Goal-Oriented Adaptive Modeling of Random Heterogeneous Media and Model-Based Multilevel Monte Carlo Methods

Laura Scarabosio\textsuperscript{1}, Danial Faghihi\textsuperscript{2}, Barbara Wohlmuth\textsuperscript{1}, J. Tinsley Oden\textsuperscript{2}

\textsuperscript{1} Technical University of Munich, Germany, Department of Mathematics, Chair of Numerical Mathematics (M2)
\textsuperscript{2} Institute for Computational Engineering and Sciences (ICES), The University of Texas at Austin

Abstract

Methods for generating sequences of surrogates approximating fine-scale models of two-phase random heterogeneous media are presented that are designed to adaptively control modeling error in key quantities of interest (QoIs). For specificity, the base models considered involve stochastic partial differential equations characterizing, for example, steady-state heat conduction in random heterogeneous materials and stochastic elastostatics problems in plane elasticity. The adaptive process involves generating a sequence of surrogate models defined on a partition of the solution domain into regular subdomains and then, based on estimates of error in the QoIs, assigning homogenized effective material properties to some subdomains and full random fine-scale properties to others, to control error so as to meet a preset tolerance. New model-based Multilevel Monte Carlo (mbMLMC) methods are presented that exploit the adaptive sequencing and are designed to reduce variances and thereby accelerate convergence of Monte Carlo (MC) sampling. Estimates of cost and mean squared error of the method are presented. The results of several numerical experiments are discussed that confirm that, after a preprocessing step, substantial saving in computer costs can be realized through the use of controlled surrogate models and the associated mbMLMC algorithms.

Keywords: Adaptive control of model error, goal-oriented a posteriori error estimation, Multi-Level Monte Carlo, random heterogeneous media.

1. Introduction

The use of high-dimensional, high-fidelity computational models to simulate complex physical phenomena in heterogeneous material systems has increased in recent years. This has been due, in part, to substantial increases in computational power that occurred during this period, but also due to the increased reliance on computer predictions as a basis for critical decisions effecting the design and performance of complex engineered systems. It is now more frequently recognized that predictive models of such system must account for uncertainties in microstructural properties and that stochastic models of random media are often required for reliable predictions. Paradoxically, models which can deliver the desired resolution of the critical quantities of interest (the QoIs), while also capturing the stochastic character, are often of such size and complexity that the desired solutions are intractable. So, methods for reducing the cost of computations by reducing the model size while also retaining accuracy in the prediction of QoIs are of great value.

In this paper, we address this broad problem by developing methods to generate sequences of tractable stochastic surrogate models less complex and of smaller dimension than fine-scale models, that deliver accurate approximations of key quantities of interest (QoIs). Our adaptive modeling processes provide a framework for developing new Monte Carlo based solution paradigms under the broad category of Multilevel Monte Carlo methods. The target application is the classical problem of analyzing the behavior of ran-
dom heterogeneous media, a perplexing problem that has been under study for decades, see e.g. [37, 8, 18].

We derive estimates of modeling error in specific QoIs and use these to control adaptively the choice of our surrogates. We consider two classes of problems: first, the problem of diffusion (e.g. heat conduction) in a random multiphase heterogeneous media modeled by elliptic, partial differential equations with random coefficients, depicting, for example, the thermal conductivity of the media material, and second, plane elastostatics of two-phase materials with randomly distributed isotropic inclusions. Among common QoIs which are relevant in the first class of applications are average temperatures over inclusions or heat transfer at inclusion boundaries characterized by local heat fluxes. In the elastostatics applications, QoIs range from average strain over an inclusion, pointwise displacement, to total strain energy in particular subdomains. In either case, the sequence of surrogate models naturally provides levels of models that fit the framework suited for Multilevel Monte Carlo implementations.

The idea of goal-oriented estimation and control of modeling error was introduced in connection with multiscale modeling of heterogeneous materials in [30, 38]. The reference to goal-oriented estimates is intended to imply local estimates of quantities of interest, generally characterized by linear functionals of the solution to the forward problem, as opposed to global estimates of errors in various norms. Modeling error, as opposed to discretization error, is the relative error in the solutions or quantities of interest between low-fidelity or coarse-scale approximations and those of a high-fidelity base model, which, while generally intractable, is assumed to provide an accurate, high-resolution characterization of the physics or phenomena of interest. In this exposition, we focus on the estimation and control of modeling error and assume that discretization errors are made negligible by using sufficiently refined meshes.

A general framework for estimating modeling error is described in [27, 35] and methods for adaptive modeling, including hybrid continuum-atomistic models, molecular dynamics models, and models of nonlinear systems, are discussed in [28, 4]. The idea of estimating and controlling modeling error for linear, multiscale models of heterogeneous materials was introduced by Oden and Vemaganti [30, 29, 38] and generalizations to nonlinear problems was given by Oden and Prudhomme [27]. Further generalizations and applications of these methods to broad classes of problems were given in Oden et al [28]. Braack and Ern [7] described techniques for a posteriori estimation and control of both modeling and discretization error. More recently, Maier and Rannacher [24, 23] continued a duality-based approach to model adaptivity for heterogeneous materials that involves a post-processing procedure for selecting optimal models for estimating errors in quantities of interest. Moreover, the use of sets of low-fidelity models correlated with a high-fidelity base model as a tool for deriving Monte Carlo solvers for large-scale simulations has led to the development of Multi-Fidelity Monte Carlo Methods (MFMC) [32, 33, 20]. In these methods, the selection of surrogate models and the optimal distribution of samples among the levels are tied to the correlation between the approximate QoIs of each surrogate and the QoI of the high-fidelity model, and by the cost of evaluating the QoI for each surrogate model.

In the present paper, we describe extensions of the notions of modeling error estimation and control in which the underlying stochastic high-fidelity model is characterized by stochastic elliptic boundary value problems. A large literature exists on the physics of such random heterogeneous media; see for example, the treatises of Torquato [37] and Buryanchenko [8], and the survey of Jeulin and Ostoja-Starzewski [18]. The central aim of most theories of random media is to derive effective properties of homogenized representation of the material [19, 36, 5], which is also a factor in the adaptive algorithms developed here.

Our approach addresses and resolves several challenging problems, including 1) the derivation of two-sided bounds on goal-oriented estimates of modeling error in QoIs in systems modeled by stochastic PDEs; 2) the construction of homogenized models of the microstructure of random two-phase media; 3) development and implementation of an adaptive modeling algorithm to control the error in QoIs delivered by
models representing a mixture of homogenized and microscale media; and 4) the solution of representative model problems to demonstrate the implementation and effective of the proposed methodology. Versions of multilevel Monte Carlo methods (MLMC) are described which, as noted, are built around the construction of sequences of surrogate approximations that provide MC levels central to MLMC. We are able to demonstrate improvements in the efficiency of these methods compared to plain Monte Carlo when estimates of modeling errors are used to guide MC-type solvers.

In section 2, following this introduction, the notation and mathematical structure of a class of stochastic boundary-value problems are presented, together with formalities on relevant function spaces and on the structure of variational formulations of the stochastic model. The idea of constructing sequences of lower-dimensional surrogate models approximating the high fidelity base model is taken up in Section 3. A posterior estimates of error in quantities of interest are presented in Section 4. Applications of the methodology to a class of heat transfer problems and the adaptive modeling algorithm for controlling the error are discussed in Section 5. The theory and algorithms underlying multilevel Monte Carlo methods based on levels generated by sequences of surrogates, which we refer to as model-based MLMC (or mbMLMC), are presented in Section 6. Extensions of the surrogate construction of Section 5 to linear elastostatics are given in Section 7. Numerical experiments involving applications of the theory and methodology to model problems are given in Section 8. It is demonstrated that the use of a posterior estimates of error in quantities of interest delivered by a sequence of surrogate models can accelerate the convergence of MC methods. Concluding comments are collected in Section 9.

2. Stochastic Models Involving Random Heterogeneous Media

We consider a class of linear stochastic elliptic boundary-value problems that model various physical phenomena in random, heterogeneous media. We assume the media occupies a bounded domain \( D \) in \( \mathbb{R}^d \), \( d = 1, 2, 3 \), with Lipschitz boundary \( \partial D \). The problem is set in a complete probability space \((\Omega, \mathcal{F}, \mathbb{P})\), the sample set \( \Omega \) of possible outcomes describes realizations of microstructural distributions of the two-phase material, \( \mathcal{F} \) the \( \sigma \)-algebra of subsets of \( \Omega \), and \( \mathbb{P} \) the probability measure. The problem is to find a stochastic field \( u : \Omega \times \bar{D} \to \mathbb{R}^m \), \( m = 1, 2, 3 \) (possibly representing a temperature field or displacement field in deformable solid over \( D \)) and compute a QoI from it. Let us first consider a model class of problems formally governed by the equations

\[
\begin{aligned}
\nabla \cdot A(\omega, x) \nabla u(\omega, x) &= f(\omega, x), \quad x \in D, \\
A(\omega, x) \nabla u(\omega, x) \cdot n &= \sigma(\omega, x), \quad x \in \Gamma_N, \\
u(\omega, x) &= 0, \quad x \in \Gamma_D,
\end{aligned}
\]  

(1)

almost surely for all \( \omega \in \Omega \), where \( \nabla \) is the spatial gradient operator, \( A(\omega, x) \) is a \( d \times d \) symmetric positive-definite matrix representing the conductivity, diffusivity, mobility, etc., of the random field, \( f(\omega, x) \) and \( \sigma(\omega, x) \) are prescribed data, \( \Gamma_N \) being a subset of \( \partial D \) on which Neumann data is prescribed, and \( \Gamma_D = \partial D \setminus \Gamma_N \). Later we focus on the special case, \( A(\omega, x) = \kappa(\omega, x) I \) and then we consider the case \( A(\omega, x) = E(\omega, x) \), \( E \) being the fourth order elasticity tensor and \( u = u(\omega, x) \) the vector-valued stochastic displacement field.

If \( \mathcal{B}(D) \) denotes the Borel \( \sigma \)-algebra generated by open sets of \( D \), then \( A, f \) and \( \sigma \) as \( (L^\infty(D))^{d \times d} \), \( H^1_\Gamma(D) = \{ v \in H^1(D), \text{tr} v|_{\Gamma_D} = 0 \} \), and \( H^{-\frac{1}{2}}(\Gamma_N) \)-valued quantities, respectively, are assumed measurable with the induced \( \sigma \)-algebra, \( \mathcal{F} \otimes \mathcal{B}(D) \) for \( A \) and \( f \), and \( \mathcal{F} \otimes \mathcal{B}(\Gamma_N) \) for \( \sigma \), \( \mathcal{B}(\Gamma_N) \) being the sigma-algebra associated with open sets in \( \Gamma_N \).
For almost every \( \omega \in \Omega \), \( A(\omega, x) \) is assumed to be uniformly bounded and coercive, with respect to \( x \) and \( \omega \), representing a possibly discontinuous, highly oscillatory function characterizing the two phases of the media. Thus, there exist \( 0 < \alpha_{\text{min}}, \alpha_{\text{max}} < +\infty \) such that

\[
\mathbb{P}(\omega \in \Omega : \alpha_{\text{min}} a^T a \leq a^T A(\omega, x) a \leq \alpha_{\text{max}} a^T a, \forall a \in \mathbb{R}^d, \text{ a.e. } x \in \bar{D}) = 1. \tag{2}
\]

2.1. Function Spaces

For an \( \mathbb{R}^N \)-valued random function \( Y \in L^2_w(\Omega) \), its expectation is denoted by

\[
\mathbb{E}[Y] := \int_{\Omega} Y(\omega) \, d\mathbb{P}(\omega) = \int_{\mathbb{R}^N} y \, d\mu_Y(y),
\]

where \( \mu_Y \) is the distribution measure for \( Y \). We assume \( \mu_Y \) is absolutely continuous with respect to the Lebesgue measure, so there exists a probability density function \( \pi : \mathbb{R} \to [0, \infty) \) such that

\[
\mathbb{E}[Y] = \int_{\mathbb{R}^N} y \, \pi(y) \, dy.
\]

If \( Y_i \in L^2_w(\Omega), i = 1, \ldots, p \), the covariance matrix of \( Y \) is defined by

\[
\text{Cov}[Y](i, j) := \mathbb{E}[(Y_i - \mathbb{E}[Y_i])(Y_j - \mathbb{E}[Y_j])], \quad i, j = 1, 2, \ldots, p.
\]

We introduce the spaces \( L^q_p(\Omega, W^{s,q}(D)) = \tilde{W}^{s,q}(D) \), defined as the spaces containing stochastic functions \( v \) on \( \Omega \times D \) that are measurable with respect to the \( \sigma \)-algebra \( \mathcal{F} \otimes \mathcal{B}(D) \), and equipped with the norms

\[
|v|_{L^q_p(\Omega, W^{s,q}(D))} := \left\{ \mathbb{E}\left[ \sum_{|\alpha| \leq s} \int_D |D^\alpha v|^q \, dx \right] \right\}^{\frac{1}{q}},
\]

\( D^\alpha v \) denoting generalized stochastic derivatives of order \( |\alpha| = \sum_{i=1}^d |\alpha_i| \). For \( q = 2 \), \( W^{s,q}(D) \) is the Hilbert space \( H^{s}(D) \), and \( \tilde{W}^{s,2}(D) \simeq L^2_w(\Omega) \otimes H^{s}(D) \). cf. [2].

2.2. Variational Form

Returning to (1), we take for simplicity in the development \( m = 1 \), and denote by \( u \) the solution \( u = u(\omega, x), \omega \in \Omega, x \in D \). The weak or variational formulation of the stochastic diffusion problem (1) is defined as follows: let \( H = L^2_p(\Omega) \otimes H^1_p(D) \) denote the tensor product Hilbert space endowed with the inner product

\[
(u, v)_H := \mathbb{E}\left[ \int_D (\nabla u \cdot \nabla v) \, dx \right] = \int_{\Omega} \int_{D} \nabla u(\omega, x) \cdot \nabla v(\omega, x) \, dx \, d\mathbb{P}(\omega).
\]

We define the bilinear form and linear form,

\[
B(u, v) = \mathbb{E}\left[ \int_D A \nabla u \cdot \nabla v \, dx \right], \quad B : H \times H \to \mathbb{R}, \tag{3}
\]

\[
F(v) = \mathbb{E}\left[ \int_D f \cdot v \, dx + \int_{\Gamma_N} \sigma \cdot v \, ds \right], \quad F : H \to \mathbb{R}. \tag{4}
\]

By virtue of (2), \( B(\cdot, \cdot) \) is continuous and coercive. We assume that \( f \in L^2(\Omega, L^2(D)) \) and \( \sigma \in L^2(\Omega, H^{-\frac{1}{2}}(\Gamma_N)) \), so that \( F(\cdot) \) is continuous.
The stochastic variational formulation of (1) is then:

\[
\begin{align*}
\text{Find } u \in H \text{ such that } \\
B(u, v) &= F(v), \quad \text{for all } v \in H.
\end{align*}
\]

(5)

Under the assumptions laid down so far, a unique solution to (5) exists by the Lax-Milgram theorem.

Upon solving (5) for the stochastic field \( u \), we wish to evaluate specific quantities of interest (QoIs), the goals of the modeling and simulation, represented by continuous linear functionals

\[ Q(u) = \mathbb{E}[q(u)] \in \mathbb{R}, \quad q : H^1(D) \to \mathbb{R}. \]

Many examples of meaningful QoIs could be cited and typical cases are given later. It should be emphasized that the calculation of QoIs is the principal goal of constructing the mathematical model (1) and of solving it numerically. The actual solution \( u(\omega, x) \) is of interest only as a step in computing the target QoIs. This fact is central in our approach to the analysis of random heterogeneous media. While we address the forward problem, goal-oriented approaches for Bayesian inverse problems can be found, for instance, in [25] for parameter and mesh adaptivity and in [34] for model adaptivity in turbulence simulations.

Corresponding to each QoI, \( Q \in \mathcal{L}(H, \mathbb{R}) \), is a unique function \( w = w(\omega, x) \in H \) that represents the stochastic field generated by the input \( Q \), and which is defined as the solution of the following adjoint or dual problem,

\[
\begin{align*}
\text{Given } Q : H \to \mathbb{R}, \text{ find } w \in H \text{ such that } \\
B(v, w) &= Q(v), \quad \text{for all } v \in H.
\end{align*}
\]

(6)

The solution \( w \) to the adjoint problem can be interpreted as the generalized Green’s function corresponding to the functional \( Q \).

3. Surrogate Problem Classes

Unfortunately, problem (5) (or (1)) and problem (6) are generally hardly tractable in real-life applications, owing to their enormous size and complexity. Realistic computational models may involve millions of degrees of freedom – and each microstructural configuration is but one of possibly hundreds or thousands of realizations. The complexity and problem size are increased further if, for example, Karhunen-Loève expansions are used to characterize the random coefficients. An alternative approach must be explored in which (5) (or (1)) and (6) are replaced by a tractable sequence of approximations that can be solved much cheaply and which produce sufficiently accurate approximations of the QoIs.

Therefore, we introduce the surrogate primal and adjoint problems

\[
\begin{align*}
\text{Find } u_0 \in H, \text{ such that } \\
B_0(u_0, v) &= F(v) \text{ for all } v \in H, \\
\text{Find } w_0 \in H, \text{ such that } \\
B_0(v, w_0) &= Q(v) \text{ for all } v \in H.
\end{align*}
\]

(7)

where

\[ B_0(u_0, v) = \mathbb{E} \left[ \int_D A_0 \nabla u_0 \cdot \nabla v \, dx \right]. \]
and $A_0$ is a suitable approximation of $A(x)$. In Section 5 we describe how to derive sequences of surrogate pairs $((u_l^0), (w_l^0))$, $l = 1, 2, \cdots, L$, of solutions to (7) designed to adaptively yield approximations of the QoI, $Q(u)$, of increasing accuracy. Generally, $A_0(x)$ will be of the form,

$$A_0(x) = \begin{cases} A(x), & x \in D_{\text{fine}}, \\ A_{00}(x), & x \in D \setminus D_{\text{fine}}, \end{cases}$$

where $A(x)$ is the “fine-scale” or heterogeneous coefficient matrix in (1). $D_{\text{fine}}$ is a subset of $D$ containing fine-scale random features of the microstructure, and $A_{00}(x)$ characterizes homogenized material features.

4. Goal-Oriented A Posteriori Estimates of Modeling Error

In this section, we present an extension of the analysis in [30, 38] to the stochastic systems (1) (or (5) and (6)). We note that the positive-definite bilinear form $B(\cdot, \cdot)$ of (3) generates an inner product and an energy norm on the space $H = L^2(\Omega) \otimes H^1_D(D)$, given by

$$\|v\|_B := \sqrt{B(v, v)},$$

which, by virtue of (2), is equivalent to the norm $\|v\|_H = \sqrt{(v, v)_H}$, with equivalence constants independent of $\omega \in \Omega$. Also, as noted earlier, we consider only modeling error under the assumptions that discretization errors are negligible.

As in [30], we first derive rigorous, two-sided bounds, see Theorem 1 below. Then, we use these results to derive approximate error indicators which are used in the numerical experiments.

The following result establishes computable two-sided error bounds on the quantity of interest. We make use of the notation $I_0 = I_0(x) := I - A^{-1}(x)A_0(x)$.

**Theorem 1.** Let $u_0$ and $w_0$ be solutions to the surrogate primal and adjoint problems (7), respectively. Then the modeling error in the quantity of interest is bounded above and below as follows:

$$\eta_{\text{low}} \leq Q(e_0) \leq \eta_{\text{app}},$$

where

$$\eta_{\text{low}} := \frac{1}{4}(\eta_{\text{low}})^2 - \frac{1}{4}(\eta_{\text{app}})^2 + R_{w_0}(w_0),$$

$$\eta_{\text{app}} := \frac{1}{4}(\eta_{\text{app}})^2 - \frac{1}{4}(\eta_{\text{low}})^2 + R_{w_0}(w_0),$$

with

$$\eta_{\text{low}}^2 := \frac{|R_{s w_0 x^w}(u_0 + \theta^w w_0)|}{\|u_0 + \theta^w w_0\|_B},$$

$s \in \mathbb{R}$ being real numbers, and

$$\eta_{\text{app}}^2 := s^2 \xi_{\text{app}}^2 + s^{-2} \zeta_{\text{app}}^2 \pm \mathbb{E} \left[ 2 \int_D I_0 \nabla u_0 \cdot A I_0 \nabla w_0 \, dx \right],$$

$$\theta^w = \frac{B(u_0, w_0)R_{w_0}(su_0 \pm s^{-1} w_0) - B(u_0, u_0)R_{w_0}(su_0 \pm s^{-1} w_0)}{B(u_0, w_0)R_{w_0}(su_0 \pm s^{-1} w_0) - B(w_0, w_0)R_{w_0}(su_0 \pm s^{-1} w_0)},$$

$$\xi_{\text{app}} = \|u_0\|_{I_0 B}, \quad \xi_{\text{app}} = \|w_0\|_{I_0 B}.$$
∥ \cdot ∥_{I_0} being the $I_0$-weighted norm,

$$
||u_0||_{I_0} = \mathbb{E} \left[ \int_D A_{I_0} \nabla u_0 \cdot \nabla u_0 \, dx \right],
$$

and $R_{m_0}(\cdot)$ is the linear functional, defined, for any $g \in H$, by

$$
R_g(v) = \mathbb{E} \left[ \int_D A_{I_0} \nabla g \cdot \nabla v \, dx \right].
$$

The proof follows from straightforward generalization of the proof in [29].

It has been noted in the numerical experiments of [30] that $\eta_{upp}$ and $\eta_{low}$ can provide poor estimates of the error. Simpler computable but effective approximate error estimators can be derived from the estimators in (8) and (9) and used to guide the adaptive processes described in the next sections. As the surrogates are refined with the addition of fine-scale information, the accuracy of these estimators improves.

Considering the estimate of the modeling error in the QoI based only on the upper bounds, we obtain the approximate estimator

$$
\eta_{est} := \frac{1}{4}(\eta_{upp})^2 - \frac{1}{4}(\eta_{upp})^2 + R_{m_0}(w_0) = \mathbb{E} \left[ - \int_D A^{-1} A_{I_0} \nabla u_0 \cdot \nabla w_0 \, dx \right].
$$

(10)

A similar second error estimator is obtained as $
\eta_{est,low} := \frac{1}{4}(\eta_{low})^2 - \frac{1}{4}(\eta_{low})^2 + R_{m_0}(w_0)$.

5. Heat Transfer in Two-Phase Materials with Random Microstructure

In this section, we focus on the case of a two-phase isotropic random media, in which $A$ is of the form

$$
A(\omega, x) = \kappa(\omega, x) I,
$$

$I$ being the identity, and $\kappa$ the random field of, for example, the thermal conductivity of the medium. Then the solution $u(\omega, x)$ to (5) is the corresponding temperature field.

The formal stochastic forward problem is to find a temperature field $u = u(\omega, x)$ such that, for $\mathbb{P}$-a.e. $\omega \in \Omega$,

$$
\begin{aligned}
-\nabla \cdot \kappa(\omega, x) \nabla u(\omega, x) &= f(\omega, x) \quad \text{in } D, \\
u(\omega, x) &= 0 \quad \text{on } \Gamma_D, \\
\kappa(\omega, x) \nabla u(\omega, x) \cdot n &= \sigma(\omega, x) \quad \text{on } \Gamma_N,
\end{aligned}
$$

(11)

for given $f(\omega, x)$, $\sigma(\omega, x)$, with $\kappa(\omega, x)$ a random field describing material conductivities varying between two phases, one with conductivity $\kappa_M$ for the matrix material and the other $\kappa_I$ for the inclusions. Formally,

$$
\kappa(\omega, x) = \kappa_M X_M(\omega, x) + \kappa_I X_I(\omega, x), \quad \omega \in \Omega, \quad x \in \bar{D},
$$

(12)

where $X_M$ and $X_I$ are characteristic functions for the two phases. If $\mathcal{V}_D$ is the total volume of the domain $D$, and if $\mathcal{V}_M$ and $\mathcal{V}_I$ are respectively the volumes of the phases, then $\phi_M = \mathcal{V}_M/\mathcal{V}$ and $\phi_I = \mathcal{V}_I/\mathcal{V}$ are the volume fractions of the material phases and, obviously, $\phi_M + \phi_I = 1$. The extension of our development to multiple phases $n > 2$ is straightforward. In this work, we allow for sample-dependent volume fractions, that is $\phi_M = \phi_M(\omega)$ and $\phi_I = \phi_I(\omega), \omega \in \Omega$. 7
For (11), the bilinear and linear forms of (3) and (4) reduce to,

\[
B(u, v) = \mathbb{E} \left[ \int_D \kappa(\cdot, x) \nabla u(\cdot, x) \cdot \nabla v(\cdot, x) \, dx \right],
\]

\[
F(v) = \mathbb{E} \left[ \int_D f(\cdot, x) \, dx + \int_{\Gamma_N} \sigma(\cdot, x) \cdot n \, dx \right].
\]

5.1. Effective Properties: Hashin-Shtrikman Bounds

The conventional approach used to analyze heterogeneous media is to replace problem (11) with a deterministic homogenized problem involving effective material parameters; in this case, an effective conductivity \( \kappa_{\text{eff}} \):

\[
-\kappa_{\text{eff}} \Delta \bar{u}(x) = \bar{f}(x) \quad \text{in } D, \\
\bar{u}(x) = 0 \quad \text{on } \Gamma_D, \\
\kappa_{\text{eff}} \nabla \bar{u}(x) \cdot n = \bar{\sigma} \quad \text{on } \Gamma_N,
\]

where, for example, \( \bar{f} = \mathbb{E}[F] \) and \( \bar{\sigma} = \mathbb{E}[\sigma] \).

Among the most widely used approaches for obtaining bounds on effective properties of random heterogeneous multiphase materials are those provided by variational methods of Hashin and Shtrikman [40, 39]. The original version of these principles involved determining the tightest possible bounds on effective material parameters (moduli or conductivities) in materials in which the volume fractions of all phases remain constant as the volume \( V \) of the material is allowed to get infinitely large and in which the inclusions are isotropic throughout the material domain. Many generalizations on this setting have been proposed, and we mention the work of Willis [40], which takes into account inclusion shape and distribution, Ponte Castaneda and Willis [10], which accounts for spatial correlations of inclusions of ellipsoidal form, Calvo-Jurado and Parnell [9], who considered transversely isotropic two-phase materials, and Liu [21] who explored the attainability of Hashin-Shtrikman bounds.

The Hashin-Shtrikman (HS) bounds on the effective conductivity of a two-phase isotropic material are

\[
\kappa_{\text{HSL}} \leq \kappa_{\text{eff}} \leq \kappa_{\text{HSU}},
\]

where

\[
\kappa_{\text{HSL}} = \kappa_{\text{arith}} - \frac{(\kappa_M - \kappa_I)^2 \phi_M \phi_I}{\kappa_M \phi_M + \kappa_I \phi_I + \kappa_{\text{min}}},
\]

\[
\kappa_{\text{HSU}} = \kappa_{\text{arith}} - \frac{(\kappa_M - \kappa_I)^2 \phi_M \phi_I}{\kappa_M \phi_M + \kappa_I \phi_I + \kappa_{\text{max}}},
\]

with \( \kappa_{\text{arith}} = \phi_M \kappa_M + \phi_I \kappa_I \), \( \kappa_{\text{min}} = \min(\kappa_M, \kappa_I) \), and \( \kappa_{\text{max}} = \max(\kappa_M, \kappa_I) \). We obviously find \( \kappa_{\text{harm}} \leq \kappa_{\text{HSL}} \leq \kappa_{\text{HSU}} \leq \kappa_{\text{arith}} \), where \( \kappa_{\text{harm}} = \frac{\kappa_{\text{HSL}} + \kappa_{\text{HSU}}}{2} \) is the harmonic mean.

The Hashin-Shtrikman bounds are the best possible bounds given volume fractions only. They are attained by coated-sphere models: when \( \kappa_M \leq \kappa_I \), then the lower bound is attained; viceversa, when \( \kappa_M \geq \kappa_I \), the upper bound is reached [37, Ch.16, p.406].

Since we consider sample-dependent volume fractions, we use the Hashin-Shtrikman bounds to obtain \( \kappa_{\text{eff}} \) for each sample, and therefore \( \kappa_{\text{eff}} = \kappa_{\text{eff}}(\omega), \omega \in \Omega \). In our case, such a coefficient has to be understood as a "surrogate" for the actual homogenized coefficient.
5.2. Adaptive Modeling and Construction of Surrogate Models

The QoI is always assumed to be a local feature of the physical system fully characterized by the solution of the fine-scale forward problem. Examples include the average temperature over a subdomain \( A_q \subset D \),

\[
Q(u(\omega, x)) = \mathbb{E} \left[ \frac{1}{|A_q|} \int_{A_q} u(\omega, x) dx \right],
\]

\( |A_q| \) being the area of \( A_q \), or the heat flux through the boundary of a subdomain.

Following the ideas in [30], we partition the domain \( D \) into blocks \( \{B_1, \ldots, B_{K_b}\} \), \( K_b \in \mathbb{N} \), so that \( D = \bigcup_{k=1}^{K_b} B_k \); see for instance Figure 1a in Section 8. As an approximation of the conductivity \( \kappa \), we consider then

\[
\kappa_0(\omega, x) = \begin{cases} 
\kappa(\omega, x) & \text{if } x \in B_k \text{ such that } B_k \subseteq D_{\text{fine}}, \\
\kappa_{\text{eff}}(\omega, k) & \text{if } x \in B_k \text{ such that } B_k \subseteq D \setminus D_{\text{fine}},
\end{cases}
\]

where \( D_{\text{fine}} \) consists of a collection of blocks around the QoI and \( \kappa_{\text{eff}}(\omega, \cdot), \omega \in \Omega \), is a blockwise-homogenized coefficient. More precisely, let \( \phi_{f,k} = \phi_{f,k}(\omega) \) and \( \phi_{M,k} = \phi_{M,k}(\omega), \omega \in \Omega \), denote respectively the volume fractions of the inclusions and the matrix in the block \( B_k, k = 1, \ldots, K_b \). For each realization and in each block \( B_k \subseteq D \setminus D_{\text{fine}} \), the coefficient \( \kappa_{\text{eff}}(\omega, k) \) is then computed using the volume fractions \( \phi_{f,k} \) and \( \phi_{M,k} \) in formulas (14). We use blockwise-homogenized coefficients instead of globally homogenized coefficients because, for the later application of the MLMC strategy, they provide better variance reduction at the same computational cost.

For heat conduction, the homogenized problem (7) involves the form

\[
B_0(u_0, v) = \mathbb{E} \left[ \int_D \kappa_0 \nabla u_0 \cdot \nabla v \, dx \right]
\]

and the error estimator (10) is

\[
\eta_{\text{est}} = \mathbb{E} \left[ - \int_{D \setminus D_{\text{fine}}} \kappa_0 \left( 1 - \frac{\kappa_0}{\kappa} \right) \nabla u_0 \cdot \nabla v_0 \, dx \right].
\]

It remains to define the subdomain \( D_{\text{fine}} \) where the microstructure is resolved. This can be constructed adaptively using local error indicators. Namely, for each block \( B_k, k = 1, \ldots, K_b \) the quantity

\[
\eta_{\text{est},k} := \mathbb{E} \left[ - \int_{B_k} \kappa_0 \left( 1 - \frac{\kappa_0}{\kappa} \right) \nabla u_0 \cdot \nabla v_0 \, dx \right]
\]

is an indicator for the contribution of the block \( B_k \) to the total modeling error \( \eta_{\text{est}} \). Clearly, \( \eta_{\text{est},k} \) is not zero only if \( \kappa_0 \neq \kappa \) in the \( k \)-th block, and

\[
\sum_{k=1}^{K_b} \eta_{\text{est},k} = \eta_{\text{est}}.
\]

In the deterministic setting, that is for a fixed \( \omega \in \Omega \), a surrogate model such that \( |q(u) - q(u_0)| \leq \text{tol} \) can be constructed with a strategy similar to the one in [30]: starting from \( D_{\text{fine}} = \emptyset \) and the corresponding surrogate model (7), blocks that give highest contribution to the modeling error are identified by using (16) and added to \( D_{\text{fine}} \), with corresponding update of the surrogate model; the latter step is iterated until the condition \( |q(u) - q(u_0)| \leq \text{tol} \) is satisfied. In the stochastic case, due to the presence of the average operator in the definitions (15) and (16), this adaptive procedure can still be applied by using error estimates which are
averaged over an ensemble of realizations \[13\]. Such approach is the core of our model selection strategy described in Subsection 6.2.1, for which we refer for the full description of the adaptive algorithm in the stochastic case.

For later application of MLMC, we not only need the model fulfilling the required tolerance, but also a sequence of coarser models. Therefore, to define the coarsest possible model, before the blockwise-homogenized one with \(D_{\text{fine}} = \emptyset\), we consider the model with \(\kappa_0(\omega) \equiv \kappa_{\text{eff}}(\omega), \omega \in \Omega\), with \(\kappa_{\text{eff}}(\omega)\) constant over \(D\) and coinciding with the lower (resp. upper) Hashin-Shtrikman bound computed with the volume fractions \(\phi_I, \phi_M\) over the whole domain.

**Remark 1 (Cost of globally homogenized model).** For equation (11), the cost of computing one realization of the QoI using the globally homogenized coefficient is \(O(1)\). Indeed, since the conductivity in this case is constant over the whole domain, it is sufficient to solve the forward problem once for a fixed value \(\kappa_{\text{fix}}\) of the conductivity, obtaining the value \(q_{\text{fix}}\) for the QoI. Due to linearity, each realization \(q\) for the surrogate model with coefficient \(\kappa_{\text{eff}}(\omega)\) can then be obtained by a scaling of \(q_{\text{fix}}\). In particular, for (13), \(q(\omega) = q_{\text{fix}} \frac{\kappa_{\text{fix}}}{\kappa_{\text{eff}}(\omega)}, \omega \in \Omega\).

6. Multilevel Monte Carlo on a Sequence of Surrogate Models

In this section, we describe a MLMC strategy for the efficient approximation of \(Q = \mathbb{E}[q(u)], \mathbb{E} : H \rightarrow \mathbb{R}\) being the localized QoI and \(u\) the solution to (5). We first present, in Subsection 6.1, a brief review of MLMC. Then, in Subsection 6.2, we illustrate how the key features of MLMC can be combined with the model adaptive strategy described in the previous sections to improve the efficiency of Monte Carlo sampling for local QoIs.

6.1. The Multilevel Monte Carlo method

The MLMC method was first introduced by Heinrich [16] in the framework of parametric integration, and was further developed by Giles in [14] for functionals of the solution to a stochastic differential equation (SDE). It is basically a method for variance reduction: it considers approximations of a QoI with different accuracies and leverages on the fact that differences of the QoI between consecutive levels have much smaller variance than the QoI itself. More precisely, let \(q\) be the functional of interest, which is, in general, not accessible from computations, and let \((q_l)_{l=1}^{L}\) be a sequence of approximations to \(q\), ordered from the least to the most accurate. Furthermore, we denote \(q_0 := 0\). The MLMC method exploits the identity

\[
\mathbb{E}[q_L] = \sum_{l=1}^{L} \mathbb{E}[Y_l], \quad \text{with } Y_l := q_l - q_{l-1}, l = 1, \ldots, L,
\]

to estimate \(Q = \mathbb{E}[q]\) by

\[
E^L[q_L] := \sum_{l=1}^{L} E_M[Y_l], \quad (17)
\]

where

\[
E_M[Y_l] := \frac{1}{M_l} \sum_{i=1}^{M_l} Y_l^i, \quad \text{for } l = 1, \ldots, L. \quad (18)
\]

In the equation above, for each \(l = 1, \ldots, L, \{Y_l^i\}_{i=1}^{M_l}\) denote \(M_l \in \mathbb{N}\) i.i.d. (independent, identically distributed) samples of \(Y_l\).
For every realization of (17), the error can be decomposed as
\[ |\mathbb{E}[q] - E^L[q_L]| \leq |\mathbb{E}[q - q_L]| + |\mathbb{E}[q_L] - E^L[q_L]|. \]

The first term on the right-hand side is deterministic, and it is called bias error. The second summand is a random variable, and corresponds to the statistical error. To compute the MLMC estimator up to a tolerance \( TOL \), we can decompose such a tolerance into a bias tolerance \( TOL_{bias} \) and a statistical error tolerance \( TOL_{stat} \) [12, 17]:

\[ TOL = (1 - \vartheta)TOL_{bias} + \vartheta TOL_{stat}, \]

for some \( \vartheta \in (0, 1] \). The tolerance on the bias is imposed by selecting \( TOL_{bias} \) and requiring that
\[ |\mathbb{E}[q - q_L]| \leq TOL_{bias}. \]

Due to the stochastic nature of the statistical error, there are multiple possibilities of prescribing a tolerance on it. In most of the literature (see, for instance, [14, 15, 11]), such a tolerance is imposed in the mean squared sense, requiring that
\[ \mathbb{E}\left[ |\mathbb{E}[q_L] - E^L[q_L]|^2 \right] \leq TOL_{stat}^2. \]

This, together with (20), ensures that
\[ ||\mathbb{E}[q] - E^L[q_L]]||_{L^2(\Omega, \mathbb{R})} \leq TOL. \]

The requirement in (21) is adopted in this paper. Alternatively, one could impose the bound on the statistical tolerance in a probabilistic sense, using confidence intervals; cf. [12, 17].

The bias error is controlled by choosing, for the level \( L \), a sufficiently good approximation of the QoI \( q \) [15]. As described in the next subsection, this means choosing a sufficiently good surrogate model. The statistical error is controlled by
\[ \mathbb{E}\left[ |\mathbb{E}[q_L] - E^L[q_L]|^2 \right] \leq \sum_{l=1}^{L} \frac{\mathbb{V}[q_l - q_{l-1}]}{M_l}, \]

(where \( \mathbb{V}[\cdot] \) denotes the variance) [15]. The estimated total work of the MLMC estimator is then [15]
\[ \mathbb{E}[W] = TOL_{stat}^{-2} \sum_{l=1}^{L} \sqrt{V_l W_l}, \]

where \( V_l = \mathbb{V}[Y_l] \), and \( W_l \) is the estimated cost of one realization of \( Y_l \), for \( l = 1, \ldots, L \). In this respect, the MLMC method can be seen as a variance reduction technique [15, 6]: the sequence \( (q_l)_{l=1}^{L} \) must be such that, as \( l \) increases and the cost of generating a sample increases, the variance of \( Y_l \) decreases. This makes it possible to evaluate many samples on coarser levels and fewer on the finer ones, obtaining an estimator that provides the same accuracy as the Monte Carlo estimator but with lower computational effort. However, the complexity theorem for MLMC [15, Thm. 2.1] clearly establishes that the gain in efficiency of MLMC with respect to plain Monte Carlo is based on a delicate, far from arbitrary, tradeoff between computational cost and accuracy across the levels.
In the classical MLMC method, the hierarchy of levels is constructed from iterative refinement of a discretization parameter, usually the mesh size or the time step. The delicate balance between cost and accuracy is therefore often guaranteed by convergence theorems for the discretized solution. This is not the case when using a sequence of surrogate models, where the unavailability of convergence rates (at least a priori) constitutes the main challenge in the construction of an efficient multilevel estimator. Defining a geometric sequence of tolerances as in [17] is also not applicable, as having a discrete set of models does not ensure that for each level of tolerance a model exists and, also if it does, the relationship between computational costs and tolerances among the levels may be highly not optimal.

One of the first algorithms addressing such issues is the Multifidelity Monte Carlo method, MFMC, see e.g. [32, 33]. There, the model selection from a given set of models and the number of samples per level are computed according to the correlation between each model and the most accurate model (the high-fidelity model) and to the difference between these correlations. The procedure proposed in [32] is independent of the availability of a priori or a posteriori error estimators. Instead, in light of the theory developed in the previous sections, our goal is to design a MLMC algorithm that exploits the information provided by an a posteriori error estimator. Our goal is to show that an (a posteriori) error estimator allows one to fit a model hierarchy into the framework of standard MLMC. We use the error estimator to order the models with respect to their accuracy, which enables us to use the standard MLMC estimator as defined in (17)-(18) instead of constructing a new estimator as in [32]. We also remark that, in the case of a hierarchy based on mesh refinement, the MFMC estimator of [32] has been shown to provide the same performance as the MLMC estimator, also when the correlation among the levels is significantly lower than 1, while distributing differently the samples across the levels [31]. A combination of low and high-fidelity models has also been explored in the stochastic collocation framework in [26, 41].

6.2. Model-based Multilevel Monte Carlo

Our aim is to use the same estimator as in (17) with \( q_l := q(u^l_0) \), where \( u^l_0 \) is the solution to a surrogate model as from Section 3 and thus

\[
Y_l = q(u^l_0) - q(u^{l-1}_0),
\]

for \( l = 1, \ldots, L \) and \( L \in \mathbb{N} \).

The strategy that we use can be subdivided into three steps:

1. model selection,
2. level selection,
3. computation of MLMC estimator.

The model selection procedure leverages the a posteriori error estimator to construct a sequence of surrogate models. Out of this sequence, the level selection step selects \( L \) models as levels for the MLMC estimator. Finally, the MLMC estimate of \( \mathbb{E}[q] \) as in (17)-(18) is computed.

6.2.1. Model selection

For a given bias tolerance \( TOL_{\text{bias}} \), the aim of the model selection procedure is twofold: to select the cheapest possible model \( M_J \) such that

\[
\mathbb{E} \left[ \left| q(u^J_0) - q(u) \right| \right] \leq TOL_{\text{bias}},
\]

and to construct a sequence of surrogate models with tolerance larger than the bias tolerance as candidates for the MLMC levels.
We denote by $S$ such a sequence, including $M_f$ as the last model. Let $M_1$ be the model in which the globally homogenized coefficient is used and $M_0$ the model with blockwise-homogenized coefficient. We start with $S = \{M_1, M_0\}$ if the scaling argument of Remark 1 applies (e.g. for the heat equation and linear QoI), with $S = \{M_0\}$ otherwise. Starting then from $M_0$, we proceed similarly to [13] and [17], and let the selection be guided by the error estimator. Let $\eta$ be an error estimator as the ones described in Section 4, and let $\eta_k$ be local error estimator for the $k$-th block, $k = 1, \ldots, K_b$. We use a greedy algorithm where, at each step $j \in \mathbb{N}$, we refine the subpatch $\mathcal{P}^j$ of smallest cardinality such that

$$
\sum_{k=1, \ldots, K_b \in \mathcal{P}^j} \mathbb{E}[|\eta^j_k|^2] \geq \gamma \sum_{k=1, \ldots, K_b} \mathbb{E}[|\eta^j_k|^2].
$$

(24)

for some a priori chosen $\gamma \in (0, 1)$ (where $\eta^j_k$, $\eta^j_k$ are the estimators at the $j$-th step). We add blocks until the global error estimator satisfies $\mathbb{E}[|\eta^j_k|^2] \leq TOL_{bias}$ for some step $J \in \mathbb{N}$. Each time we add a subpatch, the model where all previously added blocks and the new subpatch are refined is added to the set $S$. At the end of the algorithm, we correct the statistical tolerance by

$$
TOL_{stat}^2 = TOL^2 - \mathbb{E}[|\eta^j_k|^2]^2.
$$

This procedure is presented in Algorithm 1. Since the exact averages are not available in practice, they are replaced in lines 7 and 12 of Algorithm 1 by sample averages (there, $M_j$ denotes the number of samples at step $j \in \mathbb{N}$). In order to ensure that the sample estimates are sufficiently accurate, we adopt the criterion suggested in [17], operating a control on the sample variance of the estimators (line 6 of Algorithm 1):

$$
\sqrt{\frac{1}{M_j} V_{M_j}[\eta^j_k]} \leq R |E_{M_j}[\eta^j_k]|, \quad \text{for } k = 1, \ldots, K_b,
$$

(25)

for some $0 < R < 1$. Here, $V_{M_j}$ and $E_{M_j}$ denote, respectively, the sample variance and sample average at the step $j \in \mathbb{N}$ of the algorithm, using $M_j$ i.i.d. realizations $\{\eta^{j,i}\}_{i=1}^{M_j}$ of the error estimator:

$$
E_{M_j}[\eta^j_k] := \frac{1}{M_j} \sum_{i=1}^{M_j} \eta^{j,i}_k,
$$

$$
V_{M_j}[\eta^j_k] := \frac{1}{M_j - 1} \sum_{i=1}^{M_j} (\eta^{j,i}_k - E_{M_j}[\eta^j_k])^2.
$$

Lines 14-18 of the algorithm are to avoid that, in case line 24 is executed for some iterations, the initial number of samples for the subsequent iterations is not too large than necessary.

In our experiments, the error indicator (10) is used as well as the local indicators (16). We remark that, since we use greedy algorithm with approximate error indicators (see (10)), there is no theoretical guarantee that the optimal sequence of models will be generated. However, our numerical experiments show the effectiveness of our approach.

It remains to determine how to choose the bias tolerance, that is $\vartheta$ in (19). In the cases that the error estimator is sufficiently accurate or overestimates the error, then a positive bias tolerance can be used ($\vartheta < 1$). In principle, one should start with $M_{-1}$ only and add $M_0$ if the bias tolerance is not fulfilled; but we add $M_0$ for ease of presentation of the algorithm. Moreover, it has never been observed in our simulations that $M_{-1}$ satisfied the bias tolerance.
1). Otherwise, $\eta_{est}$ has to be regarded as a qualitative rather than quantitative error indicator for guiding the adaptive process. In this case, we set $TOL_B = 0$ and always choose the fine-scale model as model $M_J$. For a positive bias tolerance, we proceed as in [14] and opt for an a priori equal splitting $TOL^2_{bias} = TOL^2_{stat} = \frac{TOL^2}{2}$ and modify the statistical tolerance at the end of the model selection procedure according to the tolerance of the last model selected, cf. line 27. For MLMC levels based on mesh refinement, it has been shown in [12] that a non-even splitting of the tolerance might be more efficient than an even splitting. However, the algorithm described in [12] relies on availability of convergence rates, and therefore cannot be straightforwardly extended to our case of model hierarchy. We also note that, when the statistical tolerance is corrected as in line 27 of Algorithm 1, then an a posteriori uneven splitting is used that tries to reduce the computational cost of the multilevel estimator for the prescribed total tolerance.

Algorithm 1 Error Estimator-driven Model Selection

Input: $TOL, TOL_{bias}, M_0, R, \gamma$.
Output: $S, TOL_{stat}$.

1: $S = \{M_{-1}, M_0\}$ if Remark 1 applies, $S = \{M_0\}$ otherwise
2: $D_{fine} = \emptyset$
3: $j = 0, E_{M_{-1}}[\eta^{(0)}] = \text{Inf}, M_{-1} = \text{Inf}$
4: while $|E_{M_j}[\eta^{(j)}]| > TOL_{bias}$ do
5:   Compute $E_{M_j}[\eta^{(j)}], V_{M_j}[\eta^{(j)}], E_{M_j}[\eta^{(j)}_k]$ for $k = 1, \ldots, K_b$
6:   if (25) is fulfilled then
7:     if $|E_{M_j}[\eta^{(j)}]| > TOL_{bias}$ then
8:       if all blocks have been added then
9:         $TOL_{bias} = 0$
10:      break
11:     end if
12:     $D_{fine} = D_{fine} \cup P^j$, with $P^j$ subpatch of smallest cardinality s.t. (24) holds $\rightarrow$ model $M_j$
13:     $S = S \cup \{M_j\}$
14:     if $M_j > M_{j-1}$ then
15:       $M_{j+1} = \frac{M_j}{2}$
16:     else
17:       $M_{j+1} = M_j$
18:     end if
19:     $j = j + 1$
20:   else
21:     break
22:   end if
23: else
24:   $M_j = 2M_{j-1}$, go to 5
25: end if
26: end while
27: $TOL_{stat} = \sqrt{TOL^2 - (E_{M_j}[\eta^{(j)}])^2}$
28: return $S, TOL_{stat}$
6.2.2. Level selection

Given a number of levels $L$, the goal of the level selection step is to extract the subsequence $S_L := \{M_{j_1}, \ldots, M_{j_k}\} \subseteq S$ that minimizes the cost of the MLMC estimator for the prescribed tolerance $TOL$. Here we assume $L$ to be chosen a priori.

To select the subsequence $S_L$, we proceed in the following way. For level $L$, we choose as model $M_{j_k}$ the last element of $S$, so that the bias tolerance is fulfilled. For the other levels, we perform an exhaustive search over all ordered $(L-1)$-tuples in $S$: for each of these $(L-1)$-tuples, we compute the estimated cost as from (23) and choose the $(L-1)$-tuple that minimizes it. The procedure for the level selection is shown in Algorithm 2. There, at line 4 $E[W]$ denotes the sample estimate of $\mathbb{E}[W]$ in (23), where the sampling involves the variances $V_l$ and the average costs $W_l$, $l = 1, \ldots, L$. Samples from the model selection can be reused for estimating the costs and variances at the first level. We note that for the level selection the computation of the error estimators is not needed.

The exhaustive search requires the comparison of $S(L-1)$ $L$-tuples, where $S$ is the cardinality of $S$. Therefore, its cost is $O(M(S-L+1))$, with $M$ the number of sampled used (as input in Algorithm 2). For fixed $L$, the cost of the exhaustive search grows as $O((S-1)^{L-1})$.

Remark 2 (Reducing cost of level selection). If the error estimator is sufficiently accurate, then for any $L \in \mathbb{N}$ it would be possible to reduce the computational cost of the level selection to $O(M(S-L+1))$. Indeed, we can write (23) as

$$TOL_j^2 \mathbb{E}[W] = \sqrt{\mathbb{V}[q_1]W_1} + \sum_{l=2}^{L} \sqrt{V_lW_l}.$$  

Then, for $l = 2, \ldots, L$, it is possible to estimate $V_l$ as

$$V_l \lesssim \mathbb{V}[W_1] + \sum_{j=1}^{L} \sqrt{\mathbb{V}[q_j]}.$$

If $\mathbb{V}[\eta^{(j)}]$ and the costs $W_j$ are estimated while computing $E[\eta^{(j)}]$ ($j = 1, \ldots, J$) during the model selection, one could perform the level selection in $O(M(S-L+1))$ operations in the following way. First, one fixes a model $M_j$, $j = 1, \ldots, J - L + 1$, and, considering all ordered $(L-2)$-tuples $M_{i_1}, \ldots, M_{i_{L-2}}$ such that $j < i_1 < \ldots < i_{L-2} < J$, selects the $(L-2)$-tuple that minimizes $\sum_{l=2}^{L} \sqrt{\mathbb{V}[W_l]}$, computed by (26) with the samples from the model selection procedure. Second, for all models $M_j$, $j = 1, \ldots, J - L + 1$, one estimates $\mathbb{V}[q_j]$ and the cost of one realization of $q_j$ using $M$ samples and, adding the cost $\sum_{l=2}^{L} \sqrt{V_lW_l}$ of the $(L-2)$-tuple associated to $M_j$, one has an approximation of the cost of the corresponding $(L-1)$-tuple. Among these $(L-1)$-tuples, the one with minimum cost has to be selected.

For the described procedure to be effective, the error estimator has to be very accurate. Since this was not the case for the error estimator used in our numerical experiments, we stuck to the exhaustive search approach.

6.2.3. Computation of the Multilevel Monte Carlo estimator

Using the sequence $S_L$ from the model selection, the standard MLMC estimator can be computed. For completeness, we report the steps in Algorithm 3 [14, 11]. At line 4 of the algorithm, the sample averages of the variances and costs (resp. $V_{l,M}, W_{l,M}$) can be recycled from the level selection procedure. Using sample averages could be avoided by using more sophisticated techniques as in [12], but, as noted earlier,
Algorithm 2 Level Selection

**Input:** $S, M$

**Output:** $S_L$

1. Set $M_{j_L}$ as the last element of $S$.
2. $W_{S_L} = \infty$.
3. for $\{M_i, \ldots, M_{l-1}\} \subseteq S \setminus \{M_{j_L}\}$ do
4. Use $M$ samples to compute $TOL^2_{stat} E_M[W]$ for $\{M_i, \ldots, M_{l-1}, M_{j_L}\}$ according to (23).
5. if $TOL^2_{stat} E_M[W] < W_{S_L}$ then
6. $W_{S_L} = TOL^2_{stat} E_M[W]$
7. $S_L = \{M_i, \ldots, M_{l-1}, M_{j_L}\}$
8. end if
9. end for
10. return $S_L$

the algorithm in [12] cannot be used directly for model hierarchies. At line 10 of Algorithm 3, if $\tilde{M}_l = 0$ then $\{Y_i\}_{i=1}^{\tilde{M}_l}$ denotes an empty sum.

Algorithm 3 Multilevel Monte Carlo

**Input:** $L, S_L, TOL_{stat}$

**Output:** $E^L[q_L]$ such that (22) is fulfilled.

1. $E^L[q_L] = 0$
2. $E_{M_l}[Y_l] = 0, l = 1, \ldots, L$
3. for $l = 1, \ldots, L$ do
4. Use $M$ samples $\{Y_i\}_{i=1}^{M}$ to estimate $V_l, W_l, \rightarrow V_{l,M}, W_{l,M}$.
5. $E_{M_l}[Y_l] = E_{M_l}[Y_l] + \sum_{i=1}^{M} Y_i$
6. end for
7. Compute $\lambda = TOL^2_{stat} \sum_{i=1}^{L} \sqrt{V_{l,M} W_{l,M}}$
8. Set number of samples per level: $M_l = \left[ \frac{\lambda \sqrt{V_{l,M} W_{l,M}}}{W_{l,M}} \right], l = 1, \ldots, L$
9. for $l = 1, \ldots, L$ do
10. Compute remaining samples $\{Y_i\}_{i=1}^{\tilde{M}_l}, \tilde{M}_l = \max\{0, M_l - M\}$
11. $E_{M_l}[Y_l] = E_{M_l}[Y_l] + \{Y_i\}_{i=1}^{\tilde{M}_l}$
12. $E_{M_l}[Y_l] = E_{M_l}[Y_l] / \max\{M_l, M_l\}$
13. $E^L[q_L] = E^L[q_L] + E_{M_l}[Y_l]$
14. end for
15. return $E^L[q_L]$

7. Extension to Elastostatics of Random Materials

While algebraically more tedious, the full theory and adaptive algorithms described earlier are readily extendable to linear elasticity problems in the plane. Dealing then with the vector-valued displacement field $u = u(\omega, x), u : \Omega \times D \rightarrow \mathbb{R}^2, (u \in H := (L^2(\Omega, \mathbb{P}) \times H^1_0(D))^2)$, the model forward problem for a two-phase
random elastic material with isotropic phases is of the form, for a.e. $\omega \in \Omega$,

$$\nabla : T(\omega, \mathbf{u}) = f(\omega, \mathbf{x}), \quad \mathbf{x} \in D, \quad T(\omega, \mathbf{u}) \mathbf{n} = \sigma(\omega, \mathbf{x}), \quad \mathbf{x} \in \Gamma_N, \quad \mathbf{u}(\omega, \mathbf{x}) = 0, \quad \mathbf{x} \in \Gamma_D,$$

(27)

with $T$ the Cauchy stress, given by

$$T(\omega, \mathbf{u}) = 2\mu(\omega, \mathbf{x}) \varepsilon(\mathbf{u}) + \lambda(\omega, \mathbf{x}) \text{tr}(\varepsilon(\mathbf{u})) \mathbf{I},$$

$$\varepsilon(\mathbf{u}) = \frac{1}{2} (\nabla \mathbf{u} + (\nabla \mathbf{u})^\top).$$

Here $\lambda$ and $\mu$ are the Lamé constants for the materials, which take on different values on each phase, $M$ and $I$, e.g. in analogy with (12),

$$(\lambda, \mu)(\omega, \mathbf{x}) = (\lambda, \mu)_M X_M(\omega, \mathbf{x}) + (\lambda, \mu)_I X_I(\omega, \mathbf{x}), \quad \omega \in \Omega, \quad \mathbf{x} \in D.$$

They are related to Young’s modulus $E$ and Poisson’s ratio $\nu$ by

$$\lambda = \frac{E \nu}{(1 + \nu)(1 - 2\nu)}, \quad \mu = \frac{E}{2(1 + \nu)}.$$

The weak or variational form of (27) can be written

Find $\mathbf{u} \in \mathbf{H}$ such that for all $\mathbf{v} \in \mathbf{H}$

$$B(\mathbf{u}, \mathbf{v}) := F(\mathbf{v})$$

$$B(\mathbf{u}, \mathbf{v}) := \int_\Omega \int_D 2\mu(\omega, \mathbf{x}) \varepsilon(\mathbf{u}) : \varepsilon(\mathbf{v}) + \lambda(\omega, \mathbf{x}) \text{div} \mathbf{u} \text{div} \mathbf{v} \ dx \ d\mathcal{P}(\omega)$$

$$F(\mathbf{v}) := \int_\Omega \int_D f(\omega, \mathbf{x}) \cdot \mathbf{v}(\omega, \mathbf{x}) \ dx \ d\mathcal{P}(\omega) + \int_\Omega \int_{\Gamma_N} \sigma(\omega, \mathbf{x}) \cdot \mathbf{v}(\omega, \mathbf{x}) \ dx \ d\mathcal{P}(\omega),$$

with an analogous adjoint problem. Among QoIs are the ensemble average of strains in a subdomain of area $A$ or over an inclusion, e.g. $\bar{Q}(\mathbf{u}) = \mathbb{E} \left[ A^{-1} \int_0^1 \varepsilon_{11}(\mathbf{u}(\cdot, \mathbf{x})) \ dx \right].$

For the surrogate models, we use the same construction as in Section [6], where now the Hashin-Shtrikman bounds on the conductivity are replaced by the Hashin-Shtrikman bounds on the bulk modulus $K$ and shear modulus $\mu$ (in two dimensions) [37, Sect. 21.2.2]

$$K_{\text{LHS}} = K_{\text{arith}} - \frac{(K_M - K_I)^2 \phi_I \phi_M}{K_I \phi_M + K_M \phi_I + \mu_{\text{min}}},$$

$$K_{\text{UHS}} = K_{\text{arith}} - \frac{(K_M - K_I)^2 \phi_I \phi_M}{K_I \phi_M + K_M \phi_I + \mu_{\text{max}}},$$

$$\mu_{\text{LHS}} = \mu_{\text{arith}} - \frac{(\mu_M - \mu_I)^2 \phi_I \phi_M}{\mu_I \phi_M + \mu_M \phi_I + \mu_{\text{min}} \frac{K_{\text{min}}}{K_{\text{min}} + 2\mu_{\text{min}}}},$$

$$\mu_{\text{UHS}} = \mu_{\text{arith}} - \frac{(\mu_M - \mu_I)^2 \phi_I \phi_M}{\mu_I \phi_M + \mu_M \phi_I + \mu_{\text{max}} \frac{K_{\text{max}}}{K_{\text{max}} + 2\mu_{\text{max}}}}$$

(28)

(with $K_{\text{arith}} = K_I \phi_I + K_M \phi_M$, $K_{\text{min}} = \min \{ K_I, K_M \}$, $K_{\text{max}} = \max \{ K_I, K_M \}$ and similarly for $\mu$), from which the corresponding values for $\lambda$ can be derived by the relationship $\lambda = K - \frac{2}{3} \mu$. 

17
Remark 3 (Comment on Remark 1 for elasticity). If Poisson’s ratio is the same in both material, then one may reach the wrong conclusion that the scaling argument of Remark 1 still holds for linear elasticity. This is not the case when using homogenized coefficients, as the nonlinear relationships in (28) lead to an effective Poisson’s ratio which is in general different from the Poisson’s ratio of the two materials.

8. Numerical Experiments

In this section, we present the results for the algorithms and theory given thusfar applied to representative model problems in heat conduction and in elastostatics in random, two-phase media.

For all cases, we have tested the MLMC algorithm for $L = 2$ and $L = 3$ levels. In each test case, the model selection has been performed once for both $L = 2$ and $L = 3$. In Algorithm 1, we have used $M_0 = 100$, $R = 0.5$ and $\gamma = 0.25$ as inputs and the error indicator $\eta_{est}$ of (10). The latter provides a fairly rough estimate of the error, but it showed good performance in driving the adaptive process while having relatively low computational cost.

For the spatial discretization, conforming linear finite elements have been used. In our test cases, we consider a length scale in the microstructure which is about ten percent of the length scale in the homogenized part and therefore using non-matching grids in the two regions would have not brought significant cost savings. However, when the ratio between the scales is higher, then using a fixed non-matching grid in the homogenized region brings significant cost savings. The works $W_l$, $l = 1, \ldots, L$, are computed as $N_{dofs}^l + N_{dofs}^{l-1}$, where $N_{dofs}^l$ is the number of degrees of freedom at level $l$. These correspond to the standardised costs of solving the linear algebraic system from the PDE discretization, assuming an algebraic multigrid solver. Here we assume that the cost of solving the system dominates on the cost of meshing one sample.

8.1. Heat Conduction in Random Media

We consider a square domain $D = (0, 1)^2(m)$ with a random number of inclusions of circular shape, with stochastically perturbed radius and position. The domain $D$ is partitioned into $N \times N$ blocks. Each block contains a $n \times n$ tensorial subgrid, denoting the possible positions of the circular inclusions. The subgrid is such that the distance between its points is $h := 1/(nN)$, and the grid starts at distance $\frac{h}{2}$ from the boundary of the block, see Figure 1a. We use $N = 4$ and $n = 5$. The generation of the inclusions in each block is performed as follows:

- the number of inclusions per block, $n_{bl}$, is distributed according to a discrete uniform distribution between some values $0 \leq n_{min}$ and $n_{max} \leq n^2$; we use $n_{min} = 0$ and $n_{max} = n^2$;
- the average positions of the centers of the inclusions are selected taking the first $n_{bl}$ entries of a random permutation of all possible indices $\{1, 2, \ldots, n^2\}$;
- the $x$ and $y$ coordinates of the center of each inclusion are perturbed, each with a uniform random variable $U\left([\frac{-h}{8}, \frac{h}{8}]\right)(m)$;
- the nominal radius of an inclusion, set to $r = \frac{h}{4}$, is perturbed by a uniform random variable $U\left([\frac{-h}{16}, \frac{h}{16}]\right)(m)$.

A sample of the geometry is shown in Figure 1b.

In these numerical experiments, the fine-scale model is given by (11) with $D = (0, 1)^2$, where the boundary $\Gamma_N$ coincides with the top and bottom boundaries of $D$, and $\Gamma_D$ are the left and right boundaries.
The diffusion coefficient $\kappa$ is given by

$$\kappa(\omega, x) = \begin{cases} 
10,000 \, (W/(mK))^{-1} & \text{in the inclusions}, \\
100 \, (W/(mK))^{-1} & \text{in the matrix},
\end{cases}$$

and the applied flux $\sigma$ by

$$\sigma(x) = \begin{cases} 
16,000 \, (Wm^{-2}) & \text{on the top boundary}, \\
0 & \text{on the bottom boundary}.
\end{cases}$$

The surrogate models are as described in Section 5. The error estimator $\eta_{est}$ has shown to overestimate the error in our test cases; therefore, we proceed with an a priori equal splitting of bias and statistical tolerance and do the modification at line 27 of Algorithm 1.

Since, as already mentioned, we focus on the modeling error, we need to use sufficiently fine meshes to ensure that the discretization error is negligible with respect to the modeling error. In this respect, we highlight that the performance of the model-based MLMC algorithm depends on the ratio between the mesh size in the homogenized region and the mesh size in the blocks which are refined, rather than the values of the mesh sizes themselves. We use unstructured, conformal grids with a mesh size of 0.005 in the refined blocks and 0.05 in the homogenized region, leading to about 62,500 degrees of freedom for the fine-scale model (the actual number depends on the sample) and 594 degrees of freedom for the blockwise-homogenized model. The solution $q_{fix}$ of the globally homogenized model is computed with $\kappa_{fix} = 100$ on a regular grid with 40,401 degrees of freedom. Since Remark 1 applies, in Algorithm 1 with start with $S = \{M_{-1}, M_0\}$. For solving the algebraic system of equations stemming from the PDE discretization, we use the sparse LU solver of FEniCS [22, 1].

8.1.1. Results for local average of the solution

In this first test, the QoI is the average of the solution in the circle of centered at $x_q = (0.9, 0.9)$ and radius $r_q = 0.05$, smoothed by a cutoff function:

$$q_1(u) = \frac{1}{A_q} \int_D \chi(x) u(x) \, dx,$$
with
\[ \chi(x) = \begin{cases} 
1, & \text{for } ||x - x_q|| \leq r_q, \\
\cos^2\left(\frac{||x - x_q|| - 0.05}{0.05}\right), & \text{for } r_q < ||x - x_q|| \leq 2r_q, \\
0, & \text{for } ||x - x_q|| > 2r_q.
\end{cases} \]

We consider the tolerances \( TOL_0 = 0.04, TOL_1 = 0.02, TOL_2 = 0.01, TOL_3 = 0.005 \) for \( L = 2 \) and \( TOL_1 = 0.02, TOL_2 = 0.01, TOL_3 = 0.005 \) for \( L = 3 \). In Algorithm 2 and Algorithm 3, we have used \( M = 150 \) to compute the sample averages.

The models selected as levels by the level selection procedure for \( L = 2 \) are depicted in Figure 2. The last level corresponds to the final model selected by the model selection algorithm. For \( L = 3 \), the level selection algorithm selects the two lines of Figure 2 for levels \( l = 2 \) and \( l = 3 \), respectively, and for \( l = 1 \) and all tolerances it selects the model \( M_0 \) with patchwise-homogenized coefficient.

Tables 1 and 2 show, for each tolerance, the number of samples per level (averaged over 15 repetitions of the algorithm). The numbers in parenthesis denote the number of samples that would have been required if the minimum amount of samples required to estimate the variances (150 samples) was neglected. Figure 3 shows the variance reduction across the levels.

The convergence plot is depicted in Figure 4. The reference solution has been computed with a three-level Monte Carlo with tolerance 0.0025. The root mean squared error has been computed averaging over 15 repetitions. The cost for the standard Monte Carlo algorithm on the fine grid is also reported. From Figure 4, we see that the two-level Monte Carlo results in considerable cost savings: it delivers the same accuracy as the plain Monte Carlo at about one eight of its computations cost. The three-level Monte Carlo also results in some, although much more moderate, savings compared to the two-level Monte Carlo.

8.1.2. Results for local average of gradient component

We now consider the QoI
\[ q_2(u) = \frac{1}{\tilde{A}_q} \int_D \chi(x)(-u_y)(x) \, dx, \]
Table 1: Number of samples per level for the two-level MC for $Q_1$, averaged over 15 repetitions. The numbers in parenthesis are the number of samples required if the minimum amount of samples needed to estimate the variances (150 samples) is neglected.

| Tolerance | $\#$samples $l = 1$ | $\#$samples $l = 2$ |
|-----------|----------------------|----------------------|
| 0.04      | 1889.60              | 150 (25.73)          |
| 0.02      | 5054.46              | 150 (67.4)           |
| 0.01      | 20108.27             | 256.87               |
| 0.005     | 81547.67             | 1016.73              |

Table 2: Number of samples per level for the three-level MC for $Q_1$, averaged over 15 repetitions. The numbers in parenthesis are the number of samples required if the minimum amount of samples needed to estimate the variances (150 samples) is neglected.

| Tolerance | $\#$samples $l = 1$ | $\#$samples $l = 2$ | $\#$samples $l = 3$ |
|-----------|----------------------|----------------------|----------------------|
| 0.02      | 12587.73             | 1657.53              | 150 (55.67)          |
| 0.01      | 50774.87             | 6756.67              | 211.87               |
| 0.005     | 205131.33            | 27378.07             | 829.60               |

Figure 3: Sample variances of $q_1$ and of the MC levels for $L = 2$ and $L = 3$, averaged over 15 repetitions.

Figure 4: Convergence plot for the two-level and three level Monte Carlo for $Q_1$, with comparison with the cost of the plain Monte Carlo on the fine-scale model.
where $u_y$ denotes the derivative of the solution in the direction of the second spatial coordinate, while $\chi$ and $\tilde{A}_q$ are as in (29) but now with $x_q = (0.4, 0.4)$.

We consider the tolerances $TOL_0 = 0.1, TOL_1 = 0.05, TOL_2 = 0.025, TOL_3 = 0.0125$ for $L = 2$ and $TOL_1 = 0.05, TOL_2 = 0.025, TOL_3 = 0.0125$ for $L = 3$. In this example we expect a higher variance and therefore, to be on the safe side, we have used $M = 200$ for computing the sample averages in Algorithm 2 and Algorithm 3.

The models selected by the algorithm for $L = 3$ (for different tolerances) are depicted in Figure 5b. For $L = 2$, the algorithm selected, for $TOL_0$, the models in Figure 5a. For $L = 2$ and all other tolerances, the algorithm selected the first line in Figure 5b as first level and the last line in Figure 5b as second level. Although the level selection is independent of the statistical tolerance, we have performed it separately for $TOL_1$ and $TOL_2$ which have the same model at the last level, in order to check for the robustness of the procedure. We can observe from Figure 5b that for the two tolerances the sampling of the level selection affects the choice of the model for $l = 2$, but from the convergence plot of Figure 7 we see that indeed the two different choices of the levels lead to a similar computational cost when normalized to the statistical tolerance.

Tables 3 and 4 show, for each tolerance, the number of samples per level (averaged over 10 repetitions of the algorithm). The numbers in parenthesis denote the number of samples that would have been required if the minimum amount of samples needed to estimate the variances (200 samples) is neglected. Figure 6 shows the variance reduction across the levels. The convergence plot is given in Figure 7. The reference solution has been computed with a three-level Monte Carlo with tolerance 0.00625, and the root mean squared error has been computed averaging over 10 repetitions. The cost for the plain Monte Carlo on the fine grid is also reported. The two-level Monte Carlo allows to reduce the costs of the plain Monte Carlo by a factor greater than 2.5. The three-level Monte Carlo brings also significant savings with respect to the two-level Monte Carlo, almost halving the costs.

### 8.2. Plane Strain Elasticity in Random Media

In this numerical experiment, we test the performance of the error estimator-driven MLMC for plane strain elasticity. The setting is depicted in Figure 8. The PDE is (27) with $f = 0$. We consider homogeneous Dirichlet boundary conditions on the bottom boundary, Neumann boundary conditions with applied traction...
Figure 5: QoI $Q_2$, levels for $TOL_0$ and $L=2$ (left) and levels for $L=3$ (right). The tolerances $TOL_i$, $i=0,\ldots,3$, indicated under the blocks correspond to the total tolerances (bias tolerances set to $TOL_i/\sqrt{2}$, $i=0,\ldots,3$). For all tolerances, the coarsest model corresponds to the blockwise-homogenized solution.

Figure 6: Sample variances of $q_2$ and of the MC levels, averaged over 10 repetitions.
\[ \sigma = (500, 500) \] at the rightmost boundary and homogeneous Neumann boundary conditions elsewhere; see Figure 8.

The Lamé constants corresponds to a Young’s modulus of

\[ E(\omega, x) = \begin{cases} 1000 & \text{in the inclusions,} \\ 100 & \text{in the matrix,} \end{cases} \]

and a Poisson’s ratio of \( \nu = 0.2 \) everywhere, as in [30].

The random inclusions are generated as follows. The domain is divided into blocks as in the left plot of Figure 8. For blocks from 1 to 16, the inclusions are distributed as described in Subsection 8.1 with \( N = 16, n = 4, n_{\min} = 0 \) and \( n_{\max} = n^2 \). For the block 17, we generate the inclusions as for the other blocks, as if it was a square block; then, we retain the inclusion at the top left corner if this inclusion appears in the square block.

The surrogate models are as described in Section 5 with the Hashin-Shtrikman bounds given in Section 7, equation (28).

The QoI that we consider is the average trace of the strain around the point \( x_q = (0.4586, 0.5412) \) in the left plot of Figure 8

\[ q_3 = \frac{1}{\tilde{A}_q} \int_D \chi(x) (\varepsilon_{11}(u(x)) + \varepsilon_{22}(u(x))) \, dx, \]

with

\[ \chi(x) = \begin{cases} 1, & \text{for } ||x - x_q|| \leq r_q, x \in D, \\ \cos^2 \left( \frac{\pi}{2} \frac{||x - x_q|| - 0.05}{0.05} \right), & \text{for } r_q < ||x - x_q|| \leq 2r_q, x \in D, \\ 0, & \text{for } ||x - x_q|| > 2r_q, x \in D \end{cases} \]

and \( r_q = 0.05 \).

As for the heat conduction problem, we need sufficiently fine discretizations for the modeling error to be the main source of error. We use unstructured, conformal grids with a mesh size of 0.006 in the refined blocks and 0.06 in the homogenized region, leading to about 58000 degrees of freedom for the fine-scale model and 1108 degrees of freedom for the blockwise-homogenized model. For solving the algebraic
Figure 8: Setting used for the numerical experiment of Subsection 8.2. Left: subdivision of the domain into blocks. Right: boundary conditions and point $P$ around which the QoI is computed.

| Tolerance | \#samples $l = 1$ | \#samples $l = 2$ |
|-----------|------------------|-------------------|
| 0.03      | 1672.67          | 150 (61.93)       |
| 0.015     | 6835.53          | 249.73            |
| 0.0075    | 25748.73         | 957.73            |
| 0.00375   | 106672.40        | 3838.20           |

Table 5: Number of samples per level for the two-level MC for $Q_3$, averaged over 15 repetitions. The numbers in parenthesis are the number of samples required if the minimum amount of samples needed to estimate the variances (150 samples) is neglected.

system of equations stemming from the PDE discretization, the PETSc \[3\] GMRES solver with algebraic multigrid preconditioner has been used. Due to Remark 3 in Algorithm 1 we start with $S = \{M_0\}$.

We consider the tolerances $TOL_0 = 0.03, TOL_1 = 0.015, TOL_2 = 0.0075, TOL_3 = 0.00375$ for $L = 2$ and $L = 3$. In Algorithm 2 and Algorithm 3 we have used $M = 150$ to compute the sample averages.

The error indicator $\eta_{est}$ has shown to underestimate the error for this test case. Therefore, we have set $TOL_{bias} = 0$ and always used the fine-scale model on the last level, running Algorithm 1 once for all tolerances. Algorithm 2 has been run once for $L = 2$ and once for $L = 3$. The models selected by Algorithm 1 and Algorithm 2 for $L = 2$ and $L = 3$ are depicted in Figures 9 and 10 respectively.

Tables 5 and 6 show, for each tolerance, the number of samples per level (averaged over 15 repetitions of the algorithm). The numbers in parenthesis denote the number of samples that would have been required if the minimum amount of samples required to estimate the variances (150 samples) was neglected. Figure 11 shows the variance reduction across the levels.

The convergence plot is given in Figure 12. The reference solution has been computed averaging over four repetitions of a three-level Monte Carlo with tolerance $0.0015$ and the root mean squared error has been computed averaging over 15 repetitions. The cost for the plain Monte Carlo on the fine grid is also reported. The two-level Monte Carlo allows to reduce the cost of plain Monte Carlo by a factor greater than 2. The three-level Monte Carlo does not bring, in this case, relevant savings compared to the two-level Monte Carlo.
Figure 9: QoI $Q_3$, levels for $L = 2$: level 1 (left) and level 2 (right). For all tolerances, the same levels have been used.

Figure 10: QoI $Q_3$, levels for $L = 3$: level 1 (left), level 2 (center) and level 3 (right). For all tolerances, the same levels have been used.

| Tolerance | $\#$ samples $l = 1$ | $\#$ samples $l = 2$ | $\#$ samples $l = 3$ |
|-----------|----------------------|----------------------|----------------------|
| 0.03      | 3039.33              | 594.07               | 150 (55.80)          |
| 0.015     | 11847.87             | 2300.33              | 217.33               |
| 0.0075    | 50773.93             | 9964.13              | 896.67               |
| 0.00375   | 200587.87            | 38772.20             | 3693.07              |

Table 6: Number of samples per level for the three-level MC for $Q_3$, averaged over 15 repetitions. The numbers in parenthesis are the number of samples required if the minimum amount of samples needed to estimate the variances (150 samples) is neglected.
9. Closing Comments

In this work, a posteriori estimates of modeling error in models of random heterogeneous material are derived and used to construct sequences of surrogate approximations of increasing accuracy of quantities of interest of a fine-scale base model. This framework provides a basis for new model-based Multilevel Monte Carlo (mbMLMC) methods. Algorithms for implementing these methods are described and applied to representative examples in stochastic heat conduction and plane elasticity. Numerical experiments indicate that substantial reduction in computational costs can be realized by mbMLMC over standard MC methods. They also indicate that a construction of surrogate models which reflects features of the QoI is essential for a good performance of mbMLMC.

Acknowledgments

The work of LS and BW was supported by the German Science Foundation, DFG, WO-671 11-1. That of DF and JTO was supported by the U.S. Department of Energy, Office of Science; Office of Advanced Scientific Computing Research under Award DE-SC0009286. We benefited from discussion with Robert.
Lipton on homogenization of random heterogenous media, and we thank Daniel Drzisga for extensive support in the parallelization of the MLMC code. We also acknowledge with gratitude the early discussions with Kumar Vemaganti of the University of Cincinnati.

References

[1] M. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells. The FEniCS project version 1.5. Archive of Numerical Software, 3(100):9–23, 2015.

[2] I. Babuška, R. Tempone, and G. E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. SIAM Journal on Numerical Analysis, 42(2):800–825, 2004.

[3] S. Balay, S. Abhyankar, M. F. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, D. A. May, L. C. McInnes, R. T. Mills, T. Munson, K. Rupp, P. Sanan, B. F. Smith, S. Zampini, H. Zhang, and H. Zhang. PETSc Web page. http://www.mcs.anl.gov/petsc 2018.

[4] P. T. Bauman, J. T. Oden, and S. Prudhomme. Adaptive multiscale modeling of polymeric materials with Arlequin coupling and Goals algorithms. Computer Methods in Applied Mechanics and Engineering, 198(5):799–818, 2009.

[5] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. Asymptotic analysis for periodic structures, volume 5. North-Holland Publishing Company Amsterdam, 1978.

[6] Z. Botev and A. Ridder. Variance reduction. Wiley StatsRef: Statistics Reference Online.

[7] M. Braack and A. Ern. A posteriori control of modeling errors and discretization errors. Multiscale Modeling & Simulation, 1(2):221–238, 2003.

[8] V. Buryachenko. Micromechanics of heterogeneous materials. Springer Science & Business Media, 2007.

[9] C. Calvo-Jurado and W. J. Parnell. Hashin–Shtrikman bounds on the effective thermal conductivity of a transversely isotropic two-phase composite material. Journal of Mathematical Chemistry, 53(3):828–843, 2015.

[10] P. P. Castañeda and J. R. Willis. The effect of spatial distribution on the effective behavior of composite materials and cracked media. Journal of the Mechanics and Physics of Solids, 43(12):1919–1951, 1995.

[11] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. Computing and Visualization in Science, 14(1):3, 2011.

[12] N. Collier, A.-L. Haji-Ali, F. Nobile, E. von Schwerin, and R. Tempone. A continuation multilevel Monte Carlo algorithm. BIT Numerical Mathematics, 55(2):399–432, 2015.

[13] M. Eigel, C. Merdon, and J. Neumann. An adaptive multilevel Monte Carlo method with stochastic bounds for quantities of interest with uncertain data. SIAM/ASA Journal on Uncertainty Quantification, 4(1):1291–1245, 2016.

[14] M. B. Giles. Multilevel monte carlo path simulation. Operations Research, 56(3):607–617, 2008.

[15] M. B. Giles. Multilevel Monte Carlo methods. Acta Numerica, 24:259328, 2015.

[16] S. Heinrich. Monte Carlo complexity of global solution of integral equations. Journal of Complexity, 14(2):151–175, 1998.

[17] H. Hoel, E. Von Schwerin, A. Szepessy, and R. Tempone. Implementation and analysis of an adaptive multilevel Monte Carlo algorithm. Monte Carlo Methods and Applications, 20(1):1–41, 2014.

[18] D. Jeulin and M. Ostoa-Starzewski. Mechanics of random and multiscale microstructures. Springer, 2001.

[19] V. V. Jikov, S. M. Kozlov, and O. A. Oleinik. Homogenization of differential operators and integral functionals. Springer Science & Business Media, 2012.

[20] H. Li, V. Garg, and K. Willcox. Model Adaptivity for Goal-Oriented Inference using Adjoints. Preprint, 2017.

[21] L. Liu. Hashin–Shtrikman bounds and their attainability for two-phase composite material. Journal of Mathematical Chemistry, 53(3):828–843, 2015.

[22] A. Logg, K.-A. Mardal, and G. Wells. Automated solution of differential equations by the finite element method: The FEniCS book, volume 84. Springer Science & Business Media, 2012.

[23] M. Maier and R. Rannacher. Duality-based adaptivity in finite element discretization of heterogeneous multiscale problems. Journal of Numerical Mathematics, 24(3):167–187, 2016.

[24] M. Maier and R. Rannacher. A duality-based optimization approach for model adaptivity in heterogeneous multiscale problems. arXiv preprint arXiv:1611.09437, 2016.

[25] S. A. Mattis and B. Wohlmuth. Goal-oriented adaptive surrogate construction for stochastic inversion. Computer Methods in Applied Mechanics and Engineering, 339:36–60, 2018.

[26] A. Narayan, C. Gittelson, and D. Xiu. A stochastic collocation algorithm with multifidelity models. SIAM Journal on Scientific Computing, 36(2):A495–A521, 2014.

[27] J. T. Oden and S. Prudhomme. Estimation of modeling error in computational mechanics. Journal of Computational Physics, 182(2):496–515, 2002.
[28] J. T. Oden, S. Prudhomme, A. Romkes, and P. T. Bauman. Multiscale modeling of physical phenomena: Adaptive control of models. *SIAM Journal on Scientific Computing*, 28(6):2359–2389, 2006.

[29] J. T. Oden and K. Vemaganti. Adaptive modeling of composite structures: Modeling error estimation. In *Texas Institute for Computational and Applied Mathematics*. Citeseer, 1999.

[30] J. T. Oden and K. S. Vemaganti. Estimation of local modeling error and goal-oriented adaptive modeling of heterogeneous materials: I. Error estimates and adaptive algorithms. *Journal of Computational Physics*, 164(1):22–47, 2000.

[31] B. Peherstorfer, M. Gunzburger, and K. Willcox. Convergence analysis of multifidelity Monte Carlo estimation. *Numerische Mathematik*, pages 1–25, 2016.

[32] B. Peherstorfer, K. Willcox, and M. Gunzburger. Optimal model management for multifidelity Monte Carlo estimation. *SIAM Journal on Scientific Computing*, 38(5):A3163–A3194, 2016.

[33] B. Peherstorfer, K. Willcox, and M. Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *Preprint*, pages 1–57, 2016.

[34] S. Prudhomme and C. M. Bryant. Adaptive surrogate modeling for response surface approximations with application to bayesian inference. *Advanced Modeling and Simulation in Engineering Sciences*, 2(1):22, 2015.

[35] S. Prudhomme and J. T. Oden. On goal-oriented error estimation for elliptic problems: application to the control of pointwise errors. *Computer Methods in Applied Mechanics and Engineering*, 176(1-4):313–331, 1999.

[36] E. Sanchez-Palencia. Homogenization method for the study of composite media. *Asymptotic analysis II*, 985:192–214, 1983.

[37] S. Torquato. *Random heterogeneous materials: microstructure and macroscopic properties*, volume 16. Springer Science & Business Media, 2013.

[38] K. S. Vemaganti and J. T. Oden. Estimation of local modeling error and goal-oriented adaptive modeling of heterogeneous materials: Part II: a computational environment for adaptive modeling of heterogeneous elastic solids. *Computer Methods in Applied Mechanics and Engineering*, 190(46):6089–6124, 2001.

[39] L. Walpole. Elastic behavior of composite materials: theoretical foundations. *Advances in applied mechanics*, 21:169–242, 1981.

[40] J. Willis. Bounds and self-consistent estimates for the overall properties of anisotropic composites. *Journal of the Mechanics and Physics of Solids*, 25(3):185–202, 1977.

[41] X. Zhu, E. M. Linebarger, and D. Xiu. Multi-fidelity stochastic collocation method for computation of statistical moments. *Journal of Computational Physics*, 341:386–396, 2017.