

Portland State University

PDXScholar

---

Mathematics and Statistics Faculty  
Publications and Presentations

Fariborz Maseeh Department of Mathematics  
and Statistics

---

6-2021

# Multilevel Hierarchical Decomposition of Finite Element White Noise with Application to Multilevel Markov Chain Monte Carlo

Hillary R. Fairbanks

*Lawrence Livermore National Laboratory*

Umberto E. Villa

*University of Texas at Austin*

Panayot S. Vassilevski

*Portland State University, panayot@pdx.edu*

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/mth\\_fac](https://pdxscholar.library.pdx.edu/mth_fac)



Part of the [Physical Sciences and Mathematics Commons](#)

Let us know how access to this document benefits you.

---

## Citation Details

Published as: Fairbanks, H. R., Villa, U., & Vassilevski, P. S. (2021). Multilevel Hierarchical Decomposition of Finite Element White Noise with Application to Multilevel Markov Chain Monte Carlo. *SIAM Journal on Scientific Computing*, (0), S293-S316.

This Pre-Print is brought to you for free and open access. It has been accepted for inclusion in Mathematics and Statistics Faculty Publications and Presentations by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

# MULTILEVEL HIERARCHICAL DECOMPOSITION OF FINITE ELEMENT WHITE NOISE WITH APPLICATION TO MULTILEVEL MARKOV CHAIN MONTE CARLO

HILLARY R. FAIRBANKS<sup>1</sup> AND UMBERTO VILLA<sup>2</sup> AND PANAYOT S. VASSILEVSKI<sup>1,3</sup>

<sup>1</sup>*Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA, USA*

<sup>2</sup>*Electrical & Systems Engineering Department, Washington University in St. Louis, St. Louis, MO, USA*

<sup>3</sup>*Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland, OR, USA*

**ABSTRACT.** In this work we develop a new hierarchical multilevel approach to generate Gaussian random field realizations in an algorithmically scalable manner that is well-suited to incorporate into multilevel Markov chain Monte Carlo (MCMC) algorithms. This approach builds off of other partial differential equation (PDE) approaches for generating Gaussian random field realizations; in particular, a single field realization may be formed by solving a reaction-diffusion PDE with a spatial white noise source function as the right-hand side. While these approaches have been explored to accelerate forward uncertainty quantification tasks, e.g. multilevel Monte Carlo, the previous constructions are not directly applicable to multilevel MCMC frameworks which build fine scale random fields in a hierarchical fashion from coarse scale random fields. Our new hierarchical multilevel method relies on a hierarchical decomposition of the white noise source function in  $L^2$  which allows us to form Gaussian random field realizations across multiple levels of discretization in a way that fits into multilevel MCMC algorithmic frameworks. After presenting our main theoretical results and numerical scaling results to showcase the utility of this new hierarchical PDE method for generating Gaussian random field realizations, this method is tested on a four-level MCMC algorithm to explore its feasibility.

**Keywords.** Gaussian random field, nonlinear Bayesian inference, Markov chain Monte Carlo, multilevel Markov chain Monte Carlo, high-dimensional uncertainty quantification, algebraic multigrid

## 1. INTRODUCTION

Spatially correlated random fields are commonly used in the numerical simulation of partial differential equations (PDEs) with variable coefficients. In the case where these coefficients are not well known, as is typically the case in many geophysics applications where the coefficient describes a physical parameter, the coefficient is modeled as a random field, and uncertainty quantification (UQ) may be applied as a tool to assess the reliability of the model as well as the sensitivity to changes in this parameter. To reduce the uncertainty in

---

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 (LLNL-JRNL-820098).

the system, we may further improve the model by utilizing observational data in a Bayesian framework. That is, data related to the model output, as well as information about the model may be combined to learn the probability distribution of the variable coefficient.

For large-scale applications, common methods to perform Bayesian inference are infeasible. With the refinement of the spatial discretization scheme, both forming realizations of these random fields and performing forward PDE simulations are computationally demanding, as many approaches do not scale with the increase in problem size. Furthermore, Bayesian inference approaches are typically limited to Markov chain Monte Carlo (MCMC) [40, 31, 44], and its variants, which require a large number of simulations as the parameter space is explored. However infeasible this approach may be, MCMC methods still lie at the root of many nonlinear Bayesian inference algorithms due to ease of implementation as well as the ability to be applied in a blackbox fashion.

Over the recent decades new MCMC approaches have been developed to accelerate the parameter space exploration, in some cases allowing to perform nonlinear Bayesian inference on large-scale applications. Notable approaches include those which utilize local approximations of the Hessian and gradient to modify the MCMC proposal [39, 43, 18, 19, 12, 10]. This method, dubbed, stochastic Newton MCMC, was shown in [43] to accelerate mixing; however, it requires gradient and Hessian information in addition to having solvers for the forward PDE and adjoint PDE models, as opposed to simply having the forward PDE model.

Another class of approaches includes delayed acceptance MCMC algorithms, which utilize cheaper model approximations to accelerate the parameter search (also via proposal distribution modification) [13]. Several works have been completed that develop and explore the use of cheaper models with coarser spatial discretizations in a two-stage or multilevel framework. Early works that employ coarser spatial discretizations include [33] and [24]; the former utilizes a Metropolis coupled MCMC to swap proposals between coarse and fine chains, and the latter performs a delayed acceptance where sample proposals are only completed with the fine grid solver if their associated coarse grid solutions have been accepted. More recent works have investigated multilevel MCMC approaches. In particular, in [22], the authors developed an approach to both accelerate the mixing of the MCMC chain by using multiple levels, each with coarser spatial discretizations, and accelerate the sampling by performing variance reduction via multilevel Monte Carlo following the ideas of [32, 28, 8, 14, 48]. Analysis of a multilevel MCMC was completed in [34]. While promising speed up results have been shown, numerical testing has been limited to 2D spatial domains and structured meshes.

While scalable solvers are available for several classes of PDE forward models, the sampling of large-scale Gaussian random field on unstructured meshes in an algorithmically scalable manner is still a challenging task. The use of a Karhunen-Loève (KL) expansion to form Gaussian random field realizations requires calculating the eigenvalues and eigenfunctions of the covariance function [27]. A straightforward, though perhaps naïve implementation will have a cost that grows cubically with the degrees of freedom associated with the spatial discretization of the random field, i.e., the mesh size. While there are tools to improve this scaling, e.g., hierarchical matrix formations [9] or Nyström methods [51], storage and the ability to calculate the KL expansions for unstructured meshes are roadblocks to large-scale and extreme-scale applications. Other approaches to sampling, such as circulant embedding [30], are not directly applicable to problems with unstructured meshes.

An alternative scalable approach to generate random field realizations is via the stochastic reaction-diffusion PDE formulated in [50] and solved with finite elements in [38]. Using this approach, each independent realization requires solving the stochastic PDE with an independent realization of spatial white noise function as forcing term. Applying this approach in a multilevel setting, such as multilevel Monte Carlo or multilevel MCMC requires forming coupled realizations of Gaussian random fields on multiple levels of discretization. A few works have completed this, including [23] where fine and coarse level realizations are coupled together to perform massively parallel multilevel Monte Carlo. In [41, 42] the authors solve a mixed PDE on the space of piecewise constants, and generate matching fine and coarse realizations by restricting the fine grid spatial white noise to the coarse level, using operators and solvers from element agglomerated algebraic multigrid (AMGe). In [16] the authors couple the coarse and fine level realizations using the primal formulation of the PDE.

While these approaches have been incorporated successfully into the multilevel Monte Carlo framework, they are not useful in the multilevel MCMC framework. This is because, in the multilevel MCMC approach, we must first sample from the coarse level, and then form a fine level random field in a hierarchical manner from the coarse realization. As this sampling approach has not yet been developed (to the best of the authors' knowledge), this paper seeks to fill this void.

**1.1. Contributions of this Work.** In this work, we develop an algorithmically scalable, hierarchical Gaussian random field sampling method that can be used to construct proposals distributions in the multilevel MCMC framework. Specifically, we plug our sampling method into the multilevel MCMC framework of [22], though it is also applicable to other two-level MCMC or delayed acceptance MCMC approaches discussed earlier. To do this, we utilize the finite element solvers from [42] to map an independent realization of spatial white noise to a Gaussian random field realization. Within this mapping, we incorporate a new component: a hierarchical decomposition of the white noise (via  $L^2$  projection operators) across discretization levels. This new feature allows us to perform MCMC stepping on coarse level white noise, extend it to a finer level, and then perform an MCMC step on independent white noise in the complementary space.

The remainder of this paper is organized as follows. In Section 2, mathematical notation relevant to Gaussian random fields is presented. Our new hierarchical approach is presented in Section 3; this includes the theoretical aspects of performing a hierarchical direct decomposition of white noise – in a two-level and multilevel framework – resulting in a hierarchical approach to form Gaussian random field realizations. The numerical implementation is discussed in Section 4, in the form of algorithms, as well as visualizations of the random field hierarchical decomposition. Section 5 explores the cost and scaling of our multilevel hierarchical sampling technique applied to the Egg Model [35] using three levels; in particular, we show that the algorithm is scalable. Section 6 incorporates this new hierarchical sampling technique into a four-level MCMC following the approach of [22], and shows that we obtain similar improvements in the multilevel acceptance rate, variance decay, and total computational cost when compared to the single-level approach.

**1.2. Mathematical Notation.** As a reference to the reader, we define the majority of this paper's notation in Table 1. The first section of the table introduces general variables that provide a basis for the majority of the mathematical notation. The second section of the

table refers to discrete variables that are used in various finite element representations, and that are frequently referred to throughout this work.

## 2. GAUSSIAN RANDOM FIELDS

In this work we consider a particular class of random fields, that is, spatially correlated Gaussian random fields, which in this context, will be used to describe an uncertain physical process. Define the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , with sample space  $\Omega$ ,  $\sigma$ -algebra  $\mathcal{F}$ , and probability  $\mathbb{P}$ . Given the spatial domain of interest  $D \subset \mathbb{R}^d$ , with  $d = 2, 3$ , we seek to form random field realizations of  $\{u(\mathbf{x}, \omega) \in L^2(D) : \mathbf{x} \in D, \omega \in \Omega\}$ , that follow a Gaussian prior density  $u \sim \mathcal{N}(0, \mathcal{C})$  with zero mean and covariance operator  $\mathcal{C}$ . To ensure the mesh independent statistics of the random field  $u$ ,  $\mathcal{C}$  is a trace-class operator [47]. Specifically, we define the covariance operator as the squared inverse elliptic operator (see e.g. [25, 11, 43]). That is,

$$(1) \quad \mathcal{C} = \mathcal{A}^{-2} \text{ with } \mathcal{A}u := -\operatorname{div} \left( \frac{1}{g} \nabla u \right) + \frac{\kappa^2}{g} u,$$

where  $\kappa$  denotes the inverse of the correlation length and  $g$  controls the marginal variance of the field. Using the above notation, we then define the probability density function as

$$(2) \quad d\mu(u) \propto \exp \left( -\frac{1}{2} \langle \mathcal{A}u, \mathcal{A}u \rangle \right),$$

where  $\langle \mathcal{A}u, \mathcal{A}u \rangle = \int_D (\mathcal{A}u)^2 d\mathbf{x}$ .

TABLE 1. Mathematical Notation.

Variable	Description
$\mathbf{x} \in D \subset \mathbb{R}^d$	Point in spatial domain, $d = 2$ or $3$
$\omega \in \Omega$	Outcome of Sample Space
$u \in \Theta := L^2(D)$	$L^2$ function defined over $D$
$\zeta := \zeta(\mathbf{x}, \omega)$	White noise function in $D$
$\mathbf{q} \in \mathbf{R} := H(\operatorname{div}; D)$	$H(\operatorname{div})$ function defined over $D$
Discrete Variable	
$h, H$	Subscripts to denote fine and coarse level objects
$\ell$	Subscript denoting running level $\ell$ index, with $\ell = 0$ as finest
$k$	Subscript denoting target level of an algorithm
$\mathcal{T}_\ell$	Level $\ell$ finite element triangulation
$u_\ell \in \Theta_\ell$	Piecewise constant function defined on $\mathcal{T}_\ell$
$\mathcal{Q}_\ell$	Orthogonal projection from $L^2$ to $\Theta_\ell$
$P_\ell$	Interpolation operator mapping between $\Theta_{\ell+1}$ and $\Theta_\ell$
$\Pi_\ell$	Restriction operator mapping between $\Theta_\ell$ and $\Theta_{\ell+1}$
$\zeta_\ell$	White noise representation in $\Theta_\ell$
$\zeta_\ell$	Coefficient vector of white noise finite element representation in $\Theta_\ell$
$\xi_\ell$	Vector of random elements
$b_\ell$	Vector representation of white noise in $\Theta_\ell$
$\mathbf{q}_\ell \in \mathbf{R}_\ell$	Function of the lowest order Raviat-Thomas space on $\mathcal{T}_\ell$
$M_\ell, B_\ell, W_\ell$	Mass matrices for various level $\ell$ spaces

As described in [50, 38], for unbounded domains  $D := \mathbb{R}^d$ , the covariance operator in (1) leads to a Gaussian random field of the Matérn family with smoothness parameter  $\nu$  and marginal variance  $\sigma^2$  respectively given by

$$\nu = 2 - \frac{d}{2} \text{ and } \sigma^2 = \frac{g^2 \Gamma(\nu)}{\Gamma(\nu + d/2) (4\pi)^{d/2} \kappa^{2\nu}}.$$

In particular, in three-spatial dimensions, this gives the well-known exponential covariance operator

$$(\mathcal{C}u)(\mathbf{x}) := \int_D \text{cov}(\mathbf{x}, \mathbf{y}) u(\mathbf{y}) d\mathbf{y}, \text{ with } \text{cov}(\mathbf{x}, \mathbf{y}) := \frac{g^2}{8\pi\kappa} \exp(-\kappa \|\mathbf{x} - \mathbf{y}\|_2).$$

For a finite domain  $D \subset \mathbb{R}^d$ , suitable boundary conditions need be stipulated to reduce boundary artifacts, see e.g. [45, 36, 20]. In this work, we choose to extend the domain  $D$  to a larger domain  $\bar{D} \subset \mathbb{R}^d$  and equip  $\mathcal{A}$  with homogeneous Neumann boundary conditions on  $\partial\bar{D}$ , as described in [42].

For sample-based UQ approaches—such as standard Monte Carlo—we desire to generate samples of this random field  $u(\mathbf{x}, \omega)$  to serve as input field data to a model of interest. In our application (which we further detail in Section 6), we wish to generate permeability field realizations,  $k = \exp(u(\mathbf{x}, \omega))$ , each of which serves as an input to Darcy’s equations.

**2.1. A Stochastic PDE Approach for Finite Element Random Fields.** As presented in [25, 11, 43], a realization of a Gaussian random field, with covariance operator  $\mathcal{C}$  given by (1), can be generated by solving the stochastic reaction-diffusion PDE

$$\mathcal{A}u = \zeta,$$

where  $\zeta := \zeta(\mathbf{x}, \omega)$  is spatial Gaussian white noise. The spatial Gaussian white noise  $\zeta$  is an  $L^2(D)$ -bounded generalized function [38, Appendix B], such that

$$(3) \quad \langle \zeta, v \rangle \sim \mathcal{N}(0, \|v\|_{L^2(D)}^2) \quad \forall v \in L^2(D).$$

In the following, we consider a particular PDE-based approach that uses a mixed formulation to generate field realizations. That is, we follow the approach of [41, 42], which allows us to work in the space of piecewise constants. For large-scale applications this is beneficial as it provides a natural way to define spatial white noise, and the associated mass matrix is easily diagonalizable.

**2.1.1. A Mixed Formulation.** For a fixed  $\omega \in \Omega$ , a Gaussian random field realization  $u := u(\mathbf{x}, \omega)$  is calculated by solving the stochastic PDE:

$$(4) \quad \begin{aligned} (\boldsymbol{\rho}, \mathbf{s}) + (\text{div } \mathbf{s}, u) &= 0 & \forall \mathbf{s} \in H(\text{div}) \\ (\text{div } \boldsymbol{\rho}, v) - \kappa^2 (u, v) &= -g \langle \zeta, v \rangle & \forall v \in L^2, \end{aligned}$$

where  $(\cdot, \cdot)$  denotes the  $L^2(D)$  inner product [41, 42]. Above, the spatial Gaussian white noise  $\zeta := \zeta(\mathbf{x}, \omega)$  is a zero-mean random Gaussian field on  $D$  such that  $\langle \zeta, v \rangle \sim \mathcal{N}(0, \|v\|_{L^2(D)}^2)$ , for any function  $v \in L^2(D)$  (see (3)). Note, properties of finite element white noise will be discussed in the following section.

Define the spaces  $\Theta = L^2(D)$  with inner product  $(u, v) = \int_D uv d\mathbf{x}$  for all  $u, v \in \Theta$  and  $\mathbf{R} = H(\text{div}; D) := \{\mathbf{q} \in [L^2(D)]^d \mid \text{div } \mathbf{q} \in L^2(D), \mathbf{q} \cdot \mathbf{n} = 0 \text{ on } \partial D\}$  with inner product

$(\mathbf{q}, \mathbf{s}) = \int_D \mathbf{q} \cdot \mathbf{s} dx$  for all  $\mathbf{q}, \mathbf{s} \in \mathbf{R}$ . Let  $\mathbf{R}_h, \Theta_h$  be the pair of the lowest order Raviart-Thomas and piecewise constant finite element spaces associated with the given triangulation  $\mathcal{T}_h$ .

For a fixed  $\omega \in \Omega$ , discrete solutions  $\boldsymbol{\rho}_h \in \mathbf{R}_h$  and  $u_h \in \Theta_h$  are calculated from the mixed system,

$$(5) \quad \begin{aligned} (\boldsymbol{\rho}_h, \mathbf{s}_h) + (\operatorname{div} \mathbf{s}_h, u_h) &= 0 & \forall \mathbf{s}_h \in \mathbf{R}_h \\ (\operatorname{div} \boldsymbol{\rho}_h, v_h) - \kappa^2 (u_h, v_h) &= -g \langle \zeta, v_h \rangle & \forall v_h \in \Theta_h. \end{aligned}$$

2.1.2. *Finite Element Representation of White Noise.* Since moments of  $\zeta$  are well-defined for functions in  $\Theta_h$ , we can define the mapping  $\mathcal{Q}_h : \zeta \mapsto \mathcal{Q}_h \zeta \in \Theta_h$  using the identity

$$(6) \quad (\mathcal{Q}_h \zeta, v_h) = \langle \zeta, v_h \rangle \quad \forall v_h \in \Theta_h.$$

That is, a realization of white noise on a given finite element mesh  $\mathcal{T}_h$  can be represented in  $\Theta_h$  using the mapping  $\mathcal{Q}_h$  as follows

$$(7) \quad \zeta_h := \mathcal{Q}_h \zeta = \sum_{\tau \in \mathcal{T}_h} \zeta_\tau \chi_\tau,$$

where  $\{\chi_\tau\}$  is an  $L^2$ -orthogonal basis of piecewise constants spanning  $\Theta_h$ .

Using the expansion in (7) and the equivalence in (6), it follows that the righthand side of (5) will have the coefficient vector  $b_h \equiv ((\zeta_h, \chi_\tau))_{\tau \in \mathcal{T}_h}$ . As a consequence of using piecewise constant basis functions, each inner product simplifies as

$$b_h = (\zeta_h, \chi_\tau) = \zeta_\tau \|\chi_\tau\|^2.$$

In other words,  $b_h = W_h \boldsymbol{\zeta}_h$ , where  $W_h$  the diagonal mass matrix for the space  $\Theta_h$  and  $b_h$  is the vector collecting the coefficients  $\zeta_\tau$  in the expansion (6).

To generate realizations of white noise in  $\Theta_h$ , we consider the following properties (see [6, Section 1.4.3] and [7, Section 2.4.5] for details). We note that, while we present these properties with respect to an  $L^2$ -orthogonal basis of piecewise constants, they can be generalized to the situation of a non-orthogonal basis.

**Property 2.1** (White noise in  $\Theta_h$ ). *Let  $\zeta$  be white noise in  $D$ . Then, for the projection of  $\zeta$  onto the basis  $\{\chi_\tau\}_{\tau \in \mathcal{T}_h}$  of  $\Theta_h$ , denoted  $\zeta_h$  as in (7), it follows that,*

$$\mathbb{E}[(\zeta_h, \chi_\tau)] = \mathbb{E}[\langle \zeta, \chi_\tau \rangle] = 0,$$

and

$$\mathbb{E}[(\zeta_h, \chi_{\tau_i})(\zeta_h, \chi_{\tau_j})] = \mathbb{E}[\langle \zeta, \chi_{\tau_i} \rangle \langle \zeta, \chi_{\tau_j} \rangle] = (\chi_{\tau_i}, \chi_{\tau_j}),$$

which implies

$$\mathbb{E}[(\zeta_h, \chi_{\tau_i})_{\tau_i \in \mathcal{T}_h} ((\zeta_h, \chi_{\tau_i})_{\tau_j \in \mathcal{T}_h})^T] = ((\chi_{\tau_i}, \chi_{\tau_i})_{\tau_i, \tau_j \in \mathcal{T}_h}) = W_h,$$

where  $W_h$  is the (diagonal) mass matrix for the space  $\Theta_h$ .

These properties follow from the theoretical aspects of white noise. Specifically, the covariance between two volumes  $A$  and  $B$  (within  $D$ ) is equivalent to the mass of the intersection of the two volumes (further theoretical aspects of Gaussian white noise may be found in [6, 7]), and for finite element white noise this implies that the covariance is equivalent to the mass matrix. Using the above properties, we can show that for  $W_h \boldsymbol{\zeta}_h$  to be a realization of Gaussian white noise, we require  $\boldsymbol{\zeta}_h = (\zeta_\tau)_{\tau \in \mathcal{T}_h} \sim \mathcal{N}(0, W_h^{-1})$ .

**Lemma 2.1.** *Given the basis  $\{\chi_\tau\}_{\tau \in \mathcal{T}_h}$  of  $\Theta_h$ , associated mass matrix  $W_h = ((\chi_{\tau_i}, \chi_{\tau_j}))_{\tau_i, \tau_j \in \mathcal{T}_h}$ , and  $\zeta_h = (\zeta_\tau)_{\tau \in \mathcal{T}_h}$  sampled from  $\mathcal{N}(0, W_h^{-1})$ , it follows that  $W_h \zeta_h$  is a realization of white noise in  $\Theta_h$ .*

*Proof.* Following Property 2.1, it is sufficient to show that  $\mathbb{E}[W_h \zeta_h] = 0$  and  $\mathbb{E}[W_h \zeta_h (W_h \zeta_h)^T] = W_h$ . As  $\zeta_h \sim \mathcal{N}(0, W_h^{-1})$ , it is clear that  $\mathbb{E}[W_h \zeta_h] = 0$ . As for the covariance, we have

$$\begin{aligned} \mathbb{E}[W_h \zeta_h (W_h \zeta_h)^T] &= W_h \mathbb{E}[\zeta_h \zeta_h^T] W_h \\ &= W_h. \end{aligned}$$

□

2.1.3. *Finite Element Representation of Gaussian Random Fields.* In the actual computation of  $u_h$ , we use the equivalent vector representation for the righthand side of (5), defined as

$$(8) \quad -gb_h = -gW_h^{1/2} \xi_h,$$

with  $\xi_h \sim \mathcal{N}(0, I)$ . We note this equivalence is made clearer in Section 4. As we are in the space of piecewise constants in  $L^2$ , the square root of the (diagonal) mass matrix  $W_h$  is easily calculated. Let  $M_h$  be the mass matrix associated with inner product  $(\rho_h, \mathbf{s}_h)$  and  $B_h$  the mass matrix associated with the bilinear form  $(\text{div } \mathbf{s}_h, u_h)$ . Then the matrix representation of (5) is given as

$$(9) \quad \begin{bmatrix} M_h & B_h^T \\ B_h & -\kappa^2 W_h \end{bmatrix} \begin{bmatrix} \rho_h \\ u_h \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ -gb_h \end{bmatrix},$$

with  $b_h$  defined by (8).

For ease of notation, we introduce the scaled negative Schur Complement of (9) defined by

$$(10) \quad A_h := \frac{\kappa^2}{g} W_h + \frac{1}{g} B_h M_h^{-1} B_h^T.$$

As demonstrated in [41], solutions  $u_h$  of the mixed system in (9) are discrete realizations of a Gaussian random field with density  $\mu_h \sim \mathcal{N}(\mathbf{0}, C_h)$ , where  $C_h = A_h^{-1} W_h A_h^{-1}$ . It then follows that the corresponding probability density function is

$$(11) \quad \mu_h(u_h) \propto \exp(-u_h^T A_h W_h^{-1} A_h u_h) = \exp(-b_h^T W_h^{-1} b_h).$$

### 3. MULTILEVEL HIERARCHICAL DECOMPOSITION OF FINITE ELEMENT WHITE NOISE

In what follows we study the computational aspects of sampling the righthand side in (5) from a coarse finite element space  $\Theta_H \subset \Theta_h$ , and its (direct) hierarchical complement space  $(I - \mathcal{Q}_H)\Theta_h$ , where  $\mathcal{Q}_H : L^2 \mapsto \Theta_H$  is the corresponding  $L^2$ -projection. For any  $\zeta_h \in \Theta_h$ , we use the two-level hierarchical decomposition

$$\zeta_h = \mathcal{Q}_H \zeta_h + (I - \mathcal{Q}_H) \zeta_h$$

to decompose  $\zeta_h$  into the spaces  $\Theta_H$  and  $\Theta_h \setminus \Theta_H$ . Since we work with spaces of discontinuous (piecewise constant) functions  $\Theta_h$  and  $\Theta_H$  with associated mass matrices  $W_h$  and  $W_H$ , respectively, the projection  $\mathcal{Q}_H$  is easily implemented (by inverting a diagonal (coarse) mass matrix).



Define  $P$  to be the interpolation matrix that relates the coarse coefficient vector  $\zeta_H$  of  $\zeta_H$  (expanded in terms of the basis of  $\Theta_H$ ) and the fine coefficient vector  $\zeta_h$  of  $\zeta_h \in \Theta_H \subset \Theta_h$  expanded in terms of the basis of  $\Theta_h$ . That is,

$$\zeta_h = P\zeta_H.$$

Let  $\Pi = W_H^{-1}P^TW_h$  denote the restriction operator, then  $P\Pi$  is the matrix representation of  $\mathcal{Q}_H$  and  $\Pi P = I$ . That is, we have  $P\zeta_H = P\Pi\zeta_h$ , or  $\zeta_H = \Pi\zeta_h$ .

In what follows, we first seek to show that  $\mathcal{Q}_H\zeta_h$  gives rise to a coarse random coefficient vector  $\zeta_H \sim \mathcal{N}(0, W_H^{-1})$ .

**Lemma 3.1.** *Let  $\zeta_h \in \Theta_h$ , with coefficient vector  $\zeta_h \sim \mathcal{N}(0, W_h^{-1})$ . Then  $\zeta_H \equiv \mathcal{Q}_H\zeta_h$  has coefficient vector  $\zeta_H \sim \mathcal{N}(0, W_H^{-1})$ .*

*Proof.* Given the associated coarse coefficient  $\zeta_H = \Pi\zeta_h$  with  $\zeta_h \sim \mathcal{N}(0, W_h^{-1})$ , it is clear the mean is zero. For the covariance matrix, we have

$$\begin{aligned} \mathbb{E}[\zeta_H\zeta_H^T] &= \mathbb{E}[\Pi\zeta_h(\Pi\zeta_h)^T] \\ &= W_H^{-1}P^TW_h\mathbb{E}[\zeta_h\zeta_h^T]W_hPW_H^{-1} \\ &= W_H^{-1}. \end{aligned}$$

Above, we use  $\mathbb{E}[\zeta_h\zeta_h^T] = W_h^{-1}$  and the Galerkin relation between the coarse and the fine level mass matrices,  $W_H = P^TW_hP$ .  $\square$

Next we present our main lemma, which allows us to utilize this two-level, hierarchical decomposition to form a realization of white noise on  $\Theta_h$ .

**Lemma 3.2.** *Let  $\zeta_H \in \Theta_H$  be a coarse representation of white noise with a coarse coefficient vector  $\zeta_H \sim \mathcal{N}(0, W_H^{-1})$ , and let  $\zeta_h \in \Theta_h$  be a fine representation of white noise with fine coefficient vector  $\zeta_h \sim \mathcal{N}(0, W_h^{-1})$ , such that  $\zeta_H$  and  $\zeta_h$  are independent. Then the fine level function*

$$(12) \quad \zeta'_h = \zeta_H + (I - \mathcal{Q}_H)\zeta_h$$

*is a representation of the white noise in  $\Theta_h$ .*

*Proof.* First, consider the coefficient vector of  $\zeta'_h$ , given as

$$(13) \quad \zeta'_h = P\zeta_H + (I - P\Pi)\zeta_h.$$

To prove  $\zeta'_h$  is a representation of Gaussian white noise, we must show Definition 2.1 holds, that is  $\mathbb{E}[\zeta'_h] = 0$ , and  $\mathbb{E}[\zeta'_h(\zeta'_h)^T] = W_h^{-1}$ . We assume that  $\zeta_H$  and  $\zeta_h$  are independent, which implies that  $\mathbb{E}[\zeta_H\zeta_h^T] = \mathbb{E}[\zeta_H]\mathbb{E}[\zeta_h^T] = 0$ . Hence for the covariance matrix, we have

$$\begin{aligned} \mathbb{E}[\zeta'_h(\zeta'_h)^T] &= \mathbb{E}[(P\zeta_H + (I - PW_H^{-1}P^TW_h)\zeta_h)(P\zeta_H + (I - PW_H^{-1}P^TW_h)\zeta_h)^T] \\ &= P\mathbb{E}[\zeta_H\zeta_H^T]P^T + (I - PW_H^{-1}P^TW_h)\mathbb{E}[\zeta_h\zeta_h^T](I - W_hPW_H^{-1}P^T) \\ &= PW_H^{-1}P^T + (I - PW_H^{-1}P^TW_h)W_h^{-1}(I - W_hPW_H^{-1}P^T) \\ &= W_h^{-1}. \end{aligned}$$

That is,  $\zeta'_h \sim \mathcal{N}(0, W_h^{-1})$ ; hence  $\zeta'_h$  is a fine finite element representation of white noise. It is clear also that  $\zeta_H = \mathcal{Q}_H\zeta'_h$  and  $(I - \mathcal{Q}_H)\zeta_h = (I - \mathcal{Q}_H)\zeta'_h$ .  $\square$

In conclusion, the finite element hierarchical (direct) decomposition based on  $\mathcal{Q}_H$  provides a hierarchical decomposition of the fine finite element white noise into a coarse finite element representation of white noise plus a computational hierarchical (direct) complement which also involves fine finite element representation of white noise.

**3.1. The multilevel hierarchical decomposition.** To extend the the above two-level hierarchical decomposition of Gaussian white noise to a multilevel hierarchical decomposition, we introduce the following notation. Let  $\mathcal{T}_0 \equiv \mathcal{T}_h$  denote the finest level triangulation of  $D$ , with a hierarchy of  $L$  coarser levels given as  $\{\mathcal{T}_\ell\}_{\ell=1}^L$ , such that  $\mathcal{T}_L$  represents the coarsest triangulation. We consider the finite element space  $\Theta_\ell$  to be the space of piecewise constant functions associated with the triangulation  $\mathcal{T}_\ell$ , for  $\ell = 0, \dots, L$ , and with mass matrix  $W_\ell$ ; and  $\mathbf{R}_\ell$  the lowest order Raviart-Thomas space associated with the triangulation  $\mathcal{T}_\ell$ . Additionally, define the sequence of  $L^2$ -projections  $\mathcal{Q}_\ell : L^2 \mapsto \Theta_\ell$  with  $\ell = 0, \dots, L$ .

In what follows, we construct the multilevel hierarchical decomposition of white noise for a given level  $k < L$ .

**Theorem 3.1.** *Consider the representations of white noise, given as  $\zeta_\ell \in \Theta_\ell$  with associated coefficient vectors  $\zeta_\ell \sim \mathcal{N}(0, W_\ell^{-1})$ , for  $\ell = k, \dots, L$ , such that each  $\zeta_\ell$  is independent. Then the level  $k$  function*

$$(14) \quad \zeta'_k = \zeta_L + \sum_{\ell=k}^{L-1} (I - \mathcal{Q}_{\ell+1})\zeta_\ell,$$

with  $k < L$ , is a representation of the white noise in  $\Theta_k$ .

*Proof.* From Lemma 3.1 the result is clear for  $k = L - 1$ , i.e., the two-level case. For additional levels, the result follows by applying Lemma 3.1 in a recursive manner.  $\square$

The associated coefficient representation is defined similarly to (13); however, we replace the subscript  $h$  with  $k$  to denote the level, and let  $P_k$  be the interpolation matrix that maps the level  $k + 1$  coefficient vector  $\zeta_{k+1}$  of  $\zeta_{k+1}$  to the level  $k$  coefficient vector  $\zeta_k$  of  $\zeta_k$ , such that  $\zeta_k = P_k \zeta_{k+1}$ . This hierarchical coefficient representation (in two-level form) is given as

$$(15) \quad \zeta'_k = P_k \zeta_{k+1} + (I - P_k W_{k+1}^{-1} P_k^T W_k) \zeta_k.$$

Just as in the proof, we can hierarchically build a level  $k$  coefficient vector by starting on the coarsest level and adding on coefficients projected onto the complimentary spaces, as will be further detailed in the next section (see, e.g., Algorithm 4.1).

#### 4. IMPLEMENTATION OF THE HIERARCHICAL SAMPLER

Recall from Lemma 2.1 that we may sample (single-level) finite element white noise on level  $k$  via  $b_k = W_k \zeta_k$  with  $\zeta_k = (\zeta_i)_{i=1}^n \sim \mathcal{N}(0, W_k^{-1})$ . While we can use the decomposition in (15) for our hierarchical implementation, we instead alter this representation to accomplish two additional goals: first, that on each level we sample from a  $\mathcal{N}(0, I)$  distribution, and second, that we utilize the interpolation and restriction operators  $P_k$  and  $\Pi_k = W_{k+1}^{-1} P_k^T W_k$ .

By multiplying (15) by  $W_k$ , and after simple algebraic manipulation, we obtain the following hierarchical representation of the right hand side of (5):

$$b'_k = \Pi_k^T b_{k+1} + (I - \Pi_k^T P_k^T) b_k,$$

where  $b'_k = W_k \zeta'_k$ .

For algorithmic efficiency (and to meet our additional two goals), we simplify the above using the fact that  $b_\ell = W_\ell^{1/2} \xi_\ell$  with  $\xi_\ell \sim \mathcal{N}(0, I)$  for  $\ell = k, k + 1$  and write

$$b'_k = W_k^{1/2} \xi'_k = \Pi_k^T (W_{k+1}^{1/2} \xi_{k+1}) + (I - \Pi_k^T P_k^T) (W_k^{1/2} \xi_k).$$

In practice, we use this finite element white noise formulation of  $b_k$ , where we may construct a realization using multiple coarser levels, beyond that of level  $k+1$ . This process is described in Algorithm 4.1. For a given level  $k$ , we first calculate finite element white noise on the coarsest level  $L$ , denoted  $b_L$ . Then we iterate through each finer level, where we calculate  $b_\ell$  by first interpolating the coarser  $b_{\ell+1}$  (which was previously calculated), and then adding a spatial white noise realization that is complementary to the coarser  $\ell + 1$  space – this is accomplished by multiplying level  $\ell$  spatial white noise,  $W_\ell^{1/2}\boldsymbol{\xi}_\ell$ , with  $(I - \Pi_\ell^T P_\ell^T)$ , which projects the level  $\ell$  spatial white noise orthogonal to the coarser space(s). After each iterate, we refine a level (decrease  $\ell$  by 1), and repeat this process. This done until we reach level  $k$ , and the resulting  $b_k$  provides us with our hierarchically generated realization of spatial white noise, which can then be used in the righthand side of the discrete problem (5).

**Algorithm 4.1:** Form finite element white noise via new hierarchical approach.

**Input:** Current level  $k$  (with  $0 \leq k \leq L$ ),  $L$ , independent  $\{\boldsymbol{\xi}_L, \boldsymbol{\xi}_{L-1}, \dots, \boldsymbol{\xi}_k\}$  with each  $\mathcal{N}(\mathbf{0}, I_k)$

$\ell = L$

$b_L = W_L^{1/2}\boldsymbol{\xi}_L$

$\ell = \ell - 1$

**while**  $\ell \geq k$  **do**

$b_\ell = \Pi_\ell^T b_{\ell+1} + (I - \Pi_\ell^T P_\ell^T)W_\ell^{1/2}\boldsymbol{\xi}_\ell$   
     $\ell = \ell - 1$

**end**

**Output:**  $\{b_L, b_{L-1}, \dots, b_k\}$

In this work we employ the same linear system of [41, Section 2.2], but instead of spatial white noise generated strictly on the fine level, we use our hierarchical approach. For a given level  $k$ , we seek to calculate solutions  $(\boldsymbol{\rho}_k, u_k) \in \mathbf{R}_k \times \Theta_k$  via the linear system

$$(16) \quad \begin{bmatrix} M_k & B_k^T \\ B_k & -\kappa^2 W_k \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho}_k \\ u_k \end{bmatrix} = \begin{bmatrix} 0 \\ -g b_k \end{bmatrix},$$

where  $M_k$  be the mass matrix associated with inner product  $(\boldsymbol{\rho}_k, \mathbf{s}_k)$ ,  $W_k$  with the inner product  $(u_k, v_k)$  which is diagonal,  $B_k$  with the bilinear form  $(\text{div } \mathbf{s}_k, u_k)$ , and  $b_k$  is hierarchically generated spatial white noise (generated via Algorithm 4.1). For a scalable, parallelizable implementation, we have several solvers we may consider, one of which – hybridization AMG approach from [37, 21] – is amenable to large-scale applications because the mass matrices need only be computed one time (on each level), and then may be reapplied to different realizations of  $b_k$ .

To define the Gaussian densities  $\mu_k$  at level  $k$  we proceed as in Section 2.1.3. Let us formally introduce the negative scaled Schur Complement of (16) defined by

$$A_k := \frac{\kappa^2}{g} W_k + \frac{1}{g} B_k M_k^{-1} B_k^T.$$

Then solutions of (16) are Gaussian random vectors with distribution  $\mu_k \sim \mathcal{N}(0, A_k^{-1} W_k A_k^{-1})$  and corresponding probability density

$$\mu_k(u_k) \propto \exp(-u_k^T A_k W_k^{-1} A_k u_k) = \exp(-b_k^T W_k^{-1} b_k).$$

We also define the conditionally Gaussian density  $u_k|u_{k+1}$  based on our hierarchical decomposition of white noise in Algorithm 4.1. Sampling from the prior distribution  $\mu_k$  and from the conditional distribution  $u_k|u_{k+1}$  are summarized in Algorithm 4.2 and Algorithm 4.3. These algorithms will be used to define the proposal distributions within the multilevel MCMC algorithm in Section 6.

**Algorithm 4.2:** Generate a sample  $u_k$  from the prior distribution  $\mu_k$  at level  $k$ .

**Input:** Current level  $k$  (with  $0 \leq k \leq L$ )

Sample  $\xi_k \sim \mathcal{N}(\mathbf{0}, I_k)$

Define  $b_k = W_k^{1/2} \xi_k$

Compute  $u_k$  by solving (16)

**Output:**  $u_k$

**Algorithm 4.3:** Generate a sample  $u_k$  from the conditional distribution  $u_k|u_{k+1}$

**Input:** Current level  $k$  (with  $0 \leq k < L$ ), the coarse level sample  $u_{k+1} = A_{k+1}^{-1} b_{k+1}$

Sample  $\xi_k \sim \mathcal{N}(\mathbf{0}, I_k)$

Define  $b_k = \Pi_k^T b_{k+1} + (I - \Pi_k^T P_k^T) W_k^{1/2} \xi_k$

Compute  $u_k$  by solving (16)

**Output:**  $u_k$

**4.1. Random Field Realizations Using Hierarchical Components.** To visualize the hierarchical components of a fine level solution  $u_0$ , we consider the Egg model domain [35] using three levels of refinement with 18.5K, 148K, and 1.18M elements for levels  $\ell = 2, 1$ , and 0, respectively. Here we skip over the model details, as these will be addressed in the following section, and focus on the new hierarchical sampler.

Using Algorithm 4.1 with  $k = 0$  and  $L = 2$ , we generate the three components of the righthand side given as  $b_0 = \Pi_0^T \Pi_1^T W_2^{1/2} \xi_2 + \Pi_0^T (I - \Pi_1^T P_1^T) W_1^{1/2} \xi_1 + (I - \Pi_0^T P_0^T) W_0^{1/2} \xi_0$ . For visualization purposes, we separate these three components of  $b_0$  and solve with each independently. That is, we seek solutions  $u_0^{C\ell}$  via

$$(17) \quad \begin{aligned} A_0 u_0^{C2} &= \Pi_0^T \Pi_1^T W_2^{1/2} \xi_2, \\ A_0 u_0^{C1} &= \Pi_0^T (I - \Pi_1^T P_1^T) W_1^{1/2} \xi_1, \\ A_0 u_0^{C0} &= (I - \Pi_0^T P_0^T) W_0^{1/2} \xi_0. \end{aligned}$$

Note that the fine level realization is simply  $u_0 := u_0^{C2} + u_0^{C1} + u_0^{C0}$ . Figure 1 (a)-(c) displays the solutions  $u_0^{C2}$ ,  $u_0^{C1}$ , and  $u_0^{C0}$ . These results showcase the novelty of this hierarchical approach – that is, the finite element white noise decomposition enables a realization  $u_0$  to be decomposed into independent components across multiple levels. Moreover, this hierarchical approach induces a separation of scales among the terms  $u_0^{C\ell}$ , similar to that induced by the hierarchical KL-based sampling in [22]. On the coarse levels, the terms  $u_0^{C\ell}$  capture the smooth components of  $u_0$ , while, on finer levels, the terms  $u_0^{C\ell}$  only contain the highly oscillatory components of  $u_0$ . This property plays a fundamental role in accelerating the mixing and reducing the variance of the multilevel MCMC in Section 6. This is clearly illustrated by considering the sums of the components  $u_0^{C\ell}$  shown in Figure 1 (d)-(e). In

particular, Figure 1 (d) displays  $u_0^{C2} + u_0^{C1}$ , and Figure 1 (e) displays the complete fine level realization  $u_0$ .

## 5. NUMERICAL RESULTS: MULTILEVEL HIERARCHICAL SAMPLE GENERATION

In this section, we test the hierarchical sampler scaling performance on the ‘Egg model’ [35], as it contains a large, irregular domain. The Egg domain is contained by a  $480 \text{ m} \times 480 \text{ m} \times 28 \text{ m}$  bounding box. We note that, as we are employing the approach of [42], we require performing mesh embedding, that is, the Egg model domain is embedded within a  $512 \text{ m} \times 512 \text{ m} \times 44 \text{ m}$  domain. This mitigates variance inflation along the boundary due to Neumann boundary conditions (see [38, Section 2.3] and [41, 42] for additional discussion). Figure 2 displays both the original Egg model mesh and enlarged mesh (in which it is embedded) for the coarsest level, both with hexahedral elements of size  $8 \text{ m} \times 8 \text{ m} \times 4 \text{ m}$ . Finer mesh resolutions are formed by uniformly refining by a factor of two in each direction.

We consider three levels  $\ell = 0, 1, 2$ , with degrees of freedom (corresponding to the number of unknowns in the mixed PDE system as in (16)) given in Table 2 with  $\text{NP} = 36$  total MPI processes; then, for a fixed problem size per processor, we increase the number of processes to  $\text{NP} = 288$  and then  $\text{NP} = 2304$ . Gaussian random field realizations were generated following our new hierarchical PDE sampling approach; that is, for levels  $\ell = 0, 1, 2$ , level  $\ell$  hierarchical white noise was sampled according to Algorithm 4.1, and realizations of  $u_\ell$  were formed by solving the linear system in (16) on the Egg domain. Numerical simulations were performed

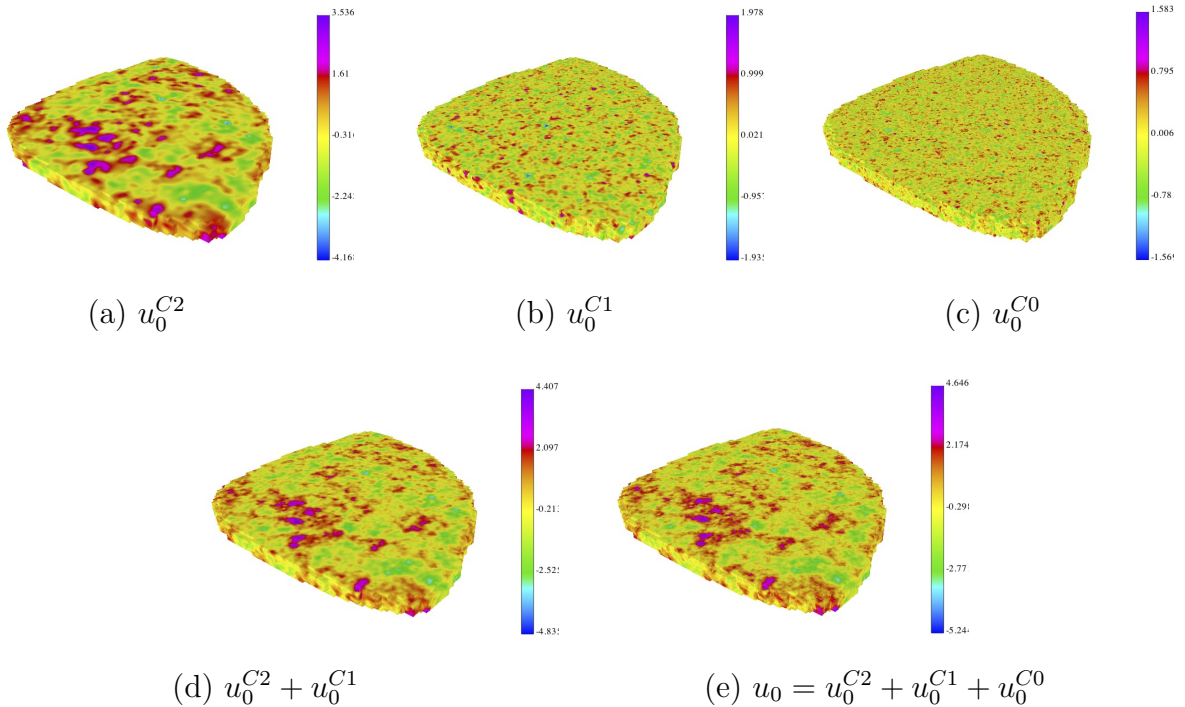


FIGURE 1. Various components of a realization  $u_0$  on three levels as defined in (17), and generated from Algorithm 4.2 and Algorithm 4.3. Visualizations are rendered with GLVis [1].

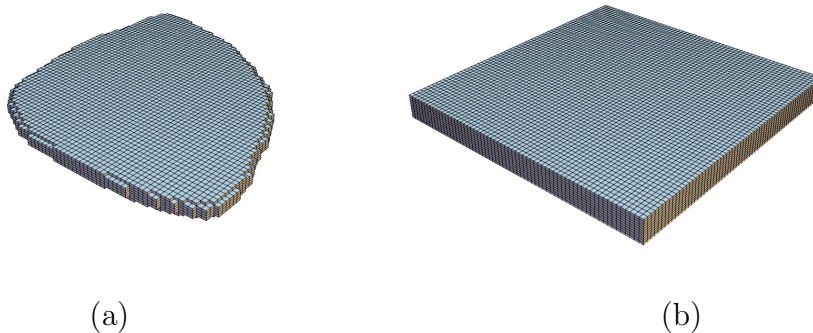


FIGURE 2. (a) Original Egg model mesh containing 18.5K elements. (b) Enlarged mesh, in which the Egg model mesh is embedded, extends two elements in each direction beyond the Egg model mesh bounding box, and contains 45K elements. Both meshes displayed correspond to the coarsest level.

TABLE 2. Number of global degrees of freedom (DOFs) associated with each level, for each set of process numbers. The DOFs here are associated with the number of unknowns in the mixed PDE system as in (16).

NP	DOFs $\ell = 0$	DOFs $\ell = 1$	DOFs $\ell = 2$
36	4.8063e+06	6.0788e+05	7.7758e+04
288	3.8223e+07	4.8063e+06	6.079e+05
2304	3.0488e+08	3.8223e+07	4.8063e+06

using tools developed in ParELAG [4], a parallel C++ library for performing numerical upscaling of finite element discretizations and AMG techniques, and ParELAGMC [5], a parallel element agglomeration MLMC library. These libraries use MFEM [3] to generate the fine grid finite element discretization and HYPRE [2] to handle massively parallel linear algebra. In particular, we employ hybridization AMG [37, 21], where the rescaled linear system is solved with conjugate gradient (CG) preconditioned by HYPRE’s BoomerAMG. Note, all timing results were executed on the Quartz cluster at Lawrence Livermore National Laboratory, consisting of 2,688 nodes where each node has two 18-core Intel Xeon E5-2695 processors. For the weak scaling results, we use 36 MPI processes per node.

Table 3 provides the average wall time for the hybridization AMG solver, where the number of preconditioned CG (PCG) iterations – to reduce the  $l^2$  norm of the residual by a factor of  $10^6$  – are provided in parentheses. These timing results indicate favorable scaling on the finest level; however, parallel efficiency does degrade on the coarser levels, which is to be expected, as we are limited to our solver and AMG’s performance on coarse levels (see [26]). Nonetheless, these results show mesh independence of our approach for all levels in the hierarchy, as the iteration count is stable with increased problem size.

In addition, Figure 3 displays these weak scaling results (for 100 simulations), as well as the efficiency decay with increased problem size. While we can claim algorithmic scalability, the

TABLE 3. Average wall time (seconds) to solve (16) using hybridization AMG (averaged over 100 realizations). The average number of PCG iterations are provided in parentheses.

Level $\ell$	Local DOFs	NP=36	NP =288	NP =2304
0	135,438	2.42 (11)	2.88 (12)	3.15 (13)
1	17,368	0.187 (10)	0.258 (11)	0.316 (12)
2	2,280	0.0152 (9)	0.0297 (11)	0.0635 (11)

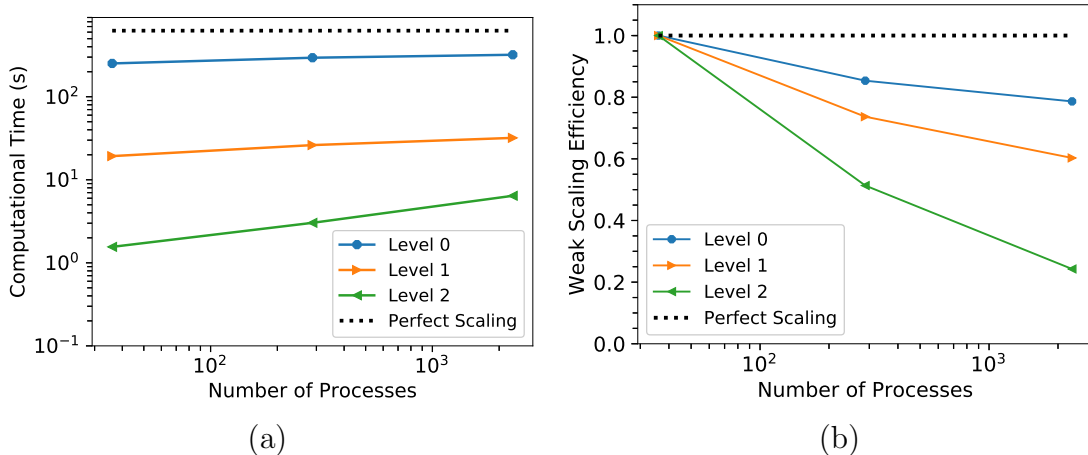


FIGURE 3. (a) Weak scaling for the three different levels; here levels refers to a fixed number of elements per processor. (b) Associated efficiency across the three levels.

degraded scaling on the coarser levels indicates a drawback with our solver implementation, that is, for coarser discretizations we require fewer processes than the fine discretizations to obtain favorable parallel efficiency on all levels. Because of this, we are unable to get complete scaling across all levels; rather each level will benefit from a different number of processes. A topic for a follow up study will include coarse-grid redistribution in order to improve the coarse level performance, see, e.g., [49]. Nonetheless, it is still clear that the coarse level samples are significantly faster to generate than the finest level samples, which is essential for multilevel MCMC performance.

## 6. HIERARCHICAL PDE APPROACH FOR MULTILEVEL MCMC

In this section, we apply the proposed hierarchical PDE-based sampling approach to solve a nonlinear Bayesian inference problem. We use the multilevel MCMC framework in [22] to explore the posterior distribution of the uncertain parameter and estimate moments (the mean) of a scalar quantity of interest  $Q$ .

In particular, we consider the problem of inferring a log-normal permeability field from cell-averaged pressure measurements for a single phase steady state subsurface problem. In what follows, we denote with  $u \in \Theta$  the uncertain parameter representing the logarithm of the permeability field, with  $(\mathbf{q}, p) \in \mathbf{R} \times \Theta$  the state variables representing the flow velocity

and pressure, and with  $p_{\text{obs}} \in \mathbb{R}^m$  the data representing cell-averaged pressure observations at  $m$  given measurement locations.

*Prior distribution.* We assume a Gaussian prior density on the spatially varying log-permeability coefficient, i.e.  $u \sim \mathcal{N}(u_*, \mathcal{C})$ , with covariance operator  $\mathcal{C}$  and mean value  $u_* = 0$ . To ensure that the inference problem is well-posed in infinite dimensions, we use a squared inverse elliptic operator in (1) to define the prior covariance operator, see for example [47, 25, 11, 43]. Samples from the prior distribution can then be drawn by solving (4) as we explained in Section 2.1, and their probability can be computed using (2).

*Forward map.* Let  $y = \mathcal{F}(u)$  denote the forward map from the uncertain field  $u \in \Theta$  to the observable  $y \in \mathbb{R}^m$ . The map  $\mathcal{F} = \mathcal{B} \circ \mathcal{H}$  is the composition of a forward PDE solve  $\mathcal{H}$  that computes the pressure field  $p$  for a given realization of log-permeability  $u$  and a linear operator  $\mathcal{B}$  that evaluates the pressure  $p$  on local cells. For a single phase porous media flow, with  $k = \exp(u)$ , we define  $p = \mathcal{H}(u)$  as the solution to the mixed formulation of the Darcy's equations

$$(18) \quad \begin{aligned} (k^{-1} \mathbf{q}, \mathbf{s}) - (\text{div } \mathbf{s}, p) &= (\mathbf{f}, \mathbf{s}) \quad \forall \mathbf{s} \in \mathbf{R} \\ (\text{div } \mathbf{q}, v) &= 0 \quad \forall v \in \Theta, \end{aligned}$$

with Dirichlet boundary condition  $p = p_D$  on  $\Gamma_D$ , enforced by the right-hand side  $\mathbf{f}$ , and Neumann boundary condition  $\mathbf{q} \cdot \mathbf{n} = 0$  on  $\Gamma_N$ , where  $\Gamma_D$  and  $\Gamma_N$  are non-overlapping partitions of  $\partial D$ .

*Likelihood function.* We assume that the measured data  $p_{\text{obs}} \in \mathbb{R}^m$  are corrupted by additive Gaussian noise  $\eta$  with zero mean and covariance  $\Gamma_\eta = \sigma_\eta^2 I_m$ , where  $I_m$  is the identity matrix in  $\mathbb{R}^m$ . That is,

$$(19) \quad p_{\text{obs}} = \mathcal{F}(u) + \eta, \quad \eta \sim \mathcal{N}(0, \Gamma_\eta).$$

From the noise model in (19), we have that the conditional probability of  $p_{\text{obs}}$  given  $u$  is also Gaussian with mean  $\mathcal{F}(u)$  and covariance  $\Gamma_\eta$ , that is

$$p_{\text{obs}} | u \sim \mathcal{N}(\mathcal{F}(u), \Gamma_\eta).$$

The likelihood function  $\pi^{\text{like}}(p_{\text{obs}} | u)$  then reads

$$(20) \quad \pi^{\text{like}}(p_{\text{obs}} | u) \propto \exp \left( -\frac{1}{2} \|p_{\text{obs}} - \mathcal{F}(u)\|_{\Gamma_\eta^{-1}}^2 \right),$$

where  $\|\cdot\|_{\Gamma_\eta^{-1}}^2$  denotes the  $\Gamma_\eta^{-1}$ -weighted  $l^2$  norm in  $\mathbb{R}^m$ .

*Posterior distribution.* By applying Bayes' theorem, the posterior density  $\nu$  in the infinite dimensional case is given by

$$(21) \quad \nu(u | p_{\text{obs}}) \propto \pi^{\text{like}}(p_{\text{obs}} | u) d\mu(u),$$

where  $\pi^{\text{like}}(p_{\text{obs}} | u)$  is the likelihood function in (20) and  $d\mu(u)$  is the prior density in (2). We note that, although the prior distribution and likelihood functions are both Gaussian, the posterior distribution  $\nu(u | p_{\text{obs}})$  is not Gaussian because of the nonlinearity introduced by the forward map  $\mathcal{F}$ . Thus, there is no closed form solution to the Bayesian inference problem and therefore we will use MCMC sampling to explore the posterior distribution.

*Quantity of interest.* Finally, let us introduce the scalar quantity of interest  $Q = Q(u)$  representing the flux across the outflow boundary  $\Gamma_{\text{out}}$ , which is defined as

$$(22) \quad Q = \frac{1}{|\Gamma_{\text{out}}|} \int_{\Gamma_{\text{out}}} \mathbf{q}(\cdot, \omega) \cdot \mathbf{n} \, dS,$$



where  $\mathbf{n}$  represents the outward unit vector normal to  $\Gamma_{\text{out}} \subset \partial D$ .

Our goal is to estimate the posterior mean of  $Q$ , defined as

$$(23) \quad \mathbb{E}_\nu[Q] = \int_{\Omega} Q(u) d\nu(u|p_{\text{obs}}),$$

by sampling the posterior distribution (21) using multilevel MCMC. As a reference to the reader, notation introduced and frequently used in this section is provided in Table 4.

**6.1. Markov Chain Monte Carlo.** For the single-level approach, the log-permeability  $u$ , associated likelihood  $\pi^{\text{like}}(p_{\text{obs}}|u)$ , and QoI  $Q$  (see (4), (20), (22), and, respectively) are approximated numerically by solving the mixed PDEs in (4) and (18) utilizing a finite element approach on triangulation  $\mathcal{T}_k$  (for the finest level  $k$ ). We denote these discrete approximations as  $u_k$ ,  $\pi_k^{\text{like}}$ , and  $Q_k = Q(u_k)$ .

MCMC, and in particular, Metropolis-Hastings, is a modified Monte Carlo approach, where samples  $Q_k$  are generated from the target (posterior) distribution via a Markov chain. Then the posterior QoI expectation  $\mathbb{E}_{\nu_k}[Q]$  may be approximated as

$$(24) \quad \hat{Q}_k^{\text{MCMC}} = \frac{1}{N} \sum_{i=n+1}^{n+N} Q_k^{(it)},$$

where  $Q_k^{(it)} = Q(u_k^{(it)})$ ,  $n$  is the number of samples discarded as burn-in, and  $t$  is the sub-sampling rate to obtain independent samples (see Appendix A). To generate subsequent samples within a chain, a new  $u_k^{\text{prop}}$  is sampled from the proposal distribution and subjected to Metropolis-Hastings acceptance/rejection criterion. In this work, we utilize a preconditioned Crank-Nicolson stepping scheme with step size  $\beta > 0$ , where samples  $\psi_k$  from  $\mu_k$  are drawn using Algorithm 4.2. Thanks to the prior-invariance of the preconditioned Crank-Nicolson proposal, the sample  $u_k^{\text{prop}}$  is then accepted with probability  $\alpha_k^{\text{SL}}$  defined in (26)

TABLE 4. Bayesian inference notation.

Infinite-Dimensional Variable	Description
$d\mu(u)$	Prior density of $u$
$p_{\text{obs}}$	Observed local pressure data in $D$
$\pi^{\text{like}}(p_{\text{obs}} u)$	Likelihood function
$\nu(u p_{\text{obs}})$	Posterior density
$\mathbb{E}_\nu[\cdot]$	Mean with respect to posterior density
$Q = Q(u)$	Quantity of interest
Finite-Dimensional Variable	
$d\mu_\ell(u_\ell)$	Prior density of $u_\ell$ on level $\ell$
$\pi_\ell^{\text{like}}(p_{\text{obs}} u_\ell)$	Likelihood function
$\nu_\ell(u_\ell p_{\text{obs}})$	Posterior density on level $\ell$
$\alpha_\ell^{\text{SL}}$	Single-level acceptance probability on level $\ell$
$\alpha_\ell^{\text{ML}}$	Multilevel acceptance probability on level $\ell$
$\mathbb{E}_{\nu_\ell}[\cdot]$	Mean with respect to level $\ell$ posterior density
$\mathbb{V}_{\nu_\ell}[\cdot]$	Variance with respect to level $\ell$ posterior density
$Q_\ell = Q(u_\ell)$	Quantity of interest on level $\ell$

[15]. The procedure is summarized in Algorithm 6.1; additional details can be found in [22].

**Algorithm 6.1:** Single level Metropolis-Hastings MCMC Algorithm with preconditioned Crank-Nicolson proposal to generate a posterior sample  $u_k^{(i)} | u_k^{(i-1)}$

- Given  $u_k^{(i-1)}$ , propose  $u_k^{\text{prop}}$  using preconditioned Crank-Nicolson:

$$(25) \quad u_k^{\text{prop}} := \sqrt{1 - \beta^2} u_k^{(i-1)} + \beta \psi_k,$$

where  $\psi_k \sim \mu_k$  is computed using Algorithm 4.2.

- Accept  $u_k^{(i)} = u_k^{\text{prop}}$  with probability

$$(26) \quad \alpha_k^{SL}(u_k^{\text{prop}} | u_k^{(i-1)}) = \min \left\{ 1, \frac{\pi_k^{\text{like}}(p_{\text{obs}} | u_k^{\text{prop}})}{\pi_k^{\text{like}}(p_{\text{obs}} | u_k^{(i-1)})} \right\}$$

- Return  $u_k^{(i)}$  and  $Q_k^{(i)} = Q_k(u_k^{(i)})$ .

The cost of performing MCMC depends on how quickly the chain mixes as well as the variance of  $\hat{Q}_k$ . The first—the mixing of the chain—is controlled by the autocorrelation of samples within the chain. As adjacent samples in the chain are correlated (and not independent), the integrated autocorrelation time  $\tau_Q$  of the chain will determine how many steps (and thus forward simulations) are required to get to the next independent sample. The second—the variance of  $\hat{Q}_k$ —is controlled by the number of independent samples used in the estimator, i.e.,  $N$ . The accuracy is similar to Monte Carlo in that the required number of (independent) samples to achieve a desired mean squared error depends on the variance of  $Q_k$  as well as the bias introduced by numerically approximating  $Q$ . If we require  $N$  independent simulations, with an integrated autocorrelation time (rounded up to an integer value) of  $t$  (see Appendix A), then we require at least  $tN$  simulations. Thus acceleration approaches should seek to reduce  $t$  and  $N$  by increasing the mixing of the chain and reducing the variance of the estimator, respectively.

**6.2. Multilevel Markov Chain Monte Carlo.** To accelerate MCMC, we consider the multilevel framework in [22], which utilizes chains at coarser spatial discretization levels to perform the majority of likelihood functions evaluations. Similar to previous sections, let us denote the log-normal permeability field, the Darcy pressure and flux, and the QoI at discretization level  $\ell$  with the symbols  $u_\ell$ ,  $(p_\ell, \mathbf{q}_\ell)$ ,  $Q_\ell := Q_\ell(u_\ell)$ , respectively, for  $k = 0 \leq \ell \leq L$ . Then the posterior mean of  $Q_0$  is equivalently written as

$$(27) \quad \mathbb{E}_{\nu_0}[Q_0] = \mathbb{E}_{\nu_L}[Q_L] + \sum_{\ell=0}^{L-1} (\mathbb{E}_{\nu_\ell}[Q_\ell] - \mathbb{E}_{\nu_{\ell+1}}[Q_{\ell+1}]),$$

where  $\nu_\ell$  is the discrete posterior measure on level  $\ell$ . Following [22], for each level  $\ell$ , we define a multilevel estimator  $\hat{Y}_\ell^{N_\ell}$  of the difference  $\mathbb{E}_{\nu_\ell}[Q_\ell] - \mathbb{E}_{\nu_{\ell+1}}[Q_{\ell+1}]$  and write

$$(28) \quad \hat{Y}_\ell^{N_\ell} = \frac{1}{N_\ell} \sum_{i=n_\ell+1}^{n_\ell+N_\ell} Y_\ell^{(it_\ell)} = \frac{1}{N_\ell} \sum_{i=n_\ell+1}^{n_\ell+N_\ell} (Q_\ell^{(it_\ell)} - Q_{\ell+1}^{(it_\ell t_{\ell+1})}).$$

Above  $n_\ell$  corresponds to the burn-in on level  $\ell$ ,  $N_\ell$  is the effective sample size on level  $\ell$  (defined later in (32)),  $t_\ell$  and  $t_{\ell+1}$  are the estimated integrated autocorrelation times of the chains at levels  $\ell$  and  $\ell + 1$ , respectively (see Appendix A for details). The key aspect of the multilevel MCMC is how to couple Markov chains at different levels so that: *i*) the variance of  $Y_\ell$  is much smaller than that of  $Q_\ell$ , *ii*) information from coarser level chains are used to accelerate mixing of finer level chains (higher acceptance rate, smaller integrated autocorrelation time). In this section, we focus on how to generalize the multilevel MCMC algorithm [22, Algorithm 3] to replace the KL decomposition-based sampling with our scalable multilevel PDE samplers described in Section 4.

As motivated in Remark 1, in what follows, we describe a two-level chain to evaluate the difference estimator  $\hat{Y}_\ell$  at a generic level  $0 \leq \ell < L$ . Given a coarse sample  $u_{\ell+1}^{(j-1)t_{\ell+1}}$ , we advance the coarse chain at level  $\ell + 1$  by  $t_{\ell+1}$  steps using single-level Metropolis-Hastings as in Algorithm 6.1. This results in a coarse sample  $u_{\ell+1}^{(jt_{\ell+1})}$  that is independent of  $u_{\ell+1}^{(j-1)t_{\ell+1}}$ .

To propose  $u_\ell$  on the finer level  $\ell$ , we use the two-level preconditioned Crank-Nicolson in (29), where  $\psi_\ell$  is sampled from the conditional distribution  $\psi_\ell | u_{\ell+1}^{(j-1)t_{\ell+1}}$  using Algorithm 4.3. Note that the independence of  $u_{\ell+1}^{(jt_{\ell+1})}$  from  $u_{\ell+1}^{(j-1)t_{\ell+1}}$  guarantees that also  $\psi_\ell$  is independent of  $u_\ell^{(j-1)}$ . That means that the two-level preconditioned Crank-Nicolson proposal in (29) satisfies the assumptions of [22, Lemma 3.1], and therefore the multilevel acceptance probability  $\alpha_\ell^{ML}(u_\ell^{\text{prop}} | u_\ell^{(i-1)})$  in (30) satisfies the detailed balance condition. Algorithm 6.2 summarizes the generation of the paired fine and coarse level samples  $u_\ell^{(j)}$  and  $u_{\ell+1}^{(jt_{\ell+1})}$ .

Note that, if  $u_\ell^{\text{prop}}$  is accepted at step  $j$ , then  $u_\ell^{(j)}$  and  $u_{\ell+1}^{(jt_{\ell+1})}$  are correlated. Specifically, both  $u_\ell^{(j)}$  and  $u_{\ell+1}^{(jt_{\ell+1})}$  are generated from the same coarse level white noise functional  $b_{\ell+1}$ , and thus the difference  $u_\ell^{(j)} - P_\ell u_{\ell+1}^{(jt_{\ell+1})}$  is small. This is observed in Section 4.1, where Figure 1 (b)-(c) display these differences, defined as solutions  $u_0^{C\ell}$  (see (17)). This means that one should expect  $Y_\ell^{(j)}$  to be small when step  $j$  is accepted, which is a necessary condition to achieve multilevel acceleration. The numerical results presented next demonstrate that our algorithm is indeed able to achieve multilevel acceleration of the chain mixing and variance reduction.

**Remark 1.** *Another small difference with respect to the work in [22] is that each  $\hat{Y}_\ell^{N_\ell}$  estimate uses only two levels. Specifically, Algorithm 6.2 uses a single auxiliary chain on the coarser level  $\ell + 1$  to estimate  $\hat{Y}_\ell$ , while [22] runs auxiliary chains on all coarser levels. Our decision to do this is based on algorithmic simplicity and scalability. That said, utilizing all coarser levels is a detail that may be considered in future work.*

**6.3. Four-Level Markov Chain Monte Carlo Results.** To demonstrate our method is well-suited for multilevel MCMC algorithms, we test a four-level MCMC with the hierarchical stochastic PDE solver by utilizing Algorithm 6.2. The computational domain is a unit cube discretized using tetrahedral elements, with about 1.57M elements on the finest level. Each coarser level is formed by uniformly coarsening by a factor of 8, until obtaining 3,072 elements on the coarsest level. For the prior, we consider Gaussian random fields with correlation length  $\lambda = 0.3$  and marginal variance of  $\sigma^2 = 0.5$ . For the observational pressure data, we synthetically generate a realization  $p_{\text{obs}} \in \mathbb{R}^{25}$  with  $\sigma_\eta^2 = 0.005$ ; this is done via (19) on a reference mesh with approximately 12.5M elements.

**Algorithm 6.2:** Two-level Metropolis-Hastings MCMC Algorithm to generate paired samples  $u_{\ell+1}^{(jt_{\ell+1})}$  and  $u_{\ell}^{(j)}|u_{\ell}^{(j-1)}$

Part I. Advance the coarse chain at level  $\ell + 1$  by  $t_{\ell+1}$  steps:

- Given  $u_{\ell+1}^{(j-1)t_{\ell+1}}$ , apply Algorithm 6.1 on level  $\ell + 1$  for  $t_{\ell+1}$  steps
- Store  $u_{\ell+1}^{(jt_{\ell+1})}$  and  $Q(u_{\ell+1}^{(jt_{\ell+1})})$

Part II. Advance the fine chain at level  $\ell$  by one step:

- Given  $u_{\ell}^{(j-1)}$  and  $u_{\ell+1}^{(jt_{\ell+1})}$ , propose  $u_{\ell}^{\text{prop}}$ :

$$(29) \quad u_{\ell}^{\text{prop}} := \sqrt{1 - \beta^2} u_{\ell}^{(j-1)} + \beta \psi_{\ell},$$

where  $\psi_{\ell}|u_{\ell+1}^{(jt_{\ell+1})}$  is sampled using Algorithm 4.3.

- Accept  $u_{\ell}^{(j)} = u_{\ell}^{\text{prop}}$  with probability

$$(30) \quad \alpha_{\ell}^{ML}(u_{\ell}^{\text{prop}}|u_{\ell}^{(j-1)}) = \min \left\{ 1, \frac{\pi_{\ell}^{\text{like}}(p_{\text{obs}}|u_{\ell}^{\text{prop}})\pi_{\ell+1}^{\text{like}}(p_{\text{obs}}|u_{\ell+1}^{(j-1)t_{\ell+1}})}{\pi_{\ell}^{\text{like}}(p_{\text{obs}}|u_{\ell}^{(j-1)})\pi_{\ell+1}^{\text{like}}(p_{\text{obs}}|u_{\ell+1}^{(jt_{\ell+1})})} \right\}.$$

- Return  $u_{\ell}^{(j)}$ ,  $u_{\ell+1}^{(jt_{\ell+1})}$ ,  $Y_{\ell}^{(j)} = Q(u_{\ell}^{(j)}) - Q(u_{\ell+1}^{(jt_{\ell+1})})$

We run five independent two-level chains (as in Algorithm 6.2) using step size  $\beta^2 = 0.3$  to estimate  $\hat{Y}_{\ell}$  at levels  $\ell = 0, 1, 2$ . For each of these chains, the first 1,000 samples of  $Y_{\ell}$  are discarded as burn-in; then the following 1,000 independent samples—properly subsampled according to integrated autocorrelation time estimates—are used in our statistical approximations. Similarly we estimate  $\hat{Q}_3$  with five independent single-level chains (as in Algorithm 6.1); however on this coarsest level, we use a longer burn-in of 3,000 samples.

From the five resulting chains of  $Y_{\ell} = Q_{\ell} - Q_{\ell+1}$  (for each  $\ell = 0, 1, 2$ ), we estimate the autocorrelation as a function of lag time. In Figure 4, the top row displays these estimates for  $Q_{\ell+1}$ , which indicate the mixing of the coarse chain in the estimation of  $Y_{\ell}$ . The bottom row of Figure 4 displays these estimates for  $Y_{\ell}$ , where the faster decay in autocorrelation is the result of the coarse level being subsampled. We further note that the autocorrelation time of  $Y_{\ell}$  decays faster for smaller  $\ell$ , as we expect the acceptance rate to increase with mesh refinement as shown in Figure 5 (a).

Figure 5 (a) displays the average acceptance rate from the five chains, on the four different levels. The small error bars indicate the range of acceptance rate values from the five chains. This increase in the acceptance rate with the refinement of levels is similar to that reported in [22], numerically demonstrating that indeed our method of using multilevel stochastic PDE samplers is a computationally efficient alternative to KL-decomposition based sampling for multilevel MCMC. Figure 5 (b) provides average variance estimates for  $Q_{\ell}$  on each level, as well as average variance estimates for the correction terms  $Y_{\ell} = Q_{\ell} - Q_{\ell+1}$ . Error bars indicate the range of variance values from the five chains. The decay in the  $Y_{\ell}$  variance estimate indicates that fewer samples are required on the finer levels (relative to the coarsest) to obtain a target mean square error. This result is similar to that of [22]; however, our decay is not quite as rapid. This is likely due to the difference in problem setup ( $d = 3$ ,  $\lambda = 0.3$ ), as well as the fact that we're doing inference in a higher-dimensional space. More specifically, we have one DOF per element with 196K elements on the finest level, while the work in

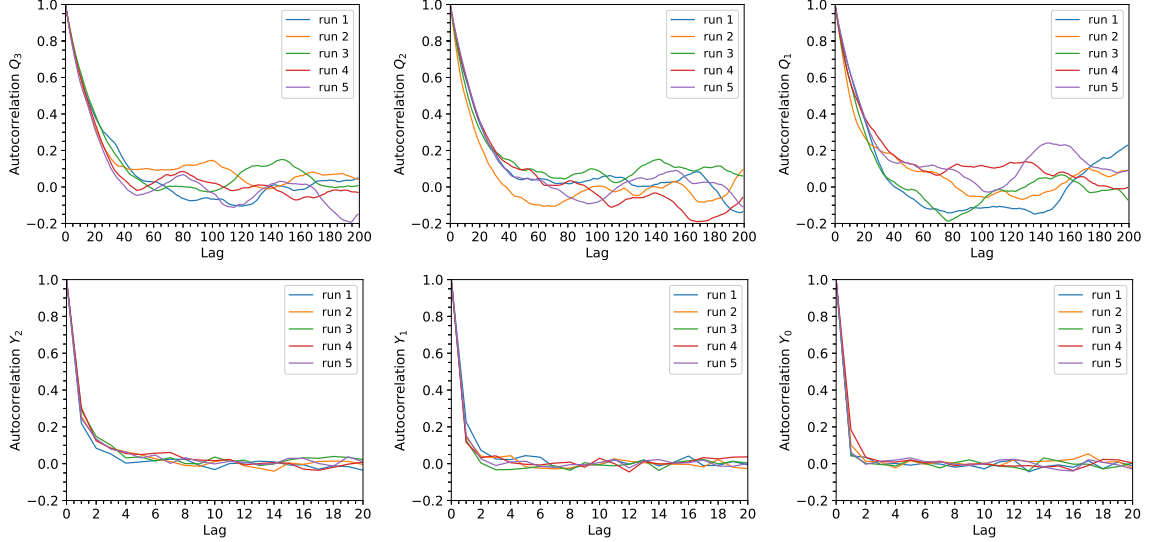


FIGURE 4. Autocorrelation estimates for increasing lag time for 5 two-level chains used to estimate  $Y_\ell = Q_\ell - Q_{\ell+1}$  at levels  $\ell = 0, 1, 2$ . Top row displays the autocorrelation for the coarse  $Q_{\ell+1}$  samples, while the bottom row shows the autocorrelation for corresponding correction chain  $Y_\ell$ .

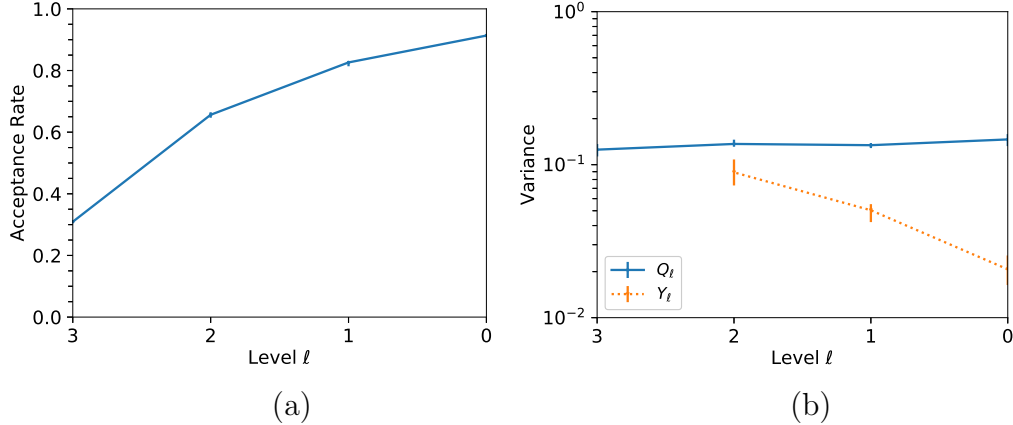


FIGURE 5. (a) Average Acceptance rate for the two-level MCMC Algorithm 6.2 as a function of the level  $\ell$ . (b) Average variance estimates for  $Q_\ell$  and  $Y_\ell$ . Averages are taken over 5 independent runs, with error bars indicating minimum and maximum values.

[22] has a KL expansion truncated at 150 DOFs on the finest level. Furthermore, we note this decay is dependent on the multilevel acceptance rates for each level. In Algorithm 6.2, upon rejection of a fine sample, the realization  $Y_\ell^{(j)} = Q(u_\ell^{(j)}) - Q(u_{\ell+1}^{(j\ell+1)})$  is calculated from unrelated realizations of  $u_\ell$  and  $u_{\ell+1}$ . This feature results in a variance that decays slower than in typical multilevel Monte Carlo (see, e.g., [28, 29, 14, 48]).

In Table 5, we provide the averaged statistical estimates derived from these five chains (on each level). In particular, we provide the estimated integrated autocorrelation times for each level, as well as the estimated mean and variance values for  $Q_\ell$  and  $Y_\ell$ . We note

TABLE 5. Multilevel MCMC average estimates from 5 chains. Values of the effective cost and sample size are calculated via (31) and (32), with  $\varepsilon = 0.01$ .

Level $\ell$	$\mathcal{C}_\ell$	$t_{\ell+1}$	$t_\ell$	$\mathcal{C}_\ell^{\text{eff}}$	$\mathbb{E}[Q_\ell]$	$\mathbb{E}[Y_\ell]$	$\mathbb{E}[ Y_\ell ]$	$\mathbb{V}[Q_\ell]$	$\mathbb{V}[Y_\ell]$	Predicted $N_\ell^{\text{eff}}$
0	494.08	44	4	12902.4	1.22	0.0054	0.0366	0.15	0.0207	773
1	62.08	36	3	1033.0	1.20	0.0046	0.0808	0.13	0.0503	4265
2	7.84	40	5	239.2	1.21	0.0169	0.1509	0.14	0.0891	11788
3	1.00	—	45	45.0	1.17	1.1709	1.1709	0.13	0.1254	32245

that, while the  $\mathbb{V}[Y_\ell]$  and  $\mathbb{E}[|Y_\ell|]$  approximations decay with mesh refinement, the  $\mathbb{E}[Y_\ell]$  approximations are not monotonic. However, this result does not conflict with the expected results, as the variance indicates error in our estimates. Using these multilevel estimates of  $\mathbb{E}[Y_\ell]$  and  $\mathbb{E}[Q_3]$  we predict a posterior mean of 1.198, with a variance (based on the number of effective samples defined in (32)) of  $5 \cdot 10^{-5}$ . With an equivalent cost, we expect the single-level estimator to have a variance of  $1.74 \cdot 10^{-4}$ .

Finally we compare the predicted cost to run this approach with that of single-level MCMC, based on the calculations in Table 5. As the scalability of solvers was investigated in [37, 21, 42], which show cost per iteration is proportional to the DOFs, and we demonstrate the number of iterations stays stable with mesh refinement, we define the cost per simulation on level  $\ell$ , denoted  $\mathcal{C}_\ell$ , to be the number of global DOFs in our linear system, normalized with respect to the coarsest level. Subsequently, the optimal effective cost per independent sample of  $Y_\ell$  is defined as

$$(31) \quad \mathcal{C}_\ell^{\text{eff}} := t_\ell(\mathcal{C}_\ell + t_{\ell+1}\mathcal{C}_{\ell+1})$$

for  $\ell = 0, \dots, L-1$ , and  $\mathcal{C}_L^{\text{eff}} = t_L\mathcal{C}_L$  for  $\ell = L$ . Then, from [22], the effective sample size (for a mean square error tolerance of  $\varepsilon^2$ ) on each level is calculated via

$$(32) \quad N_\ell^{\text{eff}} = \frac{2}{\varepsilon^2} \left( \sum_{k=0}^L \sqrt{\mathbb{V}_{\nu_k, \nu_{k+1}}[Y_k] \mathcal{C}_k^{\text{eff}}} \right) \sqrt{\frac{\mathbb{V}_{\nu_\ell, \nu_{\ell+1}}[Y_\ell]}{\mathcal{C}_\ell^{\text{eff}}}},$$

where  $\mathbb{V}_{\nu_\ell, \nu_{\ell+1}}[Y_\ell]$  is the variance of  $Y_\ell$  with respect to the joint distribution of  $u_\ell$  and  $u_{\ell+1}$ . For single-level MCMC, the number of effective samples for the target mean square error  $\varepsilon^2$  is  $N_{sl}^{\text{eff}} = 2\mathbb{V}_{\nu_0}[Q_0]/\varepsilon^2$ , and the effective cost is  $\mathcal{C}_{sl}^{\text{eff}} = t_L\mathcal{C}_0$ . Using these results, we estimate that performing our four-level MCMC is about 3.5 times faster than the single-level approach. In comparison the hierarchical four-level approach of [22] is about 5 times faster than the single-level approach. Aside from different statistical estimates, a key difference of these approaches (that impacts the cost) lies in the number of levels used to estimate each  $Y_\ell$  chain. As noted before, the approach of [22] utilizes all coarser levels in a hierarchical manner to estimate these differences. Although this component of the algorithm was not implemented in our numerical results, it is a feature that we would like to include in future work.

## 7. CONCLUSION

In this work we develop a novel, (algorithmically) scalable, hierarchical PDE-based approach to generate Gaussian random field realizations that is well-suited for multilevel MCMC on large-scale three-dimensional problems. The novelty and advantages of our two-level preconditioned Crank-Nicolson proposal in Algorithm 6.2 lies in the use of a scalable,

memory efficient stochastic PDE sampler in lieu of a computationally and memory expensive KL-decomposition in [22]. Similarly to the proposals in [22], our proposals are linear transformations of independent Gaussian vectors defined on the coarse and fine grids: the coarser-level random variables define the smooth components of the random field  $u_\ell$ , while the finer-level random variables control the high frequency components of  $u_\ell$ . However, our method uses sparse finite element interpolation operators and scalable fast PDE linear solvers to define such linear transformation, while the one in [22] uses dense matrices whose columns represents the dominant eigenvectors of the covariance method.

As our numerical result showed, Algorithm 6.2 offers comparable multilevel acceleration to that presented in [22]. First, the great majority of likelihood evaluations are done the coarse levels of the hierarchy, where evaluating the forward model is inexpensive. Second, the acceptance rate improves as the mesh is refined thus reducing the variance of the estimator  $\hat{Y}_\ell$  at finer levels. Third, the auxiliary coarse level chain allows for drastically reducing the integrated autocorrelation time  $t_\ell$  thanks to the use of independent samples from the coarse chain. As numerically illustrated in Section 4.1, our hierarchical sampler induces a multiscale decomposition of the random field  $u$ , where the finer-level proposal  $u_\ell$  shares the same smooth components of the corresponding sample from the posterior distribution  $\nu_{\ell+1}$  at the coarser-level  $\ell + 1$ . As we move to finer and finer levels we expect the likelihood function to become insensitive to the difference  $u_\ell - Pu_{\ell+1}$ , thus drastically increasing the acceptance rate. The increased mixing of the chain is then a direct consequence of the increased acceptance rate and of the independence of the coarse grid samples used in the two-level preconditioned Crank-Nicolson proposal.

The next stage of research will include investigating the overall scaling of multilevel MCMC with this new hierarchical sampler. In particular, an important—and necessary—component will be coarse grid redistribution for improved performance on all levels in the sampling hierarchy. In addition, possible future directions of this work include performing this multilevel MCMC approach with a derivative enhanced proposal, e.g., local Hessian information, as in [17], which combines the multilevel approach of [22] with dimension-independent likelihood-informed MCMC samplers of [18] to further accelerate multilevel MCMC.

#### ACKNOWLEDGEMENTS

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

#### REFERENCES

- [1] GLVis: Opengl finite element visualization tool. [glvis.org](http://glvis.org).

- [2] HYPRE: High performance preconditioners. <http://www.llnl.gov/CASC/hypre/>.
- [3] MFEM: Modular finite element methods library. [mfem.org](http://mfem.org).
- [4] ParELAG: Element-agglomeration algebraic multigrid and upscaling library, version 2.0. <http://github.com/LLNL/parelag>, 2015.
- [5] ParELAGMC: Parallel element agglomeration multilevel Monte Carlo library. <http://github.com/LLNL/parelagmc>, 2018.
- [6] R.J. Adler and J.E. Taylor. *Random fields and geometry*. Springer Science & Business Media, 2009.
- [7] R.J. Adler, J.E. Taylor, and K.J. Worsley. Applications of random fields and geometry: foundations and case studies. 2007. In preparation.
- [8] A. Barth, C. Schwab, and N. Zollinger. Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients. *Numerische Mathematik*, 119(1):123–161, 2011.
- [9] M. Bebendorf. *Hierarchical matrices*. Springer, 2008.
- [10] A. Beskos, M. Girolami, S. Lan, P.E. Farrell, and A.M. Stuart. Geometric MCMC for infinite-dimensional inverse problems. *Journal of Computational Physics*, 335:327–351, 2017.
- [11] T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler. A computational framework for infinite-dimensional Bayesian inverse problems part I: The linearized case, with application to global seismic inversion. *SIAM Journal on Scientific Computing*, 35(6):A2494–A2523, 2013.
- [12] T. Bui-Thanh and M.A. Girolami. Solving large-scale PDE-constrained Bayesian inverse problems with Riemann manifold Hamiltonian Monte Carlo. *Inverse Problems*, 30:114014, 2014.
- [13] J.A. Christen and C. Fox. Markov chain Monte Carlo using an approximation. *Journal of Computational and Graphical Statistics*, 14(4):795–810, 2005.
- [14] 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):3–15, 2011.
- [15] S.L. Cotter, G.O. Roberts, A.M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science*, pages 424–446, 2013.
- [16] M. Croci, M.B. Giles, M.E. Rognes, and P.E. Farrell. Efficient white noise sampling and coupling for multilevel Monte Carlo with nonnested meshes. *SIAM/ASA Journal on Uncertainty Quantification*, 6(4):1630–1655, 2018.
- [17] T. Cui, G. Detommaso, and R. Scheichl. Multilevel dimension-independent likelihood-informed mcmc for large-scale inverse problems. *arXiv preprint arXiv:1910.12431*, 2019.
- [18] T. Cui, K.J.H. Law, and Y.M. Marzouk. Dimension-independent likelihood-informed MCMC. *Journal of Computational Physics*, 304:109–137, 2016.
- [19] T. Cui, J. Martin, Y.M. Marzouk, A. Solonen, and A. Spantini. Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Problems*, 30(11):114015, 2014.
- [20] Y. Daon and G. Stadler. Mitigating the influence of the boundary on PDE-based covariance operators. *Inverse Problems & Imaging*, 12(5):1083–1102, 2018.
- [21] V. Dobrev, T. Kolev, C.S. Lee, V. Tomov, and P.S. Vassilevski. Algebraic hybridization and static condensation with application to scalable  $\mathbf{H}(\text{div})$  preconditioning. *SIAM Journal on Scientific Computing*, 41(3):B425–B447, 2019.
- [22] T.J. Dodwell, C. Ketelsen, R. Scheichl, and A.L. Teckentrup. A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):1075–1108, 2015.
- [23] D. Drzisga, B. Gmeiner, U. Rude, R. Scheichl, and B. Wohlmuth. Scheduling massively parallel multigrid for multilevel Monte Carlo methods. *SIAM Journal on Scientific Computing*, 39(5):S873–S897, 2017.
- [24] Y. Efendiev, T. Hou, and W. Luo. Preconditioning Markov chain Monte Carlo simulations using coarse-scale models. *SIAM Journal on Scientific Computing*, 28(2):776–803, 2006.
- [25] H.P. Flath, L.C. Wilcox, V. Akcelik, J. Hill, B. van Bloemen Waanders, and O. Ghattas. Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial hessian approximations. *SIAM Journal on Scientific Computing*, 33(1):407–432, 2011.
- [26] H. Gahvari, A.H. Baker, M. Schulz, U.M. Yang, K.E. Jordan, and W.D. Gropp. Modeling the performance of an algebraic multigrid cycle on hpc platforms. In *25th ACM International Conference on Supercomputing, ICS 2011*, pages 172–181, 2011.



- [27] R.G. Ghanem and P.D. Spanos. *Stochastic finite elements: a spectral approach*. Courier Corporation, 2003.
- [28] M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.
- [29] M.B. Giles. Multilevel Monte Carlo methods. In *Monte Carlo and Quasi-Monte Carlo Methods 2012*, pages 83–103. Springer, 2013.
- [30] I.G. Graham, F.Y. Kuo, D. Nuyens, R. Scheichl, and I.H. Sloan. Analysis of circulant embedding methods for sampling stationary random fields. *SIAM Journal on Numerical Analysis*, 56(3):1871–1895, 2018.
- [31] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. 1970.
- [32] S. Heinrich. Multilevel Monte Carlo methods. In *Large-scale scientific computing*, pages 58–67. Springer, 2001.
- [33] D. Higdon, H. Lee, and Z. Bi. A Bayesian approach to characterizing uncertainty in inverse problems using coarse and fine-scale information. *IEEE Transactions on Signal Processing*, 50(2):389–399, 2002.
- [34] V.H. Hoang, J.H. Quek, and C. Schwab. Analysis of multilevel MCMC-FEM for Bayesian inversion of log-normal diffusions. *Inverse Problems*, 2019.
- [35] J.D. Jansen, R.M. Fonseca, S. Kahrobaei, M.M. Siraj, G.M. Van Essen, and P.M.J. Van den Hof. The egg model—a geological ensemble for reservoir simulation. *Geoscience Data Journal*, 1(2):192–195, 2014.
- [36] U. Kristenko, L. Scarabosio, P. Swierczynski, E. Ullmann, and B. Wohlmuth. Analysis of boundary effects on PDE-based sampling of Whittle–Matérn random fields. *SIAM/ASA Journal on Uncertainty Quantification*, 7(3):948–974, 2019.
- [37] C.S. Lee and P.S. Vassilevski. Parallel solver for  $\mathbf{H}(\text{div})$  problems using hybridization and AMG. In *Domain Decomposition Methods in Science and Engineering XXIII*, pages 69–80. Springer, 2017.
- [38] F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.
- [39] J. Martin, L.C. Wilcox, C. Burstedde, and O. Ghattas. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing*, 34(3):A1460–A1487, 2012.
- [40] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.
- [41] S. Osborn, P.S. Vassilevski, and U. Villa. A multilevel, hierarchical sampling technique for spatially correlated random fields. *SIAM J. Sci. Comput.*, 39(5):S543–S562, 2017.
- [42] S. Osborn, P. Zulian, T. Benson, U. Villa, R. Krause, and P.S. Vassilevski. Scalable hierarchical PDE sampler for generating spatially correlated random fields using nonmatching meshes. *Numerical Linear Algebra with Applications*, 25(3):e2146, 2018.
- [43] N. Petra, J. Martin, G. Stadler, and O. Ghattas. A computational framework for infinite-dimensional Bayesian inverse problems, part II: Stochastic Newton MCMC with application to ice sheet flow inverse problems. *SIAM Journal on Scientific Computing*, 36(4):A1525–A1555, 2014.
- [44] C. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.
- [45] L. Roininen, J.M.J. Huttunen, and S. Lasanen. Whittle–Matérn priors for bayesian statistical inversion with applications in electrical impedance tomography. *Inverse Problems & Imaging*, 8(2):561, 2014.
- [46] A. Sokal. Monte Carlo methods in statistical mechanics: foundations and new algorithms. In *Functional integration*, pages 131–192. Springer, 1997.
- [47] A.M. Stuart. Inverse problems: a Bayesian perspective. *Acta numerica*, 19:451–559, 2010.
- [48] A.L. Teckentrup, R. Scheichl, M.B. Giles, and E. Ullmann. Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients. *Numerische Mathematik*, 125(3):569–600, 2013.
- [49] The Trilinos Project Team. *The Trilinos Project Website*.
- [50] P. Whittle. Stochastic processes in several dimensions. *B. Int. Statist. Inst.*, 40:974–994, 1963.
- [51] C. Williams, K.I. Christopher, and M. Seeger. Using the nyström method to speed up kernel machines. In *Advances in neural information processing systems*, pages 682–688, 2001.

## APPENDIX A. INTEGRATED AUTOCORRELATION TIME

To obtain independent samples for unbiased estimates of QoI moments from the chain  $\{Q_0^{(i)}\}_{i>0}$ , we subsample the chain according to its integrated autocorrelation time  $\tau_Q$ . In this work, we estimate  $\tau_Q$  as

$$(33) \quad \hat{\tau}_Q = 1 + 2 \sum_{\tau=1}^M \hat{\rho}_Q(\tau)$$

where the normalized autocorrelation function is estimated as

$$(34) \quad \hat{\rho}_Q(\tau) = \frac{1}{N - \tau} \sum_{i=1}^{N-\tau} \frac{(Q_0^{(i)} - \hat{\mu}_Q)(Q_0^{(i+\tau)} - \hat{\mu}_Q)}{\hat{\sigma}_Q^2},$$

with  $\hat{\mu}_Q$  and  $\hat{\sigma}_Q^2$  as the estimated mean and variance (respectively) of the data  $\{Q_0^{(i)}\}_{i=1}^N$ , and  $M \ll N$  (see [46] for more information on integrated autocorrelation time).

In practice we subsample at a rate of  $t := \lceil \hat{\tau}_Q \rceil$ . We denote  $t_\ell$  (with  $\ell < L$ ) as the estimate for the multilevel chains  $\{Y_\ell^{(i)}\}_{i>0}$ .