International Journal for Uncertainty Quantification, 10(4):351–373 (2020)

YIELD OPTIMIZATION BASED ON ADAPTIVE NEWTON-MONTE CARLO AND POLYNOMIAL SURROGATES

Mona Fuhrländer,[1,2,*] Niklas Georg,[2,3] Ulrich Römer,[3] & Sebastian Schöps[1,2]

[1]*Institut für Teilchenbeschleunigung und Elektromagnetische Felder (TEMF), Technische Universität Darmstadt, Schlossgartenstr. 8, 64289 Darmstadt, Germany*

[2]*Centre for Computational Engineering, Technische Universität Darmstadt, Dolivostr. 15, 64293 Darmstadt, Germany*

[3]*Institut für Dynamik und Schwingungen, Technische Universität Braunschweig, Schleinitzstr. 20, 38106 Braunschweig, Germany*

*Address all correspondence to: Mona Fuhrländer, Institut für Teilchenbeschleunigung und Elektromagnetische Felder (TEMF), Technische Universität Darmstadt, Schlossgartenstr. 8, 64289 Darmstadt, Germany, E-mail: fuhrlaender@temf.tu-darmstadt.de

Original Manuscript Submitted: 12/20/2019; Final Draft Received: 6/27/2020

In this paper we present an algorithm for yield estimation and optimization consisting of Hessian-based optimization methods, an adaptive Monte Carlo (MC) strategy, polynomial surrogates, and several error indicators. Yield estimation is used to quantify the impact of uncertainty in a manufacturing process. Since computational efficiency is one main issue in uncertainty quantification, we propose a hybrid method, where a large part of a MC sample is evaluated with a surrogate model, and only a small subset of the sample is reevaluated with a high-fidelity finite element model. In order to determine this critical fraction of the sample, an adjoint error indicator is used for both the surrogate error and the finite element error. For yield optimization we propose an adaptive Newton-MC method. We reduce computational effort and control the MC error by adaptively increasing the sample size. The proposed method minimizes the impact of uncertainty by optimizing the yield. It allows one to control the finite element error, surrogate error, and MC error. At the same time it is much more efficient than standard MC approaches combined with standard Newton algorithms.

KEY WORDS: *adaptivity, failure probability, Monte Carlo, polynomial surrogates, stochastic optimization, stochastic sparse grid collocation, uncertainty quantification, yield analysis*

1. INTRODUCTION

There are many applications where uncertainty quantification and optimization under uncertainty is important. Uncertainty in the manufacturing process may lead to deviations in the design parameters, i.e., geometrical or material parameters, which may lead in turn to rejections due to malfunctioning. In this context, malfunctioning means that pre-defined performance feature specifications are not fulfilled. In order to quantify the impact of uncertainty we define the yield according to [1] as the percentage of functioning realizations in a manufacturing process. Thus, yield is mathematically equivalent to the concept of reliability and the relation between yield and failure probability is given in the form *yield = 1 – failure probability*. The topic of yield optimization is motivated by high-frequency electromagnetics and circuit design.

In general, it is not possible to carry out yield calculations exactly. Hence, many algorithms have been introduced to this end and the Monte Carlo (MC) method is probably the most popular one [2]. The main challenge of yield estimation is its high computational cost, since it requires numerous evaluations of the underlying model. In practice, these models are often given by partial differential equations (PDE) of high complexity and can only be solved

NOMENCLATURE

$\mathbf{A}_{\mathbf{p},r}$	general system matrix	$N_{\mathrm{MC}}^{\mathrm{start}}$	initial size of the MC sample for adaptive yield optimization
$\mathbf{A}_{\mathbf{p},r}^{\star}$	Hermitian transpose of $\mathbf{A}_{\mathbf{p},r}$		
$\mathbf{A}_{\omega}$	system matrix of waveguide model	$N_{\mathrm{MC}}^{\mathrm{new}}$	updated size of the MC sample in adaptive yield optimization
a	waveguide width		
b	waveguide height	$N_{\mathrm{MC}}^{\mathrm{old}}$	old size of the MC sample in adaptive yield optimization
c	upper bound to define the performance feature specifications	$N_{\mathbf{p}}$	number of uncertain input parameters
D	computational domain	N_{SC}	number of interpolation nodes
$\mathbf{E}'$	test functions	$N_{\Omega_{\mathrm{s}}}$	number of accepted sample points
$\mathbf{E}_{\omega}$	electric field phasor	$\mathbf{n}$	outer unit normal vector
$\mathbf{E}_{\omega}^{\mathrm{inc}}$	incident wave phasor (excitation)	$\mathbf{p}$	vector of uncertain input parameters $\left(\mathbf{p} = [p_1, ..., p_{n_{\mathbf{p}}}]\right)$
$\mathbf{E}_{\omega,h}$	finite element approximation of $\mathbf{E}_{\omega}$		
E_0	amplitude of incident wave	$\mathbf{p}_i$	realization of the input parameter vector $\mathbf{p}$
$\mathbf{E}_{10}^{\mathrm{TE}}$	fundamental transverse electric mode		
$\mathbf{e}_{\omega}$	vector of degrees of freedom	$\mathbf{p}^{(i)}$	interpolation nodes
$e_{\omega,j}$	degree of freedom	p_j	uncertain input parameter
$\mathbf{e}_y$	unit vector in y-direction	p_1, p_2	length of the inlay, length of the offset of the waveguide
eff	computational effort for yield estimation or optimization	$p_3, ..., p_{12}$	material parameters of the waveguide
err	error of the yield estimator compared to the reference solution	$\overline{\mathbf{p}}$	mean value of the uncertain input parameter vector $\mathbf{p}$
f	frequency	$\overline{\mathbf{p}}_0$	mean value of the starting point for yield optimization
$\mathbf{f}_r$	general discrete right-hand side		
$\mathbf{f}_{\omega}$	discrete right-hand side of waveguide model	$\overline{\mathbf{p}}_{\mathrm{e}}$	mean value of the considered point for yield estimation
g_r	forcing term		
$H(\mathrm{curl}, D)$	complex function space of square integrable functions with square integrable curl	$\overline{\mathbf{p}}_{\Omega_{\mathrm{s}}}$	mean value of the accepted sample points
		$\hat{\overline{\mathbf{p}}}_{\Omega_{\mathrm{s}}}$	MC approximation of the mean value of the accepted sample points
h	mesh size for FEM		
inc	incremental factor for the adaptive Newton method	pdf	probability density function
		Q	quantity of interest
j	imaginary unit	Q_h	finite element approximation of quantity of interest
$\mathbf{K}$	stiffness matrix		
k_{z10}	propagation constant	q	parameter for angular condition in Newton method
$L_{\mathbf{p},r}$	parametric differential operator		
$L_{\mathbf{p},r}^{\star}$	adjoint operator of $L_{\mathbf{p},r}$	$(q_r, \cdot)_D$	general linear functional defining the quantity of interest
$L^2(D)$	complex function space of square integrable functions on D	$(\mathbf{q}_r, \cdot)_{\mathbb{C}^{n_h}}$	general discrete linear functional defining the discrete quantity of interest
$L^{\infty}(D)$	complex function space of essentially bounded functions on D	$(\mathbf{q}_{\omega}, \cdot)_{\mathbb{C}^{n_h}}$	discrete linear functional of the waveguide model
$\mathbf{M}^{\varepsilon}$	mass matrix		
$\mathbf{M}^{\mathrm{port}}$	system-matrix contribution stemming from port boundary conditions	r	range parameter
		r_j	range parameter point
$\mathbf{N}_j$	second order, first kind Nédélec basis functions	S	scattering parameter of the fundamental transverse electric mode on Γ_{P1}
N_h	number of degrees of freedom	S_h	finite element approximation of S
N_{MC}	size of the Monte Carlo sample	s	safety factor

NOMENCLATURE (*continued*)

s^k	search direction in the k-th step of the Newton method
T_r	range to define the performance feature specifications
T_d	discretized range to define the performance feature specifications
u_r	solution of the model problem
$\mathbf{u}_r$	discrete primal solution
V	function space defined in Eq. (20)
V_h	finite-dimensional subspace of V
w	weight function
Y	yield
Y_MC	MC estimator of the yield
Y_Ref	reference value of the yield for numerical tests
$\nabla_{\overline{\mathbf{p}}} Y_\mathrm{DQ}$	gradient of the yield according to differential quotient
$\nabla_{\overline{\mathbf{p}}} Y_\mathrm{G}$	analytical gradient of the yield according to Eq. (12)
$\mathbf{z}_r$	general discrete dual solution
$\mathbf{z}_\omega$	discrete dual solution of waveguide model

Greek Symbols

α_i	coefficients for stochastic collocation
β	parameter for Armijo rule in Newton method
$\Gamma_\mathrm{P1}, \Gamma_\mathrm{P2}$	waveguide ports
Γ_PEC	waveguide walls
γ	parameter for Armijo step size in Newton method
δ	step size for differential quotient
ϵ_fe	finite element error
ϵ_sc	stochastic collocation error
$\varepsilon_0, \varepsilon_\mathrm{r}$	vacuum and relative permittivity
μ_0, μ_r	vacuum and relative permeability
Ξ	image space of uncertain parameters of waveguide model
$\pi_\mathrm{t}, \pi_\mathrm{T}$	tangential trace operators
Σ	covariance matrix of the uncertain input parameter vector $\mathbf{p}$
$\Sigma_{\Omega_\mathrm{s}}$	covariance matrix of the accepted sample points
$\hat{\Sigma}_{\Omega_\mathrm{s}}$	MC approximation of the covariance matrix of the accepted sample points
σ^k	step size in the k-th step of the Newton method
σ_Y	standard deviation of the yield estimator
$\sigma_{Y,\max}$	upper bound for the standard deviation of the yield estimator
τ	relaxation time
Φ_i	global polynomial basis functions
φ_1, φ_2	parameters for angular condition in Newton method
ω	angular frequency
Ω_1, Ω_3	domain of the vacuum of the waveguide
Ω_2	domain of the dielectrical inlay of the waveguide
Ω_s	safe domain

Other Symbols

$\tilde{X}$	stochastic collocation approximation of a function X
$\mathcal{I}_\epsilon^1$	trusted interval (with estimated FE and SC error)
$\mathcal{I}_\epsilon^2$	trusted interval (with FE error)
$\#\mathrm{HF}_h$	number of high-fidelity (FE) evaluations with grid refinement h

numerically, with the finite element method (FEM), for instance. Since each high-fidelity evaluation with FEM itself may be computationally challenging, a standard MC analysis becomes rapidly prohibitive due to limits of computational and/or time resources. In this paper we present a hybrid approach for yield estimation combining the efficiency of stochastic collocation (SC) with the accuracy of MC for probability estimation. We then present an algorithm for yield maximization, based on a globalized Newton method.

The classical MC approach consists in sampling the original high-fidelity model, i.e., the highly resolved random finite element (FE) model. The efficiency of this approach is independent of the number of uncertain parameters and the method does not suffer from the "curse of dimensionality." Still, the sample size required for accurate estimation can be quite large [3]. There is a lot of research on reducing the computational effort of failure probability or yield estimation. The common goal is to reduce the number of high-fidelity evaluations. There are sampling-free methods such as the first-order reliability method (FORM) or the second-order reliability method (SORM). These methods determine the most probable point, which is the closest point from the parameter domain origin to the separating surface between the failure region and the safe region, and employ approximations of the limit state function around this point [4,5]. Investigations in the context of sampling have led to a sample size reduction, e.g., through importance

sampling [6] or subset simulation [7,8]. Alternatively or complementarily, the computational effort has been reduced for each sample point, e.g., with surrogate based approaches. In these surrogate methods an approximation (surrogate/response surface) of the original model is built using high-fidelity evaluations of a small training set, followed by MC sampling of the surrogate model [9]. In order to build the surrogate different methods have been employed, e.g., linear regression [10], Gaussian process regression [11], or SC [12]. In [13] a combination of two surrogate models, Gaussian process regression and SC is proposed. However, the accuracy of the surrogate depends on the size of the training set and the number of uncertain parameters. For a large number of uncertain parameters, the computational costs can exceed the costs for MC [14]. Furthermore, as shown in [15], there are examples where the surrogate model is highly accurate, measured by classical norms or pointwise, but the yield estimator fails drastically. In [15] a hybrid approach is proposed. Sample points which are close to the limit state function are evaluated based on the high-fidelity model; for all remaining sample points the surrogate model is used. Here, the assessment of whether a point is *close* to the interface between failure and safe domain is crucial for the accuracy and the efficiency of the algorithm. To this end, a method using an adjoint error indicator has been presented in [16]. Yield optimization has been carried out in [1], where a Newton method for optimization was presented, which was combined with the standard MC method.

In this paper, we present an algorithm for efficient yield estimation and optimization. For yield estimation we propose a hybrid approach similar to [15,16]. Contrary to the approach presented in [15] we use an adjoint error indicator to identify the aforementioned critical MC sample points. Also, contrary to [16] we build a polynomial surrogate model based on SC. Furthermore, we consider the FE error in addition to the surrogate error as a hybrid distinction criterion. If required, we refine the FE model for a subset of sample points. We then integrate this hybrid approach into the yield estimation and optimization framework. The optimization algorithm proposed in this paper is based on a globalized Newton method reported in [17]. For yield estimation, which is necessary in each iteration, we use our previously mentioned hybrid method, and during optimization we adaptively adjust the MC sample size. To the best of our knowledge, these are new elements in the context of yield optimization and we call the resulting algorithm adaptive Newton-MC. It achieves an *a priori* defined accuracy of the result and significantly reduces computational effort. Furthermore, we show the applicability of the presented estimation and optimization approaches to problems where the performance feature specifications are restrictions involving partial differential equations describing electromagnetic fields, i.e., Maxwell's equations in frequency domain.

This paper is structured as follows. After setting up the problem in Section 2, in Section 3 we will focus on yield estimation. We briefly review standard MC and SC. We then present the hybrid approach combining the two previous ones. In Section 4 we propose the new adaptive Newton-MC method for yield optimization, including the numerical algorithm. Numerical results for the application of electromagnetic field simulation are presented in Section 5 before the paper is concluded in Section 6.

2. PROBLEM SETTING

In this paper we consider a PDE with uncertainty in the input data. Details on the differential operator, geometry and boundary conditions will be postponed to the Section 5.1, which allows us to focus on the main algorithmic aspects for yield estimation and optimization. The starting point is the parametric model problem

$$L_{\mathbf{p},r} u_r(\mathbf{p}) = g_r \quad \text{in } D, \tag{1}$$

where $L_{\mathbf{p},r}$ is a linear parametric differential operator, g_r is a forcing term, $D \subset \mathbb{R}^d$ is a simply connected bounded domain, $\mathbf{p} \in \mathbb{R}^{N_\mathbf{P}}$ is the input parameter vector, and r is the range parameter. The range parameter may refer to frequency or to a temperature, for instance, which are not affected by uncertainties. We assume that the problem is well-posed for all $\mathbf{p}$ and that $\mathbf{p} \mapsto u(\mathbf{p})$ is a smooth function, which is often reasonable for parametrized differential equations; see [18] for the case of elliptic problems and [19] for other problem classes, for instance. Design objectives are frequently expressed through global quantities, which are modeled in our case as linear functionals of the solution. More precisely, we introduce a quantity of interest (QOI) as

$$Q(\mathbf{p}, r) := (q_r, u_r(\mathbf{p}))_D,$$

where $q_r \in L^2(D)$ and $L^2(D)$ denotes the space of complex square-integrable functions with inner product $(\cdot,\cdot)_D$.

A FE approach leads to the linear parametric system

$$\mathbf{A}_{\mathbf{p},r}\mathbf{u}_r(\mathbf{p}) = \mathbf{f}_r, \tag{2}$$

where $\mathbf{A}_{\mathbf{p},r} \in \mathbb{C}^{N_h \times N_h}$ denotes the system matrix and N_h the number of degrees of freedom. We denote with $u_{h;r}$ the interpolated discrete FE solution, without explicitly introducing the underlying polynomial FE space. Furthermore, we define the discrete linear QoI by

$$Q_h(\mathbf{p}, r) = (\mathbf{q}_r, \mathbf{u}_r)_{\mathbb{C}^{N_h}}, \tag{3}$$

where $(\cdot,\cdot)_{\mathbb{C}^{N_h}}$ refers to the finite-dimensional inner product.

We assume that the uncertainties originate in the manufacturing process which lead to deviations in the design parameters. These uncertainties are often classified as aleatory. The setting could be generalized by interpreting the computed yield to be conditioned on epistemic uncertainties and by further quantifying these uncertainties as outlined, for instance, in [20,21]. However, since the focus of the present work is on adaptivity and error control in the context of yield estimation, this will not be considered here. The percentage of functioning realizations in mass production is called the yield [1]. To give a mathematical definition, we model $\mathbf{p}$ as a random *design parameter* vector, with independent distributed elements p_j, $j = 1, ..., N_{\mathbf{p}}$. Typically the p_j are assumed to follow a normal distribution, i.e., $p_j \sim \mathcal{N}(\overline{p_j}, \sigma_j)$ with mean value $\overline{p_j} \in \mathbb{R}$ and standard deviation $\sigma_j \in \mathbb{R}$ and probability density function

$$\mathrm{pdf}_{\mathcal{N}(\overline{p_j},\sigma_j)} = \frac{1}{\sqrt{2\pi\sigma_j^2}}\, \mathrm{e}^{-\left((p_j-\overline{p_j})^2/2\sigma_j^2\right)}.$$

Then, the uncertain parameter $\mathbf{p}$ follows a multivariate normal distribution, i.e., $\mathbf{p} \sim \mathcal{N}(\overline{\mathbf{p}}, \mathbf{\Sigma})$ with mean value $\overline{\mathbf{p}} \in \mathbb{R}^{N_\mathbf{P}}$ and a diagonal covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{N_\mathbf{P} \times N_\mathbf{P}}$ and probability density function

$$\mathrm{pdf}_{\mathcal{N}(\overline{\mathbf{p}},\mathbf{\Sigma})} = \frac{1}{\left(\sqrt{2\pi}\right)^{N_\mathbf{P}}\sqrt{\det \mathbf{\Sigma}}}\, \mathrm{e}^{-(1/2)\left((\mathbf{p}-\overline{\mathbf{p}})^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{p}-\overline{\mathbf{p}})\right)}.$$

The normality assumption may be justified by the central limit theorem in the presence of averaging processes. Note that, in order to simplify notation, we do not distinguish between a random vector and its realization, whenever there is no confusion in a specific context. Following [1] we further define a *range parameter* $r \in T_r = [r_1, r_2]$ and the *performance feature specification*

$$Q(\mathbf{p}, r) \leq c \;\; \forall r \in T_r, \tag{4}$$

where c is a constant and Q the QoI introduced above. Note that, without loss of generality, we defined the performance feature specification with an upper bound. For the sake of notation simplicity, we consider only one. This may be read component-wise, as is usual in optimization. The *safe domain* Ω_s is the set of all parameters, which fulfill the performance feature specifications, i.e.,

$$\Omega_\mathrm{s} := \{\mathbf{p} : Q(\mathbf{p}, r) \leq c \;\; \forall r \in T_r\}.$$

Then we can express the yield as

$$Y(\overline{\mathbf{p}}) := \mathbb{E}[\mathbb{I}_{\Omega_\mathrm{s}}(\mathbf{p})] := \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbb{I}_{\Omega_\mathrm{s}}(\mathbf{p})\, \mathrm{pdf}_{\mathcal{N}(\overline{\mathbf{p}},\mathbf{\Sigma})}(\mathbf{p})\, \mathrm{d}\mathbf{p}, \tag{5}$$

where $\mathbb{E}$ denotes the expected value and $\mathbb{I}_{\Omega_\mathrm{s}}(\mathbf{p})$ the indicator function defined by

$$\mathbb{I}_{\Omega_\mathrm{s}}(\mathbf{p}) = \begin{cases} 1 & \mathbf{p} \in \Omega_\mathrm{s}, \\ 0 & \text{else.} \end{cases}$$

Note that $\overline{\mathbf{p}}$ will be a design parameter during optimization, whereas the covariance is fixed, which is taken into account by our notation in Eq. (5).

3. YIELD ESTIMATION

We proceed by describing a numerical method for yield estimation. The starting point will be a brief description of the MC method, followed by an outline of surrogate modeling based on SC. The section will conclude with a description of a hybrid MC method.

3.1 Monte Carlo

The most straightforward approach in order to estimate the yield, i.e., compute the integrals of Eq. (5), is a MC analysis [2,22]. In a MC approach, we consider a large number of independent random variables, distributed in the same way as $\mathbf{p}$. The set $\{\mathbf{p}_i\}_{i=1}^{N_{\text{MC}}}$, where each $\mathbf{p}_i$ represents a realization of the corresponding random variable, is called a sample and N_{MC} represents the sample size. At each sample point $\mathbf{p}_i$, we evaluate the high-fidelity FE model and count the sample points, which fulfill our performance feature specifications. Then we obtain a yield estimator as

$$Y(\overline{\mathbf{p}}) \approx Y_{\text{MC}}(\overline{\mathbf{p}}) := \frac{\#\text{ sample points in } \Omega_{\text{s}}}{\text{sample size}},$$

or, equivalently,

$$Y_{\text{MC}}(\overline{\mathbf{p}}) = \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} \mathbb{I}_{\Omega_{\text{s}}}(\mathbf{p}_i). \tag{6}$$

MC estimation is based on the law of large numbers, which ensures convergence for $N_{\text{MC}} \to \infty$ under mild regularity assumptions on the integrand. Since in practice, the sample size is always finite, we need to estimate the associated error. To this end, we use an error indicator from [3]. An estimator of the approximated yield variance is derived as follows. Since all observations are independent, we obtain

$$\begin{aligned}
\mathbb{V}[Y_{\text{MC}}(\overline{\mathbf{p}})] &= \frac{1}{N_{\text{MC}}^2} \mathbb{V}\left[\sum_{i=1}^{N_{\text{MC}}} \mathbb{I}_{\Omega_{\text{s}}}(\mathbf{p}_i)\right] \\
&= \frac{1}{N_{\text{MC}}^2} \sum_{i=1}^{N_{\text{MC}}} \mathbb{V}[\mathbb{I}_{\Omega_{\text{s}}}(\mathbf{p}_i)] \\
&= \frac{1}{N_{\text{MC}}^2} N_{\text{MC}} Y(\overline{\mathbf{p}})(1 - Y(\overline{\mathbf{p}})) \\
&= \frac{Y(\overline{\mathbf{p}})(1 - Y(\overline{\mathbf{p}}))}{N_{\text{MC}}},
\end{aligned}$$

where the expectation and variance are now defined with respect to the i.i.d. observations. Then, we derive the standard deviation of the yield estimator as

$$\sigma_Y = \sqrt{\frac{Y(\overline{\mathbf{p}})(1 - Y(\overline{\mathbf{p}}))}{N_{\text{MC}}}} \leq \frac{0.5}{\sqrt{N_{\text{MC}}}}. \tag{7}$$

The standard deviation depends on the size of the yield. Its maximum is attained for a yield of 50%, which yields the upper bound for the standard deviation given in Eq. (7). Since the MC estimator is unbiased, the variance is equal to the mean-square error. In view of Eq. (7), this approach guarantees a high accuracy for a large sample size, but it converges slowly with $\mathcal{O}\big(1/\sqrt{N_{\text{MC}}}\big)$. In many cases this is unaffordable due to the large number of expensive function evaluations required [3].

3.2 Stochastic Collocation and Error Estimation

To reduce the computational complexity of sampling the underlying FE solver, surrogate models can be employed. Based on the assumption that a map $X : \mathbb{R}^{N_{\mathbf{P}}} \times T_r \to \mathbb{C}$ (where X might refer to the QoI Q_h, for instance) is well-defined and sufficiently smooth, we denote by $\tilde{X}$ the surrogate approximation defined by

$$\tilde{X}(\mathbf{p}, r) = \sum_{i=1}^{N_{\text{SC}}} \alpha_i(r)\Phi_i(\mathbf{p}), \tag{8}$$

where N_{SC} is the number of interpolation nodes, $\Phi_i : \mathbb{R}^{N_{\text{P}}} \to \mathbb{R}$ are multivariate global polynomial basis functions with respect to $\mathbf{p}$, and $\alpha_i : T_r \to \mathbb{C}$ denotes the corresponding coefficients. Such a construction is appealing, as spectral convergence with respect to the polynomial degree can be expected [23]. In this work, we compute such approximations based on the SC method [18,24]. In particular, the surrogate model is obtained by evaluating (2) for a set of multivariate interpolation nodes $\{\mathbf{p}^{(i)}\}_{i=0}^{N_{\text{SC}}-1}$ and enforcing the corresponding collocation conditions on the surrogate model. The choice of the multivariate nodes $\mathbf{p}^{(i)}$ is crucial for the efficiency of SC. To this end, we first consider the tensor grid of univariate interpolation nodes $\{p_1^{(i)}\}_i \times \{p_2^{(i)}\}_i \times \ldots \times \{p_M^{(i)}\}_i$. Employing all points of the grid is computationally intractable for many parameters. Sparse-grids [25] are a viable alternative, where a subset of points, which do not significantly contribute to the approximation accuracy, is neglected. In this work, we use an algorithm proposed in [26, Algorithm 2], which constructs the sparse-grid adaptively. For the convenience of the reader, we recall the main ideas in the following.

The algorithm is based on weighted Leja nodes [27] which are defined recursively by an optimization problem; i.e., univariate weighted Leja nodes $\{p_m^{(i)}\}_i \subset \mathbb{R}$ are obtained as

$$p_m^{(I)} = \underset{p_m \in \mathbb{R}}{\arg\max} \sqrt{w(p_m)} \prod_{i=0}^{I-1} |p_m - p_m^{(i)}|,$$

where the weight function $w(p_m)$ is typically chosen as the probability density function of the corresponding input parameter; i.e., $w(p_m) = \text{pdf}_{\mathcal{N}(\overline{p_m}, \sigma_m)}$, and for the first node we set $p_m^{(0)} = 0$. Leja nodes are well suited for adaptive approximations in higher dimensions, since they are, by construction, nested and allow for a granular refinement [27]. To steer the adaptive selection of the corresponding multivariate nodes, an adjoint error indicator [28,29] is employed. To this end, we introduce the dual problem to Eq. (2), which is given by

$$\mathbf{A}_{\mathbf{p},r}^{\star}\, \mathbf{z}_r(\mathbf{p}) = \mathbf{q}_r,$$

where $\mathbf{A}^{\star}$ denotes the Hermitian transpose of $\mathbf{A}$. In addition to the polynomial approximation of the QoI $\tilde{Q}_h$, we construct polynomial approximations of the mappings $\mathbf{u}, \mathbf{z} : \mathbb{R}^{N_{\text{P}}} \times T_{\text{r}} \to \mathbb{C}$, where the same collocation points as for the QoI are employed, cf. [30]. The resulting approximations are denoted as $\tilde{\mathbf{u}}, \tilde{\mathbf{z}}$. We are then interested in the error,

$$\begin{aligned}
\epsilon_{\text{sc}}(\mathbf{p}, r) &= \left| Q_h(\mathbf{p}, r) - \tilde{Q}_h(\mathbf{p}, r) \right| \\
&= \left| \big(\mathbf{q}_r, \mathbf{u}_r(\mathbf{p})\big)_{\mathbb{C}^{N_h}} - \big(\mathbf{q}_r, \tilde{\mathbf{u}}_r(\mathbf{p})\big)_{\mathbb{C}^{N_h}} \right| \\
&= \left| \big(\mathbf{z}_r(\mathbf{p}), \mathbf{A}_{\mathbf{p},r}\, \mathbf{u}_r(\mathbf{p})\big)_{\mathbb{C}^{N_h}} - \big(\mathbf{z}_r(\mathbf{p}), \mathbf{A}_{\mathbf{p},r}\, \tilde{\mathbf{u}}_r(\mathbf{p})\big)_{\mathbb{C}^{N_h}} \right| \\
&= \left| \big(\mathbf{z}_r(\mathbf{p}), \mathbf{f}_r - \mathbf{A}_{\mathbf{p},r}\, \tilde{\mathbf{u}}_r(\mathbf{p})\big)_{\mathbb{C}^{N_h}} \right|.
\end{aligned} \tag{9}$$

The evaluation of Eq. (9) would always require the computation of $\mathbf{z}$, i.e., the solution of the high-fidelity adjoint problem. Hence, following [30], we employ the error indicator,

$$\tilde{\epsilon}_{\text{sc}}(\mathbf{p}, r) := \left| \big(\tilde{\mathbf{z}}_r(\mathbf{p}), \mathbf{f}_r - \mathbf{A}_{\mathbf{p},r}\, \tilde{\mathbf{u}}_r(\mathbf{p})\big)_{\mathbb{C}^{N_h}} \right|. \tag{10}$$

It should be noted that, under mild assumptions, cf. [26,28], the error ocurring when $\mathbf{z}$ is replaced with $\tilde{\mathbf{z}}$ is of higher-order. The error indicator is then used to select interpolation nodes which are admissible for refinement of the approximations until a given computational budget is reached and the algorithm terminates. For further details on the employed adaptive sparse-grid interpolation scheme, we refer to [26]. Once an accurate surrogate model is available, it can then be used as an inexpensive substitute of Eq. (3) for an extensive MC analysis (6).

Adjoint techniques can further be used to estimate the FE error following [31,32]. However, in this case, the continuous adjoint equation is required, which reads

$$L_{\mathbf{p},r}^{*} z_r(\mathbf{p}) = q_r \quad \text{in } D,$$

where $L^*_{\mathbf{p},r}$ denotes the adjoint operator with respect to the inner product $(\cdot,\cdot)_D$. With this notation at hand, we derive the following identity for the FE error:

$$\begin{aligned}\epsilon_{\text{fe}}(\mathbf{p},r) &= |(q_r, u_r(\mathbf{p}) - u_{h;r}(\mathbf{p}))_D| \\ &= \left|(L^*_{\mathbf{p},r} z_r(\mathbf{p}), u_r(\mathbf{p}) - u_{h;r}(\mathbf{p}))_D\right| \\ &= |(z_r(\mathbf{p}), L_{\mathbf{p},r}(u_r(\mathbf{p}) - u_{h;r}(\mathbf{p})))_D| \\ &= |(z_r(\mathbf{p}), g_r - L_{\mathbf{p},r} u_{h;r}(\mathbf{p}))_D|.\end{aligned}$$

A computable expression can only be obtained if the adjoint is replaced with a FE approximation. However, we cannot simply employ $z_{h;r}$ as it is orthogonal to the residual. Hence, a higher-order adjoint is required for the FE error, contrary to the surrogate error (10). A discussion can be found in [28]. Hence, we approximate the adjoint solution on a refined grid, but other options, such as higher polynomial degrees or recovery techniques [33], are equally applicable.

Finally, an error identity comprising both SC and FE contribution is obtained as

$$\begin{aligned}\left|Q(\mathbf{p},r) - \tilde{Q}_h(\mathbf{p},r)\right| &\leq \left|Q(\mathbf{p},r) - Q_h(\mathbf{p},r)\right| + \left|Q_h(\mathbf{p},r) - \tilde{Q}_h(\mathbf{p},r)\right| \\ &\approx \left|\left(z_{h/2;r}(\mathbf{p}), g_r - L_{\mathbf{p},r} u_{h;r}(\mathbf{p})\right)_D\right| + \left|\left(\tilde{\mathbf{z}}_r(\mathbf{p}), \mathbf{f}_r - \mathbf{A}_{\mathbf{p},r}\,\tilde{\mathbf{u}}_r(\mathbf{p})\right)_{\mathbb{C}^{N_h}}\right|. \qquad (11)\end{aligned}$$

The second term is immediately identified as $\tilde{\epsilon}_{\text{sc}}(\mathbf{p},r)$, which uses the surrogate approximations $\tilde{\mathbf{z}}_r(\mathbf{p})$ and $\tilde{\mathbf{u}}_r(\mathbf{p})$, and can therefore easily be evaluated for all $\mathbf{p}$. However, the first term is identified as a computable approximation to $\epsilon_{\text{fe}}(\mathbf{p},r)$, which we use, along with Eq. (8), to build the surrogate approximation $\tilde{\epsilon}_{\text{fe}}(\mathbf{p},r)$. The separation of the FE error and the SC error by the triangle inequality is a rather conservative choice to define the total error. We come back to this point at a later stage. We note, that the combined estimation of deterministic and stochastic discretization errors, has, for example, also been considered in [29], in the context of the stochastic Galerkin method for time-dependent forward and inverse problems.

3.3 Hybrid Approach

The number of collocation points N, for which the high-fidelity FE model needs to be solved, depends on the number of uncertain parameters and the polynomial degree the surrogate model is supposed to have. This number grows rapidly with the number of parameters ("curse of dimensionality") [34]. For adaptive sparse grids the required FE solver calls can be reduced significantly. However, we know from [15] that yield estimation may produce erroneous results even though the surrogate model may be highly accurate.

The aim of the hybrid approach is to restore the accuracy of the MC method while relying on surrogate modeling as much as possible to enhance the numerical efficiency. We propose a particular hybrid approach, which is an extension of the one presented in [15]. The main difference lies in the selection of sample points which have to be reevaluated with the high-fidelity model. These points are referred to as critical sample points in the following. In [15] a tube around the boundary of the failure domain is defined, where the tube size is either fixed in advance, or determined iteratively by an algorithm which adds critical samples points until some error bound is satisfied. In comparison to [16] the method we propose is using SC with Leja nodes as a surrogate model (see Section 3.2). Also, in addition to the surrogate model error (SC error), we also consider the FE error in order to determine the critical sample points. Both error contributions are estimated by the adjoint error indicator, according to Eq. (11).

In the following we assume for simplicity of notation the QoI to be real valued. Our procedure is summarized in Fig. 1. The first step is to build a surrogate model and to carry out a MC analysis with it. Then, we use an adjoint error indicator to quantify both the FE and surrogate error as

$$\tilde{\epsilon}_{\text{sc}}(\mathbf{p}_i, r_j) \;\text{ and }\; \tilde{\epsilon}_{\text{fe}}(\mathbf{p}_i, r_j) \;\; \forall i = 1,\ldots,N_{\text{MC}},\; \forall j = 1,\ldots,|T_{\text{d}}|,$$

where T_{d} is a discrete subset of T_r. We then verify whether the approximated QoI value, taking into account the aforementioned errors, meets the requirements. To this end, we define the interval

$$\mathcal{I}^1_{\epsilon}(\mathbf{p}_i, r_j) = \Big[\tilde{Q}_h(\mathbf{p}_i, r_j) - s(\tilde{\epsilon}_{\text{sc}}(\mathbf{p}_i, r_j) + \tilde{\epsilon}_{\text{fe}}(\mathbf{p}_i, r_j)), \tilde{Q}_h(\mathbf{p}_i, r_j) + s(\tilde{\epsilon}_{\text{sc}}(\mathbf{p}_i, r_j) + \tilde{\epsilon}_{\text{fe}}(\mathbf{p}_i, r_j))\Big],$$

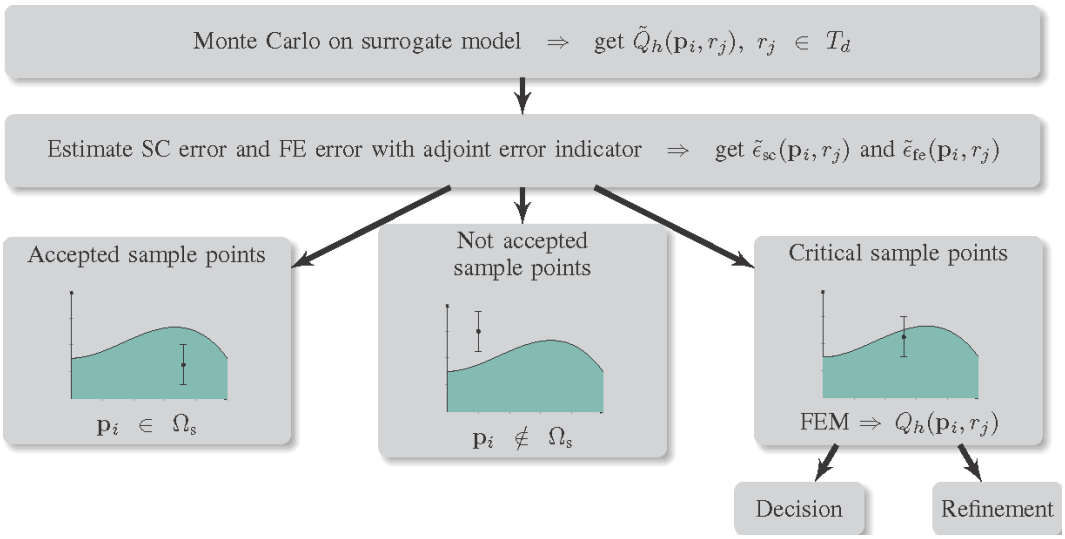

FIG. 1: Scheme of the hybrid approach

where $s \geq 1$ indicates a safety factor. If the performance feature specifications are fulfilled (or not fulfilled) for the whole interval $\mathcal{I}_\epsilon^1$, we can classify the sample point $\mathbf{p}_i$ as accepted (or not accepted). If the performance feature specifications are fulfilled only for a subset of the interval $\mathcal{I}_\epsilon^1$, we classify the sample point as critical.

For all critical sample points the high-fidelity FE model will be evaluated; hence, we obtain $Q_h(\mathbf{p}_i, r_j)$. For these points, the surrogate error is zero, however, the FE error remains unchanged. The new interval we have to examine is given by

$$\mathcal{I}_\epsilon^2(\mathbf{p}_i, r_j) = [Q_h(\mathbf{p}_i, r_j) - s(0 + \tilde{\epsilon}_{\text{fe}}(\mathbf{p}_i, r_j)), Q_h(\mathbf{p}_i, r_j) + s(0 + \tilde{\epsilon}_{\text{fe}}(\mathbf{p}_i, r_j))].$$

Applying the same rules as above, the sample points are again classified either as *accepted* or *not accepted*. If the sample point is not identified as critical, we continue with the next sample point. Otherwise, we refine the mesh of the FE model and reevaluate $Q_h(\mathbf{p}_i, r_j)$ and the FE error $\epsilon_{\text{fe}}^h(\mathbf{p}_i, r_j)$. We continue this procedure until the sample point is not critical anymore or a maximal number of refinement steps is reached. In this manner we obtain an accuracy comparable to the pure MC approach, using the finest refinement. The only difference would occur for sample points whose errors were greatly underestimated with the adjoint error indicators and which were therefore wrongly accepted or rejected instead of being classified as critical sample points. The decision process for one sample point $\mathbf{p}_i$ in one range parameter point r_j is reported in Algorithm 1.

The following paragraph is dedicated to the choice of the safety factor s. The FE error indicator $\tilde{\epsilon}_{\text{fe}}(\mathbf{p}_i, r_j)$ and the SC error indicator $\tilde{\epsilon}_{\text{sc}}(\mathbf{p}_i, r_j)$ defined in Section 3.2 are not strict upper bounds. Therefore, we introduce the safety factor. To determine the size of the safety factor, we generate a small random sample and evaluate it on the surrogate model $\tilde{Q}_h(\mathbf{p}_i, r_j)$ and on the original model $Q_h(\mathbf{p}_i, r_j)$ (with the finest mesh examined). Next we consider the maximum of the ratios of $\tilde{\epsilon}_{\text{fe}}(\mathbf{p}_i, r_j) + \tilde{\epsilon}_{\text{sc}}(\mathbf{p}_i, r_j)$ and $|Q_h(\mathbf{p}_i, r_j) - \tilde{Q}_h(\mathbf{p}_i, r_j)|$ to derive the safety factor. As with the computation of the total error (11), we choose the safety factor rather conservatively. This may result in too many sample points being classified as critical, thus increasing the computational effort of the hybrid approach. However, it avoids the misclassification of sample points and thus leads to a higher accuracy. Here, the safety factor has been set to $s = 2$.

The performance feature specifications have to be fulfilled for all $r \in T_r$, or at least for all test range parameter points $r_j \in T_{\text{d}}$. Thus, if one sample point $\mathbf{p}_i$ fulfills the requirements for a specific range parameter point, the test needs to be carried out for the remaining range parameter points as well. However, if $\mathbf{p}_i$ fails to fulfill the requirements for a single arbitrary range parameter point, it is immediately classified as *not accepted*. Thereby, we can avoid the computational effort of evaluating the remaining range parameter points. This strategy is also applied for the standard MC method and the SC surrogate-based MC method. In the hybrid method we can further benefit from the fact that

Algorithm 1: Hybrid decision

Input: sample point $\mathbf{p}_i$, range parameter point r_j, safety factor s
Evaluate surrogate model and set
  $Q = \tilde{Q}_h(\mathbf{p}_i, r_j)$
  $\epsilon = \tilde{\epsilon}_{\text{sc}}(\mathbf{p}_i, r_j) + \tilde{\epsilon}_{\text{fe}}(\mathbf{p}_i, r_j)$
while max. refinement not reached **do**
  if $Q - s\,\epsilon > c$ **then**
    classify $\mathbf{p}_i$ as *not accepted*, i.e., $\mathbf{p}_i \notin \Omega_{\text{s}}$ (middle picture in Fig. 1)
    continue with next sample point $\mathbf{p}_{i+1}$
  else if $Q + s\,\epsilon \leq c$ **then**
    sample point $\mathbf{p}_i$ accepted for this range parameter point r_j
    if all r_j checked **then**
      classify $\mathbf{p}_i$ as *accepted*, i.e., $\mathbf{p}_i \in \Omega_{\text{s}}$ (left picture in Fig. 1)
      continue with next sample point $\mathbf{p}_{i+1}$
    else
      check next range parameter point r_{j+1}
    end if
  else
    sample point $\mathbf{p}_i$ is *critical*
    if first loop **then**
      Evaluate FE model and set
        $Q = Q_h(\mathbf{p}_i, r_j)$
        $\epsilon = \tilde{\epsilon}_{\text{fe}}(\mathbf{p}_i, r_j)$
    else
      Refine the mesh with $h = h/2$
      Evaluate FE model and set
        $Q = Q_h(\mathbf{p}_i, r_j)$
        $\epsilon = \tilde{\epsilon}^h_{\text{fe}}(\mathbf{p}_i, r_j)$
    end if
  end if
end while
if sample point $\mathbf{p}_i$ still *critical* with last refinement **then**
  classify $\mathbf{p}_i$ according to Q with the finest mesh into *accepted* or *not accepted*
end if

we know the SC results for the QoI. As the performance feature specification is defined as an upper bound, we assume that for larger values of $\tilde{Q}_h(\mathbf{p}_i, r_j)$ it is more likely that $Q_h(\mathbf{p}_i, r_j)$ does not fulfill the requirements. Hence, we order the range parameter points according to $\tilde{Q}_h(\mathbf{p}_i, r_j)$ and start examining the range parameter point satisfying

$$\arg\max_{r_j \in T_{\text{d}}} \tilde{Q}_h(\mathbf{p}_i, r_j).$$

In total, three different errors have to be considered within the yield estimation process: the MC error, the FE error, and the error of the surrogate model, in our case the SC error. The hybrid approach proposed in this paper takes into account the surrogate and FE error. The FE error depends on the refinement of the mesh. Instead of evaluating the entire MC sample (or all critical sample points in a hybrid approach) with the finest mesh, we start with a coarse mesh, calculate the error indicator and refine the mesh if necessary. Thereby, the FE error is controlled and reduced if required and unnecessary computational effort avoided. The SC error is controlled by calculating an adjoint error indicator after building the surrogate model. If the sum of both indicators is too large, a sample point may be classified as critical. In this case, we evaluate the FE model and the associated SC error vanishes. In order to control the MC

error, we define a target accuracy by a maximum value of the standard deviation σ_Y and determine the minimum sample size needed by Eq. (7).

4. YIELD OPTIMIZATION

4.1 General Newton Approach

The idea of yield optimization is to change the mean value of the uncertain parameter, i.e., $\overline{\mathbf{p}}$, in order to maximize the yield. We can formulate the optimization problem as follows:

$$\max_{\overline{\mathbf{p}}} Y(\overline{\mathbf{p}}) = \max_{\overline{\mathbf{p}}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbb{I}_{\Omega_s}(\mathbf{p})\, \mathrm{pdf}_{\mathcal{N}(\overline{\mathbf{p}},\boldsymbol{\Sigma})}(\mathbf{p})\, \mathrm{d}\mathbf{p}.$$

Let the uncertain parameter $\mathbf{p}$ be modeled as a normally distributed random variable. Then, since only the probability density function of the uncertain parameter $\mathbf{p}$ depends on the optimization variable $\overline{\mathbf{p}}$, from Eq. (5) we can derive the gradient and the Hessian of the yield according to [1]. To this end, we first introduce the mean and covariance of $\mathbf{p}$, conditional to the event $\mathbf{p} \in \Omega_s$, given as

$$\begin{aligned}
\overline{\mathbf{p}}_{\Omega_s} &= \mathbb{E}_{\mathrm{pdf}_{\Omega_s}}[\mathbf{p}] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbf{p}\, \mathrm{pdf}_{\Omega_s}(\mathbf{p})\, \mathrm{d}\mathbf{p}, \\
\boldsymbol{\Sigma}_{\Omega_s} &= \mathbb{E}_{\mathrm{pdf}_{\Omega_s}}\left[(\mathbf{p} - \overline{\mathbf{p}}_{\Omega_s})\,(\mathbf{p} - \overline{\mathbf{p}}_{\Omega_s})^{\mathrm{T}}\right] \\
&= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (\mathbf{p} - \overline{\mathbf{p}}_{\Omega_s})\,(\mathbf{p} - \overline{\mathbf{p}}_{\Omega_s})^{\mathrm{T}}\, \mathrm{pdf}_{\Omega_s}(\mathbf{p})\, \mathrm{d}\mathbf{p},
\end{aligned}$$

where

$$\mathrm{pdf}_{\Omega_s}(\mathbf{p}) = \frac{1}{Y(\overline{\mathbf{p}})}\, \mathbb{I}_{\Omega_s}(\mathbf{p})\, \mathrm{pdf}_{\mathcal{N}(\overline{\mathbf{p}},\boldsymbol{\Sigma})}(\mathbf{p}).$$

These conditional moments can be estimated by

$$\begin{aligned}
\hat{\overline{\mathbf{p}}}_{\Omega_s} &= \frac{1}{N_{\Omega_s}} \sum_{i=1}^{N_{\mathrm{MC}}} \mathbb{I}_{\Omega_s}(\mathbf{p}_i)\, \mathbf{p}_i, \\
\hat{\boldsymbol{\Sigma}}_{\Omega_s} &= \frac{1}{N_{\Omega_s} - 1} \sum_{i=1}^{N_{\mathrm{MC}}} \mathbb{I}_{\Omega_s}(\mathbf{p}_i)\, \left(\mathbf{p}_i - \hat{\overline{\mathbf{p}}}_{\Omega_s}\right)\left(\mathbf{p}_i - \hat{\overline{\mathbf{p}}}_{\Omega_s}\right)^{\mathrm{T}},
\end{aligned}$$

where $\mathbf{p}_i, i = 1, ..., N_{\mathrm{MC}}$ are independent observations of the random variable $\mathbf{p}$ and N_{Ω_s} indicates the number of sample points within the safe domain. Using these formulations, the gradient and the Hessian of the yield with respect to $\overline{\mathbf{p}}$ can be written as

$$\nabla_{\overline{\mathbf{p}}} Y(\overline{\mathbf{p}}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbb{I}_{\Omega_s}(\mathbf{p})\, \nabla_{\overline{\mathbf{p}}} \mathrm{pdf}_{\mathcal{N}(\overline{\mathbf{p}},\boldsymbol{\Sigma})}(\mathbf{p})\, \mathrm{d}\mathbf{p} = Y(\overline{\mathbf{p}})\, \boldsymbol{\Sigma}^{-1}\, (\overline{\mathbf{p}}_{\Omega_s} - \overline{\mathbf{p}}) \tag{12}$$

$$\begin{aligned}
\nabla^2_{\overline{\mathbf{p}}} Y(\overline{\mathbf{p}}) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbb{I}_{\Omega_s}(\mathbf{p})\, \nabla^2_{\overline{\mathbf{p}}} \mathrm{pdf}_{\mathcal{N}(\overline{\mathbf{p}},\boldsymbol{\Sigma})}(\mathbf{p})\, \mathrm{d}\mathbf{p} \\
&= Y(\overline{\mathbf{p}})\, \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\Sigma}_{\Omega_s} + (\overline{\mathbf{p}}_{\Omega_s} - \overline{\mathbf{p}})\,(\overline{\mathbf{p}}_{\Omega_s} - \overline{\mathbf{p}})^{\mathrm{T}} - \boldsymbol{\Sigma}\right) \boldsymbol{\Sigma}^{-1}.
\end{aligned} \tag{13}$$

A detailed derivation can be found in [1]. It should be mentioned that we first differentiate and then discretize. Hence, this gradient does not necessarily coincide with the gradient obtained by differentiating after discretization.

The fact that we have given the gradient and the Hessian in analytical form allows us to use a Hessian based optimization algorithm, such as the globalized Newton method [17] as proposed in [1]. A pseudo code is given in Algorithm 2. The associated parameters have been set as follows:

$$\beta = \frac{1}{2}, \;\; \gamma = \frac{1}{100}, \;\; \varphi_1 = \varphi_2 = 10^{-6}, \;\; q = \frac{1}{10}.$$

Algorithm 2: Globalized Newton method

1: **Input:** Starting point $\overline{\mathbf{p}}^0 \in \mathbb{R}^{N_\mathrm{P}}$, $\beta \in (0,1)$, $\gamma \in (0,1)$, $\varphi_1, \varphi_2 > 0$, $q > 0$
2: **Output:** Optimal solution $\overline{\mathbf{p}}^\star$
3: **while** $\nabla Y\big(\overline{\mathbf{p}}^k\big) \neq 0$ **and** $\big\|\overline{\mathbf{p}}^k - \overline{\mathbf{p}}^{k-1}\big\| > 0$ **do**
4: Calculate $\mathbf{d}^k$ by solving Newton's equation $\nabla^2 Y\big(\overline{\mathbf{p}}^k\big)\mathbf{d}^k = -\nabla Y\big(\overline{\mathbf{p}}^k\big)$.
5: **if** "Calculation of $\mathbf{d}^k$ possible" **and** $-\nabla Y\big(\overline{\mathbf{p}}^k\big)\mathbf{d}^k \geq \min\big(\varphi_1, \varphi_2\big\|\mathbf{d}^k\big\|^q\big)\big\|\mathbf{d}^k\big\|^2$ **then**
6: Set search direction $\mathbf{s}^k = \mathbf{d}^k$.
7: **else**
8: Set search direction $\mathbf{s}^k = -\nabla Y\big(\overline{\mathbf{p}}^k\big)$.
9: **end if**
10: Determine step size with Armijo rule, i.e., search for largest $\sigma^k \in \big\{\beta^0, \beta^1, \beta^2, ...\big\}$ such that: $Y\big(\overline{\mathbf{p}}^k + \sigma^k\mathbf{s}^k\big) - Y\big(\overline{\mathbf{p}}^k\big) \leq \sigma^k\gamma\nabla Y\big(\overline{\mathbf{p}}^k\big)^{\mathrm{T}}\mathbf{s}^k$.
11: Set $\overline{\mathbf{p}}^{k+1} = \overline{\mathbf{p}}^k + \sigma^k\mathbf{s}^k$ and $k = k+1$.
12: **end while**

In this paper we assume that all uncertain parameters are optimization variables and vice versa. Little modifications in the algorithm also cover other cases. Additional deterministic optimization variables would appear in the indicator function. Thus, the analytical formulation of the gradient (and the Hessian) does not hold. Instead a finite difference approximation can be used or a negligible uncertainty (noise) can be assigned. If, instead, there are uncertain parameters $\mathbf{u}$, which are not optimization variables, they have to be considered during yield estimation, which can be achieved by setting $\mathbf{p}' = [\mathbf{p}, \mathbf{u}]^{\mathrm{T}}$. Nevertheless, during optimization we only use $\mathbf{p}$, e.g., to calculate $\boldsymbol{\Sigma}$, $\overline{\mathbf{p}}_{\Omega_\mathrm{s}}$, $\boldsymbol{\Sigma}_{\Omega_\mathrm{s}}$, etc.

4.2 Adaptive Newton-MC

The size of the MC sample is crucial, not only for accuracy but also for the efficiency of the algorithm. According to Eq. (7), for yield estimation we can use the MC error indicator to determine the sample size depending on the desired accuracy. For yield optimization, the situation is more involved. The accuracy of yield estimators at intermediate steps of the Newton algorithm is not essential to obtain a satisfying final result. In each individual iteration, it is sufficient to obtain a gradient that indicates the right direction. The stochastic gradient approach also deals with approximated or inexact gradients, used during the optimization process; see [35], for example. However, our approach uses more sample points than usual in the stochastic gradient approach, but we also calculate the objective function with the reduced sample. Only towards the termination of the algorithm, a very accurate gradient may be decisive to accurately determine the optimal solution. Our algorithmic construction ensures that the high, predefined, accuracy requirements at the final stages of the algorithm are fullfilled. More precisely, we propose the following adaptive Newton-MC approach. The optimization method is based on a globalized Newton method, as described in Algorithm 2. We start with a very small sample size and proceed with a few *fast* initial Newton iterations. If no further yield improvement is observed during the iteration process, the globalized Newton method described in Algorithm 2 would stop. Here, instead, we increase the number of MC observations until an improved yield is observed or a target accuracy is reached, then we start the next Newton iteration. Only when the target accuracy has been reached and the yield is not improving anymore, the algorithm terminates.

A pseudo code for the adaptive Newton-MC is given in Algorithm 3. First, we need to define a target accuracy in form of a maximal standard deviation $\sigma_{Y,\max}$ for our terminal solution. Furthermore, we have to define the size of the initial MC sample $N_\mathrm{MC}^\mathrm{start}$ and an incremental factor inc > 0 such that

$$N_\mathrm{MC}^\mathrm{new} = N_\mathrm{MC}^\mathrm{old} + \mathrm{inc}\, N_\mathrm{MC}^\mathrm{start}.$$

The sample size is increased until the target accuracy is reached (see line 14 in Algorithm 3), and the standard globalized Newton method terminates because no further yield improvement can be obtained; i.e., the difference between $\overline{\mathbf{p}}^k$ and $\overline{\mathbf{p}}^{k-1}$ tends to zero (see line 15). In line 3 we can see the rules for a sample size increment. This loop is

Algorithm 3: Adaptive Newton-MC

1: **Input:** Starting point $\overline{\mathbf{p}}^0 \in \mathbb{R}^{N_\mathrm{P}}$, max. std. $\sigma_{Y,\max}$, starting sample size $N_\mathrm{MC}^\mathrm{start}$, $\beta \in (0,1)$, $\gamma \in (0,1)$, $\varphi_1, \varphi_2 > 0$, $q > 0$
2: **Output:** Optimal solution $\overline{\mathbf{p}}^\star$
3: **while** $\nabla Y\big(\overline{\mathbf{p}}^k\big) \neq 0$ **and** $\big\|\overline{\mathbf{p}}^k - \overline{\mathbf{p}}^{k-1}\big\| > 0$ **do**
4: Calculate $\mathbf{d}^k$ by solving Newton's equation $\nabla^2 Y\big(\overline{\mathbf{p}}^k\big)\mathbf{d}^k = -\nabla Y\big(\overline{\mathbf{p}}^k\big)$.
5: **if** "Calculation of $\mathbf{d}^k$ possible" **and** $-\nabla Y\big(\overline{\mathbf{p}}^k\big)\mathbf{d}^k \geq \min\big(\varphi_1, \varphi_2\big\|\mathbf{d}^k\big\|^q\big)\big\|\mathbf{d}^k\big\|^2$ **then**
6: Set search direction $\mathbf{s}^k = \mathbf{d}^k$.
7: **else**
8: Set search direction $\mathbf{s}^k = -\nabla Y\big(\overline{\mathbf{p}}^k\big)$.
9: **end if**
10: Determine step size with Armijo rule, i.e., search for largest $\sigma^k \in \big\{\beta^0, \beta^1, \beta^2, \beta^3\big\}$ such that: $Y\big(\overline{\mathbf{p}}^k + \sigma^k\mathbf{s}^k\big) - Y\big(\overline{\mathbf{p}}^k\big) \leq \sigma^k\gamma\nabla Y\big(\overline{\mathbf{p}}^k\big)^\mathrm{T}\mathbf{s}^k$, else set $\sigma^k = \beta^3$.
11: Set $\overline{\mathbf{p}}^{k+1} = \overline{\mathbf{p}}^k + \sigma^k\mathbf{s}^k$ and $k = k+1$.
12: **end while**
13: Calculate standard deviation $\sigma_Y = \sqrt{\frac{Y(1-Y)}{N_\mathrm{MC}}}$
14: **if** $\sigma_Y > \sigma_{Y,\max}$ **then**
15: **while** $\sigma_{Y'} > \sigma_{Y,\max}$ **and** $\big|Y\big(\overline{\mathbf{p}}^k\big) - Y'\big(\overline{\mathbf{p}}^k\big)\big| < \sigma_{Y,\max}$ **do**
16: Increase sample size $N_\mathrm{MC}^\mathrm{new} = N_\mathrm{MC} + \mathrm{inc}\, N_\mathrm{MC}^\mathrm{start}$.
17: Calculate $Y'\big(\overline{\mathbf{p}}^k\big)$ and $\sigma_{Y'}$ with $N_\mathrm{MC}^\mathrm{new}$.
18: Set $N_\mathrm{MC} = N_\mathrm{MC}^\mathrm{new}$.
19: **end while**
20: Set $Y\big(\overline{\mathbf{p}}^k\big) = Y'\big(\overline{\mathbf{p}}^k\big)$.
21: Go back to line 3.
22: **else**
23: Stop with $\overline{\mathbf{p}}^\star = \overline{\mathbf{p}}^k$
24: **end if**

activated, if the two previous mentioned conditions are fulfilled. Then, we increase the sample size stepwise (see line 16) and reevaluate the yield with the new size $Y'(\overline{\mathbf{p}}^k)$ and its new standard deviation $\sigma_{Y'}$ (see line 17). Note that in order to estimate $Y'(\overline{\mathbf{p}}^k)$ it is not necessary to evaluate $N_\mathrm{MC}^\mathrm{new}$ new sample points. Only the inc $N_\mathrm{MC}^\mathrm{start}$ additional points have to be evaluated and can then be fused with the $N_\mathrm{MC}^\mathrm{old}$ old points to obtain the new yield estimator. This procedure is repeated until the new standard deviation $\sigma_{Y'}$ reaches the target accuracy (i.e., $\sigma_{Y'} \leq \sigma_{Y,\max}$) or the improvement of the yield is large enough (i.e., the difference between the actual yield $Y\big(\overline{\mathbf{p}}^k\big)$ and the yield with the increased sampling $Y'\big(\overline{\mathbf{p}}^k\big)$ is larger than the target accuracy $\sigma_{Y,\max}$). In that case we start a new iteration of the Newton algorithm, with updated yield and sample size (see line 21). If the target accuracy is fulfilled after a regular Newton procedure (after line 12), the algorithm terminates (see line 23).

The parameters are chosen as for Algorithm 2; additionally we set the maximal standard deviation, the starting sample size, and the incremental factor as follows:

$$\sigma_{Y,\max} = 0.01, \quad N_\mathrm{MC}^\mathrm{start} = 100, \quad \mathrm{inc} = 1.$$

Another difference in comparison to Algortihm 2 is that we bound the number of Armijo backward steps. If the inequality in line 10 is not fulfilled after three steps, we set $\sigma^k = \beta^3$ and proceed with the next iteration.

5. NUMERICAL RESULTS

We apply the methods for yield estimation and optimization discussed in the previous sections to a benchmark problem in the context of electromagnetic field simulation. In particular, we employ the model of a rectangular waveguide

with a dielectric inset, similarly to the one used in [36]. This model is well suited for validation purposes, as a closed-form solution is available [37]. Additionally, it fulfills the assumption of a smooth input-to-output behavior made in Section 3.2. In the following, we first introduce the problem setting before numerical results for yield estimation as well as yield optimization are presented.

5.1 Problem Setting

Starting from the time-harmonic Maxwell's equation on a computational domain $D \subset \mathbb{R}^3$, one can derive the curl-curl equation,

$$\nabla \times \left(\mu^{-1} \nabla \times \mathbf{E}_\omega\right) - \omega^2 \varepsilon \mathbf{E}_\omega = 0 \quad \text{in } D, \tag{14}$$

to be solved for the electric field phasor $\mathbf{E}_\omega$, where ω denotes the angular frequency, $\mu = \mu_\mathrm{r}\mu_0 \in L^\infty(D)$ the dispersive complex magnetic permeability, and $\varepsilon = \varepsilon_\mathrm{r}\varepsilon_0 \in L^\infty(D)$ the dispersive complex electric permittivity, with the vacuum permeability μ_0 and the relative permeability μ_r, respectively vacuum and relative permittivity ε_0 and ε_r. Further we have assumed the absence of charges and source currents. Relating (14) to the general problem (1) introduced in the beginning, we note that the angular frequency ω corresponds to the range parameter r.

The boundary of the domain D is split into three parts, i.e., $\partial D = \Gamma_\mathrm{PEC} \cup \Gamma_\mathrm{P1} \cup \Gamma_\mathrm{P2}$, since we consider the model of an electric waveguide with two ports $\Gamma_\mathrm{P1}, \Gamma_\mathrm{P2}$ and assume perfect electric conductor (PEC) boundary conditions at the waveguide walls, i.e.,

$$\mathbf{n} \times \mathbf{E}_\omega = 0 \quad \text{on } \Gamma_\mathrm{PEC}. \tag{15}$$

At the waveguide ports $\Gamma_\mathrm{P1}, \Gamma_\mathrm{P2}$ we impose lowest-order waveguide boundary conditions [38, Chapter 8.5],

$$\mathbf{n} \times (\nabla \times \mathbf{E}_\omega) - jk_{z10}(\mathbf{n} \times \mathbf{E}_\omega) \times \mathbf{n} = -2jk_{z10}\mathbf{E}_\omega^\mathrm{inc} \qquad \text{on } \Gamma_\mathrm{P1}, \tag{16a}$$

$$\mathbf{n} \times (\nabla \times \mathbf{E}_\omega) - jk_{z10}(\mathbf{n} \times \mathbf{E}_\omega) \times \mathbf{n} = 0 \qquad \text{on } \Gamma_\mathrm{P2}, \tag{16b}$$

where $\mathbf{n}$ denotes the outer unit normal vector and j the imaginary unit. The propagation constant k_{z10} is given by $k_{z10} = \sqrt{\omega^2\mu_0\varepsilon_0 - (\pi/a)^2}$, where, in turn, a denotes the width of the waveguide, as depicted in Fig. 2. According to [39], the boundary conditions (16) can be derived based on the assumption that the rectangular waveguide is excited at Γ_P1 by an incident TE_{10} wave,

$$\mathbf{E}_\omega^\mathrm{inc} = E_0 \mathbf{E}_{10}^\mathrm{TE} e^{-jk_{z10}z} \quad \text{with } \mathbf{E}_{10}^\mathrm{TE} := \sin\left(\frac{\pi x}{a}\right)\mathbf{e}_y,$$

where E_0 refers to the amplitude of the incident wave and $\mathbf{e}_y$ denotes the unit vector in the y-direction. Additionally it is assumed that the waveguide dimensions are chosen s.t. only the TE_{10} mode is propagating without attenuation, that the ports are placed sufficiently far from any obstacles in the waveguide which might excite higher-order modes, and

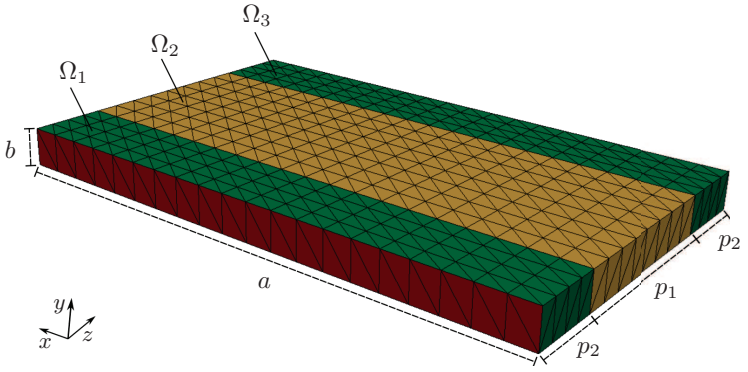

FIG. 2: Finite element model of a rectangular waveguide with dielectric inset of length p_1. The waveguide is excited at the port Γ_P1 by an incident TE_{10} wave.

that the homogeneous material at the ports $\Gamma_{\mathrm{P1}} \cup \Gamma_{\mathrm{P2}}$ fulfills $\epsilon_{\mathrm{r}} = \mu_{\mathrm{r}} = 1$. For further details on waveguide boundary conditions, we refer to [38].

As QoI we consider the fundamental scattering parameter (S-parameter) of the TE_{10} mode on Γ_{P1},

$$S := \frac{2}{E_0 ab}\big(\mathbf{E}_\omega - \mathbf{E}_\omega^{\mathrm{inc}}, \mathbf{E}_{10}^{\mathrm{TE}}\big)_{\Gamma_{\mathrm{P1}}}, \tag{17}$$

where we assumed that Γ_{P1} is located at $z = 0$ for simplicity (without loss of generality). Note that the QoI (17) is, in this case, an affine-linear functional of $\mathbf{E}_\omega$.

5.2 Weak Formulation and Discretization

In order to solve the boundary value problem (14)–(16) numerically by the FEM, we devise the corresponding weak formulation. Therefore, we build the inner products of Eq. (14) with test functions $\mathbf{E}' \in V$, where V is to be determined, and integrate by parts employing [40, Theorem 3.31]

$$\big(\mu_{\mathrm{r}}{}^{-1}\nabla \times \mathbf{E}_\omega, \nabla \times \mathbf{E}'\big)_D - \omega^2 \mu_0 (\varepsilon\, \mathbf{E}_\omega, \mathbf{E}')_D + \big(\pi_{\mathrm{t}}[\mu_{\mathrm{r}}^{-1}\nabla \times \mathbf{E}_\omega], \pi_{\mathrm{T}}[\mathbf{E}']\big)_{\partial D} = 0. \tag{18}$$

Note that we introduced the trace operators

$$\begin{aligned} \pi_{\mathrm{t}}[\mathbf{u}] &:= \mathbf{n} \times \mathbf{u}|_{\partial D} \\ \pi_{\mathrm{T}}[\mathbf{u}] &:= \big(\mathbf{n} \times \mathbf{u}|_{\partial D}\big) \times \mathbf{n} \end{aligned}$$

for brevity of notation. The boundary integral in Eq. (18) vanishes on Γ_{PEC}, since we impose PEC boundary conditions (15) for the test functions $\mathbf{E}'$ as well. On $\Gamma_{\mathrm{P1}} \cup \Gamma_{\mathrm{P2}}$ we employ the boundary conditions (16) and obtain the weak formulation: find $\mathbf{E} \in V$ s.t.

$$\begin{aligned} &\big(\mu_{\mathrm{r}}{}^{-1}\nabla \times \mathbf{E}_\omega, \nabla \times \mathbf{E}'\big)_D - \omega^2 \mu_0 (\varepsilon\, \mathbf{E}_\omega, \mathbf{E}')_D + jk_{z10}(\pi_{\mathrm{T}}[\mathbf{E}_\omega], \pi_{\mathrm{T}}[\mathbf{E}'])_{\Gamma_{\mathrm{P1}} \cup \Gamma_{\mathrm{P2}}} \\ &= 2jk_{z10}\big(\mathbf{E}_\omega^{\mathrm{inc}}, \pi_{\mathrm{T}}[\mathbf{E}']\big)_{\Gamma_{\mathrm{P1}}} \qquad \forall \mathbf{E}' \in V. \end{aligned} \tag{19}$$

The appropriate function space V is a subspace of

$$H(\mathrm{curl}, D) := \Big\{\mathbf{u} \in \big(L^2(D)\big)^3 : (\nabla \times \mathbf{u}, \nabla \times \mathbf{u})_D < \infty\Big\},$$

where, in turn, $\big(L^2(D)\big)^3$ denotes the complex vector function space of square integrable functions, i.e.,

$$\big(L^2(D)\big)^3 := \{\mathbf{u} : (\mathbf{u}, \mathbf{u})_D < \infty\},$$

cf. [40]. To account for the PEC boundary conditions (15) and obtain a well-defined boundary integral in Eq. (19), V is chosen as

$$V := \Big\{\mathbf{u} \in H(\mathrm{curl}, D) :\ \pi_{\mathrm{T}}[\mathbf{u}]\big|_{\Gamma_{\mathrm{P1}}} \in \big(L^2(\Gamma_{\mathrm{P1}})\big)^3 \ \wedge\ \pi_{\mathrm{T}}[\mathbf{u}]\big|_{\Gamma_{\mathrm{P2}}} \in \big(L^2(\Gamma_{\mathrm{P2}})\big)^3 \ \wedge\ \pi_{\mathrm{t}}[\mathbf{u}]\big|_{\Gamma_{\mathrm{PEC}}} = 0\Big\}. \tag{20}$$

In order to solve (19) with FEM, we introduce a finite-dimensional function space $V_h \subset V$ and express the electric field as

$$\mathbf{E}_{\omega,h} = \sum_{j=1}^{N_h} e_{\omega,j}\, \mathbf{N}_j,$$

where $e_{\omega,j} \in \mathbb{C}$ are the degrees of freedom (DoF), N_h is the number of DoFs, and $\mathbf{N}_j \in V_h$ denotes second-order, first-kind Nédélec basis functions defined on a tetrahedral mesh of the domain D. For further details on the

curl-conforming discretization, we refer to [41]. The discrete solution $\mathbf{e}_\omega = [e_{\omega,1}, \dots, e_{\omega,N_h}]^{\mathrm{T}}$ is then obtained by solving the linear system

$$\underbrace{\left(\mathbf{K} - \omega^2 \mathbf{M}^{\varepsilon} + jk_{z10}\mathbf{M}^{\text{port}}\right)}_{\mathbf{A}_\omega} \underbrace{\mathbf{e}_\omega}_{\mathbf{e}_\omega} = \underbrace{\mathbf{f}(\mathbf{e}^{\text{inc}})}_{\mathbf{f}_\omega}, \tag{21}$$

where $\mathbf{A}_\omega \in \mathbb{C}^{N_h \times N_h}$ is the system matrix and $\mathbf{f}_\omega \in \mathbb{C}^{N_h}$ is the discretized right-hand side. The stiffness matrix $\mathbf{K}$, the mass-matrix $\mathbf{M}^{\varepsilon}$, the matrix $\mathbf{M}^{\text{port}}$, and the right-hand side $\mathbf{f}_\omega$ in the above expression are given by

$$\begin{aligned} \mathbf{K}_{ij} &= (\mu_{\mathrm{r}}^{-1} \nabla \times \mathbf{N}_j, \nabla \times \mathbf{N}_i)_D, & \mathbf{M}^{\varepsilon}_{ij} &= \mu_0(\varepsilon \mathbf{N}_j, \mathbf{N}_i)_D, \\ \mathbf{M}^{\text{port}}_{ij} &= (\pi_{\mathrm{T}}[\mathbf{N}_j], \pi_{\mathrm{T}}[\mathbf{N}_i])_{\Gamma_{\mathrm{P1}} \cup \Gamma_{\mathrm{P2}}}, & [\mathbf{f}_\omega]_i &= 2jk_{z10}\big(\mathbf{E}^{\text{inc}}_\omega, \pi_{\mathrm{T}}[\mathbf{N}_i]\big)_{\Gamma_{\mathrm{P1}}}. \end{aligned}$$

The S-parameter can then be obtained from the discrete counterpart of Eq. (17)

$$S_h(\omega) = \big(\mathbf{q}_\omega, \mathbf{e}_\omega - \mathbf{e}^{\text{inc}}_\omega\big)_{\mathbb{C}^{N_h}}.$$

As discussed in the previous sections, we then introduce a parameter vector $\mathbf{p} \in \Xi \subset \mathbb{R}^M$ to account for variations in the design parameters, which, in this case, might represent changes in the domain D or in the material parameters ε, μ. Hence, we obtain the parametrized discrete system

$$\begin{aligned} \mathbf{A}_\omega(\mathbf{p})\, \mathbf{e}_\omega(\mathbf{p}) &= \mathbf{f}_\omega, \\ S_h(\mathbf{p}, \omega) &= \big(\mathbf{q}_\omega, \mathbf{e}_\omega(\mathbf{p})\big)_{\mathbb{C}^{N_h}} - \big(\mathbf{q}_\omega, \mathbf{e}^{\text{inc}}_\omega\big)_{\mathbb{C}^{N_h}}. \end{aligned} \tag{22}$$

We note that the S-parameter is an affine-linear functional in this case, while we only considered linear functionals in Section 3.2 for brevity of notation. However, the method can be straightforwardly adapted to address the constant offset $\big(\mathbf{q}_\omega, \mathbf{e}^{\text{inc}}_\omega\big)_{\mathbb{C}^{N_h}}$ such that the adjoint-based error indicators remain valid.

We proceed with a few details on the implementation of the numerical model. To assemble the linear system (22), we employ the FE library FENICS [42]. As FENICS 2017.2.0 does not support complex numbers, we assemble real and imaginary parts of the matrices separately. We then use NUMPY to impose the PEC boundary condition (15) and SCIPY to solve the resulting linear system of equations with a sparse-LU decomposition. Employing the readily available LU decomposition, the corresponding dual solution $\mathbf{z}_\omega(\mathbf{p})$ can then also be obtained with negligible costs, since the dual problem,

$$\mathbf{A}^{\star}_\omega(\mathbf{p})\mathbf{z}_\omega(\mathbf{p}) = \mathbf{q}_\omega,$$

can again be solved by forward-backward substitution.

5.3 Numerical Results

We consider 12 uncertain parameters

$$\mathbf{p} = [p_1, \dots, p_{12}]^{\mathrm{T}}.$$

Two of them are geometrical parameters given in mm (length of the dielectrical inlay p_1 and length of the vacuum offset p_2) and ten are material parameters with effect on the relative permittivity $\varepsilon_{\mathrm{r}}|_{\Omega_2}$ and permeability $\mu_{\mathrm{r}}|_{\Omega_2}$ on the dielectrical inlay

$$\begin{aligned} \varepsilon_{\mathrm{r}}|_{\Omega_2} &= p_5 + (p_3 - p_5)(1 + j\omega p_6 \tau)^{-1} + (p_4 - p_5)(1 + j\omega p_7 \tau)^{-1}, \\ \mu_{\mathrm{r}}|_{\Omega_2} &= p_{10} + (p_8 - p_{10})(1 + j\omega p_{11} \tau)^{-1} + (p_9 - p_{10})(1 + j\omega p_{12} \tau)^{-1}, \end{aligned}$$

where

$$\begin{aligned} \omega &= 2\pi f, \\ \omega_0 &= 2\pi\big(20 \times 10^9\,\mathrm{Hz}\big), \\ \tau &= \frac{1}{\omega_0} \end{aligned}$$

with frequency f (in Hertz). In order to consider the influence of the number of uncertain parameters, tests with four uncertain parameters are also performed. For this purpose we consider a modified parameter vector,

$$\mathbf{p}^{(\mathrm{m})} = [p_1, p_2, p_{13}, p_{14}]^{\mathrm{T}},$$

where p_1 and p_2 are the geometrical parameters from above, and p_{13} and p_{14} are material parameters with the following effect on relative permeability and permittivity:

$$\varepsilon_{\mathrm{r}}^{(\mathrm{m})}|_{\Omega_2} = 1 + p_{13} + (1 - p_{13})\Big(1 + j\omega\big(2\pi 5 \times 10^9\big)^{-1}\Big)^{-1},$$
$$\mu_{\mathrm{r}}^{(\mathrm{m})}|_{\Omega_2} = 1 + p_{14} + (2 - p_{14})\Big(1 + j\omega\big(1.1 \cdot 2\pi 20 \times 10^9\big)^{-1}\Big)^{-1}.$$

For yield optimization we set the starting point $\overline{\mathbf{p}}_0$ for 12 parameters to

$$\overline{\mathbf{p}}_0 = [9, 5, 2, 0.5, 1, 1, 1.1, 2.5, 1, 1, 1, 2]^{\mathrm{T}}.$$

The estimation tests are done for a reference value $\overline{\mathbf{p}}_{\mathrm{e}}$ close to one optimal solution:

$$\overline{\mathbf{p}}_{\mathrm{e}} = [8.6, 3.8, 2, 0.5, 0.7, 0.6, 1.4, 2.8, 1.7, 0.8, 0.3, 1.4]^{\mathrm{T}}.$$

For the tests with four parameters we set the starting points to

$$\overline{\mathbf{p}}_0^{(\mathrm{m})} = [9, 5, 1, 1]^{\mathrm{T}},$$
$$\overline{\mathbf{p}}_{\mathrm{e}}^{(\mathrm{m})} = [10.36, 4.76, 0.58, 0.64]^{\mathrm{T}}.$$

The standard deviation is set to $\sigma = 0.7^2$ mm for geometrical, and $\sigma = 0.3^2$ for material parameters. In order to avoid unphysical values, instead of a normal distribution we use a truncated normal distribution for the MC sample generation. We truncate with an offset t of $\pm$ 3 mm and $\pm$ 0.3 for the geometrical and material parameters, respectively. The performance feature specifications are

$$|S(\mathbf{p}, \omega)| \overset{!}{\leq} -24\,\mathrm{dB} \;\; \forall \omega \in T_\omega = [2\pi f_1, 2\pi f_2] = [2\pi 6.5, 2\pi 7.5] \text{ in GHz}.$$

Related to the setup of the performance feature specifications in Eq. (4) this means $c = -24\,\mathrm{dB}$ and $Q(\mathbf{p}, r) = |S(\mathbf{p}, \omega)|$ with the frequency ω as range parameter. We consider eleven equidistant frequency points $\omega_j \in T_{\omega_j}$ in the frequency range. The reference solution for yield estimation is

$$Y_{\mathrm{Ref}}(\overline{\mathbf{p}}) = 74.60\,\%,$$

for 12 uncertain parameters, and

$$Y_{\mathrm{Ref}}^{(\mathrm{m})}(\overline{\mathbf{p}}) = 95.44\,\%,$$

for four uncertain parameters. Both reference solutions have been calculated with a closed-form solution of the E-field formulation and standard MC method with $N_{\mathrm{MC}} = 2500$, which is the smallest sample size fulfilling $\sigma_{Y,\max} = 0.01$ for all sizes of the yield, according to Eq. (7).

5.3.1 Quality of the Gradient

As mentioned in Section 4 there is a difference between differentiating or discretizing first. Furthermore, for the sample generation we use a truncated normal distribution instead of a normal distribution. Thus, the gradient we use for optimization deviates from the exact gradient, which can be thought of as an inexact Newton method [43], with approximations in the root-finding problem itself, i.e., here the gradient (12), and the Jacobian which is in our case the Hessian (13).

To ensure that the yield is optimal at the end and no further improvement is possible, an extension can be added to the optimization algorithm. At the optimal solution, the gradient can be calculated with a finite difference quotient $\nabla_{\overline{\mathbf{p}}} Y_{\mathrm{DQ}}$. The gradient from Eq. (12) will be denoted $\nabla_{\overline{\mathbf{p}}} Y_{\mathrm{G}}$, to avoid any confusion. We consider the difference between the two gradients and expect it to be smaller than a constant η,

$$|\nabla_{\overline{\mathbf{p}}} Y_{\mathrm{DQ}}(\overline{\mathbf{p}}) - \nabla_{\overline{\mathbf{p}}} Y_{\mathrm{G}}(\overline{\mathbf{p}})| \leq \eta. \tag{23}$$

Figure 3 compares the two gradients $\nabla_{\overline{\mathbf{p}}} Y_{\mathrm{G}}$ and $\nabla_{\overline{\mathbf{p}}} Y_{\mathrm{DQ}}$ for the waveguide example where the only uncertain design parameter is the length of the inlay p_1. On the left, we see the yield over the parameter p_1, on the right we see the graphs of the gradients over the parameter p_1. For this calculation we set the sample size to $N_{\mathrm{MC}} = 10^6$ and the step size in the difference quotient to $\delta = 10^{-3}$. The two gradients show a similar behavior, especially near the optimum the gradients agree well. Figure 4 shows how the two gradients approach each other for large N_{MC}. Thus, if Eq. (23) is not fulfilled the number of sample points to calculate the gradients can be increased until (23) is fulfilled or an upper bound for N_{MC} is reached. In the former case, the applied gradient $\nabla_{\overline{\mathbf{p}}} Y_{\mathrm{G}}$ is accurate and the optimal solution reliable. In the latter case, a further improvement of the yield would still be possible due to the limited gradient accuracy. In

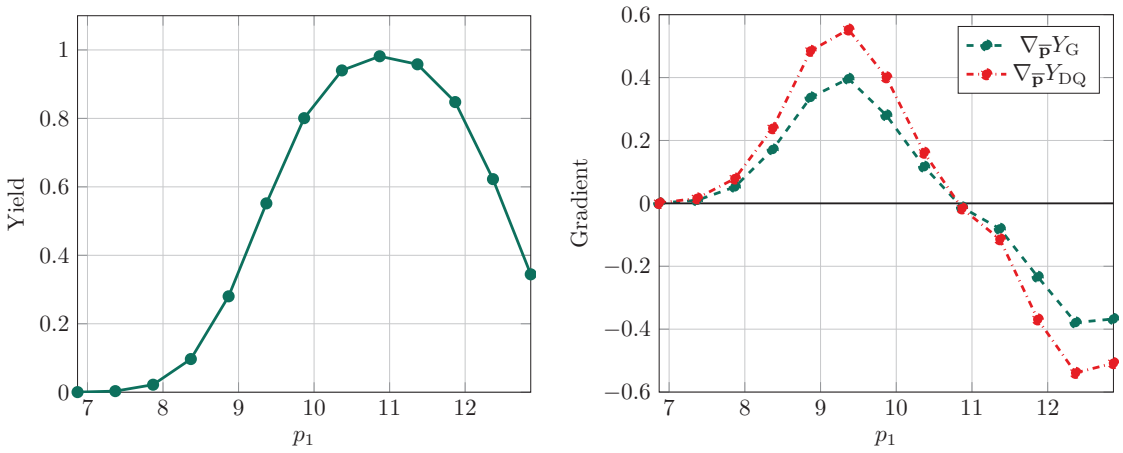

FIG. 3: Comparision of the gradients $\nabla_{\overline{\mathbf{p}}} Y_{\mathrm{G}}$ and $\nabla_{\overline{\mathbf{p}}} Y_{\mathrm{DQ}}$ for $N_{\mathrm{MC}} = 10^6$ and finite difference step size $\delta = 10^{-3}$ for truncated normal distribution

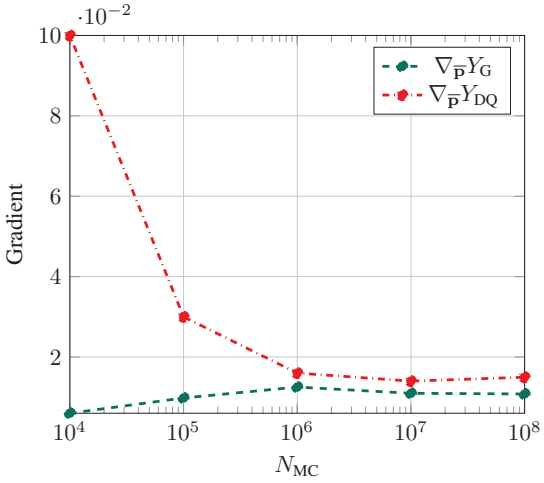

FIG. 4: Convergence of the gradients $\nabla_{\overline{\mathbf{p}}} Y_{\mathrm{G}}$ and $\nabla_{\overline{\mathbf{p}}} Y_{\mathrm{DQ}}$ for increasing N_{MC}. Calculated in $p = 10.8685\,\mathrm{mm}$ and with finite difference step size $\delta = 10^{-3}$.

this case the yield optimization could be continued with the gradient $\nabla_{\overline{\mathbf{p}}} Y_{\mathrm{DQ}}$. However, this would require additional computational effort, especially for a large number of uncertain parameters. The optimal solution can also be used as a starting point for an alternative optimization procedure.

5.3.2 Yield Estimation

We proceed by comparing the proposed hybrid approach with standard MC and a surrogate-based MC approach without hybridization. The surrogate model is constructed based on sparse-grid interpolation as explained in Section 3.2. In order to achieve a high accuracy in the L^∞ norm, we employ, in this work, uniform weight functions w_m in the ranges given by the nominal parameter values $\overline{\mathbf{p}} \pm$ truncation offset t. The comparison is based on both the computational effort and the accuracy. For the accuracy we use the relative error between the reference solution and the solution of the considered method, i.e., for the hybrid approach,

$$\mathrm{err}_{\mathrm{H}} = \frac{|Y_{\mathrm{Ref}} - Y_{\mathrm{H}}|}{Y_{\mathrm{Ref}}}, \tag{24}$$

$\mathrm{err}_{\mathrm{SC}}$, $\mathrm{err}_{\mathrm{MC}}$ for SC and MC, respectively. We measure the computational effort with the number of high-fidelity evaluations (i.e., matrix factorizations in FEM). Here we have different levels of high-fidelity evaluations due to mesh refinement within the proposed hybrid approach. We start with a mesh size h, and if necessary divide it by 2. The difference in the computational effort for each refinement level depends on the model structure and the solver used. Assuming an optimal solver with an effort which is linear in the number of degrees of freedom, the effort increases by a factor of 4 in the case of a 2D problem and by a factor of 8 in the case of a 3D problem. Since in our case the E-field is constant in the y-direction, the grid is only refined in x- and z-directions. Thus, the computational effort of a method is measured through

$$\mathrm{eff} = \#\mathrm{HF}_h + 4\,\#\mathrm{HF}_{h/2} + 16\,\#\mathrm{HF}_{h/4}, \tag{25}$$

which adds up the number of high-fidelity evaluations on the different levels, each multiplied by the factors mentioned above.

The standard approach to carry out yield estimation, with the same accuracy as with the proposed hybrid approach, would be a MC analysis with the finest mesh used within the hybrid approach, referred to as $\mathrm{MC}_{\mathrm{fine}}$. If the mesh refinement strategy is additionally applied, the method is denoted as $\mathrm{MC}_{\mathrm{refine}}$. In order to build the surrogate model both for SC and for the hybrid approach, we use the first grid with mesh size h without further refinement to evaluate the model at the Leja nodes. In Table 1 we see the results of the comparison. We consider two versions of SC, each with different accuracy (number of Leja nodes). The surrogate model used for the hybrid approach is the same as for SC with 90 Leja nodes. For each approach we use the same MC sample as for the reference solution. With the hybrid approach and MC we achieve the same result as with the closed-form reference solution. Out of these three, the hybrid method requires the least computational effort. Compared to $\mathrm{MC}_{\mathrm{refine}}$, we can save 73% computing time compared to $\mathrm{MC}_{\mathrm{fine}}$, even 98%. Comparing the hybrid and $\mathrm{MC}_{\mathrm{refine}}$ approach, we observe that most of the MC

TABLE 1: Comparison of different yield estimation approaches for 12 uncertain parameters. Different methods: MC, SC, and hybrid (H). # Leja indicates the number of Leja nodes for one frequency point ω_j, # HF the number of high-fidelity evaluations to build the surrogate model (surr.), to evaluate (critical) MC samples (MC) with indicated refinement, eff the measurement for computational effort according to Eq. (25), and err the relative error according to Eq. (24)

Approach	# Leja	# HF_h surr.	# HF_h MC	# $\mathrm{HF}_{h/2}$ MC	# $\mathrm{HF}_{h/4}$ MC	eff	err (%)
$\mathrm{MC}_{\mathrm{fine}}$	—	0	0	0	22,705	363,280	0.0000
$\mathrm{MC}_{\mathrm{refine}}$	—	0	22,705	25	6	22,901	0.0000
H	90	990	4812	25	6	**5998**	**0.0000**
SC	90	990	0	0	0	990	6.2235
SC	1600	17,600	0	0	0	17,600	0.4290

sample points are evaluated on the coarsest FE grid. Only for a few points, a refinement of the grid to $h/2$ (25 sample points) or $h/4$ (6 sample points) is necessary. Using the same surrogate model as for the hybrid approach, pure SC has much less computational effort with eff $= 990$, but the error is larger than 6%. Increasing the number of Leja nodes to 1600 results in three times higher computational effort compared to the hybrid approach, with an error still larger than 0.4%.

Table 2 shows the results for the same waveguide with only four uncertain parameters. The statement remains unchanged. However, the influence of the number of parameters can be seen. With four uncertain parameters, also with SC, we can reduce the error to zero, with only about two and a half times the computational effort compared to the hybrid method. The higher the number of uncertain parameters, the more gain can be expected from the hybrid approach compared to a SC method. Compared to $\text{MC}_{\text{refine}}$, with the hybrid approach, we can save almost 98% computing time compared to MC_{fine}, even 99.8%. This means that the advantage of the hybrid approach over MC increases as the number of parameters decreases. Nevertheless, we know by construction, that even for high numbers of uncertain parameters, the hybrid method can never become worse than MC, excluding the computational effort to build the surrogate model (which could scale poorly for a high-dimensional problem) and evaluate the error indicator.

5.3.3 Yield Optimization

We compare the proposed adaptive Newton-MC from Algorithm 3 with the standard Newton method from Algorithm 2, both with the same limited number of Armijo backsteps and the presented hybrid approach for the yield estimation. In both cases we set the target accuracy to $\sigma_{Y,\max} = 0.01$. The adaptive approach starts with 100 sample points and increases this number adaptively until optimality and a target accuracy are achieved. In the nonadaptive approach, we specify a fixed sample size so that the target accuracy is guaranteed at all times during the optimization process. This fixed sample size is $N_{\text{MC}} = 2500$. In Table 3 we see the results for tests with 12 or 4 uncertain parameters. The number of iterations of single yield estimations within the optimization process, the computational effort (eff) and the optimal yield value ($Y^\star$) are given. Note that during the optimization process the surrogate model is only built once, for the starting point. Accepting higher computational effort, the surrogate model can also be recalculated in each iteration step for the current solution, or built at the beginning in a larger interval than $\overline{\mathbf{p}}_0 \pm t$.

With 12 uncertain parameters, we started with a yield of 15%. The adaptive and the nonadaptive approach lead to different local optima with similar yield values. Both take a bit more than 30 iterations. On average, two and a half yield estimations are performed per iteration using the adaptive approach. This is due to multiple evaluations by Armijo backsteps. The nonadaptive approach has only 1.2 estimations per iteration. This can be explained by the fact that the adaptive approach performs several Newton optimizations with different sample sizes one after the other. Shortly before a Newton procedure is terminated, there is usually no further improvement, which is why Armijo backsteps increase and so does the number of yield estimations. This is the case every time before the sample size is increased in the adaptive algorithm. In the nonadaptive approach, this behavior occurs only once at the end. Potential for improvement in the adaptive approach lies in further reducing the number of yield evaluations through smoother

TABLE 2: Comparison of different yield estimation approaches for 4 uncertain parameters. Different methods: MC, SC, and hybrid (H). # Leja indicates the number of Leja nodes for one frequency point ω_j, # HF the number of high-fidelity evaluations to build the surrogate model (surr.), to evaluate (critical) MC samples (MC) with indicated refinement, eff the measurement for computational effort according to Eq. (25) and err the relative error according to Eq. (24)

Approach	# Leja	# HF_h surr.	# HF_h MC	# $\text{HF}_{h/2}$ MC	# $\text{HF}_{h/4}$ MC	eff	err (%)
MC_{fine}	—	0	0	0	26,360	421,760	0.0000
$\text{MC}_{\text{refine}}$	—	0	26,360	5	1	26,396	0.0000
H	30	330	165	5	1	**531**	**0.0000**
SC	30	330	0	0	0	330	0.1257
SC	120	1320	0	0	0	1320	0.0000

transitions from one sample size to the other. Nevertheless, the adaptive approach reduces the computational effort by a factor of 2 compared to standard Newton; see column eff in Table 3. In tests with only 4 uncertain parameters, the computing effort was even reduced to 10%. In this case, the adaptive approach resulted in significantly fewer iterations. The ratio between iterations and yield estimations remains unchanged.

For the case with 4 uncertain parameters we also draw a comparison to standard procedures. Standard procedure means, in this case, combining a standard MC analysis for the yield estimation with a standard Newton method for the optimization. On the coarsest grid (h), 816,816 evaluations with FEM were necessary to optimize the yield; i.e., eff $=$ 816,816. Thus, with the proposed adaptive Newton-MC, a saving of 98.3% in computing effort could already be achieved compared to the standard procedure mentioned above. However, in order to achieve the same accuracy as with the proposed method, the finest grid $(h/4)$ has to be used for all sample points. We assume that the number of function evaluations does not change significantly due to the grid refinement. This can be motivated by the fact that a similar number of iterations, yield estimations, and function evaluations were needed for calculation with the closed-form solution as for the FE model with coarser grid. Under this assumption we obtained an effort factor of eff $\approx 13 \times 10^6$. Thus the saving of computational effort is even 99.9%.

For 12 uncertain parameters, in Table 4 we see how many MC sample points have been used in which iteration. For most of the iterations a low number of MC sample points is sufficient; only in the last iterations we need to expend more computational effort in order to guarantee the predefined target accuracy.

6. CONCLUSION

In this paper we proposed an adaptive method for yield estimation and optimization. For yield estimation we developed a hybrid approach combining reliability and accuracy of a high-fidelity MC analysis and the efficiency of

TABLE 3: Comparision of adaptive and nonadaptive Newton's method with 12 and 4 uncertain parameters: # Leja indicates the number of Leja nodes for one frequency point ω_j, # param. the number of uncertain parameters, optimization the method used, # It the number of iterations, # YE the number of yield estimations, eff the computational effort, and $Y^\star$ the optimal yield value

Estimation	# Leja	# param.	Optimization	# It	# YE	eff	$Y^\star$ (%)
H	90	12	adapt. Newton-MC	33	86	376,073	74.84
H	90	12	Newton	37	42	682,745	78.20
H	30	4	adapt. Newton-MC	12	27	13,716	95.44
H	30	4	Newton	30	34	138,158	97.92

TABLE 4: Progress of yield optimization with adaptive Newton-MC for 12 uncertain parameters. Number of MC sample points for each iteration of the optimization algorithm

Iteration	N_{MC}
0–12	100
13–14	200
15–18	300
19	500
20–22	600
23	900
24–30	1000
31	1800
32–33	1900

surrogate-based techniques such as stochastic collocation. In case the accuracy of the surrogate model is not sufficient, sample points are reevaluated employing the high-fidelity FE model. Mesh refinement is applied if the accuracy of the FE model itself is too low. This guarantees error control while only a very small subset of the MC sample is evaluated based on expensive high-fidelity evaluations. Adjoint error indicators were applied to estimate the errors of the surrogate model and the FE model. For yield optimization we proposed an adaptive Newton-MC method, based on a globalized Newton method. During the optimization process, numerous yield estimations are performed. In order to control the MC error and at the same time save computational effort, we adaptively increase the number of MC sample points used during the optimization. Thus, the adaptive Newton-MC in combination with the hybrid approach allows us to control the FE error, the MC error, and the surrogate error. At the same time it is much more efficient than conventional MC approaches with a standard Newton method. Numerical tests on a dielectrical waveguide confirm the benefits of the presented method. Future research will deal with the transitions in the adaptive Newton-MC when the MC sample size is increased. Furthermore, although we already use a hierarchical model for Monte Carlo analysis within the optimization, we plan to explore a combination of this with a multilevel Monte Carlo approach [3].

ACKNOWLEDGMENTS

This work is supported by the Graduate School CE within the Centre for Computational Engineering at Technische Universität Darmstadt. N. Georg and U. Römer acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) RO4937/1-1. The authors would like to thank two anonymous reviewers for their helpful comments that improved the manuscript.

REFERENCES

1. Graeb, H.E., *Analog Design Centering and Sizing*, Dordrecht: Springer, 2007.
2. Hammersley, J., *Monte Carlo Methods*, New York: Springer Science & Business Media, 2013.
3. Giles, M.B., Multilevel Monte Carlo Methods, *Acta Numer.*, **24**:259–328, 2015.
4. Choi, C.K. and Yoo, H.H., Uncertainty Analysis of Nonlinear Systems Employing the First-Order Reliability Method, *J. Mech. Sci. Technol.*, **26**(1):39–44, 2012.
5. Breitung, K., Asymptotic Approximations for Multinormal Integrals, *J. Eng. Mech.*, **110**(3):357–366, 1984.
6. Gallimard, L., Adaptive Reduced Basis Strategy for Rare-Event Simulations, *Int. J. Numer. Methods Eng.*, **1**:1–20, 2019.
7. Au, S.K. and Beck, J.L., Estimation of Small Failure Probabilities in High Dimensions by Subset Simulation, *Probab. Eng. Mech.*, **16**:263–277, 2001.
8. Bect, J., Li, L., and Vazquez, E., Bayesian Subset Simulation, *SIAM/ASA J. Uncertainty Quantif.*, **5**(1):762–786, 2017.
9. Bogoclu, C. and Roos, D., A Benchmark of Contemporary Metamodeling Algorithms, in *VII Eur. Cong. Comput. Methods in Appl. Sci. Eng.*, Crete, Greece, 2016.
10. Rao, C.R. and Toutenburg, H., *Linear Models: Least Squares and Alternatives*, 2nd ed., New York: Springer, 1999.
11. Rasmussen, C.E. and Williams, C.K., *Gaussian Processes for Machine Learning*, Cambridge: The MIT Press, 2006.
12. Babuška, I., Nobile, F., and Tempone, R., A Stochastic Collocation Method for Elliptic Partial Differential Equations with Random Input Data, *SIAM J. Numer. Anal.*, **45**(3):1005–1034, 2007.
13. Leifsson, L., Du, X., and Koziel, S., Efficient Yield Estimation of Multiband Patch Antennas by Polynomial Chaos-Based Kriging, *Int. J. Numer. Modell. Electron. Networks Dev. Fields*, pp. 1–10, 2020.
14. Nobile, F., Tempone, R., and Webster, C.G., A Sparse Grid Stochastic Collocation Method for Partial Differential Equations with Random Input Data, *SIAM J. Numer. Anal.*, **46**(5):2309–2345, 2008.
15. Li, J. and Xiu, D., Evaluation of Failure Probability via Surrogate Models, *J. Comput. Phys.*, **229**(23):8966–8980, 2010.
16. Butler, T. and Wildey, T., Utilizing Adjoint-Based Error Estimates for Surrogate Models to Accurately Predict Probabilities of Events, *Int. J. Uncertainty Quantif.*, **8**(2):143–159, 2018.
17. Ulbrich, M. and Ulbrich, S., *Nichtlineare Optimierung*, Basel, Switzerland: Birkhäuser, 2012.

18. Babuška, I., Nobile, F., and Tempone, R., A Stochastic Collocation Method for Elliptic Partial Differential Equations with Random Input Data, *SIAM J. Numer. Anal.*, **45**(3):1005–1034, 2007.
19. Chkifa, A., Cohen, A., and Schwab, C., Breaking the Curse of Dimensionality in Sparse Polynomial Approximation of Parametric PDEs, *J. Math. Pures Appl.*, **103**(2):400–428, 2015.
20. Papaioannou, I., Daub, M., Drieschner, M., Duddeck, F., Ehre, M., Eichner, L., Eigel, M., Götz, M., Graf, W., and Grasedyck, L., Assessment and Design of an Engineering Structure with Polymorphic Uncertainty Quantification, *GAMM*, **42**(2):e201900009, 2019.
21. Adams, B., Bauman, L., Bohnhoff, W., Dalbey, K., Ebeida, M., Eddy, J., Eldred, M., Hough, P., Hu, K., and Jakeman, J., Dakota: A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, etc.: Version 6.0 User's Manual, Sandia Technical Report SAND2014-4633, 2015.
22. Caflisch, R.E., Monte Carlo and Quasi-Monte Carlo Methods, *Acta Numer.*, **7**:1–49, 1998.
23. Xiu, D. and Karniadakis, G.E., The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations, *SIAM J. Sci. Comput.*, **24**(2):619–644, 2002.
24. Xiu, D. and Hesthaven, J.S., High-Order Collocation Methods for Differential Equations with Random Inputs, *SIAM J. Sci. Comput.*, **27**(3):1118–1139, 2005.
25. Bungartz, H.J. and Griebel, M., Sparse Grids, *Acta Numur.*, **13**:147–269, 2004.
26. Georg, N., Loukrezis, D., Römer, U., and Schöps, S., Enhanced Adaptive Surrogate Models with Applications in Uncertainty Quantification for Nanoplasmonics, *Int. J. Uncertainty Quantif.*, **10**(2):165–193, 2020.
27. Narayan, A. and Jakeman, J.D., Adaptive Leja Sparse Grid Constructions for Stochastic Collocation and High-Dimensional Approximation, *SIAM J. Sci. Comput.*, **36**(6):A495–A521, 2014.
28. Butler, T., Constantine, P., and Wildey, T., A Posteriori Error Analysis of Parameterized Linear Systems Using Spectral Methods, *SIAM J. Matrix Anal. Appl.*, **33**(1):195–209, 2012.
29. Butler, T., Dawson, C., and Wildey, T., Propagation of Uncertainties Using Improved Surrogate Models, *SIAM/ASA J. Uncertainty Quantif.*, **1**(1):164–191, 2013.
30. Jakeman, J.D. and Wildey, T., Enhancing Adaptive Sparse Grid Approximations and Improving Refinement Strategies Using Adjoint-Based a Posteriori Error Estimates, *J. Comput. Phys.*, **280**:54–71, 2015.
31. Becker, R. and Rannacher, R., An Optimal Control Approach to a Posteriori Error Estimation in Finite Element Methods, *Acta Numer.*, **10**:1–102, 2001.
32. Eriksson, K., Estep, D., Hansbo, P., and Johnson, C., Introduction to Adaptive Methods for Differential Equations, *Acta Numer.*, **4**:105–158, 1995.
33. Zienkiewicz, O.C. and Zhu, J.Z., The Superconvergent Patch Recovery and a Posteriori Error Estimates. Part 1: The Recovery Technique, *Int. J. Numer. Methods Eng.*, **33**(7):1331–1364, 1992.
34. Bellman, R., *Curse of Dimensionality, Adaptive Control Processes: A Guided Tour*, Princeton, NJ: Princeton University Press, 1961.
35. Geiersbach, C. and Wollner, W., A Stochastic Gradient Method with Mesh Refinement for PDE Constrained Optimization under Uncertainty, *Math. Opt. Control*, arXiv:1905.08650, 2019.
36. Loukrezis, D., Römer, U., and De Gersem, H., Assessing the Performance of Leja and Clenshaw-Curtis Collocation for Computational Electromagnetics with Random Input Data, *Int. J. Uncertainty Quantif.*, **9**(1):33–57, 2019.
37. Loukrezis, D., Adaptive Approximations for High-Dimensional Uncertainty Quantification in Stochastic Parametric Electromagnetic Field Simulations, PhD, Technische Universität, 2019.
38. Jin, J.M., *The Finite Element Method in Electromagnetics*, New York: Wiley & Sons, 2015.
39. Jackson, J.D., *Classical Electrodynamics*, 3rd ed., New York: Wiley & Sons, 1998.
40. Monk, P., *Finite Element Methods for Maxwell's Equations*, Oxford: Oxford University Press, 2003.
41. Nédélec, J.C., Mixed Finite Elements in R3, *Numer. Math.*, **35**(3):315–341, 1980.
42. Alnæs, M.S., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M.E., and Wells, G.N., The FEniCS Project Version 1.5, *Arch. Numer. Software*, **3**(100):9–23, 2015.
43. Dembo, R.S., Eisenstat, S.C., and Steihaug, T., Inexact Newton Methods, *SIAM J. Numer. Anal.*, **19**(2):400–408, 1982.