Jinming Wan
Department of Systems Science
and Industrial Engineering,
State University of New York at Binghamton,
4400 Vestal Pkwy E,
Binghamton, NY 13902
e-mail: jwan8@binghamton.edu

Yiming Che
Department of Systems Science
and Industrial Engineering,
State University of New York at Binghamton,
4400 Vestal Pkwy E,
Binghamton, NY 13902
e-mail: yche1@binghamton.edu

Zimo Wang
Assistant Professor
Department of Systems Science
and Industrial Engineering,
State University of New York at Binghamton,
4400 Vestal Pkwy E,
Binghamton, NY 13902
e-mail: zimowang@binghamton.edu

Changqing Cheng[1]
Assistant Professor
Department of Systems Science
and Industrial Engineering,
State University of New York at Binghamton,
4400 Vestal Pkwy E,
Binghamton, NY 13902
e-mail: ccheng@binghamton.edu

Uncertainty Quantification and Optimal Robust Design for Machining Operations

In this study, we carry out robust optimal design for the machining operations, one key process in wafer polishing in chip manufacturing, aiming to avoid the peculiar regenerative chatter and maximize the material removal rate (MRR) considering the inherent material and process uncertainty. More specifically, we characterize the cutting tool dynamics using a delay differential equation (DDE) and enlist the temporal finite element method (TFEM) to derive its approximate solution and stability index given process settings or design variables. To further quantify the inherent uncertainty, replications of TFEM under different realizations of random uncontrollable variables are performed, which however incurs extra computational burden. To eschew the deployment of such a crude Monte Carlo (MC) approach at each design setting, we integrate the stochastic TFEM with a stochastic surrogate model, stochastic kriging, in an active learning framework to sequentially approximate the stability boundary. The numerical result suggests that the nominal stability boundary attained from this method is on par with that from the crude MC, but only demands a fraction of the computational overhead. To further ensure the robustness of process stability, we adopt another surrogate, the Gaussian process, to predict the variance of the stability index at unexplored design points and identify the robust stability boundary per the conditional value at risk (CVaR) criterion. Therefrom, an optimal design in the robust stable region that maximizes the MRR can be identified.
[DOI: 10.1115/1.4055039]

Keywords: uncertainty quantification, robust optimal design, stochastic kriging, Gaussian process, conditional value-at-risk, computational foundations for engineering optimization, data-driven engineering, machine learning

1 Introduction

The ongoing devastating COVID-19 pandemic has triggered turmoil in global supply chains, manifesting in shipping delays, rising costs for raw materials, shortages of finished components and parts, and as well as capacity crunches in transportation networks. In particular, a shortfall in semiconductors throws a wrench into a host of industries, such as the automobile, home appliances, and computer gadgets. In wake of this supply upheaval, automakers have scrambled to stockpile chips and other key parts, in antithesis of the "Just in Time" philosophy [1]. As the coronavirus contagion subsides and we gradually return to normalcy, the demand for chips is roaring up, which further complicates the logistical snags in chip acquisition. Against this backdrop, the scaleup of chip manufacturing has been brought to sharp relief. Quality assurance is the linchpin in scaleup manufacturing for the sought-after chips, which calls for optimal process design.

Remarkably, machining is a fundamental procedure in chip fabrication, including ultra-precision machining and chemical mechanical polishing [2–4]. Also known as material removal or subtractive manufacturing, the unwanted parts from a workpiece or wafer are scraped by cutting tools to achieve the desired surface roughness. The surface roughness here determines the chip performance: while the run-of-the-mill chips are made from the 28 nm wafers, chips in advanced AI settings require the 5 nm ones. One detrimental effect in a variety of machining operations (e.g., turning, drilling, and milling) is the self-exciting dynamics or regenerative chatter

arising from the intrinsic time delay phenomenon, which results in exceedingly fierce tool vibration, premature tool wear/failure, and scrap parts due to unacceptable surface finish. It is noted that a high depth of cut usually induces a large cutting force, which in turn, magnifies the chance of chatter. On the other front, large-scale fabrication enforces a high production rate and hence a high depth of cut. Therefore, optimal design for machining operations hinges on the selection of process parameters (e.g., spindle speed and depth of cut) to achieve high productivity (or equivalently, material removal rate (MRR)) and avert the adverse chatter. Identification of the boundary that separates the stable from unstable design configurations in machining processes is a profound quest.

Craftsman knowledge and trial-and-error tests are traditionally used in machining process design, which, however, incurs tedious efforts and enormous costs [5]. Alternatively, computer simulations have been enlisted to characterize the machining process dynamics. This includes the delay differential equation (DDE) models that capture the quintessential phenomenon of time delay [6,7]. In contrast to ordinary differential equations, solutions to DDEs are of infinite dimensions, and sophisticated numerical approaches, such as the temporal finite element method (TFEM), are carried out to construct the approximate solutions. Thus, chatter avoidance boils down to discerning optimal machine settings that induce stable solutions of the DDEs. In addition, the inherent uncertainty in machining operations (e.g., material and process calibration) begs the question of robust process design. Indeed, a workshop on "Uncertainty in Machining" sponsored by the National Science Foundation in 2010 underscored uncertainty quantification and risk mitigation for machining and related manufacturing operations [8]. More recently, the Bayesian network was utilized to evaluate the impact of the random residual stress on machining performance [9]. The challenge resides in how to efficiently navigate the design space to delineate the contour of the stable regimes, graphically

represented as a stability lobe diagram (SLD), and pinpoint the optimal design that maximizes the MRR in a robust sense considering the underlying uncertainty.

In this exposition, we propose an active learning multi-fidelity approach that integrates a high-fidelity simulation oracle (TFEM here) to solve the DDE model and a low-fidelity surrogate to mimic this simulator. More specifically, stochastic kriging (SK) is used as the surrogate to elucidate the relationship between the stochastic response (here, the stability index) and the design parameters and predict the response distribution at unexplored design points. On the back of such predictive distributions, active learning provides guidance on the sequential batch selection of the most informative design points that will be further evaluated by the high-fidelity TFEM. This considerably trims the computational cost compared to the high-fidelity simulator-only experimental design, and the nominal stability boundary can be attained cost-effectively. To tackle the uncertainty, we also incorporate the Gaussian process (GP) to model the heterogeneous variance of stability index across the design space and predict the variance at unexplored design points. This avoids the crude Monte Carlo (MC) sampling at each design point to derive the variance and further ameliorates the computational overhead. Subsequently, a risk measure, conditional value-at-risk (CVaR), is imposed on each candidate design point to construct the robust stability boundary. Next, design points within the robust stable region will be queried to maximize MRR.

The remainder of this article is organized as follows: related works on stability modeling in machining processes are presented in Sec. 2; the problem description, including the DDE model and TFEM, is illustrated in Sec. 3; the methodology, including sequential design, SK-active learning, Gaussian process, and CVaR are explained in Sec. 4; numerical results are be demonstrated in Sec. 5; and finally the last section concludes this study.

2 Related Work

Regenerative chatter is a major destructive phenomenon in machining operations, the avoidance of which via optimal design is indispensable for quality assurance and scaleup manufacturing. Numerical analytical approaches [10–12] have long been sought after to predict the regenerative chatter, including the finite element method (FEM) and Nyquist plots. In FEM, the governing equations of system dynamics are discretized into multiple elements to predict the behavior of cutting tools or other components in the design stage [13,14]. Nyquist plots are commonly used in the frequency response of the dynamical system, and the stability boundary is derived by repeatedly adjusting the cutting conditions (i.e., depth of cut or spindle speed) in the feasible range [15]. The numerical analytical approaches can also be classified into time and frequency domain methods. Landers and Ulsoy [16] adopted time domain simulation (TDS) with a nonlinear force model to track the evolution of the cutting tool dynamics in a face-turning process. Whereas the TDS is a straightforward simulation approach in chatter analysis, it entails tremendous computation effort, particularly considering a large design space. Greis et al. [17] applied the frequency domain stability analysis to identify the SLD by integrating physical experimental study and numerical analysis. In the frequency domain stability analysis reported in Ref. [18], Fourier transformation was implemented to transform the time domain force signal into the frequency domain displacement-to-force ratio signal to determine the analytical stability limit [18].

With the recent leap forward in sensing techniques to monitor the machining processes, data-driven methods have been extensively investigated. For instance, Khasawneh and Munch applied topological data analysis on vibration signals from a turning process to predict chatter: the time series was first transformed into a point cloud, and persistent homology was then applied to quantify the topological difference under stable and unstable cuttings [19]. Schmitz et al. [20] extracted the statistical variance of the once-per-revolution sampled audio signal to detect chatter based on the fact that regenerative chatter leads to a larger variance for the audio signal due to

asynchronous motion. Axinte et al. [21] suggested that cutting forces were more sensitive to the burr formation and chatter marks caused by regenerative instability. Denkena et al. [22] treated the identification of SLD as a classification problem and compared the performance of kernel interpolation, support vector machine (SVM), and artificial neural network to pinpoint the stability boundary with training data from simulation and experimental studies. With in situ acceleration signals, continuous learning models, such as SVM, were developed to identify the SLD under different process dynamics [23]. In a similar vein, Friedrich et al. [24] combined reinforcement learning and nearest neighbor classification to extend the SLD obtained at a certain time frame to changing machining dynamics based on in situ vibration signal. Ringgaard et al. [5] formulated a constrained optimization problem to maximize MRR, under the penalty of chatter and forced vibration, and capture the trade-off between robustness and optimality for this optimization problem. Yet, uncertainty was not explicitly included in this study.

To quantify the impact of such uncertainty on machining stability, the probabilistic characterization of the regenerative chatter in turning was studied using both the MC method and the advanced first-order second moment method, and the reliability of stability was determined by comparing the actual depth of cut and the limiting values [25]. Park and Qin [26] applied the edge theorem to predict the stability boundary considering epistemic uncertainty of process parameters for both single- and multiple-degrees-of-freedom milling systems. Totis [27] treated the model parameters as random variables to obtain the probabilistic instead of deterministic stability lobes. A new criterion for system stability based on level curves and gradient of the probabilistic lobes was then applied to identify optimal robust stable cutting conditions. Che and Cheng applied generalized polynomial chaos to quantify the uncertainty propagation from the material properties to process stability via a TFEM solver [6]. Similarly, the fuzzy mathematical theory was adopted to account for the fuzzy stability lobes, each of which was assigned a membership value for the chatter status [28,29]. Löser et al. [28] investigated the robust SLD via the frequency domain approach. The tedious MC method was only implemented at a few key design points for stability assessment, and the solution was only approximated for other design points. However, how to select those critical design points and build the approximation solutions remain elusive.

Following the principle of Bayesian optimization, active learning has gained enormous traction in the literature on optimal engineering design. Active learning aims to select the most informative data points for further investigation and has been adopted in semi-supervised learning for data labeling, particularly in data imbalance settings. To explore the design space in a cost-effective manner, the probability of improvement [30,31], expected improvement (EI) [32,33], lower confidence bound [34], and knowledge gradient [35] have been attempted. For instance, the acquisition of high-dimensional data (e.g., imaging) in advanced manufacturing prevails for process quality monitoring, which demands, however, labor-intensive data annotation. An adaptive weighted uncertainty sampling was investigated in Ref. [36] that reduces the amount of necessary annotations by 20–70% in the data query of additive manufacturing. Che and Cheng [37] integrated active learning and relevance vector machine to sequentially probe the critical design points to locate the decision boundary in stability analysis of power systems. Botcha et al. [38] developed a query-by-committee active learning approach for selecting the next optimal experimental point to increase the prediction accuracy of surface roughness in a grinding process. More recently, deep learning (e.g., convolutional neural network (CNN)) has been adopted in active learning to handle wafer map pattern classification [39]. The CNN platform has been extensively used for spatial map data, and a softmax layer is added to represent the classification uncertainty, which is amenable for active learning. Similarly, Bayesian deep learning has been designed to automatically handle data and model uncertainty in active learning for labeling of high-dimensional image data [40]. Compared to those existing efforts, our research explicitly considers the intrinsic uncertainty at each design point to identify

the nominal decision boundary, and a reliability-based design is further pursued using CVaR criteria.

3 Problem Description

The peculiar time delay effect in machining processes is portrayed in Fig. 1, in which the cutting tool continuously removes material from a rotating workpiece. The dynamics of the cutting tool can be represented via a DDE [7]:

$$\ddot{z}(t) + 2\xi\omega_n^2\dot{z}(t) + \omega_n^2 z(t) = -kb(f_0 - z(t) + z(t - \tau)) \quad (1)$$

Here, the time delay $\tau = 1/\Omega$ is the time period for one revolution of the workpiece rotation determined by the spindle speed Ω (rounds/min). $z(t)$ and $z(t - \tau)$ are the cutting tool displacement at time t and the previous revolution $t - \tau$, respectively. The nominal feed rate $f_0 = 5 \times 10^{-2}$ mm is kept constant in this study. The tunable parameters, depth of cut b, and spindle speed Ω, are deemed as the design variables. For illustration, we choose the range of $b \in [0.0, 0.4]$mm and $\Omega \in [1.5 \times 10^3, 7.5 \times 10^3]$ rounds/min in this study, which covers broad machine settings. Thanks to this intrinsic time delay, the effective feed rate $f_1 = f_0 - z(t) + z(t - \tau)$, consequently, $b(f_0 - z(t) + z(t - \tau))$ signifies the instantaneous cutting area. Thus, $-kb(f_0 - z(t) + z(t - \tau))$ represents the effective cutting force, proportional to the instantaneous cutting area at time t. That said, the phase difference of surface finish at two successive revolutions could amplify the cutting force and magnify fierce tool vibration, also called chatter, which typically results in inferior machined surfaces and premature tool wear. Moreover, with randomness in material and process calibration, we impose a normal distribution on each of the uncontrollable variables, including damping ratio ξ, natural frequency ω_n, and force coefficient k: $\xi \sim \mathbb{N}(0.02, 0.002^2)$, $\omega_n \sim \mathbb{N}(600\pi, (2\pi)^2)$ rounds/s, and $k \sim \mathbb{N}(2 \times 10^{11}, 5^2)$ N/m^2. The distributions of parameters were obtained from our prior investigation, and a comparatively small variability of the force coefficient k was reported [2]. To achieve high productivity, a certain level of MRR is oftentimes desired. Here we use the nominal MRR $= \pi(r^2 - (r - b)^2) \times f_0 \times \Omega$ and r is the radius of the workpiece. Yet, increase in the depth of cut inevitably leads to a high chance of chatter. Hence, it is indispensable to design the machining process to circumvent the chatter and simultaneously maximize the MRR.

The DDE of Eq. (1) does not admit a closed-form solution, and approximate algorithms, including Chebyshev collocation and TFEM, have been extensively adopted in literature to approximate $z(t)$ [41]. To make this article self-contained, we elaborate on the TFEM solution for machining stability analysis. Concretely, the time delay or workpiece rotation period is divided into E elements, such that each element amounts to a time interval of $T_E = \tau/E$. Then for the nth revolution, a polynomial representation for tool displacement in the jth element is given as

$$z(t) = \sum_{i=1}^{4} a_{ji}^n H_i(\rho(t)) \quad (2)$$
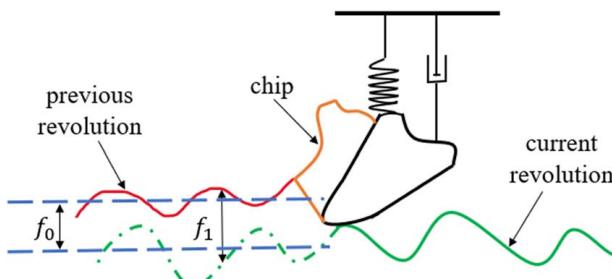
Fig. 1 Schematic diagram of time delay effect in a machining process

where $\rho(t) = t - (n-1)\tau - (j-1)T_E$, and $\rho(t) \in [0, T_E]$ represents local time in the jth element in the nth period. $H_i(\rho)$ are the cubic Hermite polynomial functions [41]:

$$H_1(\rho) = 1 - 3\left(\frac{\rho}{T_E}\right)^2 + 2\left(\frac{\rho}{T_E}\right)^3 \quad (3a)$$

$$H_2(\rho) = T_E\left[\left(\frac{\rho}{T_E}\right) - 2\left(\frac{\rho}{T_E}\right)^2 + \left(\frac{\rho}{T_E}\right)^3\right] \quad (3b)$$

$$H_3(\rho) = 3\left(\frac{\rho}{T_E}\right)^2 - 2\left(\frac{\rho}{T_E}\right)^3 \quad (3c)$$

$$H_4(\rho) = T_E\left[-\left(\frac{\rho}{T_E}\right)^2 + \left(\frac{\rho}{T_E}\right)^3\right] \quad (3d)$$

The coefficient a_{ji} characterizes the displacement $z(t)$ and velocity $\dot{z}(t)$ at the two end nodes of each temporal element (i.e., $\rho = 0$ or $\rho = T_E$). As depicted in Fig. 2 with $E = 2$ temporal elements for demonstration, the upper cuve represents the tool displacement $z(t)$ and the lower curve represents the velocity $\dot{z}(t)$ at time t. a_{11}^{n-1} and a_{12}^{n-1} signify the displacement $z(t)$ and velocity $\dot{z}(t)$ at the starting node ($\rho = 0$) of the first element in the $(n-1)$th revolution; a_{13}^{n-1} and a_{14}^{n-1} represent $z(t)$ and $\dot{z}(t)$ at the ending node ($\rho = T_E$) of the first temporal element in the $(n-1)th$ revolution. Likewise, a_{21}^n and a_{22}^n (a_{23}^n and a_{24}^n) stand for $z(t)$ and $\dot{z}(t)$ at the starting (ending) node for the second temporal element in the nth revolution. As the feasible solution complies with the continuity and boundary conditions, consistent $z(t)$ and $\dot{z}(t)$ at the common node between two adjacent elements are imposed to regulate the TFEM solution, namely, $a_{13}^{n-1} = a_{21}^{n-1}$, $a_{14}^{n-1} = a_{22}^{n-1}$, as well as $a_{23}^{n-1} = a_{11}^n$ and $a_{24}^{n-1} = a_{12}^n$. In the computational implementation, a large E is the requisite for accurate approximation, and we use $E = 100$ in this study.

Next, we plug the TFEM solution Eq. (2) into Eq. (1) to solve the coefficient a_{ji}, and this inevitably leads to the approximation residual. To minimize this approximation residual, all terms on both sides of Eq. (2) are projected onto a set of test functions $\Psi_1(\rho) = 1$ and $\Psi_2(\rho) = \rho/T_E - 1/2$, in a Galerkin projection [42]. Therefore, discretization of the cutting tool dynamics for the jth element in the nth revolution is expressed as

$$\int_0^{T_E}\left\{\left(\sum_{i=1}^4 a_{ji}^n \ddot{H}_i(\rho)\Psi_u(\rho)\right) + 2\xi\omega_n^2\left(\sum_{i=1}^4 a_{ji}^n \dot{H}_i(\rho)\Psi_u(\rho)\right)\right.$$
$$+ \omega_n^2\left(\sum_{i=1}^4 a_{ji}^n H_i(\delta(t))\Psi_u(\rho)\right)\right\}d\rho$$
$$+ \int_0^{T_E} kb\left\{f_0 - \left(\sum_{i=1}^4 a_{ji}^n H_i(\rho)\Psi_u(\rho)\right)\right.$$
$$\left. + \left(\sum_{i=1}^4 a_{ji}^{n-1} H_i(\rho)\Psi_u(\rho)\right)\right\}d\rho = 0, \quad u = 1, 2 \quad (4)$$

In a compact matrix representation, we have

$$\begin{bmatrix} N_{11} & N_{12} & N_{13} & N_{14} \\ N_{21} & N_{22} & N_{23} & N_{24} \end{bmatrix}\begin{bmatrix} a_{j1}^n & a_{j2}^n & a_{j3}^n & a_{j4}^n \end{bmatrix}^T = \begin{bmatrix} D_1 & D_2 \end{bmatrix}^T$$
$$+ \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \end{bmatrix}\begin{bmatrix} a_{j1}^{n-1} & a_{j2}^{n-1} & a_{j3}^{n-1} & a_{j4}^{n-1} \end{bmatrix}^T$$
$$(5)$$

where $N_{ui} = \int_0^{T_E}(\ddot{H}_i(\rho) + 2\xi\omega_n^2\dot{H}_i(\rho) + (\omega_n^2 - kb)H_i(\rho))\Psi_u(\rho)d\rho$, $D_u = \int_0^{T_E}(-kb)\Psi_u(\rho)d\rho$, $P_{ui} = \int_0^{T_E}(-kb)H_i(\rho)\Psi_u(\rho)d\rho$ The two successive revolutions are included in this discretization representation to account for the time delay τ in Eq. (1). Moreover, according
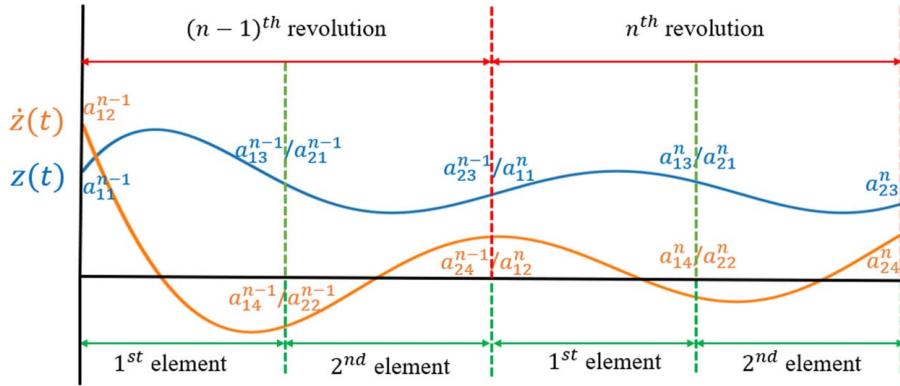
Fig. 2 Illustration of cutting tool displacement and velocity with two temporal elements

to the aforementioned continuity and boundary conditions,

$$\begin{bmatrix} a_{j1}^n & a_{j2}^n \end{bmatrix} = \begin{bmatrix} a_{(j-1)3}^n & a_{(j-1)4}^n \end{bmatrix} \quad (6)$$

$$\begin{bmatrix} a_{11}^n & a_{12}^n \end{bmatrix} = \begin{bmatrix} a_{E3}^{n-1} & a_{E4}^{n-1} \end{bmatrix} \quad (7)$$

Consequently, we assemble the local matrix equation for the jth element, Eq. (5) here, into a global matrix equation for all the E elements:

$$
\begin{bmatrix}
I & 0 & \cdots & \cdots & 0 \\
N_1 & N_2 & \ddots & \ddots & \vdots \\
0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & N_1 & N_2
\end{bmatrix}
\begin{Bmatrix}
\begin{bmatrix} a_{11}^n \\ a_{12}^n \end{bmatrix} \\
\vdots \\
\vdots \\
\begin{bmatrix} a_{E1}^n \\ a_{E2}^n \\ a_{E3}^n \\ a_{E4}^n \end{bmatrix}
\end{Bmatrix}
$$

$$
=
\begin{bmatrix}
0 & \cdots & \cdots & 0 & I \\
P_1 & P_2 & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & P_1 & P_2
\end{bmatrix}
\begin{Bmatrix}
\begin{bmatrix} a_{11}^{n-1} \\ a_{12}^{n-1} \end{bmatrix} \\
\vdots \\
\vdots \\
\begin{bmatrix} a_{E1}^{n-1} \\ a_{E2}^{n-1} \\ a_{E3}^{n-1} \\ a_{E4}^{n-1} \end{bmatrix}
\end{Bmatrix}
$$

$$
+
\begin{Bmatrix}
\begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\begin{bmatrix} D_1 \\ D_2 \end{bmatrix} \\
\vdots \\
\vdots \\
\begin{bmatrix} D_1 \\ D_2 \end{bmatrix}
\end{Bmatrix}
\quad (8)
$$

where the submatrix $N_1 = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}$, $N_2 = \begin{bmatrix} N_{13} & N_{14} \\ N_{23} & N_{24} \end{bmatrix}$,

$P_1 = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$, $P_2 = \begin{bmatrix} P_{13} & P_{14} \\ P_{23} & P_{24} \end{bmatrix}$. In a more compact representation, we have

$$Na_n = Pa_{n-1} + D \quad (9)$$

Denote λ_G as the maximum absolute eigenvalue of the transition matrix $G = N^{-1}P$. According to the finite-dimensional Monodromy operator theory [6], the stability condition of Eq. (1) is $\lambda_G < 1$. This is equivalent to stability index $\lambda = \max(\text{Re}(\log(\lambda_G)))/\tau < 0$, which is also the criterion to gauge stability of the machining process in this study. In the deterministic case, the stability boundary is the contour of $\lambda = 0$ in the design space of $x = [\Omega, b]^T$. Yet, the uncontrollable variables $\zeta = [\xi, \omega_n, k]^T$ render the quest of stability boundary a stochastic experimental design problem. This enforces a large number of replications of TFEM according to $\zeta = [\xi, \omega_n, k]^T$ at each $x = [\Omega, b]^T$ to derive the average response $\lambda(x) = \mathbb{E}_\zeta[\lambda(x, \zeta)]$ and quantify the intrinsic variance $\mathbb{V}(x) = \text{Var}(\lambda(x, \zeta))$, incurring the onerous computational cost. As shown in Sec. 5, $\mathbb{V}(x)$ is heterogeneous in the design space. The objective of this study is to seek the optimal machine settings (the design variables) that can achieve MRR above a certain level and, at the same time, avert the regenerative chatter in a robust sense.

4 Methodology

We adopt the SK surrogate to emulate the TFEM under uncertainty to capture how $\lambda(x)$ varies across x, and an active learning scheme is further developed to seek the trade-off between the high-fidelity TFEM and low-fidelity SK. Thus, the nominal stability boundary, or the contour of $\lambda(x) = 0$, can be sequentially estimated by SK. We note that SK generates a predictive distribution for $\lambda(x)$ as $\hat{\lambda}(x) \sim \mathbb{N}(\bar{\lambda}(x), V(\lambda(x)))$, where $\bar{\lambda}(x)$ is the predictive mean and $V(\lambda(x))$ is the corresponding predictive variance. It bears mentioning that $\mathbb{V}(x)$ and $V(\lambda(x))$ are different. Furthermore, the intrinsic uncertainty $\mathbb{V}(x)$ associated with $\lambda(x, \zeta)$ entails a robust quantification of the nominal stability boundary, which necessitates further quantification of $\mathbb{V}(x)$. To avoid brute force replications at each x to estimate $\mathbb{V}(x)$, we fit a GP surrogate using existing TFEM simulations and MC runs and predict $\mathbb{V}(x)$ for design points near the nominal stability boundary estimated from SK-active learning. A CVaR measure is then imposed on those design points, and a robust boundary can be derived.

The overall flowchart for the optimal design of machining processes which seeks to simultaneously maximize the MRR and eschew the adverse regenerative chatter in a robust sense is depicted in Fig. 3. In this iterative process, design points from the initial slice of batch sampling and their MC replications at uncontrollable variables are evaluated via the TFEM, which constitutes the initial training set for SK. This statistical learning model is then adopted to predict the mean response of the stability index $\lambda(x)$ at design points from the next slice. To ameliorate the prohibitive computational cost associated with TFEM and MC, we leverage the EI criterion in active learning to pick out the most informative design points in this new slice, and only those selected informative design points will be further evaluated by MC-TFEM. The evaluation results will be annexed into the training set. Therefrom, the SK
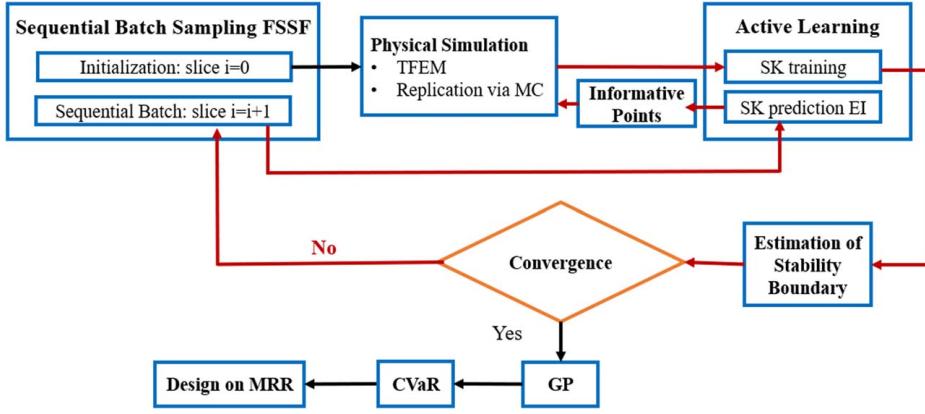
Fig. 3 Flowchart of reliable estimation of optimal design

model along with the estimated stability boundary is updated. This iteration continues until the estimated nominal stability boundary converges in the average sense. To ensure a robust stability boundary, we adopt risk measurement CVaR on design points close to the nominal stability boundary, which entails the valuation of the average response $\lambda(x)$ and variance $\mathbb{V}(x)$ of the stability index $\lambda(x, \zeta)$. GP is utilized to estimate the variance $\mathbb{V}(x)$ for design points near the stability boundary. Therefrom, a grid search is conducted in the robust stability region to seek the design that maximizes the MRR. We note that this active learning approach, which reconciles the high-fidelity stochastic physical simulation and low-fidelity surrogate, only incurs a fraction of the computational cost for the stochastic simulation.

4.1 Fully-Sequential Space-Filling Design. An efficient sequential space-filling sampling algorithm that effectively explores the design space at each iteration is the premise for active learning. Here, we employ a backward fully-sequential space-filling (FSSF-b) design introduced by Shang and Apley [43], which outstrips Sobol sequence and more complex batch-sequential methods, such as the slice Latin hypercube design [44] and nested Latin hypercube design [45] in terms of both minimum pairwise distance and maximum hole size. It starts from a candidate set of the Sobol sequence $C_N = \{x_1, x_2, \ldots, x_N\}$ with N design points, which is a quasi-random fully-sequential sequence to uniformly cover a rectangular design region, albeit with poor space-filling properties. Then one design point is removed from C_N at each stage, and we denote the sequence of indices for the removed points as $\{i_N, i_{N-1}, \ldots, i_1\}$. i_N is the index for the first removed point and i_1 is for the last removed one. At each stage $m = N-1$, $N-2, \ldots, 0$, we have the design points $S_m = \{x_{i_1}, \ldots, x_{i_m}\}$ with index set $I_m = \{i_1, \ldots, i_m\}$. Thus, a sequence of nested design points with size from 1 to N_c are formed as $S_1 \subset S_2 \subset \ldots \subset S_{N_c} \subset C_N$. At the core of FSSF-b, given I_{m+1} and S_{m+1} from stage $m+1$, we seek to find the index $i_{m+1} \in I_{m+1}$ for the next point $x_{i_{m+1}} \in S_{m+1}$ to remove to form $S_m = S_{m+1} \setminus x_{i_{m+1}}$ for each stage m. The criterion to find the index i_{m+1} is the maximum of the minimal pairwise distance among the design S_{m+1}, i.e.,

$$i_{m+1} = \arg \max_{i \in I_{m+1}} d_{\min}(S_{m+1} \setminus x_i) \qquad (10a)$$

$$d_{\min}(S_m) = \min_{\substack{j,k \in I_m \\ j \neq k}} d(x_j, x_k) \qquad (10b)$$

where $d(x_j, x_k) = x_j - x_k$ is the Euclidean distance between point x_j and x_k. The removed points are recorded reversely to explicate the sequential addition of design points. Generally, a large size N for the Sobol sequence is required for achieving a number N_c of space-filling design points, and the relationship is empirically

expressed as $N = 1000 d_m + 2N_c$ according to [43], where d_m is the dimensionality of the design points. Denote the N_c space-filling design points as $B = \{x_{i_1}, \ldots, x_{i_{N_c}}\}$, then the sequential batch sampling is given by design points on each of the n slices as $B_1 = \{x_{i_1}, \ldots, x_{i_\varphi}\}$, $B_2 = \{x_{i_{\varphi+1}}, \ldots, x_{i_{2 \times \varphi}}\}$, $\ldots, B_k = \{x_{i_{(k-1) \times \varphi+1}}, \ldots, x_{i_{k \times \varphi}}\}$, $k = 1, \ldots, n$. Here, $\varphi = N_c/n$ is the number of design points in each slice, and n is the total number of slices. For demonstration purposes, we showcase the procedure of sequential batch sampling in Fig. 4, with $\varphi = 20$.

4.2 Stochastic Kriging. SK is deemed as an extension of GP or kriging with a heterogeneous error term, which implies the intrinsic uncertainty at each design point. Thus, it inherits the extrinsic uncertainty embedded in the response surface as in GP for experimental design and also preserves the intrinsic uncertainty of stochastic simulations [46–48]. Quantification of such intrinsic uncertainty necessitates a number of replications at each design point $x = [\Omega, b]^T$. Therefore, for $x_i = [\Omega_i, b_i]^T$ the simulation output at the jth replication can be estimated as

$$\lambda_j(x_i) = f^T(x_i)\beta + M(x_i) + \varepsilon_j(x_i) \qquad (11)$$

Here, $\lambda_j(x_i) = \lambda(x_i, \zeta_j)$ denotes the stability index from TFEM in the jth replication. f is a predefined basis function that captures the overall trend (e.g., $f(x) = [1, x, x^2, \ldots]^T$), and β represents the basis coefficients. M is a zero-mean Gaussian random field and encapsulates the extrinsic uncertainty, characterized by a covariance function $\mathrm{Cov}(x, x'; L) = \sigma_f^2 \exp\left(-\frac{1}{2}(x - x')L^{-1}(x - x')^T\right)$, where σ_f^2 is the signal variance and L is the diagonal length-scale matrix. $\varepsilon_j(x_i) \sim \mathbb{N}(0, \mathbb{V}(x_i))$. $\mathbb{V}(x_i)$ is the intrinsic variance, and it is independent of the extrinsic uncertainty induced by M. With N_s design points $X = [x_1, \ldots, x_{N_s}]^T$ and n_i replications at x_i, $i = 1, 2, \ldots, N_s$, the response, namely, the average stability index $\lambda(x)$ and the intrinsic variance $\mathbb{V}(x)$, are acquired from TFEM and MC runs:

$$\lambda(x_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} \lambda_j(x_i) \qquad (12a)$$

$$\mathbb{V}(x_i) = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\lambda_j(x_i) - \lambda(x_i))^2 \qquad (12b)$$

Therefrom, the predictive $\hat{\lambda}(x_0)$ at a testing point x_0 is expressed in terms of its first two moments:

$$\bar{\lambda}(x_0) = f^T(x_0)\beta + \Sigma_M^T(x_0, X)[\Sigma_M + \Sigma_\varepsilon]^{-1}(\lambda - \boldsymbol{F}\beta) \qquad (13)$$

$$V(\lambda(x_0)) = \Sigma_M(x_0, x_0) - \Sigma_M^T(x_0, X)[\Sigma_M + \Sigma_\varepsilon]^{-1}\Sigma_M(x_0, X) \qquad (14)$$
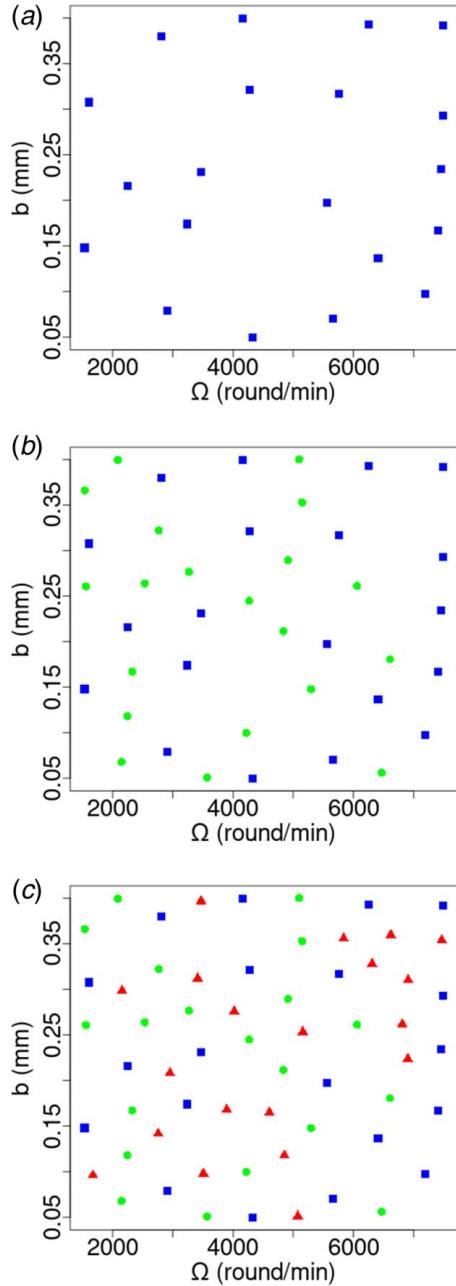
Fig. 4 Illustration of sequential addition of space-filling design points via FSSF-b: 20 sampling points are selected in each slice, indicated by the different shapes: (*a*) the square shapes are the selected sampling points in the first slice, (*b*) dot shapes represent the selected sampling points in the second slice and separate with sampling points of the first slice, and (*c*) the triangle shapes represent the selected sampling points in third slice and separate with sampling points of previous two slices

where $\Sigma_M \in \mathbb{R}^{N_s \times N_s}$ is the extrinsic covariance matrix, with entity $\Sigma_M^{(i,j)} = cov(M(x_i), M(x_j))$, and $\Sigma_M(x_0, X) = [cov(M(x_0), M(x_1)), \ldots, cov(M(x_0), M(x_{N_s}))]^T$. $\Sigma_\varepsilon = diag\{\mathbb{V}(x_i)/n_i\}$ is the intrinsic uncertainty matrix and $\boldsymbol{F} = [f(x_1), \ldots, f(x_{N_s})]^T$. The hyperparameters $\theta = [\beta, \sigma_f^2, L]$ can be trained by the maximum log likelihood [7,49]. It is noteworthy that in SK, we estimate the mean of the stability index $\hat{\lambda}(x_0)$, rather than the stability index itself $\lambda_j(x_i)$. Whereas the variance of $\lambda_j(x_i)$ is irreducible as shown in Eq. (11), entries in the intrinsic variance matrix tend to vanish with a large number of replications.

4.3 Active Learning. Given the initial training data $X = [x_1, \ldots, x_{N_s}]^T$ from Latin hypercube design (LHD) and their response $\lambda(X)$ and $\mathbb{V}(X)$, we obtain the predictive distribution of $\hat{\lambda}(x)$ according to Eqs. (13) and (14) from SK. The nominal stability boundary is the contour of $\lambda(x) = 0$. Intuitively, design points far away from this nominal boundary barely contribute to refining the estimation. Therefore, active learning is promising to single out the most informative design points that will be further evaluated by the simulation oracle or the TFEM simulator with MC runs in this study. We adopt the EI criterion on estimation for sampling points. Specifically, new sampling points in each FSSF-b slice are fed into SK to estimate their response $\bar{\lambda}(x)$ and the uncertainty $V(\lambda(x))$, and then only the most informative ones according to the EI criterion will be further evaluated by the TFEM and MC runs to obtain $\lambda(x)$ and $\mathbb{V}(x)$. Those selected points along with their responses will be annexed into the training set to update the SK model and the nominal boundary. The EI criterion [47,50,51] involves a utility function at a new sampling point x_0:

$$I(x_0|X) = \max(0, \gamma_{\min} - \gamma_e) \qquad (15)$$

where $\gamma_{\min} = \min(|\lambda_0 - \lambda|)$, $\gamma_e = |\lambda_0 - \hat{\lambda}(x_0)|$, and $\lambda_0 = 0$ is the target value. $\lambda = [\lambda(x_1), \ldots, \lambda(x_{N_s})]^T$ signifies the set of $\lambda(x)$ computed from TFEM. $\hat{\lambda}(x_0)$ is the estimated stability index for input x_0 from SK. Considering the uncertainty associated with SK prediction, the EI is defined as:

$$\mathbb{E}[I(x_0|X)] = \sigma \phi(\gamma) + ((\gamma_{\min} - \bar{\lambda}(x_0))\Phi(\gamma) \qquad (16)$$

where $\gamma = \gamma_{\min} - |\lambda_0 - \bar{\lambda}(x_0)|/\sigma$, $\sigma = \sqrt{\Sigma_M(x_0, x_0) - \Sigma_M^T(x_0, X)[\Sigma_M + \Sigma_\varepsilon]^{-1}\Sigma_M(x_0, X)}$ is the standard deviation of the estimated stability index $\hat{\lambda}(x_0)$ at design point x_0. $\phi(\gamma)$ and $\Phi(\gamma)$ are the probability density function and cumulative distribution function for the standard normal distribution, respectively. Intuitively, the EI tends to select x_0 with mean response $\bar{\lambda}(x_0)$ close to the target λ_0 or with large variance. Fifteen points with the highest EI scores are selected from each slice B_k in this study, and they will be annexed into the existing training set to update the design points and intrinsic variance. This iterative process is halted when the estimated nominal stability boundary converges. Here, we evaluate the convergence of the hyperparameters $\theta = [\beta, \sigma_f^2, L]$ of SK. If the change of θ is within 1%, then both the SK model and the fitted nominal boundary converge. For convenience, we denote the set of all design points selected for TFEM and MC as X_G.

4.4 Gaussian Process. Yet, owing to the intrinsic uncertainty, the machining process may still lose stability for design points below the nominal boundary from the sequential SK algorithm. Thereby, it is necessary to assess the intrinsic uncertainty $\mathbb{V}(x)$ for design points in the vicinity of the nominal boundary to seek a more reliable boundary, which can be attained from TFEM and MC runs in the same way as we assemble the training set for SK. To avoid this laborious process, we draw on GP to predict the intrinsic variance for un-investigated design points in this vicinity, based upon the existing TFEM simulation and MC runs.

GP is a widely used surrogate model that conventionally accommodates prior knowledge in the form of covariance functions [52,53]. Here, we represent the intrinsic variance in a GP framework as

$$\mathbb{V}(x) = f_G^T(x)\beta_G + M_G(x) \qquad (17)$$

Again, $\mathbb{V}(x)$ is the variance of $\lambda(x, \zeta)$ at design point $x = [\Omega, b]^T$ associated with uncontrollable variables $\zeta = [\xi, \omega_n, k]^T$. f_G, β_G, and M_G are similar to those terms defined in SK. Thus, given the training set $X_G = [x_1, \ldots, x_{N_G}]^T$ and $\mathbb{V}_G = [\mathbb{V}(x_1), \ldots, \mathbb{V}(x_{N_G})]^T$, the predictive distribution of the response at a test point x_0 is
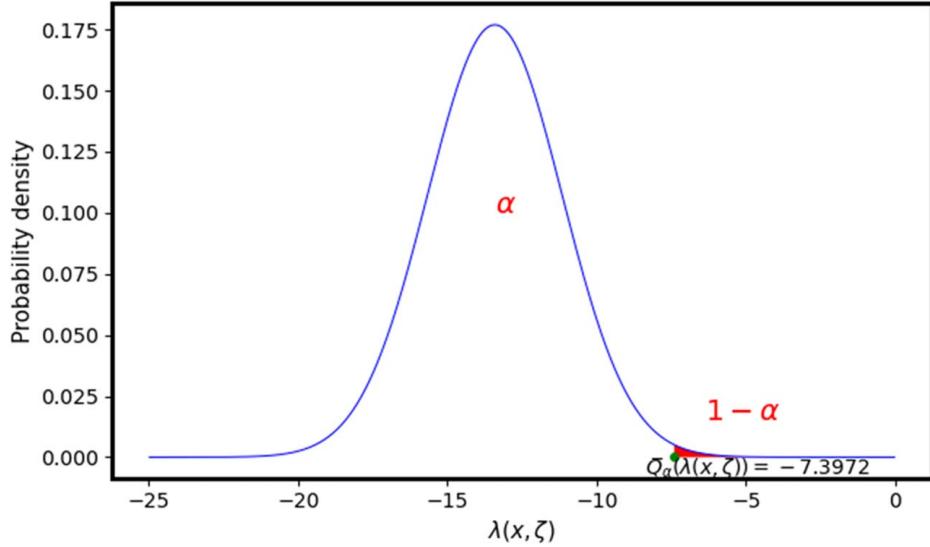
Fig. 5 Illustration of CVaR $\bar{Q}_\alpha(\lambda(x, \zeta))$ at $x = [4.53 \times 10^3, 2.22 \times 10^{-1}]$

given in terms of the first two moments

$$\bar{\mathbb{V}}(x_0) = f_G^T(x_0)\beta_G + \Sigma_{M_G}^T(x_0, X_G)\Sigma_{M_G}^{-1}(V_G - \mathbf{F_G}\beta_G) \quad (18)$$

$$\text{MSE}(\mathbb{V}(x_0)) = \Sigma_{M_G}(x_0, x_0) - \Sigma_{M_G}^T(x_0, X_G)\Sigma_{M_G}^{-1}\Sigma_{M_G}(x_0, X_G) \quad (19)$$

Here, $\Sigma_{M_G} \in \mathbb{R}^{N_G \times N_G}$ is the extrinsic covariance matrix with entity $\Sigma_{M_G}^{(i,j)} = \text{cov}(M_G(x_i), M_G(x_j))$, and $\Sigma_{M_G}(x_0, X_G) = [\text{cov}(M_G(x_0),\ldots M_G(x_1)),\ldots. \text{cov}(M_G(x_0), M_G(x_{N_G}))]^T$. $\mathbf{F_G} = [f_G(x_1), \ldots, f_G(x_{N_G})]^T$. We use the first moment $\bar{\mathbb{V}}(x_0)$ to characterize the variance of stability index $\lambda(x_0, \zeta)$, at a design point x_0.

4.5 Conditional Value-at-Risk. CVaR is widely used in uncertainty quantification [54–56], and we adopt this concept to ensure the reliability of the estimated SLD considering the inherent material and calibration uncertainty. We recall that the nominal stability regime is captured by the contour of $\lambda(x) = 0$. With the uncontrollable variables $\zeta = [\xi, \omega_n, k]^T$, $\lambda(x, \zeta) \sim \mathbb{N}(\lambda(x), \mathbb{V}(x))$, where $\lambda(x) = \mathbb{E}_\zeta[\lambda(x, \zeta)]$ and the intrinsic variance $\mathbb{V}(x) = \text{Var}(\lambda(x, \zeta))$. We define the probability of instability as $p_c = P[\lambda(x, \zeta) \geq 0]$ at design point $x = [\Omega, b]^T$ during the machining process. Given a target reliability or confidence level α, the probability of instability $p_c \leq 1 - \alpha$ is the shaded area under the distribution density curve of $\lambda(x, \zeta)$ in Fig. 5. In this illustrative example, we set $\alpha = 99\%$ and $x = [4.53 \times 10^3, 2.22 \times 10^{-1}]^T$, then the predicted stability index $\lambda(x)$ from SK is $\bar{\lambda}(x) = -13.4088$, and the predicted variance of $\lambda(x, \zeta)$ from GP is $\mathbb{V}(x) = 5.0876$. The α-quantile $Q_\alpha(\lambda(x, \zeta))$ of $\lambda(x, \zeta)$ at x is given as

$$Q_\alpha(\lambda(x, \zeta)) = F_\zeta^{-1}(\alpha) \quad (20)$$

where $F_\zeta(\lambda(x, \zeta))$ is the cumulative distribution function for $\lambda(x, \zeta)$. The CVaR denoted as $\bar{Q}_\alpha(\lambda(x, \zeta))$ is

$$\bar{Q}_\alpha(\lambda(x, \zeta)) = Q_\alpha(\lambda(x, \zeta)) + \frac{1}{1 - \alpha}\mathbb{E}_\zeta[[\lambda(x, \zeta) - Q_\alpha(\lambda(x, \zeta))]^+] \quad (21)$$

where $[\lambda(x, \zeta) - Q_\alpha(\lambda(x, \zeta))]^+ = \max(0, \lambda(x, \zeta) - Q_\alpha(\lambda(x, \zeta)))$. Following this, the CVaR can be derived at each design point $x = [\Omega, b]^T$, and the reliable stability regime of the machining process is $R = \{x | \bar{Q}_\alpha(\lambda(x, \zeta)) < 0\}$.
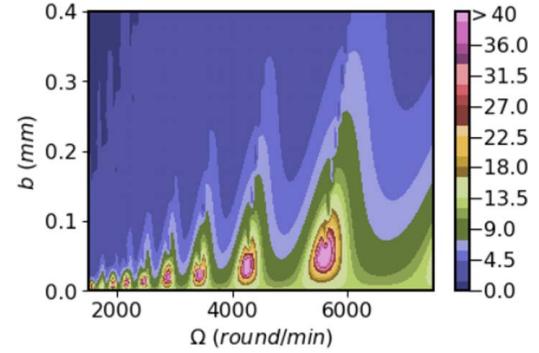
Fig. 6 Heterogeneous variance of $\lambda(x, \zeta)$ in the design space

5 Numerical Result

We first conduct a full factorial design of 100 levels of spindle speed $\Omega \in [1.5 \times 10^3, 7.5 \times 10^3]$ round/min and depth of cut $b \in [0.0, 0.4]$mm. At each of the 100×100 design points of $x = [\Omega, b]^T$, we enlist 500 crude MC runs of the TFEM for the uncontrollable variables $\zeta = [\xi, \omega_n, k]^T$ to gauge the intrinsic uncertainty $\mathbb{V}(x)$. As depicted in Fig. 6, the uncontrollable variables induce heterogeneous variance $\mathbb{V}(x)$ of the response $\lambda(x, \zeta)$ across the design space. This further justifies the SK to accommodate the variance heterogeneity.

The true nominal stability boundary $\lambda(x) = 0$ attained from the MC runs at those 10,000 design points is showcased as the solid curves in Fig. 7, which, however, is prohibitively expensive. Next, the SK-active learning framework is called upon to ameliorate the computational expense. We start the initial slice with $N_s = 100$ space-filling points from the LHD, evaluate those design points with the TFEM and 500 MC runs to derive $\lambda(x)$ and $\mathbb{V}(x)$, and train the SK and GP. This leads to a rough estimate of the nominal stability boundary. For the sequential batch sampling, a set B of $N_c = 10,000$ space-filling design points across $n = 10$ slices are included, and their responses are estimated from SK. Hence, the active learning algorithm selects the top 15 informative design points among $N_c/n = 1,000$ space-filling design points in each slice B_k, $k = 1, \ldots, 10$ via the EI criterion. Those informative points are further fed into the TFEM and MC runs and then annexed into the training dataset to train SK and GP. Therefrom,
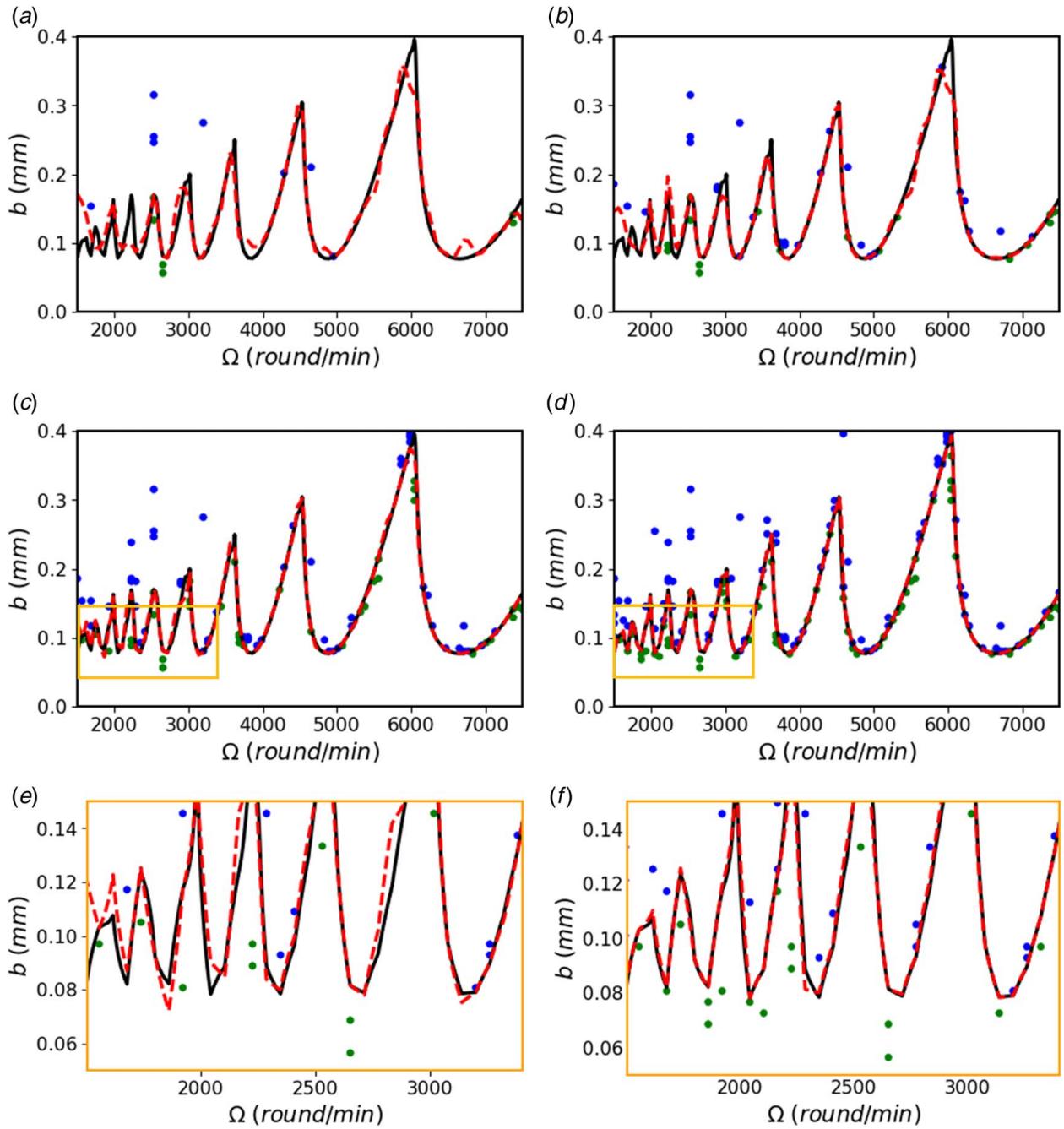
Fig. 7 Comparison of the true nominal stability boundary (the solid curve) and that from the SK-Active learning at the (a) 1st, (b) 3rd, (c) 6th, (d) 10th slice of the FSSF-b design, (e) and (f) are the insets for (c) and (d), respectively. The dots above the dashed line are the selected unstable informative points and the dots below the dashed line are the stable ones in each slice.

the SK and estimated boundary are updated. We halt this iterative process after $n = 10$ slices since the estimated boundary converges.

The sequential estimate of the nominal stability boundary via SK-active learning at different slice numbers is shown in Fig. 7. The dots above the dashed line symbolize the informative design points in the unstable region, whereas the ones below the dashed line represent the informative design points in the stable region. The dashed curve is the predictive stability boundary in each slice, which approaches the true boundary (the solid curve) with the addition of new slices. To discern the subtle difference in the boundary estimation at the 6th and 10th slices, we further zoom into Figs. 7(c) and 7(d) and manifest the local contour in Figs. 7(e) and 7(f), respectively. Therefore, with only $100 + 15 \times 10 = 250$ design points, the estimated stability contour derived from

this novel active learning framework is congruous with the nominal contour obtained from the crude MC with 10,000 design points.

As aforementioned, the SK-active learning framework only derives the nominal contour $\lambda(x) = 0$. Design points below this nominal boundary could still lead to unstable dynamics owing to the intrinsic uncertainty. In an attempt to certify the stochastic optimal design, it is imperative to appraise the risk and quantify the variance $\mathbb{V}(x)$ at design points on the underside in the vicinity of the estimated stability boundary. To shun the computation of $\mathbb{V}(x)$ via the crude MC, we apply GP, trained on the same training dataset in SK-active learning, to predict $\mathbb{V}(x)$ for design points in this critical proximity. To demonstrate the accuracy of variance prediction, 300 design points are randomly selected from the underside
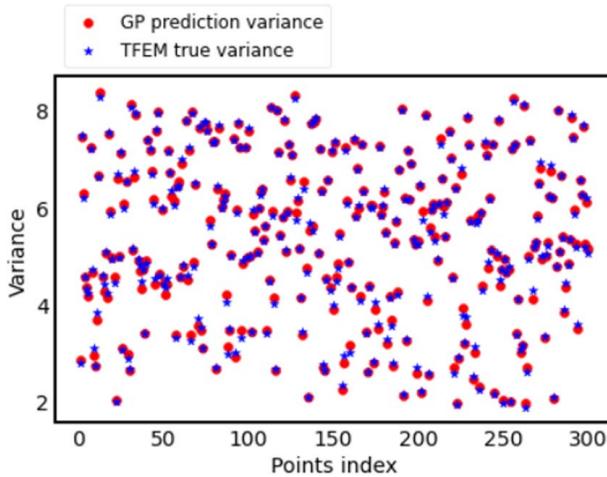
Fig. 8 Prediction and true variance of $\lambda(x, \zeta)$ for design points in the vicinity of the estimated nominal stability boundary

vicinity, and the comparison of predictive $\mathbb{V}(x)$ from GP and the valuation from the onerous TFEM and MC runs is illustrated in Fig. 8. A high predictive accuracy registered here suggests that the predictive variance can be adopted for the CVaR formulation. Consequently, the CVaR is utilized to acquire the reliable boundary under the 99% reliability constraint, shown as the blue curve in Fig. 9. That said, we can safely select the design points below this reliable stability boundary to ensure the avoidance of chatter with a probability of 99%. The dashed line here is the stability boundary estimated by the SK-active learning the same as those in Fig. 7.

To verify the robust stability boundary, we randomly select three points in the design space and evaluate the response according to the tedious TFEM and 500 MC runs of the uncontrollable variables. We also indicate the probability of instability $P[\lambda(x, \zeta) \geq 0]$ at each verifying point. As depicted in Fig. 9, point A (the square) is just above the nominal stability boundary registers $P[\lambda(x, \zeta) \geq 0] = 0.72$, point B (the triangle) residing between the nominal and

Fig. 9 The dashed line: the nominal stability boundary estimated from the SK active learning; the solid line: the robust stability boundary derived from CVaR under 99% reliability. Three verifying design points A, B, and C are randomly selected and tested with the TFEM and 500 MC runs, whose probability of instability is indicated in the figure. At design point D (the diamond), the maximal MRR of $3.26 \times 10^4\, mm^3/_{min}$ is achieved. For design points inside the shaded region, a minimal MRR of $2.50 \times 10^4\, mm^3/_{min}$ can be obtained.

robust stability boundary showcases $P[\lambda(x, \zeta) \geq 0] = 0.36$, and point C (the star) below the robust stability boundary notches $P[\lambda(x, \zeta) \geq 0] = 0.00$. This underscores the necessity to include intrinsic uncertainty to identify the robust stability boundary. Moreover, we test the running time for crude MC and our novel active learning framework on a computer with a processor of i7-8700 CPU, 3.7 GHz, and six cores. It takes 365.87 min for the crude MC to estimate the stability boundary, whereas it only costs 12.71 min in our approach. Thus, our novel approach slashes over 95% of the computational overhead required by the crude MC.

Next, we evaluate MRR at each design point within the robust stability region and identify the maximal MRR. As illustrated in Fig. 9, at point D with $\Omega = 5984.85$ rounds/min and $b = 0.35$ mm (the diamond), the maximal MRR of $3.26 \times 10^4\, mm^3/\min$ is achieved. Further, a baseline MRR is often needed in engineering practice. For demonstration, we set the baseline for MRR as $2.50 \times 10^4\, mm^3/\min$, and then seek the design that achieves an MRR above this baseline, showcased as the shaded area in Fig. 9. This gives the operators leeway to choose from a set of machine settings that avoid the adverse chatter and at the same time achieves an MRR above this baseline.

6 Discussion and Conclusions

In the wake of the ongoing COVID-19 pandemic, the chip shortfall is exacting huge economic tolls on a variety of industries. This underscores the scaleup manufacturing of chips, which necessitates optimal design for quality assurance. In this present work, we investigate the robust optimal design of machining processes, one key operation in chip fabrication, to maximize the MRR and eschew the regenerative chatter, considering the inherent uncertainty. While computer simulations have been widely used to locate the stability boundary in machining, they generally incur a huge computational cost, not to mention the replications to account for the uncertainty. We adopt an active learning framework that integrates the stochastic surrogate SK with the high-fidelity yet tedious TFEM and crude MC runs to predict the stability index at querying design points. The EI criterion is called upon to sequentially identify the critical design points that are informative to locate the stability boundary, and only those critical design points are evaluated via the TFEM. Following this, a GP surrogate is utilized to predict the variance of the stability index at design points below the nominal stability boundary, and CVaR is used to ensure a robust solution. A maximal MRR of $3.26 \times 10^4 \mathrm{mm}^3/\min$ is achieved, and a set of reliable optimal designs that attain a baseline MRR of $2.50 \times 10^4 \mathrm{mm}^3/\min$ are also derived in the illustrative example. We emphasize that the stability boundary is sequentially identified in a cost-effective manner, taking into account the uncontrollable variables ζ. Whereas numerous studies have been reported in decision boundary approximation in a stochastic setting, including efficient global reliability analysis [57,58] and SVM-based limit state analysis [59–62], very few works consider intrinsic uncertainty, particularly in the study of machine stability. The stability boundary is oftentimes bumpy and difficult to fit compared with other reliability boundaries in those existing works. We hope our novel approach can be used in a variety of machining operations (e.g., nano/micromachining) and stimulate the interest in stochastic optimal design in the broad community. We highlight the following implications:

(1) The intrinsic variance attributed to the uncontrollable variables ζ provokes the robust design problem. TFEM and MC run at different realizations of ζ are used to establish the ground truth. To ameliorate the computational overhead, SK, as opposed to the kriging or Gaussian process, is utilized to emulate the high-fidelity TFEM and MC runs. It is noted that in SK, we are modeling the average stability index $\lambda(x_i) = 1/n_i \sum_{j=1}^{n_i} \lambda_j(x_i)$ rather than the stability index $\lambda_j(x_i)$.

(2) Identification of machining stability boundary is essentially a binary classification problem. In the deterministic case, both GP [7,63–65] and SVM have been adopted to approximate the decision boundary. As reported in our earlier work, SVM fails to capture the bumpy stability boundary in a sequential framework [7]. GP has mostly been used for deterministic problems. Very few, if any, existing works consider the active learning-driven sequential design under uncertainty. Hence, we compare the result from our innovative approach to the ground truth given by the TFEM and MC runs.

(3) While we corroborate the proposed approach on a one-dimensional example, it can be easily extended to other high-dimensional machining operations (e.g., five-axis milling). Such high-dimensional processes typically incur enormous computational costs to derive the stability boundary via the high-fidelity simulation, and it is also cumbersome to explore the design space through active learning. Therefore, active learning is imperative for the optimal design for such complex processes.

Acknowledgment

Author's Statement of Original Work

This work has been neither published nor submitted for publication, in whole or in part, either in a serial, professional journal, or as a part of a book that is formally published and made available to the public.

Conflict of Interest

The authors claim that there is no conflict of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

References

[1] McLain, S., 2021, "Auto Makers Retreat From 50 Years of 'Just in Time' Manufacturing," Wall Str. J. Accessed 21 December, 2021. [Online]. Available: https://www.wsj.com/articles/auto-makers-retreat-from-50-years-of-just-in-time-manufacturing-11620051251

[2] Cheng, C., Wang, Z., Hung, W., Bukkapatnam, S. T. S., and Komanduri, R., 2015, "Ultra-Precision Machining Process Dynamics and Surface Quality Monitoring," Procedia Manuf., 1, pp. 607–618.

[3] Bukkapatnam, S. T. S., Rao, P. K., Lih, W.-C., Chandrasekaran, N., and Komanduri, R., 2007, "Process Characterization and Statistical Analysis of Oxide CMP on a Silicon Wafer With Sparse Data," Appl. Phys. A, 88(4), pp. 785–792.

[4] Shamsan, A., and Cheng, C., 2019, "Intrinsic Multiplex Graph Model Detects Incipient Process Drift in Ultraprecision Manufacturing," J. Manuf. Syst., 50, pp. 81–86.

[5] Ringgaard, K., Mohammadi, Y., Merrild, C., Balling, O., and Ahmadi, K., 2019, "Optimization of Material Removal Rate in Milling of Thin-Walled Structures Using Penalty Cost Function," Int. J. Mach. Tools Manuf., 145, p. 103430.

[6] Che, Y., and Cheng, C., 2018, "Uncertainty Quantification in Stability Analysis of Chaotic Systems With Discrete Delays," Chaos Solitons Fractals, 116, pp. 208–214.

[7] Che, Y., Liu, J., and Cheng, C., 2019, "Multi-Fidelity Modeling in Sequential Design for Stability Identification in Dynamic Time-Delay Systems," Chaos Interdiscip. J. Nonlinear Sci., 29(9), p. 093105.

[8] Schmitz, T. L., Karandikar, J., Ho Kim, N., and Abbas, A., 2011, "Uncertainty in Machining: Workshop Summary and Contributions," ASME J. Manuf. Sci. Eng., 133(5), p. 051009.

[9] Li, X., Yang, Y., Li, L., Zhao, G., and He, N., 2020, "Uncertainty Quantification in Machining Deformation Based on Bayesian Network," Reliab. Eng. Syst. Saf., 203, p. 107113.

[10] Wang, Z. C., and Cleghorn, W. L., 2002, "Stability Analysis of Spinning Stepped-Shaft Workpieces in a Turning Process," J. Sound Vib., 250(2), pp. 356–367.

[11] Baker, J. R., and Rouch, K. E., 2002, "Use of Finite Element Structural Models in Analyzing Machine Tool Chatter," Finite Elem. Anal. Des., 38(11), pp. 1029–1046.

[12] Mahdavinejad, R., 2005, "Finite Element Analysis of Machine and Workpiece Instability in Turning," Int. J. Mach. Tools Manuf., 45(7), pp. 753–760.

[13] Quintana, G., and Ciurana, J., May 2011, "Chatter in Machining Processes: A Review," Int. J. Mach. Tools Manuf., 51(5), pp. 363–376.

[14] Mahnama, M., and Movahhedy, M. R., 2010, "Prediction of Machining Chatter Based on FEM Simulation of Chip Formation Under Dynamic Conditions," Int. J. Mach. Tools Manuf., 50(7), pp. 611–620.

[15] Turkes, E., Orak, S., Neseli, S., and Yaldiz, S., 2011, "Linear Analysis of Chatter Vibration and Stability for Orthogonal Cutting in Turning," Int. J. Refract. Met. Hard Mater., 29(2), pp. 163–169.

[16] Landers, R. G., and Ulsoy, A. G., 2007, "Nonlinear Feed Effect in Machining Chatter Analysis," ASME 2007 International Manufacturing Science and Engineering Conference, Atlanta, GA, Oct. 15–18, pp. 17–26.

[17] Greis, N. P., Nogueira, M. L., Bhattacharya, S., and Schmitz, T., 2020, "Physics-Guided Machine Learning for Self-Aware Machining," 2020 AAAI Spring Symposium on AI and Manufacturing, Palo Alto, CA, Mar. 23–25.

[18] Altintaş, Y., and Budak, E., 1995, "Analytical Prediction of Stability Lobes in Milling," CIRP Ann., 44(1), pp. 357–362.

[19] Khasawneh, F. A., and Munch, E., 2016, "Chatter Detection in Turning Using Persistent Homology," Mech. Syst. Signal Process., 70–71, pp. 527–541.

[20] Schmitz, T. L., Medicus, K., and Dutterer, B., 2002, "Exploring Once-Per-Revolution Audio Signal Variance as a Chatter Indicator," Mach. Sci. Technol., 6(2), pp. 215–233.

[21] Axinte, D. A., Gindy, N., Fox, K., and Unanue, I., 2004, "Process Monitoring to Assist the Workpiece Surface Quality in Machining," Int. J. Mach. Tools Manuf., 44(10), pp. 1091–1108.

[22] Denkena, B., Bergmann, B., and Reimer, S., 2020, "Analysis of Different Machine Learning Algorithms to Learn Stability Lobe Diagrams," Procedia CIRP, 88, pp. 282–287.

[23] Friedrich, J., Hinze, C., Renner, A., Verl, A., and Lechler, A., 2015, "Estimation of Stability Lobe Diagrams in Milling With Continuous Learning Algorithms," Robot. Comput.-Integr. Manuf., 43, pp. 124–134.

[24] Friedrich, J., Torzewski, J., and Verl, A., 2018, "Online Learning of Stability Lobe Diagrams in Milling," Procedia CIRP, 67, pp. 278–283.

[25] Huang, X., Hu, M., Zhang, Y., and Lv, C., 2016, "Probabilistic Analysis of Chatter Stability in Turning," Int. J. Adv. Manuf. Technol., 87(9–12), pp. 3225–3232.

[26] Park, S. S., and Qin, Y. M., 2007, "Robust Regenerative Chatter Stability in Machine Tools," Int. J. Adv. Manuf. Technol., 33(3), pp. 389–402.

[27] Totis, G., 2009, "RCPM—A New Method for Robust Chatter Prediction in Milling," Int. J. Mach. Tools Manuf., 49(3–4), pp. 273–284.

[28] Löser, M., Otto, A., Ihlenfeldt, S., and Radons, G., 2018, "Chatter Prediction for Uncertain Parameters," Adv. Manuf., 6(3), pp. 319–333.

[29] Kong, F., and Yu, J., May 2007, "Study of Fuzzy Stochastic Limited Cutting Width on Chatter," Int. J. Adv. Manuf. Technol., 33(7–8), pp. 677–683.

[30] Emmerich, M. T. M., Giannakoglou, K. C., and Naujoks, B., 2006, "Single - and Multiobjective Evolutionary Optimization Assisted by Gaussian Random Field Metamodels," IEEE Trans. EComput., 10(4), pp. 421–439.

[31] Zhou, Z., Ong, Y. S., Nair, P. B., Keane, A. J., and Lum, K. Y., 2007, "Combining Global and Local Surrogate Models to Accelerate Evolutionary Optimization," IEEE Trans. Syst. Man Cybern. Part C Appl. Rev., 37(1), pp. 66–76.

[32] Ponweiser, W., Wagner, T., and Vincze, M., 2008, "Clustered Multiple Generalized Expected Improvement: A Novel Infill Sampling Criterion for Surrogate Models," 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence), Hong Kong, China, June 1–6, pp. 3515–3522.

[33] Lookman, T., Balachandran, P. V., Xue, D., and Yuan, R., 2019, "Active Learning in Materials Science With Emphasis on Adaptive Sampling Using Uncertainties for Targeted Design," npj Comput. Mater., 5(1).

[34] Liu, B., Zhang, Q., and Gielen, G. G. E., 2014, "A Gaussian Process Surrogate Model Assisted Evolutionary Algorithm for Medium Scale Expensive Optimization Problems," IEEE Trans. E Comput., 18(2), pp. 180–192.

[35] Scott, W., Frazier, P., and Powell, W., 2011, "The Correlated Knowledge Gradient for Simulation Optimization of Continuous Parameters Using Gaussian Process Regression," SIAM J. Optim., 21(3), pp. 996–1026.

[36] van Houtum, G. J. J., and Vlasea, M. L., 2021, "Active Learning Via Adaptive Weighted Uncertainty Sampling Applied to Additive Manufacturing," Addit. Manuf., 48(PB), p. 102411.

[37] Che, Y., and Cheng, C., 2021, "Active Learning and Relevance Vector Machine in Efficient Estimate of Basin Stability for Large-Scale Dynamic Networks," Chaos Interdiscip. J. Nonlinear Sci., 31(5), p. 053129.

[38] Botcha, B., Iquebal, A. S., and Bukkapatnam, S. T. S., 2021, "Efficient Manufacturing Processes and Performance Qualification Via Active Learning: Application to a Cylindrical Plunge Grinding Platform," Procedia Manuf., 53, pp. 716–725.

[39] Shim, J., Kang, S., and Cho, S., 2020, "Active Learning of Convolutional Neural Network for Cost-Effective Wafer Map Pattern Classification," IEEE Trans. Semicond. Manuf., 33(2), pp. 258–266.

[40] Gal, Y., Islam, R., and Ghahramani, Z., 2017, "Deep Bayesian Active Learning With Image Data," Proceedings of the 34th International Conference on Machine Learning, Sydney, NSW, Australia, Aug. 6–11, pp. 1183–1192.

[41] Bayly, P. V., Halley, J. E., Mann, B. P., and Davies, M. A., 2003, "Stability of Interrupted Cutting by Temporal Finite Element Analysis," ASME J. Manuf. Sci. Eng., 125(2), pp. 220–225.

[42] Carlberg, K., Barone, M., and Antil, H., 2017, "Galerkin v. Least-Squares Petrov–Galerkin Projection in Nonlinear Model Reduction," J. Comput. Phys., 330, pp. 693–734.

[43] Shang, B., and Apley, D. W., Mar. 2021, "Fully-Sequential Space-Filling Design Algorithms for Computer Experiments," J. Qual. Technol., 53(2), pp. 173–196.

[44] Qian, P. Z. G., 2012, "Sliced Latin Hypercube Designs," J. Am. Stat. Assoc., 107(497), pp. 393–399.

[45] Rennen, G., Husslage, B., Van Dam, E. R., and Den Hertog, D., 2010, "Nested Maximin Latin Hypercube Designs," Struct. Multidiscip. Optim., 41(3), pp. 371–395.

[46] Ankenman, B., Nelson, B. L., and Staum, J., 2010, "Stochastic Kriging for Simulation Metamodeling," Oper. Res., 58(2), pp. 371–382.

[47] Che, Y., Guo, Z., and Cheng, C., 2021, "Generalized Polynomial Chaos-Informed Efficient Stochastic Kriging," J. Comput. Phys., 445, p. 110598.

[48] Totis, G., and Sortino, M., 2020, "Polynomial Chaos-Kriging Approaches for an Efficient Probabilistic Chatter Prediction in Milling," Int. J. Mach. Tools Manuf., 157, p. 103610.

[49] Bukkapatnam, S. T. S., and Cheng, C., 2010, "Forecasting the Evolution of Nonlinear and Nonstationary Systems Using Recurrence-Based Local Gaussian Process Models," Phys. Rev. E, 82(5), p. 056206.

[50] Zhan, D., and Xing, H., 2020, "Expected Improvement for Expensive Optimization: A Review," J. Glob. Optim., 78(3), pp. 507–544.

[51] Jones, D. R., Schonlau, M., and Welch, W. J., 1998, "Efficient Global Optimization of Expensive Black-Box Functions," J. Glob. Optim., 13(4), pp. 455–492.

[52] Yin, J., Ng, S. H., and Ng, K. M., 2011, "Kriging Metamodel With Modified Nugget-Effect: The Heteroscedastic Variance Case," Comput. Ind. Eng., 61(3), pp. 760–777.

[53] Gramacy, R. B., and Lee, H. K. H., 2008, "Bayesian Treed Gaussian Process Models With an Application to Computer Modeling," J. Am. Stat. Assoc., 103(483), pp. 1119–1130.

[54] Chaudhuri, A., Kramer, B., Norton, M., Royset, J. O., and Willcox, K., 2021, "Certifiable Risk-Based Engineering Design Optimization," AIAA J., 60(2), pp. 1–15.

[55] Zhu, S., and Fukushima, M., 2009, "Worst-Case Conditional Value-at-Risk With Application to Robust Portfolio Management," Oper. Res., 57(5), pp. 1155–1168.

[56] Royset, J. O., Bonfiglio, L., Vernengo, G., and Brizzolara, S., 2017, "Risk-Adaptive Set-Based Design and Applications to Shaping a Hydrofoil," ASME J. Mech. Des., 139(10), p. 101403.

[57] Chaudhuri, A., Marques, A. N., and Willcox, K., 2021, "mfEGRA: Multifidelity Efficient Global Reliability Analysis Through Active Learning for Failure Boundary Location," Struct. Multidiscip. Optim., 64(2), pp. 797–811.

[58] Bichon, B. J., Eldred, M. S., Mahadevan, S., and McFarland, J. M., 2013, "Efficient Global Surrogate Modeling for Reliability-Based Design Optimization," ASME J. Mech. Des., 135(1), p. 011009.

[59] Li, H.-S., Zhao, A.-L., and Tee, K. F., 2016, "Structural Reliability Analysis of Multiple Limit State Functions Using Multi-Input Multi-Output Support Vector Machine," Adv. Mech. Eng., 8(10), p. 1687814016671447.

[60] Alibrandi, U., Alani, A. M., and Ricciardi, G., 2015, "A New Sampling Strategy for SVM-Based Response Surface for Structural Reliability Analysis," Probabilistic Eng. Mech., 41, pp. 1–12.

[61] Basudhar, A., and Missoum, S., 2008, "Adaptive Explicit Decision Functions for Probabilistic Design and Optimization Using Support Vector Machines," Comput. Struct., 86(19), pp. 1904–1917.

[62] Basudhar, A., Missoum, S., and Harrison Sanchez, A., 2008, "Limit State Function Identification Using Support Vector Machines for Discontinuous Responses and Disjoint Failure Domains," Probabilistic Eng. Mech., 23(1), pp. 1–11.

[63] Zhang, G., Li, J., Chen, Y., Huang, Y., Shao, X., and Li, M., 2014, "Prediction of Surface Roughness in End Face Milling Based on Gaussian Process Regression and Cause Analysis Considering Tool Vibration," Int. J. Adv. Manuf. Technol., 75(9–12), pp. 1357–1370.

[64] Kawai, K., Irino, N., Iiyama, K., Matsubara, A., Mizuyama, H., Mori, K., and Fujishima, M., 2021, "A Prediction Model for High Efficiency Machining Conditions Based on Machine Learning," Proc. CIRP, 101, pp. 54–57.

[65] Feng, J., Hou, N., Jian, Z., Sun, Z., and Zhang, J., 2020, "An Efficient Method to Predict the Chatter Stability of Titanium Alloy Thin-Walled Workpieces During High-Speed Milling by Considering Varying Dynamic Parameters," Int. J. Adv. Manuf. Technol., 106(11–12), pp. 5407–5420.