IIE Transactions (2005) **37**, 877–889

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/07408170590969889

Optimal coordinate sensor placements for estimating mean and variance components of variation sources

CHIN (QINYAN) LIU[1], YU DING[2,*] and YONG CHEN[3]

[1] *Department of Management Science & Engineering, Stanford University, Stanford, CA 94305, USA*
E-mail: chinliu@stanford.edu
[2] *Department of Industrial Engineering, Texas A&M University, 3131 TAMU, College Station, TX, 77843-3131, USA*
E-mail: yuding@iemail.tamu.edu
[3] *Department of Mechanical and Industrial Engineering, The University of Iowa, Iowa City, IA 52242, USA*
E-mail: yongchen@engineering.uiowa.edu

Received March 2003 and accepted March 2005

In-process optical coordinate measuring machines offer the potential to diagnose the sources of the variations that are responsible for product quality defects. Such a sensor system can thus help manufacturers to improve product quality and reduce process downtime. The effective use of sensor data in the diagnosis of the sources of variations depends on the optimal design of the sensor system, which is often also called the problem of sensor placement. This paper addresses coordinate sensor placement for the diagnosis of dimensional variation sources in assembly processes. Sensitivity indices for the detection of the process mean and variance components are defined as the design criteria and are derived in terms of process layout and sensor deployment information. Exchange algorithms, originally developed for optimal experimental design, are revised and then used to maximize the detection sensitivity. A sort-and-cut procedure is proposed, which is able to significantly improve the algorithm efficiency of the current exchange routine. The resulting optimal sensor layout and its implications are illustrated in the specific context of a panel assembly process.

1. Introduction

Dimensional integrity is a major quality concern in many discrete-part manufacturing processes. In the automotive and aviation industries, for instance, dimensional problems contribute to about two-thirds of the total quality problems during a new product launch (Shalon *et al.*, 1992; Ceglarek and Shi, 1995).

Coordinate Measuring Machines (CMMs) are widely used in discrete-part industries to ensure the dimensional quality of a product. The mechanism of a CMM is illustrated in Fig. 1(a). A CMM usually consists of a spatial frame that provides the coordinate reference (not shown in the figure), a mechanical arm that can move along guided tracks, and a probe that retrieves the coordinate information when its tip touches the surface of a manufactured workpiece. One disadvantage of CMMs is their low throughput. Performing the measurement jobs sequentially, a CMM with a single mechanical arm and touch probe will usually take hours to finish all measurements on a complicated product. For instance, a CMM can measure only six to eight automotive bodies per day in an automotive body shop which can fabricate 1000 units daily. In addition the high manufacturing costs of a CMM often limits the use of multiple CMMs measuring jobs in parallel.

Recent advances in sensor technology have led to the development of the optical CMM (OCMM). An OCMM replaces the mechanical arm and the touch probe in a CMM with an optical sensor unit (Fig. 1(b)), which consists of a laser source and two CCD (Charged-Coupled Device) image sensors. The laser source sends a light beam to the surface of a workpiece and the CCD sensors detect the reflected laser beam. The sensor unit is also installed within a spatial frame and is able to calculate the coordinates of the measured point relative to the frame reference using triangulation.

The OCMM frame and sensor unit are much less expensive than the CMM frame and the CMM's touch probe. It is therefore much cheaper than CMMs and it is possible to deploy multiple optical sensor units and/or build more OCMM stations, to perform the parallel measurement of multiple product characteristics. An OCMM station with multiple sensor units is capable of measuring as many as 150 product features on a car body within 1 minute (refer to Fig. 1(c)). This high throughput capability enables OCMMs to be built into the production process and obtain 100% inspection of dimensional quality characteristics (Apley and Shi, 2001).

*Corresponding author

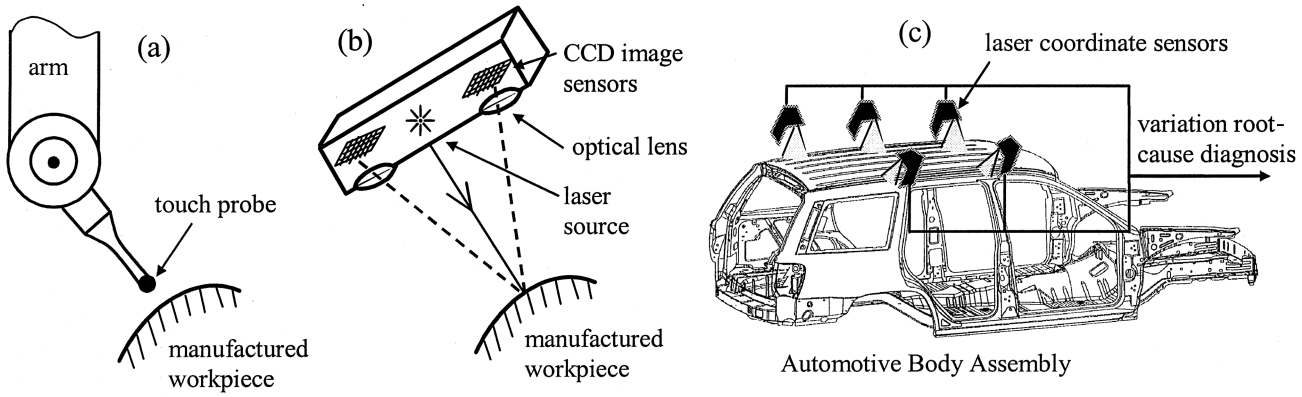

Fig. 1. Mechanism of a CMM and an OCMM.

The use of OCMMs results in a change in the quality control philosophy. With CMMs, dimensional measurements are taken offline and sampled from a large product population. In this manner the CMMs are used to inspect the key dimensional product features and ensure that they are statistically acceptable. However, the implementation of in-process OCMMs allows the continual dimensional monitoring of every manufactured product and creates the possibility of determining the underlying process variation sources that are responsible for product defects; this process is known as "root-cause diagnosis". Root-cause diagnosis is critical because the identification of variation sources will lead to corrective actions, restoring the manufacturing system to its normal condition in a timely manner.

Recent research efforts have advanced the state-of-the-art of root-cause diagnosis in complicated manufacturing systems: one focused application area is the automotive assembly process. Statistical methods employed in that area include: (i) in estimation methods (Apley and Shi, 1998; Lawless *et al.*, 1999a, 1999b; Carlson *et al.*, 2000); (ii) pattern recognition using principal components (Hu and Wu, 1992; Ceglarek and Shi, 1996; Ding *et al.*, 2002a); and (iii) factor analysis (Apley and Shi, 2001; Apley and Lee, 2003).

The effective use of product measurements in root-cause diagnosis depends to a great extent on the design of the sensor system. A poorly designed sensor system may not be able to provide the desired diagnosability or sensitivity in identifying variation sources. The design of the sensor system is realized through the design or determination of: (i) an individual sensor unit; (ii) the number of sensors needed; (iii) sensor locations; and (iv) operational strategies such as how many and how often measurements will be taken. In this paper, we consider the use of a commercially available optical sensor unit as described in Fig. 1(b). The high measurement throughput of the chosen optical sensor units makes their operations simple because in-process data are automatically collected from every product. The sensor system design considered in this paper focuses on the determination of the number and locations of the coordinate sensors; this is usually called the problem of "*sensor placement*" in engineering practice.

To some extent, coordinate sensor placement is equivalent to the selection (of the number and locations) of dimensional measurement features on a product. Relevant research on the selection of dimensional measurement features can be classified into two categories: (i) inspection-oriented; and (ii) diagnosis-oriented. Inspection-oriented feature selection is mainly based on the study of Key Product Characteristics (KPCs) because the purpose of inspection is to ensure that the KPCs meet their designed tolerances. KPCs may be decided through an empirical analysis of the product/process flow (Ceglarek *et al.*, 1994; Soderberg and Carlson, 1999) or through a more quantitative sensitivity-based design evaluation (Whitney *et al.*, 1994; Thornton, 1999; Ding *et al.*, 2002b).

Product inspection involves performing a statistical inference from the measured product features. Root-cause diagnosis, on the other hand, involves making inferences from process variation sources that are correlated with the product measurements. For this reason, diagnosis-oriented measurement feature selection aims at choosing those features that can lead to a desired optimal condition (e.g., maximum separation) to enable the identification of variation sources. This means that a criterion that characterizes the distinction between variation sources needs to be defined and then an optimization routine needs to be employed to optimize the chosen criterion. Prior research on this objective includes the papers by Khan *et al.* (1998), Khan *et al.* (1999), Wang and Nagarkar (1999), and Khan and Ceglarek (2000). The methods developed by Khan and his colleagues involve the maximization of the minimum distance between variation patterns that are computed as the eigenvectors of a measurement covariance matrix. However, due to the fact that their diagnostic procedure assumes the occurrence of a single variation source at a time, their sensor placement strategy only ensures that this individual variation source is optimally distinguished from the

others. Another limitation is that their strategy is based on a specific way to define and construct the variation patterns. Their results may no longer be optimal if a different type of variation pattern is defined and used. Wang and Nagarkar (1999) developed a sensor placement strategy for a more generic situation in which multiple simultaneous variation sources may exist. A D-optimal criterion is used, the same as that in the optimal experimental design (Fedorov, 1972; Atkinson and Donev, 1992), and Powell's direct search (Powell, 1992) is employed to find the optimal sensor placement.

In contrast with Wang and Nagarkar (1999), this paper will investigate the use of an E-optimality instead of the D-optimality as the design criterion for sensor placements. The E-optimality characterizes the minimal sensitivity of a coordinate sensor system in detecting the mean and variance components of variation sources. Optimization of such a sensitivity criterion is equivalent to ensuring a maximum separation of variation sources. This paper will develop an inequality relationship so that a unified criterion can be found for both mean-detection and variance-detection sensitivities. Exchange algorithms, initially developed for optimal experimental designs, were employed in optimizing the chosen sensitivity criterion. Following an idea proposed by Lam *et al.* (2002) in a molecule selection application, we devise a sort-and-cut procedure to address specific problems relevant to sensor system design, which considerably improves the algorithm efficiency.

This paper is structured as follows. In Section 2, we use an automotive assembly process as a case in point to explain the procedure for root-cause diagnosis and present a linear diagnostic model that links the product dimensional measurements to the process variation sources. Design criteria, i.e., the mean- and variance-detection sensitivity, and their relationship, are discussed in Section 3. The optimization algorithm and the suggested revision are presented in Section 4. We discuss the resulting optimal sensor layouts as well as a few practical issues in Section 5. Finally, we conclude the paper in Section 6.

2. Formulation for root-cause diagnosis

In this section, we consider the dimensional measurements taken in the auto-body assembly processes with the presence of fixture-related process variation sources.

Figure 2(a) shows the side panel assembly of the Sport Utility Vehicle (SUV) in Fig. 1(c). This side panel assembly is made of four components: an A-pillar panel, a B-pillar panel, a rail roof side panel, and a rear quarter panel. We simplify each panel component in the assembly as a two-dimensional (2-D) polygonal workpiece (Fig. 2(a)). In a 2-D panel assembly process, each workpiece is usually held by a set of fixtures during the assembly operation, illustrated in Fig. 2(b). A set of fixtures constitutes of a four-way locator (P_1) that controls motion in both the x and z directions ($\delta P_1(x, z)$), and a two-way locator (P_2) that controls motion only in the z-direction ($\delta P_2(z)$). The fixture location for the whole assembly is indicated in Fig. 2(a) as $P_1 - P_8$. Optic coordinate sensors are used to monitor the dimensional integrity of the eventual assembly.

The workpiece should have a small positional perturbation if all locator-contacts function properly within their designed tolerances. If there is damage to the pinholes or wear on the locators, the workpiece will then undergo a larger random deviation, as compared to the designed tolerance, from its nominal position and thus cause a mean shift or excessive dimensional variability on the final assembly. In Fig. 2(b), for instance, sensors M_1 to M_3 could detect a mean shift or excessive panel variability caused by a large z-direction deviation $\delta P_2(z)$. In this case, the process variation source is the low locator positioning repeatability. After detecting a large mean shift or variability in the product, we would like to identify the malfunctioning fixture locator responsible for the dimensional quality defect.

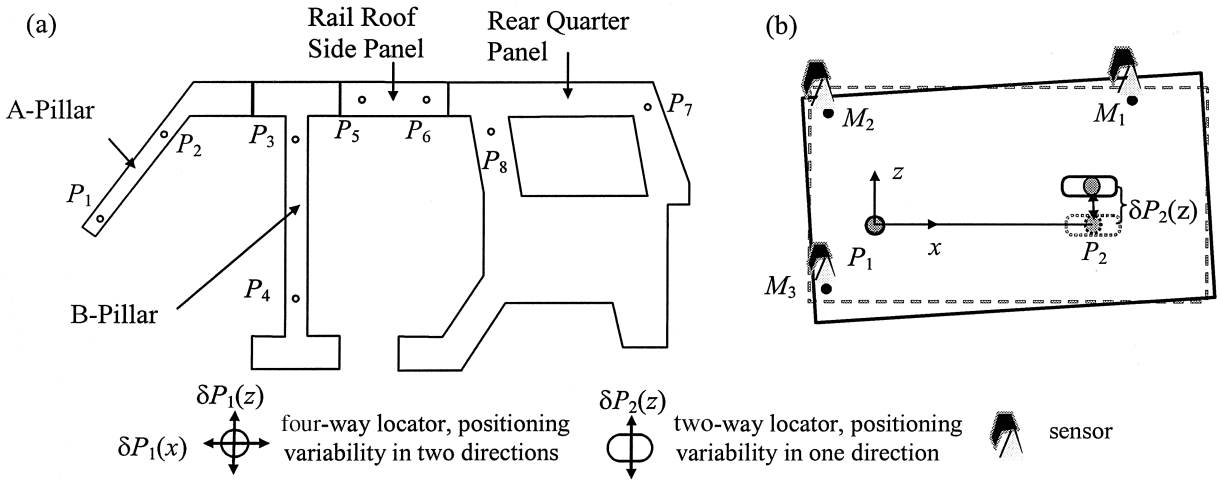

Fig. 2. Illustrative example: variation sources in panel assembly processes: (a) the assembly; and (b) the variation detection.

In order to perform root-cause diagnosis, the first step is to establish a diagnostic model that links the product measurements to the process variance sources. We here use the simple example in Fig. 2(b) to get the flavor of how the model is developed.

Denote by $\mathbf{u}$ the deviations at fixturing points such as P_1 and P_2 and by $\mathbf{y}$ the deviations measured by coordinate sensors. In Fig. 2(b), $\mathbf{u} = [\delta P_1(x)\ \delta P_1(z)\ \delta P_2(z)]^T$, where δ is the perturbation operator. When a coordinate sensor measures the position of a point, it generally returns three coordinate values of that point in space. In the 2-D assembly process discussed above, only two of the three coordinates, i.e., those of x and z directions, are of interest to us, and thus, the product measurements are $\mathbf{y} = [\delta M_1(x)\ \delta M_1(z)\ \delta M_2(x)\ \delta M_2(z)\ \delta M_3(x)\ \delta M_3(z)]^T$. However, whenever one more sensor is installed, two more measurements will be added to the vector $\mathbf{y}$ in the 2-D assembly process.

For dimensional control, the relationship between $\mathbf{y}$ and $\mathbf{u}$ can be obtained by a standard kinematics analysis (Paul, 1981). In quality control, since the deviations in product features, even when they are beyond their tolerances, are still much smaller than the distances between locators, the angles of the part rotations are in fact very small. Hence, a linear model structure or a linearization of originally nonlinear systems is often acceptable for discrete-part manufacturing processes. The linear diagnostic model can be generally expressed as:

$$\mathbf{y} = \mathbf{A}\mathbf{u} + \mathbf{v}, \tag{1}$$

where $\mathbf{v}$ is the additive sensor noise. Matrix $\mathbf{A}$ constitutes multiple row blocks, i.e., $\mathbf{A}^T = [\mathbf{A}_1^T \cdots \mathbf{A}_s^T]$ with the ith block as:

$$\mathbf{A}_i = \begin{bmatrix} 1 & \dfrac{M_i(z) - P_1(z)}{P_2(x) - P_1(x)} & \dfrac{P_1(z) - M_i(z)}{P_2(x) - P_1(x)} \\ 0 & \dfrac{P_2(x) - M_i(x)}{P_2(x) - P_1(x)} & \dfrac{M_i(x) - P_1(x)}{P_2(x) - P_1(x)} \end{bmatrix}, i = 1, \ldots, s, \tag{2}$$

where s is the number of sensors, $P_1(x)$, $M_i(z)$ etc are the nominal coordinates of the locators and the sensors, respectively. The number of variation sources (i.e., the dimension of $\mathbf{u}$) is denoted by p and the number of measurements (i.e., the dimension of $\mathbf{y}$) is denoted by n. In this particular process $p = 3$ and $n = 2s$ that is the product of the number of sensors and the measurements retrieved per sensor. For the assembly in Fig. 2(a), where more parts with complicated shapes and different orientations are involved, the $\mathbf{A}$ matrix is more complex. However, it still follows the same process as the simple example to develop the more complex linear diagnostic model. For a detailed derivation of relevant kinematic models, please refer to Jin and Shi (1999), Mantripragada and Whitney (1999) and Ding *et al.* (2000) for assembly processes or Djurdjanovic and Ni (2001) and Zhou, Huang and Shi (2003) for machining processes.

In root-cause diagnosis, one will make inferences about $\mathbf{u}$ based on a sample of measurements of $\mathbf{y}$. The following assumptions are usually made for this kind of problem:

Assumption 1. The p variation sources are independent so that $\mathbf{u}$ has a diagonal variance-covariance matrix.

Assumption 2. Sensor noise $\mathbf{v}$ is independent of $\mathbf{u}$. It is of zero-mean and has the variance-covariance matrix $\sigma_v^2\mathbf{I}_n$, where $\mathbf{I}_n$ is an $n \times n$ identity matrix and σ_v^2 is the variance of the sensor noise.

A sensor system that satisfies assumption 2 is known as a *homogeneous* sensor system. A coordinate sensor system using sensor units from the same manufacturers and in the same stage of their service lives qualifies as a homogeneous sensor system. On the other hand, a sensor system constituted by different types of sensors or the same type of sensors but with different accuracies is a *heterogeneous* sensor system. In this paper, we limit our discussion to the design of a homogenous sensor system and include a discussion on how to deal with a heterogeneous sensor system in Section 3.3. Additionally, we also assume in this paper that all sensor units function properly and thus no sensor fault is considered.

Model (1) appears to be similar to a traditional linear regression model, e.g., $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. The differences are as follows. In regression, $\mathbf{X}$ is a data matrix containing measurements of predictors, oftentimes determined through a designed experiment. In model (1), $\mathbf{A}$ is not a data matrix of predictors. Instead it is determined by system design parameters such as locator and sensor locations. The design to determine $\mathbf{A}$ is not to design an experiment but instead an engineering system. The $\mathbf{A}$ matrix is called a *system matrix* in engineering system designs. Also, $\mathbf{u}$ is not the vector of parameters but a vector of unknown *random* inputs. By contrast, $\boldsymbol{\beta}$ in the above regression model is often a *constant* unknown vector equivalent to the fixed effect in statistical inference. In fact, our model (1) fits better to a mixed linear model with both fixed effects and random effects (please refer to McCulloch and Searle (2000) for mixed linear models). If we write model (1) as:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\mu} + \mathbf{A}\tilde{\mathbf{u}} + \mathbf{v}, \tag{3}$$

where $\boldsymbol{\mu}$ is the mean vector of $\mathbf{u}$ and $\tilde{\mathbf{u}}$ is its zero-mean random part, then $\boldsymbol{\mu}$ corresponds to the fixed effects and $\tilde{\mathbf{u}}$ corresponds to the random effects. For root-cause diagnosis, one needs to detect abnormal variations of the mean components $\boldsymbol{\mu} \equiv [\mu_1 \cdots \mu_p]^T$ and the variance components $\boldsymbol{\theta} \equiv [\sigma_1^2\sigma_2^2 \cdots \sigma_p^2]^T$. Please note that we will assume in this paper that knowledge about sensor noise variance, σ_v^2, is available from the sensor vendor's calibration and specification.

3. Design criteria for root-cause diagnosis

3.1. *Diagnosability and sensitivity indices*

In the theory of mixed linear models, Rao and Kleffe (1988) defined parameter *identifiability* as the ability to verify whether or not the differences between input parameters lead to differences of the corresponding distributions of the observations $\mathbf{y}$. Following this concept and also noting that the distribution of $\mathbf{y}$ is usually specified up to the first- and second-order moments in many engineering applications, Zhou, Ding, Chen and Shi (2003) defined the *diagnosability* for the mean and variance components as: a linear parametric function $\mathbf{p}^T\boldsymbol{\mu}$ is said to be mean-diagnosable if:

$$\mathbf{p}^T\boldsymbol{\mu}_1 \neq \boldsymbol{p}^T\boldsymbol{\mu}_2 \Rightarrow \mathbf{m}_\mathbf{y}|_{\mu=\mu_1} \neq \mathbf{m}_\mathbf{y}|_{\mu=\mu_2}, \forall \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \quad (4)$$

and a linear parametric function $\mathbf{f}^T\boldsymbol{\theta}$ is said to be variance-diagnosable if:

$$\mathbf{f}^T\boldsymbol{\theta}_1 \neq \mathbf{f}^T\boldsymbol{\theta}_2 \Rightarrow \mathrm{vec}(\boldsymbol{\Sigma}_\mathbf{y})|_{\theta=\theta_1} \neq \mathrm{vec}(\boldsymbol{\Sigma}_\mathbf{y})|_{\theta=\theta_2}, \forall \quad \boldsymbol{\theta}_1,\ \boldsymbol{\theta}_2, \quad (5)$$

where $\mathbf{m}_\mathbf{y}$ and $\boldsymbol{\Sigma}_\mathbf{y}$ are the mean vector and covariance matrix of $\mathbf{y}$, respectively; $\mathrm{vec}(\cdot)$ is an operator to stack the columns of a matrix on top of one another, e.g., $\mathrm{vec}(\boldsymbol{\Sigma}) = [\sigma_{11}\ \sigma_{21}\ \sigma_{12}\ \sigma_{22}]^T$ for a 2×2 $\boldsymbol{\Sigma}$.

The above definition characterizes whether or not the sensor system provides enough information to ensure that the mean and variance components of variation sources can be separated. We can view it as: if a variation source is diagnosable, no matter how small a change it undergoes, we can theoretically find an algorithm to estimate it provided that we have a large enough amount of samples. If a sensor system does not ensure the diagnosability, no matter how much a variation source changes, we cannot uniquely pinpoint which variation source undergoes the change.

Diagnosability itself can be used as a design criterion for sensor system design. In fact, a sensor placement strategy leading to a full diagnosability was studied by Ding *et al.* (2003). On the other hand, the diagnosability condition does not make any distinction among diagnosable systems even though some sensor systems may have a superior performance to others in the sense that it could easily detect a small change in the variation sources. This difference of detection capability is characterized by the concept of "sensitivity", which may be interpreted as follows: a sensor system that has a zero sensitivity to any one of the variation sources provides no diagnosability, whereas a sensor system with a nonzero sensitivity to all variation sources possesses a certain level of diagnosability. It is desirable that a sensor system not only has a full diagnosability but also that it is sensitive to the underlying changes of variation sources. This paper will go beyond diagnosability, aiming to achieve a maximum separation of variation sources via the maximization of sensitivity indices defined below.

In this paper, we consider a discrete-part manufacturing process in the same setting as considered in Zhou, Ding, Chen and Shi (2003). Then, based on model (3), we can have:

$$\mathbf{m}_\mathbf{y} = \mathbf{A}\boldsymbol{\mu}, \quad (6)$$

and

$$\mathrm{vec}(\boldsymbol{\Sigma}_\mathbf{y}) = \boldsymbol{\pi}(\mathbf{A})\boldsymbol{\theta} + \sigma_v^2\mathrm{vec}(\mathbf{I}_n), \quad (7)$$

where $\boldsymbol{\pi}(\cdot)$ is a matrix transform defined as:

$$\boldsymbol{\pi}(\mathbf{A}) = [(\mathbf{a}^1{*}\mathbf{a}^1)^T\ \cdots\ (\mathbf{a}^1{*}\mathbf{a}^n)^T\ \cdots\ (\mathbf{a}^n{*}\mathbf{a}^1)^T \cdots\ (\mathbf{a}^n{*}\mathbf{a}^n)^T]^T, \quad (8)$$

and $\mathbf{a}^j$ is the jth row vector of $\mathbf{A}$, $j = 1, \ldots, n$, and $*$ represents the Hadamard product (Schott, 1997).

Following the same spirit in defining diagnosability, the sensitivity for detecting changes in mean and variance components can be defined as the ratio of the change in the mean or variance of $\mathbf{y}$ over a perturbation of the mean or variance of the input sources. We define the detecting sensitivity of mean and variance components as follows.

Definition 1. Given measurement $\mathbf{y}$, the mean-detecting sensitivity, denoted as S_m, is defined as:

$$S_\mathrm{m} \equiv \min_{\delta\boldsymbol{\mu}\neq\mathbf{0}} \frac{(\delta\mathbf{m}_\mathbf{y})^T(\delta\mathbf{m}_\mathbf{y})}{(\delta\boldsymbol{\mu})^T(\delta\boldsymbol{\mu})}, \quad (9)$$

and the variance-detecting sensitivity, denoted as S_v, is defined as:

$$S_\mathrm{v} \equiv \min_{\delta\boldsymbol{\theta}\neq\mathbf{0}} \frac{\mathrm{tr}\big(\delta\tilde{\boldsymbol{\Sigma}}_\mathbf{y}^T\delta\tilde{\boldsymbol{\Sigma}}_\mathbf{y}\big)}{(\delta\boldsymbol{\theta})^T(\delta\boldsymbol{\theta})}, \quad (10)$$

where $\tilde{\boldsymbol{\Sigma}}_\mathbf{y}$ is the covariance matrix contributed from the process variation sources, i.e., $\mathrm{vec}(\tilde{\boldsymbol{\Sigma}}_\mathbf{y}) = \boldsymbol{\pi}(\mathbf{A})\boldsymbol{\theta}$.

Given the linear relation in Equations (6) and (7) and utilizing the eigenvalue property of a symmetric matrix (Schott, 1997, p. 105), we can express the above-defined sensitivity indices in terms of the eigenvalue of $\mathbf{A}^T\mathbf{A}$ as follows (the proof is fairly straightforward and is thus omitted):

$$S_\mathrm{m} = \lambda_{\min}(\mathbf{A}^T\mathbf{A}) \quad \text{and} \quad S_\mathrm{v} = \lambda_{\min}(\boldsymbol{\pi}(\mathbf{A})^T\boldsymbol{\pi}(\mathbf{A})), \quad (11)$$

where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of a matrix. In deriving S_v, the relation that $\mathrm{tr}(\delta\tilde{\boldsymbol{\Sigma}}_\mathbf{y}^T\delta\tilde{\boldsymbol{\Sigma}}_\mathbf{y}) = \mathrm{vec}(\delta\tilde{\boldsymbol{\Sigma}}_\mathbf{y})^T\mathrm{vec}(\delta\tilde{\boldsymbol{\Sigma}}_\mathbf{y})$ is used.

Remark 1. The squared summations of the elements in the input/output vectors are used in the above definition so that we can have a scalar sensitivity index that is easy to interpret. The squared summations are equivalent to the Euclidean norm of the corresponding vector/matrix; $\mathrm{tr}(\delta\tilde{\boldsymbol{\Sigma}}_\mathbf{y}^T\delta\tilde{\boldsymbol{\Sigma}}_\mathbf{y})$ is the Euclidean norm of matrix $\delta\tilde{\boldsymbol{\Sigma}}_\mathbf{y}$.

Remark 2. In the variance sensitivity definition, we use $\tilde{\boldsymbol{\Sigma}}_\mathbf{y}$ rather than $\boldsymbol{\Sigma}_\mathbf{y}$ because it is assumed that the sensor noise variance σ_v^2 is known.

Remark 3. Without the minimum, the ratios in Equations (9) and (10) are input dependent. Using input-dependent indices, we will have to design a sensor system for individual changes of the input variation source and it would then be inconvenient. The minimum operator defines the sensitivity indices to be the smallest ratios given all possible combinations of input changes. Equation (11) shows that the above defined sensitivity indices are in fact input independent; they are solely determined by the system matrix $\mathbf{A}$.

Remark 4. The above definition is also consistent with the intuitive relation between sensitivity and diagnosability, which we noted before the definition. The diagnosability conditions obtained in Zhou, Ding, Chen and Shi (2003) are: the mean components are diagnosable if $\mathbf{A}^T\mathbf{A}$ is of full rank and the variance components are diagnosable if the matrix $\{(\mathbf{a}_i^T\mathbf{a}_j)^2\}_{i,j=1}^p$ is of full rank, where $\mathbf{a}_i$ is the ith column vector of $\mathbf{A}$ and $\{\cdot\}_{i,j=1}^p$ is a $p \times p$ matrix. It can be shown that $\{(\mathbf{a}_i^T\mathbf{a}_j)^2\}_{i,j=1}^p = \boldsymbol{\pi}(\mathbf{A})^T\boldsymbol{\pi}(\mathbf{A})$; the proof is included in Appendix 1. It is then apparent that a full diagnosability is guaranteed if the corresponding sensitivity index is nonzero and a system with a zero sensitivity is equivalent to the one that is not fully diagnosable.

The mean- and variance-component sensitivity indices S_v and S_m are also related to the estimation variance of the mean and variance components. To see this, consider the following. For Equation (7), if the variance components in θ are estimated using a maximum likelihood estimator, the variance-covariance matrix of $\hat{\theta}$ is approximated by the inverse of Fisher information matrix as

$$\mathrm{cov}(\hat{\theta}) \propto \left[\left\{\mathrm{tr}\left(\boldsymbol{\Sigma}_\mathbf{y}^{-1}\left(\mathbf{a}_i\mathbf{a}_i^T\right)\boldsymbol{\Sigma}_\mathbf{y}^{-1}\left(\mathbf{a}_j\mathbf{a}_j^T\right)\right\}_{i,j=1}^p\right]^{-1}. \tag{12}$$

A constant is omitted from the right-hand side in Equation (12), thus we used "$\propto$" instead of "$=$". This expression suggests that $\mathrm{cov}(\hat{\theta})$ depends on the true value of θ because $\mathrm{vec}(\boldsymbol{\Sigma}_\mathbf{y}) = \boldsymbol{\pi}(\mathbf{A})\theta + \sigma_\mathrm{v}^2\mathrm{vec}(\mathbf{I}_n)$. Under a normal process condition when there are no outstanding variation sources, we can assume that $\theta = \mathbf{0}$ and then $\boldsymbol{\Sigma}_\mathbf{y} = \sigma_\mathrm{v}^2\mathbf{I}_n$. Then Equation (12) becomes:

$$\begin{aligned}\mathrm{cov}(\hat{\theta})|_{\theta=0} &\propto \sigma_\mathrm{v}^4\left[\left\{\mathrm{tr}\left(\mathbf{a}_i\mathbf{a}_i^T\mathbf{a}_j\mathbf{a}_j^T\right\}_{i,j=1}^p\right]^{-1}\\ &= \sigma_\mathrm{v}^4\left[\left\{\left(\mathbf{a}_i^T\mathbf{a}_j\right)^2\right\}_{i,j=1}^p\right]^{-1} = \sigma_\mathrm{v}^4[\boldsymbol{\pi}(\mathbf{A})^T\boldsymbol{\pi}(\mathbf{A})]^{-1}.\end{aligned} \tag{13}$$

Thus, the variance of the linear parametric function $\mathbf{f}^T\hat{\theta}$ under a normal process condition is:

$$\mathrm{cov}(\mathbf{f}^T\hat{\theta})|_{\theta=0} \propto \sigma_\mathrm{v}^4\mathbf{f}^T[\boldsymbol{\pi}(\mathbf{A})^T\boldsymbol{\pi}(\mathbf{A})]^{-1}\mathbf{f} \tag{14}$$

Then, the maximum variance of $\mathbf{f}^T\hat{\theta}$ for any unit vector $\mathbf{f}$ is the maximum eigenvalue of $[\boldsymbol{\pi}(\mathbf{A})^T\boldsymbol{\pi}(\mathbf{A})]^{-1}$. In other words, S_v, the smallest eigenvalue of $[\boldsymbol{\pi}(\mathbf{A})^T\boldsymbol{\pi}(\mathbf{A})]$, represents the maximum variance of $\mathbf{f}^T\hat{\theta}$, $\forall\|\mathbf{f}\| = 1$, under a normal process condition. The criterion to maximize S_v is then equivalent to selecting an $\mathbf{A}$ to minimize the maximum variance of $\mathbf{f}^T\hat{\theta}$ under a normal process condition.

Similarly, it is not difficult to show that maximizing S_m is equivalent to minimizing the maximum variance of the linear parametric function $\mathbf{p}^T\hat{\boldsymbol{\mu}}$, $\forall\|\mathbf{p}\| = 1$, under a normal process condition.

Since S_m and S_v are different functions of $\mathbf{A}$, a sensor system design may end up with different results, depending on which one of the objectives is chosen, either achieving the maximum mean-detection sensitivity or the maximum variance-detection sensitivity. One can define certain weighted criteria, for instance, $c_1S_\mathrm{m}^2 + c_2S_\mathrm{v}$ as the objective function, where constants c_1 and c_2 determine the trade-off between mean- and variance-detection sensitivities.

However, further investigation found an inequality relationship between S_m and S_v; S_m^2 is a lower bound for S_v for the same $\mathbf{A}$. The result is stated in Lemma 1 and its proof can be found in Appendix 2.

Lemma 1. *For the same* $\mathbf{A}$, $S_\mathrm{v} \geq (S_\mathrm{m})^2$

Based on Lemma 1, we choose to use S_m, i.e., $\lambda_{\min}(\mathbf{A}^T\mathbf{A})$, as the unified criterion for optimal sensor placement in order to simplify the design process. In other words, we optimize S_m, while regulating S_v. Nonetheless, the optimization routines presented in the subsequent section should be equally applicable to the maximization of S_v or other combination using S_m and S_v.

3.2. *Formulations for sensor system optimization*

With the unified sensitivity index $S_\mathrm{m} = \lambda_{\min}(\mathbf{A}^T\mathbf{A})$, the problem of optimal sensor placement will be formulated as follows. The design parameters are the number and locations of sensors, denoted by $\boldsymbol{\varphi}(s) \equiv [X_1\ Z_1\ \cdots\ X_s\ Z_s]^T$. Also, certain constraints should be satisfied. One constraint is that a sensor location has to be a point on the product (geometrical constraint), represented by $G(\cdot) > 0$, where $G(\cdot)$ represents the appropriate geometry function of a manufactured product. Also, in order for OCMMs to be able to perform parallel measurements, we require a second constraint to avoid any possible optical interference among laser beams when taking measurements. Our engineering knowledge indicates that it would be sufficient to enable parallel measurements if we keep the sensor locations at least 100 mm apart from one another. For a given number of sensors, we try to find the optimal sensor locations that maximize S_m, namely:

$$\max_{\boldsymbol{\varphi}(s)}\ S_\mathrm{m} \equiv \lambda_{\min}(\mathbf{A}^T\mathbf{A}), \tag{15}$$

subject to $G(\boldsymbol{\varphi}(s)) \geq 0$ and the 100 mm apart rule.

The criterion for maximizing $\lambda_{\min}(\mathbf{A}^T\mathbf{A})$ is the same as the E-optimality in optimal experimental design, initially proposed by Ehrenfeld (1955), when the $\mathbf{A}$ matrix is considered to be mathematically equivalent to $\mathbf{X}$ in a regression model. Other optimality criteria have also been proposed for optimal experimental design, such as the D-optimality (max $\det(\mathbf{A}^T\mathbf{A})$), the A-optimality (max $\mathrm{tr}(\mathbf{A}^T\mathbf{A})$), where

tr(·) and det(·) are the trace and the determinant of a matrix, respectively. These three measures are related to one another through the eigenvalues of $\mathbf{A}^T\mathbf{A}$, $\{\lambda_i\}_{i=1}^p$:

$$D_{\text{opt}} : \det(\mathbf{A}^T\mathbf{A}) = \Pi_{i=1}^p \lambda_i;$$
$$A_{\text{opt}} : \text{tr}(\mathbf{A}^T\mathbf{A}) = \Sigma_{i=1}^p \lambda_i; \text{ and } E_{\text{opt}} : \lambda_{\min}(\mathbf{A}^T\mathbf{A}). \quad (16)$$

The D-optimality criterion is the most widely used in experimental designs mainly due to its attractive mathematical properties (Fedorov, 1972, p. 138; Atkinson and Donev, 1992, p. 107). It possesses an invariant property under scaling, i.e., optimal experiments can be designed using a group of standardized dimensionless variables instead of the original physical variables.

Those optimality criteria were also used in observer design for control systems (Muller and Weber, 1972; Patton *et al.*, 1989) as well as other engineering systems designs (e.g., the D-optimality was used by Kiridena and Ferreira (1994) and by Wang and Nagarker (1999)). In sensor system design, D-optimality may be interpreted as: if an eigenvalue of $\mathbf{A}^T\mathbf{A}$ is considered as the sensitivity index for the canonical model with a diagonalized $\mathbf{A}$ matrix, the D-optimality corresponds to the multiplication of sensitivities for all input/output pairs. Likewise, the A-optimality is the sum of those sensitivities. These two criteria attempt to optimize an aggregated (multiplication or summation) sensitivity when designing a sensor system. By contrast, the E-optimality is more conservative because it optimizes the smallest sensitivity. Our experience indicates that the E-optimality criterion is more easily accepted by practitioners.

Also, the invariant property of the D-optimality may be inapplicable to an engineering system design because engineering system designs are often accompanied by complex constraints, e.g., the geometric constraints imposed by the shape of the rear quarter panel (refer to Fig. 2(a)). Applying a standardization process will distort the design space and render the final design meaningless. So it is usually impossible to design a general engineering system based on a group of dimensionless standardized variables such as those used in the design of experiments.

The optimization in Equation (15) does not determine the number of sensors. Noting that an increase in sensor number will generally result in a larger S_{m}, people usually try to determine the appropriate sensor number by trading-off between the benefit gained from an increase in S_{m} and the cost of more sensors. However, in engineering practice, it is not easy to quantify the monetary saving associated with an increase in S_{m}. It is thus difficult to define an accurate cost function to attain this trade-off. Alternatively, we can specify a lower bound for S_{m}. Then, the second optimization formulation is to minimize the sensor number, while satisfying a lower bound constraint on S_{m} in addition to other constraints previously specified, i.e.,

$$\min_{\boldsymbol{\varphi}} \quad s, \quad (17)$$

subject to

$$S_{\text{m}} \geq c,\ G(\boldsymbol{\varphi}(s)) \geq 0, \text{ and the 100 mm apart rule.}$$

where c is the lower bound, decided based on engineering requirements.

In the next section, we will mainly study the optimization of Equation (15), which is equivalent to the "exact" design problem in optimal experimental designs. The optimization of Equation (17) can be achieved using the resulting exact design algorithm with a gradually increasing sensor number. In Section 5, we will briefly discuss other considerations in the solution of Equation (17) as well as how to select constant c.

3.3. *Alternative assumptions on sensor noises*

It has been assumed in this paper that the sensors are homogeneous, that is, the variance-covariance matrix of the sensor noise is $\sigma_{\text{v}}^2\mathbf{I}_n$. A general covariance structure for sensor noise has also been studied in the literature (Apley and Shi, 1998; Djurdjanovic and Ni, 2003, 2004). Suppose that in Equation (1), $\mathbf{v}$ is of zero mean and has the variance-covariance matrix $\boldsymbol{\Sigma}_{\text{v}}$. Following a similar treatment that presented in the aforementioned literature and also being consistent with our assumption on σ_{v}^2 in Section 2 (after Equation (3)), $\boldsymbol{\Sigma}_{\text{v}}$ is assumed to be known from the sensor vendor's specifications or it can be estimated from a gauge repeatability study. As such, Equation (1) can be transformed by pre-multiplying $\boldsymbol{\Sigma}_{\text{v}}^{-1/2}$ as:

$$\mathbf{y}^* = \mathbf{A}^*\mathbf{u} + \mathbf{v}^*, \quad (18)$$

where $\mathbf{y}^* = \boldsymbol{\Sigma}_{\text{v}}^{-1/2}\mathbf{y}$, $\mathbf{A}^* = \boldsymbol{\Sigma}_{\text{v}}^{-1/2}\mathbf{A}$, and $\mathbf{v}^* = \boldsymbol{\Sigma}_{\text{v}}^{-1/2}\mathbf{v}$. Apparently, this transformed $\mathbf{v}^*$ has an identity variance-covariance matrix, meaning that we transform the original equation into a format conforming to the homogeneity assumption. Thus, the established design criteria and the subsequent algorithm should be applied to the transformed form of Equation (18).

4. Exchange algorithm and its revision

4.1. *Exchange algorithms from optimal experimental design*

The optimization problems formulated in Section 3 are nonlinear in the design parameter $\boldsymbol{\varphi}$. Standard nonlinear programming approaches (such as quadratic programming) are usually based on a derivative calculation and they will be easily trapped in a local optimum. The derivative-based approaches will be ineffective for a nonconvex design space, imposed by the geometry of panels involved, not to mention those design spaces in the assembly that are not simply connected, for instance, the rear quarter panel, since the window-opening area is not a candidate area for sensor placement.

In the literature on optimal experimental design, exchange algorithms have been proposed to optimize the

D-, A-, and E-optimality design criteria; please see Cook and Nachtsheim (1980) and Atkinson and Donev (1992) for reviews and comparisons of exchange algorithms. According to Meyer and Nachtsheim (1995), exchange algorithms have more freedom to maneuver in a complicated design space because each of its exchanges involves only a section of the design parameters (associated with one design point). Thus, exchange algorithms can be more effective in escaping from local optima than derivative-based nonlinear programming. Additionally, exchange algorithms have other advantages that favor their use in engineering system designs and these include: their procedures are intuitive and implementation is easy; the algorithms are flexible and can easily handle complicated constraints in engineering system design; they can also be used for a wide variety of design criteria.

To use exchange algorithms, we first discretize the continuous design space. We call the resulting discretized design space with N_c candidate sensor locations the candidate space (denoted as D_c) and the space with s current sensor locations the sensor space (denoted as D_s). The basic idea of an exchange algorithm is to start with a set of s design points (i.e., the sensor location) in D_s, usually randomly selected, and exchange the current design points with those points in the much larger D_c in order to improve the chosen design criterion. In exchange algorithms, however, the exchange action is not carried out for every single point. However, every point in D_s is *tested* against a point in D_c, meaning that any improvement in the design criterion is recorded as if the point in D_s had been exchanged with a point in D_c. There are different variants to the above basic idea, depending on how often the action of exchange is actually carried out. One option is to perform the exchange action after all points in D_s have been tested against the entire set of points in D_c. It exchanges the pair of points, one in D_s and one in D_c, which made the maximum improvement in the design criterion. This option is actually the celebrated Fedorov exchange algorithm. Another option is to perform the exchange action for every point in D_s after that design has been tested against all points in D_c. In other words, point i in D_s will be exchanged with the point in D_c that maximizes the improvement in the design criterion and the same action is repeated in a sequential order for $i = 1, 2, \ldots s$. The second option is the modified Fedorov exchange. In combinatorial optimization, these are two extreme cases of a general k-exchange algorithm, with $k = s$ for the Fedorov exchange and $k = 1$ for the modified Fedorov exchange (Aarts and Lenstra, 1997).

When applying them to the sensor placement problem, we notice that exchange algorithms, especially the Fedorov algorithm, can consume a great deal of CPU time for cases containing a large number of sensors. This is not surprising since the exchange algorithm was initially developed for experimental design with a relatively small number of factors and experiments (Cook and Nachtsheim, 1980).

In the following section, we will introduce and implement a sort-and-cut procedure that will shorten the computation time without sacrificing much of the optimal value it finds. A similar procedure for the revised exchange algorithm has been implemented in the application of fixture layout design (Kim and Ding, 2004). In this paper, we need to tailor the algorithm under the specific context of sensor placement; more details are presented in the following.

4.2. *Fast exchange algorithm with a sort-and-cut procedure*

Let us first conceptually understand the factors that affect the algorithm's computation time. Define the process to pass over the entire D_c set once as a "loop". There are two major factors that affect the run time: (i) the average number of loops; and (ii) the size of the candidate space N_c. In order to reduce the computation time, we will have to reduce the average number of loops as well as the size of the candidate design space N_c. The following sort-and-cut procedure is employed to achieve both goals. The basic idea is to perform multiple exchanges in each loop to reduce the average number of loops, and after each loop, discard a subset of the candidate design points to reduce N_c.

In a design for a uniform coverage design in molecule selection, Lam *et al.* (2002) suggested that instead of exchanging one design point per loop, one may want to exchange multiple candidate points in the upper tail of the distribution of improvements in the design criterion among all the candidates. This means that the number of design points that will be exchanged during each loop will be more than one so that the average number of loops required to replace all random initial designs can be reduced.

Different to the modified Fedorov algorithm, which uses a design point in D_s as the pivoting point in each exchange, Lam *et al.* (2002) used a candidate point in D_c as the pivoting pointing. Specifically, for each point in D_c, add this point to D_s and augment an s-sensor D_s to an $(s + 1)$-sensor design. In order to maintain D_s as a s-sensor design, we will put one sensor, which makes the smallest decrease in the sensitivity index, from the augmented D_s back to D_c. The above process constitutes an exchange action.

Following the algorithm in Lam *et al.* (2002), we will record the improvement in design criterion that a candidate location can make if the corresponding exchange is indeed carried out. After each exchange action, denote by Δ the improvement in the S_m criterion, i.e., $\Delta \equiv S_m^{new} - S_m^{old}$. Record all Δ_js ($j = 1, \ldots, N_c$) when we loop through the N_c candidate locations. Sort the value of the Δ_js in a descending order as $\Delta_{(1)} \geq \Delta_{(2)} \geq \ldots$ and so on. The distribution of improvements is approximated by the sorted values $\Delta_{(j)}$. Then, we will set an integer number q so that the first q candidate locations in the upper tail of $\Delta_{(j)}$ will be exchanged in each loop.

However, we also want to reduce the total number of candidate points N_c. The sorted values of the design improvement $\Delta_{(j)}$ actually provide us with valuable information about the potential of a candidate location. Those

candidate locations with a low Δ value are less likely to be picked up by the exchange algorithm in the next iterations. Thus, we could remove a section of the candidate points after each iteration. Denote by α the portion of candidate points that will be kept after a cut.

To implement this sort-and-cut procedure, there are two parameters to be determined: q and α. In our sensor system design problem, the sensor number in D_s is usually only a small percentage of that in the D_c space (for instance, $D_s = 10$ but $D_c = 10\,000$). We recommend an aggressive choice of α, e.g., from 10 to 20% (for the above example, if $\alpha = 10\%$ for the first two iterations, the remaining locations in D_c are still about 10 times of that in D_s).

For a sort-and-cut procedure to work, the assumption is that the distribution of improvements approximated by data from the previous exchange routine can represent well enough the distribution in the next exchange. However, whenever an exchange happens, the distribution cannot be exactly the same since the D_s space generating that distribution is no longer the same. The common ground is actually constituted by the design points that are not exchanged in the last iteration. Thus, for the above assumption to hold, q should be smaller than s, namely $1 \leq q < s$. When q is close to s, however, almost all the sensors in D_s will be exchanged in one iteration and the distribution recorded in $\Delta_{(j)}$ from the previous loop does not truly represent the distribution for the new D_s space. A subsequent exchange based on the $\Delta_{(j)}$ could make a poor choice that has to be re-done in the following loops. On the other hand, too small a q will result in too few exchanges per loop and thus miss our original goal of having multiple exchanges to reduce the average number of loops. We therefore recommend selecting $q = s/2$ to strike a balance so that half of the sensors in D_s will not be exchanged to provide a common ground for distribution and half of the sensors will be exchanged to reduce the average loop number.

Here we are different to Lam *et al.* (2002). Instead of using q as the direct control on the number of exchanges, Lam *et al.* (2002) set $\Delta_{(q)}$ as the threshold to control the exchange, i.e., if there is an improvement greater than $\Delta_{(q)}$, then carry out the exchange. We find that using $\Delta_{(q)}$ in our application is not effective. The difference is due to our inclusion of a subsequent cut action, which is not included in the procedure of Lam *et al.* (2002). The effect of the cut action, as explained in the above paragraph, requires us to have more direct control over the number of exchanges, which cannot be fulfilled by using $\Delta_{(q)}$.

The algorithm for an s-sensor exact design is summarized as follows.

Step 1. The candidate design space D_c is discretized and s locations are randomly selected to form D_s. Perform exchanges for all locations in D_c and establish the initial distribution for Δ.

Step 2. In every iteration:

2.1 rank the sensor locations in D_c in a descending order according to their Δ_j values;

2.2 cut off those locations with low Δ_j values and keep the top $\alpha \times 100\%$ of candidate points in D_c;

2.3 exchange the sensor locations in D_c, which satisfy the constraint condition, with the sensor locations in the current D_s space in a sequential order, starting from the one with the largest Δ_j. Repeat this exchange for the first q sensors in the sorted D_c.

Step 3. Repeat Step 2 until the improvement in design criterion for two successive designs is smaller than a predetermined threshold (we used 0.1%).

4.3. *Implementation and comparison*

The algorithms described in the above sections were coded in Matlab and compared on the same computer. We measure the algorithm efficiency by the time that it takes to find the optimal value. We measure the algorithm effectiveness by the average value of optimal solutions (i.e., the average S_m) it finds when a group of random sensor layouts is used as the initial design. In the literature on algorithm comparison (e.g., Cook and Nachtsheim, 1980), a relative effectiveness R is often used, which is defined as the ratio of the average optimal value over the best optimal solution found by all the algorithms in the comparison under the same setting. We use both measures in this study.

For the SUV side panel assembly in Fig. 2(a), we discretize it with candidate points 10 millimeters apart. Our engineering experience indicates that this resolution of discretization is sufficient to generate a fine enough grid on a panel that has a size of over several thousands of millimeters. The discretization results in a total of $N_c = 13{,}304$ candidate positions in D_c. As for the sort-and-cut procedure, we choose $\alpha = 0.1$ and $q = s/2$. In this application, we only implement the cut action in the first iteration.

The optimization results are summarized in Table 1. The values in Table 1 are the average of 50 trials with randomly

Table 1. Comparisons of the resulting algorithms

		Average time computer (seconds)	*Average maximal* S_m	R
$s = 2$	Fedorov	27.24	1.0044	0.9995
	Modified Fedorov	19.12	0.9975	0.9926
	Fast exchange	2.47	0.9300	0.9255
$s = 4$	Fedorov	98.12	2.0105	0.9980
	Modified Fedorov	50.61	2.0088	0.9972
	Fast exchange	3.24	1.9794	0.9826
$s = 6$	Fedorov	106.50	3.0145	0.9979
	Modified Fedorov	63.50	3.0150	0.9981
	Fast exchange	4.21	2.9715	0.9837
$s = 8$	Fedorov	187.95	4.0177	0.9972
	Modified Fedorov	87.55	4.0188	0.9975
	Fast exchange	4.29	3.9913	0.9907
$s = 10$	Fedorov	376.80	5.0214	0.9978
	Modified Fedorov	142.04	5.0207	0.9976
	Fast exchange	5.11	4.9965	0.9928

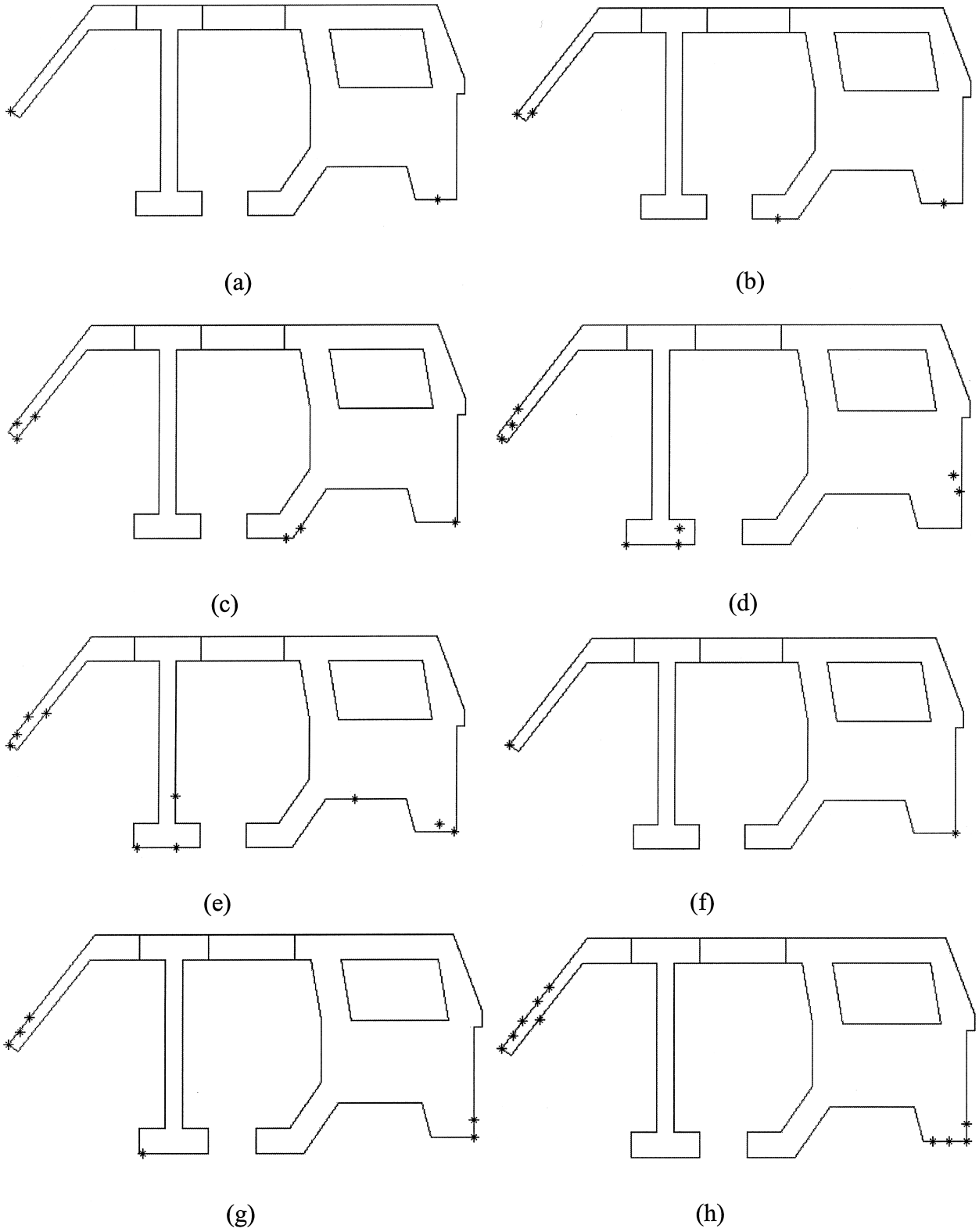

Fig. 3. Optimal sensor layouts: (a) $s = 2$; (b) $s = 4$; (c) $s = 6$; (d) $s = 8$; (e) $s = 10$; (f) $s = 2$ using S_v; (g) $s = 6$ using s_v, and (h) $s = 10$ using S_v.

generated initial designs for a sensor number ranging from $s = 2$ to $s = 10$. To save space, only those results for an even s are displayed, however, the analysis can be generally extended to odd s values. We also tried the algorithm on a problem containing a larger number of sensors; the fast exchange can solve a 60-sensor design within a minute, whereas the Fedorov algorithm will take hours. The results clearly demonstrate that the fast exchange algorithm is more efficient than the other tested algorithms without any significant loss of effectiveness.

5. Optimal sensor layout and discussion

The resulting optimal sensor layouts with an even sensor number ($s = 2, \ldots, 10$) are shown in Fig. 3(a–h), where a "*" mark indicates a sensor location. From the layouts, we observe that the sensors are located in areas close to the panel boundary and many of them are actually on an edge. This raises the question that if we can reduce our number of candidate locations by limiting our search to the geometry boundary of each part in the first place? The answer is yes. However, we should also notice that not all the sensor locations are on an edge (refer to the cases for $s = 8$ and $s = 10$). Based on empirical knowledge alone, it is nontrivial to determine a search area that contains all the potentially good sensor locations. In this study, we used the approximated distribution of design improvements in the sort-and-cut procedure, which provides more reliable information and quantitative evaluation to find the potentially good sensor locations. The algorithm is fairly general and it can be used together with a reduced candidate pool to further improve the algorithm efficiency, had the aforementioned intuitive rule been implemented before the search.

The above exact design algorithm finds an optimal sensor layout when the sensor number is specified. In order to optimally solve Equation (17), one may wonder if we can use a sequential routine, i.e., we start from an optimal two-sensor design (which is easy to find), and then sequentially add one more sensor from D_c to D_s which maximizes the resulting S_m, until eventually $S_m \geq c$. For such a sequential strategy to work well, the sensor layout for the $(s-1)$ sensors should be a subset of the optimal layout for the s sensors. This might be true when the sensor number is small ($s < 4$) but it does not stay that way when the sensor number become larger. Although we did not show the sensor layouts for odd sensor numbers, they actually agree with the phenomenon demonstrated by the layouts displayed in Fig. 3(a–h). Therefore, the sequential routine could miss the optimal layout. Nonetheless, we can combine the sequential probing and the exact design. That is, first use sequential routine to probe and find a sensor number which can yield $S_m \geq c$ and then switch to an exact design routine to find the optimal sensor locations for sensor numbers around the one found by the sequential routine. This way, we can skip a number of time-consuming exact designs, especially when the resulting sensor number is relatively large.

In the optimization of Equation (17) we specify a constant c to stop the algorithm. Usually the choice of c depends on engineering requirements and it is specified for a particular context. In this study, we can choose c based on the accuracy requirement. It is known that the OCMM, although more agile and faster, is not as accurate as the mechanical CMM: the OCMM measurement repeatability is five to 10 times lower than that of the CMM (Hu, 1990). Let us be optimistic and consider that an OCMM is five times less accurate than a CMM, namely, $\sigma^2_{v,OCMM} = 5\sigma^2_{v,CMM}$. According to the arguments after Lemma 1, under a normal process condition, the maximum variance in estimating process mean components, and also the lower bound for the maximum variance in estimating the process variance components, is:

$$\frac{\sigma^2_{v,OCMM}}{S_m}.$$

In order to achieve the same variance level as that obtained if a CMM was used to directly measure the process variation source, we require that $\sigma^2_{v,OCMM}/S_m < \sigma^2_{v,CMM}$, which translates into $S_m > 5$. Thus, we will choose $c = 5$ in this study. Certainly, this value may change when the accuracy requirement is different. However, the above logic can still be applied to determine an appropriate c. In choosing $c = 5$, we find that 10 sensors will provide a sensing capability equally as good as that of a CMM.

One may also wonder what happens if we use S_v instead of S_m as our design criterion. Examples using S_v are shown in Fig. 3(f–h) for $s = 2$, 6 and 10. Interestingly, the sensor layouts using S_v bear a strong resemblance to those using S_m, especially in terms of the areas where the sensors are located. Of course, the resulting layouts using S_v deviate to some extent from those using S_m and the deviation is less obvious for $s = 2$ but is more appreciable for a larger s.

6. Conclusions

This paper presents statistical and optimization methods for the coordinate sensor placement for the estimation of the mean and variance components of variation root causes. In an exclusion to previously defined diagnosability indices, sensitivity indices of a sensor system are defined to characterize the capability of the sensor system to detect the underlying process mean and variance changes. Mathematically, they are equivalent to the E-optimality criterion proposed in optimal experimental design studies. Optimization of the design criterion is then fulfilled by revising the exchange algorithm. We have presented a fast exchange routine with a sort-and-cut procedure, which considerably reduces the algorithm's computation time while maintaining the optimal value it can find.

It is noted that the exchange algorithm, including its variants, has been intensively studied and broadly applied in experimental designs. However, its application to engineering system design is not extensively explored. Given the flexibility of exchange algorithms and the empirical evidence that the presented fast exchange algorithm is capable of handling large-scale sensor system designs, much more needs to be done to present theoretical justifications of the algorithm's performances.

Dimensional variation reduction is critical in ensuring a high product quality in discrete-part manufacturing. The effective use of sensor data to diagnose variation sources depends to a great extent on the optimal design of a sensor system with multiple sensors. The optimal design of sensor systems will surely make the task of variation root-cause diagnosis more meaningful and efficient. The criteria and methods presented in this paper will find applications beyond coordinate sensor placement since the approach is based on a general linear system model.

References

Aarts, E. and Lenstra, J.K. (1997) *Local Search in Combinatorial Optimization*, Wiley, New York, NY.

Apley, D.W. and Lee, H.Y. (2003) Identifying spatial variation patterns in multivariate manufacturing processes: a blind separation approach. *Technometrics*, **45**, 187–198.

Apley, D.W. and Shi, J. (1998) Diagnosis of multiple fixture faults in panel assembly. *Transactions of the ASME, Journal of Manufacturing Science and Engineering*, **120**, 793–801.

Apley, D.W. and Shi, J. (2001) A factor-analysis method for diagnosing variability in multivariate manufacturing processes. *Technometrics*, **43**, 84–95.

Atkinson, A.C. and Donev, A.N. (1992) *Optimum Experimental Designs*, Oxford University Press, New York, NY.

Carlson, J.S., Lindkvist, L. and Soderberg, R. (2000) Multi-fixture assembly system diagnosis based on part and subassembly measurement data, in *Proceedings of the 2000 ASME Design Engineering Technical Conference*.

Ceglarek, D. and Shi, J. (1995) Dimensional variation reduction for automotive body assembly. *Manufacturing Review*, **8**, 139–154.

Ceglarek, D. and Shi. J. (1996) Fixture failure diagnosis for autobody assembly using pattern recognition. *Transactions of the ASME, Journal of Engineering for Industry*, **188**, 55–65.

Ceglarek, D., Shi, J. and Wu, S.M. (1994) A knowledge-based diagnostic approach for the launch of the auto-body assembly process. *Transactions of the ASME, Journal of Engineering for Industry*, **116**, 491–499.

Cook, R.D. and Nachtsheim, C.J. (1980) A comparison of algorithms for constructing exact D-optimal designs. *Technometrics*, **22**, 315–323.

Ding, Y., Ceglarek, D. and Shi, J. (2000) Modeling and diagnosis of multistation manufacturing processes: part I state space model, in *Proceedings of the 2000 Japan/ USA Symposium on Flexible Automation*.

Ding, Y., Ceglarek, D. and Shi, J. (2002a) Fault diagnosis of multi-station manufacturing processes using state space approach. *Transactions of the ASME, Journal of Manufacturing Science and Engineering*, **124**, 313–322.

Ding, Y., Ceglarek, D. and Shi, J. (2002b) Design-evaluation of multistation manufacturing processes using state space approach. *Transactions of the ASME, Journal of Mechanical Design*, **124**, 408–418.

Ding, Y., Kim, P., Ceglarek, D. and Jin, J. (2003) Optimal sensor distribution for variation diagnosis for multi-station assembly processes. *IEEE Transactions on Robotics and Automation*, **19**, 543–556.

Djurdjanovic, D. and Ni, J. (2001) Linear state space modeling of dimensional machining errors. *Transactions of NAMRI/ SME*, **XXIX**, 541–548.

Djurdjanovic, D. and Ni, J. (2003) Bayesian approach to measurement scheme analysis in multistation machining systems. *Journal of Engineering Manufacture*, **217**, 1117–1130.

Djurdjanovic, D. and Ni, J. (2004) Measurement scheme synthesis in multi-station machining systems. *Transactions of the ASME, Journal of Manufacturing Science and Engineering*, **126**, 178–188.

Ehrenfeld, S. (1955) On the efficiency of experimental designs. *Annals of Mathematical Statistics*, **26**, 247–255.

Fedorov, V.V. (1972) *Theory of Optimal Experiments*, Academic Press, New York, NY.

Hu, S.J. (1990) Impact of 100% measurement data on statistical process control (SPC) in automobile body assembly. Ph.D. dissertation, The University of Michigan, Ann Arbor, MI.

Hu, S.J. and Wu, S. (1992) Identifying sources of variation in automobile body assembly using principal component analysis. *Transactions of NAMRI/ SME*, **XX**, 311–316.

Jin, J. and Shi, J. (1999) State space modeling of sheet metal assembly for dimensional control. *Transactions of the ASME, Journal of Manufacturing Science and Engineering*, **121**, 756–762.

Khan, A. and Ceglarek, D. (2000) Sensor optimization for fault diagnosis in multi-fixture assembly systems with distributed sensing. *Transactions of the ASME, Journal of Manufacturing Science and Engineering*, **122**, 215–226.

Khan, A., Ceglarek, D. and Ni, J. (1998) Sensor location optimization for fault diagnosis in multi-fixture assembly systems, *Transactions of the ASME, Journal of Manufacturing Science and Engineering*, **120**, 781–792.

Khan, A., Ceglarek, D., Shi, J., Ni, J. and Woo, T.C. (1999) Sensor optimization for fault diagnosis in single fixture systems: a methodology. *Transactions of the ASME, Journal of Manufacturing Science and Engineering*, **121**, 109–121.

Kim, P. and Ding, Y. (2004) Optimal design of fixture layout in multistation assembly processes. *IEEE Transactions on Automation Science and Engineering*, **1**, 133–145.

Kiridena, V.S.B. and Ferreira, P.M. (1994) Parameter estimation and model verification of first-order quasi static error model for three-axis machine centers. *International Journal of Machine Tools and Manufacture*, **34**, 101–125.

Lam, R.L.H., Welch, W.J. and Young, S.S. (2002) Uniform coverage designs for molecule selection. *Technometrics*, **44**, 99–109.

Lawless, J.F., Mackay, R.J. and Robinson, J.A. (1999a) Analysis of variation transmission in manufacturing processes—part I. *Journal of Quality Technology*, **31**, 131–142.

Lawless, J.F., Mackay, R.J. and Agrawal, R. (1999b) Analysis of variation transmission in manufacturing processes—part II. *Journal of Quality Technology*, **31**, 143–154.

Mantripragada, R. and Whitney, D.E. (1999) Modeling and controlling variation propagation in mechanical assemblies using state transition models. *IEEE Transaction On Robotics and Automation*, 15, pp. 124–140.

McCulloch, C.E. and Searle, S.R. (2000) *Generalized Linear and Mixed Models*, Wiley, New York, NY.

Meyer, R.K. and Nachtsheim, C.J. (1995) The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*, **37**, 60–69.

Muller, P.C. and Weber, H.I. (1972) Analysis and optimization of certain qualities of controllability and observability for linear dynamical systems. *Automatica*, **8**, 237–246.

Patton, R., Frank, P. and Clark, R. (1989) *Fault Diagnostics in Dynamic Systems: Theory and Applications*, Prentice Hall, Upper Saddle River, NJ.

Paul, R.P. (1981) *Robot Manipulators: Mathematics, Programming, and Control*, MIT Press, Cambridge, MA.

Powell, M.J.D. (1992) A direct search optimization method that models the objective and constraint functions by linear interpolation. Numerical analysis reports DAMTP 1992/NA: The University of Cambridge, Cambridge, UK.

Rao, C.R. and Kleffe, J. (1988) *Estimation of Variance Components and Applications*, North-Holland, Amsterdam, The Netherlands.

Schott, J.R. (1997) *Matrix Analysis for Statistics*, Wiley, New York, NY.

Shalon, D., Gossard, D., Ulrich, K. and Fitzpatrick, D. (1992) Representing geometric variations in complex structural assemblies on CAD systems, in *Proceedings of the 19th Annual ASME Advances in Design Automation Conference*, **44**(2), pp. 121–132.

Soderberg, R. and Carlson, J. S. (1999) Locating scheme analysis for robust assembly and fixture design, in *Proceedings of the 1999 ASME Design Engineering Technical Conferences*, Phoenix, AZ.

Thornton, A.C. (1999) A mathematical framework for the key characteristic process. *Research in Engineering Design*, **11**, 145–157.

Wang, Y. and Nagarkar, S.R. (1999) Locator and sensor placement for automated coordinate checking fixtures. *Transactions of the ASME, Journal of Manufacturing Science and Engineering*, **121**, 709–719.

Whitney, D.E., Gilbert, O. and Jastrzebski, M. (1994) Representation of geometric variations using matrix transforms for statistical tolerance analysis in assemblies. *Research in Engineering Design*, **6**, 191–210.

Zhou, S., Ding, Y., Chen, Y., and Shi, J. (2003) Diagnosability study of multistage manufacturing processes based on linear mixed-effects models. *Technometrics*, **45**, 312–325.

Zhou, S., Huang, Q. and Shi, J. (2003) State space modeling for dimensional monitoring of multistage machining process using differential motion vector. *IEEE Transactions on Robotics and Automation*, **19**, 296–308.

Appendices

Appendix 1

Proof of $\{(\mathbf{a}_i^T\mathbf{a}_j)^2\}_{i,j=1}^p = \boldsymbol{\pi}(\mathbf{A})^T\boldsymbol{\pi}(\mathbf{A})$. Recall that $(\mathbf{a}_i^T\mathbf{a}_j)^2 = \mathrm{tr}(\mathbf{a}_i\mathbf{a}_i^T\mathbf{a}_j\mathbf{a}_j^T)$ and $\mathrm{tr}(\mathbf{AB}) = \mathrm{vec}(\mathbf{A})^T\mathrm{vec}(\mathbf{B})$ for any symmetric matrices $\mathbf{A}$ and $\mathbf{B}$. The (i,j) element in $\{\mathrm{tr}(\mathbf{a}_i\mathbf{a}_i^T\mathbf{a}_j\mathbf{a}_j^T)\}_{i,j=1}^p$ is $(\mathrm{vec}(\mathbf{a}_i\mathbf{a}_i^T))^T\mathrm{vec}(\mathbf{a}_j\mathbf{a}_j^T)$. Actually, $\mathrm{vec}(\mathbf{a}_i\mathbf{a}_i^T)$ is the ith column vector in $\boldsymbol{\pi}(\mathbf{A})$. This leads to the conclusion that $\boldsymbol{\pi}(\mathbf{A})^T\boldsymbol{\pi}(\mathbf{A}) = \{(\mathbf{a}_i^T\mathbf{a}_j)^2\}_{i,j=1}^p$. ■

Appendix 2

Proof of Lemma 1. We know that $\{\mathbf{a}_i^T\mathbf{a}_j\}_{i,j=1}^p = \mathbf{A}^T\mathbf{A}$. Then, $\{(\mathbf{a}_i^T\mathbf{a}_j)^2\}_{i,j=1}^p$ is actually $(\mathbf{A}^T\mathbf{A}) * (\mathbf{A}^T\mathbf{A})$. From the above proof, we know that $\{(\mathbf{a}_i^T\mathbf{a}_j)^2\}_{i,j=1}^p = \boldsymbol{\pi}(\mathbf{A})^T\boldsymbol{\pi}(\mathbf{A})$, which means that $\boldsymbol{\pi}(\mathbf{A})^T\boldsymbol{\pi}(\mathbf{A}) = (\mathbf{A}^T\mathbf{A}) * (\mathbf{A}^T\mathbf{A})$. Theorem 7.28 in Schott (1997, p. 276) states that $\lambda_{\min}(\mathbf{A} * \mathbf{B}) \geq \lambda_{\min}(\mathbf{AB})$ for any non-negative definite matrices $\mathbf{A}$ and $\mathbf{B}$. The matrix $\mathbf{A}^T\mathbf{A}$ is a non-negative definite matrix so that we can have $\lambda_{\min}((\mathbf{A}^T\mathbf{A}) * (\mathbf{A}^T\mathbf{A})) \geq \lambda_{\min}((\mathbf{A}^T\mathbf{A})(\mathbf{A}^T\mathbf{A}))$. Since $\lambda_{\min}((\mathbf{A}^T\mathbf{A})(\mathbf{A}^T\mathbf{A})) = \lambda_{\min}^2((\mathbf{A}^T\mathbf{A}))$, the above inequality is equivalent to $\lambda_{\min}(\boldsymbol{\pi}(\mathbf{A})^T\boldsymbol{\pi}(\mathbf{A})) \geq \lambda_{\min}^2(\mathbf{A}^T\mathbf{A})$. ■

Biographies

Qinyan Liu received B.S and M.S. degrees in Mechanical Engineering from Tsinghua University in 2000 and 2002, and a M.S. in Industrial Engineering from Texas A&M University in 2004. He is currently a graduate student in the Department of Management Science & Engineering at Stanford University. His research interests are in the area of quality engineering and applied optimization, including optimal sensor system design, industrial experiments design, simulation and optimization, and stochastic programming applied to supply chains. He is a student member of IIE, IEEE, and INFORMS.

Yu Ding received a B.S degree in Precision Engineering from the University of Science and Technology of China in 1993, an M.S. in Precision Instruments from Tsinghua University, China in 1996, an M.S. in Mechanical Engineering from the Pennsylvania State University in 1998, and a Ph.D. in Mechanical Engineering from the University of Michigan in 2001. He is currently an Assistant Professor in the Department of Industrial Engineering at Texas A&M University. His research interests are in the area of quality engineering and applied statistics, including in-process variation diagnosis, diagnosability analysis of distributed sensor systems, optimal sensor system design, and process-oriented robust design and tolerancing. His current research is sponsored by the National Science Foundation, Nokia, and the State of Texas Higher Education Coordinating Board. He has received a number of awards for his work, including a CAREER Award from the National Science Foundation in 2004 and the Best Paper Award from the ASME Manufacturing Engineering Division in 2000. He currently serves as a Department Editor of *IIE Transactions*. He is a member of IIE, ASME, SME, and INFORMS.

Yong Chen received a B. S. degree in Computer Science from Tsinghua University, China in 1998, an M.A. in Statistics and a Ph.D. in Industrial & Operations Engineering from the University of Michigan in 2003. He is currently an Assistant Professor in the Department of Mechanical and Industrial Engineering at the University of Iowa. His research interests include the fusion of advanced statistics, applied operations research, and engineering knowledge to develop in-process quality and productivity improvement methodologies to achieve automatic fault diagnosis, proactive maintenance, and integration of quality and reliability information in complex manufacturing systems. He received the Best Paper Award from *IIE Transactions on Quality and Reliability Engineering* in 2005. He is a member of IIE, ASQ, INFORMS, and SME.

Contributed by the Manufacturing Systems Department