SIAM J. SCI. COMPUT.
Vol. 42, No. 4, pp. B1067–B1091

hp-MULTILEVEL MONTE CARLO METHODS FOR UNCERTAINTY QUANTIFICATION OF COMPRESSIBLE NAVIER–STOKES EQUATIONS*

ANDREA BECK†, JAKOB DÜRRWÄCHTER†, THOMAS KUHN†, FABIAN MEYER‡, CLAUS-DIETER MUNZ†, AND CHRISTIAN ROHDE‡

Abstract. We propose a novel *hp*-multilevel Monte Carlo method for the quantification of uncertainties in the compressible Navier–Stokes equations, using the discontinuous Galerkin method as deterministic solver. The multilevel approach exploits hierarchies of uniformly refined meshes while simultaneously increasing the polynomial degree of the ansatz space. It allows for a very large range of resolutions in the physical space and thus an efficient decrease of the statistical error. We prove that the overall complexity of the *hp*-multilevel Monte Carlo method to compute the mean field with prescribed accuracy is, in the best case, of quadratic order with respect to the accuracy. We also propose a novel and simple approach to estimate a lower confidence bound for the optimal number of samples per level, which helps to prevent overestimating these quantities. The method is in particular designed for application on queue-based computing systems, where it is desirable to compute a large number of samples during one iteration without overestimating the optimal number of samples. Our theoretical results are verified by numerical experiments for the two-dimensional compressible Navier–Stokes equations. In particular we consider a cavity flow problem from computational acoustics, demonstrating that the method is suitable to handle complex engineering problems.

Key words. uncertainty quantification, multilevel Monte Carlo, discontinuous Galerkin, random Navier–Stokes equations

AMS subject classifications. 35R60, 65C05, 65M60

DOI. 10.1137/18M1210575

1. Introduction.

Due to the continuous improvement of computer-processing capacities, the demand for highly accurate numerical simulations which also account for uncertain input parameters is growing. Uncertainties might arise from limitations in measuring physical phenomena exactly or from a systematical absence of knowledge about the underlying physical processes. Uncertainty quantification (UQ) addresses this issue and provides mathematical methods to quantify the influence of uncertain input parameters on the numerical solution itself or on derived quantities of interest. There exist two major approaches for UQ. On the one hand, nonstatistical approaches like the intrusive and nonintrusive polynomial chaos expansion approximate the underlying random field by a series of polynomials and derive deterministic models for the stochastic modes. On the other hand, statistical approaches such as Monte Carlo (MC)-type methods sample the random space to obtain statistical information, like mean, variance, or higher-order moments of the corresponding random field. Especially MC-type methods are very popular as they are easy to implement

*Submitted to the journal's Computational Methods in Science and Engineering section August 30, 2018; accepted for publication (in revised form) May 29, 2020; published electronically August 13, 2020.

https://doi.org/10.1137/18M1210575

Funding: This work was supported by the Baden-Württemberg Stiftung via the project "BW-HPC2: SEAL" and by the High-Performance Computing Center Stuttgart (HLRS).

†Institute of Aerodynamics and Gas Dynamics, University of Stuttgart, Pfaffenwaldring 21, 70569 Stuttgart, Germany (beck@iag.uni-stuttgart.de, jakob.duerrwaechter@iag.uni-stuttgart.de, thomas.kuhn@iag.uni-stuttgart.de, munz@iag.uni-stuttgart.de).

‡Institute of Applied Analysis and Numerical Simulation, University of Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany (fabian.meyer@mathematik.uni-stuttgart.de, crohde@mathematik.uni-stuttgart.de).

and only require a deterministic black box solver. Moreover, in contrast to nonstatistical approaches, the MC method does not rely on the regularity of the underlying random field, rendering it a very robust method. However, the convergence of MC methods is dictated by the law of large numbers, hence relatively slow and therefore computationally expensive.

To overcome these difficulties Heinrich [15] and later Giles [12] extended the MC method to the multilevel MC (MLMC) method, where they considered different mesh hierarchies instead of one fixed mesh to discretize the deterministic equation of interest. The MLMC method relies on the idea that the global behavior of the exact expectation can be approximated by the behavior of the expectation of numerical solutions with a low spatial resolution, which can be computed at low cost. The coarse expectation is then subsequently corrected by computations on finer meshes, which are computationally more expensive per sample. The number of these simulations at full resolution is significantly reduced compared to the original MC method resulting in a considerably lower overall computational cost. Since its development the MLMC method has been very successfully applied to UQ for many different partial differential equations with uncertainties, as, for example, in [3, 4, 5, 8, 9, 17, 22, 23]. We refer to [7, 25] for applications of the MLMC method in computational fluid dynamics, especially to aerodynamics and meteorology. A generalization of the MLMC method, named multiindex MC (MIMC) method, has been presented in [13]. In contrast to the MLMC method, which computes mean and variance using first-order differences, the MIMC method uses high-order differences which allows for a faster decay of the corresponding level variances, resulting in significant efficiency gains compared to the standard MLMC method.

In [23] the authors extended the MLMC method for hyperbolic problems to a multiorder method, using an energy-preserving discontinuous Galerkin (DG) scheme for the elastic wave equation, which we will dub p-MLMC for the remainder of this article. The authors considered either mesh refinements (h-refinement) or increasing the DG polynomial degree (p-refinement) to obtain a hierarchy of different levels. Furthermore, they proved that the computational complexity to reach a prescribed accuracy is of quadratic order with respect to the accuracy. However, a proof for the computational complexity for a hierarchy of hp-refined meshes is still open. We therefore extend the h- and p-MLMC method to an hp-MLMC method, where we uniformly refine the physical mesh and at the same time uniformly increase the DG polynomial degree. As the first new contribution of this work, we extend the complexity analysis from [8, 23] to arbitrarily hp-refined meshes and show that the hp-MLMC method is—up to a constant—as efficient as the h-MLMC and the p-MLMC method. The hp mesh hierarchy enables us to cover a very large range of resolution levels, which is crucial for the efficiency of the hp-MLMC method. Furthermore, from a numerical point of view, a low polynomial degree (resulting in a more dissipative numerical scheme) might be favorable in connection with coarse meshes, where an insufficient resolution can otherwise lead to unphysical oscillations, whereas a high polynomial degree yields higher accuracy when fine meshes are employed. The hp-MLMC method becomes an attractive alternative for mesh-based MLMC methods if uniform mesh hierarchies are not available. This might occur for complex domain geometries or when using dynamical mesh adaption. In view of the recent rise of higher-order schemes it seems likely to exploit a mixture method. In contrast to h-refinement, hp-refinement introduces more degrees of freedom in designing a suitable mesh hierarchy.

For complex fluid dynamical problems, e.g., direct numerical simulations of unsteady, compressible Navier–Stokes equations, it is inevitable to use large-scale com-

puting systems. These systems are in most cases equipped with a queuing system. When using the *hp*-MLMC method on queue-based, large-scale computing systems, due to long queuing times, it is desirable to compute as many samples as possible per iteration. On the other hand, to avoid unnecessary computations and to increase the efficiency of the *hp*-MLMC method, one is interested to not overestimate the optimal number of samples per iteration. Therefore, as the second novel contribution, we show how to construct easily a robust lower confidence bound for the number of samples per level, which avoids overshooting the optimal number of samples but still yields a reasonably large number of additional samples per iteration. To demonstrate the efficiency of the *hp*-MLMC method combined with the novel sample estimator we apply our method to two different compressible flow problems, a benchmark problem with smooth solution and an open cavity flow problem. The latter is an important problem in computational acoustics that exhibits physical phenomena with high sensitivity with respect to the problem parameters [20]. Moreover, we provide for both problems a thorough comparison of the *h*-, *p*-, and *hp*-MLMC methods.

This article is structured as follows. In section 2 we introduce the necessary mathematical framework and briefly introduce the DG method. In section 3 we describe the *hp*-MLMC method and prove the stated complexity result. We also discuss the necessity of confidence intervals for the estimate of the optimal number of samples when working on queue-based large-scale computing systems. Finally, in section 4 we apply our method to two different examples and verify our theoretical results.

2. Notation and preliminaries.

2.1. A Primer on probability theory. We let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where Ω is the set of all elementary events $\omega \in \Omega$, $\mathcal{F}$ is a σ-algebra on Ω, and $\mathbb{P}$ is a probability measure. We further consider a second measurable space $(E, \mathcal{B}(E))$, where E is a Banach space and $\mathcal{B}(E)$ is the corresponding Borel σ-algebra. An E-valued random field is any mapping $X : \Omega \to E$ such that $\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$ holds for any $B \in \mathcal{B}(E)$. For $r \in [1, \infty) \cup \{\infty\}$ we consider the Bochner space $L^r(\Omega; E)$ of r-summable E-valued random variables X equipped with the norm

$$\|X\|_{L^r(\Omega;E)} := \begin{cases} \left(\int_\Omega \|X(\omega)\|_E^r \quad \mathrm{d}\mathbb{P}(\omega)\right)^{1/r} = \mathbb{E}\big[\|X\|_E^r\big]^{1/r}, & 1 \le r < \infty, \\ \operatorname{ess\,sup}_{\omega\in\Omega} \|X(\omega)\|_E, & r = \infty. \end{cases}$$

The uncertainty is introduced via a random vector $\xi(\omega) = \big(\xi_1(\omega), \dots, \xi_N(\omega)\big) : \Omega \to \Xi \subset \mathbb{R}^N$ with independent, absolutely continuous random variables as components. This means that for each random variable ξ_i there exists a density function $f_{\xi_i} : \mathbb{R} \to \mathbb{R}_+$ such that $\int_{\mathbb{R}} f_{\xi_i}(y)\, \mathrm{d}y = 1$ and $\mathbb{P}[\xi_i \in A] = \int_A f_{\xi_i}(y)\, \mathrm{d}y$ for any $A \in \mathcal{B}(\mathbb{R})$ and for all $i = 1, \dots, N$. Moreover, the joint density function f_ξ of the random vector $\xi = (\xi_1, \dots, \xi_N)$ can be written as $f_\xi(y) = \prod_{i=1}^N f_{\xi_i}(y_i)$ for all $y = (y_1, \dots, y_N)^\top \in \Xi$. The random vector induces a probability measure $\tilde{\mathbb{P}}(B) := \mathbb{P}(\xi^{-1}(B))$ for all $B \in \mathcal{B}(\Xi)$ on the measurable space $(\Xi, \mathcal{B}(\Xi))$. This measure is called the law of ξ, and in the following we work on the image probability space $(\Xi, \mathcal{B}(\Xi), \tilde{\mathbb{P}})$.

2.2. The random Navier–Stokes equations. As the physical space we consider a bounded domain $D \subset \mathbb{R}^2$, and we further define the space-time-stochastic domain $D_{T,\Xi} := (0, T) \times D \times \Xi$. We focus on the random compressible Navier–Stokes equations given by

$$U_t + \nabla_x \cdot (G(U) - H(U, \nabla_x U)) = 0 \qquad \forall\ (t, x, y) \in D_{T,\Xi}, \tag{2.1}$$

where $U(t,x,y)$ denotes the solution vector of the conserved quantities, i.e., we have $U = (\rho, \rho v_1, \rho v_2, \rho e)^\top$. G and H denote the advective and viscous fluxes, i.e.,

$$G_i(U) = \begin{pmatrix} \rho\, v_i \\ \rho\, v_1 v_i + \delta_{1i}\, p \\ \rho\, v_2 v_i + \delta_{2i}\, p \\ \rho\, e v_i + p\, v_i \end{pmatrix}, \quad H_i(U, \nabla_x U) = \begin{pmatrix} 0 \\ \tau_{1i} \\ \tau_{2i} \\ \tau_{ij} v_j - q_i \end{pmatrix}, \quad i = 1,2. \tag{2.2}$$

Here, δ_{ij} is the Kronecker delta function, and the physical quantities ρ, $v = (v_1, v_2)^\top$, p, and e represent density, the velocity vector, the pressure, and the specific total energy, respectively. With Stokes' and Fourier's hypothesis, the viscous stress tensor τ and the heat flux q reduce to

$$\tau = \mu \left(\nabla v + (\nabla v)^\top - \frac{2}{3} (\nabla \cdot v) I \right), \quad q = -k \nabla \mathcal{T}. \tag{2.3}$$

In (2.3) μ is the dynamic viscosity, k the thermal conductivity, and $\mathcal{T}$ the local temperature. In order to solve for the unknowns, the system has to be closed by appropriate equations of state. We choose for the gas constant R, the adiabatic exponent κ, and the specific heat at constant volume c_v the perfect gas law assumptions

$$p = \rho R \mathcal{T} = (\kappa - 1)\rho \left(e - \frac{1}{2} v \cdot v \right), \qquad e = \frac{1}{2} v \cdot v + c_v \mathcal{T}. \tag{2.4}$$

We augment (2.1) with suitable boundary and initial conditions, denoted by

$$B(U) = g \quad \forall\ (x,y) \in \partial D \times \Xi, \quad U(0,x,y) = U^0(x,y) \quad \forall\ (x,y) \in D \times \Xi.$$

The boundary operator B, the boundary data g, and the initial condition U^0 will be specified when we detail the settings for the numerical experiments in section 4.

Following [21] we call $U \in L^2(\Xi; C^1([0,T]; L^2(D)))$ a weak random solution of (2.1) if it is a weak solution $\tilde{\mathbb{P}}$-a.s. $y \in \Xi$ and a measurable mapping $(\Xi, \mathcal{B}(\Xi)) \ni y \to U(\cdot,\cdot,y) \in (C^1([0,T]; L^2(D)), \mathcal{B}(C^1([0,T]; L^2(D))))$.

2.3. The Runge–Kutta DG method. We shortly recall the DG spatial discretization for the initial-boundary value problem (2.1); see [16] for more details. To partition the spatial domain we subdivide D into $N \in \mathbb{N}$ quadrilateral elements D_m, $m = 1, \dots, N$ with $D = \bigcup_{m=1}^N D_m$ and define the mesh size $h := \max_{m=1,\dots,N} h_m$, where h_m is the length of D_m. Moreover, we define $h_{min} := \min_{m=1,\dots,N} h_m$. Furthermore, let us introduce the space of piecewise DG polynomial ansatz and test functions: $\mathcal{V}_h^q := \{U : D \to \mathbb{R}^4 \mid U\big|_{D_m} \in \mathbb{P}_q(D_m; \mathbb{R}^4) \text{ for } 1 \le m \le N\}$, where $\mathbb{P}_q(D_m; \mathbb{R}^4)$ is the space of polynomials of degree q on the element D_m and $U\big|_{D_m}$ denotes U restricted to D_m. The DG solution U_h is then sought in $\mathcal{V}_h^q$, i.e., $U_h(t,\cdot,y) \in \mathcal{V}_h^q$, a.e. $(t,y) \in (0,T) \times \Xi$. On each element D_m, $m = 1, \dots, N$, we use tensor products of local one-dimensional Lagrange interpolation polynomials of degree q, i.e.,

$$U_h(t,x,y)\big|_{D_m} = \sum_{i,j=0}^{q} U_{i,j}^m(t,y) l_i^m(x_1) l_j^m(x_2). \tag{2.5}$$

The interpolation nodes are chosen to be the Gauss–Legendre nodes; cf. [16]. We then consider the (spatial) weak form of (2.1) given by

$$\frac{\partial}{\partial t}\int_D U(t,\cdot,y)\Phi \ \mathrm{d}x + \oint_{\partial D}(\mathcal{G}(U(t,\cdot,y)) - \mathcal{H}(U(t,\cdot,y),\nabla_x U(t,\cdot,y)))\Phi \ \mathrm{d}s$$
$$- \int_D (G(U(t,\cdot,y)) - H(U(t,\cdot,y),\nabla_x U(t,\cdot,y)))\cdot\nabla_x\Phi \ \mathrm{d}x = 0 \tag{2.6}$$

for a.e. $t \in (0,T)$ and $\tilde{\mathbb{P}}$-a.s. $y \in \Xi$ and for all test functions Φ. Now, using the same discrete space $\mathcal{V}_h^q$ for ansatz and test functions in (2.6), we obtain the semidiscrete DG scheme for $U_h \in L^2(\Xi; C^1([0,T);\mathcal{V}_h^q))$:

$$\frac{\partial}{\partial t}\int_D U_h(t,\cdot,y)\Phi_h \ \mathrm{d}x + \oint_{\partial D}\mathcal{G}_n^*(U_h^-(t,\cdot,y),U_h^+(t,\cdot,y))\Phi_h \ \mathrm{d}s$$
$$+ \oint_{\partial D}\mathcal{H}_n^*(U_h(t,\cdot,y),\nabla_x U_h(t,\cdot,y))\Phi_h \ \mathrm{d}s - \int_D G(U_h(t,\cdot,y))\cdot\nabla_x\Phi_h \ \mathrm{d}x$$
$$+ \int_D H(U_h(t,\cdot,y),\nabla_x U_h(t,\cdot,y))\cdot\nabla_x\Phi_h \ \mathrm{d}x = 0 \tag{2.7}$$

for all $\Phi_h \in \mathcal{V}_h^q$. Here, $\mathcal{G}_n^*(U^-,U^+)$ denotes a numerical flux, which depends on values at the grid cell interface from neighboring cells. In this paper, we have chosen the approximate Roe Riemann solver with entropy fix described in [14]. The viscous fluxes $\mathcal{H}_n^*$ normal to the cell interfaces are approximated by the procedure described by Bassi and Rebay in [6]. The DG scheme (2.7) is then advanced in time by a $(q+1)$th order Runge–Kutta method [18] constrained by a CFL-type condition of the form

$$\Delta t \le \min\left\{\frac{h_{min}}{\lambda_{max}^c(2q+1)}, \left(\frac{h_{min}}{\lambda_{max}^v(2q+1)}\right)^2\right\}. \tag{2.8}$$

In (2.8) $\lambda_{max}^c := ((|v_1|+c)+(|v_2|+c))$ is an estimate for the absolute value of the largest eigenvalue of the convective flux Jacobian with $c := \sqrt{\kappa\frac{p}{\rho}}$ being the speed of sound. Moreover, $\lambda_{max}^v := (\max(\frac{4}{3\rho},\frac{\kappa}{p})\frac{\mu}{Pr})$ is an estimate for the largest eigenvalue of the diffusion matrix of the viscous flux, $Pr = \frac{c_p\mu}{k}$ being the Prandtl number. With this choice the consistency error of the numerical scheme is formally of order $\mathcal{O}(h^{q+1}+\Delta t^{q+1})$. We note that we indicate the numerical solution in the remaining part of this paper by the spatial parameter h only.

3. The *hp*-MLMC method.

3.1. Description of the *hp*-MLMC method. In this section we introduce the *hp*-MLMC method, based on the classical MLMC method from [12]. For levels $l = 0,\dots,L$, we consider spatial meshes with $N_l \in \mathbb{N}$ elements and ansatz spaces of polynomial degree $q_l \in \mathbb{N}$. We choose the number of elements N_l and the DG polynomial degrees q_l such that $N_0 < \cdots < N_L$ and $q_0 < \cdots < q_L$ holds, i.e., we simultaneously increase the mesh size and the DG polynomial degree. With $\mathcal{V}_{h_l}^{q_l}$ we denote the DG polynomial space corresponding to level $l = 0,\dots,L$. Moreover, by $U_l(t,\cdot,y) := U_{h_l}(t,\cdot,y) \in \mathcal{V}_{h_l}^{q_l}$ (a.e. $(t,y) \in (0,T)\times\Xi$) we denote the DG numerical solution associated with level $l \in \{0,\dots,L\}$. Additionally, the deterministic numerical solution of (2.7) on level $l = 0,\dots,L$ for input parameter $y_i \in \Xi$ is denoted by $U_l^i := U_l(\cdot,\cdot,y_i)$ and will be called sample for the remaining part of this paper. Since we do not enforce a global CFL time-step restriction across different discretization levels, each sample has to obey the CFL condition (2.8). Thus, coarser levels admit a bigger time-step than fine levels. We want to emphasize that, in contrast to the MIMC method from [13] we index mesh size and polynomial degree by a single parameter.

It is our goal to compute statistical moments like expected value or higher-order moments of a general quantity of interest (QoI) $Q(U)$ of the random weak solution U of (2.1). Precisely, we are interested to determine

$$\mathbb{E}\big[Q(U(t,x,y))\big] = \int_{\Xi} Q(U(t,x,y)) f_\xi(y)\ \mathrm{d}y \tag{3.1}$$

for a.e. $(t,x) \in (0,T)\times D$. Here Q can be an arbitrary nonlinear function or functional of U. To ease notation we suppress the dependence of the QoI on (t,x,y) and write $Q(U)$. We approximate (3.1) with an MC estimator. To this end, we let $\{U_L^i\}_{i=1}^M$ be $M \in \mathbb{N}$ independent, identically distributed samples. The MC estimator for (3.1) is then defined by

$$\mathrm{E}_{\mathrm{MC}}^M\big[Q(U)\big] := \frac{1}{M}\sum_{i=1}^M Q(U_L^i) \approx \mathbb{E}\big[Q(U)\big]. \tag{3.2}$$

Next we advance the MC estimator $\mathrm{E}_{\mathrm{MC}}^M\big[\,\cdot\,\big]$ to the *hp*-MLMC estimator $\mathrm{E}_{hp}^L\big[\,\cdot\,\big]$ by using the linearity of the expectation in combination with a telescoping sum. We then write (see [12])

$$\mathbb{E}\big[Q(U_L)\big] = \sum_{l=0}^L \mathbb{E}\big[Q(U_l) - Q(U_{l-1})\big], \tag{3.3}$$

where we used the definition $Q(U_{-1}) = 0$. Now, each term in (3.3) can be estimated by the MC estimator (3.2). If we let $M_l \in \mathbb{N}$ denote a level-dependent number of samples for each level $l = 0,\dots,L$ and assume that the samples $\{Q(U_l^i)\}_{i=1}^{M_l}$, $l = 0,\dots,L$, on different levels are independent from each other, we obtain the *hp*-MLMC estimator via

$$\begin{aligned}\mathrm{E}_{hp}^L\big[Q(U_L)\big] &:= \sum_{l=0}^L \frac{1}{M_l}\sum_{i=1}^{M_l}(Q(U_l^i) - Q(U_{l-1}^i)) = \sum_{l=0}^L \mathrm{E}_{\mathrm{MC}}^{M_l}\big[Q(U_l) - Q(U_{l-1})\big] \\ &\approx \sum_{l=0}^L \mathbb{E}\big[Q(U_l) - Q(U_{l-1})\big] = \mathbb{E}\big[Q(U_L)\big].\end{aligned}$$

Here, $Q(U_l^i)$ and $Q(U_{l-1}^i)$, are computed using the same sample $y_i^l \in \Xi$. The main idea of the MLMC estimator is that the global behavior of the exact expectation can be approximated by the behavior of the expectation of numerical solutions with low resolution, where each sample can be computed with low cost. Thus, M_l is supposed to be large for coarse levels. The coarse-level expectation is then successively corrected by a few computations on finer levels. Each fine-level sample is computationally expensive, and therefore, M_l is supposed to be small on fine levels. Hence, the most important aspect for the efficiency of the *hp*-MLMC estimator is the correct choice of M_l. In the following section we want to derive the best choice for M_l such that the total work is minimized under the constraint that the spatial and stochastic error satisfy a certain threshold.

3.2. Optimal number of samples. For the following analysis we set the QoI to be the solution itself at a fixed point $t \in [0,T]$ in time, i.e.,

$$Q(U) = U(t,\cdot,\cdot). \tag{3.4}$$

We note that our analysis can be analogously performed using any other QoI, including functional ones and suitable norms for Q.

Remark 3.1. We consider the QoI (3.4) because we are mainly interested in statistical quantities of the solution U itself. However, many other QoIs have a higher regularity than the solution U, especially if Q is a functional. Therefore, using such a QoI yields a faster decrease of the variance across different levels, which increases the performance of the MLMC method compared to MC. In [1, Figure 10] it has been numerically shown that for an uncertain Kelvin-Helmholtz problem the level variance of the solution does not decrease, because upon each mesh refinement additional smaller-scale structures are detected. Thus, the MLMC method provides no computational gains compared to the MC method when the QoI is the solution itself.

With the help of the following representation of the root mean square error (RMSE) we derive an optimal number of samples M_l for all $l = 0, \dots, L$:

$$\begin{aligned} \text{RMSE} &:= \|\mathbb{E}\big[U(t,\cdot,\cdot)\big] - \mathrm{E}_{hp}^{L}\big[U_L(t,\cdot,\cdot)\big]\|_{L^2(\Xi;L^2(D))} \le \epsilon_{\text{det}} + \epsilon_{\text{stat}}, \\ \epsilon_{\text{det}} &:= \|\mathbb{E}\big[U(t,\cdot,\cdot)\big] - \mathbb{E}\big[U_L(t,\cdot,\cdot)\big]\|_{L^2(D)}, \\ \epsilon_{\text{stat}} &:= \|\mathbb{E}\big[U_L(t,\cdot,\cdot)\big] - \mathrm{E}_{hp}^{L}\big[U_L(t,\cdot,\cdot)\big]\|_{L^2(\Xi;L^2(D))}. \end{aligned} \tag{3.5}$$

The term ϵ_{det} in (3.5) is the deterministic approximation error (bias). It accounts for the insufficient resolution of the deterministic system. The term ϵ_{stat} corresponds to the statistical (sampling) error. Its occurs due to the finite number of samples in (3.2). We choose the optimal number of samples M_l to minimize this term. For notational convenience we suppress the explicit dependence on $t \in [0,T]$. Using the independence of the samples, we rewrite the statistical error in (3.5) (cf. [24]):

$$\begin{aligned} \epsilon_{\text{stat}}^2 &= \mathbb{E}\left[\|\mathbb{E}\big[U_L\big] - \mathrm{E}_{hp}^{L}\big[U_L\big]\|_{L^2(D)}^2\right] \\ &= \mathbb{E}\left[\left\|\sum_{l=0}^{L} \mathbb{E}\big[U_l - U_{l-1}\big] - \mathrm{E}_{\text{MC}}^{M_l}\big[U_l - U_{l-1}\big]\right\|_{L^2(D)}^2\right] \\ &= \sum_{l=0}^{L} \frac{1}{M_l^2} \sum_{i=1}^{M_l} \mathbb{E}\left[\|\mathbb{E}\big[U_l - U_{l-1}\big] - (U_l^i - U_{l-1}^i)\|_{L^2(D)}^2\right] \\ &= \sum_{l=0}^{L} \frac{1}{M_l} \mathbb{E}\left[\|\mathbb{E}\big[U_l - U_{l-1}\big] - (U_l - U_{l-1})\|_{L^2(D)}^2\right] \\ &=: \sum_{l=0}^{L} \frac{\sigma_l^2}{M_l}. \end{aligned} \tag{3.6}$$

Remark 3.2. We use the $L^2(D)$-norm instead of the $L^1(D)$-norm for two reasons. First, the $L^1(D)$-norm is appropriate for inviscid flow problems in one spatial dimension, but we are interested in regular solutions of (2.1). Second, using the $L^2(D)$-norm has the side effect that (3.6) is satisfied as equality (cf. [24, section 5.2]).

From this representation, the optimal number of samples can be obtained by an error-complexity analysis as in [12, 23, 24]. We introduce the total work

$$\text{W}_{\text{tot}} := \text{W}_{\text{tot}}(M_0, \dots, M_L) := \sum_{l=0}^{L} M_l w_l, \tag{3.7}$$

where w_l is the work needed to create one sample $U_l^i - U_{l-1}^i$. Following [12] we obtain the optimal number of samples on different levels by considering the following minimization problem.

(3.8) For a tolerance $\epsilon > 0$, $\underset{M_0,\dots,M_L\in\mathbb{N}}{\text{minimize}}\ \mathrm{W}_{\text{tot}}$ under the constraint $\displaystyle\sum_{l=0}^{L} \frac{\sigma_l^2}{M_l} \leq \frac{1}{4}\epsilon^2$.

The minimization problem can be explicitly solved by (cf. [12, 23])

$$M_l = \left\lceil \left(\frac{\epsilon}{2}\right)^{-2} \sqrt{\frac{\sigma_l^2}{w_l}} \sum_{k=0}^{L} \sqrt{\sigma_k^2 w_k} \right\rceil . \tag{3.9}$$

In this work we consider an iterative version of the *hp*-MLMC method, which means that we initialize the algorithm with a number of warm-up samples and then estimate each quantity in (3.9) during each iteration of the *hp*-MLMC method. In order to keep track of the number of samples that have already been calculated during the iterations of the algorithm we denote this quantity by M_{tot_l} for each level $l \in \{0,\dots,L\}$. We want to emphasize that M_{tot_l} is not equal to M_l, which is the estimated optimal number of samples from (3.9).

The level variances $\sigma_0^2,\dots\sigma_L^2$ in (3.6) are not known in general, and we therefore estimate the level variance using the unbiased estimator as in [24]:

$$\hat{\sigma}_l^2 := \frac{1}{M_{\text{tot}_l}-1} \sum_{j=1}^{M_{\text{tot}_l}} \int_D \left(\left(\frac{1}{M_{\text{tot}_l}} \sum_{i=1}^{M_{\text{tot}_l}} (U_l^i - U_{l-1}^i) \right) - \left(U_l^j - U_{l-1}^j \right) \right)^2 dx. \tag{3.10}$$

The work required for the simulation of one sample can vary with an uncertain parameter (e.g., when uncertain viscosity influences the time-step restriction). Moreover, on high performance computing systems, random variations in work can occur between two executions of the same simulation. In order to account for this uncertainty, we estimate the work w_l on level $l = 0,\dots,L$ by the average work per sample U_l^i, denoted by w_l^i, and define the average work by

$$\hat{w}_l := \frac{1}{M_{\text{tot}_l}} \sum_{i=1}^{M_{\text{tot}_l}} w_l^i. \tag{3.11}$$

As a matter of fact, M_l from (3.9) is also only estimated, and we denote the estimator for M_l by $\hat{M}_l$. We have now all ingredients together to state in Algorithm *hp*-MLMC the classical MLMC algorithm proposed by Giles [12]. Based on this algorithm we want to discuss several important aspects of the *hp*-MLMC method. First, the complexity of the algorithm will be analyzed in Theorem 3.4. The choice of the maximum level L will be considered in Remark 3.8. The discussion of the number of warm-up samples $K_0,\dots,K_L$ (line two in the algorithm), respectively, the additional samples (lines six and seven in the algorithm) will be postponed to section 3.4, where we derive lower confidence bounds for the optimal number of samples M_l.

3.3. Computational complexity of the *hp*-MLMC method. In what follows we use the notation $\tilde{q}_l = q_l + 1$ for $l \in \mathbb{N}$.

For the analysis of the computational complexity of Algorithm *hp*-MLMC in Theorem 3.4 below we impose the following assumptions.

- (A1) Asymptotic work: there exist $\gamma_1, c_1 > 0$ (independent of h_l, q_l): $w_l \leq c_1(h_l^{-1}\tilde{q}_l)^{\gamma_1}$ for all $l \in \mathbb{N}$.
- (A2) Bias reduction: there exist $\kappa_1, c_2 > 0$ (independent of h_l, q_l): $\|\mathbb{E}\big[U\big] - \mathbb{E}\big[U_l\big]\|_{L^2(D)} \leq c_2 h_l^{\kappa_1 \tilde{q}_l}$ for all $l \in \mathbb{N}$.

Algorithm *hp*-MLMC

Fix a tolerance $\epsilon > 0$, the maximum level $L \in \mathbb{N}$ and set $\mathcal{L} := \{0, \dots, L\}$
Compute K_l (warm-up) samples on level $l = 0, \dots, L$ and set $M_{\mathrm{tot}_l} := K_l$
while $\mathcal{L} \neq \emptyset$ **do**
  for $l \in \mathcal{L}$ **do**
    Estimate w_l by (3.11), σ_l^2 by (3.10), and then M_l by (3.9)
    if $M_l > M_{\mathrm{tot}_l}$ **then**
      Add $(M_l - M_{\mathrm{tot}_l})$ new samples of $U_l^i - U_{l-1}^i$ and update M_{tot_l}
    else
      Set $\mathcal{L} := \mathcal{L} \setminus \{l\}$
    end if
  end for
end while
Compute $\mathrm{E}_{hp}^L[U_L]$

(A3) Variance reduction between two levels: there exists $c_3 > 0$ (independent of h_l, q_l): $\sigma_l^2 \leq c_3 h_l^{\kappa_2 \tilde{q}_l}$ for some $\kappa_2 > 0$ with $\kappa_1 \tilde{q}_0 \geq \kappa_2 \tilde{q}_0/2$ and for all $l \in \mathbb{N}$.

Remark 3.3.

1. Assumption (A1) provides a bound for the computational work in terms of the complete number of degrees of freedom on single levels. For the Runge–Kutta DG method in $d = 2$ spatial dimensions we have $h_l^{-2}\tilde{q}_l^2$ spatial degrees of freedom. This has to be multiplied with the number of time-steps, which is proportional to $h_l^{-1}\tilde{q}_l$, or $h_l^{-2}\tilde{q}_l^2$, depending on the minimum in the CFL condition (2.8). Thus, the total computational work asymptotically equals $\mathcal{O}(h_l^{-3}\tilde{q}_l^3)$, respectively, $\mathcal{O}(h_l^{-4}\tilde{q}_l^4)$. Therefore, we expect the parameter γ_1 to satisfy $\gamma_1 = 3$, or $\gamma_1 = 4$.
2. In assumption (A2) it is stated that the bias, i.e., the deterministic approximation error, converges with the order of the DG method. For smooth solutions the order of convergence is $\mathcal{O}(h_l^{q_l+1}) = \mathcal{O}(h_l^{\tilde{q}_l})$; cf. [26] and the numerical experiments in [16]. Therefore, we expect $\kappa_1 = 1$.
3. In assumption (A3) we require that the variance decays on all levels similar to the bias term. If we consider regular solutions of a random differential equation, we expect $\kappa_2 = 2$; cf. the discussion in [25, page 25].

In Theorem 3.4 below we present a complexity bound for the *hp*-MLMC method. This result generalizes [8, Theorem 1] on *h*-refined and [23, Theorem 3] on *p*-refined meshes to *hp*-refined mesh hierarchies.

We distinguish between the two different cases $\kappa_2\tilde{q}_0 > \gamma_1$ and $\kappa_2\tilde{q}_0 < \gamma_1$. For the second case $\kappa_2\tilde{q}_0 < \gamma_1$ we need to define a critical level $L^* \in \mathbb{N}$ such that $\kappa_2\tilde{q}_{L^*} < \gamma_1 \leq \kappa_2\tilde{q}_{L^*+1}$.

THEOREM 3.4 (complexity of the *hp*-MLMC method). *For $\beta \in \mathbb{N}$ and $q_0 \geq 0$ let $\{q_l := q_0 + \beta l\}_{l\in\mathbb{N}}$ be a sequence of DG polynomial degrees. Additionally, we consider a family of meshes with associated mesh size $h_l = \lambda^{-l}h_0$ for some $h_0 \in (0, 1)$ and $\lambda \geq 2$. Let $\mathcal{V}_{h_l}^{q_l}$ be the corresponding DG spaces.*

Under the assumptions (A1)–(A3)*, there exists a constant $c > 0$ such that for any tolerance $0 < \epsilon < \min\left(1, 2c_2 h_0^{\kappa_1\tilde{q}_0}\right)$, there exists a maximum level $L = L(\epsilon) \in \mathbb{N}$, $L(\epsilon) \geq 2$, and a number of samples M_l on each level $l \in \{0, \dots, L(\epsilon)\}$ such that the* RMSE *from* (3.5) *satisfies* RMSE $\leq \epsilon$ *with the computational complexity*

$$W_{\mathrm{tot}} \leq \begin{cases} c\epsilon^{-2}, & \kappa_2\tilde{q}_0 > \gamma_1, \\ c\epsilon^{-2-\frac{\gamma_1-\kappa_2\tilde{q}_0}{\min\{\kappa_1\tilde{q}_L,\kappa_1\tilde{q}_{L^*}\}}}, & \kappa_2\tilde{q}_0 < \gamma_1. \end{cases} \tag{3.12}$$

Remark 3.5.

1. The complexity bound for the case $\kappa_2\tilde{q}_0 < \gamma_1$ in Theorem 3.4 suggests that there exists a threshold for the asymptotic complexity scaling like $\mathcal{O}(\epsilon^{-2-\frac{\gamma_1-\kappa_2\tilde{q}_0}{\kappa_1\tilde{q}_{L^*}}})$. In other words, increasing the number of levels L beyond the critical level L^* does not pay off.
2. Theorem 3.4 does not include the borderline case $\kappa_2\tilde{q}_0 = \gamma_1$. Following the analysis in [8] one can make a saturated choice for all sample numbers M_l scaling with the maximum level $L(\epsilon)$. Then a slight variation of Case A in the proof of Theorem 3.4 leads to the total work estimate $W_{\mathrm{tot}} = \mathcal{O}\big(\epsilon^{-2}\log(\epsilon)^2\big)$. Note that the sample numbers M_l in Theorem 3.4 are chosen for each single level l independent of $L(\epsilon)$ (see, e.g., (3.16)).

Proof of Theorem 3.4. In the proof we determine first a natural number of levels $L = L(\epsilon) \geq 2$ such that the bias term ϵ_{det} in (3.5) is bounded by $\epsilon/2$. Based on this number L we define the sample number M_l for each level $l \in \{0,\dots,L\}$ and verify that $\epsilon_{\mathrm{stat}} \leq \epsilon/2$ and the complexity estimate (3.12) hold.

Using assumption (A2) on the bias term and the definition of h_l and q_l yields

$$\epsilon_{\mathrm{det}} = \|\mathbb{E}\big[U\big] - \mathbb{E}\big[U_L\big]\|_{L^2(D)} \leq c_2 h_l^{\kappa_1\tilde{q}_l} = c_2(\lambda^{-l}h_0)^{\kappa_1(\tilde{q}_0+\beta l)}. \tag{3.13}$$

To ensure that the latter term is bounded by $\epsilon/2$ for $l = 0,\dots,L$ it suffices to determine L such that $P(L) \leq 0$ holds with

$$P(l) := \log(2c_2\epsilon^{-1}) + \kappa_1(\tilde{q}_0 + \beta l)\big(\log(h_0) - l\log(\lambda)\big). \tag{3.14}$$

The quadratic polynom P is bounded from above and monotone decreasing on $[0,\infty)$. Since $\epsilon \leq 2c_2 h_0^{\kappa_1\tilde{q}_0}$ holds we have $P(0) > 0$. Thus, P vanishes for some $\bar{L} > 0$, and we choose the number of levels to be

$$L = L(\epsilon) = \max\{\lceil \bar{L} \rceil, 2\}. \tag{3.15}$$

We proceed considering the two different cases in (3.12).

Case A: $\kappa_2\tilde{q}_0 > \gamma_1$. We recall the definition of S from Lemma 3.6 and set $S_0 := h_0^{-\gamma_1/2} S(h_0^{\kappa_2/2}, \lceil \gamma_1/2 \rceil, \tilde{q}_0)$. Then we choose the number of samples on levels $l = 0,\dots,L$ to be

$$M_l := \Big\lceil 4\epsilon^{-2} c_3 S_0 h_l^{(\gamma_1+\kappa_2\tilde{q}_l)/2} \tilde{q}_l^{-\gamma_1/2} \Big\rceil. \tag{3.16}$$

Using (3.16) and assumption (A3) we obtain for the statistical error

$$\epsilon_{\mathrm{stat}}^2 = \sum_{l=0}^{L} \frac{\sigma_l^2}{M_l} \leq \frac{1}{4}\epsilon^2 S_0^{-1} \sum_{l=0}^{L} h_l^{(\kappa_2\tilde{q}_l-\gamma_1)/2} \tilde{q}_l^{\gamma_1/2},$$

which implies by $\kappa_2\tilde{q}_l - \gamma_1 > \kappa_2\tilde{q}_0 - \gamma_1 > 0$, Lemma 3.6, and the definition of S_0 the desired estimate

$$\epsilon_{\mathrm{stat}}^2 \leq \frac{1}{4}\epsilon^2 S_0^{-1} \sum_{l=0}^{\infty} h_0^{(\kappa_2\tilde{q}_l-\gamma_1)/2} \tilde{q}_l^{\gamma_1/2} \leq \frac{1}{4}\epsilon^2. \tag{3.17}$$

Next we derive a bound for the total work $\mathrm{W}_{\mathrm{tot}}$; see (3.7). Using assumption (A1) and the ceiling definition of M_l we obtain

$$\begin{aligned}\mathrm{W}_{\mathrm{tot}} &\le c_1 \sum_{l=0}^{L} M_l h_l^{-\gamma_1} \tilde{q}_l^{\gamma_1} \le c_1 \left(4\epsilon^{-2} S_0 c_3 \sum_{l=0}^{L} h_l^{(\kappa_2 \tilde{q}_l - \gamma_1)/2} \tilde{q}_l^{\gamma_1/2} + \sum_{l=0}^{L} h_l^{-\gamma_1} \tilde{q}_l^{\gamma_1} \right) \\ &=: c_1 \big(\mathrm{W}_{\mathrm{tot},1} + \mathrm{W}_{\mathrm{tot},2} \big). \end{aligned} \tag{3.18}$$

The term $\mathrm{W}_{\mathrm{tot},1}$ can be shown to be of order ϵ^{-2} by the same arguments as before. It remains to consider $\mathrm{W}_{\mathrm{tot},2}$. We have obviously

$$\mathrm{W}_{\mathrm{tot},2} \le h_L^{-\gamma_1} \tilde{q}_L^{\gamma_1} (1 + L) \le h_L^{-\gamma_1} \tilde{q}_L^{\gamma_1 + 1} \tag{3.19}$$

with the last inequality following from $1 + L \le \tilde{q}_0 + \beta L = \tilde{q}_L$. From (3.31) in Lemma 3.7, which we have moved to the end of this proof, we deduce

$$h_L^{-\gamma_1} \le (2c_2 \delta^{-1} \epsilon^{-1})^{\gamma_1 / \kappa_1 \tilde{q}_0}. \tag{3.20}$$

On the other hand, using the order bound in (3.31) from Lemma 3.7 and taking the logarithm yields

$$\tilde{q}_L \le \frac{\log(2c_2 \delta^{-1} \epsilon^{-1})}{\log(h_0^{-\kappa_1})} = \frac{\log(2c_2 \delta^{-1})}{\log(h_0^{-\kappa_1})} + \frac{\log(\epsilon^{-1})}{\log(h_0^{-\kappa_1})} =: \hat{c}_1 + \hat{c}_2 \log(\epsilon^{-1}).$$

Thus, $\tilde{q}_L$ grows at most logarithmically in ϵ^{-1}. Therefore we find a constant $\hat{c}_3 > 0$ which is independent of ϵ and L such that the algebraic estimate

$$\tilde{q}_L \le \hat{c}_3 \epsilon^{-\frac{2 - \frac{\gamma_1}{\kappa_1 \tilde{q}_0}}{\gamma_1 + 1}} \tag{3.21}$$

holds. Note that in Case A the term $2\kappa_1 \tilde{q}_0 - \gamma_1$ is positive due to assumption (A3). Now, using (3.20) and (3.21) in (3.19) yields for some $c > 0$ independent of ϵ the bound

$$\mathrm{W}_{\mathrm{tot}} \le c \epsilon^{-\frac{\gamma_1}{\kappa_1 \tilde{q}_0}} \epsilon^{-2 + \frac{\gamma_1}{\kappa_1 \tilde{q}_0}} = c \epsilon^{-2}. \tag{3.22}$$

Thus, Case A is proven.

Case B: $\kappa_2 \tilde{q}_0 < \gamma_1$. Recall the definition of $L^* \in \mathbb{N}$ to be such that $\kappa_2 \tilde{q}_{L^*} < \gamma_1 \le \kappa_2 \tilde{q}_{L^*+1}$. Let us assume first that we have $L^* < L$ with L from (3.15).
We choose then the number of samples M_l for $l = 0, \dots, L^*$ according to

$$M_l := \left\lceil 8\epsilon^{-2} c_3 h_l^{(\gamma_1 + \kappa_2 \tilde{q}_l)/2} h_{L^*}^{-(\gamma_1 - \kappa_2 \tilde{q}_0)/2} \big(1 - \lambda^{-(\gamma_1 - \kappa_2 \tilde{q}_0)/2)} \big)^{-1} \right\rceil \tag{3.23}$$

and for $l = L^* + 1, \dots, L$ by

$$M_l := \left\lceil 8\epsilon^{-2} c_3 S_* h_l^{(\gamma_1 + \kappa_2 \tilde{q}_l)/2} \tilde{q}_l^{-\gamma_1/2} \right\rceil. \tag{3.24}$$

Here we used $S_* = S(h_0^{\kappa_2/2}, \lceil \gamma_1/2 \rceil, \tilde{q}_{L^*+1})$. Let us consider a splitting for the statistical error given by

$$\epsilon_{\mathrm{stat}}^2 = \sum_{l=0}^{L} \frac{\sigma_l^2}{M_l} = \sum_{l=0}^{L^*} \frac{\sigma_l^2}{M_l} + \sum_{l=L^*+1}^{L} \frac{\sigma_l^2}{M_l} =: \epsilon_{\mathrm{stat},1}^2 + \epsilon_{\mathrm{stat},2}^2. \tag{3.25}$$

Using the definition (3.23) of M_l for $l \in \{0, \dots, L^*\}$ we have for the first term

$$\epsilon^2_{\text{stat},1} \le \frac{1}{8}\epsilon^2 h_{L^*}^{(\gamma_1-\kappa_2\tilde{q}_0)/2}(1-\lambda^{-(\gamma_1-\kappa_2\tilde{q}_0)/2)})\sum_{l=0}^{L^*} h_l^{(\kappa_2\tilde{q}_l-\gamma_1)/2}$$
$$\le \frac{1}{8}\epsilon^2 h_{L^*}^{(\gamma_1-\kappa_2\tilde{q}_0)/2}(1-\lambda^{-(\gamma_1-\kappa_2\tilde{q}_0)/2)})\sum_{l=0}^{L^*} h_l^{(\kappa_2\tilde{q}_0-\gamma_1)/2}.$$

For the last inequality we used the fact that $h_l^{(\kappa_2\tilde{q}_l-\gamma_1)/2} < h_l^{(\kappa_2\tilde{q}_0-\gamma_1)/2}$ holds as long as $0 > \kappa_2\tilde{q}_l - \gamma_1 > \kappa_2\tilde{q}_0 - \gamma_1$. Next we use $h_l = \lambda^{L^*-l}h_{L^*}$ to get by $\gamma_1 - \kappa_2\tilde{q}_0 > 0$ in Case B and $\lambda \ge 2$ the estimate

$$\epsilon^2_{\text{stat},1} \le \frac{1}{8}\epsilon^2 h_{L^*}^{(\gamma_1-\kappa_2\tilde{q}_0)/2}(1-\lambda^{-(\gamma_1-\kappa_2\tilde{q}_0)/2)})\sum_{l=0}^{L^*}(\lambda^{L^*-l}h_{L^*})^{-(\gamma_1-\kappa_2\tilde{q}_0)/2}$$
$$= \frac{1}{8}\epsilon^2(1-\lambda^{-(\gamma_1-\kappa_2\tilde{q}_0)/2)})\sum_{l=0}^{\infty}\left(\lambda^{-(\gamma_1-\kappa_2\tilde{q}_0)/2}\right)^l \le \frac{1}{8}\epsilon^2.$$

Since M_l is defined for $l \ge L^*$ (see (3.24)) almost as in Case A (see (3.16)) (with a level shift expressed in S_*) we can apply the same arguments as in Case A to show $\epsilon^2_{\text{stat},2} \le \epsilon^2/8$. Thus the splitting (3.25) leads to the desired result $\epsilon^2_{\text{stat}} \le \epsilon^2/4$ for the complete statistical error.

Similarly to the previous splitting (3.25) we rewrite the total work as

$$\text{W}_{\text{tot}} = \sum_{l=0}^{L} M_l w_l = \sum_{l=0}^{L^*} M_l w_l + \sum_{l=L^*+1}^{L} M_l w_l =: \text{W}_{\text{tot},1} + \text{W}_{\text{tot},2}. \tag{3.26}$$

The definition of M_l in (3.23) implies

$$\text{W}_{\text{tot},1} \le c_1 \sum_{l=0}^{L^*}\left(8\epsilon^{-2}c_3 h_l^{(\gamma_1+\kappa_2\tilde{q}_l)/2}h_{L^*}^{-(\gamma_1-\kappa_2\tilde{q}_0)/2}\left(1-\lambda^{-(\gamma_1-\kappa_2\tilde{q}_0)/2)}\right)^{-1}+1\right)h_l^{-\gamma_1}\tilde{q}_l^{\gamma_1}$$
$$=: c_1(\text{W}_{\text{tot},11} + \text{W}_{\text{tot},12}).$$

Starting with $\text{W}_{\text{tot},11}$ we use the definitions of h_l, q_l and estimate

$$\text{W}_{\text{tot},11} \le 8c_3\epsilon^{-2}h_{L^*}^{-(\gamma_1-\kappa_2\tilde{q}_0)/2}\left(1-\lambda^{-(\gamma_1-\kappa_2\tilde{q}_0)/2)}\right)^{-1}\sum_{l=0}^{L^*}(\lambda^l h_{L^*})^{-(\gamma_1-\kappa_2\tilde{q}_0)/2}\tilde{q}_l^{\gamma_1}$$
$$< 8c_3\epsilon^{-2}h_{L^*}^{-(\gamma_1-\kappa_2\tilde{q}_0)}\left(1-\lambda^{-(\gamma_1-\kappa_2\tilde{q}_0)/2)}\right)^{-1}\sum_{l=0}^{\infty}\left(\lambda^{-(\gamma_1-\kappa_2\tilde{q}_0)/2}\right)^l(\tilde{q}_0+\beta l)^{\gamma_1}.$$

Being in Case B and the condition $\lambda \ge 2$ ensures the existence of the infinite sum such that we are led for some $c > 0$ by (3.32) in Lemma 3.7 to

$$\text{W}_{\text{tot},11} \le c\epsilon^{-2-\frac{\gamma_1-\kappa_2\tilde{q}_0}{\kappa_1\tilde{q}_{L^*}}}. \tag{3.27}$$

Advancing to the total work contribution $\text{W}_{\text{tot},12}$ we estimate

$$\text{W}_{\text{tot},12} = \sum_{l=0}^{L^*} h_l^{-\gamma_1}\tilde{q}_l^{\gamma_1} \le h_{L^*}^{-\gamma_1}\tilde{q}_{L^*}^{\gamma_1}(1+L^*) \le h_{L^*}^{-\gamma_1}\tilde{q}_{L^*}^{\gamma_1+1}.$$

By the same arguments employed to derive (3.21) in Case A, but using (3.32) from Lemma 3.7 instead, we find a constant $\hat{c}_4 > 0$ such that

$$\tilde{q}_{L^*} \leq \hat{c}_4 \epsilon^{\frac{-2+\frac{\kappa_2 \tilde{q}_0}{\kappa_1 \tilde{q}_{L^*}}}{\gamma_1+1}} \tag{3.28}$$

holds. In this case we need $\kappa_1 \tilde{q}_{L^*} > \kappa_2 \tilde{q}_0/2$, which follows from assumption (A3) because $\kappa_1 \tilde{q}_{L^*} > \kappa_1 \tilde{q}_0 \geq \kappa_2 \tilde{q}_0/2$. Altogether we end up with

$$\mathrm{W}_{\mathrm{tot},12} \leq h_{L^*}^{-\gamma_1} \tilde{q}_{L^*}^{\gamma_1+1} \leq c\epsilon^{-2-\frac{\gamma_1-\kappa_2\tilde{q}_0}{\kappa_1\tilde{q}_{L^*}}},$$

which yields with (3.27) the desired bound for $\mathrm{W}_{\mathrm{tot},1}$ in (3.26).

For the second term $\mathrm{W}_{\mathrm{tot},2}$ in (3.26) we use an index shift and apply the same arguments which we have used for Case A with $\kappa_2 \tilde{q}_0 > \gamma_1$. We then obtain

$$\mathrm{W}_{\mathrm{tot},2} \leq c\epsilon^{-2} \leq c\epsilon^{-2-\frac{\gamma_1-\kappa_2\tilde{q}_0}{\kappa_1\tilde{q}_{L^*}}},$$

which concludes the proof of the theorem for Case B and $L^* < L$.

It remains to consider $L \leq L^*$. Then one defines M_l for $l = 1, \ldots, L$ like in (3.23) and gets for the statistical error and for the total work only the first sums in (3.25), (3.26), respectively. Exactly the same estimates imply the results for the statistical error and lead to

$$\mathrm{W}_{\mathrm{tot}} \leq c\epsilon^{-2-\frac{\gamma_1-\kappa_2\tilde{q}_0}{\kappa_1\tilde{q}_L}},$$

which concludes the proof. ∎

Two auxiliary lemmas have been used in the proof of Theorem 3.4. The first result refers to the convergence of some series and is taken from [23].

LEMMA 3.6 (see [23, Lemma 5.1]). *For integers $p, q \geq 1$, there are numbers $d_1, \ldots, d_p \in \mathbb{R}$ such that we have for $r \in (0,1)$*

$$\sum_{l=0}^{\infty} r^{(q+\beta l)}(q+\beta l)^p = S(r,p,q) \tag{3.29}$$

with

$$S(r,p,q) = \sum_{k=1}^{p} d_k r^k f^{(k)}(r), \quad f(r) = \frac{r^q}{1-r^\beta}. \tag{3.30}$$

The second lemma establishes estimates on ϵ exploiting the exact definition of L.

LEMMA 3.7. *Let the assumptions from Theorem 3.4 be valid. Let $L \in \mathbb{N}$, $L \geq 2$, be given as in (3.15). Then, there exists a constant $\delta > 0$ independent of ϵ such that the following inequality holds:*

$$\frac{\epsilon}{2}\delta \leq c_2 \min\left\{h_L^{\kappa_1\tilde{q}_0}, h_0^{\kappa_1\tilde{q}_L}\right\}. \tag{3.31}$$

Moreover, for $L^* \in \mathbb{N}$ satisfying $L^* < L$, we have the estimate

$$\frac{\epsilon}{2} \leq c_2 \min\left\{h_{L^*}^{\kappa_1\tilde{q}_0}, h_0^{\kappa_1\tilde{q}_{L^*}}\right\}. \tag{3.32}$$

Proof. A straightforward calculation yields

$$\begin{aligned} h_L^{\kappa_1\tilde{q}_L} &= (\lambda^{-1}h_{L-1})^{\kappa_1(\tilde{q}_{L-1}+\beta)} \\ &= h_{L-1}^{\kappa_1\tilde{q}_{L-1}} h_{L-1}^{\kappa_1\beta}\lambda^{-\kappa_1 q_L} \\ &= h_{L-1}^{\kappa_1\tilde{q}_{L-1}}\lambda^{-(L-1)\kappa_1\beta}h_0^{\kappa_1\beta}\lambda^{-\kappa_1\tilde{q}_0}\lambda^{-\kappa_1\beta L} \\ &= h_{L-1}^{\kappa_1\tilde{q}_{L-1}}\lambda^{\kappa_1(\beta-\tilde{q}_0)}h_0^{\kappa_1\beta}\lambda^{-2L\kappa_1\beta} \\ &= h_{L-1}^{\kappa_1\tilde{q}_{L-1}}\delta\lambda^{-2L\kappa_1\beta}, \end{aligned}$$

where we have set $\delta := \lambda^{\kappa_1(\beta-\tilde{q}_0)}h_0^{\kappa_1\beta}$. Let $\bar{L} \in (0,\infty)$ be the zero of P as defined in (3.14). For $\bar{L} \geq 1$ we deduce from $L \geq \bar{L} \geq L-1 \geq 1$ that

$$\frac{\epsilon}{2}\delta\lambda^{-2L\kappa_1\beta} \leq c_2 h_{L-1}^{\kappa_1\tilde{q}_{L-1}}\delta\lambda^{-2L\kappa_1\beta} = c_2 h_L^{\kappa_1\tilde{q}_L}. \tag{3.33}$$

After rearranging (3.33) we obtain

$$\begin{aligned} \frac{\epsilon}{2}\delta \leq c_2\lambda^{2\kappa_1\beta}h_L^{\kappa_1\tilde{q}_L} &= c_2\lambda^{2\kappa_1\beta L}\lambda^{-L\kappa_1\tilde{q}_L}h_0^{\kappa_1\tilde{q}_L} \\ &= c_2\lambda^{2\kappa_1\beta L}\lambda^{-L\kappa_1\tilde{q}_0}\lambda^{-\kappa_1\beta L^2}h_0^{\kappa_1\tilde{q}_L} \\ &= c_2\lambda^{\kappa_1\beta(2L-L^2)}\lambda^{-L\kappa_1\tilde{q}_0}h_0^{\kappa_1\tilde{q}_0}h_0^{\kappa_1\beta L}. \end{aligned} \tag{3.34}$$

Because of $L \geq 2$ it holds that $(2L-L^2) \leq 0$ and thus $\lambda^{\kappa_1\beta(2L-L^2)} \leq 1$. Furthermore, $h_0^{\kappa_1\beta L} \leq 1$ and $\lambda^{-L\kappa_1\tilde{q}_0} \leq 1$ which yields altogether (3.31).

If $\bar{L} \in (0,1)$ holds it follows from (3.15) that $L = 2$, and we calculate

$$h_2^{\kappa_1\tilde{q}_2} = \lambda^{-2\kappa_1\tilde{q}_2}h_0^{\kappa_1(\tilde{q}_0+2\beta)} = \delta h_0^{\kappa_1\tilde{q}_0},$$

where we have defined $\delta := \lambda^{-2\kappa_1\tilde{q}_2}h_0^{2\kappa_1\beta}$. Finally, (3.31) follows from the fact that $c_2 h_0^{\kappa_1\tilde{q}_0} \geq \frac{\epsilon}{2}$. The estimate (3.32) for $L^* < L$ follows from

$$c_2 h_{L^*}^{\kappa_1\tilde{q}_{L^*}} \geq c_2 h_{L-1}^{\kappa_1\tilde{q}_{L-1}} \geq \frac{\epsilon}{2}. \qquad \square$$

Remark 3.8 (choice of the maximum level). The number of levels L in Algorithm *hp*-MLMC can be computed a priori using (3.15). It is also possible to compute L on the fly. To this end we consider, similarly as in [12, 23],

$$\begin{aligned} \|\mathbb{E}\big[U\big] - \mathbb{E}\big[U_L\big]\|_{L^2(D)} &= \left\| \sum_{l=L+1}^{\infty} \big(\mathbb{E}\big[U_l\big] - \mathbb{E}\big[U_{l-1}\big]\big) \right\|_{L^2(D)} \\ &\leq \|\mathbb{E}\big[U_L\big] - \mathbb{E}\big[U_{L-1}\big]\|_{L^2(D)} \sum_{l=L+1}^{\infty} \frac{\|\mathbb{E}\big[U_l\big] - \mathbb{E}\big[U_{l-1}\big]\|_{L^2(D)}}{\|\mathbb{E}\big[U_L\big] - \mathbb{E}\big[U_{L-1}\big]\|_{L^2(D)}} \\ &\leq \|\mathbb{E}\big[U_L\big] - \mathbb{E}\big[U_{L-1}\big]\|_{L^2(D)} \sum_{l=1}^{\infty} (\lambda^{-\kappa_1\tilde{q}_0}h_0^{\kappa_1\beta})^l \\ &= \|\mathbb{E}\big[U_L\big] - \mathbb{E}\big[U_{L-1}\big]\|_{L^2(D)} \frac{\lambda^{-\kappa_1\tilde{q}_0}h_0^{\kappa_1\beta}}{1-(\lambda^{-\kappa_1\tilde{q}_0}h_0^{\kappa_1\beta})}, \end{aligned}$$

where we have used assumption (A2). Therefore, the condition for adding new levels becomes

$$\max_{j\in\{0,1,2\}} \frac{(\lambda^{-\kappa_1\tilde{q}_0}h_0^{\kappa_1\beta})^{(j+1)}}{1-(\lambda^{-\kappa_1\tilde{q}_0}h_0^{\kappa_1\beta})}\|\mathbb{E}\big[U_{L-j}\big]-\mathbb{E}\big[U_{L-j-1}\big]\|_{L^2(D)} \le \frac{1}{2}\epsilon. \tag{3.35}$$

This criterion ensures that the deterministic error approximated by an extrapolation from the three finest meshes is within the desired range; cf. [23].

3.4. Confidence intervals for the number of additional samples. In this section we discuss the computation of the optimal number of samples M_l based on confidence intervals, having in mind the use of queue-based HPC systems. In most modern large-scale computing systems, access to compute nodes is based on job schedulers. For the execution of a job, a certain number of nodes can be requested for a specified time slot. The job is executed after some queuing time, which can be much longer than the actual job execution time. In the context of *hp*-MLMC, it is advisable to submit a new job to the queue for each iteration of Algorithm *hp*-MLMC, since otherwise idle times of the compute nodes are very difficult to avoid. As each iteration requires its own queuing time, it is our aim to compute as many new samples as possible during one iteration of Algorithm *hp*-MLMC. On the other hand, we want to avoid computing more samples than optimal, as this would decrease the efficiency of the *hp*-MLMC method, and from an economical point of view wasted computing time is expensive. In that sense, one is facing two competing issues, namely, either to reduce queuing time by computing as many samples as possible per iteration or to reduce the number of unnecessarily computed samples, i.e., saving computing time. A straightforward approach to satisfy the first issue is to rely on the standard estimator $\hat{M}_l$. However, this approach contradicts the second aim of saving computing time. Let us recall that the quantities w_l and σ_l in (3.6) and (3.7) are not known exactly but are estimated by $\hat{w}_l, \hat{\sigma}_l$. In most cases the number of samples in the warm-up phase and after the first iteration of the *hp*-MLMC algorithm is too small to obtain a reliable estimate $\hat{M}_l$ of the optimal number of samples M_l. This wrong estimate may then lead to a severe overestimation of M_l (see, for example, Figure 3(a)) and thus spoils the goal of avoiding the computation of unnecessary samples and saving computing time. In order to satisfy the second goal of saving computing time, we want to properly account for the fact that M_l is only estimated by constructing a confidence interval for M_l.

More specifically, we want to construct a one-sided confidence interval $\mathrm{I}_{M_l} = [\underline{M_l}, \infty)$ such that $\mathbb{P}(M_l \in \mathrm{I}_{M_l}) \ge 1-\alpha$. To obtain the desired confidence interval we construct corresponding one-sided confidence intervals for σ_l, w_l denoted by $\mathrm{I}_{\underline{\sigma_l}} = [\underline{\sigma_l}, \infty)$, $\mathrm{I}_{\underline{w_l}} = [\underline{w_l}, \infty)$, $\mathrm{I}_{\overline{w_l}} = (-\infty, \overline{w_l}]$, respectively. As we do not have any information about the underlying distributions of σ_l and w_l we construct the confidence interval based on asymptotic confidence intervals, and hence our approach is heuristic because the central limit theorem implies that the number of samples needs to be sufficiently large to ensure that the estimators are asymptotically normally distributed. This seems to be a contradiction to the fact that we choose a small number of samples for the warm-up phase. However, we show in estimate (3.39) that our construction of the confidence interval is very conservative and yields a robust lower estimate on the optimal number of samples, although the number of warm-up samples scales like $\mathcal{O}(1)$. Indeed, we never overestimated the optimal number of samples in our computations, justifying our approach.

For the construction of the confidence interval for σ_l we use the method described in [2, Formula (6)], which employs an adjustment to the degrees of freedom of the χ^2-distribution. More precisely, we let

$$\hat{r}_l := \frac{2M_{\mathrm{tot}_l}}{\hat{\gamma}_{e_l} + \left(\frac{2M_{\mathrm{tot}_l}}{M_{\mathrm{tot}_l}-1}\right)},$$
$$\hat{\gamma}_{e_l} = \frac{M_{\mathrm{tot}_l}(M_{\mathrm{tot}_l}+1)}{(M_{\mathrm{tot}_l}-1)(M_{\mathrm{tot}_l}-2)(M_{\mathrm{tot}_l}-3)}\frac{\hat{\mu}_l^4}{\hat{\sigma}_l^4} - \frac{3(M_{\mathrm{tot}_l}-1)^2}{(M_{\mathrm{tot}_l}-2)(M_{\mathrm{tot}_l}-3)},$$

where $\hat{\mu}_l^4 := \| \sum_{i=1}^{M_{\mathrm{tot}_l}} ((U_l^i - U_{l-1}^i) - \frac{1}{M_{\mathrm{tot}_l}} \sum_{j=1}^{M_{\mathrm{tot}_l}} (U_l^j - U_{l-1}^j))^4 \|_{L^2(D)}$. For the lower confidence interval $\mathrm{I}_{\underline{\sigma_l}} = [\underline{\sigma_l}, \infty)$ we therefore define

$$\underline{\sigma_l} := \sqrt{\frac{\hat{r}_l \hat{\sigma}_l^2}{\chi^2_{1-\frac{\alpha}{2},\hat{r}_l}}}, \tag{3.36}$$

where for some $\alpha \in (0,1)$, $\chi^2_{1-\frac{\alpha}{2},\hat{r}_l}$ is the $(1-\frac{\alpha}{2})$-quantile of the χ^2-distribution for $\hat{r}_l$ degrees of freedom. If the random samples are normally distributed, it follows that $\hat{\gamma}_{e_l} = 0$ and thus $\hat{r}_l = M_{\mathrm{tot}_l} - 1$, i.e., we obtain the standard confidence interval for the variance of a normal distribution (cf. [2]). For w_l we compute the confidence interval using the standard asymptotic confidence interval for the mean, i.e.,

$$\underline{w_l} := \hat{w}_l - z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}_{w_l}}{\sqrt{M_{\mathrm{tot}_l}}}, \quad \overline{w_l} := \hat{w}_l + z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}_{w_l}}{\sqrt{M_{\mathrm{tot}_l}}}, \tag{3.37}$$

where $z_{1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$-quantile of the normal distribution and $\hat{\sigma}^2_{w_l}$ is the unbiased estimator for the variance of w_l. We then define

$$\underline{M_l} = \frac{1}{\epsilon^2} \frac{\underline{\sigma_l}}{\sqrt{\overline{w_l}}} \left(\sum_{k=0}^{L} \underline{\sigma_k} \sqrt{\underline{w_k}} \right) \tag{3.38}$$

and the confidence interval $\mathrm{I}_{\underline{M_l}} := [\underline{M_l}, \infty)$. Moreover, for $l = 0, \ldots, L$ we define the events

$$X_l := \{M_l \in \mathrm{I}_{\underline{M_l}}\}, \Sigma_{\epsilon,l,\mathrm{lower}} := \left\{ \frac{1}{\epsilon^2}\sigma_l \in \mathrm{I}_{\frac{1}{\epsilon^2}\underline{\sigma_l}} \right\}, \Sigma_{l,\mathrm{lower}} := \{\sigma_l \in \mathrm{I}_{\underline{\sigma_l}}\},$$
$$W_{l,\mathrm{lower}} := \left\{ \sqrt{w_l} \in \mathrm{I}_{\underline{\sqrt{w_l}}} \right\}, W_{l,\mathrm{upper}} := \left\{ \sqrt{w_l} \in \mathrm{I}_{\overline{\sqrt{w_l}}} \right\}.$$

It then follows that $Y_l \subseteq X_l$, with $Y_l := \bigcap_{k=0}^{L} (\Sigma_{k,\mathrm{lower}} \cap W_{k,\mathrm{lower}} \cap W_{l,\mathrm{upper}} \cap \Sigma_{\epsilon,l,\mathrm{lower}})$, for all $l = 0, \ldots, L$. Using elementary probability estimates and De Morgan's rule we estimate

$$\begin{aligned} \mathbb{P}(X_l) \geq \mathbb{P}(Y_l) &= 1 - \mathbb{P}(Y_l^c) \\ &\geq 1 - \sum_{k=0}^{L} \left((\mathbb{P}(\Sigma_{k,\mathrm{lower}}^c) + \mathbb{P}(W_{k,\mathrm{lower}}^c) + \mathbb{P}(W_{l,\mathrm{upper}}^c) + \mathbb{P}(\Sigma_{\epsilon,l,\mathrm{lower}}^c) \right). \end{aligned} \tag{3.39}$$

We construct the confidence intervals $\mathrm{I}_{\underline{\sigma_k}} = [\underline{\sigma_k}, \infty)$, $\mathrm{I}_{\underline{\sqrt{w_k}}} = [\underline{\sqrt{w_k}}, \infty)$, and $\mathrm{I}_{\overline{\sqrt{w_l}}} = (-\infty, \overline{\sqrt{w_l}}]$ such that

$$\mathbb{P}(\Sigma_{\epsilon,l,\mathrm{lower}}) = \mathbb{P}(\Sigma_{k,\mathrm{lower}}) = \mathbb{P}(W_{k,\mathrm{lower}}) = \mathbb{P}(W_{l,\mathrm{upper}}) = 1 - \frac{\alpha}{4L}.$$

This choice yields $\mathbb{P}(X_l) \geq 1 - \alpha$ for all $l = 0, \ldots, L$. Whereas the lower confidence bound $\underline{M_l}$ helps in saving computing time, it increases the number of queuing operations. Therefore, in order to balance the two competing issues of either reducing queuing time or saving computing time we introduce the parameter $\zeta \in [0,1]$ and let

Algorithm *hp*-MLMC with confidence intervals

Fix a tolerance $\epsilon > 0$, set $L = 2$, and set $\mathcal{L} := \{0, \dots, L\}$
Compute K_l (warm-up) samples on $l = 0, \dots, L$, set $M_{\text{tot}_l} := K_l$ and $\text{iter}_l = 1$
while $\mathcal{L} \neq \emptyset$ **do**
  for $l \in \mathcal{L}$ **do**
    Estimate $\hat{w}_l$ by (3.11), $\hat{\sigma}_l^2$ by (3.10) and then $\hat{M}_l$ by (3.9)
    Estimate $\underline{w_l}$, $\overline{w_l}$ by (3.37), $\underline{\sigma_l^2}$ by (3.36), and then $\underline{M_l}$ by (3.38)
    if $M_{\text{tot}_l} < \hat{M}_l$ **then**
      Set $\zeta_l^{\text{iter}_l}$ according to (3.41) and compute $\tilde{M}_l$ by (3.40)
      Add $\lceil \tilde{M}_l - M_{\text{tot}_l} \rceil$ new samples of $U_l^i - U_{l-1}^i$
      Set $M_{\text{tot}_l} := M_{\text{tot}_l} + \lceil \tilde{M}_l - M_{\text{tot}_l} \rceil$
      Set $\text{iter}_l := \text{iter}_l + 1$
    else
      Set $\text{iter}_l := \text{iter}_l + 1$
      Skip level l
    end if
  end for
  if the statistical tolerance (3.8) is satisfied **then**
    if the bias term satisfies (3.35) **then**
      Set $\mathcal{L} := \emptyset$
    else
      Set $\mathcal{L} := \mathcal{L} \cup \{L + 1\}$
      Compute K_{L+1} (warm-up) samples and set $M_{\text{tot}_{L+1}} := K_{L+1}$
      Set $\text{iter}_{L+1} = 1$
      Set $L := L + 1$
    end if
  end if
end while
Compute $\text{E}_{hp}^L[U_L]$

$$\tilde{M}_l := \zeta \underline{M_l} + (1 - \zeta)\hat{M}_l. \tag{3.40}$$

By setting $\zeta = 0$ we pursue the aim of reducing queuing time, and by setting $\zeta = 1$ we try to save computing time. Choosing $\zeta \in (0, 1)$ corresponds to finding a strategy in between. We choose ζ adaptively for each iteration of Algorithm *hp*-MLMC and for each level $l = 0, \dots, L$. Hence, for each level $l = 0, \dots, L$, $\zeta = \zeta_l^{\text{iter}_l}$ has its own iteration counter iter_l. We choose the following strategy which tries to realize a trade-off between reducing queuing time and saving computing time:

$$\zeta_l^{\text{iter}_l} = \begin{cases} 1 & \text{for } \text{iter}_l = 1, \\ 0.5 & \text{for } \text{iter}_l = 2, \\ 0 & \text{for } \text{iter}_l > 2. \end{cases} \tag{3.41}$$

In the first iteration we rely on the lower confidence bound $\underline{M_l}$ from (3.38) to avoid overestimating M_l. In the second iteration we start to approach $\hat{M}_l$ but still consider $\underline{M_l}$ as a safe-guard for overestimating M_l. After the second iteration, where M_{tot_l} is sufficiently large, we trust the estimate $\hat{M}_l$. It might happen that during the first two iterations we have $\underline{M_l} \leq M_{\text{tot}_l} < \hat{M}_l$. In that case we set $\zeta_l^{\text{iter}_l} = 0$.

The modified *hp*-MLMC method which adds new levels adaptively based on the bias estimate (3.35), is summarized in Algorithm *hp*-MLMC with confidence intervals. For the number of warm-up samples K_l we typically choose ten to one hundred samples

on the coarse levels and three samples on the fine levels. When we add a new level we set the number of warm-up samples also equal to three.

4. Numerical experiments. We present numerical results for the *hp*-MLMC method as introduced in Algorithm *hp*-MLMC with confidence intervals. In section 4.1, we apply the *h*-, *p*-, *hp*-MLMC method to a smooth benchmark problem to verify Theorem 3.4. In section 4.2, we apply the *h*-, *p*-, and *hp*-MLMC method to an open cavity flow problem, an important flow problem from computational acoustics. We give a detailed comparison of all methods and verify that for both problems, *h*-, *p*-, and *hp*-MLMC yield an optimal asymptotic work. This shows that all three methods under consideration are applicable for UQ of complex engineering problems in computational fluid dynamics. The computations for the second numerical experiment were performed on Cray XC40 at the High-Performance Computing Center Stuttgart. The numerical solver relies on the DG spectral element solver FLEXI [16]. The time-stepping uses a Runge–Kutta method of order four.

4.1. Smooth benchmark solution. In this numerical example we verify Theorem 3.4 by means of a smooth manufactured solution, given by

$$\begin{pmatrix} \rho(t,x_1,x_2,y) \\ (\rho v_1)(t,x_1,x_2,y) \\ (\rho v_2)(t,x_1,x_2,y) \\ E(t,x_1,x_2,y)) \end{pmatrix} = \begin{pmatrix} 2 + A\sin(4\pi((x_1+x_2) - ft)) \\ 2 + A\sin(4\pi((x_1+x_2) - ft)) \\ 2 + A\sin(4\pi((x_1+x_2) - ft)) \\ \big(2 + A\sin(4\pi((x_1+x_2) - ft))\big)^2 \end{pmatrix}. \tag{4.1}$$

The benchmark solution (4.1) is obtained by introducing an additional source term in (2.1). We choose the amplitude and frequency of (4.1) to be uncertain, i.e., we let $A \sim \mathcal{U}(0.1, 0.9)$ and $f \sim \mathcal{N}(1, 0.05^2)$. The spatial domain is $D = (-1,1)^2$, and we consider periodic boundary conditions. The setup of the mesh hierarchies for *h*-MLMC, *p*-MLMC, and *hp*-MLMC can be found in Table 1. The final computational time for this example is $T = 1$, and the QoI is the momentum in x_1-direction at final time T, i.e., $Q(U) = (\rho v_1)(T,\cdot,\cdot,\cdot)$. For the confidence intervals from section 3.4 we set α to be 0.025.

In Figure 1(a) we plot the estimated bias, i.e., the quantity $\|\mathbb{E}\big[U_l\big] - \mathbb{E}\big[U_{l-1}\big]\|_{L^2(D)}$ from Remark 3.8, where the quantities $\mathbb{E}\big[U_l\big]$ and $\mathbb{E}\big[U_{l-1}\big]$ have been estimated by the standard MC estimator (3.2). In Figure 1(b) we plot the bias term vs. $h_l^{\bar{q}_l}$ in a log-log plot. This allows us to estimate κ_1 from assumption (A2) using a linear fit over all data points. For *h*-MLMC, where we consider a DG polynomial degree of five, we estimate $\kappa_1 \approx 0.86$, and this is in accordance with the expected value of one; cf. item 2 of Remark 3.3. On the other hand, the values of κ_1 for *p*- and *hp*-MLMC are clearly smaller than one. Figure 1(c) shows the estimated level variances $\hat{\sigma}_l^2$ and its 95% confidence interval. To verify the assumptions of Theorem 3.4 we

TABLE 1
Level setup for h-, p-, and hp-MLMC. Numerical experiment from section 4.1.

level	*h*-MLMC		*p*-MLMC		*hp*-MLMC	
	N_l	q_l	N_l	q_l	N_l	q_l
0	16	5	256	3	16	3
1	64	5	256	4	64	4
2	256	5	256	5	256	5
3	1024	5	256	6	1024	6

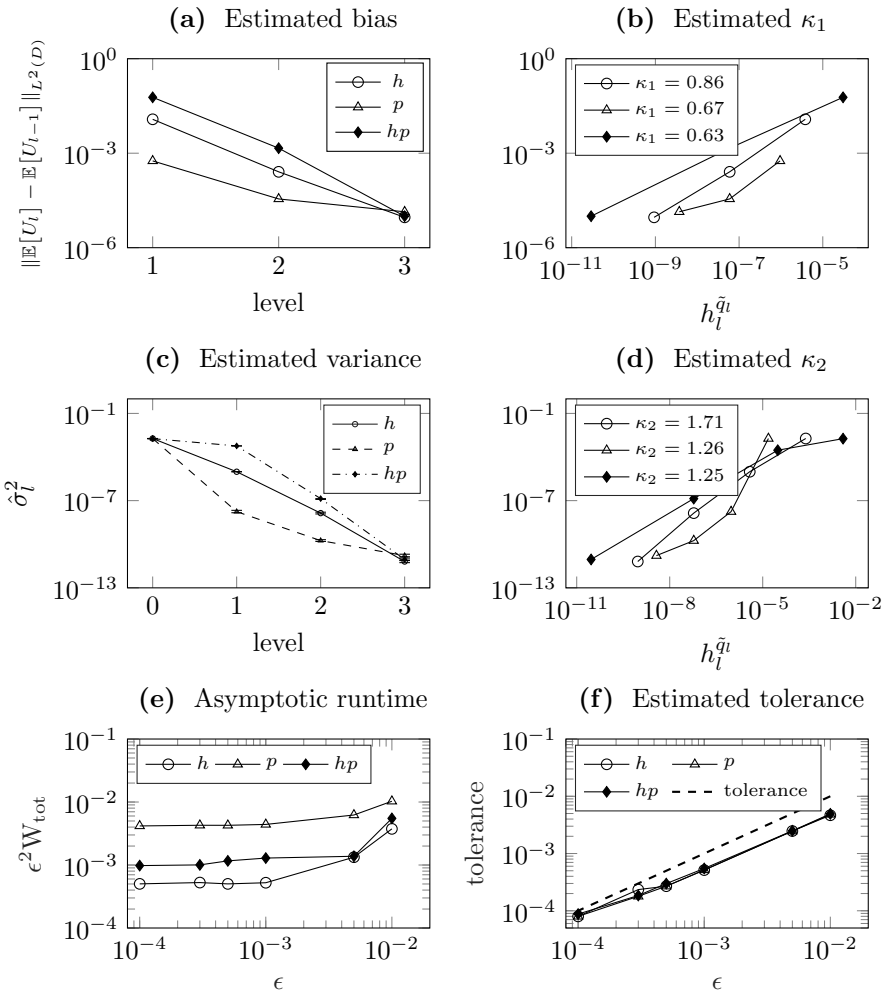

FIG. 1. *Estimated bias, variance, tolerance, and asymptotic runtime. For* $\hat{\sigma}_l^2$ *we also plot the* 95% *confidence interval. Numerical experiment from section* 4.1.

estimate κ_2 from assumption (A3) in Figure 1(d) by linearly fitting the last three data points. For h-MLMC we estimate $\kappa_2 \approx 1.71$. According to item 3 of Remark 3.3. we expect $\kappa_2 \approx 2$. Again, p- and hp-MLMC yield distinctively smaller values than two. However, all three methods satisfy $2\kappa_1 \geq \kappa_2$ from assumption (A2). Next, we check assumption (A1) for all three methods under consideration, where we estimate the average work using the sample mean from (3.11). Since we expect $\gamma_1 = 3$ and because $\mathrm{DOF}_l := (h_l^{-1}\tilde{q}_l)^2$, the average work should scale as $\hat{w}_l = \mathcal{O}(\mathrm{DOF}_l^{3/2})$ (cf. item 1 of Remark 3.3). This can be observed in Figure 2(d). Combining the estimate of κ_2 from Figure 1(d) and the fact that $\gamma_1 \approx 3$, we compute that h-, p-, and hp-MLMC satisfy $\kappa_2 \tilde{q}_0 > \gamma_1$ (see the parameters of the coarsest level given in Table 1). Therefore, according to Theorem 3.4 all three methods should yield an optimal asymptotic work of $\mathcal{O}(\epsilon^{-2})$, which can be observed in Figure 1(e). It appears that for this example h-MLMC is more efficient than p- and hp-MLMC. This is probably due to the very good variance reduction across all levels and a similar average work compared to hp-MLMC.

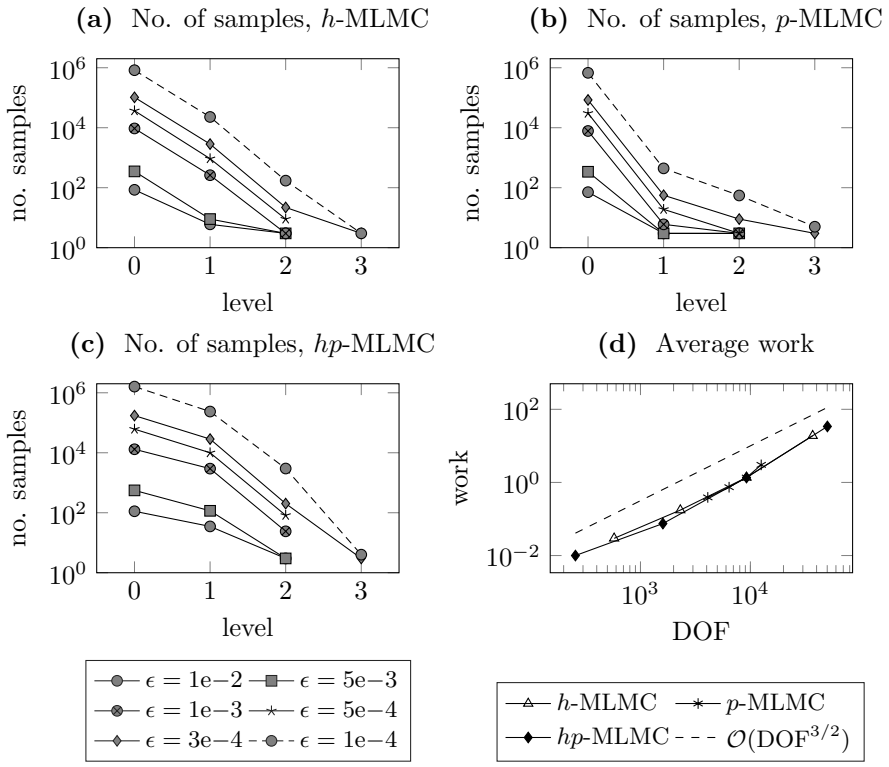

FIG. 2. *Computed number of samples and average work on each level. Numerical experiment from section* 4.1.

Finally, we check whether the prescribed tolerance ϵ is reached by all three methods. To this end we evaluate the statistical error ϵ_{stat} by (3.10) and compute the deterministic approximation error ϵ_{det} using the exact solution (4.1). Both terms ϵ_{det} and ϵ_{stat} are then summed up as in (3.5) and plotted in Figure 1(f). The computed number of samples on each level for different tolerances is shown in Figure 2. As expected, we only need a few computations on the fine levels, the majority of the computations is performed on coarse levels.

Figure 3 illustrates the advantage of computing the lower confidence bound $\underline{M_l}$ from (3.38). In this figure we plot the values of $\underline{M_l}$ from (3.38), of $\tilde{M}_l$ from (3.40), and of $\hat{M}_l$ from (3.9) for each level $l = 0, \dots, 3$, for the first three iterations of Algorithm *hp*-MLMC with confidence intervals. For this example we use *hp*-MLMC. In Figure 3(a) we observe that relying on the estimate of $\hat{M}_0$ would have led to an overestimate of the optimal number of samples by more than 8000 samples. The same holds for level two (Figure 3(b)) where we would have overestimated the optimal number of samples by more than 400. On the other hand, our proposed strategy ensures that we reach the standard estimator $\hat{M}_l$ in (at most) three iterations. This shows in particular the advantage of our approach because we avoid computing unnecessary samples; hence we save computing time while trying to keep the number of queuing operations low.

4.2. Open cavity. In this numerical example we investigate the influence of uncertain input parameters on the aeroacoustic feedback of cavity flows as in [20]. The prediction of aeroacoustic noise is an important branch of research, for example, in the

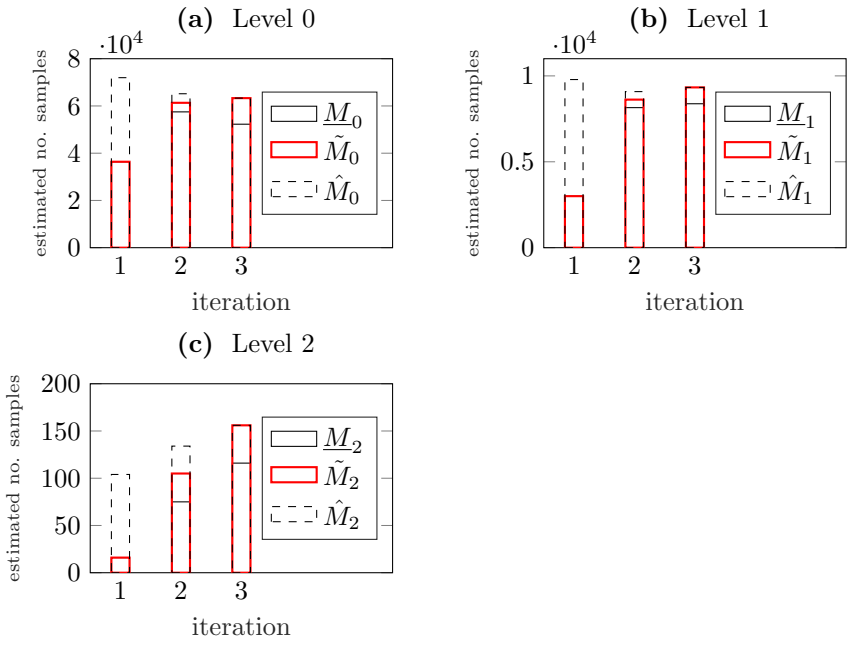

FIG. 3. *Estimated* $\underline{M}_l$, $\hat{M}_l$ *and chosen number of samples* $\tilde{M}_l$ *from* (3.40) *for different levels for the hp-MLMC method. The tolerance in this example is* $\epsilon = 5\text{e}{-4}$. *The number of warm-up samples was* $(100, 10, 3)$ *(Numerical experiment from section* 4.1*).*

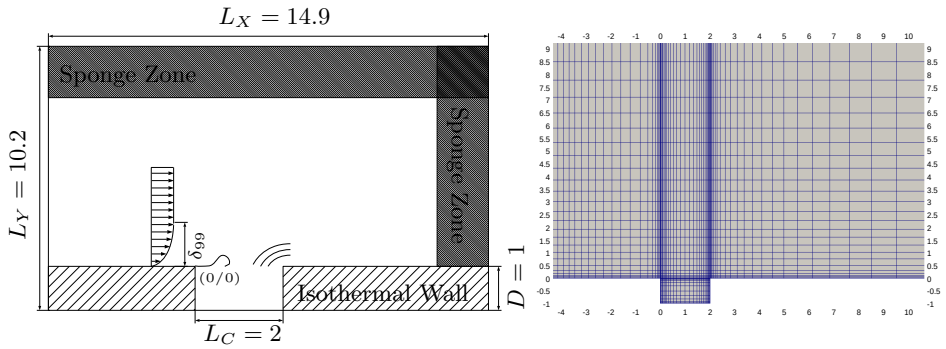

FIG. 4. *Left: Schematic sketch of the open cavity setup with a laminar inflow boundary layer. All geometric parameters are adopted from* [20] *and are nondimensionalized by the cavity depth. Right: Computational mesh on the finest level. (Numerical experiment from section* 4.2*).*

automotive industry. However, due to the large bandwidth of spatial and temporal scales, a high-order numerical scheme with low dissipation and dispersion error is necessary to preserve important small-scale information, and hence it poses a very challenging numerical problem for UQ. We consider the flow over a two-dimensional open cavity (cf. Figure 4) using *h*-, *p*-, and *hp*-MLMC.

For this flow problem we consider two uncertain parameters. The first uncertain parameter is the initial condition for pressure, i.e., we let $p^0 \sim \mathcal{N}(1.8, 0.01^2)$ be normally distributed. With this choice the Mach number $\text{Ma} = \frac{v_1}{c}$, with $c = \sqrt{\kappa \frac{p}{\rho}}$ being the speed of sound, becomes uncertain. The initial condition in primitive variables

then reads as

$$\left(\rho^0, v_1^0, v_2^0, p^0\right) = \left(1, 1, 0, p^0(y)\right).$$

As a second uncertain parameter, we let the boundary layer thickness in front of the cavity, $\delta_{99} \sim \mathcal{U}(0.28, 0.48)$, be uniformly distributed. For the boundary conditions at the inlet we employ Dirichlet boundary conditions in combination with the precomputed Blasius boundary layer profile. All wall boundaries are modeled as isothermal no-slip walls where the temperature $\mathcal{T}$ is computed from the ideal gas law $p = \rho R\mathcal{T}$, where the gas constant for air satisfies $R = 287.058$. At the end of the cavity we consider a pressure outflow boundary condition, where the pressure is specified by the initial pressure. Above the cavity we also consider a pressure outflow boundary condition, and we augment the boundary with a sponge zone (cf. Figure 4) to avoid that artificial reflections reenter the computational domain. Detailed information about the sponge zone can be found in [20]. For our QoI we record the pressure fluctuations $p(t, x, y)$ on top of the cavity at $\bar{x} = (x_1, x_2) = (1.57, 0)$ over time and then perform the discrete-time Fourier transform (DTFT) to obtain the sound pressure spectrum at $\bar{x}$, i.e., $Q(U) = \mathrm{DTFT}(p(\cdot, \bar{x}, y))$. The corresponding L^2-norm is then taken in frequency space. The mesh hierarchies for h-, p-, and hp-MLMC can be found in Table 2. For the confidence intervals from section 3.4 we set α to be 0.025.

Considering the bias error we can see that in Figure 5(a) for p-MLMC the bias estimate $\|\mathbb{E}\big[Q(U_l)\big] - \mathbb{E}\big[Q(U_{l-1})\big]\|_{L^2}$ for p-MLMC is smaller than 2.5e−5. As all three methods share the same finest level, we can assume that the bias error from (3.5) is satisfied for all three methods under consideration. Figure 5(b) shows the estimated level variance $\hat{\sigma}_l^2$ across different levels. All three methods yield a very good variance reduction; especially p-MLMC has already a very small variance on level two. However, the computation of samples on the coarse grids is extremely costly for p-MLMC. Taking a closer look at the variance on level zero, we see that hp-MLMC achieves the same variance as p-MLMC but with much less DOFs. This yields the computational advantage of h- and hp-MLMC compared to p-MLMC (see Figure 5(c)) for this open cavity problem. The asymptotic work is still optimal for all three methods, which can be seen in Figure 5(c). Finally, in Figure 6 we plot the number of computed samples for different tolerances and the average work that is needed to compute one sample on the corresponding level. Again, for all three methods most of the computations are performed on the coarse levels as they have a low computational cost. In contrast to the benchmark problem the average work does not scale as $\mathcal{O}(\mathrm{DOF}_l^{3/2})$ but more like $\mathcal{O}(\mathrm{DOF}_l^2)$, indicating that $\gamma_1 \approx 4$. The main reason for this behavior is that the uncertain Mach number influences the time-step size because the speed of sound enters the eigenvalues of the Jacobian of the advective fluxes; hence a bigger pressure leads to a smaller time-step size.

TABLE 2
Level setup for h-, p, and hp-MLMC (Numerical experiment from section 4.2).

level	h-MLMC		p-MLMC		hp-MLMC	
	N_l	q_l	N_l	q_l	N_l	q_l
0	279	7	1987	4	279	4
1	423	7	1987	5	423	5
2	957	7	1987	6	957	6
3	1987	7	1987	7	1987	7

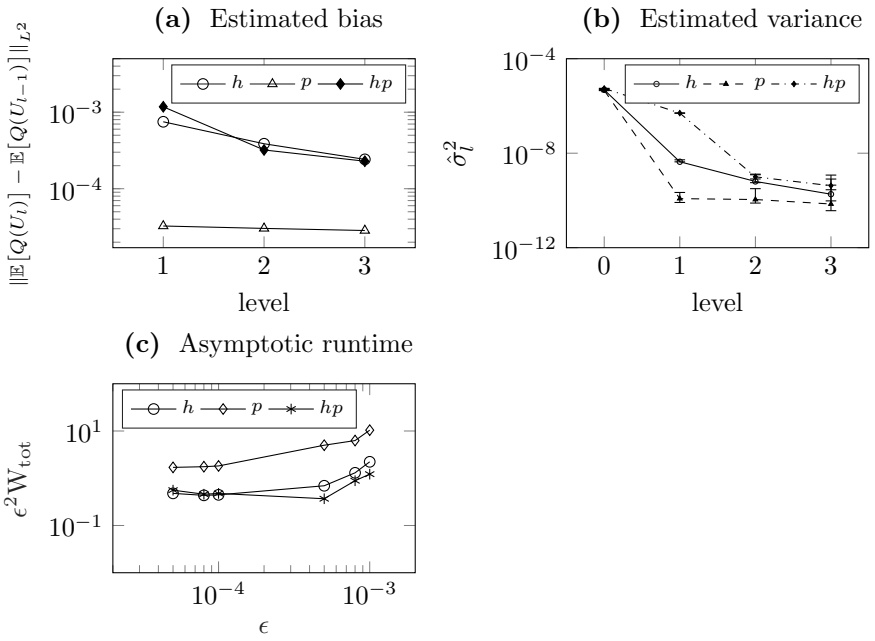

FIG. 5. *Estimated bias, variance, and asymptotic work. For $\hat{\sigma}_l^2$ we also plot the 95% confidence interval (Numerical experiment from section 4.2).*

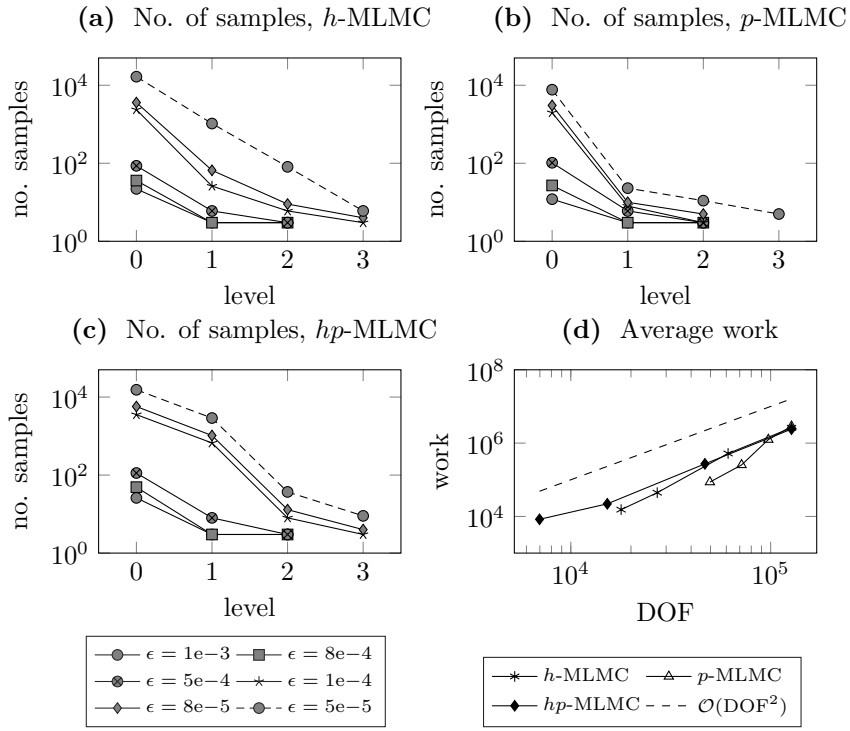

FIG. 6. *Computed number of samples and average work on each level (Numerical experiment from section 4.2).*

Summarizing the numerical experiments in sections 4.1 and 4.2 we have seen that h-, p-, and hp-MLMC yield an optimal asymptotic runtime and can be used for efficient UQ of the compressible Navier–Stokes equations. However, we also observed that the hp-MLMC method does not outperform the h-MLMC method, as long as the latter can be applied. Therefore, in order to increase the efficiency of the hp-MLMC method future work should aim at fully hp-adaptive algorithms as, for example, in [10, 11, 19].

5. Conclusions. In this article we have proposed the hp-MLMC method, a DG-based MLMC, where different levels consist of a hierarchy of uniformly refined spatial meshes in combination with a hierarchy of uniformly increasing DG polynomial degrees. We generalized the complexity results from [8, 23] to arbitrarily hp-refined meshes. To account for the uncertainty of the optimal number of samples on each level, we introduced a confidence interval, leading to a robust lower confidence bound for these quantities. Our theoretical results are confirmed by numerical experiments for the two-dimensional compressible Navier–Stokes equations, including an extensive comparison between h-, p-, and hp-MLMC methods. Finally, we applied our method to a challenging engineering problem from computational acoustics, demonstrating its capability to perform UQ for complex flow problems. In order to improve the efficiency of the hp-MLMC method it should be advanced to an hp-adaptive method.

Acknowledgments. The authors thank the anonymous referees and the associate editor for their helpful suggestions to improve the article.

REFERENCES

[1] R. Abgrall and S. Mishra, *Uncertainty quantification for hyperbolic systems of conservation laws*, in Handbook of Numerical Methods for Hyperbolic Problems, R. Abgrall and C.-W. Shu, eds., Handb. Numer. Anal. 18, Elsevier/North-Holland, Amsterdam, 2017, pp. 507–544.

[2] S. Banik, A. N. Albatineh, M. O. A. Abu-Shawiesh, and B. G. Kibria, *Estimating the population standard deviation with confidence interval: A simulation study under skewed and symmetric conditions*, Internat. J. Statist. Med. Res., 3 (2014), pp. 356–367.

[3] A. Barth and F. G. Fuchs, *Uncertainty quantification for hyperbolic conservation laws with flux coefficients given by spatiotemporal random fields*, SIAM J. Sci. Comput., 38 (2016), pp. A2209–A2231.

[4] A. Barth, A. Lang, and C. Schwab, *Multilevel Monte Carlo method for parabolic stochastic partial differential equations*, BIT, 53 (2013), pp. 3–27.

[5] A. Barth, C. Schwab, and N. Zollinger, *Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients*, Numer. Math., 119 (2011), pp. 123–161.

[6] F. Bassi and S. Rebay, *A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations*, J. Comput. Phys., 131 (1997), pp. 267–279.

[7] Q. Chen and J. Ming, *The multilevel Monte Carlo method for simulations of turbulent flows*, Monthly Weather Rev., 146 (2018), pp. 2933–2947.

[8] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup, *Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients*, Comput. Vis. Sci., 14 (2011), pp. 3–15.

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

[10] G. Detommaso, T. Dodwell, and R. Scheichl, *Continuous level Monte Carlo and sample-adaptive model hierarchies*, SIAM/ASA J. Uncertain. Quantif., 7 (2019), pp. 93–116.

[11] 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 J. Uncertain. Quantif., 4 (2016), pp. 1219–1245.

[12] M. B. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.

[13] A.-L. Haji-Ali, F. Nobile, and R. Tempone, *Multi-index Monte Carlo: When sparsity meets sampling*, Numer. Math., 132 (2016), pp. 767–806.
[14] A. Harten and J. M. Hyman, *Self-adjusting grid methods for one-dimensional hyperbolic conservation laws*, J. Comput. Phys., 50 (1983), pp. 235–269.
[15] S. Heinrich, *Multilevel Monte Carlo methods*, in Large-Scale Scientific Computing, S. Margenov, J. Waśniewski, and P. Yalamov, eds., Berlin, Heidelberg, 2001, pp. 58–67.
[16] F. Hindenlang, G. J. Gassner, C. Altmann, A. Beck, M. Staudenmaier, and C.-D. Munz, *Explicit discontinuous Galerkin methods for unsteady problems*, Comput. & Fluids, 61 (2012), pp. 86–93.
[17] H. Hoel and S. Krumscheid, *Central limit theorems for multilevel Monte Carlo methods*, J. Complexity, 54 (2019), 101407.
[18] C. A. Kennedy, M. H. Carpenter, and R. M. Lewis, *Low-storage, explicit Runge-Kutta schemes for the compressible Navier-Stokes equations*, Appl. Numer. Math., 35 (2000), pp. 177–219.
[19] R. Kornhuber and E. Youett, *Adaptive multilevel Monte Carlo methods for stochastic variational inequalities*, SIAM J. Numer. Anal., 56 (2018), pp. 1987–2007.
[20] T. Kuhn, J. Dürrwächter, F. Meyer, A. Beck, C. Rohde, and C.-D. Munz, *Uncertainty quantification for direct aeroacoustic simulations of cavity flows*, J. Theor. Comput. Acoust., 27 (2019).
[21] S. Mishra and C. Schwab, *Sparse tensor Multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data*, Math. Comp., 81 (2012), pp. 1979–2018.
[22] S. Mishra, C. Schwab, and J. Sŭkys, *Multi-level Monte Carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions*, J. Comput. Phys., 231 (2012), pp. 3365–3388.
[23] M. Motamed and D. Appelö, *A multiorder discontinuous Galerkin Monte Carlo method for hyperbolic problems with stochastic parameters*, SIAM J. Numer. Anal., 56 (2018), pp. 448–468.
[24] F. Müller, P. Jenny, and D. W. Meyer, *Multi-level Monte Carlo for two phase flow and Buckley-Leverett transport in random heterogeneous porous media*, J. Comput. Phys., 250 (2013), pp. 685–702.
[25] M. Pisaroni, F. Nobile, and P. Leyland, *A continuation multi level Monte Carlo (C-MLMC) method for uncertainty quantification in compressible inviscid aerodynamics*, Comput. Methods Appl. Mech. Engrg., 326 (2017), pp. 20–50.
[26] Q. Zhang and C.-W. Shu, *Error estimates to smooth solutions of Runge-Kutta discontinuous Galerkin methods for scalar conservation laws*, SIAM J. Numer. Anal., 42 (2004), pp. 641–666.