma^4 + 8\sigma^2 b(\Delta x_\ell)^2 + \mathcal{O}((\Delta x_\ell)^4) \\ \sigma_\ell^2 \sigma_{\ell-1}^2 &= (\sigma^2 + b(\Delta x_\ell)^2 + \mathcal{O}((\Delta x_\ell)^4)) (\sigma^2 + 4b(\Delta x_\ell)^2 + \mathcal{O}((\Delta x_\ell)^4)) \\ &= \sigma^4 + 5\sigma^2 b(\Delta x_\ell)^2 + \mathcal{O}((\Delta x_\ell)^4) \end{aligned}$$

Finally, substituting these terms into (??) results in:

$$\begin{aligned} \sigma_\ell^4 + \sigma_{\ell-1}^4 - 2\rho^2 \sigma_\ell^2 \sigma_{\ell-1}^2 &= 2\sigma^4(1 - \rho^2) + 10\sigma^2 b(\Delta x_\ell)^2(1 - \rho^2) + \mathcal{O}((\Delta x_\ell)^2) \\ &= \begin{cases} \mathcal{O}((\Delta x_\ell)^4) & \text{if } \rho = \pm 1 \\ \mathcal{O}((\Delta x_\ell)^2) & \text{otherwise} \end{cases} \end{aligned}$$

□

### 3.2.3.2 System Energy

A fundamental property of the solution is its total energy, which is defined as the squared  $L^2$ -norm of the solution  $u(t, x)$  at a final time  $T$ . The continuous form of this quantity is given by:

$$P_{\text{energy}} = \int_0^1 (u(T, x))^2 dx \quad (3.31)$$

To derive the expected energy,  $\mathbb{E}[Q_{\text{energy}}]$ , we again express the solution  $u(t, x)$  in its Fourier sine series  $u(t, x) = \sum_{n=1}^{\infty} u_n(t) \phi_n(x)$ . Substituting this into (??):

$$\begin{aligned}
\int_0^1 (u(T, x))^2 dx &= \int_0^1 \left( \sum_{n=1}^{\infty} u_n(t) \phi_n(x) \right) \left( \sum_{n=1}^{\infty} u_n(t) \phi_n(x) \right) dx \\
&= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} u_n(t) u_m(t) (\phi_n, \phi_m) \\
&= \sum_{n=1}^{\infty} (u_n(t))^2
\end{aligned}$$

Taking the expectation:

$$\begin{aligned}
\mathbb{E} \left[ \int_0^1 (u(T, x))^2 dx \right] &= \mathbb{E} \left[ \sum_{n=1}^{\infty} (u_n(t))^2 \right] \\
&= \sum_{n=1}^{\infty} \mathbb{E} [(u_n(t))^2] \\
&= \sum_{n=1}^{\infty} \frac{1 - e^{-2\lambda_n t}}{2\lambda_n} \quad \text{via (??)} \\
&= \frac{1}{12} - \sum_{n=1}^{\infty} \frac{e^{-2n^2\pi^2 t}}{2n^2\pi^2} \\
&\approx \frac{1}{12} \quad \text{for sufficiently large } t \tag{3.32}
\end{aligned}$$

### 3.2.4 Validating MLMC Stochastic Heat Implementation

We now validate our MLMC implementation for the Stochastic Heat Equation. We do this by demonstrating that our MLMC estimator converges to the correct, analytically-derived expected values for our chosen quantities of interest for all three coupling strategies, and that the observed growth and decay rates align with those derived for squared amplitudes. We also discuss the resulting validation charts as they vary for different coupling strategies.

We first consider the squared Fourier amplitude. We use the first Fourier mode, for which the true value we know to be  $\approx 0.10$  via (??). For each of the three coupling strategies (NN, CC, and FE) the following was done. A preliminary run with 10,000 samples across each of the first six levels was performed to obtain empirical estimates of the key MLMC parameters  $(\alpha, \beta, \gamma)$  via linear regression. The adaptive MLMC algorithm (Algorithm ??) was then executed using those estimates rates for a range of target RMSEs,  $\varepsilon \in \{0.1, 0.05, 0.02, 0.01, 0.005, 0.001\}$ . This is in line with the method outlined in Section ??.

The results of this analysis are presented in Figure ??, the estimated decay rates summarised in Table ??.

Coupling Method	$\alpha$	$\beta$	$\gamma$
Nearest Neighbour	2.0	2.07	3.02
Central Coupling	1.86	2.08	3.02
Finite Element	1.96	4.0	3.0

Table 3.1: Empirically estimated rates for the Squared Amplitude QoI.

We first comment on the rates observed. In line with expectations for a 1D problem, we observe  $\gamma = 3$  across all couplings. Similarly, we obtain estimates of  $\alpha = 2$  for all coupling methods, in line with Proposition ?? which was independent of coupling. Figure ?? shows this decay, including the theoretical  $\alpha = 2$  observed rate which can be seen to be in alignment with the decays observed. Notably, there is strong alignment in the magnitude of the of weak errors for the NN and CC coupling methods, while the FE method systematically has a smaller weak error, suggesting fewer samples are required to remove the discretisation error for the FE method.

The most critical distinction between the methods is revealed in the variance decay, shown in Figure ??. We observe that the FE method achieves the optimal decay rate of  $\beta = 4$ , We can infer, following Proposition ??, that FE coupling achieves perfect correlation between levels while the CC method does not. We knew the NN method would not, as it discards some noise.

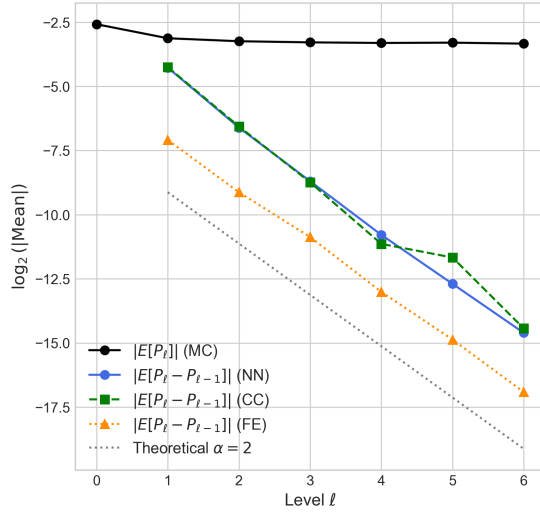
We also include the MC estimator’s means and variances, obtained also using 10,000 samples at each level reported. We observe no error or variance decay in the MC estimator which also aligns with out expectations.

Finally, we observe convergence for all coupling methods. Figure ?? shows clear convergence across the different RMSEs chosen. In Figure ?? we clearly see the benefit of the faster variance decay, FE coupling requiring only 3 levels to achieve the requisite level of accuracy, while NN and CC required 5.

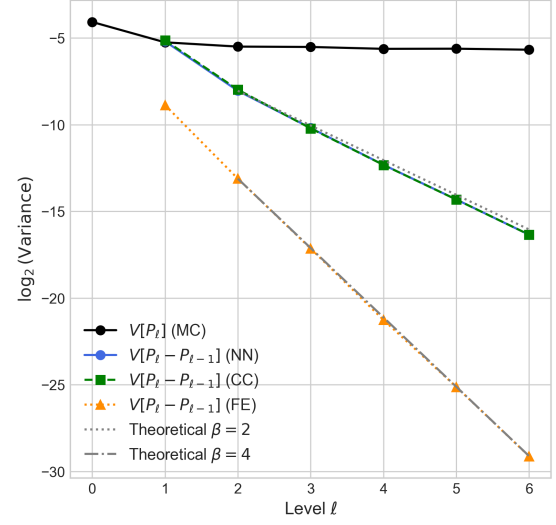
We now provide similar validation results for our SHE MLMC implementation where our QoI is the system energy. We have the true value for the problem,  $\approx \frac{1}{12}$ , via ??.

The rates shown in Table ?? and Figures ?? and ?? align with the results for the squared amplitude. We see the same approximate estimates in the decays.

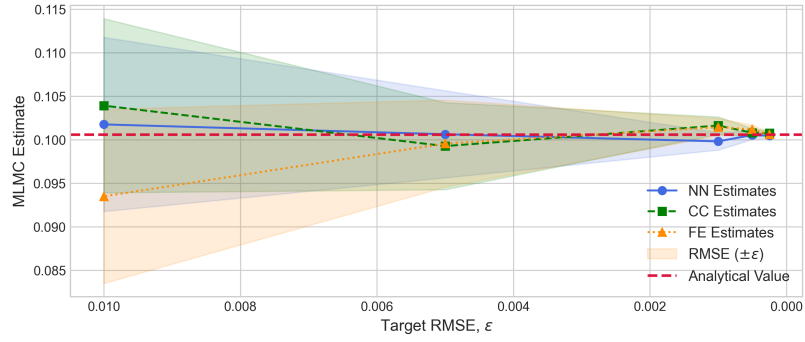
In Figure ?? we again see the benefits of the faster variance decay, with convergence being achieved at level 6 for the FE method, and level 12 for the CC and NN methods.



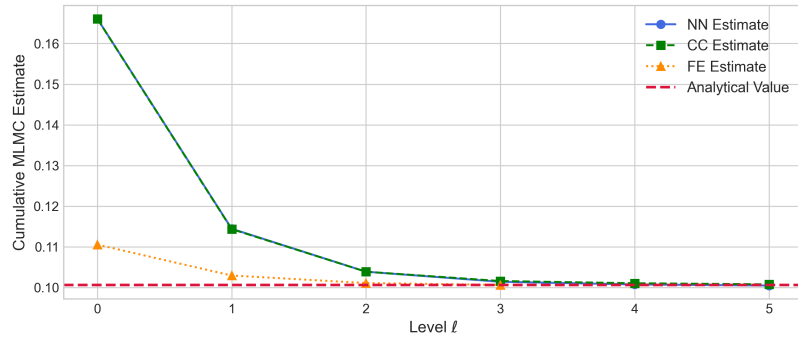
(a) Weak error convergence ( $\alpha$ ).



(b) MLMC variance decay ( $\beta$ ).

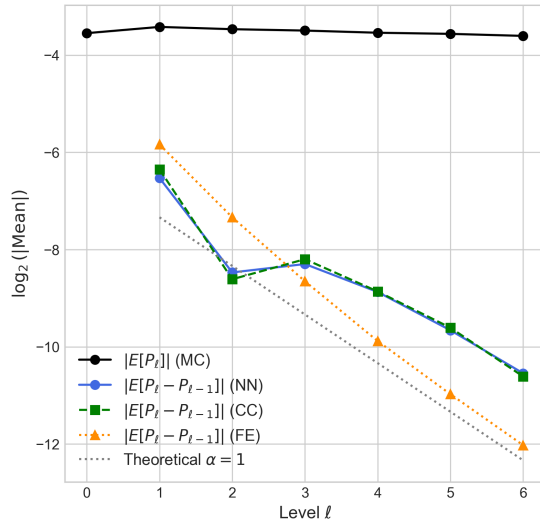


(c) Final MLMC estimate vs. target RMSE ( $\varepsilon$ ).

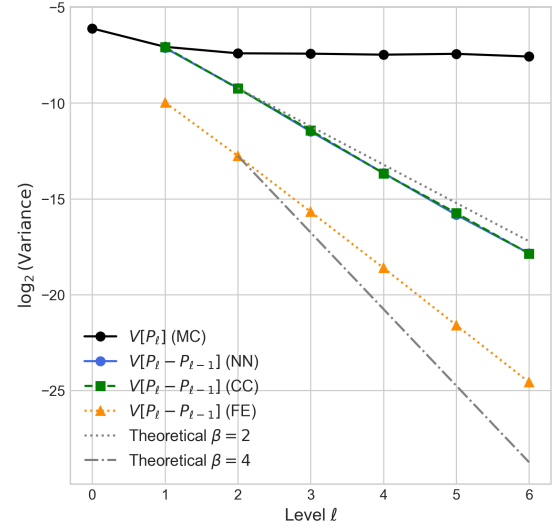


(d) Cumulative estimate vs. level ( $\ell$ ) for  $\varepsilon = 0.001$ .

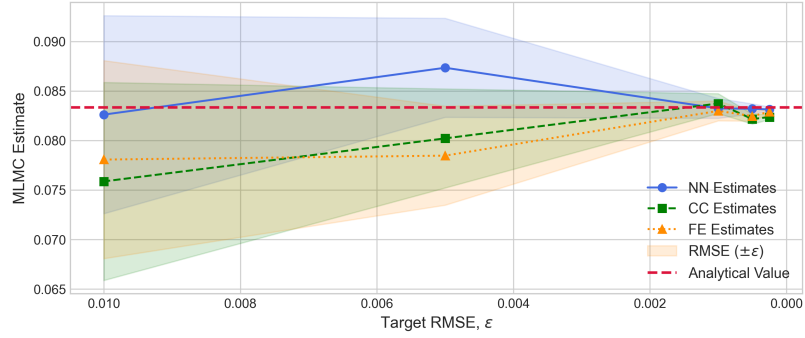
Figure 3.1: Decay and convergence plots for the MLMC implementation of the SHE, using the squared amplitude of the first Fourier mode as our QoI.



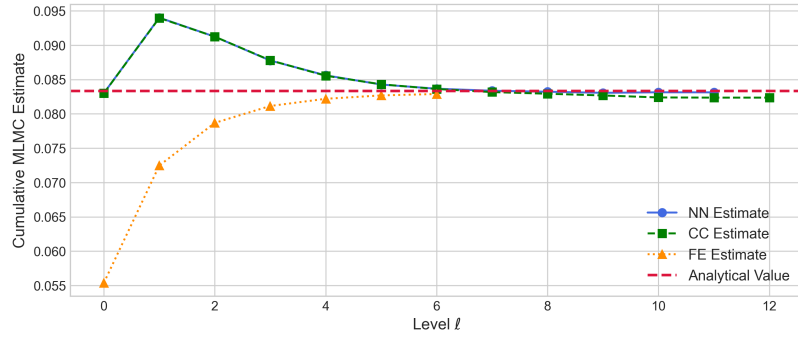
(a) Weak error convergence ( $\alpha$ ).



(b) MLMC variance decay ( $\beta$ ).



(c) Final MLMC estimate vs. target RMSE ( $\varepsilon$ ).



(d) Cumulative estimate vs. level ( $\ell$ ) for  $\varepsilon = 0.001$ .

Figure 3.2: Decay and convergence plots for the MLMC implementation of the SHE, using the system energy as our QoI.

Coupling Method	$\alpha$	$\beta$	$\gamma$
Nearest Neighbour	2.0	2.07	3.02
Central Coupling	1.86	2.08	3.02
Finite Element	1.96	4.0	3.0

Table 3.2: Empirically estimated rates for the system energy

### 3.3 Dean-Kawasaki Validation

#### 3.3.1 Dean-Kawasaki Implementation

We now present the Dean-Kawasaki problem considered and the finite-difference scheme, following [?], and our MLMC implementation. We consider the Dean-Kawasaki equation on a one dimensional torus  $\Omega = [0, 2\pi)$  with periodic boundary conditions and final time  $T > 0$ . The particle number is denoted  $N_p$ . The density field  $\rho$  evolves according to

$$\frac{\partial \rho(t, x)}{\partial t} = \frac{1}{2} \frac{\partial^2 \rho(t, x)}{\partial x^2} + N_p^{-1/2} \frac{\partial \left( \sqrt{\rho(t, x)} \xi(t, x) \right)}{\partial x}, \quad (t, x) \in (0, T] \times [0, 2\pi), \quad (3.33)$$

$$\begin{aligned} \rho(0, x) &= Z_0 \left( 1 + \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \sin^2\left(x - \frac{\pi}{2}\right)\right) \right), & x \in \Omega, \\ \rho(t, 0) &= \rho(t, 2\pi), & t \in [0, T], \end{aligned}$$

where  $\xi$  is space-time white noise and  $Z_0 \approx 0.12$  is a normalisation constant.

We discretise  $\Omega$  with a uniform grid of spacing  $\Delta x, \Delta t$  and denote by  $\rho_j^n$  the approximation to  $\rho(t_n, x_j)$  at time  $t_n = n\Delta t$  and position  $x_j = \Delta x$ .

$$\begin{aligned} \rho_j^{n+1} &= \rho_j^n + \frac{\lambda}{2} (\rho_{j+1}^n - 2\rho_j^n + \rho_{j-1}^n) + \frac{1}{\sqrt{N_p}} \frac{F_{j+1}^n - F_{j-1}^n}{2\Delta x}, \\ F_j^n &:= \sqrt{[\rho_j^n]_+} \Delta W_j^n, \quad \Delta W_j^n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \frac{\Delta t}{\Delta x}\right). \end{aligned} \quad (3.34)$$

Our quantity of interest is the expected squared mean field fluctuations in the  $\sin(x)$  mode at final time  $T$ , namely

$$P = \mathbb{E} \left[ N_p \left( \int_0^{2\pi} (\rho(T, x) - \bar{\rho}(T, x)) \sin(x) dx \right)^2 \right], \quad (3.35)$$

where  $\bar{\rho}$  denotes the deterministic mean-field solution.

In the discrete setting we approximate this by

$$P = N_p \left( (\boldsymbol{\rho}^N - \bar{\boldsymbol{\rho}}^N, \sin(x_h))_h \right)^2, \quad (3.36)$$

To compute  $P$ , we update in parallel the deterministic mean-field solution  $\bar{\rho}$  via an explicit finite difference scheme for the heat equation.

Our MLMC implementation is performed in the same manner as that used for the Stochastic Heat Equation. We set  $\Delta x_\ell = 2\pi/2^{l+1}$ ,  $\lambda = \frac{1}{4}$  and set  $\Delta t = \lambda(\Delta x)^2$ , ensuring stability [?]. For our quantity of interest, we take  $T = 0.25$ .

For coupling methods, we implement the NN and CC methods outlined in Section ???. The NN method is used to replicate the results from Cornalba and Fischer [?]. We then compare this with the proposed CC method to assess whether the symmetric coupling provides any tangible performance improvements.

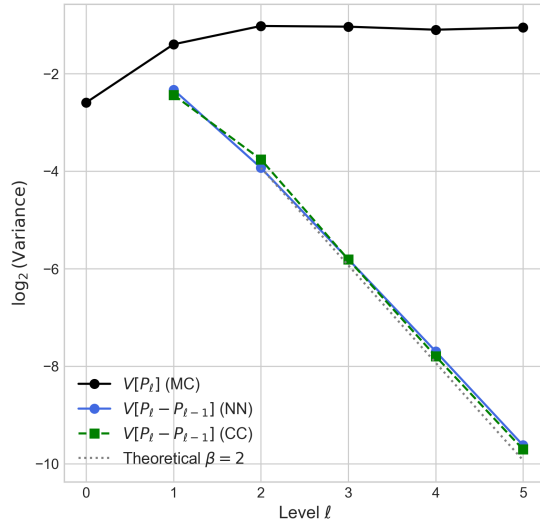
### 3.3.2 Dean-Kawasaki Validation

We now verify that our implementation converges. We obtained an estimate of the true value of (??) via MC simulation, and now verify convergence of our implementation across a range of accuracies,  $\varepsilon \in \{0.001, 0.005, 0.01, 0.05\}$ . We also again analyse variance and error decay.

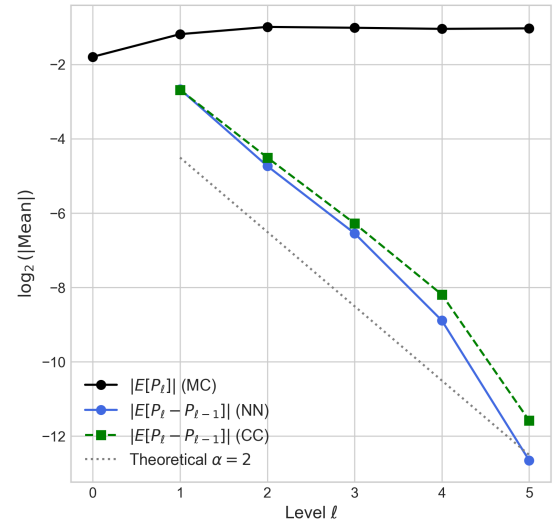
Coupling Method	$\alpha$	$\beta$	$\gamma$
Nearest Neighbour	2.11	1.95	3.0
Central Coupling	1.95	1.95	2.98

Table 3.3: Empirically estimated rates for the DK QoI.

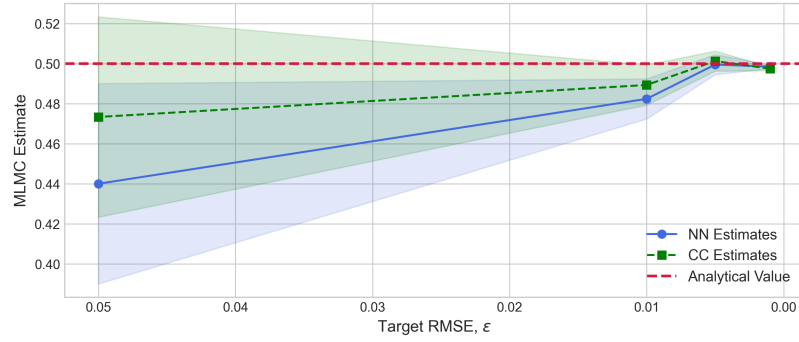




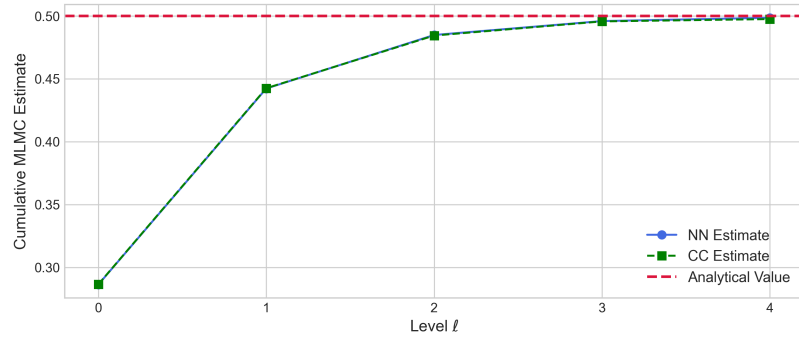
(a) MLMC variance decay ( $\beta$ ).



(b) Weak error convergence ( $\alpha$ ).



(c) Final MLMC estimate vs. target RMSE ( $\epsilon$ ).



(d) Cumulative estimate vs. level ( $\ell$ ) for  $\epsilon = 0.001$ .

Figure 3.3: Validation and convergence plots for the MLMC implementation for the Dean-Kawasaki equation.

# Chapter 4

## Performance Evaluations

Having validated the numerical implementations in the preceding chapter, we now present a comprehensive performance analysis. This chapter addresses the two of the central research questions of this dissertation by quantifying the computational savings achieved by the Multilevel Monte Carlo method compared to the standard Monte Carlo method, and further assessing the impact of the different coupling strategies.

Performance is assessed by analysing the computational cost required to achieve a target RMSE,  $\varepsilon$ . Following the methodology outlined by Giles [?] and used in the work by Cornalba and Fischer [?], we present results on a normalised log-log plot of  $\varepsilon^2 \text{Cost}$  versus  $\varepsilon$ . This metric is chosen for its direct connection to the MLMC Complexity Theorem: an optimal method with a theoretical complexity of  $\mathcal{O}(\varepsilon^{-2})$  will appear as a nearly horizontal line on this plot, providing a clear and visual confirmation of the method's efficiency.

The computational cost is measured as the total wall-clock time for each simulation. All experiments

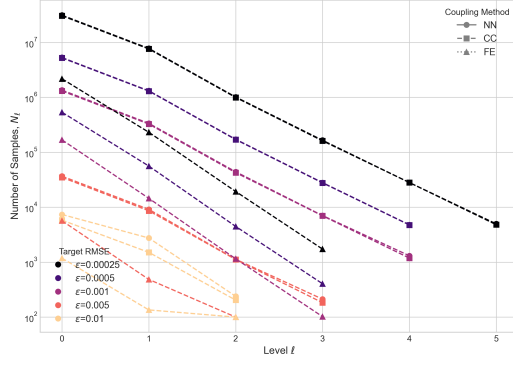
In this chapter we present an evaluation of the performance of our MLMC implementations for the Stochastic Heat Equation and Dean-Kawasaki equations outlined in the previous chapter. We aim to quantify the computational savings achieved by the MLMC method relative to the standard Monte Carlo estimator, and to compare the relative performance savings of different noise coupling strategies within the MLMC framework. Before presenting results, we outline how the performance is measured.

Performance is measured using the dimensional metric

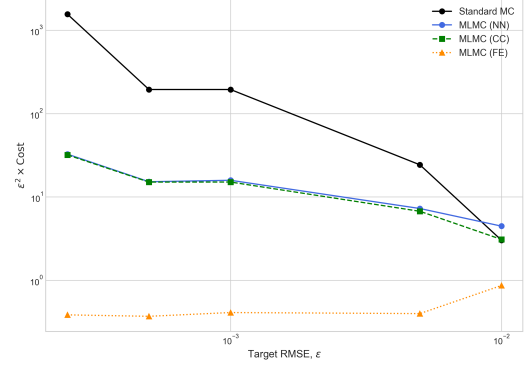
$$\text{Cost} \times \varepsilon^2,$$

where  $\varepsilon$  is a target RMSE and Cost is taken to be the measured wall-clock runtime. This is similar to the performance measurement approach used in These results were

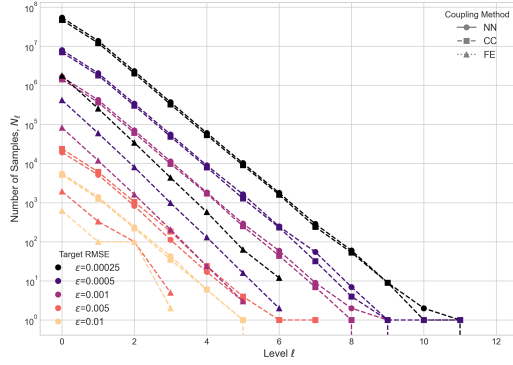
gathered using code written in C++ and used the standard C++11 Mersenne Twister PRNG [?]. All experiments were executed on Intel Gold 6538Y+ processors [?].



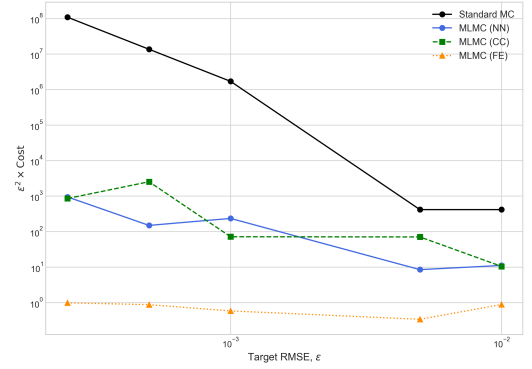
(a) First plot



(b) Second plot



(c) Third plot



(d) Fourth plot

Figure 4.1: A 2 by 2 grid of plots.

# Bibliography

- [1] Assyr Abdulle, Andrea Barth, and Christoph Schwab. Multilevel monte carlo methods for stochastic elliptic multiscale pdes. *Multiscale Modeling & Simulation*, 11(4):1033–1070, 2013.
- [2] Assyr Abdulle and Adrian Blumenthal. Stabilized multilevel monte carlo method for stiff stochastic differential equations. *Journal of Computational Physics*, 251:445–460, 2013.
- [3] Juan A Acebrón. A probabilistic linear solver based on a multilevel monte carlo method. *Journal of Scientific Computing*, 82(3):65, 2020.
- [4] Louis JM Aslett, Tigran Nagapetyan, and Sebastian J Vollmer. Multilevel monte carlo for reliability theory. *Reliability Engineering & System Safety*, 165:188–196, 2017.
- [5] Andrea Barth, Annika Lang, and Christoph Schwab. Multilevel monte carlo method for parabolic stochastic partial differential equations. *BIT Numerical Mathematics*, 53(1):3–27, 2013.
- [6] Federico Cornalba and Julian Fischer. Multilevel monte carlo methods for the dean–kawasaki equation from fluctuating hydrodynamics. *SIAM Journal on Numerical Analysis*, 63(1):262–287, 2025.
- [7] Intel Corporation. Intel xeon gold 6538y+ processor. <https://www.intel.com/content/www/us/en/products/sku/237563/intel-xeon-gold-6538y-processor-60m-cache-2-20-ghz/specifications.html>.
- [8] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*, volume 152. Cambridge university press, 2014.
- [9] Elsevier. Scopus. <https://www.scopus.com>, Accessed: August 11, 2025.

- [10] International Organization for Standardization. Iso/iec 14882:2020 - programming languages - c++, 2020.
- [11] Michael B Giles. Multilevel monte carlo path simulation. *Operations research*, 56(3):607–617, 2008.
- [12] Michael B Giles. Multilevel monte carlo methods. *Acta numerica*, 24:259–328, 2015.
- [13] Michael B Giles and Christoph Reisinger. Stochastic finite differences and multilevel monte carlo for a class of spdes in finance. *SIAM journal on financial mathematics*, 3(1):572–592, 2012.
- [14] Michael B Giles and Yuan Xia. Multilevel monte carlo for exponential lévy models. *Finance and Stochastics*, 21(4):995–1026, 2017.
- [15] Stefan Heinrich. Monte carlo complexity of global solution of integral equations. *Journal of complexity*, 14(2):151–175, 1998.
- [16] Leon Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139, 1918.
- [17] Ralf Kornhuber, Christoph Schwab, and Maren-Wanda Wolf. Multilevel monte carlo finite element methods for stochastic elliptic variational inequalities. *SIAM Journal on Numerical Analysis*, 52(3):1243–1268, 2014.
- [18] Sergey V Lototsky, Boris L Rozovsky, et al. *Stochastic partial differential equations*, volume 11. Springer, 2017.
- [19] Yan Luo and Zhu Wang. A multilevel monte carlo ensemble scheme for random parabolic pdes. *SIAM Journal on Scientific Computing*, 41(1):A622–A642, 2019.
- [20] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.
- [21] Étienne Pardoux et al. *Stochastic partial differential equations: An introduction*. Springer, 2021.
- [22] Chang-han Rhee and Peter W Glynn. Unbiased estimation with square root convergence for sde models. *Operations Research*, 63(5):1026–1043, 2015.

- [23] Gilbert Strang. *Computational science and engineering*, volume 791. Wellesley-Cambridge Press Wellesley, MA, 2007.
- [24] Walter A Strauss. *Partial differential equations: An introduction*. John Wiley & Sons, 2007.
- [25] Endre Süli. Numerical solutions to partial differential equations, September 2024.
- [26] Endre Süli. Lecture notes on finite element methods for partial differential equations, January 2025.
- [27] John B Walsh. An introduction to stochastic partial differential equations. In *École d'Été de Probabilités de Saint Flour XIV-1984*, pages 265–439. Springer, 2006.