PDE-Constrained Gaussian Process Model on Material Removal Rate of Wire Saw Slicing Process

Hongxu Zhao
Department of Industrial Engineering,
Tsinghua University,
Beijing, China

Ran Jin
H. Milton Stewart School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, GA

Su Wu
Department of Industrial Engineering,
Tsinghua University,
Beijing, China

Jianjun Shi[1]
H. Milton Stewart School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, GA
e-mail: jianjun.shi@isye.gatech.edu

Thickness uniformity of wafers is a critical quality measure in a wire saw slicing process. Nonuniformity occurs when the material removal rate (MRR) changes over time during a slicing process, and it poses a significant problem for the downstream processes such as lapping and polishing. Therefore, the MRR should be modeled and controlled to maintain the thickness uniformity. In this paper, a PDE-constrained Gaussian process model is developed based on the global Galerkin discretization of the governing partial differential equations (PDEs). Three features are incorporated into the statistical model: (1) the PDEs governing the wire saw slicing process, which are obtained from engineering knowledge, (2) the systematic errors of the manufacturing process, and (3) the random errors, including both random manufacturing errors and measurement noises. Real experiments are conducted to provide data for the validation of the PDE-constrained Gaussian process model by estimating the model coefficients and further using the model to predict the overall MRR profile. The results of cross-validation indicate that the prediction performance of the PDE-constrained Gaussian process model is better than the widely used universal Kriging model with a mean of second order polynomial functions. [DOI: 10.1115/1.4003617]

Keywords: wire saw slicing, wafer production, PDE-constrained system, Gaussian process

1 Introduction

The wire saw slicing technology has been widely used in wafer manufacturing due to its numerous advantages: It offers high throughput capacities, decreased surface damage, and lower kerf loss [1–4]. One of the critical quality concerns of the wire saw slicing process is the thickness uniformity of the sliced wafers [1,2,5]. This problem becomes more critical in wafer manufacturing today as the diameter of the wafer increases. Figure 1 shows an example of the thickness profile of a sliced wafer; the solid curve marks the contour of the wafer, and darker areas indicate larger thicknesses.

Various efforts have been reported in modeling the vibration and thickness variation of the wire saw slicing process using engineering models. Möller [2] provided the basic mechanism about the microscopic detail of the wire saw slicing process, such as micro-indentation, single particle interaction, and fracture mechanics. Sahoo et al. [3] discussed the vibration and stiffness control aspects. Kao and co-workers [1,3,5] proposed the governing partial differential equations (PDEs) for the wire saw slicing process. The vibration characteristics of the wire saw under hydrodynamic effects are studied by modeling the interactions between the dynamics of the wire saw and the hydrodynamic characteristics of slurry flow [1]. Galerkin-based discretization and general eigenvalue analysis are employed to analyze the vibration response of the system subject to simple viscous damping and harmonic external excitations [1,5,6]. Möller [4] studied the hydrodynamic behavior of the slurry and the elastic interaction with the wire saw and attributed the material removal to the indentation of the free floating abrasive particles under the pressure of the wire saw.

The fluctuation of the material removal rate (MRR) (defined in Sec. 2.2) is one of the major causes of the dimensional variation and the surface roughness of wafers [1,7]. The MRR should be carefully controlled to maintain a uniform rate, thus obtaining a uniform thickness of wafers. The objective of this paper is to predict the MRR using a PDE-constrained Gaussian process (GP) model, which sets up a basis for the effective optimization of a slicing process as future work.

The GP model has been well studied in the statistical literature [8]. A typical GP model consists of a deterministic mean part and a GP part, which stands for the random field. GP models provide a probabilistic modeling approach for nonlinear systems. It has been widely used to predict the unobserved values of a random field based on the mean part and the GP part. Compared with traditional regression models, GP models have four advantages due to their capability to model and to represent (1) the spatial configuration of observational points, (2) the position of the data points within the region of interest, (3) the distances between the data points with respect to the area of interest, and (4) the spatial continuity of the variables [9].

In a wire saw slicing process, the wafer thickness profile is influenced by the manufacturing process variables. A direct adoption of the GP model to predict the thickness profile is not effective. This can be explained by its two inherent limitations: (1) It is difficult to determine the mean structure of a common GP model to describe the process effectively as the process parameters change. (2) A common GP model does not reasonably incorporate the physical process variables into the model based on the engineering knowledge; thus, it cannot be used in process optimization effectively. On the other hand, the governing PDEs of the wire saw slicing process have been developed for the ideal equilibrium

[1]Corresponding author.

Contributed by the Manufacturing Engineering Division of ASME for publication in the JOURNAL OF MANUFACTURING SCIENCE AND ENGINEERING. Manuscript received March 9, 2010; final manuscript received January 27, 2011; published online March 23, 2011. Assoc. Editor: Steven J. Skerlos.

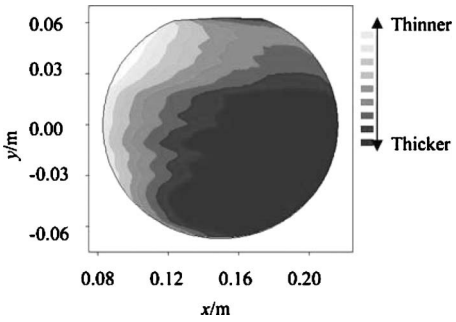

Fig. 1 Contour diagram of the thickness profile of a sliced wafer (unit: m)

conditions without consideration of the manufacturing uncertainties [1]. Thus, the PDEs have limited accuracy or applicability to represent the *real* slicing process where the manufacturing uncertainties exist. Furthermore, analytical solutions for those two coupled PDEs do not exist [1].

In this paper, an innovative modeling approach is proposed to utilize the engineering specifications in a GP model framework for the wafer profile prediction. The governing PDEs are discretized and serve as the constraints of the GP model when the coefficients are estimated based on the observational data. The obtained PDE-constrained GP model takes three features into consideration: (1) the PDEs governing the wire saw slicing process, which are developed from the engineering knowledge, (2) the systematic errors of the manufacturing process, and (3) the random errors, including both random manufacturing errors and measurement noises. The obtained PDE-constrained GP model is used to predict the overall MRR profile and consequently the overall thickness profile. This paper also demonstrates that the obtained model provides better prediction performance than the commonly used universal Kriging model in the statistics literature [8]. In addition, the PDE-constrained GP model has the potential to determine the optimal settings of the wire saw process variables because those parameters have been embedded in the mean part of the PDE-constrained GP model. The research of the optimization of the slicing process will be discussed in the future.

This paper is organized as follows: After the introduction, we illustrate the principle of the wire saw slicing process, define the MRR, and introduce the governing PDEs in Sec. 2. In Sec. 3, we discretize the governing PDEs and develop the PDE-constrained GP model. In Sec. 4, we present the validation of the developed model and the comparison of the prediction performance with a universal Kriging model through Monte Carlo cross-validation using real production data. Finally, the collusions are summarized in Sec. 5.

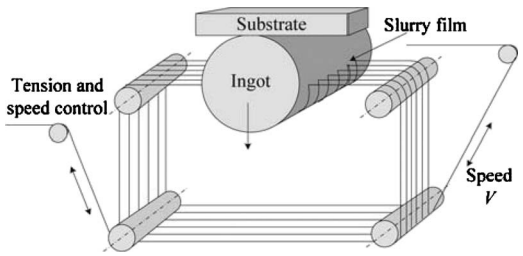

Fig. 2 Schematic of the wire saw slicing process

2 Principle of Wire Saw Slicing Process and Coupled Governing PDEs

2.1 Principle of Wire Saw Slicing Process. The principle of a wire saw slicing process is depicted in Fig. 2. In Fig. 2, a single wire is wound in a group of guides, forming a wire web consisting of parallel strands of wires [1,2]. A silicon ingot is fixed in a substrate above the wire web. During the wire saw slicing process, the wire saw is subjected to a designed tension T and a speed V. The wire saw moves back and forth between the rotating wire guides by changing the direction periodically to slice the ingot. The substrate moves down slowly and presses the ingot against the web of parallel wire saws. Meanwhile, the slurry manifolds above the ingot supply abrasive slurry to the contact zone of the wire saw and the ingot. A thin abrasive slurry film is trapped between the gap of the wire saw and the ingot due to the vertical force and the wire tension, as depicted in Figs. 3 and 4.

The gap of the wire saw and the ingot, also known as the thickness of the slurry film, is denoted by $h(x,y)$; here, (x,y) are the coordinates of a contact point in the ingot. The gap $h(x,y)$ varies along the contact span. Typically, the wire saw does not contact with the ingot directly but is supported entirely by the pressure of the slurry film [1,5], denoted by $p(x,y)$. The pressure on the wire saw comes from the contacts or semicontacts with the abrasive particles and/or other system noises [4,5]. The material removal is achieved by the abrasive particles freely dispersed in the slurry film [10,11].

2.2 Definition of Material Removal Rate. The MRR determines the overall thickness profile. Thus, a constant MRR is desired to maintain a uniform thickness of wafer. The MRR, in units of m^3/s, is defined as [12]

$$\mathrm{MRR} = \frac{\Delta M}{\rho \times \Delta T} \tag{1}$$

where ΔT is the processing time in seconds, ΔM is the mass of material lost at ΔT time interval in kg, and ρ is the density of material in kg/m^3.

In a typical wire saw slicing process, the minimum film thickness is greater than the average size of the abrasive particles, which is known as a "floating" machining condition. The slicing is achieved by fine abrasive particles freely distributed in the slurry film. In practice, this is very important since the wire saw is used to slice fragile materials with high precision requirements on the surface finish. During the wire saw slicing process, as depicted in Fig. 5, a theoretical formulation of the MRR is derived for the moment in which the wire saw process across a point (x,y) in the ingot [1,5],

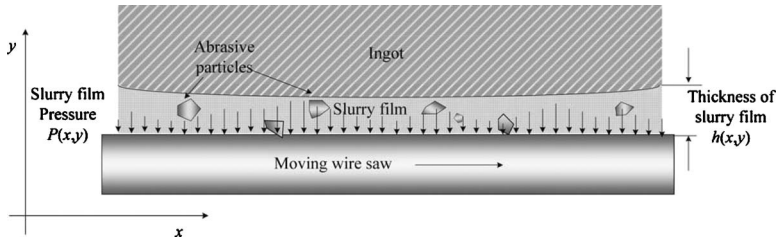

Fig. 3 Schematic of the slurry film in the contact zone of the wire saw and the ingot

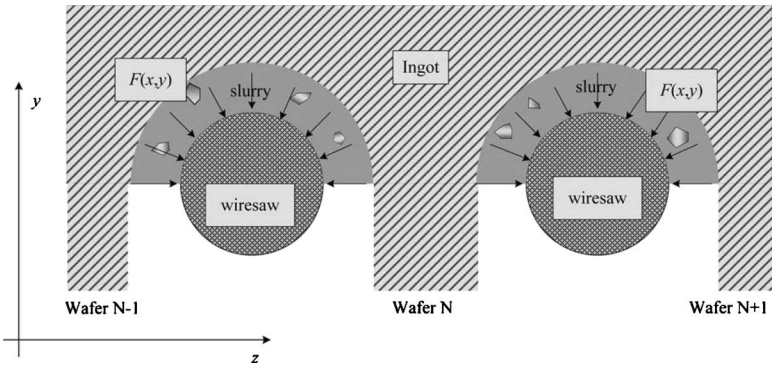

Fig. 4 Cross section of the wire saw, ingot, and slurry film in the contact zone

$$\mathrm{MRR}(x,y) = auV^2\frac{P(x,y)}{h^2(x,y)} \tag{2}$$

where a is a constant value, u is the viscosity of the slurry, V is the speed of the wire saw when the wire saw proceeds across the contact span of y, which is an online controllable variable, $P(x,y)$ is the hydrodynamic pressure between the slurry film and the wire saw at point (x,y), and $h(x,y)$ is the thickness of the slurry film (also known as the vertical displacement of the wire saw) at point (x,y). $P(x,y)$ and $h(x,y)$ are the critical variables to determine $\mathrm{MRR}(x,y)$. However, it is practically infeasible to measure $P(x,y)$ and $h(x,y)$ directly during the slicing operations. Sections 2.3 and 2.4 will introduce two PDEs governing $P(x,y)$ and $h(x,y)$.

2.3 Vibration Equation of the Wire Saw in a Slicing Process. As depicted in Fig. 5, the wires move back and forth with a mass density per unit length m in kg/m, tension $T(y)$ in N, speed V in m/s, and vertical displacement $h(x,y)$ in m. The wire saw is excited by force $F(x,y)$, which is distributed along the contact span. $F(x,y)$ is exerted by the internal interactions of the wires, slurry film, and abrasive particles contained in the slurry film and the external excitations. As illustrated in Fig. 4, $F(x,y)$ is expressed as [13]

$$F(x,y) = 2\int_0^{\pi r/2} P(x,y)\cos\left(\frac{1}{2r}\right)dl = cP(x,y) \tag{3}$$

where $c=2\int_0^{\pi r/2}\cos(1/2r)dl$ is a constant, l is the circumference of the wire saw, and r is the radius of the wire saw.

Without loss of generality, we focus on a neighborhood of contact span $y=Y$. Let a tiny time increment (or decrement) be denoted as t. As depicted in Fig. 6, a vertical neighborhood of contact span $y=Y$ is defined as

$$\{(x,y)|y = Y + y(t), -\bar{t} \le t \le \bar{t}\} \tag{4}$$

where $\bar{t}$ and $-\bar{t}$ provide the upper and lower boundaries of this neighborhood. In this neighborhood, the transient vibration of the moving wire saw is described by a PDE [1,14],

$$m\ddot{h} + 2mV\dot{h}_x = -(T - mV^2)h_{xx} = F\cdot[H(x-b_1) - H(x-b_2)] \quad (0 \le x \le L) \tag{5}$$

where h and F stand for $h(x,y)$ and $F(x,y)$, respectively, and in the following discussions, functions will be abbreviated in the same way for simplicity. $b_1(y)$ and $b_2(y)$ are the x coordinates of the left and right boundaries of the contact span, respectively, as shown in Fig. 5. L is a constant value, standing for the length of the wire saw between the guides. $\ddot{h}=\partial^2 h/\partial t^2$ is the second order partial derivative of $h(x,y(t))$ with respect to t. Similarly, there are $\dot{h}_x=\partial^2 h/\partial t\,\partial x$ and $h_{xx}=\partial^2 h/\partial x^2$, and $H(x-b_1(y))$ is a Heaviside function, defined as

$$H(x-b_1(y)) - H(x-b_2(y)) = \begin{cases} 1 & \text{if } b_1(y) \le x \le b_2(y) \\ 0 & \text{else} \end{cases} \tag{6}$$

Equation (5) is derived based on the Newtonian mechanics theory [14]. It is worth mentioning that Hamilton's principle yields the same equation [15]. Equation (5) is a transient equation because the parameters (process variables, such as contact span,

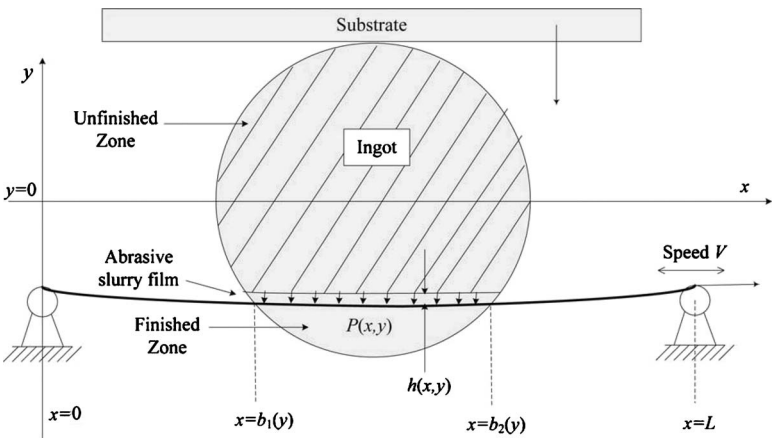

Fig. 5 Schematic of the wire saw slicing process

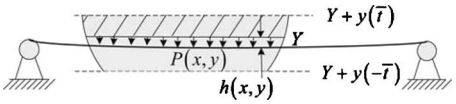

Fig. 6 Neighborhood of contact span $y=Y$

wire saw speed, and tension) are changing in the process.

As the wire saw is wound in the guides, the vibrations of the wire saw in both edges are forced to remain at zero. This consideration gives the boundary conditions of the vibration equation as follows:

$$h(0,y) = h(L,y) = 0$$

$$\dot{h}(0,y) = \dot{h}(L,y) = 0$$

$$\ddot{h}(0,y) = \ddot{h}(L,y) = 0 \tag{7}$$

A similar equation for the wire saw slicing process is also given by Kao and co-workers [1,16], while the second term of the left side of Eq. (5) is removed for numerical computational efficiency in the finite element method.

2.4 Nonlinear Reynold's Equation of the Pressure of the Slurry Film. The slurry film between the moving wire saw and the ingot surface is of pseudohydrodynamic characteristics [1,11]. In the perspective of tribology of fluid film, the interaction between the slurry film and the wire saw could be described by a nonlinear Reynold's equation [1] as

$$(h^3 P_x)_x = 6uVh_x + 12uh_t \quad (b_1 \le x \le b_2) \tag{8}$$

with boundary conditions,

$$P(b_1,y) = P(b_2,y) = P_{\text{atm}} \tag{9}$$

where $P_{\text{atm}} = 1.013 \times 10^5$ is a constant value, standing for atmospheric pressure (unit: Pa).

3 PDE-Constrained Gaussian Process Model

3.1 Fourier Series Expansions and Global Galerkin Discretization of Governing PDEs. In Sec. 2, the wire saw slicing process is described by the coupled PDEs (Eqs. (5) and (8)). However, the analytical solutions to the coupled PDEs do not exist. In this section, the global Galerkin method [6] is employed to discretize these two PDEs. The derived relationship of $h(x,y)$ and $P(x,y)$ is critical for modeling the MRR.

The global Galerkin method of the weighted residual is an effective way to obtain approximate solutions to PDEs by using a finite number of shape functions [6]. Fourier series and polynomial series are common choices of shape functions. By taking the boundary conditions Eqs. (7) and (9) into account, the pressure $P(x,y)$ and the thickness of the slurry film $h(x,y)$ are expanded into a set of Fourier series as

$$P(x,y) = P_{\text{atm}} + \sum_{k=1}^{NP} \beta_k \psi_k(x,y) \tag{10}$$

$$h(x,y) = \sum_{j=1}^{Nh} \xi_j \phi_j(x) \tag{11}$$

where NP and Nh are the degrees of the Fourier series of $P(x,y)$ and $h(x,y)$, respectively; $\{\beta_k\}_{k=1}^{NP}$ and $\{\xi_j\}_{j=1}^{Nh}$ are the unknown parameters; $\{\psi_k(x,y)\}_{k=1}^{NP}$ and $\{\phi_j(x)\}_{j=1}^{Nh}$ are two sets of basis functions in the function space, which are of unit length and orthogonal to each other in each set. These basis functions are

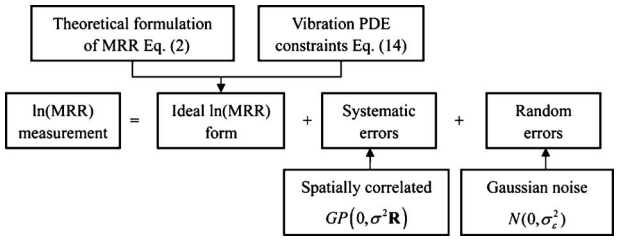

Fig. 7 Schematic of the structure of the PDE-constrained GP model

$$\psi_k(x,y) = \sqrt{\frac{2}{b_2(y)-b_1(y)}} \sin\left(\frac{2k\pi(x-b_1(y))}{b_2(y)-b_1(y)}\right), \quad b_1(y) \le x \le b_2(y) \tag{12}$$

$$\phi_j(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{2j\pi x}{L}\right), \quad 0 \le x \le L \tag{13}$$

Applying the global Galerkin method of weighted residual, a solution of $\xi_j (j=1,2,\ldots,Nh)$ is given as a linear combination of $\{\beta_k\}_{k=1}^{NP}$,

$$\xi_j = \frac{L^2}{4j^2\pi^2(T-mV^2)}\left(cP_{\text{atm}}\int_{b_1}^{b_2}\phi_j(x)dx + \mathbf{E}_{(j)}\boldsymbol{\beta}\right), \quad j = 1,2,\ldots,Nh \tag{14}$$

where $\mathbf{E} \in R^{Nh\times NP}$, with $\mathbf{E}(j,k) = c\int_{b_1}^{b_2}\phi_j(x)\psi_k(x,y)dx$ and $\mathbf{E}_{(j)}$ stands for the jth row of $\mathbf{E}$, and $\boldsymbol{\beta} = (\beta_1\beta_2\cdots\beta_{NP})^T$.

The derivation of Fourier series expansions and Eq. (14) is provided in Appendix A. Equation (14) will be employed as constraints to the modeling of the MRR in the next subsection.

3.2 PDE-Constrained Gaussian Process Model of MRR. In this section, a PDE-constrained GP model of the MRR is developed. Figure 7 shows the structure of the proposed PDE-constrained GP model.

For simplicity, the stochastic variable ln(MRR) is denoted as M. In practice, the wire speed V is not only measurable but also controllable; a uniform thickness profile is obtained by manipulating V. Therefore, V is also incorporated into the model, which is the framework for process optimization. Furthermore, taking the manufacturing errors and sensing noises into consideration, a PDE-constrained GP model is developed in the form of

$$M(x,y,V) = \text{mean}(x,y,V) + Z(x,y) + \varepsilon(x,y) \tag{15}$$

and is subject to the constraint of Eq. (14). $\text{mean}(x,y,V)$ is the deterministic mean part; $Z(x,y)$ is the spatially correlated part (GP part), standing for the systematic manufacturing errors, with $E(Z(x,y))=0$ and

$$\text{cov}(Z(x_i,y_i),Z(x_j,y_j)) = \sigma^2 R(\boldsymbol{\theta},(x_i,y_i),(x_j,y_j))$$

where σ^2 is the variance of $Z(x,y)$; $R(\cdot)$ is the Gaussian correlation function, which is widely used in geostatistics, as follows:

$$R(\boldsymbol{\theta},(x_i,y_i),(x_j,y_j)) = \exp(-\theta_x(x_i-x_j)^2 - \theta_y(y_i-y_j)^2)$$

$\boldsymbol{\theta} = (\theta_x\theta_y)^T$ is the scale parameter and $\varepsilon(x,y)$ is the random error, which is independently and identically distributed Gaussian random variable with $E(\varepsilon(x,y))=0$ and $\text{Var}(\varepsilon(x,y))=\sigma_\varepsilon^2$. In this model, $\varepsilon(x,y)$ is independent of $Z(x,y)$, and both $Z(x,y)$ and $\varepsilon(x,y)$ are not impacted by the controllable variable V.

By substituting the constraints Eq. (14) into Eq. (2), one gets the deterministic mean structure of the model. The details of derivation are given in Appendix B. Thus, one gets the PDE-constrained GP model in the form of

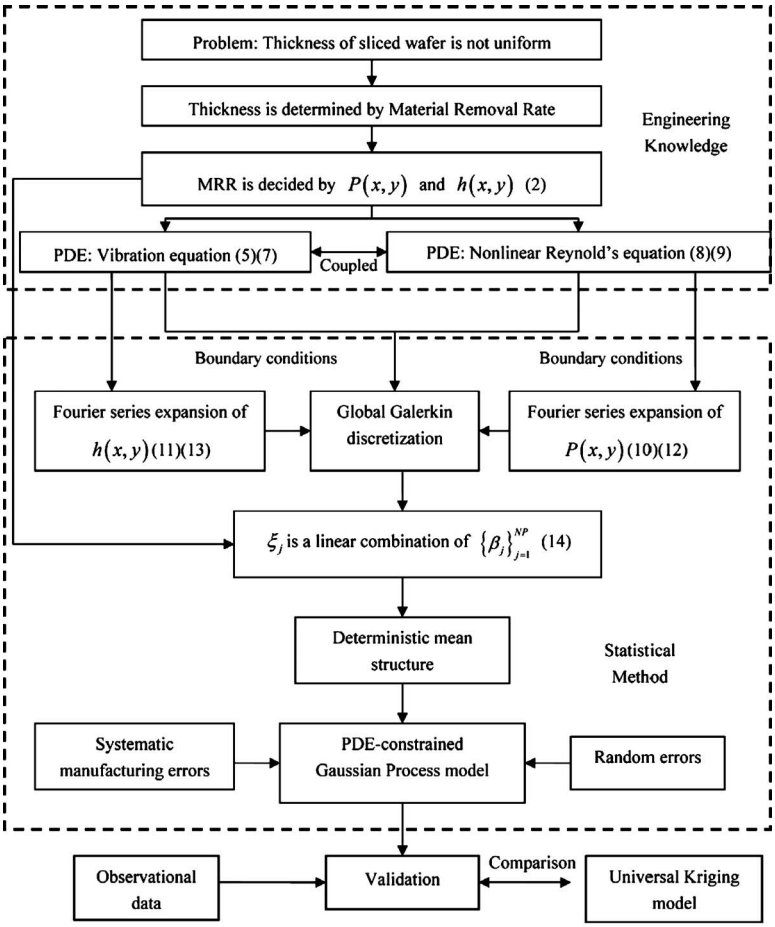

Fig. 8 Procedure of the modeling approach

$$M(x,y,V) = \sum_{k=0}^{NP} \beta_k f_k(x,y,V) + 2\ln(|V|) + \ln(u(y)) + Z(x,y) + \varepsilon(x,y) \tag{16}$$

where $\{\beta_k\}_{k=0}^{NP}$ are unknown parameters to be estimated and $\{f_k\}_{k=1}^{NP}$ are trigonometric functions as linear combinations of $\{\psi_j(x,y)\}_{j=1}^{NP}$ and $\{\phi_j(x)\}_{j=1}^{Nh}$, defined as

$$f_0(x,y,V) = 1,$$

$$f_k(x,y,V) = \left(\psi_k - \frac{L^2}{2\pi^2(T-mV^2)}\sum_{j=1}^{Nh}\frac{1}{j^2}\mathbf{E}(j,k)\phi_j\right),$$

$$k = 1, \ldots, NP \tag{17}$$

The process variables T (tension), m (mass density), u (viscosity), and V (wire saw speed) are given by the design, and the trigonometric functions $\{f_k\}_{k=1}^{NP}$ become known basis functions. On the other hand, $M(x,y,V)$ are also available by obtaining the experimental MRR profiles. The unknown parameters $\{\beta_k\}_{k=0}^{NP}$ in the Eq. (17) will be estimated by the observational data in the method of maximum marginal likelihood [17]. As a summary, Fig. 8 shows the procedure of the proposed modeling approach. Section 4 will provide a validation for the developed PDE-constrained GP model.

4 Validation and Thickness Profile Prediction

In this section, the validation is conducted to evaluate the obtained PDE-constrained GP model by using the data collected from a wafer manufacturing process in a semiconductor company. The validation is conducted through the following procedure.

- Experiments are conducted to slice the ingots in a HCT Shaping™ wire saw machine.
- The thicknesses of the sliced wafers are measured with an ADE WaferCheck™ system, and the experimental ln(MRR) profile is calculated according to Eq. (1).
- The ln(MRR) profile is used to estimate the PDE-constrained GP model. Cross-validation is conducted to compare the predicted ln(MRR) profile and the ln(MRR) profile obtained from the experiments.

Furthermore, the predictions made by the PDE-constrained GP model are compared with those of the universal Kriging model to evaluate the prediction accuracy. A smaller root mean squared prediction error (RMSPE) indicates better prediction performance.

4.1 Observational Data and System Settings. The observational data are collected in a real production line when one ingot (crystallographic orientation ⟨100⟩; diameter of 150 mm) was sliced in a HCT Shaping™ wire saw system. The ingot is set up in the upper position of the wire saw system. The process information, such as machine setup, slurry preparation, web check and replacement, and process settings, are recorded.

During the slicing process, the related process variables, such as wire speed, substrate speed, and slurry information, are recorded every 6 s automatically, and one ingot is sliced into 438 wafers. After the slicing, the thickness profiles of the sliced wafers are measured by an ADE WaferCheck™ system. The sampling points of each wafer, consisting of 5540 points, are distributed in

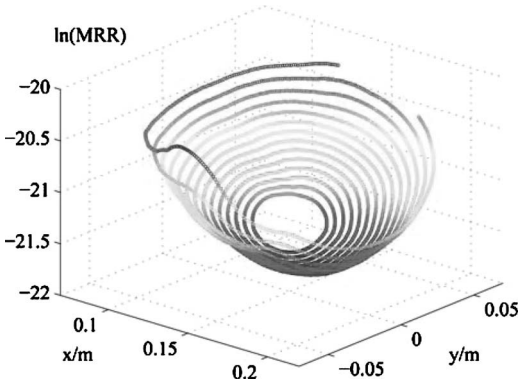

Fig. 9 Profile plot of the experimental ln(MRR) profile over all sample points

a group of concentric circles.

According to the definition given in Eq. (2), the experimental MRR(x,y) in a position (x,y) is defined in a neighborhood of (x,y), denoted as $\Omega(x,y)$. The experimental MRR(x,y) in the neighborhood $\Omega(x,y)$ is calculated according to Eq. (1) as

$$\text{MRR} = \frac{\Delta M}{\rho \Delta T} = \frac{\Delta S \cdot \Delta \text{th}}{\Delta t} \tag{18}$$

where ΔS is the area of the neighborhood, ΔT is the processing time to slice through $\Omega(x,y)$, and Δth is the removed material thickness in $\Omega(x,y)$ at the ΔT time interval.

In this validation, the parameters ΔS and ΔT are determined as follows: (1) a MRR profile is formed by the 5540 sampling data points, and the area of a wafer disk is 0.0182 m^2. Thus, each data point on the profile represents a small area in the order of 1×10^{-6} m^2. Therefore, ΔS is set as 10^{-6} m^2. (2) Based on ΔS, the $\Omega(x,y)$ is approximated as a small square with length in the order of 1×10^{-3} m, which is the distance that the wires travel in that area. Based on the instant substrate speed, ΔT is calculated as the distance 1×10^{-3} m over the instant substrate speed. (3) Δth is calculated by subtracting the thickness at the location (x,y) from the total length of the ingot before the slicing process.

Without loss of generality, the calculation procedure in a position $(x=0.16525,\ y=-0.0421)$ is shown as an example. After the slicing, the thicknesses of all the 438 sliced wafers in the position (1.6525–0.0421) are measured. The summation of thicknesses of all 438 wafers in this position is 0.3232 m, and the original thickness of the ingot in this position before the slicing process is 0.325 m. The difference yields the thickness removed during the slicing process in this position, i.e., Δth=0.0018 m. According to the recorded substrate speed at this moment $(3.01 \times 10^{-4}$ m/s), the processing time is ΔT=3.322 s. Therefore, the experimental MRR in the neighborhood $\Omega(1.6525,-0.0421)$ is estimated as $\Delta S \cdot \Delta \text{th}/\Delta T = 5.418 \times 10^{-10}$ m^3/s.

The experimental ln(MRR) profile is illustrated in Fig. 9. As depicted in Fig. 5, the coordinates of the center of the ingot are $x=L/2$ and $y=0$.

4.2 Comparison of Prediction Performance of Two Kinds of GP models. In this section, the PDE-constrained GP model (Eq. (16)) is adopted to estimate the ln(MRR) profile over the ingot. Here, *NP* is set as 5, and the selection procedure is discussed in Appendix C. Monte Carlo cross-validation is conducted to evaluate the prediction accuracy of the PDE-constrained GP model. In one cross-validation, the 5540 sample points of the experimental ln(MRR) are randomly partitioned into two complementary sets: the training set and the validation set. The parameters of the PDE-constrained GP model are estimated using the training set in the method of maximum marginal likelihood

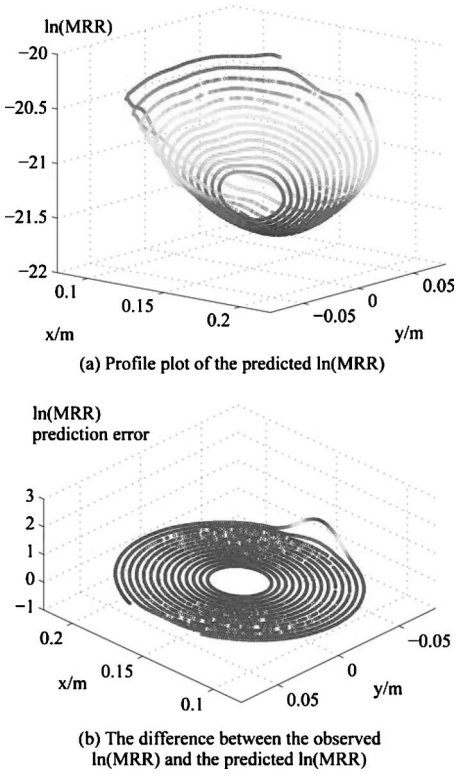

Fig. 10 Profile plot of the predicted ln(MRR) using the estimated PDE-constrained GP model and the difference

[17,18]. Then, the predictions of ln(MRR) are made in the validation set using the estimated PDE-constrained GP model. The RMSPE is calculated to evaluate the prediction performance. $\text{RMSPE}_{\text{PDE},i}$ at the ith cross-validation for the PDE-constrained GP model is defined as

$$\text{RMSPE}_{\text{PDE},i} = \sqrt{\frac{1}{N}\sum_{j=1}^{N} (M_{i,j} - \hat{M}_{\text{PDE},i,j})^2}$$

where N is the sample size of the validation set, $M_{i,j}$ is the observation of the jth point of the validation set in the ith cross-validation, and $\hat{M}_{\text{PDE},i,j}$ is the corresponding prediction made by the PDE-constrained GP model.

The RMPSE quantifies the difference between the predicted profile and the observed profile [17]. With smaller values, the prediction performance is better. To assess the variability of prediction, 300 cross-validations are performed using different partitions.

As an example, the predicted ln(MRR) profile using the estimated PDE-constrained GP model in a cross-validation is depicted in Fig. 10(*a*), and the difference between the experimental and predicted ln(MRR) profiles are illustrated in Fig. 10(*b*). It is clear that the prediction error is close to zero except for a few outliers. These outliers are mainly caused by lacking samples at the zone in the training data set. The estimated parameters of the PDE-constrained GP model using the MATLAB toolbox GPML [17] are

$$\boldsymbol{\theta} = [0.0871, 0.0139], \quad \sigma^2 = 9.3779, \quad \sigma_\varepsilon^2 = 0.0012$$

$$\boldsymbol{\beta}_{\text{PDE}} = [-13.92, 0.0278, 0.0036, 9.45e-4, 2.56e-4, 4.1e-5] \tag{19}$$

Based on the estimated PDE-constrained GP model, the average of the RMSPE in all cross-validations for the ln(MRR) profile is 0.051. Furthermore, the profile of the removed material thickness

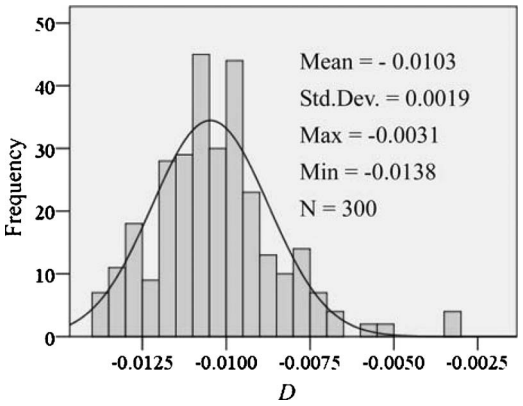

Fig. 11 Histogram of the differences *D*

Δth is predicted by Eq. (18). The standard deviation of the difference between the predicted Δth profile and the experimental Δth data is 2.9×10^{-7} m, which is an acceptable accuracy of the removed material thickness by engineering specifications in this experiment. Therefore, the developed PDE-constrained GP model provides satisfactory prediction performance to the ln(MRR) profile.

As a comparison, a universal Kriging model is also employed to estimate the ln(MRR) profile. The universal Kriging model is a kind of GP model commonly used in computer experiments [8] and geostatistics [9]. A universal Kriging model is in the form of

$$M(x,y) = \mathbf{F}_n(x,y)\boldsymbol{\beta}_{\mathrm{UK}} + Z_{\mathrm{UK}}(x,y) + \varepsilon_{\mathrm{UK}}(x,y)$$

where $\mathbf{F}_n(x,y)\boldsymbol{\beta}_{\mathrm{UK}}$ is a n order polynomial function of the coordinates, $\boldsymbol{\beta}_{\mathrm{UK}}$ is an unknown parameter vector with appropriate dimensions, $Z_{\mathrm{UK}}(x,y)$ is the Gaussian process part, and $\varepsilon_{\mathrm{UK}}(x,y)$ is the random error. Without loss of generality, $\mathbf{F}_n(x,y)\boldsymbol{\beta}_{\mathrm{UK}}$ are selected as the second order polynomial functions of the coordinates to compare with the PDE-constrained GP model. In this case, the universal Kriging model has the same number of predictors in the mean part as the PDE-constrained GP model when NP is set as 5. $\mathrm{RMSPE}_{\mathrm{UK},i}$ is defined in the same way as $\mathrm{RMSPE}_{\mathrm{PDE},i}$.

The parameters of the universal Kriging model are also estimated in the method of maximum marginal likelihood. With the same partition of the sample points, which yields the estimation given in Eq. (19), the estimated parameters of the second order polynomial universal Kriging model are

$$\boldsymbol{\beta}_{\mathrm{UK}} = [-4.088, -297.16, 49.93, 1068.5, -271.8, -1500]$$

To compare the prediction performance of the PDE-constrained GP model and the universal Kriging model, the $\mathrm{RMSPE}_{\mathrm{PDE},i}$ and $\mathrm{RMSPE}_{\mathrm{UK},i}$ are compared. Denote $\mathrm{RMSPE}_{\mathrm{UK}}=\{\mathrm{RMSPE}_{\mathrm{UK},i}\}_{i=1}^{300}$ and $\mathrm{RMSPE}_{\mathrm{PDE}}=\{\mathrm{RMSPE}_{\mathrm{PDE},i}\}_{i=1}^{300}$ as the two sets of RMSPEs; the differences between the paired RMSPEs are defined as $D=\{\mathrm{RMSPE}_{\mathrm{PDE},i}-\mathrm{RMSPE}_{\mathrm{UK},i}\}_{i=1}^{300}$. The average $\mathrm{RMSPE}_{\mathrm{PDE}}$ is 0.051, while that of $\mathrm{RMSPE}_{\mathrm{UK}}$ is 0.0612. Furthermore, as presented in Fig. 11, the results of the cross-validations indicate that the $\mathrm{RMSPE}_{\mathrm{PDE}}$ is always smaller than $\mathrm{RMSPE}_{\mathrm{UK}}$ in all cross-validations. The prediction accuracy is improved by 20.1% on average. Therefore, the proposed PDE-constrained GP model is more accurate than the second order universal Kriging model.

5 Conclusions

In this paper, the modeling of the MRR profile of the wire saw slicing process is studied. In order to model the MRR, the pressure between the wire saw and the slurry film $P(x,y)$ and the thickness of the slurry film $h(x,y)$ are described by the coupled vibration equation and the nonlinear Reynold's equation, where analytical solutions do not exist. Therefore, the MRR profile is hard to be analytically predicted by using engineering knowledge alone. In this paper, the PDE-constrained GP model is proposed to model the MRR profile, with the relation of $h(x,y)$ and $P(x,y)$ as constraints. The obtained PDE-constrained GP model has three components: (1) the mean part derived from the engineering knowledge based on the PDEs, (2) the spatially correlated part for the systematic manufacturing errors, and (3) the random errors, including the random manufacturing errors and measurement noises. In the validation, the PDE-constrained GP model is estimated from the experimental data obtained in a real production environment. The smaller RMSPEs of cross-validations indicate that the prediction performance of the PDE-constrained GP model is superior to the second order polynomial universal Kriging model.

The PDE-constrained GP model addresses three challenges to model the MRR: (1) It provides a group of basis functions as the mean structure of a GP model to describe the engineering process effectively, which is usually difficult to be estimated by using statistical methods alone; (2) it incorporates the process variables into the model, which provides potentials to improve the uniformity of the MRR profile; and (3) it includes the manufacturing uncertainty factors as the spatially correlated GP part and the uncorrelated random error part beyond the engineering models.

Appendix A: Details of Fourier Series Expansions and Derivation of Global Galerkin Discretization of Governing PDEs

We focus on a neighborhood of contact span $y=Y$, defined in Eq. (4). In this neighborhood, we rewrite $P(x,y)=P(x,Y+y(t))$ and $h(x,y)=h(x,Y+y(t))$. The parameters of Eq. (5), such as $b_1(y)$, $b_2(y)$, and V, are treated as constant values in this neighborhood.

By taking the boundary conditions Eqs. (7) and (9) into account, the pressure $P(x,Y+y(t))$ and the thickness of the slurry film $h(x,Y+y(t))$ are expanded into a set of Fourier series as

$$P(x,y) = P(x,Y+y(t)) = P_{\mathrm{atm}} + \sum_{k=1}^{NP} \beta_k(t,Y)\psi_k(x,Y) \tag{A1}$$

$$h(x,y) = h(x,Y+y(t)) = \sum_{j=1}^{Nh} \xi_j(t,Y)\phi_j(x) \tag{A2}$$

where NP and Nh are the degrees of the Fourier series of $P(x,Y+y(t))$ and $h(x,Y+y(t))$, respectively; the parameters $\{\beta_k(t,Y)\}_{k=1}^{NP}$ and $\{\xi_j(t,Y)\}_{j=1}^{Nh}$ are unknown; and

$$\psi_k(x,Y) = \sqrt{\frac{2}{b_2(Y)-b_2(Y)}}\sin\left(\frac{2k\pi(x-b_1(Y))}{b_2(Y)-b_1(Y)}\right),$$
$$b_1(Y) \le x \le b_2(Y) \tag{A3}$$

$$\phi_j(x) = \sqrt{\frac{2}{L}}\sin\left(\frac{2j\pi x}{L}\right), \quad 0 \le x \le L \tag{A4}$$

where $\{\psi_k(x,Y)\}_{k=1}^{NP}$ is a group of basis functions in the function space. They are of unit length and orthogonal to each other. The same properties also hold for $\{\phi_j(x)\}_{j=1}^{Nh}$.

Replacing $P(x,y)$ and $h(x,y)$ in the vibration equation (5) with Fourier expansion (Eqs. (A1)–(A4)), the residual of approximation is defined as the difference of the left and right sides,

$$r_h = m\left(\sum_{j=1}^{Nh} \ddot{\xi}_j \phi_j\right) + 2mV\left(\sum_{j=1}^{Nh} \dot{\xi}_j \frac{d\phi_j}{dx}\right) - (T - mV^2)\left(\sum_{j=1}^{Nh} \xi_j \frac{d^2\phi_j}{dx^2}\right) - c\left(P_{\text{atm}} + \sum_{k=1}^{NP} \beta_k \psi_k\right) \cdot [H(x-b_1) - H(x-b_2)] \quad \text{(A5)}$$

where $\ddot{\xi}_j$ stands for $d^2\xi(t,Y)/dt^2$.

By the global Galerkin method of weighted residual, multiplying Eq. (A5) by the shape functions of $h(x,y)$ given in Eq. (A4) and integrating over the interval $[0,L]$, one obtains

$$\int_0^L r_h \phi_i(x) dx = 0, \quad i = 1,2,\ldots,Nh \quad \text{(A6)}$$

Substituting Eq. (A5) into Eq. (A6), one gets

$$m\left(\sum_{j=1}^{Nh} \int_0^L \phi_i \phi_j dx \cdot \ddot{\xi}_j\right) + 2mV\left(\sum_{j=1}^{Nh} \int_0^L \phi_i \frac{d\phi_j}{dx} dx \cdot \dot{\xi}_j\right) - (T - mV^2) \times \left(\sum_{j=1}^{Nh} \int_0^L \phi_i \frac{d^2\phi_j}{dx^2} dx \cdot \xi_j\right) = cP_{\text{atm}} \int_{b_1}^{b_2} \phi_i dx + c\sum_{k=1}^{NP} \int_{b_1}^{b_2} \phi_i \psi_k dx \cdot \beta_k, \quad i = 1,2,\ldots,Nh$$

Rearranging these Nh equations into the matrix form, one obtains

$$\mathbf{Q}\ddot{\boldsymbol{\xi}} + \mathbf{C}\dot{\boldsymbol{\xi}} + \mathbf{K}\boldsymbol{\xi} = cP_{\text{atm}}\boldsymbol{\Phi} + \mathbf{E}\boldsymbol{\beta} \quad \text{(A7)}$$

where $\mathbf{Q} \in R^{Nh\times Nh}$, $\mathbf{C} \in R^{Nh\times Nh}$, $\mathbf{K} \in R^{Nh\times Nh}$, $\boldsymbol{\Phi} \in R^{Nh\times 1}$, and $\mathbf{E} \in R^{Nh\times NP}$,

$$\mathbf{Q}(i,j) = m\int_0^L \phi_i(x)\phi_j(x)dx = \begin{cases} m & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

$$\mathbf{C}(i,j) = 2mV\int_0^L \phi_i(x)\frac{d\phi_j(x)}{dx}dx = 0$$

$$\mathbf{K}(i,j) = -(T - mV^2)\int_0^L \phi_i(x)\frac{d^2\phi_j(x)}{dx^2}dx = \begin{cases} \frac{4j^2\pi^2}{L^2}(T - mV^2) & \text{if } i = j \\ 0 & \text{if } i \neq J \end{cases}$$

$$\boldsymbol{\Phi}(i) = \int_{b_1}^{b_2} \phi_i(x)dx, \quad \mathbf{E}(i,k) = c\int_{b_1}^{b_2} \phi_i(x)\psi_k(x,Y)dx$$

and

$$\boldsymbol{\xi} = (\xi_1\xi_2\cdots\xi_{Nh})^T, \quad \boldsymbol{\beta} = (\beta_1\beta_2\cdots\beta_{NP})^T$$

Notice that $\{\phi_j(x)\}_{j=1}^{Nh}$ are of unit length and are orthogonal to each other, i.e.,

$$\int_0^L \phi_i(x)\phi_j(x)dx = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \quad \int_0^L \phi_i(x)\frac{d\phi_i(x)}{dx}dx = 0$$

Therefore, $\mathbf{Q}$ and $\mathbf{K}$ are diagonal matrices; all elements of $\mathbf{C}$ are zeros. Equation (A7) is reduced to a group of linear constant coefficient second order ordinary differential equations (ODEs), which are independent of each other, in matrix form as

$$\mathbf{Q}\ddot{\boldsymbol{\xi}} + \mathbf{K}\boldsymbol{\xi} = cP_{\text{atm}}\boldsymbol{\Phi} + \mathbf{E}\boldsymbol{\beta}$$

or equivalent to

$$\ddot{\xi}_j + \frac{4j^2\pi^2}{mL^2}(T - mV^2)\xi_j = \frac{1}{m}\left(cP_{\text{atm}}\int_{b_1}^{b_2} \phi_j(x)dx + \mathbf{E}_{(j)}\boldsymbol{\beta}\right), \quad j = 1,2,\ldots,Nh \quad \text{(A8)}$$

where $\mathbf{E}_{(j)}$ stands for the jth row of $\mathbf{E}$.

A solution of ξ_1 is given as a function of $\{\beta_k\}_{k=1}^{NP}$ [19],

$$\xi_j = d_j \sin\left(t\sqrt{\frac{4j^2\pi^2}{mL^2}(T - mV^2)}\right) + \frac{L^2}{4j^2\pi^2(T - mV^2)}\left(cP_{\text{atm}}\int_{b_1}^{b_2} \phi_j(x)dx + \mathbf{E}_{(j)}\boldsymbol{\beta}\right), \quad j = 1,2,\ldots,Nh \quad \text{(A9)}$$

where $\{d_j\}_{j=1}^{Nh}$ should be fitted by the initial condition of the corresponding ODE equation (A8). Because Eq. (5) is a transient equation of the moving wire saw, when the neighborhood defined in Eq. (4) approaches the contact span $y=Y$, i.e., $\bar{t}\to 0$, one obtains a solution for the vibration of the wire saw in the contact span $y=Y$ as Eq. (14).

When set $\bar{t}\to 0$, the Fourier series expansions (Eqs. (A1) and (A2)) can be expressed as

$$P(x,y) = P_{\text{atm}} + \sum_{k=1}^{NP} \beta_k(Y)\psi_k(x,Y), \quad h(x,y) = \sum_{j=1}^{Nh} \xi_j(Y)\phi_j(x)$$

Generally speaking, $\{\beta_k(Y)\}_{k=1}^{NP}$ and $\{\xi_j(Y)\}_{j=1}^{Nh}$ are assumed to remain constant at different Ys; thus, they are rewritten as $\{\beta_k\}_{k=1}^{NP}$ and $\{\xi_j\}_{j=1}^{Nh}$. If the dependence of $\{\beta_k\}_{k=1}^{NP}$ and $\{\xi_j\}_{j=1}^{Nh}$ on Y is of interest, the method of piecewise estimation could be applied based on local data. Therefore, the common form of Fourier series expansions of $P(x,y)$ and $h(x,y)$ are given as Eqs. (10) and (11).

Appendix B: Derivation of the PDE-Constrained GP Model

Applying natural logarithm on both sides of Eq. (2), one obtains

$$\text{mean}(x,y,V) = \ln(a) + 2\ln(|V|) + \ln(u) + \ln(P(x,y)) - 2\ln(h(x,y)) \quad \text{(B1)}$$

In Eq. (B1), expanding $\ln(P(x,y))$ and $\ln(h(x,y))$ into the Taylor series, discarding high order terms and replacing $P(x,y)$ and $h(x,y)$ with Eqs. (10) and (11), one obtains an approximation equation,

$$\text{mean}(x,y,V) \approx \ln(a) + 2\ln(|V|) + \ln(u) + P_{\text{atm}} - 1 + \sum_{k=1}^{NP} \beta_k\psi_k - 2\left(\sum_{k=1}^{Nh} \xi_j\phi_j - 1\right) \quad \text{(B2)}$$

The wire saw slicing process is governed by the vibration equation; thus, Eq. (B2) has to be subjected to the constraints of Eq. (14). Inserting Eq. (14) into the Fourier expansion of $h(x,y)$ given in Eq. (11) yields

$$\sum_{j=1}^{Nh} \xi_j \phi_j = \sum_{j=1}^{Nh} \frac{L^2}{4j^2\pi^2(T-mV^2)} \left(cP_{\text{atm}} \int_{b_1}^{b_2} \phi_j(x)dx + \sum_{k=1}^{NP} \mathbf{E}_{j,k}\beta_k \right) \phi_j$$

$$= \frac{L^2 cP_{\text{atm}}}{4\pi^2(T-mV^2)} \sum_{j=1}^{Nh} \frac{1}{j^2} \left(\int_{b_1}^{b_2} \phi_j(x)dx \right) \phi_j + \frac{L^2}{4\pi^2(T-mV^2)} \sum_{k=1}^{NP} \left(\sum_{j=1}^{Nh} \frac{1}{j^2} \mathbf{E}_{j,k}\phi_j \right) \beta_k \quad \text{(B3)}$$

Substituting Eq. (B3) into Eq. (B2), one gets

$$\text{mean}(x,y,V) \approx \beta_0 + \sum_{k=1}^{NP} \beta_k f_k + \zeta + 2\ln(|V|) + \ln(u) \quad \text{(B4)}$$

where $\beta_0 = \ln(a) + P_{\text{atm}} + 1$ is an unknown constant value over the whole process and $\zeta(x,y)$ is given by the process variables and the shape functions (Eq. (12)),

$$\zeta(x,y) = -\frac{L^2 cP_{\text{atm}}}{2\pi^2(T-mV^2)} \sum_{j=1}^{Nh} \frac{1}{j^2} \left(\int_{b_1}^{b_2} \phi_j(x)dx \right) \phi_j \quad \text{(B5)}$$

$\{f_k\}_{k=1}^{NP}$ are defined by Eq. (17). Because the wire saw slicing system is symmetric about $x = L/2$, i.e., the midpoint of the contact span, there is

$$\int_{b_1}^{b_2} \phi_j(x)dx = 0$$

Therefore, there are $\zeta = 0$ and

$$\text{mean}(x,y) \approx \beta_0 + \sum_{k=1}^{NP} \beta_k f_k + \ln(u) + 2\ln(|V|)$$

Taking spatial correlation and manufacturing uncertainty factors into consideration, the PDE-constrained GP model is in the form of Eq. (16).

Appendix C: Some Details in Order Selection

In this study, $P(x,y)$ and $h(x,y)$ are assumed to mainly consist of low frequency components since both variables cannot have high frequency changes physically along the contact span. Figure 12 shows the estimations of the parameters $\{\beta_k\}_{k=1}^{NP}$ of the PDE-constrained GP model when $NP = 30$. The $|\beta_k|$ decreases evidently and is less than 0.005 when $k > 5$. In the performance comparison, the number of terms retained in the Fourier series is set as $NP = 5$. Therefore, to compare the model with the same number of the predictors in the mean part, the second order polynomial functions are selected as the mean part of the universal Kriging model, i.e.,

$$\mathbf{F}_2(x,y)\boldsymbol{\beta}_{\text{UK}} = \beta_{0,\text{UK}} + \beta_{1,\text{UK}}x + \beta_{2,\text{UK}}y + \beta_{3,\text{UK}}xy + \beta_{4,\text{UK}}x^2 + \beta_{5,\text{UK}}y^2$$

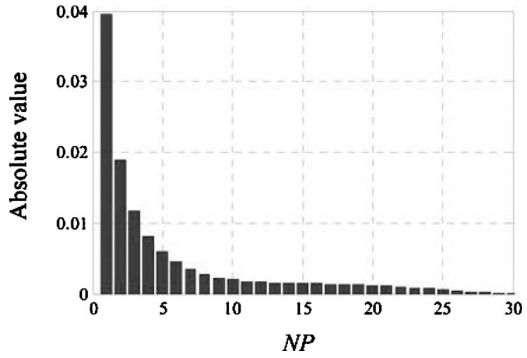

Fig. 12 The absolute values of the estimated parameters $\{\beta_k\}_{k=1}^{NP}$ when $NP = 30$

References

[1] Zhu, L., and Kao, I., 2005, "Galerkin-Based Modal Analysis on the Vibration of Wire–Slurry System in Wafer Slicing Using a Wire Saw," J. Sound Vib., **283**, pp. 589–620.

[2] Möller, H. J., 2004, "Basic Mechanisms and Models of Multi-Wire Sawing," Adv. Eng. Mater., **6**(7), pp. 501–513.

[3] Sahoo, R. K., Prasad, V., Kao, I., Talbott, J., and Gupta, K. P., 1998, "Towards an Integrated Approach for Analysis and Design of Wafer Slicing by a Wire Saw," ASME J. Electron. Packag., **120**, pp. 35–40.

[4] Möller, H. J., 2006, "Wafering of Silicon Crystals," Phys. Status Solidi A, **203**(4), pp. 659–669.

[5] Bhagavat, M., Prasad, V., and Kao, I., 2000, "Elasto-Hydrodynamic Interaction in the Free Abrasive Wafer Slicing Using a Wire Saw: Modeling and Finite Element Analysis," ASME J. Tribol., **122**(2), pp. 394–404.

[6] Zienkiewicz, O. C., and Taylor, R. L., 2000, *The Finite Element Method*, 5th ed., Butterworth-Heinemann, Oxford.

[7] Preston, F. W., 1927, "The Theory and Design of Plate Glass Polishing Machines," J. Soc. Glass Technol., **11**, pp. 214–256.

[8] Santner, T. J., 2003, *The Design and Analysis of Computer Experiments*, Springer, New York.

[9] Journel, A. G., and Huijbregts, C. J., 1978, *Mining Geostatistics*, Academic, London.

[10] Lakshmikumaran, A. V., and Wickert, J. A., 1999, "Equilibrium Analysis of Finite Width Tension Dominated Foil Bearings," ASME J. Tribol., **121**, pp. 108–113.

[11] Bhushan, B., 1999, *Principles and Applications of Tribology*, Wiley, New York.

[12] Hed, P. P., 1993, "Calculations of Material Removal, Removal Rate, and Preston Coefficient in Continuous Lapping/Polishing Machines," Materials Fabrication Division, Lawrence Livermore National Laboratory, Technical Report No. UCRL-ID-115321.

[13] Zhu, L., and Kao, I., 2001, "Equilibrium Elasto-Hydrodynamic Interaction Analysis in Wafer Slicing Process Using Wire Saw," *Proceedings of IMECE'01*, ASME, New York, EEP-Vol. 1, pp. 123–128.

[14] Swope, R. D., and Ames, W. F., 1963, "Vibrations of a Moving Threadline," J. Franklin Inst., **275**, pp. 36–55.

[15] Archibald, F. R., and Emslie, A. G., 1958, "The Vibration of a String Having a Uniform Motion Along Its Length," ASME J. Appl. Mech., **25**, pp. 347–348.

[16] Zhu, L., Bhagavat, M., and Kao, I., 2000, "Analysis of the Interaction Between Thin-Film Fluid Hydrodynamics and Wire Vibration in Wafer Manufacturing Using Wire Saw," *The Proceedings of IMECE'00*, ASME, Orlando, EEP-Vol. 28, pp. 233–241.

[17] Rasmussen, C. E., and Williams, C. I., 2006, *Gaussian Processes for Machine Learning*, MIT, Cambridge, MA.

[18] Cressie, N., 1993, *Statistics for Spatial Data*, Wiley, New York.

[19] Jeffrey, A., 2002, *Advanced Engineering Mathematics*, Harcourt/Academic, San Diego.