1-4-2017

Dirichlet Process Gaussian Mixture Models for Real-Time Monitoring and Their Application to Chemical Mechanical Planarization

Jia (Peter) Liu
Virginia Tech, jliu@vt.edu

Omer F. Beyca
Istanbul Technical University, beyca@itu.edu.tr

Prahalada K. Rao
University of Nebraska-Lincoln, rao@unl.edu

Zhenyu (James) Kong
Virginia Tech, zkong@vt.edu

Satish T. S. Bukkapatnam
Texas A&M University, satish@tamu.edu

Dirichlet Process Gaussian Mixture Models for Real-Time Monitoring and Their Application to Chemical Mechanical Planarization

Jia (Peter) Liu, Omer F. Beyca, Prahalad K. Rao, Zhenyu (James) Kong, *Member, IEEE*, and Satish T. S. Bukkapatnam

Abstract—The goal of this work is to use sensor data for online detection and identification of process anomalies (faults). In pursuit of this goal, we propose Dirichlet process Gaussian mixture (DPGM) models. The proposed DPGM models have two novel outcomes: 1) DP-based statistical process control (SPC) chart for anomaly detection and 2) unsupervised recurrent hierarchical DP clustering model for identification of specific process anomalies. The presented DPGM models are validated using numerical simulation studies as well as wireless vibration signals acquired from an experimental semiconductor chemical mechanical planarization (CMP) test bed. Through these numerically simulated and experimental sensor data, we test the hypotheses that DPGM models have significantly lower detection delays compared with SPC charts in terms of the average run length (ARL_1) and higher defect identification accuracies (F-score) than popular clustering techniques, such as mean shift. For instance, the DP-based SPC chart detects pad wear anomaly in CMP within 50 ms, as opposed to over 140 ms with conventional control charts. Likewise, DPGM models are able to classify different anomalies in CMP.

Note to Practitioners—This paper forwards novel Dirichlet process Gaussian mixture (DPGM) models for online process quality monitoring. The practical outcome is that the deleterious impact of process drifts on product quality is identified in their early stages using the presented DPGM models. For instance, sensor signal patterns from contemporary advanced manufacturing processes rarely follow distribution symmetry or normality assumptions endemic to traditional statistical process control (SPC) methods. These assumptions limit the effectiveness of traditional SPC methods for detection of process anomalies from complex heterogeneous sensor data. In comparison, the proposed DP-based SPC is capable of detecting process changes

in the sensor data notwithstanding the characteristics of the underlying distributions. Moreover, we show that the recurrent hierarchical DP clustering model identifies process anomalies with higher fidelity compared with traditional methods, such as mean shift clustering.

Index Terms—Chemical mechanical planarization (CMP), Dirichlet process (DP), DP Gaussian mixture (DPGM) models, fault detection, online process monitoring, recurrent hierarchical DP (RHDP) clustering.

I. INTRODUCTION

A. Motivation

THIS paper addresses two research questions from the perspective of sensor-based online process monitoring.

1) How to detect the onset of anomalous process states.
2) How to identify/classify different types of anomalous process states.

The first falls under the category of anomaly (fault) detection; the second under the purview of anomaly (fault) diagnosis. These questions are resolved using Dirichlet process Gaussian mixture (DPGM) models. The utility of the proposed DPGM models are demonstrated in the context of process monitoring in chemical mechanical planarization (CMP) using *in situ* wireless vibration sensor signals [1].

This research is valuable for mitigating process anomalies in the manufacture of ultraprecision high-value components. For instance, CMP-related defects, e.g., dishing and erosion, are attributed to be among the prominent reasons inhibiting semiconductor device yield rates [2], [3]. Therefore, there is a burgeoning need for *in situ* sensor-based quality monitoring approaches in ultraprecision manufacturing processes, such as CMP, so that incipient process anomalies can be identified and corrected at an earlier stage [1].

B. Challenges

Monitoring of complex manufacturing processes, e.g., CMP from sensors signals, is challenging due to the inherent nonlinear and non-Gaussian dynamics [4]. The complex process dynamics manifest prominently in the acquired sensor signals. For instance, Fig. 1 shows a representative wireless micro-electro-mechanical systems (MEMS) vibration sensor signal obtained from the experimental CMP test bed used in this paper (see Section IV). The signal time series in Fig. 1 has aspects that are evocative of nonlinear quasi-periodic dynamics, as revealed in [1].

Traditional statistical process control (SPC) charts and clustering techniques [e.g., mean shift, k-means, and expectation
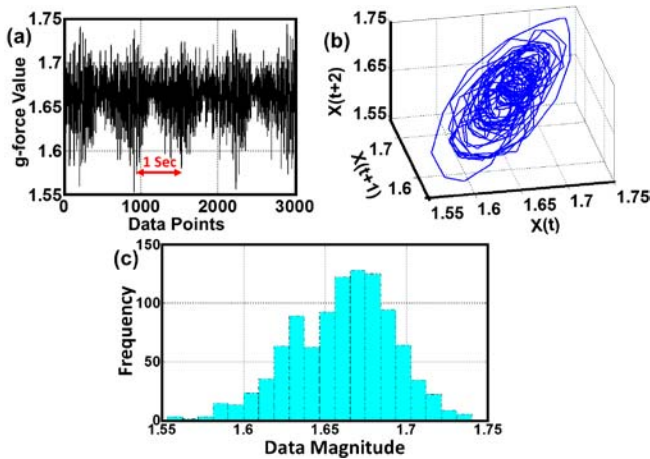
Fig. 1. Characteristics of a typical experimental CMP wireless vibration sensor signal sampled at ∼670 Hz (see Section IV). (a) Temporal signal with prominent cycles occurring at 1 s intervals and intermittent spikes. (b) State space plot of the signal data at three consecutive time instances ($X(t)$, $X(t+1)$, and $X(t+2)$) depicts quasi-periodic dynamics. (c) Histogram of vibration signal data reveals that the distribution is asymmetric and has bimodal tendency.

maximization (EM)] are ill suited for monitoring of such complex dynamics (e.g., Fig. 1) due to the following reasons.

- Parametric SPC charts have limited applicability to data with the following salient characteristics: the underlying distribution is non-Gaussian or multimodal, with persistent autocorrelation, or marked periodicity [5].
- Most distribution-free (nonparametric) SPC charts for non-Gaussian data are based on data ranking and constrained by the symmetry assumption of data distribution [6].
- Control charts cannot identify/localize the different process anomalies [7].
- Classical clustering techniques, for instance, mean shift, require sufficient high data density with clear gradients to locate the cluster centers, and *k*-means and EM methods require *a priori* assumptions regarding the number of clusters in the data [8]–[10].

C. Contributions and Novelty

We seek to overcome the aforementioned shortcomings in traditional SPC and clustering techniques using DPGM models for signal analysis [11]. DPGM models surmount limitations, such as normality and symmetry of data inherent in SPC charts. Furthermore, DPGM models inherently account for temporal autocorrelation in the data. Accordingly, the main contributions from DPGM models presented in this paper are as follows:

1) an SPC chart using DP mixture model for detection of anomalies;
2) unsupervised clustering of process states based on recurrent hierarchical DP (RHDP) clustering model, which is used for identifying/classifying specific process anomalies.

The rest of this paper is structured as follows. A review of the pertinent literature is presented in Section II. The methodology of DPGM models is detailed in Section III, which includes delineation of constitutive equations for DP-based SPC and RHDP clustering and further corroboration

with numerical studies. DPGM models are applied to experimentally acquired vibration signals from the CMP process in Section IV; and conclusions and avenues for further research are summarized in Section V.

II. REVIEW OF RELATED RESEARCH

The review of the relevant literature is conducted in the following two parts:

1) Recent developments in SPC.
2) Research in techniques such as neural networks (NNs), wavelet analysis, and Gaussian mixture models as enhancements to traditional SPC.

A. Statistical Process Control for Process Monitoring

Traditional parametric SPC charts, such as Shewhart X-bar and R, cumulative sum (CUSUM), and exponentially weighted moving average (EWMA), have been widely used in various scenarios ranging from manufacturing to service industries for process improvement [7]. Despite the underlying normality and independence assumptions, the effectiveness of Shewhart control charts have been attested; they are particularly useful for situations where subgrouped measurements can be made and the process shifts are significant (>1 standard deviation) [7]. CUSUM and EWMA control charts can be applied to both subgrouped and individual measurements, and are particularly suited for detecting small drifts. However, the latter (EWMA) are not *directionally invariant*, i.e., the control chart has a certain *inertia effect* in reacting to process drifts [7].

To overcome these restrictive assumptions with traditional parametric control charts, researchers devised nonparametric SPC charts, which are also called *distribution-free* SPC charts. Chakraborti *et al.* [12] provided a comprehensive review of nonparametric SPC charts. Although a specific type of distribution does not restrain nonparametric charts, nonetheless, most are based on data ranking methods, which entail that the data are implicitly assumed to be symmetric about the median.

To overcome this drawback, Qiu and Li [13] proposed a categorization-based nonparametric SPC chart for univariate data sequences. Their method relaxes the data symmetry assumption and is shown to be effective for non-Gaussian data. However, relying on *a priori* categorization of data for analysis results in information loss. Particularly, the selection of the number of groups for categorization, which is a heuristic parameter, is critical to the performance of control chart.

In the work of Qiu and Li [14], they devised nonparametric SPC charts leveraging Gaussian transformations, i.e., transforming data belonging to an unknown distribution to approximately Gaussian. However, the normality of transformed data cannot be universally guaranteed for cases where the data are patently multimodal and complex, such as in the CMP vibration signals used in this work.

To overcome these challenges, researchers have explored wavelet and NN-based SPC. These techniques can accommodate complex process dynamics, and have also been applied in CMP process [15].

B. Wavelet and Neural Network-Based Monitoring

Wavelet analysis has been successfully implemented in the modeling and monitoring of functional data in advanced manufacturing [16]. For instance, Ganesan *et al.* [17] developed

the wavelet-based SPC approach for real-time identification of delamination defects in CMP process.

Guo *et al.* [5] presented an approach that uses wavelet coefficients in an SPC setting for detecting process drifts. Their method involves multiscale decomposition of a signal using a predetermined Harr wavelet basis function. Subsequently, they tracked the wavelet coefficients at a predetermined optimal (wavelet) level using CUSUM and EWMA control charts. Jeong *et al.* [18] described a similar wavelet SPC procedure using the Symlet-8 wavelet basis function for functional data analysis of radio antenna reception patterns. Their approach uses a customized control chart with control limits derived from a statistic resembling multivariate Hotelling's T^2 [7].

Pugh [19] showed that feedforward NNs have significantly lower types I and II errors compared with traditional Shewhart X-bar and R charts and therefore could be valuable for process monitoring applications. Subsequently, Zorriassatine and Tannock [20] have developed methods that employ NNs for process monitoring applications. As an example of NN-based process monitoring, Rao *et al.* [21], [22] integrated a feedback delay embedded recurrent NN (RNN) with Bayesian particle filtering (PF) for real-time detection of mean shift in ultraprecision diamond turning process. The evolving surface morphology of diamond-turned workpieces is predicted in real time from *in situ* heterogeneous sensor data using PF-updated RNN weights. The network weights are subsequently monitored in an SPC setting using mean shift clustering [23].

Although these wavelet and NN-based SPC methods are applicable to complex signals without being constrained by the underlying assumptions of data distribution, they are nonetheless computationally demanding and engender a large number of variables that have to be tracked simultaneously. Moreover, these approaches require a predetermined model or basis function, such as the structure of the NN, and the basis and scaling function for wavelet decomposition. Therefore, decision uncertainty due to model selection remains a contentious challenge.

In contrast, SPC methods with Gaussian mixture modeling (GMM) overcome these aforementioned data distribution and model selection limitations. In this context, Choi *et al.* [24] and Thissen *et al.* [25] proposed PCA-based monitoring techniques, where GMM-derived models constructed via EM algorithms are used to approximate the data pattern. Similarly, Chen *et al.* [26] utilized infinite Gaussian mixture models to construct the control chart.

The DPGM models presented in this paper extend GMM-based SPC techniques toward both process anomaly detection and identification. It is noted that DP mixture models are the basis for constructing DPGM models. DP mixture modeling has been previously applied to data in several areas, such as medical images and online documents [27], [28]. DP mixture models approximate an empirical (arbitrary) data distribution via a mixture of finite Gaussian distributions without *a priori* knowledge of the number of mixture components. For instance, Rao *et al.* [29] applied DP mixture model to identify process anomalies in fused filament fabrication—an additive manufacturing process.
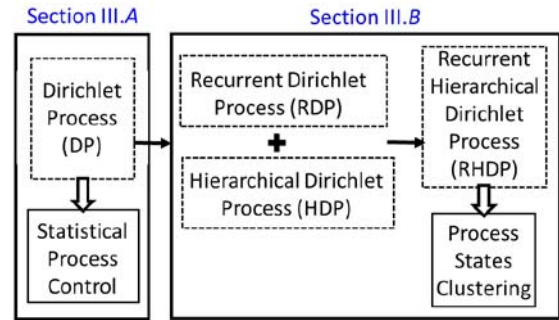

Fig. 2.   Overall methodology of DPGM modeling for online monitoring.

III. RESEARCH METHODOLOGY AND VERIFICATION WITH NUMERICAL SIMULATIONS

In this paper, we first propose DP-based SPC for detecting process anomalies. Then, toward identifying different types of process anomalies from a continuous sensor data stream, two extensions to DP models are forwarded to accommodate the following aims:

1) Estimating distribution characteristics *within* a contiguous time epoch, given the autocorrelation in the data.
2) Tracking the evolution of the data distribution *between* time epochs.

The first aim is realized using the recurrent DP (RDP) method proposed in [30]. The second involves using the hierarchical DP (HDP) developed in [31]. The integration of these two entities is accomplished in this paper, and is termed RHDP. Specifically, the RDP and HDP parts resolve the following questions.

- RDP: What are the characteristics of the data distribution at the current time epoch, given the knowledge of the distribution characteristics at the previous time epochs?
- HDP: What category is the process state (anomaly/fault) at the current time epoch, given the distribution characteristics estimated using RDP?

In other words, RDP determines the characteristics of data distribution at the current time epoch by including information from previous time epochs, while HDP enables data to be classified by allowing distributions falling under the same DP model to be clustered together. The overall framework of the proposed methodology is summarized in Fig. 2.

First, the DP mixture model is introduced, which approximates an empirical data distribution as a mixture of Gaussian component and develops DP-based SPC (Section III-A). Next, the novel RHDP model is elucidated for clustering process states and identifying process anomalies (Section III-B).

A. Dirichlet-Process-Based SPC for Monitoring Complex Non-Gaussian Waveforms

For time series data with complex nonlinear dynamics, such as the vibration sensor data acquired from the CMP process (Fig. 1), the data distribution may not be Gaussian. This poses a significant challenge for process modeling and monitoring with traditional methods based on the presumption

of normality and symmetry of data. Pertinently, since a non-Gaussian distribution can be modeled using a mixture of Gaussian distributions, such impediments could be overcome.

The essential concept of our DPGM models is to represent a non-Gaussian probability distribution as a mixture of multiple Gaussian distributions. This implication can be stated mathematically as follows:

$$p(x) = \sum_{j=1}^{k} \pi_j \mathbb{N}(x|\theta_j) \tag{1}$$

where x represents a time series collected by sensors from the process, $p(x)$ is its data distribution, and k is the number of Gaussian components $\mathbb{N}(\cdot)$ in the mixture, each of which is modeled with weight π_j and parameters θ_j (mean μ_j and variance σ_j^2). In reality, k may be unknown. We apply the DP mixture model, which is a data-driven nonparametric Bayesian approach to approximate a non-Gaussian distribution without any *a priori* knowledge of k [11]. The procedure is broadly elucidated in the forthcoming section.

1) Theoretical Development of DP Mixture Model: In DP, the limit for the number of clusters k goes to infinity [32]. In other words, when $k \to \infty$, the conditional prior distribution for the component indicators reaches its limit as follows:

$$p(c_i = j|\boldsymbol{c}_{-i}, \alpha)$$
$$\propto \begin{cases} \dfrac{n_{-i,j}}{N-1+\alpha}, & \text{if exsisting component } j \text{ is chosen} \\ \dfrac{\alpha}{N-1+\alpha}, & \text{if a new component is created} \end{cases} \tag{2}$$

where $\boldsymbol{c} = (c_1, \ldots, c_N)$ are the indicators of data points for components, α is the concentration parameter, n_j is the number of data points in Gaussian component j, and N is the number of data points, $N = \sum_{j=1}^{k} n_j$. The subscript $-i$ indicates all indices except i, and similarly, $n_{-i,j}$ indicates the number of observations in component j for all data points except point i.

For each component indicator c_i drawn conditioned on all other component indicators from the multinomial distribution, there is a corresponding component parameter θ_i drawn from a base distribution G_0. This result signifies a DP mixture model, which can be used to model a set of observations $(x_1, \ldots, x_i, \ldots, x_N)$, with latent variables of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_i, \ldots, \theta_N)$ as follows:

$$G \sim \text{DP}(\alpha, G_0)$$
$$\theta_i \sim G$$
$$x_i \sim \mathbb{N}(\cdot|\theta_i) \tag{3}$$

where $\text{DP}(\alpha, G_0)$ is the DP with base distribution G_0 and concentration parameter α, G is a random discrete distribution drawn from $\text{DP}(\alpha, G_0)$, each θ_i is drawn from the discrete distribution G, and each data point x_i (which may include statistical features, e.g., mean and variation, from the sensor data) is drawn from a normal distribution with parameter θ_i. Because the empirical distribution G is discrete, the same values can be assigned to multiple θ_i. Data points that have the same latent value belong to the same component [11], [32].

Furthermore, on integrating out G, the following conditional distribution for θ_i is obtained [33]:

$$\theta_i|\boldsymbol{\theta}_{-i}, G_0, \alpha \sim \frac{\alpha}{N-1+\alpha} G_0 + \sum_{j=1\cup j \neq i}^{N-1} \frac{1}{N-1+\alpha} \delta(\theta_j) \tag{4}$$

where $\delta(\theta_j)$ is the Dirac delta function peaked on θ_j. Subsequently, combining the prior distribution for θ_i of data i in (4) and the likelihood function in (3) results in the following posterior distribution for Gaussian component parameters:

$$p(\theta_1 = j|\boldsymbol{\theta}_{-i}, x_i)$$
$$\sim \begin{cases} \dfrac{n_{-i,j}}{N-1+\alpha} \mathbb{N}(x_i|\theta_j), & \text{if existing component} \\ & j \text{ is chosen (a)} \\ \dfrac{\alpha q}{N-1+\alpha} H(\theta|x_i), & \text{if a new component is created (b)} \end{cases} \tag{5}$$

where $q = \int G_0(\theta)\mathbb{N}(x_i|\theta)d(\theta)$, $H(\theta|x_i) = (G_0(\theta)\mathbb{N}(x_i|\theta)/\int G_0(\theta)\mathbb{N}(x_i|\theta)d(\theta))$. Equation (5a) shows the probabilities of θ_i having the same value with the existing Gaussian component parameter θ_j, and (5b) is the posterior probability of θ_i choosing a new value that is randomly generated from $H(\theta|x_i)$.

2) Dirichlet-Process-Based Statistical Process Control Chart: A control chart is a visual tool that is used for monitoring whether a process or system at a given time is under the influence of common cause (chance) variation or special cause (assignable) variation [6], [7]. The limits of the control charts represent thresholds that are obtained when a system operates wholly under common cause variation [in-control (IC) condition]. In the IC condition, the monitoring statistic falls within the control limits threshold. If special causes take effect, the control chart should presumably signal a change in terms of the monitoring statistic drifting outside the control limits [out-of-control (OOC) condition]. Thus, the control chart is effectively a two-state or binary classifier as it signals only IC or OOC process states. The control chart does not identify, explicitly, the type of anomaly/special cause.

In DP-based SPC, the likelihood values of new data are calculated under IC data distribution acquired by the DP mixture model. Process changes are detected once the likelihood values drop, indicating that such data are not likely generated under IC condition. An effective way to detect the OOC operation is by monitoring the average log-likelihood value in a subgroup of incoming data under IC data distribution as in

$$\frac{1}{w} \log[L(x_1, \ldots, x_i, \ldots, x_w|\theta_1, \theta_2, \ldots, \theta_j)]$$
$$= \frac{1}{w} \sum_{i=1}^{w} \log \left[\sum_{k=1}^{j} \pi_k \mathbb{N}(x_i|\theta_k) \right] \tag{6}$$

where x_i is the incoming data, j is the number of components for mixture distribution for IC condition, and w is the subgroup size of testing data. The larger the value of w, the more reliable the detection of OOC operation, but a longer delay is caused to detect process changes. Based on empirical results, we choose w as the minimal number of observations to achieve

the average run length (ARL$_0$) for type I error of likelihood values below a certain value, e.g., 0.05, in order to balance fast detection and detecting accuracy.

By the central limit theorem, the average log-likelihood values of incoming data are approximately normally distributed. Therefore, the problem of monitoring original complex non-Gaussian data reduces to a scheme of monitoring normally distributed average log-likelihood values. For simplicity, we construct the DP-based SPC by closely emulating the framework of the CUSUM chart with the average log-likelihood values as the monitoring target. Therefore, representing the average log-likelihood value in time epoch t as y_t, we have monitoring statistics for DP-based SPC

$$\begin{cases} C_t^+ = \max\left[0, y_t - (\mu_0 + K) + C_{t-1}^+\right] \\ C_t^- = \max\left[0, (\mu_0 - K) - y_t + C_{t-1}^-\right] \end{cases} \quad (7)$$

with the control limits (threshold) for the chart set at

$$H = L\sigma \quad (8)$$

where σ and μ_0 are the standard deviation and mean of the sequential data y_t under IC condition, respectively, the parameters K and L are adjusted for a given average run length criteria (ARL$_0$) [7], and the average log-likelihood value y_t is obtained from (6). The CUSUMs C_t^+ and C_t^- are tracked over time; if these quantities are greater than H, then OOC status is signaled. The CUSUMs C_t^+ and C_t^- are never negative. The implementation of DP-based SPC is introduced in Appendix I.

3) Application of DP-Based SPC for Process Monitoring—Numerical Case Studies: In this section, we show that DP-based SPC can capture changes in the data despite the underlying distribution being asymmetric and multimodal. We compare the results with two conventional control charts, namely, EWMA and CUSUM [7]. The traditional control charts monitor the raw data values, while the DP-based SPC uses the average log-likelihood values in (6) within the CUSUM framework in (7). For comparison purposes, we use the following two average run length (ARL) criteria as widely used for performance evaluation of control charts: 1) ARL$_0$ and 2) ARL$_1$ [7].

We now test the hypothesis that DP-based SPC has the superior ability (i.e., smaller ARL$_1$) in capturing the changes in incoming data compared with EWMA and CUSUM given identical ARL$_0$. The following three scenarios are investigated.

- Case N1: Detecting mean shifts in univariate unimodal Gaussian and non-Gaussian distributions.
- Case N2: Detecting mean shifts in univariate multimodal non-Gaussian distributions.
- Case N3: Detecting shifts in multivariate, nonlinear, and quasi-periodic data from the Rössler chaotic attractor [34].

a) Case N1 (DP-based SPC for data from univariate unimodal Gaussian and non-Gaussian distributions): The aim of this paper is to ascertain the ARL$_1$ performance of DP-based SPC toward detecting a shift in mean (location parameter) of a distribution. Furthermore, we contrast the ARL$_1$ performance of DP-based SPC with those of EWMA and CUSUM control charts.
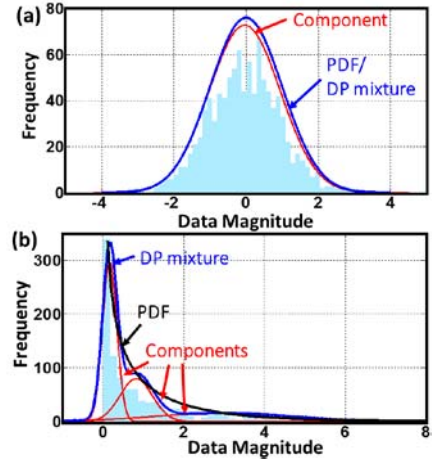

Fig. 3. True PDF, Gaussian components, and approximated distribution by DP mixture model for data generated from (a) $\mathbb{N}(\mu, 1)$ and (b) χ_1^2.
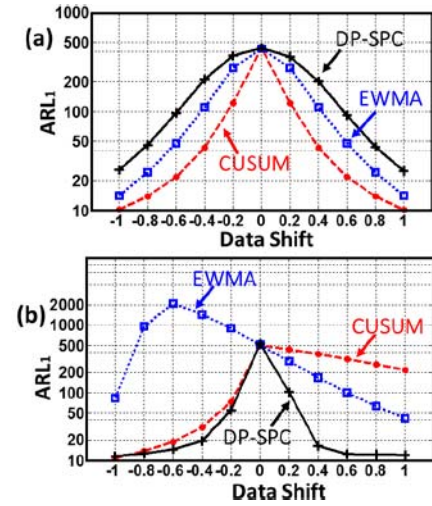

Fig. 4. Case N1 results—OOC ARL$_1$ values of different control charts when ARL$_0$ is fixed at 500 and actual IC data generated from (a) $\mathbb{N}(\mu, 1)$ and (b) χ_1^2. Scale on the y-axis is in natural logarithm.

This study is conducted with data generated from two basic univariate distributions: the Gaussian distribution with mean μ and variance 1, $\mathbb{N}(\mu, 1)$; and the Chi-squared distribution with one degree of freedom, χ_1^2. The mean of these distributions will be shifted from the IC state of zero mean, and ARL$_1$ will be evaluated for CUSUM, EWMA, and DP-based SPC.

We note that the latter distribution (χ_1^2) is inherently asymmetric (right skewed) and theoretically equivalent to a F distribution, $F(1, \infty)$. Both the Gaussian and Chi-squared probability density functions (PDF) are approximated by mixtures of Gaussian components using DP mixture model, as exemplified in Fig. 3.

The OOC data are obtained by mean shift, ranging from -1.0 to 1.0 with a step of 0.2. The control limit is acquired by adjusting parameter L in (8) to obtain average ARL$_0$ in 5000 repetitions at 500 under IC condition; the ARL$_1$ values are reported based on 10 000 replications. The average ARL$_1$ results for EWMA, CUSUM, and DP-based SPC are reported in Fig. 4. The following observations can be tendered from Fig. 4.

1) Fig. 4(a): When the normality assumption is not violated, as the case with the Gaussian distribution $\mathbb{N}(\mu, 1)$,
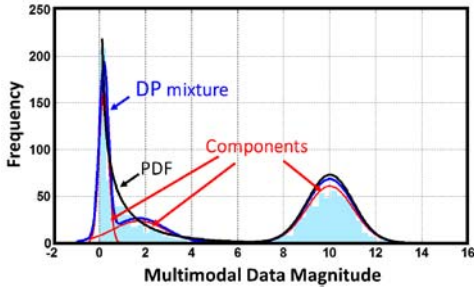
Fig. 5. True PDF, Gaussian components, and fitted distribution by DP model for data generated from a bimodal distribution consisting of $\mathbb{N}(10, 1)$ and χ_1^2.
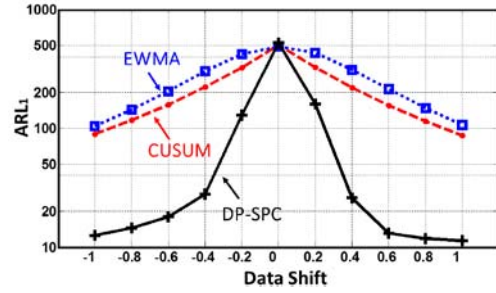

Fig. 6. Case N2 results—OOC ARL_1 values of three SPC methods when ARL_0 is 500 and actual IC data are generated from a bimodal distribution consisting of $\mathbb{N}(10, 1)$ and χ_1^2. The scale on the *y*-axis is in natural logarithm.
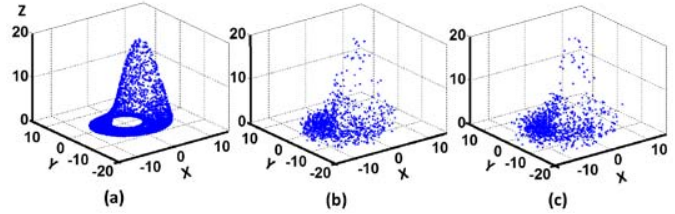

Fig. 7. (a) Rössler attractor delineated in (9). (b) Sample 1000 data points from the Rössler attractor contaminated with white noise $\mathbb{N}(0, \mathbf{I_3})$. (c) New sample from the approximated distribution of the Rössler attractor.

CUSUM and EWMA perform better (lower ARL_1) than DP-based SPC.

2) Fig. 4(b): If the data are patently non-Gaussian, i.e., the normality condition as in the case of χ_1^2 is violated, then the ARL_1 of DP-based SPC is smaller than those of EWMA and CUSUM control charts.

b) Case N2 (DP-based SPC for data from univariate multimodal non-Gaussian distributions): In this case study, the data are obtained from an underlying bimodal distribution consisting of $\mathbb{N}(10, 1)$ and χ_1^2. As evident in Fig. 5, the DP mixture model closely approximates the data distribution, which corroborates our assertion that the DP mixture model can capture complex distributions.

As in the previous case (Case N1), OOC data are obtained by shifting the mean of the data in the range of -1.0 to 1.0 with a step of 0.2. Once again, the control limit is acquired by adjusting parameter L in (8) to obtain average ARL_0 of 5000 repetitions at 500 under IC condition, and ARL_1 results from a 10 000-replication study are reported (Fig. 6).

It can be inferred from Fig. 6 that under a multimodal distribution, and when the data are patently non-Gaussian and asymmetric, the performance of EWMA and CUSUM is considerably inferior to DP-based SPC; the ARL_1 of DP-based SPC is smaller than those of EWMA and CUSUM. Indeed, the performance of the DP-based SPC is almost identical to Fig. 4(b), thus further affirming that the DP-based SPC is not influenced by symmetry and modes of the underlying data.

c) Case N3 (DP-based SPC for multivariate, nonlinear, and quasi-periodic data): Real-world signals customarily portray strong nonlinearity and high dimensionality; such a behavior has been observed in several practical instances in manufacturing processes, including CMP [22], [35], [36]. In this case study, we show that the DP mixture model can accommodate multidimensional data depicting nonlinear quasi-periodic dynamics [36].

The 3-D Rössler system, as delineated in (9), is used in this case study [34]; it consists of three coupled ordinary differential equations to define a continuous-time dynamical system, which exhibits chaotic nonlinear behavior predicated by the choice of three parameters, namely, a, b, and c in [34]

$$\frac{dx}{dt} = -y(t) - z(t)$$
$$\frac{dy}{dt} = x(t) + a \cdot y(t)$$
$$\frac{dz}{dt} = b + z(t) \cdot [x(t) - c]. \quad (9)$$

We fix the parameters as follows: $a = 0.2$, $b = 0.2$, and $c = 5$. The Rössler system depicts prominent chaotic dynamics under these conditions; the dynamics of the Rössler system has been extensively investigated in [34]. The Rössler attractor state phase diagrams obtained as a result of (9) are shown in Fig. 7(a). We note that in this simulation, data generated from the Rössler system of (9) are purposely contaminated with Gaussian white noise $\mathbb{N}(0, \sigma^2 \mathbf{I_3})$, where $\mathbf{I_3}$ is the identity matrix of order 3; the effect of variance σ^2 on ARL_1 of DP-based SPC is tested in this case study. Shown in Fig. 7(b) is a sample of 1000 data points from the contaminated Rössler attractor. Next, DP mixture model is used to approximate the data distribution of Rössler contractor using a mixture of multivariate Gaussian distributions. One thousand new data points are generated from the DP approximated distribution of the contaminated Rössler attractor, as shown in Fig. 7(c). It is apparent from Fig. 7(c) that the data generated from the DP approximated distribution closely resemble the data sampled from the contaminated Rössler attractor [Fig. 7(b)]. The Chi-square goodness of fit (GoF) test attests that there is no significant difference between the actual and DP approximated data in Fig. 7(b) and (c), respectively.

In order to detect the effects of mean and variance shifts, OOC data are generated as follows [the IC state is the data obtained from (9) with white noise $N(0, \mathbf{I_3})$].

- for the mean shift case, OOC data are obtained by translating the original data from (9) in all directions $(x(t), y(t), z(t))$ in the range of 0.5–2.5 (step size 0.5);
- for variance shifts, the OOC data are obtained by contaminating original data with different levels of Gaussian noise $\mathbb{N}(0, \sigma^2 \mathbf{I_3})$ with variance (σ^2) ranging from 1.5 to 4 (step size 0.5).

The ARL_0 values of the multivariate extension of EWMA (MEWMA), Hotelling's T^2 multivariate control chart, and DP-based SPC are maintained at 500 to obtain the control
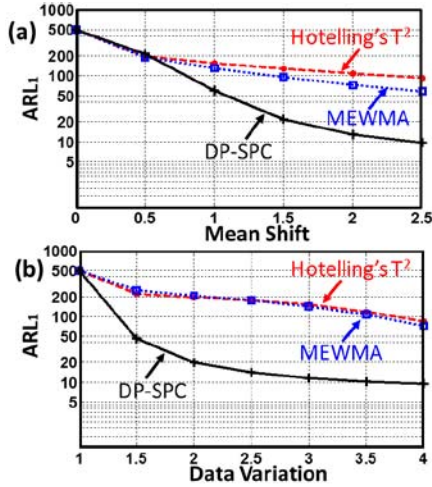
Fig. 8.   Case N3 results—OOC ARL_1 values of three SPC methods when IC ARL_0 is 500. The scale on the y-axis is in natural logarithm. (a) ARL_1 results for mean shifts. (b) ARL_1 detection for data variation σ^2.

limit, and ARL_1 is assessed [7]. We use Hotelling's T^2 instead of CUSUM, because Hotelling's T^2 is easier to implement than to extend CUSUM to the multivariate case, and it is also considered one of the standard multivariate control charts [37].

As in previous cases (Cases N1 and N2), the ARL_1 results from 10 000 replications are reported for the three control charts. The ARL_1 results of the DP-based SPC are compared with Hotelling's T^2 and MEWMA in Fig. 8. It can be inferred, based on the evidence presented in Fig. 8, that the DP-based SPC has significantly smaller ARL_1, i.e., DP-based SPC is able to detect data shifts and variability earlier than either of the conventional control charts compared (Hotelling's T^2 and MEWMA) for multivariate, nonlinear, and quasi-periodic data.

Note on Computational Time: For univariate data (Cases N1 and N2), the computational time of DP-based SPC is about 0.02 ms per data point and those of EWMA and CUSUM are about 0.005 ms per data point; for the complex multivariate data (Case N3), the computational time of DP-based SPC is about 0.2 ms per data point and those of MEWMA and Hotelling's T^2 are about 0.02 ms per data point (with Intel Core i7-4770 CPU at 3.40 GHz). Although DP-based SPC is slower than traditional SPC charts, it is fast enough (∼50 kHz for 1-D data and ∼5 kHz for 3-D multivariate data) to handle many manufacturing processes (e.g., in CMP, the sampling frequency of vibration sensors is ∼670 Hz), and it is superior in monitoring complex signal data.

B. Recurrent Hierarchical Dirichlet Process for Evolutionary Clustering of Process States

For complex manufacturing processes, each process state manifests in unique signal distributions. A control chart cannot classify the differences in process states because the control limits are estimated based on the so-called IC state alone. In order to identify the specific process anomalies (drifts), we herewith propose RHDP clustering. And we demonstrate the utility of the RHDP clustering via a simulation-based study.

1) Theoretical Development of RHDP Model: As noted previously at the beginning of Section III, there are two elements that are critical toward formulating the RHDP model, namely, RDP model, and the HDP model. We will cover these elements in the forthcoming sections and finally demonstrate the development of the RHDP model.

a) Recurrent Dirichlet process model: The basic DP modeling delineated in Section III-A assumes that data points $(x_1, x_2, \ldots, x_N)$ are fully *exchangeable*. In other words, the autocorrelation of the data is not accounted for in the basic DP modeling. This assumption is not detrimental for DP-based SPC, since the IC condition model does not rely on the temporal order of IC data. However, it is critical to overcome this *exchangeability* limitation for classifying process states from sequential data.

Specifically, the exchangeability limitation can be surmounted using the RDP model proposed in [30], which divides time series data into contiguous sequential epochs (windows); data points within the same epoch are assumed to be exchangeable, while the temporal order is maintained across epochs. Thus, the autocorrelation in consecutive epochs is accounted for in the RDP model. In the implementation of RDP, the incoming sensor data are divided using a sliding window technique; the data inside a sliding window are an epoch.

Furthermore, the time lag of two consecutive epochs is controlled by arranging the data overlap between the two windows. For instance, if the first epoch consists of a set of data points $(x_1, x_2, \ldots, x_{N-1})$, then the second epoch will consist of data points $(x_2, x_3, \ldots, x_N)$ using time lag of one.

In the RDP model, the data point x_i^t at time epoch t is generated from a distribution with latent variable θ_i^t. The conditional distribution of θ_i^t can be formulated using the information from an immediately preceding epoch as follows [30]:

$$\theta_i^t \big| \boldsymbol{\theta}_{\cdot}^{t-1}, \theta_1^t, \ldots, \theta_{i-1}^t, G_0, \alpha \sim$$

$$\frac{1}{N^{t-1}+i-1+\alpha} \left[\sum_{j \in J^{t-1} \cup J^t} \left(n_{\cdot j}^{t-1} + n_{-i,j}^t \right) \delta(\phi_j^t) + \alpha G_0 \right] \tag{10}$$

where θ_i^t is the distribution parameter for data x_i^t at time epoch t, $\boldsymbol{\theta}_{\cdot}^{t-1}$ stands for distribution parameters for all data at previous time epoch $t-1$, ϕ_j^t is the distribution parameter value of mixture component j at time epoch t, $\delta(\phi_j^t)$ is the Dirac delta function peaked on ϕ_j^t, N^{t-1} denotes the number of data points at time epoch $t-1$, $n_{\cdot j}^{t-1}$ is the number of data points associated with mixture component j at time epoch $t-1$, $n_{-i,j}^t$ is the number of data points associated with mixture component j at time epoch t with data point i excluded, and J^t includes all the unique mixture components at time epoch t.

In accordance with these conditions, in a given epoch t, an observation belonging to component j is proportional to the number of observations in component j at that epoch t, plus the number of observations in component j at previous epoch $t-1$. Subsequently, the distribution parameters of a particular epoch can be estimated by invoking the RDP model. Component assignments are made using the Gibbs sampling algorithm.
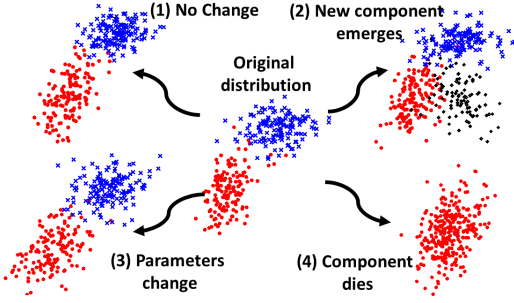
Fig. 9. Possible evolutions of data distribution in a physical process.

b) Recurrent hierarchical Dirichlet process model: Although we have estimated the data distribution at each time epoch in a sequential manner using RDP, we have not yet considered the evolution of distributions *between* time epochs. For instance, during the physical process, as depicted in Fig. 9, the following four possibilities exist:

1) the signals dynamics remain stationary (no change);
2) new components may emerge;
3) the parameters of mixture components may change over time;
4) an existing component may die.

Therefore, to classify the process, it is essential to track the evolution of mixture components between time epochs.

The concept of HDP, which is essentially a multiple-level DP, as proposed in [31], could be adopted in this context. Unlike with DP, in a two-level HDP model, the *parent* DP G_0 is a random variable distributed with concentration parameter γ and base distribution H. The so-called *child* DPs $G'_j s$ have concentration parameter α and base distribution G_0. Since G_0 is discrete, the child DPs $G'_j s$ share atoms (mixture components) with each other. Data distributions parametrized by $G'_j s$ with the same atoms will have the similar Gaussian components, and therefore could be clustered together [31].

In this way, clustering of data distributions can be achieved by HDP. In order to monitor the evolution in RDP-formulated data distributions *between* consecutive time epochs, we propose a novel approach, which integrates RDP with HDP, and consequently term the resulting approach as RHDP.

Given a temporal data set, RHDP model could be used to monitor the distribution evolution among multiple sequential epochs, and subsequently, the time epochs with similar distribution characteristics can be grouped/clustered together [31], [38]. The RHDP model is formulated as

$$G_0|\gamma \sim \mathrm{DP}(\gamma, H)$$
$$G^t|\alpha \sim \mathrm{DP}(\alpha, G_0)$$
$$\theta_i^t|G^t \sim G^t$$
$$x_i^t|\theta_i^t \sim \mathbb{N}(\cdot|\theta_i^t) \quad (11)$$

where x_i^t for $i = 1, \ldots, N^t$ are observations in time epoch t and $N(\cdot|\theta_i^t)$ denotes the Gaussian component parameterized by θ_i^t, which is sampled from the child DP G^t.

Data in different epochs are modeled by using Gaussian mixture distributions with parameters $\{\theta_1^t \ldots, \theta_{N^t}^t\}$ sampled from G^t. If the process is stationary, the parameters of the mixture distribution would remain constant. However, if there

is a change in the underlying process, entailing a change in the data distribution, the current data distribution will not suit the new data, i.e., the existing parameters drawn from G^t will not appropriately model the new data. Accordingly, new samples for G_0 will be drawn from the base function H of parent DP.

We can estimate marginal distributions of the mixture component at two levels of DP by integrating out G_0 and G^t. The conditional distribution for θ_i^t can be calculated by integrating out G^t as follows:

$$\theta_i^t|\boldsymbol{\theta}^{t-1}, \theta_1^t, \ldots, \theta_{i-1}^t, G_0, \alpha \sim$$
$$\frac{1}{N^{t-1} + i - 1 + \alpha} \left[\sum_{j \in J^{t-1} \cup J^t} (n_{\cdot j}^{t-1} + n_{-i,j}^t)\delta(\phi_j^t) + \alpha G_0 \right] \quad (12)$$

where ϕ_j^t represents the distribution parameter of the mixture component j at time epoch t.

Although (12) appears to be exactly the same as (10), they have one significant difference—note that in (12), G_0 is not fixed, but distributed as DP.

The subsequent step is to integrate out G_0 to get the conditional distribution for ϕ_j^t. Since G_0 is distributed as DP, it can be integrated out as follows:

$$\phi_j^t|\boldsymbol{\phi}^{t-1}, \phi_1^t, \ldots, \phi_{j-1}^t, H, \gamma \sim$$
$$\frac{1}{M^{t-1} + j - 1 + \gamma} \left[\sum_{l \in L^{t-1} \cup L^t} (m_{\cdot l}^{t-1} + m_{-j,l}^t)\delta(\tau_l) + \gamma H \right] \quad (13)$$

where τ_l denotes a value drawn from base distribution H, M^{t-1} is the number of all Gaussian components in epoch $t-1$, $m_{\cdot l}^{t-1}$ is the number of Gaussian components associated with τ_l at time epoch $t-1$, $m_{-j,l}^t$ is the number of the Gaussian components except components j associated with τ_l at time epoch t, and L^t denotes the collection of samples drawn from H at epoch t.

Subsequently, we obtain the posterior probability distributions for both the component values of DP G_0 in (14) and its child DP G^t in (15)

$$p(\phi_j^t = l|\boldsymbol{\phi}^{t-1}, \phi_1^t, \ldots, \phi_{j-1}^t, \{c(x_i^t) = j\}) \sim$$
$$\begin{cases} (m_{\cdot l}^{t-1} + m_{-j,l}^t)F(\phi_j^t|\tau_l), & \text{if component } l \text{ is chosen} \\ \gamma s T(\tau|\phi_j^t), & \text{if a new component is created} \end{cases}$$
$$(14)$$

where $s = \int H(\tau)F(\phi_j^t|\tau)d(\tau)$, $T(\tau|\phi_j^t) = (H(\tau)F(\phi_j^t|\tau)/ \int H(\tau)F(\phi_j^t|\tau)d(\tau))$ and $F(\phi_j^t|\tau_l)$ is the probability of ϕ_j^t getting the value of τ_l, which can be represented by likelihood of all data belonging to the component j in the mixture distribution at epoch t (i.e., all data with indicator $c(x_i^t) = j$)

$$p(\theta_i^t = j|\boldsymbol{\theta}^{t-1}, \theta_1^t, \ldots, \theta_{i-1}^t, x_i^t) \sim$$
$$\begin{cases} (n_{\cdot j}^{t-1} + n_{-i,j}^t)\mathbb{N}(x_i^t|\theta_j^t), & \text{if component } j \text{ is chosen} \\ \alpha q R(\theta|x_i^t), & \text{if a new component is created} \end{cases}$$
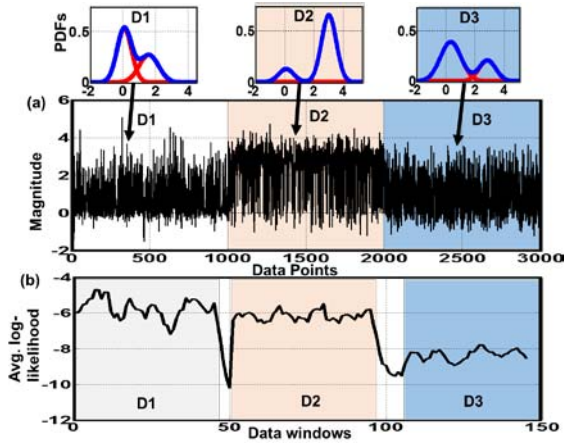$$(15)$$

Fig. 10. (a) Generated three-part data from different Gaussian mixture distributions. (b) Average log-likelihood values of data in time epochs. Three different shades indicate data from distributions D1, D2, and D3, and the white areas between different parts of data are transition periods.

where $q = \int G_0(\theta)\mathbb{N}(x_i^t|\theta)d(\theta)$, $R(\theta|x_i^t) = (G_0(\theta)\mathbb{N}(x_i^t|\theta)/ \int G_0(\theta)\mathbb{N}(x_i^t|\theta)d(\theta))$.

In (15), x_i^t represents data observation i during time epoch t, θ_i^t represents the distribution parameter for data x_i^t at time epoch t, and ϕ_j^t represents the jth atom value of child DP G^t (i.e., the mixture component j at time epoch t). If the base distributions H is Gaussian, i.e., it is conjugate with distribution of observations, then the integrals in (14) and (15) have analytical solutions.

RHDP could attain unsupervised clustering of process states by monitoring the change of mixture components among time epochs, i.e., the evolution of sequential data distributions. RHDP clustering includes the following two major steps.

1) RDP modeling is used for sequential process data, which are segregated into sliding windows. Gibbs sampling is adopted to update the data distribution, and Pearson's Chi-square GoF test is used to evaluate the accuracy of distribution modeling.

2) Cluster data of which the mixture distributions are from the same realizations in HDP. The average log-likelihood value of current data under previous distribution is continuously calculated and monitored as follows:

$$\frac{1}{w}\log\left[L\left(x_1^t,\ldots,x_i^t,\ldots,x_w^t|\theta_1^{t-1},\theta_2^{t-1},\ldots,\theta_j^{t-1}\right)\right]$$
$$= \frac{1}{w}\sum_{i=1}^{w}\log\left[\sum_{k=1}^{j}\pi_k\mathbb{N}\left(x_i^t|\theta_k^{t-1}\right)\right] \quad (16)$$

where x_i^t is the incoming data in epoch t and j is the number of components in epoch $t-1$. If the average log-likelihood values calculated as (16) remain stable and without significant drop, it indicates that the data in these consecutive windows have the similar distribution and, therefore, could be clustering as one process state. This is computationally amenable than tracking the change of mixture components. The implementation of RHDP clustering is introduced in Appendix II.

In this way, by tracking the evolution of mixture distributions at consecutive time epochs using the RHDP model,

TABLE I
F-SCORE RESULTS FOR DATA SERIES WITH THREE DISTRIBUTIONS —
COMPARISON OF RHDP CLUSTERING VERSUS MEAN SHIFT
(THE VALUES IN THE PARENTHESIS ARE
THE STANDARD DEVIATION)

	RHDP Clustering			Mean Shift		
	D_1	D_2	D_3	D_1	D_2	D_3
Precision	0.9872 (0.0117)	0.9784 (0.0150)	0.9826 (0.0097)	1 (0)	0.6573 (0.2644)	0.9825 (0.0304)
Sensitivity	0.9913 (0.0194)	1 (0)	1 (0)	0.9263 (0.0798)	0.9889 (0.0248)	0.6852 (0.3060)
F-score	0.9892	0.9891	0.9912	0.9617	0.7897	0.8074
Average F-score	0.9898			0.8592		

process drifts in complex manufacturing processes, such as semiconductor CMP, can be monitored and different process states (e.g., different anomalies) can be identified. We demonstrate this assertion herewith using a numerical example.

2) RHDP Clustering Analysis for Simulated Data in Sequential Epochs: The aim of this case study is to demonstrate the ability of the RHDP clustering to group non-Gaussian nonstationary sequentially acquired time series data. We show that using numerically generated data, the unsupervised clustering technique of RHDP identifies specific process states contingent on their data distributions.

As noted in the preceding section, we continuously monitor the average log-likelihood values of new data as in (16). For a stationary process, the data distribution does not change over time, and therefore, the average log-likelihood values remain stable. If the average log-likelihood values were to change dramatically, it indicates that the current data are not generated from the existing distribution but from a new one. Therefore, all the time epochs preceding the change of log-likelihood values are grouped into the same cluster, given their distribution similarity. In addition, we note that a transition period between two process states is inevitable, because the RHDP model splits the data into time epochs (windows), and consequently, some windows will contain data from two temporally adjacent process states.

In this paper, we define the following three mixture distributions from which the data are sequentially generated [Fig. 10(a)].

- D_1: $x_t \sim 0.5\mathbb{N}(0, 0.2) + 0.5\mathbb{N}(1.5, 1)$.
- D_2: $x_t \sim 0.2\mathbb{N}(0, 0.5) + 0.8\mathbb{N}(3, 0.5)$.
- D_3: $x_t \sim 0.5\mathbb{N}(0, 0.5) + 0.5\mathbb{N}(2, 0.7)$.

Referring to Fig. 10(a), it is observed that the data naturally cluster into three parts, as shaded by different colors. The corresponding average log-likelihood values as estimated using RHDP are shown in Fig. 10(b). The unshaded parts indicate the transition periods.

We report results from a ten-replication study and compare the clustering results from RHDP with the mean shift method [23], a frequently used unsupervised clustering method; mean shift uses the raw data as opposed to utilizing average log-likelihood values by RHDP. Since the labels of sequential data are known, in order to evaluate the effectiveness of RHDP in terms of percentage of correctly

clustering data, we use the F-score (precision and sensitivity) as the evaluation metric [39]. The higher the F-score, the more accurate the model is. The clustering results are presented in Table I, and it is evident that RHDP clustering has both higher precision and sensitivity compared with mean shift, and consequently, the F-score for RHDP clustering is significantly higher than mean shift (98% versus 85%). This is because RHDP utilizes all the characteristics of data distribution to compute average log-likelihood values [see (16)], while mean shift uses only the average values of data.

Note on Computational Time: Due to continuous updates to the distribution estimates on sequential data, the computational time of RHDP clustering is about 1.3 ms per data point (200-point window with a ten-point overlap is used in this simulation) and that of mean shift is about 0.03 ms per data point (with Intel Core i7-4770 CPU at 3.40 GHz). Still, the computational time of RHDP clustering is fast enough (with a sampling frequency of $\sim$700 Hz) for our application in CMP.

IV. APPLICATION OF DPGM MODELS FOR ONLINE MONITORING OF CMP PROCESS

The aim of this section is to verify the effectiveness of the proposed DPGM models in a practical advanced manufacturing scenario, namely, for monitoring a semiconductor CMP process [1], [40]. DPGM models will be used toward attaining two specific goals in the context of CMP:

1) Detection of process anomalies using the DP-based SPC as explained in Section III-A.
2) Identification of anomalies using RHDP clustering as described in Section III-B.

A. Experimental Setup

CMP is a vital back-end-of-line process in semiconductor manufacturing. Semiconductor wafer defects resulting from CMP process drifts can lead to high yield losses [3]. It is therefore desirable to ensure defect-free operation in CMP by employing real-time *in situ* sensor-based process monitoring approaches [1]. Various sensors, such as acoustic emission, force, and vibration sensors, have been applied to CMP process monitoring [1], [41]–[46]. Miniature wireless MEMS devices are particularly attractive for *in situ* monitoring applications due to their weight, and energy efficiency. MEMS vibration sensors have been successfully used hitherto for model-based monitoring, material removal rate estimation, and endpoint detection in CMP process [45]–[47].

In this work, we use a Buehler Automet 250 benchtop CMP apparatus for our experiments. Further details of the setup and experimental outcomes are available in [1]. A triaxis MEMS vibration sensor (ADXL 335) manufactured by Analog Devices Inc. is mounted on the apparatus to collect sensor data. The sensor signals are sampled at 670 Hz and transmitted wirelessly to a desktop computer with a matching wireless receiver unit. The CMP setup and wireless sensor network are shown in Fig. 11(a) and (b). Blanket copper wafer disks of Φ1.625 in (40.625 mm) are polished in a KOH-based alkaline colloidal silica slurry medium, which has a constant flow rate of 20 mL/min. Near-optical (arithmetic average
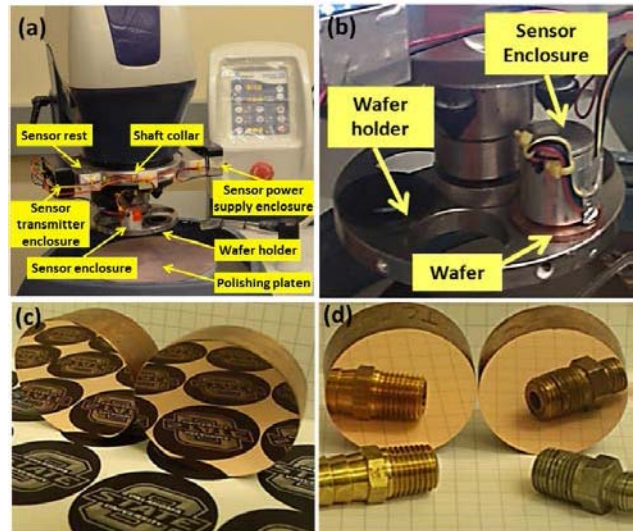

Fig. 11. (a) and (b) Buehler Automet 250 experimental CMP setup with the integrated wireless sensor. (c) and (d) Near-specular CMP finished copper wafers [1].

roughness, Sa$\sim$5 nm) quality surface finish blanket copper wafers are obtained by polishing with *a priori* identified optimal processing conditions [Fig. 11(c) and (d)].

However, as described previously in Fig. 1, sensor signals acquired from CMP process are complex; they may violate normality and linearity conditions. Consequently, traditional SPC and mean shift clustering approaches may not lend toward detection of CMP process anomalies, as demonstrated in the numerical case studies presented in Section III.

B. CMP Experimental Tests and Case Studies

In our experimental tests, certain CMP process parameters are deliberately changed to induce precisely controlled defects on the semiconductor wafer (e.g., scratches on the wafer). The following practical case studies are illustrated in this section.

1) Case E1: Changes in polishing load or downforce.
2) Case E2: Wear of the polishing pad.
3) Case E3: Sequential changes in processing conditions.

The first two of the above cases are instances where DP-based SPC will be applied for detecting process anomalies; the last case, Case E3, involves identification of specific anomalies using RHDP clustering.

1) Case E1—Capturing Changes in CMP Polishing Load (Downforce) With Data Slightly Violating Normal Assumption: The polishing load is one of the most significant factors in CMP and determines not only physical aspects, such as the nature of tribological contact, but also key process output variables, namely, material removal rate, within wafer nonuniformity, surface quality, etc. [3].

In this experiment, a change in polishing load (downforce) is monitored based on acquired vibration sensor data. As depicted in Fig. 12(a), after a low load (5 lb) is active for the first half time, the load is suddenly increased to a high-load (8 lb) condition. All other factors, namely, head speed and base speed, are maintained constant at 60 and 150 r/min, respectively. We acquire 4000 data points in total, amounting
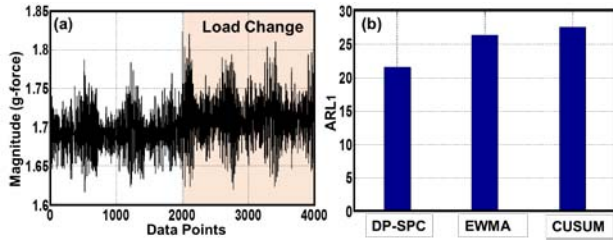
Fig. 12.   (a) Representative vibration signal patterns obtained under changing load conditions. (b) Comparison of ARL_1 in changing load conditions.

TABLE II
COMPARISON OF ARL_1 VALUES FOR TWO PREDEFINED PROCESS ANOMALIES WITH TRADITIONAL SPC AND DP-BASED SPC. THE UNITS ARE IN MILLISECONDS

	CASE	DP-SPC	EWMA	CUSUM
CASE E1:	LOAD CHANGE	21	26	28
CASE E2:	PAD WEAR	47	140	56
	PAD DEGRADATION	65	102	69

to about 6 s of polishing, during which the change of load occurs approximately midway. A visible prominent shift in signal mean, as well as variation, is evident; the signal mean and variation increase with an increase in downforce.

CUSUM, EWMA, and DP-based SPC are applied to the same time series data, allowing us to compare their ARL_1 results. The control limits are adjusted *a priori* to maintain identical type I error probabilities (α-error) at 5%, where this translates to an ARL_0 of 200. The results from a ten-replication study are presented in Fig. 12(b). Moreover, it is observed that the CMP vibration data depart from Gaussian distribution as indicated by the Anderson–Darling GoF test. Depending on the severity (p-value) of the non-Gaussian nature of the data distribution, the DP-based SPC charts are faster (low ARL_1) compared with the CUSUM and EWMA charts. For instance, referring to Table II, the DP-based SPC detects the change in polishing load within ~21 ms (14 data points) on average, whereas CUSUM and EWMA require ~27 ms (18 data points).

2) Case E2—Capturing Wear of CMP Polishing Pad With Data Severely Violating Normal Assumption: Degradation of the polishing pad is caused by wear overtime, selection of suboptimal process conditions, or improper postprocess handling [1]. For instance, inadequate postprocess cleaning allows the residual slurry to dry and coagulate on the pad. In addition, some portions of the polishing pad may be sheared away during polishing, thus exposing the underlying hard layer. Such polishing pads are *glazed*, i.e., the fibers of the polishing pad become entangled and lose the ability to retain slurry abrasives [1]. Polishing with a glazed pad leads to deep scratches and nonuniform wafer morphology [1].

The DP-based SPC aims to detect a degraded pad condition. It is constructed by training the IC mixture distribution with operational data using good pads, and is then applied to monitor CMP runs. This is akin to building a Phase I control chart based on an *a priori* IC process state [7]. The degraded pad is treated as the shifted process state.
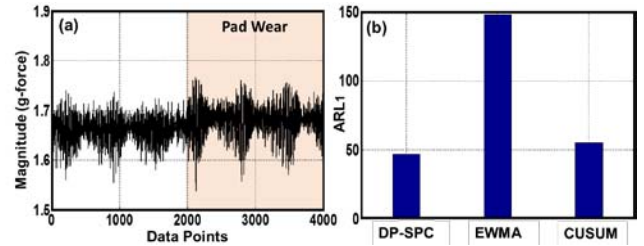

Fig. 13.   (a) Representative vibration signal patterns obtained for pad wear experiments. (b) Comparison of ARL_1 for pad wear.
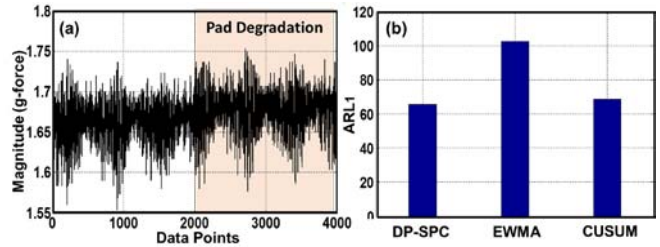

Fig. 14.   (a) Representative vibration signal patterns obtained for pad degradation experiments. (b) Comparison of ARL_1 for pad degradation.

In the case study, in order to verify the efficiency of DP-based SPC in detecting a degraded pad, we combine data from two experiments. The first half of the data (2000 data points, ~3 s) is obtained from an experiment where a new pad is used, while the second half is gathered from an experiment conducted with a glazed pad. DP-based SPC is compared with CUSUM and EWMA in terms of detection of the pad condition change.

Fig. 13(a) depicts the vibration time series data gathered under the following CMP conditions: 8 lb contact load, 150 r/min base speed and 60 r/min head speed. We discern from Fig. 13(a) that not only does the mean of the vibration signal change but also the variance of the signal slightly increases.

Moreover, compared with the previous in this instance, the vibration signals are found to depart more severely from Gaussian behavior. Therefore, DP-based SPC significantly outperforms the other two methods; it detects the pad wear earlier than CUSUM and two times quicker than EWMA control charts. Referring to Table II, it is observed that the DP-based SPC detects the change in pad wear within ~47 ms (31 data points) on average, whereas CUSUM requires ~56 ms (37 data points) and EWMA over 140 ms (99 data points).

In addition, another polishing experiment is conducted to show the effectiveness of DP-based SPC in detecting signal change with a mildly used pad (neither brand new nor glazed). From Fig. 14(a), it is observed that there is a slight mean shift of vibration signals after switching to the mildly used pad. Comparing with the results in Fig. 13(b), the time of detecting pad degradation by DP-based SPC increases to ~65 ms (43 data points) on average and CUSUM increases to ~71 ms (48 data points) on average. Yet, DP-based SPC still outperforms EWMA and CUSUM.

The ARL_1 results from the foregoing cases are summarized in Table II. The following inferences can be obtained.

- If the IC data slightly deviate from Gaussian distributed as in Case E1, then DP-based SPC detects the
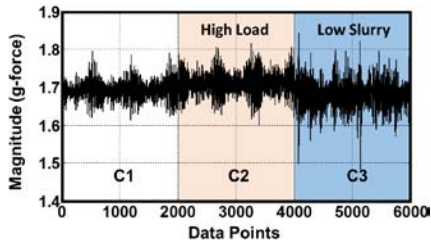
Fig. 15. Vibration data time series for the multiple process states (Case E3), including normal condition (C1), high load (C2), and low slurry (C3).

process anomalies nearly as quickly as EWMA and CUSUM.

- If the normality or symmetry conditions for the data distributions are violated severely as in Case E2, then DP-based SPC significantly outperforms CUSUM and EWMA.

These results agree closely with the implications from the numerical studies discussed in Section III-A. It is observed that the relative performance of DP-based SPC against traditional SPC drops with experimental data: while numerical case studies are generated from highly non-Gaussian and/or nonlinear systems (χ_1^2, bimodal data, or Rössler attractor), the experimental data manifest a modicum of similarity to Gaussian data [1], [21].

3) Case E3—Identifying Multiple Sequentially Occurring CMP Process Anomalies Using RHDP Clustering: After having demonstrated the utility of DP-based SPC for detecting process anomalies in the last two cases, we now apply the RHDP unsupervised evolutionary clustering approach explained in Section III-B for identifying multiple sequentially occurring faults in CMP. This is important from an application standpoint for CMP process, since if the process is OOC, it is valuable to know what type of anomaly is prevalent so that the appropriate corrective action can be taken.

In this study, three different kinds of CMP process operating conditions are sequentially activated during a single experimental run are as follows.

- The normal condition (C_1) occurs under nominally optimal process conditions, viz., 5 lb polishing load, 150 r/min base speed, and 60 r/min head speed.
- Condition C_2 occurs after 3 s of operation ($\sim$2000 data points), the polishing load is increased to 8 lb (the other settings are maintained at constant) for 3 s.
- Condition C_3 is when the slurry feed is low for 3 s (2000 data points), while the polishing load is kept at 8 lb.

Vibration signal patterns acquired for this experiment are presented in Fig. 15.

The comparison between the RHDP clustering and mean shift methods is also based on F-score (precision and sensitivity) borrowed from classification. The results for Case E3, presented in Table III, indicate that despite the continual change in CMP operating conditions, RHDP clustering identifies the different process states with higher precision and sensitivity compared with the conventional mean shift clustering method. For our three process states, RHDP clustering achieves an average F-score of 0.7923, which is about 10% higher than the mean shift method. Moreover, the sampling

TABLE III
CLUSTERING RESULTS FOR MULTIPLE PROCESS STATES IN CMP EXPERIMENT—COMPARISON OF RHDP CLUSTERING VERSUS MEAN SHIFT (THE VALUES IN THE PARENTHESIS ARE THE STANDARD DEVIATION)

	RHDP CLUSTERING			MEAN SHIFT		
	C_1	C_2	C_3	C_1	C_2	C_3
PRECISION	0.8834 (0.0746)	0.7699 (0.1942)	0.8181 (0.1696)	0.9765 (0.0132)	0.7121 (0.1639)	0.7075 (0.1888)
SENSITIVITY	0.7750 (0.1554)	0.7822 (0.1911)	0.7367 (0.2557)	0.0740 (0.0548)	0.5644 (0.1679)	0.5422 (0.2253)
F-SCORE	0.8256	0.7760	0.7752	0.8419	0.6297	0.6139
AVERAGE F-SCORE	0.7923			0.6952		

frequency of RHDP clustering is $\sim$700 Hz, faster than the sampling frequency of vibration sensors ($\sim$670 Hz) in CMP experiments.

V. CONCLUSION

In this paper, we devised DP-based SPC and RHDP unsupervised clustering for sensor-based process monitoring by the concept of DPGM modeling. We validated these approaches using numerical simulations and real-world wireless vibration signals acquired from an experimental CMP setup. Based on these studies, it is evident that DP-based SPC and RHDP clustering outperform traditional methods under conditions where the sensor signal patterns are nonlinear and non-Gaussian. Practical outcomes from this research are as follows.

1) DP-based SPC detects the onset of CMP process anomalies, such as changes in pad wear, within 50 ms of their inception. In contrast, the traditional methods, such as EWMA control chart, has a delay of over 140 ms.
2) RHDP clustering model classifies with about 80% fidelity (F-score) multiple sequential process drifts; traditional mean shift clustering accounts for an F-score under 70%.

Consequently, this paper addresses one of the significant challenges for process monitoring in ultraprecision manufacturing applications. As part of our future research, we aim to improve DPGM modeling in the following manner:

- increasing the accuracy of distribution approximation using extracted features instead of raw data in DP model;
- improving the computational tractability of RHDP clustering model for high-dimensional data by incorporating dimension reduction techniques.

APPENDIX I
(IMPLEMENTATION OF DP-BASED SPC)

Step 1: Assign conjugate prior to parameter θ_i of Gaussian component (mean μ_i and variance σ_i^2) by G_0 in DP

$$\sigma_i^2 \sim \text{InverseWishart}(S_0, n_0)$$
$$\mu_i | \sigma_i^2 \sim \text{Normal}(\mu_0, k_0 \sigma_i^2)$$
$$G_0 = p(\theta_i) = p(\mu_i, \sigma_i^2)$$

where S_0, n_0, μ_0, and k_0 are hyperparameters.

Step 2: Apply Gibbs sampling to acquire parameters of each Gaussian component from posterior conditional distributions in (5)

$$(\theta_i = |\boldsymbol{\theta}_{-i}, x_i) \sim \frac{n_{-i,j}\mathbb{N}(x_i|\theta_j)}{N-1+\alpha} + \frac{\alpha q H(\theta|x_i)}{N-1+\alpha}.$$

Step 3: Choose average log-likelihood values as monitoring statistic and calculate the average log-likelihood values y_t for IC data under the trained mixture distribution in (6)

$$y_t = \frac{1}{w} \sum_{i=1}^{w} \log \left[\sum_{k=1}^{j} \pi_k \mathbb{N}(x_i|\theta_k) \right].$$

Step 4: Establish the control limit of monitoring statistic under the scheme of CUSUM and perform anomaly detection.

APPENDIX II
(IMPLEMENTATION OF RHDP CLUSTERING)

Step 1: Initialize data distribution by using DP mixture model on the initial window of data.

Step 2: Calculate the average log-likelihood value of data in next window y^t under current mixture distribution in (16)

$$y^t = \frac{1}{w} \sum_{i=1}^{w} \log \left[\sum_{k=1}^{j} \pi_k \mathbb{N}\big(x_i^t | \theta_k^{t-1}\big) \right].$$

Step 3: Update the mixture distribution on next window of data by using RDP in (15)

$$\big(\theta_i^t = j \big| \boldsymbol{\theta}^{t-1}, \theta_1^t, \ldots, \theta_{i-1}^t, x_i^t\big)$$
$$\sim \big(n_{\cdot j}^{t-1} + n_{-i,j}^t\big)\mathbb{N}(x_i^t|\theta_j^t) + \alpha q R\big(\theta|x_i^t\big).$$

Step 4: Investigate whether consecutive mixture distributions are from the same draws in HDP by detecting a drop in average log-likelihood values.

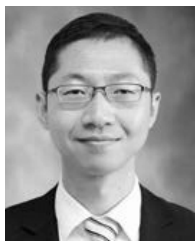Step 5: Cluster data sharing similar mixture distributions and achieve different states identification.

ACKNOWLEDGMENT

REFERENCES

[1] P. K. Rao *et al.*, "Process-machine interaction (PMI) modeling and monitoring of chemical mechanical planarization (CMP) process using wireless vibration sensors," *IEEE Trans. Semicond. Manuf.*, vol. 27, no. 1, pp. 1–15, Feb. 2014.

[2] J. D. Morillo, T. Houghton, J. M. Baue, R. Smith, and R. Shay, "Edge and bevel automated defect inspection for 300mm production wafers in manufacturing," in *Proc. IEEE/SEMI Adv. Semicond. Manuf. Conf. Workshop*, Munich, Germany, Apr. 2005, pp. 49–52.

[3] J. M. Steigerwald, S. P. Murarka, and R. J. Gutmann, *Chemical Mechanical Planarization of Microelectronic Materials*. Weinheim, Germany: Wiley, 2008.

[4] S. Bukkapatnam, S. Kamarthi, Q. Huang, A. Zeid, and R. Komanduri, "Nanomanufacturing systems: Opportunities for industrial engineers," *IIE Trans.*, vol. 44, no. 7, pp. 492–495, Apr. 2012.

[5] H. Guo, K. Paynabar, and J. Jin, "Multiscale monitoring of autocorrelated processes using wavelets analysis," *IIE Trans.*, vol. 44, no. 2, pp. 312–326, Apr. 2012.

[6] P. Qiu, *Introduction to Statistical Process Control*. Boca Raton, FL, USA: CRC Press, 2013.

[7] D. C. Montgomery, *Introduction to Statistical Quality Control*, 6th ed. New York, NY, USA: Wiley, 2008.

[8] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.

[9] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, pp. 264–323, Sep. 1999.

[10] D. Comaniciu and P. Meer, "Mean shift analysis and applications," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Kerkyra, Greece, 1999, pp. 1197–1203.

[11] C. E. Rasmussen, "The infinite Gaussian mixture model," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 12, 2000, pp. 554–560.

[12] S. Chakraborti, P. Van der Laan, and S. T. Bakir, "Nonparametric control charts: An overview and some results," *J. Quality Technol.*, vol. 33, no. 3, pp. 304–315, 2001.

[13] P. Qiu and Z. Li, "On nonparametric statistical process control of univariate processes," *Technometrics*, vol. 53, pp. 1–40, Jan. 2011.

[14] P. Qiu and Z. Li, "Distribution-free monitoring of univariate processes," *Statist. Probab. Lett.*, vol. 81, pp. 1833–1840, Dec. 2011.

[15] R. Ganesan, T. K. Das, and V. Venkataraman, "Wavelet-based multiscale statistical process monitoring: A literature review," *IIE Trans.*, vol. 36, pp. 787–806, Sep. 2004.

[16] J.-C. Lu, S.-L. Jeng, and K. Wang, "A review of statistical methods for quality improvement and control in nanotechnology," *J. Quality Technol.*, vol. 41, pp. 148–164, Apr. 2009.

[17] R. Ganesan, T. K. Das, A. K. Sikder, and A. Kumar, "Wavelet-based identification of delamination defect in CMP (Cu-low k) using nonstationary acoustic emission signal," *IEEE Trans. Semicond. Manuf.*, vol. 16, no. 4, pp. 677–685, Nov. 2003.

[18] M. K. Jeong, J.-C. Lu, and N. Wang, "Wavelet-based SPC procedure for complicated functional data," *Int. J. Prod. Res.*, vol. 44, pp. 729–744, Feb. 2006.

[19] G. A. Pugh, "A comparison of neural networks to SPC charts," *Comput. Ind. Eng.*, vol. 21, nos. 1–4, pp. 253–255, 1991.

[20] F. Zorriassatine and J. D. T. Tannock, "A review of neural networks for statistical process control," *J. Intell. Manuf.*, vol. 9, no. 3, pp. 209–224, 1998.

[21] P. K. Rao, "Sensor-based monitoring and inspection of surface morphology in ultraprecision manufacturing processes," Ph.D. dissertation, Dept. Ind. Eng. Manage., Oklahoma State Univ., Stillwater, OK, USA, 2013.

[22] P. Rao, S. Bukkapatnam, O. Beyca, Z. J. Kong, and R. Komanduri, "Real-time identification of incipient surface morphology variations in ultraprecision machining process," *J. Manuf. Sci. Eng.*, vol. 136, no. 2, pp. 021008-1–021008-11, 2014.

[23] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[24] S. W. Choi, J. H. Park, and I.-B. Lee, "Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis," *Comput. Chem. Eng.*, vol. 28, no. 8, pp. 1377–1387, Jul. 2004.

[25] U. Thissen, H. Swierenga, A. P. de Weijer, R. Wehrens, W. J. Melssen, and L. M. C. Buydens, "Multivariate statistical process control using mixture modelling," *J. Chemometrics*, vol. 19, pp. 23–31, Jan. 2005.

[26] T. Chen, J. Morris, and E. Martin, "Probability density estimation via an infinite Gaussian mixture model: Application to statistical process monitoring," *J. Roy. Statist. Soc., Ser. C (Appl. Statist.)*, vol. 55, pp. 699–715, Nov. 2006.

[27] A. R. Ferreira da Silva, "A Dirichlet process mixture model for brain MRI tissue classification," *Med. Image Anal.*, vol. 11, pp. 169–182, Apr. 2007.

[28] J. Zhang, Z. Ghahramani, and Y. Yang, "A probabilistic model for online document clustering with application to novelty detection," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 17, 2005, pp. 1617–1624.

[29] P. K. Rao, J. Liu, D. Roberson, Z. Kong, and C. Williams, "Online real-time quality monitoring in additive manufacturing processes using heterogeneous sensors," *J. Manuf. Sci. Eng.*, vol. 137, p. 061007, Sep. 2015.

[30] A. Ahmed and E. P. Xing, "Dynamic non-parametric mixture models and the recurrent Chinese restaurant process," in *Proc. SIAM Int. Conf. Data Mining*, Atlanta, GA, USA, 2008, pp. 219–230.

[31] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *J. Amer. Statist. Assoc.*, vol. 101, no. 476, pp. 1566–1581, Dec. 2006.

[32] M. D. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *J. Amer. Statist. Assoc.*, vol. 90, pp. 577–588, Jun. 1995.

[33] D. Blackwell and J. B. MacQueen, "Ferguson distributions via Pólya urn schemes," *Ann. Statist.*, vol. 1, pp. 353–355, Mar. 1973.

[34] J. Crutchfield, D. Farmer, N. Packard, R. Shaw, G. Jones, and R. J. Donnelly, "Power spectral analysis of a dynamical system," *Phys. Lett. A*, vol. 76, pp. 1–4, Mar. 1980.

[35] S. Bukkapatnam, P. Rao, and R. Komanduri, "Experimental dynamics characterization and monitoring of MRR in oxide chemical mechanical planarization (CMP) process," *Int. J. Mach. Tools Manuf.*, vol. 48, pp. 1375–1386, Oct. 2008.

[36] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, vol. 7, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[37] C. A. Lowry and D. C. Montgomery, "A review of multivariate control charts," *IIE Trans.*, vol. 27, no. 6, pp. 800–810, 1995.

[38] A. Rodríguez, D. B. Dunson, and A. E. Gelfand, "The nested Dirichlet process," *J. Amer. Statist. Assoc.*, vol. 103, no. 483, pp. 1131–1154, Sep. 2008.

[39] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.

[40] P. B. Zantye, A. Kumar, and A. K. Sikder, "Chemical mechanical planarization for microelectronics applications," *Mater. Sci. Eng., R, Rep.*, vol. 45, pp. 89–220, Oct. 2004.

[41] Z. Kong, A. Oztekin, O. F. Beyca, U. Phatak, S. T. S. Bukkapatnam, and R. Komanduri, "Process performance prediction for chemical mechanical planarization (CMP) by integration of nonlinear Bayesian analysis and statistical modeling," *IEEE Trans. Semicond. Manuf.*, vol. 23, no. 2, pp. 316–327, May 2010.

[42] H. Jeong, H. Kim, S. Lee, and D. Dornfeld, "Multi-sensor monitoring system in chemical mechanical planarization (CMP) for correlations with process issues," *CIRP Ann.-Manuf. Technol.*, vol. 55, no. 1, pp. 325–328, 2006.

[43] A. K. Sikder, F. Giglio, J. Wood, A. Kumar, and M. Anthony, "Optimization of tribological properties of silicon dioxide during the chemical mechanical planarization process," *J. Electron. Mater.*, vol. 30, no. 12, pp. 1520–1526, 2001.

[44] J. Tang, D. Dornfeld, S. K. Pangrle, and A. Dangca, "In-process detection of microscratching during CMP using acoustic emission sensing technology," *J. Electron. Mater.*, vol. 27, pp. 1099–1103, Oct. 1998.

[45] Z. Kong, O. Beyca, S. T. Bukkapatnam, and R. Komanduri, "Nonlinear sequential Bayesian analysis-based decision making for end-point detection of chemical mechanical planarization (CMP) processes," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 4, pp. 523–532, Nov. 2011.

[46] U. Phatak, S. Bukkapatnam, Z. Kong, and R. Komanduri, "Sensor-based modeling of slurry chemistry effects on the material removal rate (MRR) in copper-CMP process," *Int. J. Mach. Tools Manuf.*, vol. 49, pp. 171–181, Feb. 2009.

[47] Z. Wang, S. T. S. Bukkapatnam, S. R. T. Kumara, Z. Kong, and Z. Katz, "Change detection in precision manufacturing processes under transient conditions," *CIRP Annals-Manu. Technol.*, vol. 63, no. 1, pp. 449–452, 2014.

Omer F. Beyca received the B.S. degree from Fatih University, Istanbul, Turkey, and the Ph.D. degree in industrial engineering and management from Oklahoma State University, Stillwater, OK, USA.

He is currently an Assistant Professor with the Industrial Engineering Department, Istanbul Technical University, Istanbul, Turkey. His current research interests include dynamic modeling of nonlinear systems, quality control in micromachining, and sensor-based modeling.

Prahalad K. Rao received the B.Eng. (Hons.) degree from Veermata Jijabai Technological Institute (VJTI), Mumbai, India, in 2003, and the M.S. and Ph.D. degrees in industrial engineering from Oklahoma State University, Stillwater, OK, USA, in 2006 and 2013, respectively.

He was an Assistant Professor with the System Science and Industrial Engineering Department, State University of New York at Binghamton, Binghamton, NY, USA, from 2014 to 2016. He is currently an Assistant Professor with the Mechanical and Materials Engineering Department, University of Nebraska–Lincoln, Lincoln, NE, USA. His current research interests include sensor-based monitoring of manufacturing processes (ultraprecision diamond turning, chemical mechanical planarization, and additive manufacturing processes).

Dr. Rao is a member of the Institute of Industrial Engineers, American Society for Quality, and the Institute for Operations Research and the Management Sciences.

Zhenyu (James) Kong (M'12) received the B.S. and M.S. degrees in mechanical engineering from the Harbin Institute of Technology, Harbin, China, in 1993 and 1995, respectively, and the Ph.D. degree from the Department of Industrial and System Engineering, University of Wisconsin–Madison, Madison, WI, USA, in 2004.

He was a Faculty Member with the School of Industrial Engineering and Management, Oklahoma State University, Stillwater, OK, USA. He is currently an Associate Professor with the Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, USA. His current research interests include sensing and analytics for advanced manufacturing and automatic quality control for large and complex manufacturing systems.

Dr. Kong is a member of the Institute of Industrial Engineers and the American Society of Mechanical Engineers. He received the Halliburton Outstanding Faculty Award from the College of Engineering, Architecture and Technology at the Oklahoma State University in 2013.

Satish T. S. Bukkapatnam received the Ph.D. degree in industrial and manufacturing engineering from the Pennsylvania State University, State College, PA, USA.

He is currently a Rockwell International Professor with the Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, USA. He is also the Director of Texas A&M Engineering Experiment Station, Institute for Manufacturing Systems, and has joint appointments with the Biomedical and Mechanical Engineering Departments. His research has led to 133 peer-reviewed publications (74 published/accepted in journals and 59 in conference proceedings), five pending patents, U.S. $5 million in grants as Principal Investigator/Co-Principal Investigator from the National Science Foundation, the U.S. Department of Defense, and the private sector, and ten best-paper/poster recognitions. His current research interests include the harnessing of high-resolution nonlinear dynamic information, especially from wireless MEMS sensors, to improve the monitoring and prognostics, mainly of ultraprecision and nanomanufacturing processes and machines, and cardiorespiratory processes.

Dr. Bukkapatnam is a Fellow of the Institute for Industrial and Systems Engineers (IISE), and has been recognized by Oklahoma State University Regents Distinguished Research, Halliburton Outstanding College of Engineering Faculty, IISE Eldin Outstanding Young Industrial Engineer, and the Society of Manufacturing Engineers Dougherty outstanding young manufacturing engineer awards.

Jia (Peter) Liu received the B.S. and M.S. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2005 and 2007, respectively. He is currently pursuing the Ph.D. degree in industrial and systems engineering with Virginia Tech, Blacksburg, VA, USA.

His current research interests include the development of nonparametric Bayesian statistical models and spatial statistical models for online monitoring and heterogeneous sensor data analysis in manufacturing processes.

Mr. Liu is a member of the Institute of Industrial Engineers, and the Institute for Operations Research and the Management Sciences.