A MULTILEVEL MONTE CARLO ENSEMBLE SCHEME FOR
SOLVING RANDOM PARABOLIC PDES

YAN LUO ∗ AND ZHU WANG†

Abstract. A first-order, Monte Carlo ensemble method has been recently introduced for solving
parabolic equations with random coefficients in [26], which is a natural synthesis of the ensemble-
based, Monte Carlo sampling algorithm and the ensemble-based, first-order time stepping scheme.
With the introduction of an ensemble average of the diffusion function, this algorithm leads to a
single discrete system with multiple right-hand sides for a group of realizations, which could be
solved more efficiently than a sequence of linear systems. In this paper, we pursue in the same
direction and develop a new multilevel Monte Carlo ensemble method for solving random parabolic
partial differential equations. Comparing with the approach in [26], this method possesses a second-
order accuracy in time and further reduces the computational cost by using the multilevel Monte
Carlo sampling method. Rigorous numerical analysis shows the method achieves the optimal rate of
convergence. Several numerical experiments are presented to illustrate the theoretical results.

Key words. ensemble-based time stepping, multilevel Monte Carlo, random parabolic PDEs

1. Introduction. In this paper, we consider numerical solutions to the following
unsteady heat conduction equation in a random, spatially varying medium: to find a
random function, \( u: \Omega \times D \times [0,T] \to \mathbb{R} \) satisfying almost surely (a.s.)

\[
\begin{cases}
  u_t(\omega, x, t) - \nabla \cdot \left[(a(\omega, x)\nabla u(\omega, x, t)) \right] = f(\omega, x, t), & \text{in } \Omega \times D \times [0,T] \\
  u(\omega, x, t) = g(\omega, x, t), & \text{on } \Omega \times \partial D \times [0,T] \\
  u(\omega, x, 0) = u^0(\omega, x), & \text{in } \Omega \times D
\end{cases}
\]

where \( D \) is a bounded Lipschitz domain in \( \mathbb{R}^d \) \( (d = 1, 2, \text{ or } 3) \) and \( (\Omega, \mathcal{F}, P) \) is a
probability space with the sample space \( \Omega \), \( \sigma \)-algebra \( \mathcal{F} \), and probability measure \( P \);
diffusion coefficient \( a: \Omega \times D \to \mathbb{R} \) and body force \( f: \Omega \times D \times [0,T] \to \mathbb{R} \) are random
fields with continuous and bounded covariance functions.

Many numerical methods, either intrusive or non-intrusive, have been developed
for random partial differential equations (PDEs), see, e.g., in the review papers [16, 40]
and the references therein. For the random steady or unsteady heat equation, non-
intrusive numerical methods such as Monte Carlo methods are known for easy imple-
mentation but requiring a very large number of PDE solutions to achieve small errors;
while intrusive methods such as the stochastic Galerkin or collocation approaches can
achieve faster convergence but would require the solution of discrete systems that
couple all spatial and probabilistic degrees of freedom [2, 3, 41]. To improve the com-
putational efficiency of the non-intrusive approaches, other sampling methods such
as quasi-Monte Carlo, multilevel Monte Carlo (MLMC), Latin hypercube sampling
and Centroidal Voronoi tessellations can be used [29, 19, 8, 35]. In particular, the
MLMC method is designed to greatly reduce the computational cost by performing
most simulations at a low accuracy, while running relatively few simulations at a

∗School of Mathematical Sciences, University of Electronic Science and Technology of China,
No.2006, Xiyuan Ave, West Hi-Tech Zone, Chengdu, Sichuan 611731, China; and School of Mathe-
matics, Sichuan University, No.24 South Section 1, Yihuan Road, Chengdu, Sichuan 610064, China.
Research supported by the Young Scientists Fund of the National Natural Science Foundation of
China grant 11501088.
†Department of Mathematics, University of South Carolina, 1523 Greene Street, Columbia, SC
29208, USA (wangzhu@math.sc.edu). Research supported by the U.S. National Science Foundation
grant DMS-1522672 and the U.S. Department of Energy grant DE-SC0016540.

This manuscript is for review purposes only.
high accuracy. It was first introduced by Heinrich [18] for the computation of high-
dimensional, parameter-dependent integrals and was analyzed extensively by Giles
[7], Cliffe et al. applied the MLMC method to the elliptic PDEs with random coeffi-
cients and demonstrated its numerical superiority. Under the assumptions of uniform
coercivity and boundedness of the random parameter, numerical error of the MLMC
approximation has been analyzed in [4]. The result was extended in [5] for random
elliptic problems with weaker assumptions on the random parameter and a limited
spatial regularity.

Overall, the above mentioned sampling methods are ensemble-based. To quantify
probabilistic uncertainties in a system governed by random PDEs, an ensemble of in-
dependent realizations of the random parameters needs to be considered. In practice,
this process would involve solving a group of deterministic PDEs corresponding to all
the realizations. A straightforward solution strategy is to find numerical approximate
solutions of the deterministic PDEs from a sequence of discrete linear systems. Obvi-
ously, this approach ignores any possible relationships among the group members, thus
cannot improve the overall computational efficiency. To speed up the group of simu-
lations, current active research mainly starts from the perspective of numerical linear
algebra, and develops iterative algorithms that can take advantage of the relationship
in the sequence of discrete systems. For instance, subspace recycling techniques such
as GCRO with deflated restarting have been introduced in [33] for accelerating the
solutions of slowly-changing linear systems, which is further developed in [1] for cli-
mate modeling and uncertainty quantification applications. For sequences sharing a
common coefficient matrix, block iterative algorithms [17, 27, 31, 32, 36] have been
developed to solve the system with many right-hand sides. The algorithms have been
used to accelerate convergence even when there is only one right-hand side in [6, 32].
The block version of GCRO with deflated restarting was introduced in [34], and its
high-performance implementation is available in the Belos package of the Trilinos
project developed at US Sandia National Laboratories.

Recently, the Monte Carlo ensemble method was introduced by the authors of this
paper for solving the random heat equations in [26]. This method is motivated by
the ensemble-based time stepping algorithm, which was proposed for solving Navier-
Stokes incompressible flow ensembles in [23, 20, 22, 24, 37, 21] and for simulating
ensembles of parameterized Navier-Stokes flow problems in [14, 15]. It has been
extended to MHD flows in [28] and to low-dimensional surrogate models in [12, 13].
The main idea is to manipulate the numerical scheme so that all the simulations in
the ensemble could share a common coefficient matrix. As a consequence, simulating
the ensemble only requires to solve a single linear system with multiple right-hand
sides, which could be easily handled by a block iterative solver and, thus, improves
the overall computational efficiency. Thus, the Monte Carlo ensemble method was
proposed in [26] for synthesizing a first-order, ensemble-based time-stepping and the
ensemble-based, Monte Carlo sampling method in a natural way, which speeds up the
numerical approximation of the random parabolic PDE solutions and other possible
quantities of interest. However, it is known that the Monte Carlo method, although
easy for implementations, is a computationally expensive random sampling approach.
Therefore, we develop a new method for solving the same random heat equations
with a better accuracy and efficiency in this paper: the new method is second-order
accurate in time, which improves the temporal accuracy of our previous work; it
employs the idea of multilevel Monte Carlo methods, which improves the sampling
efficiency comparing with the Monte Carlo. We further perform theoretical analysis on
the method and present numerical tests that illustrate our theoretical findings. Upon
the completion of this paper, we found the second-order ensemble-based time-stepping
scheme had been used in [9] for solving heat equation with uncertain conductivity,
however, without discussing the sampling error in their analysis.

The rest of this paper is organized as follows. In Section 2, we present some
notation and mathematical preliminaries. In Section 3, we introduce the multilevel
Monte Carlo ensemble scheme in the context of finite element (FE) methods. In
Section 4, we analyze the proposed algorithm, prove its stability and convergence, and
discuss its computational complexity. Numerical experiments are presented in Section
5, which illustrate the effectiveness of the proposed scheme on random parabolic
problems. A few concluding remarks are given in Section 6.

2. Notation and preliminaries. Denote the \( L^2(\Omega) \) norm and inner product
by \( \| \cdot \| \) and \( \langle \cdot, \cdot \rangle \), respectively. Let \( W^{s,q}(D) \) be the Sobolev space of functions having
generalized derivatives up to the order \( s \) in the space \( L^q(D) \), where \( s \) is a nonnegative
integer and \( 1 \leq q \leq +\infty \). The equipped Sobolev norm of \( v \in W^{s,q}(D) \) is denoted
by \( \| v \|_{W^{s,q}(D)} \). When \( q = 2 \), we use the notation \( H^s(D) \) instead of \( W^{s,2}(D) \).
As usual, the function space \( H^0(D) \) is the subspace of \( H^1(D) \) consisting of functions
that vanish on the boundary of \( D \) in the sense of trace, equipped with the norm
\( \| v \|_{H^0(D)} = (\int_D |\nabla v|^2 \, dx)^{1/2} \). When \( s = 0 \), we shall keep the notation with \( L^2(D) \)
instead of \( W^{0,q}(D) \). The space \( H^{-s}(D) \) is the dual space of bounded linear functions
on \( H^0(D) \). A norm for \( H^{-1}(D) \) is defined by
\[
\| f \|_{-1} = \sup_{0 \neq v \in H^0(D)} \frac{(f,v)}{|\nabla v|}:
\]

Let \( (\Omega, \mathcal{F}, P) \) be a complete probability space. If \( Y \) is a random variable in the
space that belongs to \( L^1_p(\Omega) \), its expected value is defined by
\[
E[Y] = \int_{\Omega} Y(\omega) \, dP(\omega).
\]
With the multi-index notation, \( \alpha = (\alpha_1, \ldots, \alpha_d) \) is a \( d \)-tuple of nonnegative in-
tegers with the length \( |\alpha| = \sum_{i=1}^d \alpha_i \). The stochastic Sobolev space \( \tilde{W}^{s,q}(D) = L^q_p(\Omega, \tilde{W}^{s,q}(D)) \) containing stochastic functions, \( v : \Omega \times D \to R \), that are measurable
with respect to the product \( \sigma \)-algebra \( \mathcal{F} \otimes B(D) \) and equipped with the averaged
\( \| v \|_{\tilde{W}^{s,q}(D)} = \left( E[\| v \|_{\tilde{W}^{s,q}(D)}^q] \right)^{1/q} = \left( E[\sum_{|\alpha| \leq s} \int_D |\partial^\alpha v|^q \, dx] \right)^{1/q} \), \( 1 \leq q < +\infty \).
Observe that if \( v \in \tilde{W}^{s,q}(D) \), then \( v(\omega, \cdot) \in W^{s,q}(D) \) a.s. and \( \partial^\alpha v(\cdot, x) \in L^p(D) \) a.e.
on \( D \) for \( |\alpha| < s \). In particular, we consider the Hilbert space \( L^2(H^s(D); 0,T) \) of
stochastic functions \( v : \Omega \times D \times [0,T] \to R \), in which any element \( v \) belongs to \( H^s(D) \)
for each \( 0 \leq t \leq T \) with the property that \( \| v \|_{\tilde{W}^{s,q}(D)} \) is square integrable on \([0,T] \);
and \( \tilde{H}^s(L^2(D); 0,T) \) in which any element \( v \) belongs to \( L^2(D) \) for each \( 0 \leq t \leq T \)
with the property that \( \| v \|_{L^2(D)} \) belongs to \( H^s(0,T) \).

3. Multilevel Monte Carlo ensemble method. Given statistical information
on the inputs of a random/stochastic PDE, uncertainty quantification fulfills the task
dermore, independent realizations, that is, deterministic PDEs at randomly selected sample
values. Usually, numerical simulations are implemented separately, thus the total
computational cost is simply multiplied as the sampling set grows. To improve the
efficiency, we propose an ensemble-based multilevel Monte Carlo method in this paper,

This manuscript is for review purposes only.
which is an extension of the Monte Carlo ensemble method we introduced in [26]. The new approach outperforms the previous one in both accuracy and efficiency, which is due to the combination of a second-order, ensemble-based time stepping scheme and the multilevel Monte Carlo method.

Next, we present the algorithm in the context of numerical solutions to the random PDE (1). For the spatial discretization, we use conforming finite elements, although other numerical methods could be applied as well. To fit in the hierarchic nature of multilevel Monte Carlo methods, we consider a sequence of quasi-uniform meshes comprising a set of shape-regular triangles (or tetrahedra), \( \{ \mathcal{T}_l \}_{l=0}^L \), for a polygonal (or polyhedral) domain \( D \). Denote the mesh size of \( \mathcal{T}_l \) by

\[
    h_l = \max_{K \in \mathcal{T}_l} \text{diam } K.
\]

Assume the sequence of meshes is generated by uniform refinements satisfying

\[
    h_l = 2^{-l} h_0.
\]

Define the function space \( H^1_g(D) = \{ v \in H^1(D) : v|_{\partial D} = g \} \) and the FE space

\[
    V_l^g := \{ v \in H^1_g(D) \cap H^{m+1}(D) : v|_K \text{ is a polynomial of degree } m, \forall K \in \mathcal{T}_l \}
\]

for a non-negative integer \( m \). The sequence of finite element spaces satisfies

\[
    V_0^g \subset V_1^g \subset \cdots \subset V_l^g \subset \cdots \subset V_L^g.
\]

Denoted by \( u_l(\omega, x, t_n) \) the finite element solution in \( V_l^g \) at the time instance \( t_n \). The MLMC FE solution at the \( L \)-th level mesh can be written as

\[
    u_L(\omega, x, t_n) = \sum_{l=1}^L (u_l(\omega, x, t_n) - u_{l-1}(\omega, x, t_n)) + u_0(\omega, x, t_n).
\]

Based on linearity of the expectation operator \( E[\cdot] \), we have

\[
    E[u_L(\omega, x, t_n)] = \mathbb{E} \left[ \sum_{l=1}^L (u_l(\omega, x, t_n) - u_{l-1}(\omega, x, t_n)) + u_0(\omega, x, t_n) \right]
\]

\[
    = \sum_{l=1}^L E[u_l(\omega, x, t_n) - u_{l-1}(\omega, x, t_n)] + E[u_0(\omega, x, t_n)].
\]

Numerically, the expected value of the FE solution on the \( l \)-th level, \( E[u_l(\omega, x, t_n)] \) is approximated by the sampling average \( \Psi_{J_l}^g = \Psi_{J_l} u_l(\omega, x, t_n) = \frac{1}{J_l} \sum_{j=1}^{J_l} u_l(\omega_j, x, t_n) \), where \( J_l \) is the sample size. Correspondingly, \( E[u_L(\omega, x, t_n)] \) is approximated by an unbiased estimator:

\[
    \Psi[u_L(\omega, x, t_n)] := \sum_{l=1}^L (\Psi_{J_l} u_l(\omega, x, t_n) - u_{l-1}(\omega, x, t_n)) + \Psi_{J_0} u_0(\omega, x, t_n).
\]

It is seen that, at each mesh level, a group of simulations needs to be implemented. Thus, it is natural to extend ensemble-based time stepping to such settings for reducing the computational cost. Next, we introduce the multilevel Monte Carlo ensemble (MLMCE) method to achieve this goal.

This manuscript is for review purposes only.
For simplicity of presentation, we assume that, at the \( l \)-th level, a uniform time partition with the time step \( \Delta t_l \) is used for the simulations and further set \( N_l = T / \Delta t_l \); \( J_l \) independent, identically distributed (i.i.d.) samples are selected, and the associated random functions are denoted by \( a_j \equiv a(\omega_j, \cdot) \), \( f_j \equiv f(\omega_j, \cdot) \), \( g_j \equiv g(\omega_j, \cdot) \), and \( u_{0,j} \equiv u^0(\omega_j, \cdot) \) for \( j = 1, \ldots, J_l \), and define the ensemble mean of the diffusion coefficient functions by

\[
\bar{a}_l := \frac{1}{J_l} \sum_{j=1}^{J_l} a(\omega_j, x).
\]

Here, we note that the corresponding exact solutions \( \{u(\omega_j, x, t)\}_{j=1}^{J_l} \) are i.i.d. Let \( u_{n,j,l} = u_t(\omega_j, x, t_n) \), the finite element approximation of \( u(\omega_j, x, t_n) \) at the \( l \)-th level.

The multilevel Monte Carlo ensemble method (MLMCE) applied to (1) solves the following group of simulations at the \( l \)-th level: for \( j = 1, \ldots, J_l \), given \( u_{0,j,l} \) and \( u_{1,j,l} \), to find \( u_{n+1,j,l} \in V_l^0 \) such that,

\[
\left( 3u_{n+1,j,l} - 4u_{n,j,l} + u_{n-1,j,l} \over 2\Delta t_l, v_l \right) + (\bar{a}_l \nabla u_{n+1,j,l}, \nabla v_l)
\]

\[
= -((a_j - \bar{a}_l) \nabla(2u_{n,j,l} - u_{n-1,j,l}), \nabla v_l) + (f_{n+1,j,l}, v_l), \quad \forall v_l \in V_l^0,
\]

for \( n = 1, \ldots, N_l - 1 \). Once the numerical solutions at all the \( L \) levels are found, the MLMCE approximates the random PDE solution at the time instance \( t_n \), \( \mathbb{E}[u(t_n)] \), by (3). Meanwhile, given a quantity of interest \( Q(u) \), one can analyze the outputs from the ensemble simulations, \( Q(u_h(\omega_1, \cdot)), \ldots, Q(u_h(\omega_J, \cdot)) \), to extract the underlying stochastic information of the system.

The MLMCE naturally combines the ensemble-based sampling method and the ensemble-based time stepping algorithm, and inherits advantages from both sides. As the MLMC, the method can reduce the computational cost by balancing the time step size, mesh size, and the number of samples at each level. Meanwhile, the ensemble-based time stepping algorithm leads to a discrete linear system (4) whose coefficient matrix is independent of \( J \). Indeed, denote the mass matrix by \( M_l \) that is associated with \( (v_l, v_l) \) and the stiffness matrix \( S_l \) that is related to \( \bar{a}_l \nabla v_l, \nabla v_l \), the coefficient matrix of (4) is \( \frac{3}{2\Delta t_l} M_l + S_l \). Hence, for evaluating \( J_l \) realizations, one only needs to solve one linear system with \( J_l \) right-hand sides, which leads to great computational savings comparing with a sequence of individual simulations: when the number of degrees of freedom is small, one only need to perform the LU factorization once instead of \( J_l \) times; when the number of degrees of freedom is large, one can use the block iterative algorithms to accelerate solutions. Next, we will analyze the stability and asymptotic error estimate of the MLMCE method.

4. Stability and error estimate. To simplify the presentation, we only consider equation (1) with the homogeneous boundary condition (that is, \( g = 0 \) and \( u_{n+1,j,l}^0 \in V_l^0 \) in the FE weak form (4)), while the nonhomogeneous cases can be similarly analyzed by incorporating the method of shifting. Meanwhile, we will include numerical test cases with nonhomogeneous boundary conditions in Section 5. As the MLMCE approximation is based on the MC solutions at various levels, we first analyze the ensemble-based single-level Monte Carlo in Subsection 4.1 and derive the error estimate for MLMCE in Subsection 4.2.

Assume the exact solution of (1) is smooth enough, in particular,

\[
u_j \in \tilde{L}^2(H^1_0(D) \cap H^{m+1}(D); 0, T) \cap \tilde{H}^1(H^{m+1}(D); 0, T) \cap \tilde{H}^2(L^2(D); 0, T)\]
and suppose
\[ f_j \in L^2(H^{-1}(D); 0, T). \]

Here we use the notation introduced in Section 2. We emphasize the assumed regularity only requires the random fields to be square integrable. Assume the following two conditions hold:

(i) There exists a positive constant \( \theta \) such that
\[ P\{\omega \in \Omega; \min_{x \in \mathcal{D}} a(\omega, x) > \theta\} = 1. \]

(ii) There exists a positive constant \( \theta_+ \), for \( l = 0, \ldots, L \), such that
\[ P\{\omega_j \in \Omega; |a(\omega, x) - \overline{a}_l|_\infty \leq \theta_+\} = 1. \]

Here, condition (i) guarantees the uniform coercivity a.s. and condition (ii) gives an upper bound of the distance from coefficient \( a(\omega, x) \) to the ensemble average \( \overline{a}_l \) a.s.

4.1. Single-level Monte Carlo ensemble finite element method. When \( \mathbb{E}[u(t_n)] \) is numerically approximated by \( \Psi^n_j \), the associated approximation error can be separated into two parts:
\[ \mathbb{E}[u(t_n)] - \Psi^n_j = (\mathbb{E}[u_j(t_n)] - \mathbb{E}[u^n_j]) + (\mathbb{E}[u^n_j] - \Psi^n_j) := \mathcal{E}^n + \mathcal{E}^n_S, \]
where we use the fact that \( \mathbb{E}[u(t_n)] = \mathbb{E}[u_j(t_n)] \). The finite element discretization error, \( \mathcal{E}^n = \mathbb{E}[u_j(t_n)] - u^n_j \), is controlled by the size of spatial triangulations \( T_l \) and time step; while the statistical sampling error, \( \mathcal{E}^n_S = \mathbb{E}[u^n_j] - \Psi^n_j \), is dominated by the number of realizations and variance. Next, we will first discuss the stability of the ensemble scheme (4) at the \( l \)-th level (Theorem 1), derive the bounds for \( \mathcal{E}^n_S \) (Theorem 3) and \( \mathcal{E}^n_t \) (Theorem 4), and then obtain the asymptotic error estimation (Theorem 5).

Theorem 1. Under conditions (i) and (ii), the scheme (4) is stable provided that
\[ \theta > 3\theta_+. \]

Furthermore, the numerical solution to (4) satisfies
\[ \frac{1}{4} \mathbb{E}[\|u^{N_i}_j\|^2] + \frac{1}{4} \mathbb{E}[\|2u^{N_i}_j - u^{N_i-1}_j\|^2] + \left( \frac{\theta}{3} - \theta_+ \right) \Delta t_i \sum_{n=1}^{N_i} \mathbb{E}[\|\nabla u^n_j\|^2] \]
\[ \leq \frac{\Delta t_i}{2(\theta - 3\theta_+)} \sum_{n=1}^{N_i-1} \mathbb{E}[\|\theta_j^{n+1}\|^2] + \frac{1}{4} \mathbb{E}[\|\theta_j^{n}\|^2] + \frac{1}{4} \mathbb{E}[\|2\theta_j^{n} - \theta_j^{n-1}\|^2] \]
\[ + \frac{\theta}{2} \Delta t_i \mathbb{E}[\|\nabla u_j^{n}\|^2] + \frac{\theta}{6} \Delta t_i \mathbb{E}[\|\nabla u_j^{n-1}\|^2]. \]

Proof. Choosing \( v_h = u_j^{n+1} \) in (4), we obtain
\[ \frac{3u^{n+1}_{j,l} - 4u^n_{j,l} + u^{n-1}_{j,l}}{2\Delta t} + \left( \overline{a}_l \nabla u^{n+1}_j, \nabla u^{n+1}_j \right) \]
\[ = - \left( (a_j - \overline{a}_l) \nabla (2u^n_{j,l} - u^{n-1}_{j,l}), \nabla u^{n+1}_j \right) + \left( f^{n+1}_j, u^{n+1}_j \right). \]
Multiplying both sides by $\Delta t_i$, integrating over the probability space and considering the coercivity, we get

$$
\frac{1}{4}E\left[\|u^{n+1}_{j,i}\|^2 + \|2u^{n+1}_{j,i} - u^n_{j,i}\|^2\right] - \frac{1}{4}E\left[\|u^n_{j,i}\|^2 + \|2u^n_{j,i} - u^{n-1}_{j,i}\|^2\right]
$$

$$
+ \frac{1}{4}E\left[\|u^{n+1}_{j,i} - 2u^n_{j,i} + u^{n-1}_{j,i}\|^2\right] + \Delta t_i \theta E\left[\|\nabla u^{n+1}_{j,i}\|^2\right]
$$

$$
\leq \Delta t_i E\left[\|(f_j^{n+1}, u^{n+1}_{j,i})\| + \Delta t_i \theta + E\left[\|(\nabla (2u^n_{j,i} - u^{n-1}_{j,i}), \nabla u^{n+1}_{j,i})\|\right].
$$

Apply Young’s inequality to the terms on the right-hand side (RHS), we have, for any $\beta_i > 0, i = 1, 2, 3$,

$$
E\left[\|(f_j^{n+1}, u^{n+1}_{j,i})\|\right] \leq \frac{\beta_1}{4}E\left[\|\nabla u^{n+1}_{j,i}\|^2\right] + \frac{1}{\beta_1}E\left[\|f_j^{n+1}\|^2\right],
$$

and

$$
\frac{\beta_2 + \beta_3}{2}E\left[\|\nabla u^{n+1}_{j,i}\|^2\right] + \frac{2}{\beta_2}E\left[\|\nabla u^n_{j,i}\|^2\right] + \frac{1}{2\beta_3}E\left[\|\nabla u^{n-1}_{j,i}\|^2\right].
$$

The term $\Delta t_i \theta E\left[\|\nabla u^{n+1}_{j,i}\|^2\right]$ on the left-hand side (LHS) can be split into several parts, for any $C_1 \in (0,1)$:

$$
\Delta t_i \theta E\left[\|\nabla u^{n+1}_{j,i}\|^2\right] = C_1 \Delta t_i \theta E\left[\|\nabla u^{n+1}_{j,i}\|^2\right] + (1 - C_1) \Delta t_i \theta E\left[\|\nabla u^{n+1}_{j,i}\|^2 - \|\nabla u^n_{j,i}\|^2\right]
$$

$$
+ (1 - C_1) \Delta t_i \theta E\left[\|\nabla u^n_{j,i}\|^2\right].
$$

Substituting (9)-(11) into (8), we get

$$
\frac{1}{4}(E\left[\|u^{n+1}_{j,i}\|^2\right] + E\left[\|2u^{n+1}_{j,i} - u^n_{j,i}\|^2\right]) - \frac{1}{4}(E\left[\|u^n_{j,i}\|^2\right] + E\left[\|2u^n_{j,i} - u^{n-1}_{j,i}\|^2\right])
$$

$$
+ \frac{1}{4}E\left[\|u^{n+1}_{j,i} - 2u^n_{j,i} + u^{n-1}_{j,i}\|^2\right] + (C_1 \theta - \frac{\beta_1}{4} - \frac{\beta_2 + \beta_3}{2} \theta) \Delta t_i E\left[\|\nabla u^{n+1}_{j,i}\|^2\right]
$$

$$
+ (1 - C_1) \Delta t_i \theta E\left[\|\nabla u^{n+1}_{j,i}\|^2 - \|\nabla u^n_{j,i}\|^2\right] + \frac{2}{3} \left(1 - C_1\right) \theta - \frac{2 \theta_+}{\beta_2} \Delta t_i E\left[\|\nabla u^n_{j,i}\|^2\right]
$$

$$
+ \left(\frac{1}{3} - C_1\right) \theta - \frac{\theta_+}{2 \beta_3} \Delta t_i E\left[\|\nabla u^{n-1}_{j,i}\|^2\right] \leq \frac{\Delta t_i}{\beta_1} E\left[\|f_j^{n+1}\|^2\right].
$$

Selecting $\beta_1 = 4 \delta \theta_+, \beta_2 = 2$, and $\beta_3 = 1$ for some positive $\delta$, (12) becomes

$$
\frac{1}{4}E\left[\|u^{n+1}_{j,i}\|^2 + \|2u^{n+1}_{j,i} - u^n_{j,i}\|^2\right] - \frac{1}{4}E\left[\|u^n_{j,i}\|^2 + \|2u^n_{j,i} - u^{n-1}_{j,i}\|^2\right]
$$

$$
+ \frac{1}{4}E\left[\|u^{n+1}_{j,i} - 2u^n_{j,i} + u^{n-1}_{j,i}\|^2\right] + \left(C_1 \theta - \frac{2 \delta + 3 \theta_+}{2}\right) \Delta t_i E\left[\|\nabla u^{n+1}_{j,i}\|^2\right]
$$

$$
+ \left(\frac{1}{3} - C_1\right) \theta - \theta_+ \Delta t_i E\left[\|\nabla u^{n+1}_{j,i}\|^2 - \|\nabla u^n_{j,i}\|^2\right]
$$

$$
+ \left(\frac{1}{3} - C_1\right) \theta - \frac{\theta_+}{2} \Delta t_i E\left[\|\nabla u^{n-1}_{j,i}\|^2\right] \leq \frac{\Delta t_i}{4 \delta \theta_+} E\left[\|f_j^{n+1}\|^2\right].
$$
Stability follows if the following conditions hold:

\begin{align}
(14) \quad C_1 \theta - \frac{2\delta + 3}{2} \theta_+ &\geq 0, \\
(15) \quad \frac{1}{3} (1 - C_1) \theta - \frac{\theta_+}{2} &\geq 0.
\end{align}

By taking \( C_1 = \frac{1}{2} \) and \( \delta = \frac{\theta - 3 \theta_+}{2 \theta_+} \), under the assumption (5), we have

\[ C_1 \theta - \frac{2\delta + 3}{2} \theta_+ = \frac{\theta}{2} - \frac{\theta_+}{2} = 0 \quad \text{and} \quad \frac{\theta}{3} - \theta_+ > 0. \]

Then, by dropping a positive term, (13) becomes

\[ \frac{1}{4} \mathbb{E}[\|u_{j,l}^{n+1}\|^2 + \|2u_{j,l}^{n+1} - u_{j,l}^{n-1}\|^2] - \frac{1}{4} \mathbb{E}[\|u_{j,l}^{n}\|^2 + \|2u_{j,l}^{n} - u_{j,l,t}^{n-1}\|^2] + \frac{\theta}{2} \Delta t \mathbb{E}[\|\nabla u_{j,l}^{n+1}\|^2 - \|\nabla u_{j,l}^{n}\|^2] + \left( \frac{\theta}{3} - \theta_+ \right) \Delta t \mathbb{E}[\|\nabla u_{j,l}^{n}\|^2] \]

\begin{align}
(16) \quad \frac{\theta}{6} \Delta t &\mathbb{E}[\|\nabla u_{j,l}^{n+1}\|^2 - \|\nabla u_{j,l}^{n}\|^2] + \left( \frac{\theta}{6} - \frac{\theta_+}{2} \right) \Delta t \mathbb{E}[\|\nabla u_{j,l}^{n}\|^2] \\
&\leq \frac{\Delta t}{2(\theta - 3 \theta_+)} \mathbb{E}[\|f_{j,l}^{n+1}\|^2].
\end{align}

Summing (16) from \( n = 1 \) to \( n = N_t - 1 \) and dropping two positive terms gives

\[ \frac{1}{4} \mathbb{E}[\|u_{j,l}^N\|^2] + \frac{1}{4} \mathbb{E}[\|2u_{j,l}^N - u_{j,l}^{N-1}\|^2] + \left( \frac{\theta}{3} - \theta_+ \right) \Delta t \sum_{n=1}^{N_t} \mathbb{E}[\|\nabla u_{j,l}^n\|^2] \]

\begin{align}
(17) \quad \leq \frac{\Delta t}{2(\theta - 3 \theta_+)} \sum_{n=1}^{N_t-1} \mathbb{E}[\|f_{j,l}^{n+1}\|^2] + \frac{1}{4} \mathbb{E}[\|u_{j,l}^1\|^2] + \frac{1}{4} \mathbb{E}[\|2u_{j,l}^1 - u_{j,l,0}\|^2] + \frac{\theta}{2} \Delta t \mathbb{E}[\|\nabla u_{j,l}^1\|^2] + \theta_+ \frac{\Delta t}{6} \mathbb{E}[\|\nabla u_{j,l}^0\|^2],
\end{align}

which completes the proof. \( \square \)

Remark 2. The ensemble-based time stepping scheme (4) is stable if condition (5) is satisfied. Moreover, it becomes to be unconditionally stable when the size of ensemble equals one since \( \theta_+ \) would shrink to zero. Thus, given a group of problems, one can use condition (5) as a guideline to divide problems into subgroups so that condition (5) holds in each of them. The smallest subgroup could contain only one member for that no stability condition is required.

Next, by using the standard error estimate for the Monte Carlo method (e.g., [25]), we can bound the statistical error \( \mathcal{E}_S^n \) as follows.

Theorem 3. Let \( \mathcal{E}_S^n = \mathbb{E}[u_{j,l}^n] - \Psi_{j,l}^n \), where \( u_{j,l}^n \) is the result of scheme (4) and \( \Psi_{j,l}^n = \frac{1}{J} \sum_{j=1}^{J} u_{j,l}^n \), Suppose conditions (i) and (ii), and the stability condition (5) hold, there is a generic positive constant \( C \) independent of \( J_t, h_t, \) and \( \Delta t \) such that

\[ \frac{1}{4} \mathbb{E}[\|\mathcal{E}_S^N\|^2] + \frac{1}{4} \mathbb{E}[\|2\mathcal{E}_S^N - \mathcal{E}_S^{N-1}\|^2] + \left( \frac{\theta}{3} - \theta_+ \right) \Delta t \sum_{n=1}^{N_t} \mathbb{E}[\|\nabla \mathcal{E}_S^n\|^2] \]

\begin{align}
(18) \quad \leq C &\left( \frac{\Delta t}{J_t} \sum_{n=1}^{N_t} \mathbb{E}[\|f_t^n\|^2] + \Delta t \mathbb{E}[\|\nabla u_{j,l}^1\|^2] + \mathbb{E}[\|\nabla u_{j,l}^0\|^2] \\
&+ \mathbb{E}[\|u_{j,l}^1\|^2] + \mathbb{E}[\|2u_{j,l}^1 - u_{j,l,0}\|^2] \right). \nonumber
\end{align}
Proof. First, we estimate $E[||\nabla E_i^n||^2]$. 

$$E[||\nabla E_i^n||^2] = E \left( \frac{1}{J_l} \sum_{j=1}^{J_l} (\nabla E[u_{i,l,j}^n] - \nabla u_{i,l,j}^n) \right)^2$$

$$= \frac{1}{J_l^2} \sum_{j=1}^{J_l} E \left( \left( \nabla E[u_{i,l,j}^n] - \nabla u_{i,l,j}^n, \nabla E[u_{i,l,j}^n] - \nabla u_{i,l,j}^n \right) \right)$$

$$= \frac{1}{J_l^2} \sum_{j=1}^{J_l} E \left[ \left( \nabla E[u_{i,l,j}^n] - \nabla u_{i,l,j}^n, \nabla E[u_{i,l,j}^n] - \nabla u_{i,l,j}^n \right) \right].$$

The last equality is due to the fact that $u_{i,l,j}^n, \ldots, u_{i,l,j}^n$ are i.i.d., and thus the expected value of $(\nabla E[u_{i,l,j}] - \nabla u_{i,l,j}, \nabla E[u_{i,l,j}] - \nabla u_{i,l,j})$ is a zero for $i \neq j$. We now expand $E[\nabla E[u_{i,l,j}^n] - \nabla u_{i,l,j}^n, \nabla E[u_{i,l,j}^n] - \nabla u_{i,l,j}^n]$ and use the fact that $E[\nabla u_{i,l,j}^n] = \nabla E[u_{i,l,j}^n]$ and $E[u_{i,l,j}^n] = E[u_{i,l,j}^n]$ to obtain

$$E[||\nabla E_i^n||^2] = -\frac{1}{J_l} \||\nabla E[u_{i,l,j}^n]||^2 + \frac{1}{J_l} E[||\nabla u_{i,l,j}^n||^2],$$

which yields

$$E[||\nabla E_i^n||^2] \leq \frac{1}{J_l} E[||\nabla u_{i,l,j}^n||^2].$$

With the help of Theorem 1, we have

$$\left( \frac{\theta}{3} - \theta_+ \right) \Delta t \sum_{n=1}^{N_j} E[||\nabla E_i^n||^2] \leq C \left( \frac{\Delta t}{\theta - 3\theta_+} \sum_{n=1}^{N_j} E[||f_i^n||^2] \right)$$

$$+ \theta \Delta t E[||\nabla u_{i,j}^0||^2 + ||\nabla u_{i,j}^0||^2] + E[||u_{i,j}^0||^2 + ||u_{i,j}^0 - u_{i,j}^0||^2].$$

The other terms on the LHS of (18) can be treated in the same manner. This completes the proof.

Next, we estimate the finite element discretization error $\mathcal{E}_i^n$.

**Theorem 4.** Let $\mathcal{E}_i^n = E[u_j(t_n) - u_{i,j}^n]$, where $u_j(t_n)$ is the solution to equation (1) when $\omega = \omega_j$ and $t = t_n$ and $u_{i,j}^n$ is the result of scheme (4). Assume that the initial errors $||u_j(t_0) - u_{i,j}^0||$, $||u_j(t_1) - u_{i,j}^0||$, $||\nabla(u_j(t_0) - u_{i,j}^0)||$ and $||\nabla(u_j(t_1) - u_{i,j}^0)||$ are all at least $O(h^m)$. Suppose conditions (i) and (ii), and the stability condition (5) hold, there exists a generic constant $C$ independent of $J_l$, $h_l$ and $\Delta t_l$ such that

$$\frac{1}{4} E[||\mathcal{E}_i^{N_j}||^2] + \frac{1}{4} E[||2\mathcal{E}_i^{N_j} - \mathcal{E}_i^{N_j-1}||^2] + \left( \frac{\theta}{3} - \theta_+ \right) \Delta t \sum_{n=1}^{N_j} E[||\nabla \mathcal{E}_i^n||^2] \leq C(\Delta t_l^2 + h_l^2).$$

Proof. We first derive the error equation for (4). Equation (1) evaluated at $t_{n+1}$ and tested by $\forall v_i \in V_l^n$ yields

$$\left( \frac{3u_j(t_{n+1}) - 4u_j(t_n) + u_j(t_{n-1})}{2\Delta t_l}, v_i \right) + \left( a_j \nabla u_j(t_{n+1}), \nabla v_i \right)$$

$$= \left( f_j^{n+1}, v_i \right) - \left( R_j^{n+1}, v_i \right).$$
where \( f_j^{n+1} = f_j(t_{n+1}) \) and \( R_j^{n+1} = u_j,t(t_{n+1}) - \frac{3u_j(t_{n+1}) - 4u_j(t_n) + u_j(t_{n-1})}{2\Delta t} \). Denoted by \( e_j^n := u_j(t_n) - u_j^n \), the approximation error at the time \( t_n \). Subtracting (4) from (21) produces

\[
(22) \quad \left( \frac{3e_j^{n+1} - 4e_j^n + e_j^{n-1}}{2\Delta t}, v_t \right) + (\overline{a}_t \nabla e_j^{n+1}, \nabla v_t) + \left( (a_j - \overline{a}_t) \nabla e_j^{n+1} - (2e_j^n - e_j^{n-1}), \nabla v_t \right) + \left( (a_j - \overline{a}_t) \nabla (e_j^{n+1} - 2u_j^n + u_j^{n-1}), \nabla v_t \right) + (R_j^{n+1}, v_t) = 0.
\]

Let \( P_t(u_j(t_n)) \) be the Ritz projection of \( u_j(t_n) \) onto \( V_t^0 \) satisfying

\[
(\overline{a}_t(\nabla(u_j(t_n) - P_t(u_j(t_n))), \nabla v_t) = 0, \quad \forall v_t \in V_t^0.
\]

The error can be decomposed as

\[
e_j^n = \rho_j^n - \phi_j^n \quad \text{with} \quad \rho_j^n = u_j(t_n) - P_t(u_j(t_n)) \quad \text{and} \quad \phi_j^n = u_j^n - P_t(u_j(t_n)).
\]

By substituting this decomposition into (22) and choosing \( v_t = \phi_j^{n+1} \), we obtain

\[
(23) \quad \left( \frac{3\rho_j^{n+1} - 4\rho_j^n + \rho_j^{n-1}}{2\Delta t}, \phi_j^{n+1} \right) + (\overline{a}_t \nabla \phi_j^{n+1}, \nabla \phi_j^{n+1})
\]

\[
= - \left( (a_j - \overline{a}_t) \nabla (2\phi_j^n - \rho_j^{n-1}), \nabla \phi_j^{n+1} \right) + \left( \frac{3\rho_j^{n+1} - 4\rho_j^n + \rho_j^{n-1}}{2\Delta t}, \phi_j^{n+1} \right)
\]

\[
+ (\overline{a}_t \nabla \rho_j^{n+1}, \nabla \phi_j^{n+1}) + \left( (a_j - \overline{a}_t) \nabla (2\rho_j^n - \rho_j^{n-1}), \nabla \phi_j^{n+1} \right)
\]

\[
+ \left( (a_j - \overline{a}_t) \nabla (u_j^{n+1} - 2u_j^n + u_j^{n-1}), \nabla \phi_j^{n+1} \right) + (R_j^{n+1}, \phi_j^{n+1}).
\]

After integrating over probability space, we have, for the LHS,

\[
(24) \quad \text{LHS} \geq \frac{1}{4\Delta t} \mathbb{E} \left[ ||\phi_j^{n+1}||^2 + ||2\phi_j^{n+1} - \phi_j^n||^2 \right] - \frac{1}{4\Delta t} \mathbb{E} \left[ ||\phi_j^n||^2 + ||2\phi_j^n - \phi_j^{n-1}||^2 \right]
\]

\[
+ \frac{1}{4\Delta t} \mathbb{E} \left[ ||\phi_j^{n+1} - 2\phi_j^n + \phi_j^{n-1}||^2 \right] + \theta \mathbb{E} \left[ ||\nabla \phi_j^{n+1}||^2 \right].
\]

We then bound the terms on the RHS of (23) one by one. By applying the Cauchy-Schwarz and Young’s inequalities, we have

\[
\mathbb{E} \left[ \left( (a_j - \overline{a}_t) \nabla (2\phi_j^n - \rho_j^{n-1}), \nabla \phi_j^{n+1} \right) \right]
\]

\[
\leq \theta \mathbb{E} \left[ ||2\nabla \phi_j^n, \nabla \phi_j^{n+1}||^2 \right] + \theta \mathbb{E} \left[ ||\nabla \phi_j^{n-1}, \nabla \phi_j^{n+1}||^2 \right]
\]

\[
\leq \theta \mathbb{E} \left[ ||\nabla \phi_j^n||^2 \right] + \frac{\theta}{2} \mathbb{E} \left[ ||\nabla \phi_j^{n-1}||^2 \right] + \frac{3\theta}{2} \mathbb{E} \left[ ||\nabla \phi_j^{n+1}||^2 \right].
\]
We further use the Poincaré inequality and have

\[
E \left[ \left( \frac{3\rho_{j,t}^{n+1} - 4\rho_j^n + \rho_{j,t}^{n-1}}{2\Delta t} \right)^2 \right] 
\leq \frac{C}{4C_0} E \left[ \left( \frac{3\rho_{j,t}^{n+1} - 4\rho_j^n + \rho_{j,t}^{n-1}}{2\Delta t} \right)^2 \right] + C_0 \theta E \left[ \| \nabla \phi_{j,t}^{n+1} \|^2 \right]
\]

\[
\leq \frac{C}{4C_0} E \left[ \left( \frac{1}{\Delta t} \int_{t_{n-1}}^{t_{n+1}} \rho_{j,t} dt \right)^2 \right] + C_0 \theta E \left[ \| \nabla \phi_{j,t}^{n+1} \|^2 \right]
\]

\[
\leq \frac{C}{4C_0} \theta \Delta t E \left[ \int_{t_{n-1}}^{t_{n+1}} \| \rho_{j,t} \|^2 dt \right] + C_0 \theta E \left[ \| \nabla \phi_{j,t}^{n+1} \|^2 \right],
\]

where \( C \) is the Poincaré coefficient and \( C_0 \) is an arbitrary positive constant. The rest
of terms can be bounded as follows.

\[
E \left[ \langle \phi_{j,t}^{n+1}, \nabla \phi_{j,t}^{n+1} \rangle \right] = 0.
\]

\[
E \left[ \left( \frac{(\rho_j - \bar{\rho}) \nabla (2\rho_{j,t} - \rho_j^{n-1})}{\rho_{j,t}^{n+1}}, \nabla \phi_{j,t}^{n+1} \right) \right]
\leq \frac{\theta}{C_0} E \left[ \| \nabla \rho_j^{n-1} \|^2 \right] + \frac{\theta^2}{C_0} E \left[ \| \nabla \rho_j^{n-1} \|^2 \right] + 2C_0 \theta E \left[ \| \nabla \phi_{j,t}^{n+1} \|^2 \right].
\]

\[
E \left[ \left( ((\rho_j - \bar{\rho}) \nabla (2\rho_{j,t} - \rho_j^{n-1})), \nabla \phi_{j,t}^{n+1} \right) \right]
\leq \frac{\theta^2}{4C_0} E \left[ \| \nabla \rho_j^{n-1} \|^2 \right] + C_0 \theta E \left[ \| \nabla \phi_{j,t}^{n+1} \|^2 \right]
\]

\[
\leq \frac{C_0 \theta^2}{4C_0} E \left[ \int_{t_{n-1}}^{t_{n+1}} \| \nabla u_{j,tt} \|^2 dt \right] + C_0 \theta E \left[ \| \nabla \phi_{j,t}^{n+1} \|^2 \right],
\]

and

\[
E \left[ \left( R_{j,t}^{n+1}, \phi_{j,t}^{n+1} \right) \right] \leq C_0 \theta E \left[ \| \nabla \phi_{j,t}^{n+1} \|^2 \right] + \frac{C_0 \theta^3}{C_0} E \left[ \int_{t_{n-1}}^{t_{n+1}} \| u_{j,tt} \|^2 dt \right].
\]

Substituting (24) to (31) into (23), we get

\[
idm \frac{1}{4\Delta t} \left( E \left[ \| \phi_{j,t}^{n+1} \|^2 \right] + E \left[ \| 2\phi_{j,t}^{n+1} - \phi_{j,t}^{n-1} \|^2 \right] \right) - \frac{1}{4\Delta t} \left( E \left[ \| \phi_{j,t}^{n+1} \|^2 \right] + E \left[ \| 2\phi_{j,t}^{n+1} - \phi_{j,t}^{n-1} \|^2 \right] \right)
+ \frac{1}{4\Delta t} E \left[ \| \phi_{j,t}^{n+1} - 2\phi_{j,t}^{n-1} + \phi_{j,t}^{n-1} \|^2 \right] + \theta (1 - 5C_0 - \frac{3\theta}{2}) E \left[ \| \nabla \phi_{j,t}^{n+1} \|^2 \right]
\]

\[
- \theta E \left[ \| \nabla \phi_{j,t}^{n+1} \|^2 \right] - \frac{\theta}{2} E \left[ \| \nabla \phi_{j,t}^{n+1} \|^2 \right]
\]

\[
\leq \frac{C}{4C_0 \theta \Delta t} E \left[ \int_{t_{n-1}}^{t_{n+1}} \| \rho_{j,t} \|^2 dt \right] + \frac{\theta^2}{C_0 \theta} E \left[ \| \nabla \rho_{j,t} \|^2 \right] + \frac{\theta^2}{4C_0 \theta} E \left[ \| \nabla \rho_{j,t} \|^2 \right]
\]

\[
+ \frac{C_0 \theta^2}{4C_0 \theta} E \left[ \int_{t_{n-1}}^{t_{n+1}} \| u_{j,tt} \|^2 dt \right] + \frac{C_0 \theta^3}{C_0 \theta} E \left[ \int_{t_{n-1}}^{t_{n+1}} \| u_{j,tt} \|^2 dt \right].
\]
Now we split the term $\theta E[\|\nabla \phi_{j,t}^n\|^2]$, and choose $C_0 = \frac{1}{30}(1 - \frac{3\theta}{\theta_+})$:

\[\frac{1}{4\Delta t} (E[\|\phi_{j,t}^{n+1}\|^2] + E[\|2\phi_{j,t}^n - \phi_{j,t}^{n-1}\|^2]) - \frac{1}{4\Delta t} (E[\|\phi_{j,t}^{n-1}\|^2] + E[\|2\phi_{j,t}^n - \phi_{j,t}^{n-1}\|^2]) + \theta (\frac{1}{3} - \theta_+) E[\|\nabla \phi_{j,t}^{n+1}\|^2]
+ \theta (\frac{1}{3} - \theta_+) E[\|\nabla \phi_{j,t}^{n-1}\|^2] + \frac{\theta}{6} E[\|\nabla \phi_{j,t}^{n}\|^2] + \frac{\theta}{6} E[\|\nabla \phi_{j,t}^{n-1}\|^2] \leq \frac{C}{(\theta - 3\theta_+)} \left\{ \frac{1}{\Delta t} E \left[ \int_{t_{n-1}}^{t_n} \|\rho_{j,t}\|^2 dt \right] + \Delta t \theta_2^2 E[\|\nabla \phi_j^n\|^2] + \Delta t \theta_1^2 E[\|\nabla \phi_j^{n-1}\|^2] + \Delta t^2 \theta_2^2 E \left[ \int_{t_{n-1}}^{t_n} \|\nabla u_{j,t}\|^2 dt \right] \right\}.

Summing (32) from $n = 1$ to $N_t - 1$, multiplying both sides by $\Delta t$, and dropping several positive terms, we have

\[\frac{1}{4} E[\|\phi_{j,t}^1\|^2] + \frac{1}{4} E[\|2\phi_{j,t}^0 - \phi_{j,t}^{-1}\|^2] + (\theta - \theta_+) \Delta t \sum_{n=1}^{N_t} E[\|\nabla \phi_{j,t}^n\|^2] \leq \frac{C}{(\theta - 3\theta_+)} \sum_{n=1}^{N_t} \left\{ E \left[ \int_{t_{n-1}}^{t_n} \|\rho_{j,t}\|^2 dt \right] + \Delta t \theta_2^2 E[\|\nabla \phi_j^n\|^2] + \Delta t \theta_1^2 E[\|\nabla \phi_j^{n-1}\|^2] + \Delta t \theta_2^2 E \left[ \int_{t_{n-1}}^{t_n} \|\nabla u_{j,t}\|^2 dt \right] \right\}.

By the regularity assumption and standard finite element estimates of Ritz projection error (see, e.g., Lemma 13.1 in [39]), namely, for any $u_j^n \in H^{m+1}(D) \cap H_0^1(D)$,

\[\|\rho_{j,t}^n\|^2 \leq C h_t^{2m+2}\|u_j(t_n)\|^2_{L^2(D)} \quad \text{and} \quad \|\nabla \phi_{j,t}^n\|^2 \leq C h_t^{2m}\|u_j(t_n)\|^2_{L^2(D)}.

and use the assumption that $\|e_{j,t}^0\|$, $\|e_{j,t}^1\|$, $\|\nabla e_{j,t}^0\|$, and $\|\nabla e_{j,t}^1\|$ are at least $O(h^m)$, we have

\[\frac{1}{4} E[\|\phi_{j,t}^0\|^2] + \frac{1}{4} E[\|2\phi_{j,t}^{-1} - \phi_{j,t}^{-2}\|^2] + (\theta - \theta_+) \Delta t \sum_{n=1}^{N_t} E[\|\nabla \phi_{j,t}^n\|^2] \leq \frac{C}{(\theta - 3\theta_+)} \left\{ h_t^{2m+2} + \theta_2^2 h_t^{2m} + \Delta t \theta_2^2 E \left[ \int_0^T \|\nabla u_{j,t}\|^2 dt \right] \right\} + \Delta t \theta_1^2 E \left[ \int_0^T \|u_{j,t}\|^2 dt \right] + h_t^{2m} + \theta \Delta t h_t^{2m},

where $C$ is a generic constant independent of the sample size $J_t$, time step $\Delta t$ and

This manuscript is for review purposes only.
mesh size $h_l$. By the triangle inequality, we have

$$
\frac{1}{4}E[\|u_j(t_{N_l}) - u_j^{N_l}\|^2] + \frac{1}{4}E[\|2(u_j(t_{N_l}) - u_j^{N_l}) - (u_j(t_{N_l-1} - u_j^{N_l-1})\|^2]
$$

$$
+ \left(\frac{\theta}{3} - \theta_+\right)\Delta t_l \sum_{n=1}^{N_l} E[\|\nabla (u_j(t_n - u_j^{N_l})\|^2) \leq C(\Delta t_l^4 + h_l^{2m}).
$$

Applying Jensen’s inequality to terms on the LHS leads to the error estimate (20).

This completes the proof.

The combination of the error contributions from the Monte Carlo sampling and finite element approximation leads to the following estimate for the $l$-th level Monte Carlo ensemble approximation.

**Theorem 5.** Let $u(t_n)$ be the solution to equation (1) and $\Psi^{n}_{J_l} = \frac{1}{J_l} \sum_{j=1}^{J_l} u^{n}_{j,l}$.

Suppose conditions (i) and (ii) hold, and suppose the stability condition (5) is satisfied, then

$$
\frac{1}{4}E[\|E[u(t_{N_l})] - \Psi^{N_l}_{J_l}\|^2] + \frac{1}{4}E[\|2(E[u(t_{N_l})] - \Psi^{N_l}_{J_l}) - (E[u(t_{N_l-1})] - \Psi^{N_l-1}_{J_l})\|^2] 
$$

$$
+ \left(\frac{\theta}{3} - \theta_+\right)\Delta t_l \sum_{n=1}^{N_l} E[\|\nabla(E[u(t_n)]) - \Psi^{n}_{J_l})\|^2] 
$$

$$
\leq C \left(\frac{\Delta t_l}{J_l} \sum_{n=1}^{N_l} E[\|f^{n}_{J_l}\|^2] + \Delta t_l E[\|\nabla u^{j,0}_{l}\|^2 + \|\nabla u^{j,0}_{l}\|^2] 
$$

$$
+ E[\|u^{j,1}_{l}\|^2 + \|2u^{j,0}_{l} - u^{j,0}_{l}\|^2]] + C(\Delta t_l^4 + h_l^{2m}),
$$

where $C$ is a positive constant independent of $J_l, \Delta t_l$ and $h_l$.

**Proof.** Consider the first term on the LHS of (36). By the triangle and Young’s inequalities, we get

$$E[\|E[u(t_{N_l})] - \Psi^{N_l}_{J_l}\|^2] \leq 2(E[\|E[u(t_{N_l})] - E[u^{N_l}_{j,l}]\|^2] + E[\|E[u^{N_l}_{j,l}] - \Psi^{N_l}_{J_l}\|^2]).
$$

Then the conclusion follows from Theorems 3-4. The other terms on the LHS of (36) can be estimated in the same manner.

**4.2. Multilevel Monte Carlo ensemble finite element method.** Now, we derive the error estimate for the MLMCE method.

**Theorem 6.** Suppose conditions (i) and (ii) and the stability condition (5) hold, then the MLMCE approximation error satisfies

$$
\frac{1}{4}E[\|E[u(t_{N_L})] - \Psi^{N_L}_{J_L}\|^2] + \frac{1}{4}E[\|E[u^{N_L}_{j,L}] - \Psi[u^{L}_{J_L}] - (E[u^{N_L}_{j,L}])\|^2] 
$$

$$
- \Psi[u^{L}_{J_L} - (E[u^{L}_{J_L}])]\|^2] + \left(\frac{\theta}{3} - \theta_+\right)\Delta t_L \sum_{n=1}^{N_L} E[\|\nabla E[u(t_n)] - \nabla \Psi[u^{L}_{J_L}]\|^2] 
$$

$$
\leq C \left(h_L^{2m} + \Delta t_L^4 + \sum_{l=1}^{L-1} \frac{1}{J_l} (h_l^{2m} + \Delta t_l^4) \right) + \frac{C}{J_0} \left(\Delta t_0 \sum_{n=1}^{N_0} E[\|f^{n}_{J_0}\|^2] 
$$

$$
+ \Delta t_0 E[\|\nabla u^{j,0}_{l,0}\|^2 + \|\nabla u^{j,0}_{l,0}\|^2] + E[\|u_{j,l,0}\|^2 + \|2u_{j,0} - u_{j,0}\|^2] \right),
$$

This manuscript is for review purposes only.
where $C > 0$ is a constant independent of $J_l, \Delta t_l$ and $h_l$.

Proof. We only analyze the first term on the LHS because the other terms can be treated in the same manner. First, we introduce $u_{-1}(t) = 0$.

\[
E \left[ \left\| E[u(t_{N_L})] - \Psi[u_L(t_{N_L})] \right\|^2 \right]
\]

\[
= E \left[ \left\| E[u(t_{N_L})] - E[u_L(t_{N_L})] + E[u_L(t_{N_L})] - \sum_{l=0}^{L} \Psi J_l [u_L(t_{N_L}) - u_{l-1}(t_{N_L})] \right\|^2 \right]
\]

(38)

\[
\leq C \left( E \left[ \left\| E[u(t_{N_L})] - E[u_L(t_{N_L})] \right\|^2 \right] + \sum_{l=0}^{L} E \left[ \left\| (E[u(t_{N_L})] - u_{l-1}(t_{N_L})) - \Psi J_l [u_L(t_{N_L}) - u_{l-1}(t_{N_L})] \right\|^2 \right] \right).
\]

By Jensen’s inequality and Theorem 4, we get

(39)

\[
E \left[ \left\| E[u(t_{N_L})] - E[u_L(t_{N_L})] \right\|^2 \right] \leq E \left[ \left\| u(t_{N_L}) - u_L(t_{N_L}) \right\|^2 \right] \leq C(\Delta t_f^4 + h_f^{2m}).
\]

By Theorems 3-4 and the triangle inequality, we have

\[
E \left[ \left\| (E - \Psi J_l)[u_L(t_{N_L}) - u_{l-1}(t_{N_L})] \right\|^2 \right]
\]

\[
= E \left[ \left\| (E - \Psi J_l)[u(t_{N_L}) - u_{l-1}(t_{N_L})] \right\|^2 \right]
\]

(40)

\[
\leq \frac{1}{J_l} E \left[ \left\| u_L(t_{N_L}) - u_{l-1}(t_{N_L}) \right\|^2 \right] \leq \frac{2}{J_l} \left( E \left[ \left\| u(t_{N_L}) - u_L(t_{N_L}) \right\|^2 \right] + E \left[ \left\| u_L(t_{N_L}) - u_{l-1}(t_{N_L}) \right\|^2 \right] \right)
\]

\[
\leq \frac{C}{J_l} (\Delta t_f^4 + h_f^{2m} + \Delta t_{l-1}^4 + h_f^{2m}) \leq \frac{C}{J_f} (\Delta t_f^4 + h_f^{2m}).
\]

Meanwhile, based on Theorem 5, we have

\[
E \left[ \left\| E[u_0(t_{N_L})] - \Psi J_0 [u_0(t_{N_L})] \right\|^2 \right]
\]

(41)

\[
\leq \frac{C}{J_0} \left( \Delta t_0 \sum_{n=1}^{N_0} E \left[ \left\| f_j^n \right\|^2 \right] + \Delta t_0 E \left[ \left\| \nabla u_{j,0} \right\|^2 \right] + \left\| \nabla u_{j,0} \right\|^2 \right)
\]

\[
+ \left( \left\| \nabla u_{j,0} \right\|^2 + \left\| 2u_{j,0} - u_{j,0} \right\|^2 \right).
\]

Plugging (39), (40) and (41) into (38), we have

\[
\frac{1}{4} E \left[ \left\| E[u(t_{N_L})] - \Psi[u_L(t_{N_L})] \right\|^2 \right] \leq C(\Delta t_f^4 + h_f^{2m} + \sum_{l=1}^{L} \frac{1}{J_l} (\Delta t_f^4 + h_f^{2m}))
\]

(42)

\[
+ \frac{C}{J_0} \left( \Delta t_0 \sum_{n=1}^{N_0} E \left[ \left\| f_j^n \right\|^2 \right] + \Delta t_0 E \left[ \left\| \nabla u_{j,0} \right\|^2 \right] + \left\| \nabla u_{j,0} \right\|^2 \right)
\]

\[
+ \left( \left\| \nabla u_{j,0} \right\|^2 + \left\| 2u_{j,0} - u_{j,0} \right\|^2 \right).
\]

This manuscript is for review purposes only.
The other terms on the LHS of (37) can be treated in the same manner. This completes
the proof.

Since, in general, the finite element simulation cost increases as the mesh is refined,
we can balance the time step size $\Delta t_l$, mesh size $h_l$ and sampling size $J_l$ in the
preceding error estimation for achieving an optimal rate of convergence.

**Corollary 7.** By taking

$$\Delta t_l = O(\sqrt{h_l^m}) \quad \text{and} \quad J_l = O(l^{1+\varepsilon} 2^{2m(L-l)})$$

for an arbitrarily small positive constant $\varepsilon$ and $l = 0, 1, \cdots, L$, the MLMCE approxi-
mation satisfies

$$\frac{1}{4} \mathbb{E} \left[ \left\| \mathbb{E} [u(t_{N_l})] - \Psi[u_L(t_{N_l})] \right\|^2 \right] + \frac{1}{4} \mathbb{E} \left[ \left\| \mathbb{E} [u^{N_{l-1}}] - \mathbb{E} [u^{N_l} - 1] - \Psi[u_L(t_{N_l})] \right\|^2 \right]$$

where $C > 0$ are constants independent of $J_l, \Delta t_l$ and $h_l$.

Similar to the MLMC method [7, 38, 16], one can choose the sample size in
MLMCE by minimizing the total computational cost while achieving a desired error.
Take $\Delta t_l = O(\sqrt{h_l^m})$ to match the spatial and temporal errors, and suppose that, as
the mesh size decreases, the average cost of solving the PDE at level $l$ increases and
the average variance decreases in the following relations:

$$C_l = C h_l^{-\gamma_1} \quad \text{and} \quad \sigma_l = C_{\sigma} h_l^{\beta},$$

where $C, C_{\sigma}, \gamma_1$ and $\beta$ are some positive constants. One can optimize the number of
samples at the $l$-th level, $J_l$, by minimizing the total sampling cost while ensuring the
statistical error stays at the user-defined tolerance $\varepsilon$. This can be formulated as an
unconstrained optimization problem using the Lagrangian approach:

$$\min_{J_l} \sum_{l=0}^{L} J_l C_l + \lambda \left[ (L+1) \sum_{l=0}^{L} \frac{\sigma_l}{J_l} - \frac{\varepsilon^2}{4} \right].$$

Applying the Euler-Lagrange condition, we get

$$J_l = \frac{4(L+1)}{\varepsilon^2} \left( \sum_{l=0}^{L} \sqrt{\sigma_l C_l} \right) \frac{\sigma_l}{C_l}$$

and the associated total cost is

$$C = \frac{4(L+1)}{\varepsilon^2} \left( \sum_{l=0}^{L} \sqrt{\sigma_l C_l} \right)^2.$$

Note that, in this setting, the MLMCE shares the same expression of optimal sample
size and total cost as those of the MLMC. However, the use of scheme (4) in MLMCE
leads to smaller average cost for solving the PDE than the MLMC. Denote the average
cost of MLMC at level $l$ to be $C h_l^{-\gamma_2}$, we have $\gamma_1 < \gamma_2$ when either direct or block
iterative methods are used in the linear solver. Let $C_{\text{MLMCE}}$ and $C_{\text{MLMC}}$ be the total costs of MLMCE and MLMC methods, respectively, we have
\[
\frac{C_{\text{MLMCE}}}{C_{\text{MLMC}}} = \left( \frac{\sum_{l=0}^{L} \sqrt{\sigma_l h_l^{-\gamma_1}}}{\sum_{l=0}^{L} \sqrt{h_l^{-\gamma_2}}} \right)^2.
\]
Then
\[
\frac{C_{\text{MLMCE}}}{C_{\text{MLMC}}} = \begin{cases} 
\frac{h_0^{\beta-\gamma_1}}{h_0^{\beta-\gamma_2}} = h_0^{\gamma_2-\gamma_1} & \text{if } \gamma_2 < \beta, \\
\frac{h_L^{\beta-\gamma_1}}{h_L^{\beta-\gamma_2}} = 2L^{(\beta-\gamma_2)}h_0^{\gamma_2-\gamma_1} & \text{if } \gamma_1 < \beta < \gamma_2, \\
\frac{h_L^{\beta-\gamma_1}}{h_L^{\beta-\gamma_2}} = h_L^{\gamma_2-\gamma_1} & \text{if } \gamma_2 < \beta.
\end{cases}
\]
It is seen the total computational complexity of the MLMCE is lower than standard MLMC in any case. In particular, when the standard LU factorization is used in the linear solver, we can derive a more concrete computational complexity. Let $d$ be the dimension of domain. The complexity for LU factorization is $Ch^{-3d}$ and that for solving triangular systems is $Ch^{-2d}$. Then the total computational cost for sampling is $L \sum_{l=0}^{L} (J_l h_l^{-2d} + h_l^{-3d})$ since only one LU factorization is needed at each level. The corresponding optimal sample size is
\[
J_l = \frac{4(L+1)}{\epsilon^2} \left( \sum_{l=0}^{L} \sqrt{\sigma_l h_l^{-2d}} \right) \sqrt{\sigma_l h_l^{2d}}
\]
by minimizing the total cost while achieving error $\epsilon$. The associated computational complexity is
\[
C_{\text{MLMCE}} = \frac{4(L+1)}{\epsilon^2} \left( \sum_{l=0}^{L} \sqrt{\sigma_l h_l^{-2d}} \right)^2 + \sum_{l=0}^{L} h_l^{-3d}.
\]
That of the optimized MLMC complexity is
\[
C_{\text{MLMC}} = \frac{4(L+1)}{\epsilon^2} \left( \sum_{l=0}^{L} \sqrt{\sigma_l \left( h_l^{-2d} + h_l^{-3d} \right)} \right)^2.
\]

### 5. Numerical Experiments.
In this section, we apply the proposed ensemble-based multilevel Monte Carlo algorithm to two numerical tests for solving the random parabolic equation (1). The goal is two-fold: to illustrate the theoretical results in Test 1; and to show the efficiency of the proposed method in Test 2.

#### 5.1. Test 1.
We first check the convergence rate of the MLMCE method numerically by considering a problem with an \emph{a priori} known exact solution. The diffusion coefficient and the exact solution of equation (1) are selected as follows.
\[
a(\omega, x) = 8 + (1 + \omega) \sin(xy),
\]
\[
u(\omega, x, t) = (1 + \omega)[\sin(2\pi x) \sin(2\pi y) + \sin(4\pi t)],
\]
where $\omega$ obeys a uniform distribution on $[\sqrt{3}, \sqrt{3}]$, $t \in [0, 1]$, and $(x, y) \in [0, 1]^2$.
The initial condition, inhomogeneous Dirichlet boundary condition and source term

This manuscript is for review purposes only.
are chosen to match the prescribed exact solution. Therefore, the expectation of the
solution is
\[ E[u] = \sin(2\pi x) \sin(2\pi y) + \sin(4\pi t). \]

For the spatial discretization, we use quadratic finite elements on uniform triangulations, that is, \( m = 2 \). To verify the analysis given in (7), we fix \( L \) and choose
the mesh size \( h_L = \sqrt{2} \cdot 2^{-2-l} \), time step size \( \Delta t_L = 2^{-3-l} \), and number of samples \( J_l = 2^{4(L-l)+1} \) at the \( l \)-th level of the MLMCE simulation for \( l = 0, \ldots, L \). The experiment is repeated for \( R = 10 \) times. Let
\[ E_L^2 = \sqrt{\frac{1}{R} \sum_{r=1}^{R} \| \mathbb{E}[u(T)] - \Psi[u_L^{(r)}(t_{N_L})] \|^2}, \]
\[ E_{H^1} = \sqrt{\frac{1}{RM} \sum_{r=1}^{R} \sum_{m=1}^{M} \| \mathbb{E}[\nabla u(t_m)] - \Psi[\nabla u_L^{(r)}(t_m)] \|^2}, \]
where \( u \) is the exact solution and \( u_L^{(r)} \) is the MLMCE solution of the \( r \)-th replica. Hence, \( E_L^2 \) and \( E_{H^1} \) represent the numerical error in \( L^2 \) and \( H^1 \) norms, respectively.

With the above choice of discretization and sampling strategy, we expect both quantities converge quadratically with respect to \( h_L \) as indicated in Corollary 7.

Table 1: Numerical errors of the MLMCE.

<table>
<thead>
<tr>
<th>( L )</th>
<th>( E_{L^2} )</th>
<th>rate</th>
<th>( E_{H^1} )</th>
<th>rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( 6.11 \times 10^{-2} )</td>
<td>-</td>
<td>( 5.60 \times 10^{-1} )</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>( 1.43 \times 10^{-4} )</td>
<td>2.10</td>
<td>( 1.50 \times 10^{-1} )</td>
<td>1.90</td>
</tr>
<tr>
<td>3</td>
<td>( 3.60 \times 10^{-6} )</td>
<td>1.99</td>
<td>( 3.81 \times 10^{-2} )</td>
<td>1.98</td>
</tr>
</tbody>
</table>

The MLMCE numerical errors as \( L \) varies from 1 to 3 are listed in Table 1. It is observed that both \( E_{L^2} \) and \( E_{H^1} \) converge at the order of nearly 2 with respect to \( h_L \), which matches our expectation.

5.2. Test 2. Next, we use a test problem to demonstrate the effectiveness of the MLMCE method. The same test problem was considered in [26] for testing the first-order, ensemble-based Monte Carlo method and a similar computational setting was used in [30] to compare numerical approaches for parabolic equations with random coefficients.

The test problem is associated with the zero forcing term \( f \), zero initial conditions, and homogeneous Dirichlet boundary conditions on the top, bottom and right edges of the domain but inhomogeneous Dirichlet boundary condition, \( u = y(1-y) \), on the left edge. The random coefficient varies in the vertical direction and has the following form

\[ a(\omega, x) = a_0 + \sigma \sqrt{\lambda_0} Y_0(\omega) + \sum_{i=1}^{n_f} \sigma \sqrt{\lambda_i} [Y_i(\omega) \cos(i\pi y) + Y_{n_f+i}(\omega) \sin(i\pi y)] \]

with \( \lambda_0 = \sqrt{\pi L_c} \), \( \lambda_i = \sqrt{\pi L_c} e^{-i(x_n, y)^2} \) for \( i = 1, \ldots, n_f \) and \( Y_0, \ldots, Y_{2n_f} \) are uncorrelated random variables with zero mean and unit variance. In the following numerical
test, we take \( a_0 = 1, L_c = 0.25, \sigma = 0.15, n_f = 3 \) and assume the random variables \( Y_0, \ldots, Y_{2n_f} \) are independent and uniformly distributed in the interval \([-\sqrt{3}, \sqrt{3}]\). We use quadratic finite elements for spatial discretization and simulate the system over the time interval \([0, 0.5]\).

We use the MLMCE method to analyze some stochastic information of the system such as the expectation of the solution at final time. More precisely, we apply the MLMCE with the maximum level \( L = 2 \), the mesh size \( h_l = \sqrt{2} \cdot 2^{-3-l} \) and time step size \( \Delta t_l = 2^{-4-l} \). Due to the small size of the problem, we apply LU factorization in solving linear systems. Targeting a numerical error \( \epsilon = 10^{-3} \), we choose the number of samples \( J_l = 2^{4(L-l)+1} \) at the \( l \)-th level, for \( l = 0, \ldots, L \) based on (44) with \( d = 2 \) and \( \beta = 4 \). Note that if the samples does not satisfy the stability condition (5), we will divide the sample set into small subsets so that (5) holds on each smaller group. Since the diffusion coefficient function is independent of time, such a process can be efficiently implemented for ensemble calculations at each level. The MLMCE solution at the final time \( T \) is

\[
\Psi^E_{t, h}(x) = \Psi[u^E_{t, N^L}(t_{N^L})],
\]

which is shown in Figure 1 (left).

Since the exact solution is unknown, to quantify the performance of the MLMCE method, we compare the result with that of the standard MLMC finite element simulations using the same computational setting. The same set of sample values is used, thus, the only difference is that individual finite element simulations are implemented at each level in the latter. Denote the approximated expected value of the latter approach by

\[
\Psi^I_{t, h}(x) = \Psi[u^I_{t, N^L}(t_{N^L})],
\]

which is shown in Figure 1 (middle). Note that for a fair comparison, we also use the LU factorization in solving all the linear systems in individual simulations. The difference between \( \Psi^E_{t, h} \) and \( \Psi^I_{t, h} \), \( |\Psi^E_{t, h} - \Psi^I_{t, h}| \), is shown in Figure 1 (right). It is observed that the difference is on the order of \( 10^{-4} \), which indicates the MLMCE method is able to provide the same accurate approximation as individual simulations. However, the computational complexity of the MLMCE simulation is smaller than that of the individual MLMC simulations. By (45)-(46), we have the complexity estimations of both approaches as follows:

\[
C_{\text{MLMCE}} = \frac{4(L + 1)^3}{\epsilon^2} + \sum_{l=0}^{2} h_l^{-6} \approx 1.39 \times 10^9
\]
and

\[ C^{\text{MLMC}} = \frac{4(L + 1)}{\epsilon^2} \left( \sum_{l=0}^{L} h_l^{-1} \right) \approx 5.37 \times 10^9. \]

Meanwhile, the CPU time for the ensemble simulation in this numerical test is \(2.65 \times 10^3\) seconds and that of the MLMC finite element simulations is \(1.01 \times 10^4\) seconds, which matches our complexity estimations.

6. Conclusions. A multilevel Monte Carlo ensemble method is developed in this paper to solve second-order random parabolic partial differential equations. This method naturally combines the ensemble-based, multilevel Monte Carlo sampling approach with a second-order, ensemble-based time stepping scheme so that the computational efficiency for seeking stochastic solutions is improved. Numerical analysis shows the numerical approximation achieves the optimal order of convergence. As a next step, we will investigate performance of the method on large-scale, nonlinear problems, in which we will deal with nonlinearity of the system and use block iterative solvers to treat high-dimensional linear systems.

REFERENCES


This manuscript is for review purposes only.


