Department of Mathematics, ETH Zürich

Multi-Level Monte Carlo Methods for Stochastic Parabolic and 2nd Order Hyperbolic PDEs

Master Thesis
by
D. Bernhardsgrütter

supervised by
Prof. Ch. Schwab

Zürich, February 25, 2011

Contents

1 Introduction

Diffusion is a physical process that describes the spread of particles from regions of higher concentrations to areas of lower concentrations. The mathematical formulation of this process dates back to 1855, when Adolf Fick introduced the particle diffusion equation in Fick's Second Law. In the domain $D \subset \mathbb{R}^d$, we consider the problem

$$\begin{aligned} &\tfrac{\partial u}{\partial t} - \operatorname{div}(a(x)\nabla u(t,x)) = f(t,x) && x \in D,\ 0 < t < T, \\ &u(t,x) = 0 && x \in \partial D,\ 0 < t < T, \\ &u(0,x) = u_0(x) && x \in D. \end{aligned}$$

Various numerical methods for approximating solutions of parabolic partial differential equations are available, cf. [13]. Those schemes have been designed for solving PDEs, where the diffusion coefficient a, source term f and initial condition u_0 are known exactly. However, for many problems of practical interest, no exact models determining the diffusion coefficient are available. Thus, measurements are needed to specify the parameter, which involves uncertainty. This results in a stochastic model for the diffusion coefficient rather than exact deterministic input data.
A consistent mathemtatical formulation of the diffusion equation with a random diffusion coefficient has been given in [4]. We focus here on elaborating an efficient numerical algorithm, that deals with uncertainity in the input data.
A widely used technique to treat randomness are Monte Carlo methods. Here, "sampling" entails the numerical solution of many deterministic parabolic PDEs. We use efficient numerical schemes available in the literature to compute each sample. It is well-known that MC type methods reduce the statistical error at rate $1/2$ as the number of samples is increased. This results in a high computational cost of this compound algorithm.
In order to improve efficiency, Multi-Level Monte Carlo methods have recently been proposed for the numerical solution of stochastic elliptic PDEs [1] and for hyperbolic conservation laws with random input data [7]. We adopt the underlying concepts for the solution of stochastic parabolic PDEs. The MLMC type methods make use of the fact, that computing a rough approximate solution sample is cheap on a coarse grid with a large timestep, whereas obtaining a good approximation on a fine grid with small timestep is expensive. By choosing the optimally distributed sample size on each grid, we prove that the contribution to the computational cost due to sampling is only logarithmic in the spatial degrees of freedom and thus negligible.
The same concepts may then be applied to the wave equation in random medium

$$\begin{aligned} &\tfrac{\partial^2 u}{\partial t^2} - \operatorname{div}(a\nabla u) = f && x \in D,\ 0 < t < T, \\ &u(t,x) = 0 && x \in \partial D,\ 0 < t < T, \\ &u(0,x) = u_0(x),\ \tfrac{\partial u}{\partial t}(0,x) = u_1(x) && x \in D. \end{aligned}$$

Again, the source of randomness lies in the coefficient a, whereas deterministic source term and initial condition are assumed. Only minor modifications of the algorithm are required.

The choice of the time stepping scheme needs to be adjusted to the second order evolution problem, we discuss Newmark's scheme here.
Numerical experiments of the stochastic wave equation in one space dimension confirm the theory.

2 Stochastic Diffusion Equation

Let $D \subset \mathbb{R}^d$, $d = 1, 2, 3$, be an open bounded Lipschitz polyhedron and let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. For $0 < T < \infty$ the bounded time interval is denoted by $I := (0, T)$. The space-time cylinder is denoted by $Q_T := I \times D$ and we set $\Sigma_T = I \times \partial D$. Consider the stochastic diffusion equation

$$\frac{\partial u}{\partial t} - \operatorname{div}(a(x,\omega)\nabla u) = f(t,x) \quad \text{in } Q_T,\ \omega \in \Omega, \tag{1}$$

$$u(t,x,\omega) = 0 \quad \text{on } \Sigma_T, \tag{2}$$

$$u(0,x,\omega) = u_0(x) \quad \text{in } D. \tag{3}$$

We assume deterministic source term and initial condition, and that the stochastic diffusion coefficient $a(x,\omega) : \Omega \to L^\infty(D)$ is measurable and independent of time t.

Assumption 2.1. There exist constants $0 < a_- \le a_+ < \infty$ such that

$$0 < a_- \le a(x,\omega) \le a_+ < \infty \quad \text{for all } x \in D, \omega \in \Omega. \tag{4}$$

2.1 Weak Form and Well-Posedness

We follow [4] to obtain the variational formulation of the stochastic diffusion equation (1) - (3). To this end, we set $V := H_0^1(D)$ and $H := L^2(D)$. Identifying H with its dual, we obtain the Gelfand triple $V \subset H \simeq H' \subset V'$. Let

$$\mathcal{X} := L^2(I;V) \cap H^1(I;V') \quad \text{and} \quad \mathcal{Y} := L^2(I;V) \times H, \tag{5}$$

equipped with the norms

$$\|u\|_{\mathcal{X}} := (\|u\|^2_{L^2(I;V)} + \|\tfrac{du}{dt}\|^2_{L^2(I;V')})^{1/2} \text{ and } \|(v_1,v_2)\|_{\mathcal{Y}} := (\|v_1\|^2_{L^2(I;V)} + \|v_2\|^2_H)^{1/2}. \tag{6}$$

It is well-known that $\mathcal{X} \hookrightarrow C([0,T];H)$ (see e.g. [6, Theorem 3.1]). Thus, for $w(t,x) \in \mathcal{X}$ the function $w(0,x) \in L^2(D)$ is well-defined and therefore the initial condition (3) is satisfied in the sense of $L^2(D)$.
For a given realization $\omega \in \Omega$, the weak solution $u(\cdot,\cdot,\omega) \in \mathcal{X}$ of the initial-boundary value problem (1) - (3) is characterized by

$$b(\omega; u, v) = l(v) \quad \text{for all } v = (v_1, v_2) \in \mathcal{Y}, \tag{7}$$

where

$$b(\omega; u, v) := \int_I (\tfrac{\partial u}{\partial t}, v_1)_H dt + \int_I \int_D a(x, \omega) \nabla u \nabla v_1 dx dt + (u(0, x), v_2)_H, \tag{8}$$

and

$$l(v) := \int_I (f, v_1)_H dt + (u_0(x), v_2)_H. \tag{9}$$

Assumption 2.1 ensures well-posedness of the weak form (7) for all $\omega \in \Omega$.

Proposition 2.2. *Under Assumption 2.1, for every $\omega \in \Omega$ and given $f \in L^2(I; V')$, $u_0 \in H$ the space-time variational formulation (7) of the stochastic diffusion equation (1) - (3) admits a unique solution $u(\cdot, \cdot, \omega) \in \mathcal{X}$. It holds*

$$\|u\|_{\mathcal{X}} \leq C(\|f(t, x)\|_{L^2(I;V')} + \|u_0\|_{L^2(D)}), \tag{10}$$

where C is bounded uniformly for all $\omega \in \Omega$.

Proof. The bilinear form $b(\omega; \cdot, \cdot) : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ induces a linear operator $B_\omega : \mathcal{X} \to \mathcal{Y}'$ through ${}_{\mathcal{Y}'}\langle Bw, v\rangle_{\mathcal{Y}} = b(\omega; w, v)$. We can formally write $u(\cdot, \cdot, \omega) = B_\omega^{-1}(f, u_0)$. Bounded invertibility of $B_\omega \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ is characterized in terms of three conditions on the associated bilinear form (see [2, Theorem 3.6]). A verification of those conditions for the bilinear form defined in (8) can be found in Appendix A of [11]. Inspecting the proof given there, one obtains that under Assumption 2.1 the norm of B_ω^{-1} is uniformly bounded for all $\omega \in \Omega$. □

We introduce the Bochner spaces $\underline{\mathcal{X}} := L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathcal{X})$ and $\underline{\mathcal{Y}} := L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathcal{Y})$. We then define the bilinear form $B(\cdot, \cdot) : \underline{\mathcal{X}} \times \underline{\mathcal{Y}} \to \mathbb{R}$ and the "load functional" $F(\cdot) : \underline{\mathcal{Y}} \to \mathbb{R}$ by

$$B(u, v) := \mathbb{E}[b(\omega; u, v)] \quad \text{and} \quad F(v) := \mathbb{E}[l(v)] \quad \text{for } u \in \underline{\mathcal{X}},\ v \in \underline{\mathcal{Y}}. \tag{11}$$

The variational formulation of problem (1) - (3) is then: Find $u \in \underline{\mathcal{X}}$ such that

$$B(u, v) = F(v) \quad \text{for all } v \in \underline{\mathcal{Y}}. \tag{12}$$

Theorem 2.3. *Under Assumption 2.1 for given $f \in L^2(I; V')$ and $u_0 \in H$, the variational formulation (12) of the stochastic diffusion equation (1) - (3) admits a unique solution $u \in \underline{\mathcal{X}}$. It holds the apriori estimate*

$$\|u\|_{\underline{\mathcal{X}}} \leq C(\|f(t, x)\|_{L^2(I;V')} + \|u_0\|_{L^2(D)}). \tag{13}$$

Proof. According to Proposition 2.2 there exists a solution $u \in L^\infty(\Omega;\mathcal{X}) \subset L^2(\Omega;\mathcal{X})$ satisfying (13). To show uniqueness, let $v(t,x,\omega) = (v_1(t,x)\varphi(\omega), v_2(x)\varphi(\omega)) \in \underline{\mathcal{Y}} \simeq L^2(\Omega)\otimes\mathcal{Y}$, where $\varphi \in L^2(\Omega,\mathcal{A},\mathbb{P})$. Then there holds

$$B(u,v) = \mathbb{E}[b(\omega;u,(v_1,v_2))\varphi(\omega)] = \mathbb{E}[l(v_1,v_2)\varphi(\omega)] = F(v).$$

As $\varphi \in L^2(\Omega,\mathcal{A},\mathbb{P})$ is arbitrary, it holds for $\mathbb{P}$-a.e. $\omega \in \Omega$

$$b(\omega;u,(v_1,v_2)) = l(v_1,v_2) \quad \text{for all } v = (v_1,v_2) \in \mathcal{Y}.$$

Thus, the solution $u(t,x,\omega) \in \mathcal{X}$ is unique by Proposition 2.2. □

2.2 Regularity of the Solution

For the numerical approximation we need additional regularity of the solution. To establish a regularity theory for the random solution under certain smoothness conditions on the input data is not within the scope of this master thesis. We therefore make the following

Assumption 2.4. We assume that the stochastic diffusion equation (1) - (3) admits a unique solution

$$u \in L^2(\Omega;C^r([0,T];H^{p+1}(D)) \cap C^{r+1}([0,T];L^2(D))) \tag{14}$$

for $p \geq 1$ and $r = 1,2$.

Note that we implicitly assume compatible initial condition $u_0(x) = u(0,x) \in H^{p+1}(D)$. We refer to [15] for regularity theory of deterministic parabolic PDEs. Regularity of the random solution $u \in L^2(\Omega;\mathcal{X})$ has been discussed in [4], however the smoothness in space and time derived therein is not sufficient for our purposes.

3 Numerical Approximation

For a given realization $\hat{a}(x) \in L^\infty(D)$ of the random diffusion coefficient we face the parabolic PDE

$$\frac{\partial u}{\partial t} - \operatorname{div}(\hat{a}(x)\nabla u) = f(t,x) \quad \text{in } Q_T, \tag{15}$$

$$u(t,x) = 0 \quad \text{on } \Sigma_T, \tag{16}$$

$$u(0,x) = u_0(x) \quad \text{in } D. \tag{17}$$

There are several numerical methods to solve this problem. We discuss a Finite Element method of order $p \geq 1$ for spatial approximation and the implicit Euler and Crank-Nicolson time stepping schemes here.

The stochastic solution $u(t,x,\omega)$ of the diffusion equation (1) - (3) is characterized by its moments. We estimate the mean field $\mathbb{E}[u]$ by a Monte Carlo method, i.e. by averaging over a large number of solution samples $\hat{u}^i, i = 1,\dots,M$ corresponding to the solution for M independent, identically distributed realizations $\hat{a}^i \in L^\infty(D)$ of the stochastic coefficient $a(x,\omega)$.
By combining these techniques, we pick up three contributions to the overall error induced by the spatial approximation, time stepping and an additional statistical error resulting from the Monte Carlo estimator. The equilibration of these errors by choosing the optimal sample size and timestep for a given grid size is crucial.
The computational cost is considerably reduced by using different sample sizes on different levels of spatial and temporal approximation without a significant loss of accuracy. The Multi-Level Monte Carlo method uses a number of samples which is inversely proportional to the finess of the grid and timestep. It has the same convergence rate as the single-level Monte Carlo method on the finest level, however the efficiency is improved.

4 Space Discretization

Let $\mathcal{T}_h$ be a regular simplicial mesh of the polyhedral domain $D \subset \mathbb{R}^d$ with meshwidth $h = \max_{K\in\mathcal{T}_h}\{\mathrm{diam}(K)\}$. We denote the space of continuous, piece-wise polynomial functions on the simplicial triangulation $\mathcal{T}_h$ by

$$\mathcal{S}_0^{p,1}(D,\mathcal{T}_h) := \{v \in H_0^1(D) : v|_K \in \mathcal{P}_p \text{ for all } K \in \mathcal{T}_h\},$$

where $\mathcal{P}_p(K) = \mathrm{span}\{x^\alpha; |\alpha| \le p\}$ are the polynomials of degree p on the simplex K.
In the following, we write $V_h := \mathcal{S}_0^{p,1}(D,\mathcal{T}_h)$ for the Finite Element space. Although the space $H_0^1(D)$ generally does not ensure continuity in more than one spatial dimension, we have $\mathcal{S}_0^{p,1}(D,\mathcal{T}_h) \subset C^0(D)$ (cf. [2, Theorem 5.2]).
For a given realization $\hat{a}(x) \in L^\infty(D)$ of the stochastic diffusion coefficient $a(x,\omega)$, the semidiscrete problem reads: For given $f \in L^2(I;H)$ and for an approximation $u_{0,h} \in V_h$ of the initial data $u_0 \in H$, find $u_h \in H^1(I;V_h)$ such that

$$u_h(0) = u_{0,h} \qquad \text{and} \tag{18}$$

$$(\tfrac{du_h}{dt}, v_h) + a(u_h, v_h) = (f(t,x), v_h) \quad \text{for all } v_h \in V_h. \tag{19}$$

Here, the bilinear form $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ is defined by $a(v,w) := \int_D \hat{a}(x)\nabla v \nabla w dx$ and $(\cdot,\cdot) : L^2(D) \times L^2(D) \to \mathbb{R}$ denotes the inner product $(u,v) = \int_D uv dx$.
Let $\{\Phi_j\}_{j=1}^{N_h}$ be a basis of V_h, where $N_h = \dim V_h$ is the number of degrees of freedom of the space discretization. Then $u_h \in H^1(I;V_h)$ admits the unique representation $u_h(t,x) = \sum_{j=1}^{N_h} u_j(t)\Phi_j(x)$ and the initial data can be written as $u_{0,h} = \sum_{j=1}^{N_h} u_j^0 \Phi_j$. The semidiscrete problem (18), (19) may then be reformulated as a system of ordinary differential equations

$$\boldsymbol{M}\dot{\underline{u}}(t) + \boldsymbol{A}\underline{u}(t) = \underline{f}(t), \quad \underline{u}(0) = \underline{u}_0. \tag{20}$$

Here, $\boldsymbol{M}_{ij} = (\Phi_i, \Phi_j)$ is the mass matrix, $\boldsymbol{A}_{ij} = a(\Phi_i, \Phi_j)$ is the stiffness matrix, $\underline{f}(t) = (f_j) = (f(t), \Phi_j)$ is the load vector, $\underline{u}_0 = (u_j^0)$ is the vector of coefficients of the initial data and $\underline{u} = (u_j)$ is the seeked vector of solution coefficients.
For the error analysis, we briefly review the theory of the corresponding stationary problem.

4.1 Analysis of the Stationary Problem

We consider the elliptic problem

$$-\mathrm{div}(\hat{a}\nabla u(x)) = f(x) \text{ in } D, \tag{21}$$

$$u = 0 \text{ on } \partial D. \tag{22}$$

As before, the diffusion coefficient $\hat{a} \in L^\infty$ is assumed to satisfy (4).
The weak formulation reads: Given $f \in H^{-1}(D)$ find $u \in H_0^1(D)$ such that

$$a(u, v) = (f, v) \text{ for all } v \in H_0^1. \tag{23}$$

We define the Ritz projection $R_h : V \to V_h$ as the orthogonal projection with respect to the bilinear form $a(\cdot, \cdot)$, i.e.

$$a(R_h u, v_h) = a(u, v_h) \text{ for all } v_h \in V_h. \tag{24}$$

The Ritz projection is quasi optimal,

$$\|u - R_h u\|_{H_0^1} \leq \frac{a_+}{a_-} \inf_{v_h \in V_h} \|u - v_h\|_{H_0^1}. \tag{25}$$

We have the following approximation property of the Finite Element spaces (see eg. [2]):

Proposition 4.1. *Let $D \subset \mathbb{R}^d$ be a polyhedron and let $\mathcal{T}_h$ be a regular, simplicial mesh with meshwidth h. Let $V_h = \mathcal{S}_0^{p,1}(D, \mathcal{T}_h)$ with $N_h = \dim V_h$. Given $u \in X_{s^*} = V \cap H^{1+s^*}(D)$, $s^* > 0$, it holds for $0 \leq s \leq \min\{s^*, p\}$*

$$\inf_{v_h \in V_h} \|u - v_h\|_V \leq C h^s \|u\|_{X_s} \sim C N_h^{-s/d} \|u\|_{X_s}. \tag{26}$$

Thus, for $u \in H^{p+1}(D)$, we have

$$\|u - R_h u\|_{H_0^1(D)} \leq C h^p \|u\|_{H^{p+1}(D)}.$$

Later, in the parabolic setting, we estimate the error in the L^2-norm. We obtain a scheme of order $p+1$ in the L^2-norm for the elliptic problem by a duality argument (cf. [2, Theorem II.7.6]). To this end, we first recapitulate regularity of the solution of the elliptic problem.

Theorem 4.2. *For $f \in L^2(D)$ and $\hat{a} \in W^{1,\infty}(D)$ as in (4), the variational formulation (23) admits a unique solution $u \in W$, where*

$$W := \{w \in H_0^1(D); \Delta w \in L^2(D)\}$$

is equipped with the norm $\|w\|_W := \|\Delta w\|_{L^2(D)} + \|w\|_{L^2(D)}$. There holds the a-priori estimate

$$\|u\|_W \leq C(\hat{a})\|f\|_{L^2(D)}, \tag{27}$$

where $C(\hat{a})$ depends on a_-, a_+ and $\|\hat{a}\|_{W^{1,\infty}(D)}$.

Proof. Existence and uniqueness of the solution is a consequence of the Lax-Milgram Lemma. The regularity of the solution is a consequence of the $W^{1,\infty}$-regularity of the diffusion coefficient and the regularity of the data f, there holds

$$-\hat{a}\Delta u = f + \nabla\hat{a}\nabla u \quad \text{in } L^2(D).$$

Therefore we may estimate

$$\|\Delta u\|_{L^2(D)} \leq C(\hat{a}) \left(\|f\|_{L^2(D)} + \|\nabla u\|_{L^2(D)}\right) \leq C(\hat{a}) \|f\|_{L^2(D)}.$$

□

Remark 4.3. Generally, it holds $W \supsetneq H^2 \cap H_0^1(D)$ for polyhedral domains D. The corners spoil global $H^2(D)$ regularity of the solution, it merely holds $u \in H^2_{\text{loc}}(D)$. However, for convex domains we have $W = H^2 \cap H_0^1(D)$, cf. [3, Theorem 2.4.3, Theorem 2.6.3].
In view of Proposition 4.1, the Riesz projection defines an approximation of first order in the H^1-norm in convex polyhedral domains D. The situation is more delicate for nonconvex domains, as singularities at the nonconvex corners spoil linear convergence rates on quasiuniform meshes. However, one can restore $O(h)$ convergence by using meshes which are suitably refined towards the nonconvex corncers. The two-dimensional case with one nonconvex corner has been exemplified in [13, Lemma 19.10].

We can now prove, that the Ritz projection admits an error of order $p+1$ in the L^2-norm.

Proposition 4.4. *Assume $u \in H^{p+1}(D)$, $\hat{a} \in W^{1,\infty}(D)$ as in (4) and a suitably refined mesh $\mathcal{T}_h$ as in Remark 4.3. Then the Ritz projection admits the error*

$$\|u - R_h u\|_{L^2(D)} \leq Ch^{p+1}\|u\|_{H^{p+1}(D)}, \tag{28}$$

where C depends on a_-, a_+ and $\|\hat{a}\|_{W^{1,\infty}(D)}$.

Proof. In view of (25) and the approximation property (26) we have $\|u - R_h u\|_{H^1(D)} \leq Ch^p \|u\|_{H^{p+1}(D)}$. Let $\varphi \in L^2(D)$ and let $\psi \in W$ be the solution of

$$-\mathrm{div}(\hat{a}\nabla\psi) = \varphi \text{ in } D, \quad \psi = 0 \text{ on } \partial D. \tag{29}$$

For arbitrary $\psi_h \in V_h$ there holds

$$\begin{aligned} (u_h - u, \varphi) &= -(u_h - u, \mathrm{div}(\hat{a}\nabla\psi)) = a(u_h - u, \psi) = a(u_h - u, \psi - \psi_h) \\ &\leq a_+ \|\nabla(u_h - u)\|_{L^2} \|\nabla(\psi - \psi_h)\|_{L^2}. \end{aligned} \tag{30}$$

Taking the infimum over $\psi_h \in V_h$ we obtain by Proposition 4.1, Theorem 4.2 and Remark 4.3

$$(R_h u - u, \varphi) \leq Ch^p \|u\|_{H^{p+1}(D)} h \|\psi\|_W \leq Ch^{p+1} \|u\|_{H^{p+1}(D)} \|\varphi\|_{L^2(D)}. \tag{31}$$

Choosing $\varphi = u_h - u$ implies the assertion. □

We proceed with the error analysis of the parabolic problem.

4.2 Error Analysis of the Semidiscrete Problem

We return to the parabolic setting: Let u_h be the solution of the semidiscrete problem (18), (19) and let u be the solution of (15) - (17). We have the following error bound.

Theorem 4.5. *Under Assumption 2.4, for $\hat{a} \in W^{1,\infty}(D)$ satisfying (4) and for $u_{0,h} := R_h u_0 \in V_h$, the solution u_h of the semidiscrete problem admits the error bound*

$$\|u_h(t) - u(t)\|_{L^2(D)} \leq C(\hat{a}) h^{p+1} \left(\|u(t)\|_{H^{p+1}(D)} + \int_0^t \left\| \tfrac{\partial u(s)}{\partial t} \right\|_{H^{p+1}(D)} ds \right). \tag{32}$$

Here, the constant $C(\hat{a})$ depends on a_-, a_+ and $\|\hat{a}\|_{W^{1,\infty}(D)}$.

Proof. We follow closely the proof of [13, Theorem 1.2], where a constant diffusion coefficient $a(x) \equiv 1$ and smooth boundaries $\partial D \in C^\infty$ have been assumed. The dependencies on the diffusion coefficient are important for the analysis of the random PDE later on.
We split the error $e_h(t) := u(t) - u_h(t)$ as follows

$$e_h(t) = (u(t) - R_h u(t)) + (R_h u(t) - u_h(t)) =: \eta(t) + \xi(t), \tag{33}$$

where $\eta(t) = u(t) - R_h u(t)$ and $\xi(t) = R_h u(t) - u_h(t)$.
Both terms can be bounded seperately. The first term $\eta(t)$ is for every t the error of the corresponding elliptic problem. We have by Proposition 4.4

$$\|\eta(t)\|_{L^2(D)} = \|u(t) - R_h u(t)\|_{L^2(D)} \leq Ch^{p+1} \|u(t)\|_{H^{p+1}(D)}. \tag{34}$$

In order to bound $\xi(t)$, observe that for $\chi \in V_h$, $t > 0$ holds

$$\begin{aligned}(\tfrac{\partial \xi}{\partial t}, \chi) + a(\xi, \chi) &= (R_h \tfrac{\partial u}{\partial t}, \chi) - (\tfrac{\partial u_h}{\partial t}, \chi) + a(R_h \tfrac{\partial u}{\partial t}, \chi) - a(\tfrac{\partial u_h}{\partial t}, \chi) \\ &= -(f, \chi) + (R_h \tfrac{\partial u}{\partial t}, \chi) + a(R_h \tfrac{\partial u}{\partial t}, \chi) \\ &= -(f, \chi) + (R_h \tfrac{\partial u}{\partial t}, \chi) + a(\tfrac{\partial u}{\partial t}, \chi) \\ &= (R_h \tfrac{\partial u}{\partial t} - \tfrac{\partial u}{\partial t}, \chi) = -(\tfrac{\partial \eta}{\partial t}, \chi).\end{aligned}$$

Thus,

$$(\tfrac{\partial \xi}{\partial t}, \xi) + a(\xi, \xi) = -(\tfrac{\partial \eta}{\partial t}, \xi) \tag{35}$$

and with coercivity of the billinear form $a(\cdot, \cdot)$ follows

$$\tfrac{1}{2} \tfrac{d}{dt} \|\xi\|^2_{L^2(D)} = \|\xi\|_{L^2(D)} \tfrac{d}{dt} \|\xi\|_{L^2(D)} \leq \left\| \tfrac{\partial \eta}{\partial t} \right\|_{L^2(D)} \|\xi\|_{L^2(D)} . \tag{36}$$

We obtain with $\xi(0) = 0$

$$\|\xi(t)\|_{L^2(D)} \leq \int_0^t \left\| \tfrac{\partial \eta}{\partial t} \right\|_{L^2(D)} ds \tag{37}$$

and therefore the desired estimate. □

Remark 4.6. We can choose the approximate initial condition $u_{0,h} := R_h u_0$ as we assumed compatible initial conditions in (14).

Remark 4.7. The regularity of the diffusion coefficient $\hat{a} \in W^{1,\infty}$ is merely needed for the approximation order $p + 1$, cf. Proposition 4.4. If it was dropped, we would obtain a slightly lower convergence rate $O(h^p)$ by Proposition 4.1. However, in view of Assumption 2.4, the assumed regularity of the diffusion coefficient does not seem to be an additional constraint.

5 Time Discretization

We need to solve the initial value problem (20) numerically. We discuss two common time stepping schemes for solving the system of ordinary differential equations.

5.1 Implicit Euler

For $N_t \in \mathbb{N}$ we define the time step

$$k := \frac{T}{N_t}$$

and $t_m := mk$, $m = 0, \ldots, N_t$. We obtain the fully discrete implicid Euler method by replacing the time derivative in (19) by a backward difference quotient: For given $f \in$

$C^0([0,T];H)$ and for an approximation $u_{0,h} \in V_h$ of the initial data $u_0 \in H$, find $u_h^m \in V_h$ such that for $m = 1, \ldots, N_t$ and for all $v_h \in V_h$

$$u_h^0 = u_{0,h} \qquad \text{and} \tag{38}$$

$$\left(\tfrac{u_h^m - u_h^{m-1}}{k}, v_h\right) + a\left(u_h^m, v_h\right) = \left(f(t_m), v_h\right). \tag{39}$$

As for the semidiscrete situation (20), using a basis of the FE space V_h we can rewrite the fully discrete problem as

$$(\boldsymbol{M} + k\boldsymbol{A})\underline{u}^m = \boldsymbol{M}\underline{u}^{m-1} + k\underline{f}(t_m), \quad \underline{u}^0 = \underline{u}_0. \tag{40}$$

It is well-known that the Euler method is of first order accuracy in time. We have the following error estimate.

Theorem 5.1. *Under Assumption 2.4, for $\hat{a} \in W^{1,\infty}(D)$ satisfying (4), $f \in C^0([0,T];H)$ and for $u_{0,h} := R_h u_0 \in V_h$, there holds for $m \geq 1$*

$$\begin{aligned} \|u_h^m - u(t_m)\|_{L^2(D)} \;\leq\; & C(\hat{a})h^{p+1}\left(\|u(t_m)\|_{H^{p+1}(D)} + \int_0^{t_m}\left\|\tfrac{\partial u(s)}{\partial t}\right\|_{H^{p+1}(D)} ds\right) \\ & + k\int_0^{t_m}\left\|\tfrac{\partial^2 u}{\partial t^2}\right\|_{L^2(D)} ds. \end{aligned} \tag{41}$$

The constant $C(\hat{a})$ depends on a_-, a_+ and $\|\hat{a}\|_{W^{1,\infty}(D)}$.

Proof. We follow closely the proof of [13, Theorem 1.5], considering a non-constant diffusion coefficient a and a polyhedral domain D. It is important to know the exact dependencies on the diffusion coefficient in order to obtain uniform bounds in the fully random case later on.
We rewrite the error as

$$u_h^m - u(t_m) = (u_h^m - R_h u(t_m)) + (R_h u(t_m) - u(t_m)) = \xi^m + \eta^m. \tag{42}$$

As in the proof of Theorem 4.5 we realize that η^m is a best approximation error and can be bounded according to Proposition 4.4,

$$\|\eta^m\|_{L^2(D)} \leq Ch^{p+1}\|u(t_m)\|_{H^{p+1}(D)}. \tag{43}$$

A straightforward calculation yields

$$\left(\tfrac{\xi^m - \xi^{m-1}}{k}, v_h\right) + a\left(\xi^m, v_h\right) = -\left(\omega^m, v_h\right) \qquad \text{for all } v_h \in V_h \text{ and } m \geq 1, \tag{44}$$

where

$$\omega^m = R_h\tfrac{u(t_m)-u(t_{m-1})}{k} - \tfrac{\partial u}{\partial t}(t_m) = (R_h - I)\tfrac{u(t_m)-u(t_{m-1})}{k} + \left(\tfrac{u(t_m)-u(t_{m-1})}{k} - \tfrac{\partial u}{\partial t}(t_m)\right) = \omega_1^m + \omega_2^m.$$

Choosing $v_h = \xi^m$ in (44) yields

$$\|\xi^m\|_{L^2(D)} \leq \left\|\xi^{m-1}\right\|_{L^2(D)} + k\left\|\omega^m\right\|_{L^2(D)},$$

and with $\xi^0 = 0$ we obtain

$$\|\xi^m\|_{L^2(D)} \leq k\sum_{j=1}^{m}\left\|\omega_1^j\right\|_{L^2(D)} + k\sum_{j=1}^{m}\left\|\omega_2^j\right\|_{L^2(D)}.$$

Observe that

$$\omega_1^j = (R_h - I)k^{-1}\int_{t_{j-1}}^{t_j}\frac{\partial u}{\partial t}ds = k^{-1}\int_{t_{j-1}}^{t_j}(R_h - I)\frac{\partial u}{\partial t}ds,$$

which yields

$$k\sum_{j=1}^{m}\left\|\omega_1^j\right\|_{L^2(D)} \leq \sum_{j=1}^{m}\int_{t_{j-1}}^{t_j} Ch^{p+1}\left\|\frac{\partial u}{\partial t}\right\|_{H^{p+1}(D)}ds = Ch^{p+1}\int_0^{t_m}\left\|\frac{\partial u(s)}{\partial t}\right\|_{H^{p+1}(D)}ds.$$

Analogously, we rewrite

$$k\omega_2^j = u(t_j) - u(t_{j-1}) - k\frac{\partial u}{\partial t}(t_j) = -\int_{t_{j-1}}^{t_j}(s - t_{j-1})\frac{\partial^2 u}{\partial t^2}ds$$

so that

$$k\sum_{j=1}^{m}\left\|\omega_1^j\right\|_{L^2(D)} \leq \sum_{j=1}^{m}\left\|\int_{t_{j-1}}^{t_j}(s - t_{j-1})\frac{\partial^2 u}{\partial t^2}ds\right\|_{L^2(D)} \leq k\int_0^{t_m}\left\|\frac{\partial^2 u}{\partial t^2}\right\|_{L^2(D)}ds.$$

□

5.2 Crank-Nicolson Scheme

Discretizing the derivative in a symmetric way around the midpoint $t_{m-1/2} = (m - \frac{1}{2})k$ results in a second order method in time. The Crank-Nicolson scheme reads: For an approximation $u_{0,h} \in V_h$ of the initial data $u_0 \in H$, find $u_h^m \in V_h$ such that for $m = 1, \ldots, N_t$ and for all $v_h \in V_h$

$$u_h^0 = u_{0,h} \qquad \text{and} \tag{45}$$

$$\left(\frac{u_h^m - u_h^{m-1}}{k}, v_h\right) + a\left(\tfrac{1}{2}(u_h^m - u_h^{m-1}), v_h\right) = \left(f(t_{m-1/2}), v_h\right). \tag{46}$$

We rewrite the fully discrete problem in matrix form as

$$(\boldsymbol{M} + \tfrac{1}{2}k\boldsymbol{A})\underline{u}^m = (\boldsymbol{M} - \tfrac{1}{2}k\boldsymbol{A})\underline{u}^{m-1} + k\underline{f}(t_{m-1/2}), \quad \underline{u}^0 = \underline{u}_0. \tag{47}$$

It is well-known that the Crank-Nicolson method is second order accurate in time. We have the following error estimate.

Theorem 5.2. *Under Assumption 2.4, for $\hat{a} \in W^{1,\infty}(D)$ satisfying (4), $f \in C^0([0,T];H)$ and for $u_{0,h} := R_h u_0 \in V_h$, there holds for $m \geq 1$*

$$\begin{aligned} \|u_h^m - u(t_m)\|_{L^2(D)} \quad &\leq \quad Ch^{p+1}\left(\|u(t_m)\|_{H^{p+1}(D)} + \int_0^{t_m} \left\|\tfrac{\partial u(s)}{\partial t}\right\|_{H^{p+1}(D)} ds\right) \\ &\quad +Ck^2 \int_0^{t_m} (\left\|\tfrac{\partial^3 u}{\partial t^3}\right\|_{L^2(D)} + \left\|\tfrac{\partial^2 u}{\partial t^2}\right\|_W) ds. \end{aligned} \tag{48}$$

The constant C depends on a_-, a_+ and $\|\hat{a}\|_{W^{1,\infty}(D)}$.

Proof. The proof of [13, Theorem 1.6] may be adapted in the same spirit as in the proof of Theorem 5.1. □

Remark 5.3. The implicit Euler and Crank-Nicolson timetepping schemes are both unconditionally stable. However, the Crank-Nicolson is the "limiting case" of unconditional stability, which may cause artificial oscillations of the approximate solutions.

6 Monte Carlo Method

We are interested in the approximation of the statistical moments of the solution $u \in L^2(\Omega;\mathcal{X})$. We will here focus on estimating the mean field. For M independent, identically distributed realizations $\hat{a}^i$ of the diffusion coefficient, we denote by ${}^i\hat{u}$ the corresponding (exact) solution samples. The Monte Carlo estimator of $\mathbb{E}[u] \in \mathcal{X}$ is then given by

$$E_M[u] := \frac{1}{M}\sum_{i=1}^{M} {}^i\hat{u}. \tag{49}$$

We have the following standard result for the purely statistical error resulting from this Monte Carlo estimator.

Lemma 6.1. *For any $M \in \mathbb{N}$ and for $u \in L^2(\Omega;\mathcal{X})$, $0 < t < T$ holds*

$$\|\mathbb{E}[u(t)] - E_M[u(t)]\|_{L^2(\Omega;L^2(D))} \leq M^{-1/2}\,\|u(t)\|_{L^2(\Omega;L^2(D))}\,. \tag{50}$$

Proof. As we have the continuous embedding $\mathcal{X} \subset C([0,T];L^2(D))$, the solution $u(t,\cdot,\cdot)$ defines a random field on $L^2(D)$. Independence of the samples is preserved under the

continuous embedding and it follows

$$\begin{aligned}
\|\mathbb{E}[u(t)] - E_M[u(t)]\|^2_{L^2(\Omega;L^2(D))} &= \mathbb{E}\Big[\|\mathbb{E}[u(t)] - \frac{1}{M}\sum_{i=1}^{M} {}^i\hat{u}(t)\|^2_{L^2(D)}\Big] \\
&= \frac{1}{M^2}\sum_{i=1}^{M}\mathbb{E}\big[\|\mathbb{E}[u(t)] - {}^i\hat{u}(t)\|^2_{L^2(D)}\big] \\
&= \frac{1}{M}\mathbb{E}\big[\|\mathbb{E}[u(t)] - u(t)\|^2_{L^2(D)}\big] \\
&= \frac{1}{M}(\mathbb{E}\|u(t)\|^2_{L^2(D)} - \|\mathbb{E}[u(t)]\|^2_{L^2(D)}) \\
&\leq \frac{1}{M}\|u(t)\|^2_{L^2(\Omega;L^2(D))}.
\end{aligned}$$

□

6.1 Single-Level Monte Carlo Method

For the sake of actual computations, we need to replace the exact solution samples by some approximation. We therefore choose the spatial approximation as discussed in Section 4 combined with one of the time stepping schemes discussed in Section 5. On a given mesh $\mathcal{T}_h$ with meshwidth h and a fixed timestep k, for a given realization $\hat{a}^i \in W^{1,\infty}$ of the stochastic diffusion coefficient let ${}^i\hat{u}^m_{h,k}$ denote the approximation of the solution sample ${}^i\hat{u}(t_m)$ at time $t_m = mk$ according to (40) or (47), respectively. We estimate the mean field of the solution at time t_m by

$$E_M[u^m_{h,k}] := \frac{1}{M}\sum_{i=1}^{M} {}^i\hat{u}^m_{h,k}. \tag{51}$$

There are three contributions to the overall error: The spatial approximation, time stepping and the purely statistical error from the Monte Carlo estimator. We have the following error bound.

Theorem 6.2. *Assume* (4), (14) *and let* $a(\omega,\cdot) \in W^{1,\infty}(D)$ *for every* $\omega \in \Omega$. *Then the Monte Carlo estimator* (51) *with solution samples given by* (39) *for* $r = 1$ *and* (46) *for* $r = 2$, *respectively, admits the error bound*

(52)
$$\begin{aligned}
\Big\|\mathbb{E}[u(t_m)] - E_M[u^m_{h,k}]\Big\|_{L^2(\Omega;L^2(D))} &\leq M^{-1/2}\|u(t_m)\|_{L^2(\Omega;L^2(D))} \\
&\quad + C(h^{p+1} + k^r)\|u\|_{L^2(\Omega;C^r([0,T];H^{p+1}(D))\cap C^{r+1}([0,T];L^2(D)))}.
\end{aligned}$$

The constant C *depends on* a_-, a_+ *and* $\|a(x,\omega)\|_{L^\infty(\Omega;W^{1,\infty}(D))}$.

Proof. We split the error as follows

$$\begin{aligned}\left\|\mathbb{E}[u(t_m)] - E_M[u_{h,k}^m]\right\|_{L^2(\Omega;L^2(D))} &\leq \|\mathbb{E}[u(t_m)] - E_M[u(t_m)]\|_{L^2(\Omega;L^2(D))} \\ &\quad + \left\|E_M[u(t_m) - u_{h,k}^m]\right\|_{L^2(\Omega;L^2(D))} =: I + II\end{aligned}$$

The first term is a pure statistical error and is estimated with Lemma 6.1,

$$I \leq M^{-1/2}\|u(t_m)\|_{L^2(\Omega;L^2(D))}.$$

The second term is a discretization error and bounded by the theory for deterministic parabolic PDEs. Using Theorem 5.1 or Theorem 5.2 respectively, we obtain

$$\begin{aligned}II &\leq M^{-1}\sum_{i=1}^{M}\left\|{}^i\hat{u}(t_m) - {}^i\hat{u}_{h,k}^m\right\|_{L^2(\Omega;L^2(D))} \\ &\leq C(h^{p+1} + k^r)\|u\|_{L^2(\Omega;C^r([0,T];H^{p+1}(D))\cap C^{r+1}([0,T];L^2(D)))}.\end{aligned}$$

□

We obtain optimal convergence rates, when the errors are equilibrated. On a given grid $\mathcal{T}_h$ with meshwidth h, we choose a stepsize of the order $k = O(h^{(p+1)/r})$ for time stepping. The optimal choice of sample size is then given by $M = O(h^{-2(p+1)})$.
To compute the Monte Carlo estimator (51), we have to solve the initial value problem for M samples. The computation of each approximate solution sample corresponds to solving an elliptic problem for each timestep, cf. the fully discrete problem (40) or (47). We make the following assumption on the elliptic solver:

Assumption 6.3. For each timestep t_l, $l = 1, \ldots, m$, an approximate solution $\tilde{u}^l$ of the deterministic elliptic problem (40) or (47) is obtained in $O(N_h^\gamma)$ work, $1 \leq \gamma \leq 3$, such that the exact solution $\underline{u}^m$ of the fully discrete problem at final time t_m is approximated to order $O(h^{p+1} + k^r)$ in L^2-norm, i.e.

$$\|u_h^m - \tilde{u}_h^m\|_{L^2(D)} \leq C(h^{p+1} + k^r).$$

Clearly, this results in an approximation of the exact solution at time t_m of the same order,

$$\|u(t_m) - \tilde{u}_h^m\|_{L^2(D)} \leq \|u(t_m) - u_h^m\|_{L^2(D)} + \|u_h^m - \tilde{u}_h^m\|_{L^2(D)} \leq C(h^{p+1} + k^r).$$

There are several solution techniques which satisfy Assumption 6.3. The Cholesky decomposition of the left-hand side in (40) or (47) is computed in $O(N_h^3)$ work. For constant timesteps, the Cholesky decomposition is computed only once. However, forward and backward substitution needs to be performed for each timestep, which is of quadratic complexity. Drawbacks are the very high computational cost and the loss of sparsity in the Cholesky

factorization.
There is no need to solve the fully discrete equations exactly. As the solution of (40) or (47) is an approximation of the exact solution, it is sufficient to find an approximation of the same order. Iterative schemes to approximate $\underline{u}^m$ can thus be exploited. We refer to [2] for a detailed analysis of the Multigrid and preconditioned conjugate gradients method. As these techniques admit an error in every timestep, the amplification of these errors in time needs to be analyzed carefully.

Under Assumption 6.3, the computational cost of the Monte Carlo estimator (51) is thus given by $O(M \cdot N_t \cdot N_h^\gamma) = O(N_h^{\gamma+(p+1)/(rd)+2(p+1)/d})$. The contribution due to Monte Carlo sampling to the overall complexity can be reduced, as we will see in the following.

6.2 Multi-Level Monte Carlo Method

For the single-level Monte Carlo estimator (51) the same mesh $\mathcal{T}_h$ and a fixed timestep k is used for all samples. We can considerably improve the efficiency of the algorithm by using different sample sizes on different levels of spatial approximation and for different timesteps.
Solving the fully discrete problem is cheap on a coarse grid. Equilibrating the error of the spatial resolution and of the time stepping scheme, we use a large timestep on a coarse grid. Therefore, small systems for a small number of timesteps need to be solved and we can use a large sampling size M. However, accuracy of every approximate solution sample is relatively poor. On the other hand, we get fast convergence on a fine grid. But we need a small timestep in order to avoid a dominating error from temporal approximation. Thus, we need to solve a large system for many timesteps and we use a reduced number of samples in this case.

By $\{\mathcal{T}_l\}_{l=0}^\infty$ we denote a sequence of regular simplicial meshes obtained by uniform mesh refinement. For $l \geq 0$ the meshwidth of $\mathcal{T}_l$ is then

$$h_l := \max_{K \in \mathcal{T}_l} \{\mathrm{diam}(K)\} = 2^{-l} h_0.$$

Let $V_l := \mathcal{S}_0^{p,1}(D, \mathcal{T}_l)$ be the Finite Element space with $N_l^x := \mathrm{dim} V_l = O(2^{dl})$ degrees of freedom. We introduce a level-dependent timestep $k_l = T/N_l^t$. In order to equilibrate the spatial error and time stepping error, we choose $k_l = O(h_l^{(p+1)/r}) = O(2^{-l(p+1)/r})$, cf. Theorem 5.1, Theorem 5.2. The number of timesteps is then $N_l^t = O(h_l^{-(p+1)/r}) = O(2^{l(p+1)/r}) = O(N_l^{x\,(p+1)/(rd)})$. It is a measure for the computational cost of the timestepping scheme.
We then denote by $u_l := u_{h_l,k_l}^{m_l} \in V_l$ the approximation of $u(t_m)$ according to (40) or (47) on mesh $\mathcal{T}_l$ with timestep k_l at time $t_m = m_l k_l$, $l \geq 1$.

As we assumed a nested family of meshes $\{\mathcal{T}_l\}_{l=0}^{\infty}$, we obtain a hierarchic sequence of finite dimensional subspaces

$$V_0 \subset V_1 \subset \ldots \subset V_l \subset \ldots \subset H_0^1(D).$$

We may thus write the approximate solution at time t_m as

$$u_L = \sum_{l=1}^{L} (u_l - u_{l-1}),$$

where $u_0 := 0$ and $u_l - u_{l-1} \in V_l$. By linearity of the expectation, we obtain

$$\mathbb{E}[u_L] = \sum_{l=1}^{L} \mathbb{E}[u_l - u_{l-1}].$$

We estimate each addend $\mathbb{E}[u_l - u_{l-1}]$ by a level dependent number of samples M_l. The Multi-Level Monte Carlo estimator is then given by

$$E^L[u(t_m)] := \sum_{l=1}^{L} E_{M_l}[u_l - u_{l-1}] = \sum_{l=1}^{L} E_{M_l}[u_l] - E_{M_l}[u_{l-1}]. \tag{53}$$

We have the following error estimate.

Lemma 6.4. *Under Assumption 2.4 and for $a(\omega, \cdot) \in W^{1,\infty}(D)$ satisfying (4), the MLMC estimator (53) admits the error bound*

(54)

$$\begin{aligned} \left\|\mathbb{E}[u(t_m)] - E^L[u(t_m)]\right\|_{L^2(\Omega;L^2(D))} \quad &\leq \quad C\left(h_L^{p+1} + \textstyle\sum_{l=1}^{L} h_l^{p+1} M_l^{-1/2}\right) \\ &\quad \times \|u\|_{L^2(\Omega;C^r([0,T];H^{p+1}(D))\cap C^{r+1}([0,T];L^2(D)))}. \end{aligned}$$

Here, the constant C depends on p, a_-, a_+ and $\|a(x,\omega)\|_{L^\infty(\Omega;W^{1,\infty}(D))}$.

Proof. We rewrite the error as

$$\begin{aligned} \left\|\mathbb{E}[u(t_m)] - E^L[u(t_m)]\right\|_{L^2(\Omega;L^2(D))} &= \left\|\mathbb{E}[u(t_m)] - \mathbb{E}[u_L] + \mathbb{E}[u_L] - E^L[u(t_m)]\right\|_{L^2(\Omega;L^2)} \\ &\leq \|\mathbb{E}[u(t_m)] - \mathbb{E}[u_L]\|_{L^2(\Omega;L^2(D))} + \sum_{l=1}^{L} \|\mathbb{E}[u_l - u_{l-1}] - E_{M_l}[u_l - u_{l-1}]\|_{L^2(\Omega;L^2(D))} \\ &=: I + II. \end{aligned}$$

We estimate each term separately. The first term is bounded by Jensen's inequality, Theorem 5.1, Theorem 5.2 and Hölder's inequality,

$$I \leq \mathbb{E}[\|u(t_m) - u_L\|_{L^2(D)}] \leq C(h_L^{p+1} + k_L^r)\|u\|_{L^2(\Omega;C^r([0,T];H^{p+1}(D))\cap C^{r+1}([0,T];L^2(D)))},$$

where C depends on a_-, a_+ and $\|a(x,\omega)\|_{L^\infty(\Omega;W^{1,\infty}(D))}$.
For the second term, we estimate each addend as follows:

$$\begin{aligned}
\|\mathbb{E}[u_l - u_{l-1}] - E_{M_l}[u_l - u_{l-1}]\|_{L^2(\Omega;L^2(D))} &\leq M_l^{-1/2}\,\|u_l - u_{l-1}\|_{L^2(\Omega;L^2(D))} \\
&\leq M_l^{-1/2}\big(\,\|u(t_m) - u_l\|_{L^2(\Omega;L^2(D))} + \|u(t_m) - u_{l-1}\|_{L^2(\Omega;L^2(D))}\,\big) \\
&\leq C(a) M_l^{-1/2}(h_l^{p+1} + k_l^r + h_{l-1}^{p+1} + k_{l-1}^r)\|u\|_{L^2(\Omega;C^r([0,T];H^{p+1}(D))\cap C^{r+1}([0,T];L^2(D)))} \\
&\leq (2^{p+1}+1) C(a) M_l^{-1/2}(h_l^{p+1} + k_l^r)\|u\|_{L^2(\Omega;C^r([0,T];H^{p+1}(D))\cap C^{r+1}([0,T];L^2(D)))}.
\end{aligned}$$

Here we used Lemma 6.1, Theorem 5.1 and Theorem 5.2. The constant depends on a_-, a_+ and $\|a(x,\omega)\|_{L^\infty(\Omega;W^{1,\infty}(D))}$. A summation over $l = 1,\dots,L$ and observing that $k_l = O(h_l^{(p+1)/r})$ completes the proof. □

The preceding error estimate for the MLMC approximation holds for any distribution $\{M_l\}_{l=1}^L$ of samples over mesh levels. By optimizing the ratio between sample size and meshlevel, we recover the same convergence rate as single-level MC on the finest level. However, the computational cost is only a fraction of the latter.

Theorem 6.5. *Under Assumptions 2.4 and 6.3, for $a(\omega,\cdot) \in W^{1,\infty}(D)$ satisfying (4), the MLMC estimator (53) of the expectation of the solution of the stochastic parabolic PDE (1) - (3) computed with M_l samples on mesh level h_l and timestep $k_l = O(h_l^{(p+1)/r})$, where the samples are distributed as*

$$M_l = l^{2+\epsilon} 2^{2(L-l)(p+1)}, \quad l = 1, 2, \dots, L,$$

with some $\epsilon > 0$ arbitrarily small, admits the error bound

$$\big\|\mathbb{E}[u(t_m)] - E^L[u(t_m)]\big\|_{L^2(\Omega;L^2(D))} \leq C h_L^{p+1} \|u\|_{L^2(\Omega;C^r([0,T];H^{p+1}(D))\cap C^{r+1}([0,T];L^2(D)))}.$$

Here, the constant C depends on ϵ, p, a_-, a_+ and $\|a(x,\omega)\|_{L^\infty(\Omega;W^{1,\infty}(D))}$. The computational cost is bounded by

$$\mathit{Work}(L) \leq C(\delta) \begin{cases} (N_L^x)^\gamma N_L^t (\log N_L^x)^{2+\epsilon} & \text{if } \delta < 0, \\ (N_L^x)^\gamma N_L^t (\log N_L^x)^{3+\epsilon} & \text{if } \delta = 0, \\ (N_L^x)^{\gamma+\delta/d} N_L^t (\log N_L^x)^{2+\epsilon} & \text{if } \delta > 0, \end{cases}$$

where $\delta = (p+1)(2 - 1/r) - d\gamma$.

Remark 6.6. We mention again that the regularity assumption $a(\omega,\cdot) \in W^{1,\infty}(D)$ could be dropped and instead just assume (4). Then the constant C obviously does not depend on $\|a(x,\omega)\|_{L^\infty(\Omega;W^{1,\infty}(D))}$, but the convergence rate is reduced to order p, cf. Proposition 4.1 and Proposition 4.4. However, the regularity assumption on the diffusion coefficient does not seem to be an additional constraint in view of Assumption 2.4.

Proof. We need to find the correct distribution of samples M_l for an overall convergence rate $O(h_L)$ in Lemma 6.4. We obtain the asserted error bound, if we can show $\sum_{l=1}^{L} h_l^{p+1} M_l^{-1/2} = O(h_L)$. Choosing $M_l = l^{2+\epsilon} 2^{2(L-l)(p+1)}$ for $l = 1, \ldots, L$ and some $\epsilon > 0$, we obtain

$$\sum_{l=1}^{L} h_l^{p+1} M_l^{-1/2} \sim \sum_{l=1}^{L} 2^{-l(p+1)} l^{-(1+\epsilon/2)} 2^{(l-L)(p+1)} = 2^{-L(p+1)} \sum_{l=1}^{L} l^{-(1+\epsilon/2)} = C(\epsilon) h_L^{p+1}.$$

To estimate the work, we observe that under Assumption 6.3 the computational cost for M_l samples on level l is bounded by $O(M_l N_l^t (N_l^x)^\gamma)$. Note that $N_l^x = O(h^{-d}) = O(2^{ld})$ and $N_l^t = O(h^{-(p+1)/r}) = O(2^{l(p+1)/r}) \sim (N_l^x)^{(p+1)/(rd)}$. We get the following bound for the overall work of the MLMC estimator (53) on finest level L:

$$\begin{aligned}
\text{Work}(L) &= \sum_{l=1}^{L} M_l (N_l^t (N_l^x)^\gamma + N_{l-1}^t (N_{l-1}^x)^\gamma) \leq 2 \sum_{l=1}^{L} M_l N_l^t (N_l^x)^\gamma \\
&\lesssim \sum_{l=1}^{L} l^{2+\epsilon} 2^{2(L-l)(p+1)} 2^{l(p+1)/r} 2^{ld\gamma} \\
&= 2^{Ld(\gamma+(p+1)/(rd))} \sum_{l=1}^{L} l^{2+\epsilon} 2^{2(L-l)(p+1)} 2^{(l-L)(p+1)/r} 2^{(l-L)d\gamma} \\
&= (N_L^x)^\gamma N_L^t \sum_{l=1}^{L} l^{2+\epsilon} 2^{(l-L)(-2(p+1)+(p+1)/r+d\gamma)} \\
&= (N_L^x)^\gamma N_L^t \sum_{l'=0}^{L-1} (L-l')^{2+\epsilon} 2^{l'\delta}.
\end{aligned}$$

We obtain the asserted bounds on the computational cost by straightforward estimates:
$\delta < 0$:

$$\text{Work}(L) \leq (N_L^x)^\gamma N_L^t L^{2+\epsilon} \sum_{l'=0}^{\infty} 2^{l'\delta} \leq C(\delta) (N_L^x)^\gamma N_L^t (\log N_L^x)^{2+\epsilon}.$$

$\delta = 0$:

$$\text{Work}(L) \leq (N_L^x)^\gamma N_L^t L L^{2+\epsilon}.$$

$\delta > 0$:

$$\text{Work}(L) \leq (N_L^x)^\gamma N_L^t L^{2+\epsilon} 2^{L\delta} (2^\delta - 1)^{-1} \leq C(\delta) (N_L^x)^\gamma N_L^t (\log N_L^x)^{2+\epsilon} (N_L^x)^{\delta/d}.$$

□

Remark 6.7. For low order approximations in space and time, the contribution to the overall computational cost due to sampling is logarithmic in the spatial degrees of freedom and thus negligible. The computational cost of the MLMC estimator (53) is comparable to one solve of a deterministic parabolic PDE.
However, in order to preserve increased convergence rates induced by high order approximations in space and time, the number M_l of samples on each level need to be increased. This results in an algebraic contribution of sampling to the overall complexity.
Therefore, the proposed MLMC approximation of the mean field of random solutions of stochastic parabolic PDEs is efficient with low order approximations in space and time, particularly suitable for solutions that admit low spatial and temporal regularity. This is in correspondence with the MLMC estimator for stochastic elliptic PDEs [1].

7 Generalizations

7.1 Discontinuous Galerkin Time Stepping

So far, we discussed time stepping schemes of first and second order only. They are suitable for solutions with low regularity in time. However, for smooth solutions in space and time, in view of the high-order approximation in space, we need a small timestep to equilibrate the error of spatial and temporal approximation. Thus, we need to solve the elliptic problem (40) or (47) many times, which is suboptimal.
Typically the solution of deterministic parabolic problems of the type (15) - (17) with a time-independent diffusion coefficient admit smooth solutions in time for sufficiently regular initial condition u_0 and analytic source term $f(t,x)$. Even for incompatible initial conditions, the solution is smooth after a startup singularity. It is thus reasonable to assume high temporal regularity of the random solution $u \in L^2(\Omega;\mathcal{X})$.
Time discretization of deterministic parabolic problems by the *hp* Discontinuous Galerking method has been discussed in [8],[9],[13],[14]. We follow [9] since the error analysis therein is explicit in the dependence on the elliptic differential operator in terms of a_- and a_+, which ensures uniform bounds for all realizations of the random input data.
We semidiscretize the parabolic problem (15) - (17) in time by Discontinuous Galerkin time stepping, the fully discrete problem is then attained by a Finite Element method in space.

7.1.1 Semidiscretization in Time

Let $\mathcal{M} = \{I_m\}_{m=1}^{\mathfrak{M}}$ be a partition of $I = (0,T)$ into $\mathfrak{M}$ subintervals $I_m = (t_{m-1}, t_m)$. The timestep is again denoted by $k_m := t_m - t_{m-1}$. We introduce a vector $\underline{r} := (r_m)_{m=1}^{\mathfrak{M}}$ of temporal approximation orders $r_m \geq 0$ on each time interval I_m. The exact solution is then approximated by piecewise polynomials in time with coefficients in $V = H_0^1(D)$. To this end, we define the test space

$$\mathcal{V}^{\underline{r}}(\mathcal{M};V) = \{u : I \to V; u|_{I_m} \in \mathcal{P}^{r_m}(I_m;V), 1 \leq m \leq \mathfrak{M}\}.$$

We set $N_t := \sum_{m=1}^{\mathfrak{M}}(r_m+1)$ for the number of degrees of freedom of the temporal semidiscretization.
The approximate solution is allowed to be discontinuous across time nodes. We therefore define the left and right limits at time t_m by

$$u_m^+ = \lim_{s\downarrow 0} u(t_m+s), \qquad u_m^- = \lim_{s\downarrow 0} u(t_m-s).$$

The jump across time node t_m is then defined as $[u]_m := u_m^+ - u_m^-$.
We can now formulate the Discontinuous Galerkin time stepping method: Find $U \in \mathcal{V}^r(\mathcal{M};V)$ such that

$$B_{DG}(U,V) = F_{DG}(V) \quad \text{for all } V \in \mathcal{V}^r(\mathcal{M};V), \tag{55}$$

where

$$\begin{aligned} B_{DG}(u,v) &:= \sum_{m=1}^{\mathfrak{M}} \int_{I_m} \{(\tfrac{\partial u}{\partial t}, v) + a(u,v)\}dt + \sum_{m=2}^{\mathfrak{M}} ([u]_{m-1}, v_{m-1}^+) + (u_0^+, v_0^+), \\ F_{DG}(v) &:= \sum_{m=1}^{\mathfrak{M}} \int_{I_m} (f(t),v)dt + (u_0, v_0^+). \end{aligned}$$

The DG time stepping method admits the following error bound.

Theorem 7.1. *Assume that the exact solution has the regularity $u \in L^2(\Omega; H^{s+1}(I;V))$ for some $s \geq 0$. For a given realization $\hat{a} \in L^\infty(D)$ of the stochastic diffusion coefficient satisfying (4) let $U \in \mathcal{V}^r(\mathcal{M};V)$ be the semidiscrete solution of (55) with constant polynomial order $r_m = r$. Then there holds with $k = \max\{k_m\}$ the a priori error estimate*

$$\|u-U\|_{L^2(I;V)} \leq C k^{\min(r,s)+1} r^{-(s+1)} \|u\|_{H^{s+1}(I;V)}, \tag{56}$$

where the constant C depending on a_- and a_+ is uniformly bounded for all $\omega \in \Omega$.

Proof. This is a direct consequence of [9, Theorem 3.11]. The dependence of the constant C on the elliptic spatial differential operator is given explicitly in [9, Proposition 3.3] and is thus uniformly bounded under Assumption 2.1. □

We can therefore combine high-order approximations in space and time. In contrast to low-order time stepping schemes discussed in Section 5 we obtain a reduced computational cost of the temporal approximation, as fewer timesteps are needed to equilibrate spatial and temporal errors. The adjustment of the Multi-Level Monte Carlo method as in Section 6 is then straightforward.
If the stochastic diffusion equation (1) - (3) admits a solution which is analytic in time, we obtain exponential rates of convergence.

Theorem 7.2. *Assume that the stochastic diffusion equation* (1) - (3) *admits a solution* $u \in \underline{\mathcal{X}}$ *which is analytic in the time interval* $[0,T]$ *for every* $\omega \in \Omega$. *Under Assumption 2.1, for a given realization* $\hat{a} \in L^\infty(D)$ *of the stochastic diffusion coefficient, let* $U \in \mathcal{V}^r(\mathcal{M};V)$ *be the semidiscrete solution of* (55) *with polynomial order* $r_m = r$ *on a fixed partition* $\mathcal{M}$ *of the time interval. Then there holds*

$$\|u - U\|_{L^2(I;V)} \leq C \exp(-br), \tag{57}$$

where C *depending on* a_- *and* a_+ *is uniformly bounded for all* $\omega \in \Omega$ *and* $b > 0$ *is independent of* $\omega \in \Omega$.

Proof. Exponential convergence rates of the p-version of Discontinuous Galerkin time stepping have been proven in [9]. Again, uniform boundedness of the constant C follows from [9, Proposition 3.3]. Exponential convergence follows then from [9, Lemma 3.6] with standard approximation theory for analytic functions. Clearly, $b > 0$ is independent of $\omega \in \Omega$, as it stems from general approximation theory. □

Remark 7.3. We note that this result holds true even for more realistic regularity assumptions on the solution $u \in \underline{\mathcal{X}}$. Exponential convergence of the Discontinuous Galerkin time stepping has been proven for solutions that admit time singularities. Singularities may be induced through non-smooth initial data or discontinuities in the source term f. Using a partition $\mathcal{M}$ which is suitably refined towards the singularity and a time approximation order vector $\underline{r}$ which increases away from the singularity, one reobtaines exponential convergence [9]. However, for the sake of simplicity and notational convenience, we stick to the restrictive regularity assumption in Theorem 7.2 but keep in mind that this assumption could be relaxed.

7.1.2 The Fully Discrete Scheme

On each time interval $I_m = (t_{m-1}, t_m)$, we need to solve successively a problem of the form: Find $U_m \in \mathcal{P}^{r_m}(I_m;V)$ such that for all $V \in \mathcal{P}^{r_m}(I_m;V)$ holds

$$\int_{I_m} \{(U_m', V) + a(U_m, V)\}dt + (U_{m-1}^+, V_{m-1}^+) = \int_{I_m} (f(t), V)dt + (U_{m-1}^-, V_{m-1}^+). \tag{58}$$

U_{m-1}^- is given by preceding computations and we set $U_0^- = u_0$.
Choosing a basis $\{\varphi_{i,m}\}_{i=0}^{r_m}$ for the polynomial space $\mathcal{P}^{r_m}(I_m;V)$ we rewrite the semidiscrete solution in every time step as $U_m = \sum_{j=0}^{r_m} u_{j,m}\varphi_{j,m}$. Problem (58) can then be reformulated as a system of $r_m + 1$ coupled elliptic reaction diffusion equations for the coefficients $u_{j,m}$: Find $\{u_{j,m}\}_{j=0}^{r_m} \subset V$ such that for all $\{v_i\}_{i=0}^{r_m} \subset V$ holds

$$\sum_{i,j=0}^{r_m} \mathbb{A}_{ij}^m(u_{j,m}, v_i) + \mathbb{B}_{ij}^m a(u_{j,m}, v_i) = \sum_{i=0}^{r_m} f_{i,m}(v_i). \tag{59}$$

We introduced the matrices

$$\mathbb{A}_{ij}^m := \int_{I_m} \varphi'_{j,m}\varphi_{i,m}dt + \varphi_{j,m}^+(t_{m-1})\varphi_{i,m}^+(t_{m-1}), \qquad \mathbb{B}_{ij}^m := \int_{I_m} \varphi_{j,m}\varphi_{i,m}dt,$$

and the right-hand side is defined by

$$f_{i,m}(v_i) := \Big(\int_{I_m} f\varphi_{i,m}dt, v_i\Big) + (U_{m-1}^-, v_i)\varphi_{i,m}^+(t_{m-1}).$$

We need to solve the coupled system (59) numerically. The optimal choice for the basis functions $\{\varphi_{i,m}\}_{i=0}^{r_m}$ would be the one that diagonalizes the matrices $\mathbb{A}_{ij}^m$ and $\mathbb{B}_{ij}^m$ simultaneously. The system would then decouple in $r_m + 1$ independent equations. Unfortunately, it does not seem to be possible to find such diagonalizations over $\mathbb{R}$ [9].
We *choose* normalized Legendre polynomials as basis functions. Due to orthogonality of the Legendre polynomials with respect to the L^2 inner product, the matrix $\mathbb{B}_{ij}^m = \frac{k_m}{2}\delta_{ij}$ is then diagonalized. We have two options to discretize problem (59) in space, cf. [9], [14] for details.

Direct Spatial Discretization: We directly apply standard techniques analogous to the spatial discretization presented in Section 4 to approximate the solution of (59) in space. We choose a finite dimensional subspace $V_h \subset V$ with dimension $N_h = \dim(V_h)$ and basis $\{\Phi_l\}_{l=1}^{N_h}$. Each coefficient $u_{j,m} \in V$ is then approximated by $u_{j,m}^{FE} = \sum_{l=1}^{N_h} u_{j,m}^l \Phi_l \in V_h$.
We denote the mass matrix again by $\boldsymbol{M}_{lk} = (\Phi_l, \Phi_k)$ and $\boldsymbol{A}_{lk} = a(\Phi_l, \Phi_k)$ is the stiffness matrix. The fully discrete system for the coefficient vector $\underline{u}_{j,m}^{FE} = (u_{j,m}^1, \dots, u_{j,m}^{N_h}) \in \mathbb{R}^{N_h}$ is then for $m = 1, \dots, \mathfrak{M}$:

$$\begin{bmatrix} \mathbb{A}_{00}^m\boldsymbol{M} + \frac{k_m}{2}\boldsymbol{A} & \dots & \mathbb{A}_{0r}^m\boldsymbol{M} \\ \vdots & \ddots & \vdots \\ \mathbb{A}_{r0}^m\boldsymbol{M} & \dots & \mathbb{A}_{rr}^m\boldsymbol{M} + \frac{k_m}{2}\boldsymbol{A} \end{bmatrix} \begin{bmatrix} \underline{u}_{0,m}^{FE} \\ \vdots \\ \underline{u}_{r,m}^{FE} \end{bmatrix} = \begin{bmatrix} \underline{f}_{0,m} \\ \vdots \\ \underline{f}_{r,m} \end{bmatrix} \tag{60}$$

with load vector

$$\underline{f}_{i,m} = (f_{i,m}(\Phi_1), \dots, f_{i,m}(\Phi_{N_h})).$$

Note that the system (60) is of dimension $(r_m + 1)N_h$. The source of randomness lies in the stiffness matrix in the diagonal elements. Efficient solvers for sparse linear systems of the type (60) need to be used. Nevertheless, in view of the computational cost, it is advantageous to decouple the problem (59) into $(r_m + 1)$ scalar problems.

Decoupling: Numerical calculations have shown that the matrix $\mathbb{A}^m$ is diagonalizable over $\mathbb{C}$ for $0 \le r_m \le 100$, cf. [9]. There exists a matrix $\boldsymbol{Q}_m \in \mathbb{C}^{(r_m+1)\times(r_m+1)}$ such that $\boldsymbol{Q}_m^{-1}\mathbb{A}^m\boldsymbol{Q}_m = \mathrm{diag}(\lambda_1, \dots, \lambda_{r_m+1})$. The coupled system (59) transforms into a decoupled

system of $r_m + 1$ reaction-diffusion equations, which can be solved independently: Find $w_{j,m} \in V$ such that for all $v \in V$ holds

$$\lambda_j(w_{j,m}, v) + \frac{k_m}{2} a(w_{j,m}, v) = g_{j,m}(v), \qquad j = 0, \ldots, r_m. \tag{61}$$

The vector $\vec{w}_m = (w_{0,m}, \ldots, w_{r_m,m})$ is related to the solution $\vec{u}_m = (u_{0,m}, \ldots, u_{r_m,m})$ of the coupled system (59) by $\vec{w}_m = \boldsymbol{Q}_m^{-1}\vec{u}_m$ and the right-hand side is given by $\vec{g}_m = \boldsymbol{Q}_m^{-1}\vec{f}_m(v)$. We use Finite Elements to approximate the solution of the decoupled system (61). For some triangulation $\mathcal{T}_h$ of the domain D with meshwidth h we introduce the Finite Element space $V_h := \mathcal{S}_0^{p,1}(D, \mathcal{T}_h) \subset V$ of dimension $N_h = \dim(V_h)$. The Finite Element approximation $w_{j,m}^{FE} \in V_h$ is then given by

$$\lambda_j(w_{j,m}^{FE}, v_h) + \frac{k_m}{2} a(w_{j,m}^{FE}, v_h) = g_{j,m}(v_h), \quad \forall v_h \in V_h,\ j = 0, \ldots, r_m.$$

In a particular basis $\{\Phi_l\}_{l=1}^{N_h} \subset V_h$ this translates into the following linear system

$$(\lambda_j \boldsymbol{M} + \frac{k_m}{2}\boldsymbol{A})\underline{w}_{j,m}^{FE} = \underline{g}_{j,m}. \tag{62}$$

This linear system of dimension N_h need to be solved $r_m + 1$ times for every timestep $1 \le m \le \mathfrak{M}$. For a given realization $\hat{a} \in L^\infty(D)$ of the stochastic diffusion coefficient, we obtain the Discontinuous Galerkin Finite Element approximation of (15) - (17) by applying the backtransformation $\vec{u}_m^{FE} = \boldsymbol{Q}_m\vec{w}_m^{FE}$, $1 \le m \le \mathfrak{M}$.
For the error analysis of the fully discrete method, we need the following conjecture.

Conjecture 7.4. *The matrix $\mathbb{A}^m$ diagonalizable for any $r_m \ge 0$ and the squared norm of the transformation matrix $\boldsymbol{Q}_m$ grows at most linearly in r_m, i.e.*

$$\|\boldsymbol{Q}_m\|^2 \le C r_m^\alpha, \qquad 0 \le \alpha \le 1. \tag{63}$$

As all matrix norms are equivalent, we do not specify the norm in (63). Numerical experiments sustain this conjecture. The matrix l^2-norm of the transformation $\boldsymbol{Q}_m$ has been computed in [9] for $0 \le r_m \le 50$ and these numerical experiments suggest that indeed $\alpha = 1$.
We obtain the following error estimate of the fully discrete DGFEM scheme:

Theorem 7.5. *Assume that the stochastic diffusion equation* (1) - (3) *admits a solution $u \in L^2(\Omega; C^\infty([0,T]; H^{p+1}(D)))$ which is analytic in the time. Under Assumption 2.1, for a given realization $\hat{a} \in L^\infty(D)$ of the stochastic diffusion coefficient, let $u^{DG} \in \mathcal{V}^r(\mathcal{M}; V_h)$ be the approximate solution given by $u^{DG}|_{I_m} = \sum_{j=0}^r u_{j,m}^{FE}\varphi_{j,m}$, where $\varphi_{j,m}$ are normalized Legendre polynomials and the coefficients $\vec{u}_m^{FE} = \boldsymbol{Q}_m\vec{w}_m^{FE}$ are computed by solving the decoupled system* (62) *with constant polynomial order $r_m = r$ on a fixed partition $\mathcal{M}$ of the*

time interval.
If Conjecture 7.4 is valid, we obtain the following error bound:

$$\|u - u^{DG}\|_{L^2(I;V)} \leq C\big(\exp(-br) + r^2 h^p\big), \tag{64}$$

where the constant C depending on T, a_-, a_+ is uniformly bounded for all $\omega \in \Omega$ and $b > 0$ is independent of $\omega \in \Omega$.

Proof. Let $U \in \mathcal{V}^r(\mathcal{M}; V)$ be the semidiscrete solution given by (55). We have

$$\|u - u^{DG}\|^2_{L^2(I;V)} \leq 4\big(\|u - U\|^2_{L^2(I;V)} + \|U - u^{DG}\|^2_{L^2(I;V)}\big).$$

The first term on the right hand side is bounded exponentionally in the polynomial degree r by Theorem 7.2. We can therefore focus on the second term. We write the semidiscrete solution in each timestep as $U|_{I_m} = \sum_{j=0}^{r} u_{j,m}\varphi_{j,m}$. The orthogonality properties of the Legendre polynomials imply (cf. [9, Proposition 6.1])

$$\|U - u^{DG}\|^2_{L^2(I;V)} = \sum_{m=1}^{\mathfrak{M}} \frac{k_m}{2} \sum_{j=0}^{r} \|u_{j,m} - u^{FE}_{j,m}\|^2_V.$$

By Conjecture 7.4 we have

$$\|u - u^{DG}\|^2_{L^2(I;V)} \leq C\Big(\exp(-br) + \sum_{m=1}^{\mathfrak{M}} \frac{k_m}{2} \sum_{j=0}^{r} r\|w_{j,m} - w^{FE}_{j,m}\|^2_V\Big). \tag{65}$$

We follow [9] for the error analysis of the Finite Element approximation of the singularly perturbed reaction-diffusion equation (61). For $\varepsilon \in \mathbb{C}$ we define the bilinear form $b_\varepsilon(\cdot,\cdot) := \varepsilon^2 a(\cdot,\cdot) + (\cdot,\cdot)$ on the vector space V. Under Assumption 2.1, the bilinear form b_ε is continuous and coercive on V equipped with the energy norm $\|\cdot\|^2_\varepsilon := |\varepsilon|^2\|\cdot\|^2_V + \|\cdot\|^2_H$. We have (cf. [9, Lemma 6.4])

$$|b_\varepsilon(w,w)| \geq \sqrt{2}\min(1, a_-)\|w\|^2_\varepsilon \quad \text{for all } w \in V.$$

Note that the coercivity constant is uniformly bounded away from zero for all $\omega \in \Omega$. As a direct consequence, we obtain quasioptimality of the Finite Element approximation $w^{FE}_{j,m} \in V_h$ given by (62) in the energy norm:

$$\|w_{j,m} - w^{FE}_{j,m}\|_\varepsilon \leq C \inf_{v_h \in V_h} \|w_{j,m} - v_h\|_\varepsilon,$$

where the constant C is uniformly bounded for all $\omega \in \Omega$. The construction of suitable finite dimensional subspaces V_h (or the creation of suitable meshes, respectively) for solutions that admit boundary layers due to the small parameter ε or singularities near corners has been exemplified in case of the two dimensional heat equation in [9]. As we assumed additional

regularity of the solution, we conclude with standard approximation theory of the FE space $V_h := \mathcal{S}_0^{p,1}(D, \mathcal{T}_h)$:

$$\inf_{v_h \in V_h} \|w_{j,m} - v_h\|_\varepsilon \leq Ch^p.$$

In (61) we have $\varepsilon^2 = \frac{k_m}{2\lambda_j}$. By [9, Lemma 6.3] we have $|\varepsilon|^2 \geq C\frac{k_m}{r^2}$ and hence

$$\|w_{j,m} - w_{j,m}^{FE}\|_V^2 \leq |\varepsilon|^{-2}\|w_{j,m} - w_{j,m}^{FE}\|_\varepsilon^2 \leq C\frac{r^2}{k_m}h^{2p}.$$

Combining this result with (65) we obtain

$$\|u - u^{DG}\|_{L^2(I;V)}^2 \leq C\Big(\exp(-br) + Tr^4h^{2p}\Big).$$

The assertion follows. □

7.1.3 Single-Level Monte Carlo Method

We adopt the concepts of Section 6 to estimate the mean field of the random solution $u \in L^2(\Omega; \mathcal{X})$ of the stochastic diffusion equation (1) - (3). Since the Discontinuous Galerking time stepping method provides solution samples which are piece-wise continuous in time rather than just point values, we estimate the error in a different norm. We therefore need slight adjustments of the theory for the Monte Carlo estimator.
For M independent, identically distributed realizations $\hat{a}^i$ of the diffusion coefficient, we denote by ${}^i\hat{u}$ the corresponding exact solution samples. The Monte Carlo estimator is then given by

$$E_M[u] := \frac{1}{M}\sum_{i=1}^{M} {}^i\hat{u} \approx \mathbb{E}[u]. \tag{66}$$

We have the following analogon to Lemma 6.1.

Lemma 7.6. *For any $M \in \mathbb{N}$ and for $u \in L^2(\Omega; \mathcal{X})$ holds*

$$\|\mathbb{E}[u] - E_M[u]\|_{L^2(\Omega;L^2(I;V))} \leq M^{-1/2}\,\|u\|_{L^2(\Omega;L^2(I;V))}\,. \tag{67}$$

Proof. Obviously the solution $u \in \underline{\mathcal{X}}$ defines a random field on $L^2(I;V)$. The proof of Lemma 6.1 can thus be followed step by step. □

We replace the exact solution samples by the approximation obtained from the DGFEM scheme. For a given realization $\hat{a}^i \in L^\infty(D)$ of the stochastic diffusion coefficient let ${}^i\hat{u}_{h,r}$ denote the approximation of the solution sample ${}^i\hat{u}$ obtained by DGFEM on a given mesh $\mathcal{T}_h$ of the physical domain D with meshwidth h, with a fixed partition $\mathcal{M}$ of the time

interval and constant polynomial order $r_m = r$.
The single-level Monte Carlo estimator is then given by

$$E_M[u_{h,r}] := \frac{1}{M}\sum_{i=1}^{M} {}^i\hat{u}_{h,r} \approx \mathbb{E}[u]. \tag{68}$$

We have the following error bound, that comprises spatial approximation, time stepping and the statistical error from the MC estimator.

Theorem 7.7. *Assume that the stochastic diffusion equation* (1) - (3) *admits a solution* $u \in L^2(\Omega; C^\infty([0,T]; H^{p+1}(D)))$ *which is analytic in time and that the stochastic diffusion coefficient satisfies* (4). *If Conjecture 7.4 is valid, then the Monte Carlo estimator* (68) *with solution samples generated by the DGFEM scheme with constant polynomial order* $r_m = r$ *on a fixed partition* $\mathcal{M}$ *of the time interval and a given mesh* $\mathcal{T}_h$ *with meshwidth* h *admits the error bound*

$$\|\mathbb{E}[u] - E_M[u_{h,r}]\|_{L^2(\Omega;L^2(I;V))} \leq C\big(\exp(-br) + r^2h^p + M^{-1/2}\big). \tag{69}$$

The constant C *depends on* T, a_- *and* a_+.

Proof. As in the proof of Theorem 6.2 we split the error as follows

$$\begin{aligned}\|\mathbb{E}[u] - E_M[u_{h,r}]\|_{L^2(\Omega;L^2(I;V))} \;\leq\; & \|\mathbb{E}[u] - E_M[u]\|_{L^2(\Omega;L^2(I;V))} \\ & + \|E_M[u - u_{h,r}]\|_{L^2(\Omega;L^2(I;V))} =: I + II.\end{aligned}$$

The first term is estimated with Lemma 7.6, we have $I \leq CM^{-1/2}$. The second term is then bounded by Theorem 7.5

$$II \leq M^{-1}\sum_{i=1}^{M} \big\|{}^i\hat{u} - {}^i\,\hat{u}_{h,r}\big\|_{L^2(\Omega;L^2(I;V))} \leq C\big(\exp(-br) + r^2h^p\big).$$

Note that it is crucial, that the constant C in (64) is uniformly bounded for all $\omega \in \Omega$. □

As the errors of spatial and temporal approximation are not strictly separated for the DGFEM scheme, the error equilibration is more challenging in this case. We start with a given grid $\mathcal{T}_h$ with meshwidth h for the spatial approximation of the decoupled system (61). For a given partition $\mathcal{M}$ of the time interval we need to choose the appropriate polynomial degree r on each timestep and the optimal sampling size M to equilibrate the overall error (69).
Let $W(x)$ denote the Lambert W function, also called Omega function, which is defined as the solution of $x = W(x)\exp(W(x))$. One easily checks that $r := \frac{2}{b}W(\frac{1}{2}\sqrt{b^2h^{-p}})$ is the solution of the equation $r^2\exp(br) = h^{-p}$ and that the error of spatial and temporal

approximation is thus equilibrated. The derivation of the optimal sampling size is then straightforward. However, in view of the asymptotical analysis of the computational cost and for the sake of simplicity, we choose a slightly suboptimal temporal polynomial degree. Using $r = -(p/b)\log h = O(|\log h|)$ for the semidiscretization in time, we obtain a temporal approximation error of order h^p. The spatial error then decreases asymptotically as $O(|\log h|^2 h^p)$. The sampling size is then chosen as $M = O(h^{-2p})$ which results in a statistical error of order h^p. As the overall error is then dominated by the spatial error of order $|\log h|^2 h^p$, one could also choose $M = O(h^{-2p}|\log h|^{-4})$ which would slightly reduce the computational cost due to sampling. In case of the single-level Monte Carlo method, we are merely interested in rough work estimates, but we will consider this factor for the MLMC estimator, where optimal work estimates are sought. For the sake of simplicity, we therefore drop the logarithmic factor in the sampling size.
The choice $r = O(|\log h|)$ has merely been justified by intuition for the purpose of simplicity. However, the asymptotic behaviour of the Lambert W function suggests, that this choice is "almost" optimal. It holds $W(x) = \log x - \log\log x + O(\frac{\log\log x}{\log x})$ for $x \gg 1$. The optimal polynomial degree is thus

$$\begin{aligned} r &:= \tfrac{2}{b} W(\tfrac{1}{2}\sqrt{b^2 h^{-p}}) = \tfrac{2}{b}\log(\tfrac{1}{2}\sqrt{b^2 h^{-p}}) - \log\log(\tfrac{1}{2}\sqrt{b^2 h^{-p}}) + \dots \\ &\sim -\tfrac{p}{b}\log h - \log|\log h| + \dots \end{aligned}$$

The difference is iterated logarithmic and therefore negligible.
For fixed meshwidth h, the SLMC error is equilibrated with temporal polynomial degree $r = O(|\log h|)$ and sampling size $M = O(h^{-2p})$.
We can now turn to the analysis of the computational cost of the single-level Monte Carlo estimator (68). The computation of each approximate solution sample inherits solving the linear system (62) $r+1$ times for each timestep $1 \le m \le \mathfrak{M}$, i.e. we need to solve (62) $M \cdot N_t$ times, where $N_t = \sum_{m=1}^{\mathfrak{M}}(r+1) = O(r)$. We make the following assumption on the solver of the linear system:

Assumption 7.8. In each time interval I_m, $1 \le m \le \mathfrak{M}$, for $j = 0, \dots, r$ an approximate solution $\tilde{w}_{j,m}^{FE}$ of the linear problem (62) is obtained in $O(N_h^\gamma)$ work, $1 \le \gamma \le 3$, such that the exact solution u^{DG} of the fully discrete problem is approximated to order $O(\exp(-br) + r^2 h^p)$ in $L^2(I;V)$-norm, i.e.

$$\|u^{DG} - \tilde{u}\|_{L^2(I;V)} \le C(\exp(-br) + r^2 h^p),$$

where $\tilde{u}|_{I_m} = \sum_{j=0}^{r} \tilde{u}_{j,m}^{FE} \varphi_{j,m}$ and $\underline{\tilde{u}}_m^{FE} = \boldsymbol{Q}_m \underline{\tilde{w}}_m^{FE}$.

The work needed to solve the decoupled system (62) for each approximate solution sample is thus given by $O(M \cdot N_t \cdot N_h^\gamma)$. We have not yet considered the work needed for decoupling and the backward transformation.
Note that for constant polynomial degree $r_m = r$, the matrix $\mathbb{A}_m$ need to be diagonalized

only once. The computational cost for diagonalization is thus negligible. However, the backward transformation $\underline{\tilde{u}}_m^{FE} = \boldsymbol{Q}_m \underline{\tilde{w}}_m^{FE}$ in each timestep $1 \leq m \leq \mathfrak{M}$ is needed for each sample. As every entry $\tilde{w}_{j,m}$, $0 \leq j \leq r$ of the vector $\underline{\tilde{w}}_m^{FE}$ is in fact a vector $(\tilde{w}_{j,m}^l)_{l=1}^{N_h} \in \mathbb{R}^{N_h}$, this backward transformation needs $O(r^2 N_h)$ operations.
The overall computational cost of the single-level Monte Carlo estimator (68) is thus given by $O(M \cdot \{N_t \cdot N_h^\gamma + r^2 N_h\}) = O(N_h^{2p/d+\gamma}(\log N_h)^2)$. In view of the results derived in Section 6, the contribution to the overall computational cost due to sampling is unchanged for the single-level MC DGFEM estimator. However, the computational cost due to time stepping is logarithmic in the spatial degrees of freedom.

7.1.4 Multi-Level Monte Carlo Method

We improve the efficiency of the algorithm by using different sample sizes on different levels of spatial and temporal approximation. We introduce a level-dependent temporal polynomial degree r_l, $l \geq 0$. Let $\{\mathcal{T}_l\}_{l=0}^{\infty}$ be a sequence of regular simplicial meshes of the domain D obtained by uniform mesh refinement. For $l \geq 0$ the meshwidth of $\mathcal{T}_l$ is then $h_l = 2^{-l} h_0$. Let $V_l := \mathcal{S}_0^{p,1}(D, \mathcal{T}_l)$ be the Finite Element space with $N_l^x := \dim V_l = O(2^{dl})$ degrees of freedom. On each level l, the DGFEM method consists of solving $r_l + 1$ decoupled linear systems of the type (62) in each timestep. We choose $r_l = O(|\log h_l|)$ to equilibrate spatial and temporal error.
We then denote by $u_l := u_{h_l, r_l}$ the approximate solution sample obtained by DGFEM. On a fixed partition $\mathcal{M}$ of the time interval, for the temporal polynomial order r_l, the solution is then given by $u_l|_{I_m} = \sum_{j=0}^{r_l} u_{j,m}^{FE} \varphi_{j,m}$, where $\varphi_{j,m}$ are normalized Legendre polynomials and the coefficients $\vec{u}_m^{FE} = \boldsymbol{Q}_m \vec{w}_m^{FE}$ are computed by solving the decoupled system (62) on Finite Element space V_l.
The Multi-Level Monte Carlo estimator is then defined as

$$E^L[u] := \sum_{l=1}^{L} E_{M_l}[u_l - u_{l-1}] \approx \mathbb{E}[u]. \tag{70}$$

We have the following error estimate.

Lemma 7.9. *Assume that the stochastic diffusion equation* (1) - (3) *admits a solution* $u \in L^2(\Omega; C^\infty([0,T]; H^{p+1}(D)))$ *which is analytic in time and assume that the stochastic diffusion coefficient satisfies* (4). *If Conjecture 7.4 is valid, then the multi-level Monte Carlo estimator* (70) *admits the error bound*

$$\left\| \mathbb{E}[u] - E^L[u] \right\|_{L^2(\Omega; L^2(I;V))} \leq C\Big(|\log h_L|^2 h_L^p + \sum_{l=1}^{L} |\log h_l|^2 h_l^p M_l^{-1/2} \Big). \tag{71}$$

The constant C depends on p, T, a_- and a_+.

Proof. Analogously to the proof of Lemma 6.4, we rewrite the error as

$$\begin{aligned}\left\|\mathbb{E}[u] - E^L[u]\right\|_{L^2(\Omega;L^2(I;V))} &\leq \|\mathbb{E}[u] - \mathbb{E}[u_L]\|_{L^2(\Omega;L^2(I;V))} \\ &\quad + \sum_{l=1}^{L} \|\mathbb{E}[u_l - u_{l-1}] - E_{M_l}[u_l - u_{l-1}]\|_{L^2(\Omega;L^2(I;V))} \\ &=: I + II.\end{aligned}$$

The first term is bounded by Jensen's inequality and Theorem 7.5,

$$I \leq \mathbb{E}[\|u - u_L\|_{L^2(I;V)}] \leq C\big(\exp(-br_L) + r_L^2 h_L^p\big) \leq C|\log h_L|^2 h_L^p.$$

For the second term, we estimate each addend as follows:

$$\begin{aligned}\|\mathbb{E}[u_l - u_{l-1}] - E_{M_l}[u_l - u_{l-1}]\|_{L^2(\Omega;L^2(I;V))} &\leq M_l^{-1/2} \|u_l - u_{l-1}\|_{L^2(\Omega;L^2(I;V))} \\ &\leq M_l^{-1/2}\big(\|u - u_l\|_{L^2(\Omega;L^2(I;V))} + \|u - u_{l-1}\|_{L^2(\Omega;L^2(I;V))}\big) \\ &\leq C M_l^{-1/2}\big(\exp(-br_l) + r_l^2 h_l^p + \exp(-br_{l-1}) + r_{l-1}^2 h_{l-1}^p\big) \\ &\leq C M_l^{-1/2}\big(|\log h_l|^2 h_l^p + |\log h_{l-1}|^2 h_{l-1}^p\big) \\ &\leq C M_l^{-1/2} |\log h_l|^2 h_l^p.\end{aligned}$$

Here we used Lemma 7.6 and Theorem 7.5. Summation over $l = 1, \ldots, L$ completes the proof. □

By finding the optimal ratio between sample size and meshlevel, we recover the same convergence rate as single-level Monte Carlo on the finest level with a reduced computational cost.

Theorem 7.10. *Assume that the stochastic diffusion equation* (1) - (3) *admits a solution* $u \in L^2(\Omega; C^\infty([0,T]; H^{p+1}(D)))$ *which is analytic in time and that the stochastic diffusion coefficient satisfies* (4). *If Conjecture 7.4 is valid, then the multi-level Monte Carlo estimator* (70) *with* M_l *samples given by*

$$M_l = l^{2+\varepsilon} 2^{2p(L-l)} (l/L)^4, \qquad l = 1, 2, \ldots, L, \tag{72}$$

with $\varepsilon > 0$ *arbitrarily small, admits the error bound*

$$\left\|\mathbb{E}[u] - E^L[u]\right\|_{L^2(\Omega;L^2(I;V))} \leq C|\log h_L|^2 h_L^p. \tag{73}$$

The constant C *depends on* ε, p, T, a_- *and* a_+. *Under Assumption 7.8, the computational cost is bounded by*

$$\mathit{Work}(L) \leq C(\delta^*) \begin{cases} (N_L^x)^\gamma (\log N_L^x)^{4+\epsilon} & \text{if } \delta^* < 0, \\ (N_L^x)^\gamma (\log N_L^x)^{5+\epsilon} & \text{if } \delta^* = 0, \\ (N_L^x)^{\gamma + \delta^*/d} (\log N_L^x)^{4+\epsilon} & \text{if } \delta^* > 0, \end{cases}$$

where $\delta^* = 2p - d\gamma$.

Remark 7.11. We mention again that the error estimate holds true under weaker regularity assumptions on the solution $u \in \underline{\mathcal{X}}$. Analyticity of the solution is not realistic for most applications. Startup-singularities may be induced through non-smooth initial data. Using a partition $\mathcal{M}$ which is suitably refined towards the singularity and a time approximation order vector $\underline{r}$ which increases away from the singularity, one reobtaines exponential convergence of the Discontinuous Galerkin time stepping scheme [9]. Due to the nonconstant approximation order in time, the system (59) needs to be decoupled in every timestep. Therefore the computational cost of the diagonalization process needs to be considered.

Proof. With the samples M_l distributed as (72) we have

$$\begin{aligned}\sum_{l=1}^{L} M_l^{-1/2} h_l^p |\log h_l|^2 &\sim \sum_{l=1}^{L} l^{-(1+\varepsilon/2)} 2^{(l-L)p} (L/l)^2 2^{-lp} l^2 = 2^{-Lp} L^2 \sum_{l=1}^{L} l^{-(1+\varepsilon/2)} \\ &\leq C(\varepsilon) h_L^p |\log h_L|^2.\end{aligned}$$

We obtain the asserted error bound from Lemma 7.9.
We proceed with the analysis of the computational cost. Note that for each level $l = 1, \ldots, L$ the matrix $\mathbb{A}_m$ needs to be diagonalized only once. The computational cost of decoupling is therefore negligible. For each sample, we need to solve $r_l + 1$ linear systems of the form (62) in each timestep and then perform backward transformation. Under Assumption 7.8, the work is therefore bounded on each level by $O(M_l r_l^2 (N_l^x)^\gamma)$. We obtain the following bound for the overall work of the MLMC estimator (70) on the finest level:

$$\begin{aligned}Work(L) &\leq \sum_{l=1}^{L} M_l \left(r_l^2 (N_l^x)^\gamma + r_{l-1}^2 (N_{l-1}^x)^\gamma\right) \leq 2 \sum_{l=1}^{L} M_l r_l^2 (N_l^x)^\gamma \\ &\lesssim \sum_{l=1}^{L} l^{2+\varepsilon} 2^{2p(L-l)} (l/L)^4 l^2 2^{dl\gamma} \\ &= 2^{dL\gamma} L^2 \sum_{l=1}^{L} l^{2+\varepsilon} (l/L)^6 2^{(l-L)(d\gamma-2p)} \\ &\lesssim (N_L^x)^\gamma |\log N_L^x|^2 \sum_{l=1}^{L} l^{2+\varepsilon} 2^{(L-l)\delta^*} \\ &= (N_L^x)^\gamma |\log N_L^x|^2 \sum_{l'=0}^{L-1} (L-l')^{2+\varepsilon} 2^{l'\delta^*}.\end{aligned}$$

The sum is then estimated as in the proof of Theorem 6.5 which leads to the asserted bound of the computational cost. □

We observe that for low-order approximations in space, the MLMC DGFEM algorithm (70) for stochastic parabolic PDEs is of log-linear complexity and therefore comparable to

one deterministic solve of an elliptic PDE.
Time stepping and sampling contributes merely logarithmically, the overall computational cost is dominated by solving the decoupled elliptic system (62).
However, for high order schemes, when $\delta^* = 2p - d\gamma > 0$, more samples are needed, which increases the complexity of the MLMC DGFEM method. It is nevertheless superior to the single-level MC approach. However, "polynomial chaos" type discretizations of the probability space may be advantageous in that case.

7.2 Heat Equation

We use the concepts derived in the previous sections to solve a class of stochastic parabolic PDEs which is more general. We introduce a random density field $\rho(x,\omega) : \Omega \to L^\infty(D)$ which is independent of time. Consider the stochastic heat equation

$$\rho(x,\omega)\tfrac{\partial u}{\partial t} - \operatorname{div}(a(x,\omega)\nabla u) = f(t,x) \quad \text{in } Q_T,\ \omega \in \Omega, \tag{74}$$

$$u(t,x,\omega) = 0 \quad \text{on } \Sigma_T, \tag{75}$$

$$u(t=0,x,\omega) = u_0(x) \quad \text{in } D. \tag{76}$$

As for the stochastic diffusion coefficient we assume boundedness of the density.

Assumption 7.12. There exist constants $0 < \rho_- \le \rho_+ < \infty$ such that

$$0 < \rho_- \le \rho(x,\omega) \le \rho_+ < \infty \quad \text{for all } x \in D, \omega \in \Omega. \tag{77}$$

The unique weak solution $u \in \underline{\mathcal{X}}$ is then given by

$$\tilde{B}(u,v) = \tilde{F}(v) \quad \text{for all } v \in \underline{\mathcal{Y}}, \tag{78}$$

where

$$\tilde{B}(u,v) = \mathbb{E}\Big[\int_I (\rho(x,\omega)\tfrac{\partial u}{\partial t}, v_1)_H dt + \int_I \int_D a(x,\omega)\nabla u \nabla v_1 dx dt + (u(0,x), v_2)_H\Big],$$

and

$$\tilde{F}(v) = \mathbb{E}\Big[\int_I (f, v_1)_H dt + (u_0(x), v_2)_H\Big].$$

We assume again additional regularity (14) of the stochastic solution.
It is straightforward to check that all results derived in the previous sections are applicable. For fixed $\omega \in \Omega$ we use Finite Elements of order $p \ge 1$ for spatial discretization as in Section 4. We arrive at a system of ODEs

$$\boldsymbol{M}\dot{\underline{u}}(t) + \boldsymbol{A}\underline{u}(t) = \underline{f}(t), \quad \underline{u}(0) = \underline{u}_0.$$

Note that the mass matrix is now given by $\boldsymbol{M}_{ij} = \int_D \rho(\cdot,\omega)\Phi_i\Phi_j dx$. The semidiscrete solution $u_h(t)$ admits the same error bound (32) as in the settings of Section 2, here the

constant C depends on a_-, a_+, $\|\hat{a}\|_{W^{1,\infty}(D)}$, ρ_- and ρ_+. The proof of Theorem 4.5 needs minor modifications; (35) is replaced by

$$(\rho(\omega,\cdot)\tfrac{\partial \xi}{\partial t},\xi) + a(\xi,\xi) = -(\rho(\omega,\cdot)\tfrac{\partial \eta}{\partial t},\xi)$$

and (77) is then used.
We obtain a fully discrete problem by using implicit Euler or Crank-Nicolson time stepping. The same error bounds as in Theorem 5.1 and Theorem 5.2 hold, where the constant C depends additionally on ρ_- and ρ_+. Again, minor modifications of the proofs are required, analogous to the semidiscrete case.
As there are essentially no changes in the error analysis of the spatial and temporal discretization, we can adopt the Monte Carlo theory of Section 6 without any modifications. The main results, in particular Theorem 6.5, are thus applicable for the generalized problem (74) - (76). The only additional difficulty lies in the implementation of the mass matrix.

8 Stochastic Wave Equation

We discuss a different type of stochastic PDEs in this section. Techniques analogous to the concepts presented in the previous sections are used for the numerical solution of second order hyperbolic partial differential equations with random coefficients.
For $0 < T < \infty$ we define again the space-time cylinder $Q_T = I \times D$, where $I = (0,T)$ is the bounded time interval and $D \subset \mathbb{R}^d$ is a bounded Lipschitz polyhedron. We consider the stochastic wave equation

$$\frac{\partial^2 u}{\partial t^2} - \operatorname{div}(a(x,\omega)\nabla u) = f(t,x) \quad \text{in } Q_T,\ \omega \in \Omega, \tag{79}$$

$$u(t,x,\omega) = 0 \quad \text{on } \Sigma_T, \tag{80}$$

$$u(0,x,\omega) = u_0(x) \quad \text{in } D, \tag{81}$$

$$\frac{\partial u}{\partial t}(0,x,\omega) = u_1(x) \quad \text{in } D. \tag{82}$$

We assume again $\mathbb{P}$-a.s. boundedness of the random field $a(x,\omega) : \Omega \to L^\infty(D)$.

Assumption 8.1. There exist constants $0 < a_- \le a_+ < \infty$ such that

$$0 < a_- \le a(x,\omega) \le a_+ < \infty \quad \text{for all } x \in D, \omega \in \Omega. \tag{83}$$

We follow [5] for the variational formulation and well-posedness. We set $V := H_0^1(D)$, $H := L^2(D)$ and introduce the Bochner spaces

$$\underline{\mathcal{X}}^* := L^2(\Omega,\mathcal{A},\mathbb{P}; L^2(I;V) \cap H^1(I;H) \cap H^2(I;V')) \quad \text{and} \tag{84}$$

$$\underline{\mathcal{Y}}^* := L^2(\Omega,\mathcal{A},\mathbb{P}; L^2(I;V) \times V \times H). \tag{85}$$

The variational formulation of the stochastic wave equation (79) - (82) is then: Find $u \in \underline{\mathcal{X}}^*$ such that

$$B^*(u,v) = F^*(v) \quad \text{for all } v = (v_0, v_1, v_2) \in \underline{\mathcal{Y}}^*, \tag{86}$$

where

$$\begin{aligned} B^*(u,v) &= \mathbb{E}\Big[\int_I (\tfrac{\partial^2 u}{\partial t^2}, v_0)_H dt + \int_I \int_D a(x,\omega)\nabla u \nabla v_0 dx dt + (u(0), v_1)_V + (u_t(0), v_2)_H\Big], \\ F^*(v) &= \mathbb{E}\Big[\int_I (f, v_0)_H dt + (u_0, v_1)_V + (u_1, v_2)_H\Big]. \end{aligned}$$

For fixed $\omega \in \Omega$, the mappings $u(\cdot,\cdot,\omega) : [0,T] \to V$ and $\frac{\partial u}{\partial t}(\cdot,\cdot,\omega) : [0,T] \to H$ are continuous (cf. [6, Theorem 8.2]) so that the initial conditions (81),(82) incorporated in the variational formulation (86) are indeed meaningful.

Theorem 8.2. *Under Assumption 8.1, for $f \in L^2(I;H)$, $u_0 \in V$ and $u_1 \in H$ the variational formulation* (86) *of the stochastic wave equation* (79)-(82) *admits a unique solution $u \in \underline{\mathcal{X}}^*$. It holds the apriori estimate*

$$\|u\|_{\underline{\mathcal{X}}^*} \leq C(\|f\|_{L^2(I;H)} + \|u_0\|_V + \|u_1\|_H), \tag{87}$$

where C depends only on T, a_- and a_+.

Proof. The proof is given in [5, Proposition 1.6]. For fixed $\omega \in \Omega$, existence of a weak solution and uniform boundedness for all realizations is a consequence of [15, Theorem 29.1]. Measurability of the random solution is more demanding in the hyperbolic case [5, Proposition 1.5]. Uniqueness is then shown analogously to Theorem 2.3. □

For the numerical analysis, we assume additional regularity of the random solution.

Assumption 8.3. We assume that the stochastic wave equation (79) - (82) admits a unique solution

$$u \in L^2(\Omega; C^2([0,T]; H^{p+1}(D)) \cap C^{2+j}([0,T]; L^2(D))), \tag{88}$$

for $p \geq 1$, $j = 1, 2$.

8.1 Space Semidiscretization

The spatial semidiscretization of the stochastic wave equation is analogous to the spatial approximation presented in Section 4 for the parabolic case. Let $\mathcal{T}_h$ be a regular simplicial mesh of the polyhedral domain $D \subset \mathbb{R}^d$ with meshwidth $h = \max_{K \in \mathcal{T}_h}\{\text{diam}(K)\}$. The Finite Element space is then given by $V_h := \mathcal{S}_0^{p,1}(D, \mathcal{T}_h)$ with $N_h = \dim V_h$ degrees of freedom.

For a given realization $\hat{a}(x) \in L^\infty(D)$ of the stochastic diffusion coefficient $a(x,\omega)$, the semidiscrete problem reads: Given $f \in L^2(I;H)$, for an approximation $u_{0,h} \in V_h$, $u_{1,h} \in V_h$ of the initial data, find $u_h \in H^2(I;V_h)$ such that for all $v_h \in V_h$ holds:

$$(\tfrac{d^2u_h(t)}{dt^2}, v_h) + a(u_h, v_h) = (f(t,x), v_h), \tag{89}$$

$$u_h(0) = u_{0,h}, \quad \tfrac{du_h}{dt}(0) = u_{1,h}. \tag{90}$$

In an arbitrary basis $\{\Phi_j\}_{j=1}^{N_h}$ of the finite dimensional subspace V_h, the semidiscrete problem is reformulated as a system of ODEs

$$\boldsymbol{M}\underline{\ddot{u}}(t) + \boldsymbol{A}\underline{u}(t) = \underline{f}(t), \quad \underline{u}(0) = \underline{u}_0, \ \underline{\dot{u}}(0) = \underline{u}_1. \tag{91}$$

As before, $\boldsymbol{M}_{ij} = (\Phi_i, \Phi_j)$ is the mass matrix, $\boldsymbol{A}_{ij} = a(\Phi_i, \Phi_j)$ is the stiffness matrix, $\underline{f}(t) = (f_j) = (f(t), \Phi_j)$ is the load vector, $\underline{u}_0$ and $\underline{u}_1$ are the vectors of coefficients of the initial data and $\underline{u}$ is the seeked vector of solution coefficients.

Theorem 8.4. *Under Assumptions 8.1 and 8.3 , the solution $u_h(t)$ of the semidiscrete problem* (89), (90) *admits the error bound*

$$\begin{aligned} \|u_h(t) - u(t)\|_V + \|\dot{u}_h(t) - \dot{u}(t)\|_H \ \le C\Big(& \|u_{0,h} - R_h u_0\|_V + \|u_{1,h} - R_h u_1\|_H \\ & + \|(I - R_h)u(t)\|_V + \|(I - R_h)\dot{u}(t)\|_H \\ & + \textstyle\int_0^t \|(I - R_h)\ddot{u}(s)\|_H ds \Big). \end{aligned} \tag{92}$$

Here, the constant C depends on a_-, a_+ and the Ritz projection $R_h : V \to V_h$ is defined by (24).

Proof. Inspecting the proof of Theorem 1.8 in [10], it can be deduced that the constant C depends only on a_-, a_+ and is thus uniformly bounded for all realizations of the stochastic coefficient $a(x,\omega)$. □

8.2 Newmark Time Stepping

We solve the system of ODEs (91) by Newmark's time stepping scheme. For $N_t \in \mathbb{N}$ we define the timestep $k = T/N_t$ and $t_m = mk$, $m = 0, \dots, N_t$. Newmark's scheme provides a sequence of tuples $(\underline{y}_m, \underline{z}_m)_{m=0}^{N_t} \approx (\underline{u}(t_m), \underline{\dot{u}}(t_m))_{m=0}^{N_t}$ approximating the solution and its derivative. For a pair of parameters $(\beta, \gamma) \in \mathbb{R} \times \mathbb{R}$ the fully discrete problem reads:

$$\begin{aligned} \tfrac{1}{k^2}\boldsymbol{M}(\underline{y}_{m+1} - \underline{y}_m - k\underline{z}_m) + \boldsymbol{A}(\beta\underline{y}_{m+1} + (\tfrac{1}{2} - \beta)\underline{y}_m) \\ = \beta\underline{f}(t_{m+1}) + (\tfrac{1}{2} - \beta)\underline{f}(t_m), \quad 0 \le m \le N_t - 1 \end{aligned} \tag{93}$$

$$\begin{aligned} \tfrac{1}{k}\boldsymbol{M}(\underline{z}_{m+1} - \underline{z}_m) + \boldsymbol{A}(\gamma\underline{y}_{m+1} + (1 - \gamma)\underline{y}_m) \\ = \gamma\underline{f}(t_{m+1}) + (1 - \gamma)\underline{f}(t_m), \quad 0 \le m \le N_t - 1 \end{aligned} \tag{94}$$

$$\underline{y}_0 = \underline{u}_0, \quad \underline{z}_0 = \underline{u}_1. \tag{95}$$

We need some preliminaries to state the convergence results. For fixed $\omega \in \Omega$, by the Spectral Theorem, there exists a finite sequence of eigenvalues $0 < \lambda_{1,h}(\omega) \leq \lambda_{2,h}(\omega) \leq \ldots \leq \lambda_{N_h,h}(\omega)$ and an H-orthonormal basis of eigenfunctions $\{w_{m,h}(\omega)\}_{m=1}^{N_h} \subset V_h$ such that for all $v_h \in V_h$ holds

$$a(w_{m,h}(\omega), v_h) = \lambda_{m,h}(\omega)(w_{m,h}(\omega), v_h). \tag{96}$$

We need uniform boundedness of the eigenvalues for all realizations in order to obtain convergence:

Proposition 8.5. *Under Assumption 8.1, for a given Finite Element space $V_h \subset V$, there exist constants $0 < \lambda_- \leq \lambda_+ < \infty$ such that*

$$0 < \lambda_- \leq \lambda_{1,h}(\omega) \leq \cdots \leq \lambda_{N_h,h}(\omega) \leq \lambda_+ < \infty \qquad \text{for all } \omega \in \Omega. \tag{97}$$

Proof. Observe that

$$a(w_{m,h}(\omega), w_{m,h}(\omega)) = \lambda_{m,h}(\omega)\|w_{m,h}(\omega)\|_H^2,$$

and

$$a_-\|w_{m,h}(\omega)\|_V^2 \leq a(w_{m,h}(\omega), w_{m,h}(\omega)) \leq a_+\|w_{m,h}(\omega)\|_V^2.$$

We first show, that the eigenvalues are bounded away from zero: By Poincaré's inequality we have

$$a_-\|w_{m,h}(\omega)\|_V^2 \leq a(w_{m,h}(\omega), w_{m,h}(\omega)) = \lambda_{m,h}(\omega)\|w_{m,h}(\omega)\|_H^2 \leq C\lambda_{m,h}(\omega)\|w_{m,h}(\omega)\|_V^2,$$

and thus

$$\lambda_{1,h}(\omega) \geq C^{-1}a_- > 0 \qquad \text{for all } \omega \in \Omega.$$

For the boundedness of the largest eigenvalue, we use the fact that all norms are equivalent in finite dimensional vector spaces: There exists a constant $C > 0$ such that $C^{-1}\|v_h\|_H \leq \|v_h\|_V \leq C\|v_h\|_H$ for all $v_h \in V_h$. In particular, we obtain for $w_{m,h}(\omega) \in V_h$

$$\lambda_{m,h}(\omega)\|w_{m,h}(\omega)\|_H^2 = a(w_{m,h}(\omega), w_{m,h}(\omega)) \leq a_+\|w_{m,h}(\omega)\|_V^2 \leq Ca_+\|w_{m,h}(\omega)\|_H^2,$$

and thus

$$\lambda_{N_h,h}(\omega) \leq Ca_+ < \infty \qquad \text{for all } \omega \in \Omega.$$

□

We can now state the convergence results for the Newmark scheme.

Theorem 8.6. *Under Assumptions 8.1 and 8.3, for*

$$\beta \geq 0 \quad \text{and} \quad \vartheta := \gamma - \tfrac{1}{2} \geq 0,$$

the approximate solution tuples $(u_h^m, z_h^m)_{m=0}^{N_t} \subset V_h \times V_h$ *corresponding to* $(\underline{y}_m, \underline{z}_m)_{m=0}^{N_t} \subset \mathbb{R}^{N_h} \times \mathbb{R}^{N_h}$ *generated by Newmark's time stepping scheme* (93) - (95) *satisfy the following error bound for* $m = 0, 1, \ldots, N_t$*:*

$$\begin{aligned} \|u_h^m - u(t_m)\|_H \leq \; & C\Big(\|u_{0,h} - R_h u_0\|_H + \|u_{1,h} - R_h u_1\|_H + \|(I - R_h)u(t)\|_H \\ & + \textstyle\int_0^{t_m} \big\{\|(I - R_h)\ddot{u}(s)\|_H + k^j \|\frac{d^{2+j}u}{dt^{2+j}}(s)\|_H\big\} ds\Big), \end{aligned} \tag{98}$$

provided that the CFL condition

$$k^2 \lambda_{N_h,h}(\omega) \leq \begin{cases} K & \text{if } \beta \geq (1+\vartheta)^2/4, \\ \frac{4(1-\epsilon)}{(1+\vartheta)^2 - 4\beta} & \text{if } \beta < (1+\vartheta)^2/4, \end{cases} \tag{99}$$

holds for all $\omega \in \Omega$ *and some* $K > 0$, $0 < \epsilon < 1$. *In* (98) *we have* $j = 2$ *if* $\vartheta = 0$ *and if the solution is sufficiently regular; we have* $j = 1$ *else. The constant* C *is independent of* $\omega \in \Omega$ *but depends on* K, λ_- *and* ϵ *in the sense that* $C \to \infty$ *as* $K \to \infty$, $C \to \infty$ *as* $\lambda_- \downarrow 0$ *and* $C \to \infty$ *for* $\epsilon \downarrow 0$.

Proof. According to Proposition 8.5 it is possible to satisfy the CFL condition (99). Inspecting the proof of Theorem 1.13 in [10] it can be deduced that the CFL condition (99) is sufficient for uniform boundedness of the constant C for all realizations of the stochastic coefficient $a(x, \omega)$. Note that all eigenvalues are bounded away from zero according to Proposition 8.5. This fact, and (99), is needed for stability of Newmark's scheme [10, Theorem 1.12]. □

As we implicitly assume compatible initial data in (88), we may choose

$$u_{0,h} := R_h u_0 \qquad \text{and} \qquad u_{1,h} := R_h u_1 \tag{100}$$

as approximate initial data. We use Proposition 4.4 to estimate the spatial error and immediately obtain the following corollary.

Corollary 8.7. *Assume* $\hat{a}(x) \in W^{1,\infty}(D)$ *as in* (83), $f \in C^0([0,T]; H)$ *and initial data given by* (100). *Under Assumption 8.3 and CFL condition* (99), *for* $\beta \geq 0$ *and* $\vartheta = \gamma - \frac{1}{2} \geq 0$, *the approximate solution* $(u_h^m)_{m=0}^{N_t} \subset V_h$ *generated by the fully discrete scheme* (93) - (95) *admits the error bound*

$$\|u_h^m - u(t_m)\|_H \leq C(h^{p+1} + k^j) \|u\|_{C^2([0,T];H^{p+1}(D)) \cap C^{2+j}([0,T];L^2(D))}, \tag{101}$$

where $j = 2$ *if* $\vartheta = 0$ *and if the solution is sufficiently regular and* $j = 1$ *else.*

8.3 Single-Level Monte Carlo Method

We can now use the concepts presented in Section 6 to estimate the mean field of the random solution $u \in \mathcal{X}^*$ of the stochastic wave equation (79) - (82). On a given mesh $\mathcal{T}_h$ with meshwidth h, for a fixed timestep k and for M i.i.d. realizations $\hat{a}^i \in W^{1,\infty}(D)$ we denote by ${}^i\hat{u}_{h,k}^m$ the approximation of the exact solution sample ${}^i\hat{u}(t_m)$ at time $t_m = mk$ according to (93) - (95). We estimate the mean field of the solution at time t_m by

$$E_M[u_{h,k}^m] := \frac{1}{M}\sum_{i=1}^{M} {}^i\hat{u}_{h,k}^m \approx \mathbb{E}[u(t_m)]. \tag{102}$$

We pick up an error induced by spatial approximation, time stepping and from the Monte Carlo estimator. We have the following error bound.

Theorem 8.8. *Assume $a(\omega,\cdot) \in W^{1,\infty}(D)$ as in (83), $f \in C^0([0,T];H)$ and initial data given by (100). Under Assumption 8.3 and CFL condition (99), for $\beta \geq 0$ and $\vartheta = \gamma - \frac{1}{2} \geq 0$, the Monte Carlo estimator (102) admits the error bound*

$$\begin{aligned}\left\|\mathbb{E}[u(t_m)] - E_M[u_{h,k}^m]\right\|_{L^2(\Omega;L^2(D))} &\leq M^{-1/2}\|u(t_m)\|_{L^2(\Omega;L^2(D))} \\ &\quad + C(h^{p+1}+k^j)\|u\|_{L^2(\Omega;C^2([0,T];H^{p+1}(D))\cap C^{2+j}([0,T];L^2(D)))}.\end{aligned} \tag{103}$$

Here, $j = 2$ if $\vartheta = 0$ and if the solution is sufficiently regular and $j = 1$ else.

Proof. The proof is completely analogous to the proof of Theorem 6.2. We split the error into a purely statistical part and a discretization error:

$$\begin{aligned}\left\|\mathbb{E}[u(t_m)] - E_M[u_{h,k}^m]\right\|_{L^2(\Omega;L^2(D))} &\leq \|\mathbb{E}[u(t_m)] - E_M[u(t_m)]\|_{L^2(\Omega;L^2(D))} \\ &\quad + \left\|E_M[u(t_m) - u_{h,k}^m]\right\|_{L^2(\Omega;L^2(D))} =: I + II\end{aligned}$$

The first term is estimated with Lemma 6.1,

$$I \leq M^{-1/2}\|u(t_m)\|_{L^2(\Omega;L^2(D))}.$$

The second term is bounded by Corollary 8.7

$$\begin{aligned}II &\leq M^{-1}\sum_{i=1}^{M}\left\|{}^i\hat{u}(t_m) - {}^i\hat{u}_{h,k}^m\right\|_{L^2(\Omega;L^2(D))} \\ &\leq C(h^{p+1}+k^j)\|u\|_{L^2(\Omega;C^2([0,T];H^{p+1}(D))\cap C^{2+j}([0,T];L^2(D)))}.\end{aligned}$$

□

On a given grid $\mathcal{T}_h$ with meshwidth h, we chose the stepsize $k = O(h^{(p+1)/j})$ and the sample size $M = O(h^{-2(p+1)})$ to equilibrate the errors.
We proceed with the analysis of the computational cost of the Monte Carlo estimator (102). Eliminating $\underline{z}_m$ in Newmark's scheme (93)-(95), we obtain for $0 \leq m \leq N_t - 2$:

$$(104)\ \tfrac{1}{k^2}\boldsymbol{M}(\underline{y}_{m+2} - 2\underline{y}_{m+1} + \underline{y}_m) + \boldsymbol{A}(\beta\underline{y}_{m+2} + (\tfrac{1}{2} - 2\beta + \gamma)\underline{y}_{m+1} + (\tfrac{1}{2} + \beta - \gamma)\underline{y}_m) \\ = \beta\underline{f}(t_{m+2}) + (\tfrac{1}{2} - 2\beta + \gamma)\underline{f}(t_{m+1}) + (\tfrac{1}{2} + \beta - \gamma)\underline{f}(t_m),$$

$$(105)\ \tfrac{1}{k^2}\boldsymbol{M}(\underline{y}_1 - \underline{u}_0 - k\underline{u}_1) + \boldsymbol{A}(\beta\underline{y}_1 + (\tfrac{1}{2} - \beta)\underline{u}_0) = \beta\underline{f}(t_1) + (\tfrac{1}{2} - \beta)\underline{f}(t_0).$$

Thus, for each timestep, we must solve a linear system of the form

$$(\boldsymbol{M} + \beta k^2 \boldsymbol{A})\underline{y}_{m+1} = \underline{\zeta}_m.$$

We make the following assumption on the solver for this linear system:

Assumption 8.9. For each timestep t_l, $l = 1, \ldots, m$, an approximate solution $\tilde{u}_h^l$ of the deterministic problem (104) - (105) is obtained in $O(N_h^\gamma)$ work, $1 \leq \gamma \leq 3$, such that the exact solution $u_h^m \in V_h$ corresponding to $\underline{y}_m \in \mathbb{R}^{N_h}$ is approximated to order

$$\|u_h^m - \tilde{u}_h^m\|_{L^2(D)} \leq C(h^{p+1} + k^j).$$

Under Assumption 8.9, the computational cost of the Monte Carlo estimator (102) is then given by $O(M \cdot N_t \cdot N_h^\gamma) = O(N_h^{2(p+1)/d+(p+1)/(jd)+\gamma})$. This is in exact accordance with the parabolic case. The contribution due to Monte Carlo sampling to the overall complexity is reduced by a multi-level approach analogous to Section 6.2.

8.4 Multi-Level Monte Carlo Method

By $\{\mathcal{T}_l\}_{l=0}^\infty$ we denote a sequence of regular simplicial meshes obtained by uniform mesh refinement. For $l \geq 0$ the meshwidth of $\mathcal{T}_l$ is then $h_l = 2^{-l}h_0$. Let $V_l := \mathcal{S}_0^{p,1}(D, \mathcal{T}_l)$ be the Finite Element space with $N_l^x := \dim V_l = O(2^{dl})$ degrees of freedom. We introduce a level-dependent timestep $k_l = T/N_l^t$. In order to equilibrate the spatial error and time stepping error, we choose $k_l = O(h_l^{(p+1)/j})$. The number of timesteps is then given by $N_l^t = O(h_l^{-(p+1)/j}) = O(N_l^{x(p+1)/(jd)})$.
We then denote by $u_l := u_{h_l,k_l}^{m_l} \in V_l$ the approximation of $u(t_m)$ according to (93) - (95) on mesh $\mathcal{T}_l$ with timestep k_l at time $t_m = m_l k_l$, $l \geq 1$.
The Multi-Level Monte Carlo estimator is then defined by

$$(106)\qquad E^L[u(t_m)] := \sum_{l=1}^{L} E_{M_l}[u_l - u_{l-1}] \approx \mathbb{E}[u(t_m)].$$

We have the following error estimate.

Lemma 8.10. *Assume $a(\omega,\cdot) \in W^{1,\infty}(D)$ as in (83), $f \in C^0([0,T];H)$ and initial data given by (100). Under Assumption 8.3 and CFL condition (99), for $\beta \geq 0$ and $\vartheta = \gamma - \frac{1}{2} \geq 0$, the Multi-Level Monte Carlo estimator (106) admits the error bound*

(107)

$$\left\|\mathbb{E}[u(t_m)] - E^L[u(t_m)]\right\|_{L^2(\Omega;L^2(D))} \leq C\left(h_L^{p+1} + \sum_{l=1}^L h_l^{p+1} M_l^{-1/2}\right) \times \|u\|_{L^2(\Omega;C^2([0,T];H^{p+1}(D))\cap C^{2+j}([0,T];L^2(D)))}.$$

Proof. Analogously to the proof of Lemma 6.4, we rewrite the error as

$$\begin{aligned}\left\|\mathbb{E}[u(t_m)] - E^L[u(t_m)]\right\|_{L^2(\Omega;L^2(D))} &\leq \|\mathbb{E}[u(t_m)] - \mathbb{E}[u_L]\|_{L^2(\Omega;L^2(D))} \\ &\quad + \sum_{l=1}^L \|\mathbb{E}[u_l - u_{l-1}] - E_{M_l}[u_l - u_{l-1}]\|_{L^2(\Omega;L^2(D))} \\ &=: I + II.\end{aligned}$$

The first term is bounded by Jensen's inequality, Corollary 8.7 and Hölder's inequality,

$$I \leq \mathbb{E}[\|u(t_m) - u_L\|_{L^2(D)}] \leq C(h_L^{p+1} + k_L^j)\|u\|_{L^2(\Omega;C^2([0,T];H^{p+1}(D))\cap C^{2+j}([0,T];L^2(D)))}.$$

For the second term, we estimate each addend as follows:

$$\begin{aligned}\|\mathbb{E}[u_l - u_{l-1}] - E_{M_l}[u_l - u_{l-1}]\|_{L^2(\Omega;L^2(D))} &\leq M_l^{-1/2}\|u_l - u_{l-1}\|_{L^2(\Omega;L^2(D))} \\ &\leq M_l^{-1/2}\left(\|u(t_m) - u_l\|_{L^2(\Omega;L^2(D))} + \|u(t_m) - u_{l-1}\|_{L^2(\Omega;L^2(D))}\right) \\ &\leq C M_l^{-1/2}(h_l^{p+1} + k_l^j + h_{l-1}^{p+1} + k_{l-1}^j)\|u\|_{L^2(\Omega;C^2([0,T];H^{p+1}(D))\cap C^{2+j}([0,T];L^2(D)))} \\ &\leq C(2^{p+1}+1)M_l^{-1/2}(h_l^{p+1} + k_l^j)\|u\|_{L^2(\Omega;C^2([0,T];H^{p+1}(D))\cap C^{2+j}([0,T];L^2(D)))}.\end{aligned}$$

Here we used Lemma 6.1 and Corollary 8.7. A summation over $l = 1,\dots,L$ and observing that $k_l = O(h_l^{(p+1)/j})$ completes the proof. □

By optimizing the sample sizes on each level, we recover the same convergence rate as single-level MC on the finest level with a reduced computational cost.

Theorem 8.11. *Assume $a(\omega,\cdot) \in W^{1,\infty}(D)$ as in (83), $f \in C^0([0,T];H)$ and initial data given by (100). Let the approximate solution samples be generated by the fully discrete system (93) - (95) with a pair of parameters $\beta \geq 0$ and $\vartheta = \gamma - \frac{1}{2} \geq 0$ and the timestep satisfying the CFL condition (99) on each mesh level. Under Assumptions 8.3 and 8.9 the MLMC estimator (106) of the expectation of the solution of the stochastic wave equation (79)-(82) computed with M_l samples on mesh level h_l and with $k_l = O(h_l^{(p+1)/j})$ timesteps, where the samples are distributed as*

$$M_l = l^{2+\epsilon}2^{2(L-l)(p+1)}, \quad l = 1,2,\dots,L,$$

with some $\epsilon > 0$ arbitrarily small, admits the error bound

$$\left\|\mathbb{E}[u(t_m)] - E^L[u(t_m)]\right\|_{L^2(\Omega;L^2(D))} \leq Ch_L^{p+1}\|u\|_{L^2(\Omega;C^2([0,T];H^{p+1}(D))\cap C^{2+j}([0,T];L^2(D)))}.$$

The computational cost is bounded by

$$Work(L) \leq C(\delta)\begin{cases} (N_L^x)^\gamma N_L^t(\log N_L^x)^{2+\epsilon} & \text{if } \delta < 0, \\ (N_L^x)^\gamma N_L^t(\log N_L^x)^{3+\epsilon} & \text{if } \delta = 0, \\ (N_L^x)^{\gamma+\delta/d} N_L^t(\log N_L^x)^{2+\epsilon} & \text{if } \delta > 0, \end{cases}$$

where $\delta = (p+1)(2-1/j) - d\gamma$.

Proof. The proof corresponds one-to-one the proof of Theorem 6.5. As the MLMC estimator (106) admits the same convergence rate in the hyperbolic case, we can adopt the optimal distribution of samples from Theorem 6.5. Under Assumption 8.9 the estimator admits the same computational cost as solving the stochastic diffusion equation and the work estimate remains therefore valid for the hyperbolic case. □

Remark 8.12. All conclusions from the parabolic section hold true for the hyperbolic case: For low order approximations the contribution to the overall computational cost due to sampling is logarithmic in the spatial degrees of freedom and thus negligible. However, in order to preserve increased convergence rates induced by high order approximations in space and time, the number M_l of samples on each level need to be increased. This results in an algebraic contribution of sampling to the overall complexity.
The MLMC approximation of the mean field of random solutions is thus particularly efficient with low order approximations in space and time.

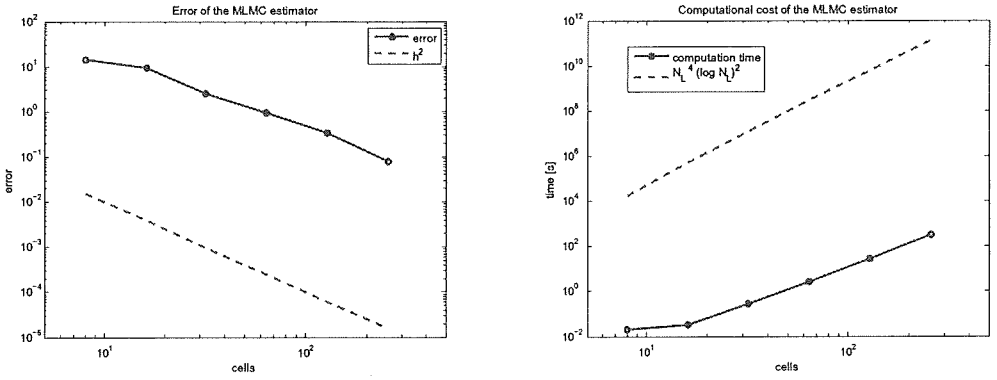
9 Numerical Results

We discuss implementation issues and validate the theory based on a one dimensional example of the stochastic wave equation. In the unit interval $D = (0,1)$ we consider

$$\begin{array}{ll} \frac{\partial^2 u}{\partial t^2} - \frac{\partial}{\partial x}(a(x,\omega)\frac{\partial}{\partial x}u) = 0 & x \in D,\ 0 < t < 1, \\ u(t,x,\omega) = 0 & x \in \partial D,\ 0 < t < 1, \\ u(0,x,\omega) = \sin(\pi x),\ \frac{\partial u}{\partial t}(0,x,\omega) = 0 & x \in D. \end{array}$$

The stochastic coefficient $a(x,\omega) := 2 + Y(\omega)\sin(\pi x)$ with a uniformly distributed random variable $Y(\omega) \sim \mathcal{U}(-1,1)$ satisfies (83).
For the approximation in space, we initially choose a uniform partition $\mathcal{T}_0$ of the domain D with meshwidth $h_0 := 2^{-3}$. Uniform mesh refinement provides a nested family of regular meshes

$$\mathcal{T}_l = \{I_{l,j}\}_{j=1}^{N_l},\ I_{l,j} = (x_{j-1}^l, x_j^l),\ x_j^l = jh_l,\ h_l = 2^{-l}h_0 = 2^{-(l+3)},\ l \geq 0.$$



(a) Error

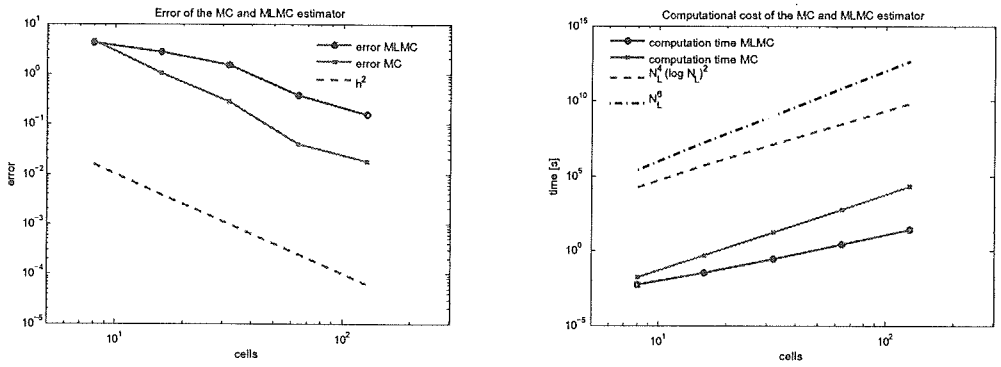
(b) Work

Figure 1: Rate of convergence of the MLMC estimator for the approximation of $\mathbb{E}[u(T)]$ at final time $T = 1$ in the L^2-norm in space and the total CPU-time needed to compute $E^L[u(T)]$.

We use piecewise linear polynomials $V_l := \mathcal{S}_0^{1,1}(D, \mathcal{T}_l)$ with strongly imposed Dirichlet boundary conditions. The standard hat basis $(b_l)_i$ is given by

$$(b_l)_i(x) = \begin{cases} 1 - |x - x_i^l|/h_l & \text{for } x_{i-1}^l \leq x \leq x_{i+1}^l, \\ 0 & \text{else.} \end{cases}$$

This leads to second order approximations in the L^2-norm in space.
We use Newmark's scheme for discretization in time. The linear system (93) - (95) is solved using the LU matrix factorization provided by Matlab. The hat basis $(b_l)_i$ induces a tridiagonal mass matrix $\boldsymbol{M}$ and stiffness matrix $\boldsymbol{A}$. The computational cost is therefore linear in the system size, we have $\gamma = 1$ in Assumption 8.9. As Newmark's scheme is second order in time, we choose the timestep $k = O(h_l) = O(2^{-l})$.
To compute the MLMC estimator (106) we scale the optimal sample distribution such that four samples are used on the finest grid, we use $M_l = 4(l/L)^{2+\varepsilon}2^{4(L-l)}$. Note that the solution samples of coarse levels need to be interpolated onto the finest grid.
For the error analysis, we compute a reference solution by adding two additional refinement steps. The rate of convergence of the MLMC approximation is shown in Figure 1a. The results coincide with the predicted rate of convergence $O(h_L^2)$ derived in Theorem 8.11.
Figure 1b shows the CPU-time needed to compute the MLMC estimator. From theoretical considerations we expect the computational cost to behave as $O(N_L^4(\log N_L)^2)$. To see this, note that Newmark time stepping is of second order and that the linear system is solved in linear complexity in each time step. We therefore have $\delta = (p+1)(2-1/j) - d\gamma = 2$ in



(a) Error

(b) Work

Figure 2: Rate of convergence and complexity of the single-level MC estimator versus multi-level MC.

Theorem 8.11. The predicted complexity is in accordance with numerical experiments as shown in Figure 1b.
We compare these results to the single-level Monte Carlo estimator in Figure 2. Again, the plain MC algorithm provides an approximation which is second order in the meshwidth. The error analysis given in Section 8 suggests to use $M = O(N_l^4)$ samples to equilibrate statistical and spatial errors. The overall complexity of the single-level Monte Carlo method is thus $O(N_l^6)$ which matches with numerical results given in Figure 2b. The multi-level Monte Carlo approach is thus clearly favorable.

10 Summary and Conclusions

We consider a class of either parabolic or second order hyperbolic partial differential equations with random coefficients. Assuming adequate regularity of the random solution, we analyse numerical schemes to approximate the mean field of the stochastic solution.
We propose Monte Carlo methods together with standard Finite Element methods in space and various time stepping schemes. We show these compound Monte Carlo type methods to converge, but the overall efficiency (i.e. error versus computational cost) is relatively poor.
We therefore introduce multi-level Monte Carlo methods and prove them to admit the same convergence rate as single-level MC algorithms, whereas the computational work is considerably reduced. In fact, our error and complexity analysis reveals that for low order Finite Element approximations in space, the contribution due to sampling to the overall compu-

tational cost is only logarithmic in the spatial degrees of freedom. Hence, the proposed MLMC algorithm provides an approximation of the mean field of the stochastic solution with an efficiency which is comparable to one solve of a deterministic partial differential equation of the same type.
However, in order to preserve increased convergence rates induced by high order approximations in space and time, the number of samples needs to be increased on each level. This results in an algebraic contribution of sampling to the overall complexity.
Therefore, the proposed MLMC method is particularly suitable for the approximation of solutions that admit low spatial and temporal regularity.
Numerical experiments for the stochastic wave equation in one space dimension confirm the theory.

11 References

[1] A. Barth, Ch. Schwab and N. Zollinger, Multi-Level Monte Carlo FEM for Elliptic PDE's with stochastic Coefficients. Report 2010-18, Seminar for Applied Mathematics, ETH Zürich, 2010 (in review).

[2] D. Braess, Finite elements, 3rd ed., Cambridge University Press, Cambridge, 2007.

[3] P. Grisvard, Singularities in boundary value problems, RMA 22, Masson, Paris, 1992.

[4] V.H. Hoang and C. Schwab, Sparse Tensor Galerkin Discretizations for parametric and random parabolic PDEs I: Analytic regularity and gpc-Approximation. Report 2010-11, Seminar for Applied Mathematics, ETH Zürich, 2010 (in review).

[5] V.H. Hoang and C. Schwab, Analytic regularity and gpc approximation for parametric and random 2nd order hyperbolic PDEs. Report 2010-19, Seminar for Applied Mathematics, ETH Zürich, 2010 (in review).

[6] J.-L. Lions and E. Magenes, Non-homogeneous boundary value problems and applications. Vol. I, Springer-Verlag, New York, 1972. Translated from French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181.

[7] S. Mishra and Ch. Schwab, Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data. Report 2010-24, Seminar for Applied Mathematics, ETH Zürich, 2010 (in review).

[8] D. Schötzau, hp-DGFEM for parabolic evolution problems. Doctoral Dissertation No. 13041, ETH Zürich, 1999.

[9] D. Schötzau and C. Schwab, Time Discretization of Parabolic Problems by the *hp*-Version of the Discontinuous Galerkin Finite Element Method. SIAM Journal on Numerical Analysis, Vol. 38, No. 3 (2000), pp. 837 - 875.

[10] C. Schwab, Numerical Analysis of Hyperbolic PDEs, Lecture Notes for FS 2008.

[11] C. Schwab and R. Stevenson, Space-time adaptive wavelet methods for parabolic evolution problems, Math. Comp. 78 (2009), 1293-1318.

[12] T. von Petersdorff and C. Schwab, Wavelet-discretizations of parabolic integro-differential equations, SIAM Journal on Numerical Analysis, Vol. 41, No. 1 (2004), pp. 159-180

[13] V. Thomée, Galerkin finite element methods for parabolic problems, 2nd ed., Springer, Berlin, 2006.

[14] T. Werder, K. Gerdes, D. Schötzau and C. Schwab, *hp* Discontinuous Galerkin Time Stepping for Parabolic Problems. Comput. Methods Appl. Mech. Engrg. 190 49-50 (2001), pp. 6685-6708.

[15] J. Wloka, Partial differential equations, Cambridge University Press, Cambridge, 1987. Translated from German by C. B. Thomas and M. J. Thomas.