IIE Transactions (2009) **41**, 309–322

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/07408170802510408

Process capability surrogate model-based tolerance synthesis for multi-station manufacturing systems

WENZHEN HUANG[1,*], TIRAWAT PHOOMBOPLAB[2,3] and DARIUSZ CEGLAREK[2,3]

[1]*Department of Mechanical Engineering, University of Massachusetts, Dartmouth, MA 02747, USA*
E-mail: whuang@umassd.edu
[2]*Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA*
[3]*International Digital Laboratory, WMG, University of Warwick, Coventry CV4 7AL, UK*
E-mail: {T.Phoomboplab, d.j.ceglarek}@warwick.ac.uk

Received February 2007 and accepted November 2007

The main challenges in tolerance synthesis for complex assembly design currently are: (i) to produce a simplified deterministic model that is able to formulate general statistic models in complex assembly problems; (ii) to lower the high computation intensity required in optimization studies when the process capability (yield) model is used for key product characteristics. In this paper, tolerance synthesis for complex assemblies is defined as a probabilistic optimization problem which allows the modeling of assemblies with a general multivariate statistical model and complex tolerance regions. An approach is developed for yield surrogate model generation based on an assembly model in multi-station manufacturing systems, computer experiments, multivariate distribution transformation and regression analysis. Therefore, efficient gradient-based approaches can be applied to avoid the intensive computation in direct optimization. Industrial case studies are presented to illustrate and validate the proposed methodology and compared with the existing tolerance synthesis methods.

[Supplementary materials are available for this article. Go to the publisher's online edition of *IIE Transactions* for the following free supplemental resources: Appendix]

Keywords: Tolerance synthesis, manufactoring, surrogate model, process capability

1. Introduction

Dimensional variation control is one of the most challenging problems in mechanical assemblies, in particular, in multi-station assembly systems, such as in automotive and aerospace industries. A typical automotive body assembly system involves several hundreds of parts, about 50 stations, thousands of locators and hundreds of measuring points. The models for assemblies in automotive and aerospace industries may have thousands of design variables or Key Control Characteristics (KCCs), and hundreds of design responses (Key Process Characteristics (KPCs)). It incorporates both product and process factors. The latest research advances have provided models for controlling variation in complex assembly systems (Chase *et al.*, 1990; Jin and Shi, 1999; Ding *et al.*, 2002; Ceglarek *et al.*, 2004; Ding *et al.*, 2005; Shi, 2006; Huang, Lin, Bezdecny, Kong and Ceglarek, 2007; Huang, Lin, Kong and Ceglarek, 2007). The so-called Stream of Variation (SOVA) model for an assembly system can be mathematically expressed as a Multi-Input-Multi-Output (MIMO) system (Shi, 2006; Huang, Lin, Bezdecny, Kong and Ceglarek, 2007; Huang, Lin, Kong and Ceglarek, 2007). However, neither well established nor recently emerging techniques are capable of effectively dealing with the tolerance synthesis problems of such complex systems.

The tolerance design synthesis for an assembly was formulated as a probabilistic optimization problem by Lee and Woo (1989). The objective is designed to optimally meet functional and economic requirements through properly assigning tolerances to component dimensions. The functional requirements are defined as the process capability to make quality products or yield which defines an implicit constraint function. In multivariate cases the yield assessment is difficult and relies exclusively on Monte Carlo simulations. When embedded in internal iteration loops of optimization algorithms, it can be extremely computationally intensive. Simplifications have been introduced in the literature to convert the probabilistic optimizations into deterministic problems. Despite the simplicity, the major concerns in the simplification of multivariate problems are threefold: (i) difficulty achieving a specified system process

*Corresponding author

capability through individual tolerance assignment; (ii) creation of conservative solutions (Shiu *et al.*, 2003); and (iii) lack of generality (e.g., for non-normal KPCs and asymmetric or irregular tolerance regions).

Targeting these challenges, a new technique is proposed in this paper. The uniqueness of the approach is due to its ability in: (i) the formulation of the synthesis problem in a general framework in terms of process capability; (ii) the construction of a yield surrogate model using candidate subspace searching and a novel sampling technique (computer experiment); and (iii) surrogate-model-based optimization.

The outline of this paper is as follows. Section 2 presents a literature review, followed by a briefing on three-dimensional assembly models in Section 3. Section 4 formulates the tolerance synthesis. The proposed yield Surrogate Model (SM) technique is presented in Section 5. A validation case and comparative studies are presented in Section 6. Conclusions are drawn in Section 7.

2. Related works

Tolerance synthesis represents the computer-aided design technique of assigning tolerances to individual components to minimize the manufacturing cost and ensure that all the function requirements of an assembly are met. The function requirements are represented as conformity to specifications of key dimensions of an assembly. Various cost functions, such as reciprocal, exponential, etc., are available (Wu *et al.*, 1988). These functions are designed to characterize the trend that the cost is inversely affected by tolerances.

2.1. *Tolerance synthesis for manufacturing and design*

There are two types of synthesis problems: *synthesis for manufacturing* (tolerance transfer) in process planning and *synthesis for assembly* in design.

In order to reduce the manufacturing cost, tolerance allocation based on sequences, set-ups and multiple process alternatives have been extensively investigated by various authors. Tolerance charting, developed in the 1950s and popularized in the 1960s as a simple manual tool, receives considerable interest from industrial practitioners. It involves converting design tolerances into tolerances for the manufacturing processes and is a part of process design. Extensive efforts have been devoted to automating the manual tolerance charting and create a computer-aided charting procedure. A comprehensive review on tolerance charting can be found in Ngoi and Kuan (1995). Considerable effort has been devoted to relating tolerance allocation with process planning (Zhang and Wang, 1993a, 1993b; Roy and Feng, 1997). The purpose of process planning in terms of tolerance allocation is to minimize cost through the selection of set-ups and operation sequences. Related process planning topics, such as set-ups in NC machining, fixture planning and sequence of operations were extensively investigated by Zhang and Wang (1993a). The simultaneous synthesis of design and process tolerances has also been investigated in the last decade. Zhang and Wang (1993b) developed a general mathematical model of optimal tolerancing supporting concurrent engineering to determine optimal machining tolerances in product design. Comprehensive reviews on synthesis for manufacturing can be found in Ngoi and Ong (1998) and Hong and Chang (2002).

One of the primary concerns in tolerance synthesis in design is the assembly model which relates component tolerances to assembly key dimension tolerances. An assembly model characterizes the dimensional variation flow in an assembly system. Several variation propagation (stack-up) models have been developed and are summarized in Hong and Chang (2002).

Comprehensive research has been conducted in the last two decades on assembly modeling for tolerancing and variation analysis (Chase *et al.*, 1990; Jin and Shi, 1999; Ding *et al.*, 2002; Ceglarek *et al.*, 2004; Ding *et al.*, 2005; Shi, 2006; Huang, Lin, Bezdecny, Kong and Ceglarek, 2007; Huang, Lin, Kong and Ceglarek, 2007). Efforts have also been made in the last decade to characterize variation propagation in multi-station assembly processes. Jin and Shi (1999) and Ding *et al.* (2000) initiated a State Space Model (SSM) for variation modeling, wherein a spatial indexed model and observation equation were established. Ding *et al.* (2002), Ceglarek *et al.* (2004) Ding *et al.* (2005), Chen *et al.* (2006), and Shi (2006) developed a two dimensional SSM for automotive body assemblies and applied it to process-oriented tolerancing problems. More recently, a SSM-based three-dimensional assembly model was developed, which integrates both part errors and fixture errors and thus enables integration of product and process factors in a tolerance synthesis (Huang, Lin, Bezdecny, Kong and Ceglarek, 2007; Huang, Lin, Kong and Ceglarek, 2007; Loose *et al.*, 2007). Efforts have also been made to model compliant assemblies (Liu *et al.*, 1995; Chang and Gossard, 1997; Rong *et al.*, 2000; Shiu *et al.*, 2003; Camelio *et al.*, 2003). Elastic deformation and locked-in stresses created by closing the gaps between parts were analyzed using mechanics principles and finite element analysis. Shiu *et al.* (2003) introduced a beam compliant assembly model for tolerance allocation. For complex assemblies the quality characteristics are usually composed of multiple responses, e.g., an automotive body uses hundreds of measurement points to ensure dimensional integrity. Therefore, these methods, when applied to industrial problems, will result in high-dimensional MIMO models.

The complexity and high dimension of these models pose tremendous challenges in optimization modeling and algorithms in tolerance synthesis in both manufacturing and design.

2.2. *Optimization models and algorithms*

Tolerance synthesis in manufacturing and design is primarily an optimization problem which involves the following strategies.

1. Optimization formulation: deterministic models or probabilistic optimization models.
2. Optimization algorithms.

One of the key issues in problem formulation is how to define a constraint model. The primary goals in a tolerance synthesis are to achieve: (i) *a minimum cost*; (ii) *a specified capability* of producing quality product, which is represented by process capability (*e.g., yield* (Lee and Woo, 1989)) for given specifications. The constraint function should represent this capability requirement. The yield was introduced in tolerance synthesis (Lee and Woo, 1989):

$$yield = \text{Prob}\,(\cap\{y_i \in [-T_{si}, T_{si}]\}),\ i = 1, 2, \ldots, n.$$

where y_i and T_{si} are ith quality and its specification limit, respectively.

In more general settings with an irregular tolerance region (Spence and Soin, 1997; Kotz and Johnson, 2002):

$$yield = \text{Prob}\,(\cap\{y_i \in R_n\}).$$

For a univariate normal KPC, yield is easy to assess and can be directly translated into a constraint on the tolerance or variance. In multivariate cases, for given KPC distributions and tolerance region $\Re_n$ the yield can be exclusively assessed by numerical sampling methods, e.g., Monte Carlo. Although the algorithm is straightforward the computation can be extremely intensive, which motivated the simplifications and led to various deterministic models, for example the reliability index model (Lee and Woo, 1989) and multivariate capability index models (Kotz and Johnson, 2002).

The deterministic constraint model was formulated by Shi *et al.* (2003) and Ding *et al.* (2005). Using a linear model approximation, the root sum of squares (*RSS*) and *worst case* represent two simplified deterministic and widely used constraints to a design response:

$$RSS: \quad \sum_{i=1}^{n} \left|\frac{\partial f}{\partial x_i}\right|^2 t_{id}^2 \le T_f^2;$$
$$worst\ case: \sum_{i=1}^{n} \left|\frac{\partial f}{\partial x_i}\right| t_{id} \le T_f, \qquad (1)$$

where f, x_i, t_{id} and T_f are the assembly function, ith component dimension, x_i's tolerance and the tolerance of the design response (KPC), respectively. The implication of Equation (1), under normality assumption, is yield = 99.73% for a 3σ tolerance on each individual KPC with the hope of achieving the same total yield. When using *RSS* in Equation (1) for a multivariate case, it is equivalent to the Bonferroni method and may result in a much lower yield than expected. For example, suppose there are 100 KPCs and constraints are set individually on the variance of KPCs as $T_{si} \ge 3\sigma_i$, $i = 1, \ldots, 100$, if all KPCs are independent, then what we eventually achieved is

$$yield = \text{Prob}\,(\cap\{y_i \in [-T_{si}, T_{si}]\}) = \Pi p_i = 0.9973^{100} = 0.7631.$$

It is possible to achieve a yield of 0.9973 by setting more conservative tolerances, e.g., one needs to set an unnecessarily conservative individual $p_i = 0.999\,973$ or 27 ppm or equivalently $T_{si} \ge 4$–$5\,\sigma_i$ for each KPC. This is actually the Bonferroni method which results in conservative results in high-dimension MIMO problems.

Another deterministic model is

$$\sigma_s^2 - ||\text{diag}\,(\mathbf{\Sigma}_Y)||_k \ge 0, \qquad (2)$$

where $\mathbf{\Sigma}_Y$ and $\boldsymbol{\sigma}_s^2$ are the covariance matrix of multiple responses, and a specified value to ensure conformity to the specified tolerances, respectively; $||\cdot||_k$ represents the k-norm $k = 1, 2$ or ∞. There seems to be no available method for determining σ_s^2 in Equation (2) to ensure a specified process capability. It can be a challenging topic *per se* to have a one-to-one correspondence between σ_s^2 and a specified yield. Additional limitations of these models include the assumptions on distributions and simplified tolerance region.

Other alternatives include specifying Prob$[(\mathbf{X} - \mu)'\mathbf{\Sigma}^{-1}(\mathbf{X} - \mu) \le U]$, and using normality and χ^2 distribution (Kotz and Johnson, 2002; Shiu *et al.*, 2003). It is equivalent to the reliability index model in Lee and Woo (1989). In Shiu *et al.* (2003) a nice two-dimensional interpretation of this model was given using a distribution ellipsoid and a critical point on the tolerance region. Kotz and Johnson (2002) reviewed similar methods for defining multivariate process capability indices. However, this model also demands normality and relatively simple tolerance regions e.g., symmetric tolerances and hyper-rectangles.

Applicability and generality of these simplified models are limited due to the: (i) ambiguity of yield related to the assigned tolerance limits in Equations (1) and (2); and (ii) the assumptions on normality and tolerances. Non-normality and irregular tolerance regions are common in industry. Examples include circuit design (Spence and Soin, 1997), gap control in automotive body assembly (e.g., to prevent a V-gap one needs $|y_i - y_j| \le \Delta$ type of constraints) and composite tolerance on KPCs etc. The normal approximation cannot ensure the accurate tail property of a distribution which has a significant impact in capability analysis (in terms of ppm). Despite the computational complexity, the yield model is appealing in its generality, flexibility and unambiguity in terms of process capability. Therefore, it is very desirable in applications.

Tremendous efforts have also been devoted to highly efficient optimization algorithms and these efforts were motivated by the complexity of the models in optimization for real-world problems. Early attempts can be traced back to Speckhart (1972) and Ostwald and Huang (1977). Various optimization techniques have been developed along with tolerance allocation in manufacturing. Mathematic programming was introduced, such as, 0–1 integer programming, exhaustive search, univariate search and sequential quadratic programming (Ostwald and Huang, 1977; Ceglarek *et al.*, 2004). Lagrange-multiplier-based non-linear and geometric programming approaches were also developed (Lee and Woo, 1990; Lee *et al.*, 1993). Sampling algorithms represent another more recent group of efforts which can be found (Lee and Johnson, 1993; Kopardekar and Anand, 1995; Dupinet *et al.*, 1996; Li *et al.*, 2000; Zhou *et al.*, 2001; Huang *et al.*, 2004; Prabhaharam *et al.*, 2004). These methods allow near-optimal solutions within a reasonable computational time for problems with a small or moderate complexity. Dupinet *et al.* (1996) proposed the use of fuzzy logic and simulated annealing method to obtain global optima in tolerance synthesis. The variations in dimension chains were used in the model to represent KPCs of the assembly. Lee and Johnson (1993) applied Genetic Algorithm (GA) and Monte Carlo (MC) to non-linear tolerance synthesis which was formulated as a probabilistic optimization problem. MC simulation and a reliability-index-based integration technique were used for yield assessment. In a similar vein Li *et al.* (2000) and Prabhaharam *et al.* (2004) represent recent application of GA in tolerance synthesis. Zhou *et al.* (2001) and Huang *et al.* (2004) introduced a quasi-MC algorithm to improve the computation efficiency. Kopardekar and Anand (1995) adopted a neural network method in tolerance design.

These sampling-based algorithms may encounter tremendous computational difficulties in complex tolerance synthesis problems and tend to be impractical.

Since the mid-1990s efforts have also been made to improve the computation by using quality engineering techniques. Kapur (1993) was one of the first attempts to apply analysis of variance and design of experiments techniques in concurrent parameter and tolerance design problems. In a similar vein, Gadallah and El Maraghy (1994) and Kusiak and Feng (1996) reported similar works in tolerance research or robust parameter design. These works were summarized in Hong and Chang (2002). More recently, Jordaan and Ungerer (2002) proposed a Response Surface (RS) for yield approximation which allows for optimization with Sequential Quadratic Programming (SQP). The second-order model without the interaction was used in the RS model. With these techniques, the tolerance design synthesis may still be intractable for complex problems, (e.g., the number of a two-level full factorial experiment is 2^n that can be enormous when n is large, say, >30). These techniques tend to be more appropriate to the problems with lower complexity (e.g., <20 parameters).

Recent advances in multi-station assembly systems modeling have posed a tremendous challenge to the current tolerance techniques (Jin and Shi, 1999; Ding *et al.*, 2002; Shi, 2006; Huang, Lin, Bezdecny, Kong, Ceglarek, 2007; Huang, Lin, Kong, and Ceglarek, 2007). It is common in industries such as automotive and aerospace that a highly complex assembly model and tolerance requirements are involved. It is, therefore, very desirable to accommodate such complex situations and develop highly efficient techniques for tolerance design.

3. Three-dimensional multi-station SOVA model in multi-station manufacturing systems

A brief review of the recent advances in three-dimensional SOVA modeling techniques is given below. This provides a basis for subsequent problem formulation and yield analysis.

Assembly Model: The stream of variation in an N-station process is modeled as (Ding *et al.*, 2000; Shi, 2006):

$$\mathbf{X}(i) = \mathbf{A}(i-1)\mathbf{X}(i-1) + \mathbf{B}(i)\mathbf{U}(i) + W(i) \quad \text{and}$$
$$\mathrm{Y}(i) = \mathbf{C}(i)\mathbf{X}(i) + \mathbf{V}(i),\ i = 1, 2, \ldots, N \tag{3}$$

where $\mathbf{X}(i) \in \mathrm{R}^{n_i \times 1}$ is the accumulated deviation; $\mathbf{U}(i) \in \mathrm{R}^{m(i)\times 1}$ for a three-dimensional model is the fixture/part errors at station i; $\mathrm{Y}(i) \in R^{q(i)\times 1}$ is the measurement at station i; superscripts n_t, $m(i)$, $q(i)$ are dimensions of the three vectors, respectively; $W(i)$, $V(i)$ are unmodeled noise. Details of the **A**, **B** and **C** matrices are given in Huang, Lin, Bezdecny, Kong and Ceglarek (2007) and Huang, Lin, Kong and Ceglarek (2007). Defining $\mathbf{\Phi}(N,i) = \mathbf{A}(N-1)\ldots\mathbf{A}(i), \gamma(i) = \mathbf{C}\mathbf{\Phi}(N,i)\,\mathbf{B}(i)$ and $\gamma(0) = \mathbf{C}\mathbf{\Phi}(N,0)$, Equation (3) gives:

$$\mathbf{Y} = \sum_{i=1}^{N} \gamma(i)U(i) + \gamma(0)X(0) + \varepsilon. \tag{4}$$

The first equation in Equation (3) can also be expressed as:

$$\begin{bmatrix} \mathbf{X}(1) \\ \mathbf{X}(2) \\ \mathbf{X}(3) \\ \vdots \\ \mathbf{X}(N) \end{bmatrix} = \begin{bmatrix} \mathbf{B}(1) & 0 & 0 & 0 & 0 \\ \mathbf{A}(1)\mathbf{B}(1) & \mathbf{B}(2) & 0 & 0 & 0 \\ \mathbf{A}(2)\mathbf{A}(1)\mathbf{B}(1) & \mathbf{A}(2)\mathbf{B}(2) & \mathbf{B}(3) & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \mathbf{A}(N-1)\ldots\mathbf{A}(1)\mathbf{B}(1) & \mathbf{A}(N-1)\ldots\mathbf{A}(2)\mathbf{B}(2) & \ldots & \ldots & \mathbf{B}(N) \end{bmatrix} \begin{bmatrix} \mathbf{U}(1) \\ \mathbf{U}(2) \\ \vdots \\ \vdots \\ \mathbf{U}(N) \end{bmatrix}. \tag{5}$$

Automatic SOVA Model Generation: Equations (3) and (5) can be automatically generated; and thus the applicability of SOVA in complex systems can be greatly enhanced. The idea is to use the numerical simulation to transform a non-linear assembly model into an explicit SOVA model. The coefficients of the first-order Taylor expansion of an assembly model can be generated by any simulation-based off-the-shelf software for variation analysis such as 3DCS. GeoFactor in 3DCS represents the linear relationship between input and output. To generate the SOVA model, we firstly build a 3DCS model. At each station, the input tolerances and measurement points are defined and then GeoFactor is obtained and recorded station for each in matrices $\gamma_{ij}(j = 1 - i)$:

$$\begin{bmatrix} \mathbf{X}(1) \\ \mathbf{X}(2) \\ \vdots \\ \vdots \\ \mathbf{X}(N) \end{bmatrix} = \begin{bmatrix} \bar{\gamma}_{11} & 0 & 0 & 0 & 0 \\ \bar{\gamma}_{21} & \bar{\gamma}_{22} & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \bar{\gamma}_{N1} & \bar{\gamma}_{N2} & \cdots & \cdots & \bar{\gamma}_{NN} \end{bmatrix} \begin{bmatrix} \mathbf{U}(1) \\ \mathbf{U}(2) \\ \vdots \\ \vdots \\ \mathbf{U}(N) \end{bmatrix},$$

$$\bar{\gamma}_{11} = \begin{bmatrix} \gamma_{ij} \\ 0_{6(n_t - k_i)u_i} \end{bmatrix}, \tag{6}$$

where n_{t} is total number of parts after station N, k_i is total number of parts assembled from station 1 to i and u_i is the total number of inputs from station 1 up to station i. Using Equations (3) and (6) the design response function can be formulated as a linear SOVA model (Huang, Lin, Bezdecny, Kong and Ceglarek (2007) and Huang, Lin, Kong, and Ceglarek (2007)):

$$\mathbf{Y} = \mathbf{\Psi U}. \tag{7}$$

4. Problem formulation

The tolerance design synthesis is formulated as a probabilistic optimization problem in this section. It involves the definitions of both an objective function and constraints. The synthesis problem is defined either as minimizing cost in manufacturing subject to functional requirements on process capability or as maximizing the process capability with allowed cost.

The challenges in the tolerance synthesis are two-fold: cost and process capability. The former in manufacturing is affected by many factors and can be a challenging topic *per se* which is beyond the scope of this paper. The latter has been defined as the conformity probability (yield) in the literature and needs to be assessed by simulation-based methods, which poses a computational challenge for complex problems. The exponential cost function that is widely used in the literature is adopted in the following discussion. The tolerance synthesis with n design parameters t_i, $i = 1, \ldots, n$ and m design responses y_k, $k = 1, \ldots, m$ is formulated as

$$\min\left(\sum_{i=1}^{n} C_i(t_i) \right), \tag{8}$$

subject to

$$yield = \Pr\left\{ \bigcap_{k=1}^{m} (\mathrm{L}_k \le \mathrm{y}_k \le U_k) \right\} \ge P_{\text{threshold}} \quad t_i^{\mathrm{L}} \le t_i \le t_i^{\mathrm{U}}, \ t_j > 0, \tag{9}$$

where the cost function is $C_i(t_i) = A_i e^{-B_i(t_i)} + G_i$, A_i, B_i and G_i are model constants associated with the manufacturing cost for dimension i; L_k and U_k are lower and upper specification limits on response y_k; $P_{\text{threshold}}$ denotes a specified yield; t_i^{L}, t_i^{U} are process precision limits on tolerance t_i. The cost $C_i(t_i)$ is inversely affected by component tolerances that can be obtained by model fitting if the cost–tolerance data are available. Other types of cost functions can also be used (Wu *et al.*, 1988). The type of cost function does not affect the procedure and optimality because of their common monotonic property. The model coefficients give the flexibility for tolerance–cost data fitting or process–cost knowledge inclusion, e.g., different weights can be assigned to processes to represent different cost contributions. Under the normality assumption, the distribution of u_i in Equation (7) is

$$f_i(u_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} t((u_i - \mu_i)/\sigma_i)^2} \quad \text{and } \sigma_i = \frac{t_i}{3}, \ \ \mu_i = 0. \tag{10}$$

The yield is defined as.

$$Yield = \Pr\left\{ \bigcap_{k=1}^{m} (L_k \le y_k \le U_k) \right\} = \int_{L_m}^{U_m} \cdots \int_{L_1}^{U_1} q(y_1, y_2, \ldots, y_m) \varphi(y_1, y_2, \ldots, y_m) \mathrm{d}y_1 \mathrm{d}y_2 \ldots \mathrm{d}y_m, \tag{11}$$

where φ is a multivariate probability density function of m random KPCs variables $\{y_i\}$. $q(y_1, \ldots, y_m)$ is a test function which checks whether $\{y_i\}$ falls in the tolerance region. $q(y_1, \ldots, y_m) = 1$ if all of y_i in $\{y_i\}$ falls in the region simultaneously, otherwise $q = 0$. Yield, by definition, is a function of all KCC and KPC tolerances. Notice that Equation (11) is distribution independent. Thus, for a non-normal case, the yield analysis shares the same format and procedure. The differences are only in design response models and random number generation engines.

For given tolerance design and specifications, yield can be assessed numerically. Figure 1 illustrates the yield estimation through sampling in a two-dimensional case. In a typical industrial case study (Huang, Lin, Kong and Ceglarek, 2007) the dimension of the assembly model $\mathbf{\Psi}$ is about 390×1200, for a single yield analysis with 20000 samples and the total number of algebraic operations (multiplications and additions after the random number generation) is in the order of 10^{10} with an additional 10^7 logic operations. Since Equation (9) is a numerical model highly efficient algorithms such as SQP are not valid. Sampling algorithms

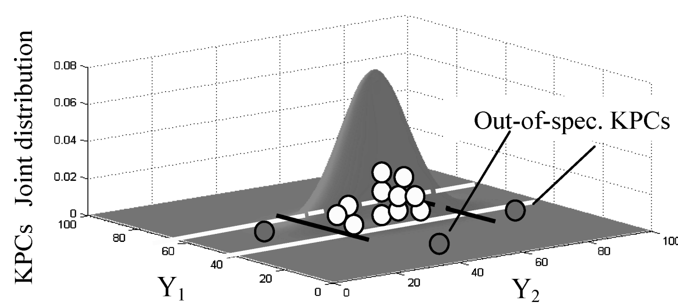

Fig. 1. In-spec. and out-of-spec. KPCs points.

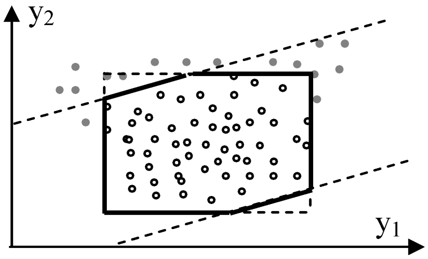

Fig. 2. An irregular tolerance region ($|y_1 - y_2| \leq \Delta$ and $L_i \leq y_i \leq U_i$) and a non-normal process.

such as MC, GA etc. become the only options and large numbers of yield assessments are necessary to ensure an optimum. Once a given number of samples $\mathbf{X}_k = \{x_1, \ldots, x_i, \ldots, x_n\}_k$, $k = 1, \ldots, N$ are generated the respective KPCs $\mathbf{Y}_k = \{y_1, \ldots, y_m\}_k$ can be obtained from Equation (9). For simple symmetric tolerances or general tolerance regions, we define the conformity of $\mathbf{Y}_k$ as

$$\mathbf{Y}_k \in \mathcal{R}_{\mathrm{m}}, \mathcal{R}_{\mathrm{m}} \text{ is defined as } \bigcap_{k=1}^{m} (L_k \leq y_k \leq U_k)$$
$$\text{or } \bigcap_{k=1}^{m} (L_k(y) \leq y_k \leq U_k(y)), \quad (12)$$

where L_k, U_k, $L_k(y)$ and $U_k(y)$ are boundaries of a hyper-rectangle and a general tolerance region, respectively. The conformity check involves only logic operations as shown in Equation (12). The yield can thus be approximated as

$$\text{Yield} = \frac{\textit{Number of conforming samples}}{\textit{Total number of samples}}. \quad (13)$$

Figure 2 illustrates a case with a composite tolerance $|y_1 - y_2| \leq \Delta$ and usual tolerance boundaries $L_i \leq y_i \leq U_i$. The combined effect is to give an irregular tolerance region. This is a typical case in gap or flushness control in automotive body manufacturing. A more complicated tolerance region can be found in Spence and Soin (1997). The case in Fig. 2 can also be treated as a non-normal problem if one defines a new KPC as $y' = |y_1 - y_2|$. In general, if the *function-oriented* quality characteristics are represented by general functions of variables $\{y_i\}$: $\mathrm{Z} = g(\mathrm{y}_1, \ldots, \mathrm{y}_m)$, $\mathbf{Z}$ may have a non-normal distribution. Or regular tolerances on $\mathbf{Z}$ will give an irregular tolerance region on the KPCs $\{y_i\}$. For example, in automotive assemblies the functional quality characteristics, such as wind noise, water leakage and door closing problems etc. can be expressed as complex functions of the dimensional KPCs by using experimental or numerical methods.

Equations (8), (9) and (12) provide a very general framework for tolerance synthesis formulation. It may open up new research opportunities in *function-oriented process capability and tolerancing* beyond *current dimensional tolerancing.*

5. Proposed methodology

Direct optimization in synthesis involves two sampling processes, one is sampling the design space for an optimum and the other is a sampling on each trial design point for yield calculation and constraint check. The basic idea behind the proposed method is to reduce the sample size in the first sampling. The method is to use a computer experiment and designed strategy for the first sampling, calculate yields on the samples, and then fit a surrogate yield model for optimization. The key is to use an SM model and gradient-based algorithm to avoid the large sample yield calculations in direct optimization. Once an SM model, e.g., *yield* $= \hat{y} = b_0 + \sum b_i t_i + \sum b_{ij} t_i t_j$ is established, Equation (9) becomes an analytic function and many efficient gradient-based algorithms become available to use for its solution.

To facilitate model fitting processes, as an initial step a partitioning strategy is developed for design space compression which will be detailed in Section 5.3. It helps to identify the candidate design space inside the initial design space within which qualified design candidate points are most likely to reside. Then, a three-step procedure is developed for yield model approximation: (i) sampling (in Section 5.1); (ii) multivariate distribution transformation (MDT) for design response simulation (in Section 5.2); and (iii) regression analysis for model fitting and variable screening which is presented step by step together with the partitioning strategy in Section 5.3. Finally, the fitted yield model is incorporated with a gradient-based optimization technique such as SQP for optimization which is presented in Section 5.4.

5.1. *Sampling strategy*

The first step in the SM model fitting procedure is to sample the design space for yield calculations. We denote a design space with n design variables as C^n (n-dimensional half space) below. Each sample point for a tolerance design represents a computer experiment design point and the corresponding yield represents an outcome of the experiment. Selection of an appropriate sampling strategy is key for highly efficient computation. To reduce computation intensity a quasi-MC approach Number-theoretical net method (NT-net), is introduced.

Space-filling approaches are popular in computer experiments such as Latin hypercube sampling, random sampling and uniform design etc. (Santner et al., 2002). To obtain complete information on an unknown design function with a limited sample size one always attempts to spread out design points as uniformly as possible. One criterion for this is the so-called discrepancy (D_n) which is a measure of the uniformity of the scattered sampling points in design space C^n for different sampling strategies (Fig. 3) (Fang and wang, 1996; Santner *et al.*, 2002). We introduce below a *good point set* (*gps*) (with lower D_n compared with MC (Fang and wang, 1996)), because of: (i) simplicity; (ii) a *gps*; and (iii) sample size independency, i.e., we can augment a sample by

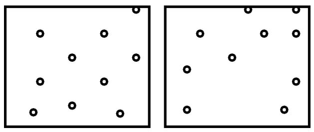
(a) (b)

Fig. 3. Discrepancies $D_a < D_b$.

adding extra points without redesigning the whole sample points set (and avoid yield recalculation). This gives flexibility in the model fitting process when more extra sample points are needed to improve model quality.

In Huang *et al.* (2004) a NT-net with $n \leq 18$ was introduced for tolerance analysis. For $n > 18$, there are some alternative *gps* such as, Circle Division Method (CDM), square root sequence *gps* etc. (Fang and Wang, 1996). We introduce CDM below for sampling. Given a sample size N the CDM is expressed as

$$\gamma = (\gamma_1, \ldots, \gamma_n) = \left(\left\{2\cos\frac{2\pi}{p}\right\}, \left\{2\cos\frac{4\pi}{p}\right\}, \ldots, \left\{2\cos\frac{2n\pi}{p}\right\}\right),$$

$$\mathbf{X}_k = (\{k\gamma_1\}, \ldots \{k\gamma_n\}) = \left(\left\{2k\cos\frac{2\pi}{p}\right\}, \left\{2k\cos\frac{4\pi}{p}\right\}, \ldots, \left\{2k\cos\frac{2n\pi}{p}\right\}\right), \quad \mathbf{X}_k \in C^n, \tag{14}$$

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix} = \begin{bmatrix} \left\{2\cos\frac{2\pi}{p}\right\} & \left\{2\cos\frac{4\pi}{p}\right\} & \cdots & \left\{2\cos\frac{2n\pi}{p}\right\} \\ \left\{4\cos\frac{2\pi}{p}\right\} & \left\{4\cos\frac{4\pi}{p}\right\} & \cdots & \left\{4\cos\frac{2n\pi}{p}\right\} \\ \vdots & \vdots & \vdots & \vdots \\ \left\{2N\cos\frac{2\pi}{p}\right\} & \left\{2N\cos\frac{4\pi}{p}\right\} & \cdots & \left\{2N\cos\frac{2n\pi}{p}\right\} \end{bmatrix}, \tag{15}$$

where bracket {.} denotes an operator, $\{W\}$ means taking the fixed point part of W (e.g., $\{1.23\} = 0.23$), and p is a selected arbitrary prime number satisfying $p > 2n + 3$. $\gamma = (\gamma_1, \ldots \gamma_n)$ is called a *gps* and designed as a tool to generate *gps* $\mathbf{X}$ in the literature. We also suggest using the alternative of the square root sequence *gps* (Feng and Wang, 1996) for sampling because it shares the similar performance and simplicity.

Equation (15) shows that new points $(X_{N+1}, X_{N+2}, \ldots)$ can be directly generated without affecting the results $(X_1, \ldots X_N)$ that have been obtained. For example, if the initial sample size N needs to be augmented to $N + k$, the extra sample points can be generated by using $N + 1, \ldots, N + k$ in Equation (15) with the initially generated points $X_1, \ldots, X_N$ being unaffected and then obtain the corresponding yields. A simple scale transformation can be used to project $\mathbf{X}_k$ into any tolerance design space. Once the sampling point set is generated, the respective design responses $\mathbf{Y}_k = \{y_1, \ldots, y_m\}_k$ can be correspondingly obtained from Equation (9) for yield calculation.

5.2. *MDT*

MDT is a valid method for multivariate normal distributions. The purpose is to generate random samples in compressed dimensions. Each sample point generated in a design space represents a trial tolerance design, which allows the statistical models of design variables to be determined from Equation (10). Thus, Equations (9) and (10) provide tools for simulating design responses $\mathbf{Y}_k = y_1, \ldots, y_{mk}$. There are two ways for simulating the design responses: (i) directly using the statistical model in Equation (10) to generate samples of $\mathbf{X}$ by the MC method and then using Equation (9) to calculate design responses $\mathbf{Y}_k = y_1, \ldots, y_{mk}$; and (ii) indirectly simulating design responses by using the proposed MDT. The latter approach is adopted because of its higher computation efficiency.

In the assembly model of Equation (10) it is common that the number of design variables is greater than that of responses i.e., $n \gg m$. This implies, when method (i) is adopted, that most of the effort will be wasted in repeatedly calculating design responses through Equation (9). A transformation is introduced below to reduce the computational effort. Under a normality assumption it is easy to show from the central limit theorem that $y = (y_1, y_2, \ldots\ldots, y_m)$ is normal and can be represented by the following multivariate distribution:

$$f(y) = \frac{1}{(2\pi)^{m/2}|\boldsymbol{\Sigma}_y|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}_y^{-1}(\mathbf{y} - \boldsymbol{\mu})\right], \tag{16}$$

which can be denoted by $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_y)$. Here $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_m)$ is mean vector, and $\boldsymbol{\Sigma}_y$ is the covariance matrix of design responses $y = (y_1, y_2, \ldots\ldots, y_m)$. $|\boldsymbol{\Sigma}_y|$ is the determinant of $\boldsymbol{\Sigma}_y$.

Because $\boldsymbol{\Sigma}_y = [\sigma_{ij}]$ is positive and symmetric, there exists a unique lower triangular matrix $\mathbf{C} = [c_{ij}]$, $c_{ij} = 0$ if $j > i$ and $\boldsymbol{\Sigma}_y = \mathbf{CC}^{\mathrm{T}}$. Then vector $\mathbf{y}$ can be represented as $y = \mathbf{Cz} + \boldsymbol{\mu}$, where $\mathbf{z} = (z_1, z_2, \ldots, z_m)$ has a standard multivariate normal distribution $N(\mathbf{0}, \mathbf{I})$. In order to obtain $\mathbf{C}$ from $\boldsymbol{\Sigma}_y = \mathbf{CC}^{\mathrm{T}}$ the so-called "square root method" can be used. The c_{ij} in $\mathbf{C}$ can be obtained from $\boldsymbol{\Sigma}_y = \mathbf{CC}^{\mathrm{T}}$ and the lower triangular property of $\mathbf{C}$:

$$c_{ij} = \frac{\sigma_{ij} - \sum_{k=1}^{j-1} c_{ik}c_{jk}}{\sqrt{\sigma_{jj} - \sum_{k=1}^{j-1} c_{jk}^2}}, \quad 1 \leq j \leq i \leq m. \tag{17}$$

Denote $\boldsymbol{\Sigma}_{\mathrm{x}}$ as the covariance matrix of design variables. From independency assumption of KCCs, $\boldsymbol{\Sigma}_{\mathrm{x}} = \mathrm{diag}\,\{\sigma_{\mathrm{x}i}^2\}$ which can be obtained from Equation (10) and $\boldsymbol{\Sigma}_y = \mathbf{A}\,\Sigma_{\mathrm{x}}\mathbf{A}^{\mathrm{T}}$ (Equation (9)).

Once $\mathbf{C}$ is obtained the simulation can be directly conducted using $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I})$ which provides a more efficient procedure for yield analysis. In the case study presented in this section, the efficacy of the MDT method is proved. It shows about a 99% improvement in computation time (4.093 versus 339.296 s per yield with a sample size of 20 000 in an industrial case study).

5.3. *Regression analysis and SM model fitting*

The procedure of fitting the yield model consists of two major steps:

1. Partitioning the design space and searching for a candidate subspace.
2. Sampling the candidate subspace and model fitting.

Based on the independency of KCCs the tolerance region for components is a hyper-rectangle. For convenience we define a feasible point in the design space as the point that satisfies the constraint in Equation (9). In the design space only a subspace that contains feasible points is worth considering. Since the yield is monotonically inversely affected by component tolerances the maximum subspace with the upper-right corner point satisfying this condition is further defined as a candidate space. Thus, the model fitting in the candidate space is more accurate and thus less sampling points are required. A procedure is proposed for identifying the candidate space and fitting the yield in it.

In the literature on computer experiments a variety of metamodels are available, such as Kriging, polynomials, splines, neural network etc. (Santner *et al.*, 2002). It is easy to show the monotonicity of the yield to tolerances, so the yield has a simple surface property. For simplicity we adopt the most popular second-order polynomials for yield SM model fitting; it is a special case of Kriging when the sample point's variations are independent and identically distributed (iid) (it is easy to show that the yields on different tolerance design points are iid and approximately normal). In the compressed candidate space a simple polynomial model can serve as a good approximation for the local model fitting.

The procedure starts from the initial design space or the feasible region for tolerances design. It is initially partitioned into sequentially embedded subspaces. Trial samples are then generated, by using the *gps* in Section 5.1, in these spaces sequentially. Respective yields are calculated to identify the candidate space. Once the candidate space is identified, more sample points are generated in it by the *gps* for yield calculation and model fitting. The selection of the sample size in the candidate space needs to be balanced between the model accuracy requirements and computational effort required. Suggested sample size is at least $>2n$ which can ensure fitting a model with linear and second-order terms (without interaction terms). Inclusion of interaction terms in the fitted yield model is possible when needed by adding new extra sample points for extra yield calculations. Finally, the goodness of fit is checked. This process may be repeated to fine tune the model when a higher model accuracy is required. The step-by-step procedure is presented below.

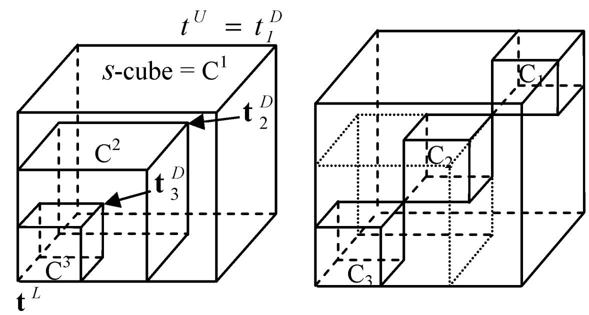

Fig. 4. Nested sub-cubes and diagonal cubes.

Step 1. Design space definition.
Tolerances that designers can assign to the corresponding components are selected in the design space which is bounded by allowable upper/lower limits. The minimum allowable tolerances depend on the achievable process precisions. The upper limits are the maximum allowable deviations. A design space with *n* KCCs forms a hyper-rectangular cube (s-cube). An s-cube has two unique vertices, denoted as *D-Vertices (DVs)*. The tolerances at *DVs of the s-cube* are denoted as $\mathbf{t}^L$ and $\mathbf{t}^U$ as shown in Fig. 4.

Step 2. Design space partitioning.
The s-cube is partitioned into k (=3–4) nested sub-cubes as shown in Fig. 4. The partition is conducted by dividing the diagonal line of the s-cube into k sections whereby k nodes are generated. All the nested cubes share the common vertex of $\mathbf{t}^L$ which is the vector of the smallest KCC tolerances in the s-cube. The other *DV* of each nested cube is one of the k nodes. The steps for generating the k nodes are as follows:

1. Define the range from $\mathbf{t}^L$ to maximum boundary $\mathbf{t}^U$ for n KCC tolerances in the s-cube.
2. Calculate the range of each KCC variable ($d = \mathbf{t}^U - \mathbf{t}^L$), and divide d into k sections, whereby a node at the end of each section is denoted as $\mathbf{t}_j^D$; $\mathbf{t}_j^D = \mathbf{t}^U - d(j-1)/k$, $j = 1, \ldots, k$. These nodes are generated between $\mathbf{t}^L$ and $\mathbf{t}^U$. The first node $\mathbf{t}_1^D$ is coincident with $\mathbf{t}^U$ which is the maximum allowable tolerance in the design.
3. The nested cubes are defined as sub-cubes with $\mathbf{t}^L$ and $\mathbf{t}_j^D$ as *DV*s. These cubes are shown as C^1, $C^2, \ldots, C^k$ in Fig. 4.

Step 3. Candidate space identification.
Diagonal cubes are created and denoted as C_1, $C_2, \ldots, C_k$ (Fig. 4). Each diagonal cube has *DV*s of $\mathbf{t}_j^D$ and $\mathbf{t}_{j+1}^D$. These diagonal cubes are used to

identify the candidate space. The steps for selecting the interesting nested cubes are as follows:

3.1 Set $i = 1$.
3.2 Generate n_0 tolerance vectors (NT-net points); e.g., $n_0 = 30$ in this paper, in the cube C_i.
3.3 Calculate yields on all n_0 sample points in C_i (MDT-MC simulation) and check the maximum yield.

- If maximum yield $\leq$ Threshold yield, go to Step 3.2 with $i = i + 1$.
- If maximum yield $\geq$ Threshold yield, select C^i as the candidate space and go to Step 4.

Step 4. Sampling in the candidate space.
CDM *gps* is used to generate n_1 samples with $n_1 \geq 2n + 1$ C^i.

Step 5. Yield analysis and model fitting.
Yields are calculated using MDT and Equation (12) in cube C^i for model fitting. A linear model is initially selected for yield function approximation. If the first-order linear model is not sufficient to predict an observed response as indicated by the goodness-of-fit test, the quadratic and interaction terms that include only significant factors will be added to update the model as described below in Steps 6 and 7. The initial model is

$$\hat{y} = b_0 + \sum_{i=1}^{n} b_i t_i. \tag{18}$$

The coefficient is calculated by $\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$, where $\mathbf{b}$ is a coefficient vector. $\mathbf{y}$ is a yield vector: $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n1})^T$, and $\mathbf{X}$ is a design matrix consisting of $n + 1$ columns and n_1 rows of KCC samples.

Step 6. Variable screening.
p-values are used to indicate the significances of variables for screening.

Step 7. Model test and updating.
The adjusted R^2 (R^2_{adjusted}) value is used. If R^2_{adjusted} indicates an inadequate fit, the quadratic terms of significant KCC variables can be added into the yield model to improve the R^2_{adjusted} value. The above procedure will provide an approximate yield or SM model:

$$\hat{y} = b_0 + \sum b_i t_i + \sum b_{ij} t_i t_j. \tag{19}$$

We suggest including more terms in model (19) to improve the model predictability. The SM in Equation (19) will be integrated with a standard optimization algorithm for tolerance synthesis in Section 5.4.

Step 8. "Model bending."
Since the yield must be less than one, if predicted $\hat{y} > 1.0$ set yield $= 1$ (*bending*).

5.4. *Tolerance synthesis*

The SM provides a polynomial constraint function in optimization which can avoid a tedious direct yield assessment. The proposed procedure for tolerance synthesis is presented below.

1. Assign the largest tolerances in the feasible design space to all insignificant KCC variables.
2. Solve the optimization problem below by using gradient-based algorithms:

$$\min\left\{\sum C_i(t_i)\right\}$$

subject to

$$\hat{y} = b_0 + \sum b_i t_i + \sum b_{ij} t_i t_j \geq P_{\text{threshhold}}. \tag{20}$$

The feasible space is defined by $t_i^L \leq t_i \leq t_i^U$ where $i \in$ {significant KCC variables}. Widely used gradient-based algorithms such as SQP can be applied to Equation (20).

5.5. *Discussion*

Multivariate probabilistic optimization in a tolerance design (Equations (8) and (9)) requires tedious sampling with dense points for yield calculation. Given the complexity of the problem, no available method can ensure a global optimum which, we believe, is also unnecessary. It is more appropriate to balance the computation cost and accuracy. Instead of exhaustive sampling, e.g., MC that requires several thousands or more samples, the proposed method uses a *design space compression*, *gps sampling* and *SM-model-based optimization* strategy. It can expedite the process and achieve near-optimal solutions with affordable computations (e.g., dozens to hundreds of samples). An important implication of the SM model in Equation (20) is that it reveals the principal contributors (main effects) to process capability

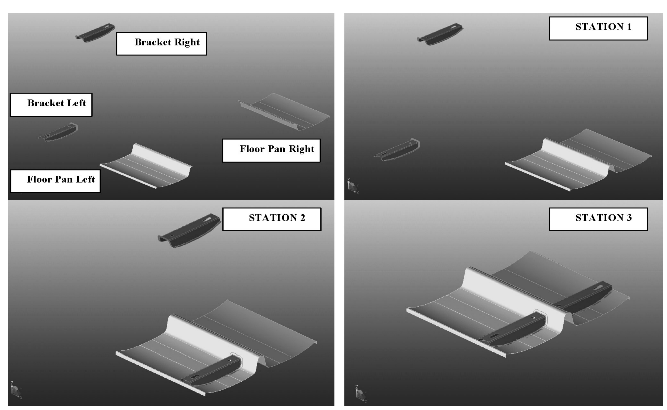

Fig. 5. A floor pan assembly.

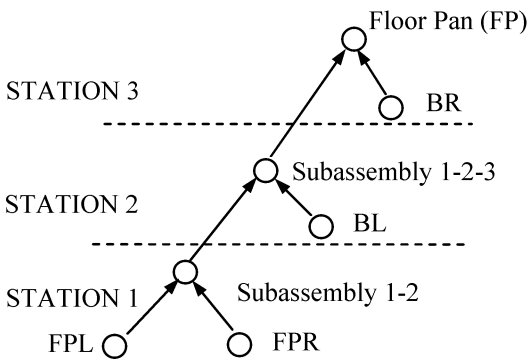

Fig. 6. Floor pan assembly tree.

6. Case study

6.1. *Case I*

A floor pan assembly (Fig. 5), which is a subassembly in an automotive underbody assembly, is used for validation. It consists of four parts: Floor Pan Left (FPL), Floor Pan Right (FPR), Bracket Left (BL) and Bracket Right (BR). They are assembled in three stations as illustrated in Figs. 5 and 6. A total of 12 measurements (KPCs) are defined for dimensional quality inspection. Cost functions are assumed to be exponential. In each station, there are 21 KCCs which can be categorized into three groups as follows.

1. The first nine KCCs are on fixtures. The first six KCCs constrain a root part and the other three KCCs constrain the mating plane. The feasible tolerance region of nine fixtures is assumed to be ±0.06–±0.15 mm. The cost function is $C(t) = 2 + 10/e^{2t}$.
2. The KCCs indexed from 13 to 15 and 19 to 21 are defined on linear mating features with three of them being located on a root part or a subassembly and the other three located on a mating part. The feasible region of the linear mating features is ±0.10–±0.25 mm. The cost function is $C(t) = 1{+}15/e^{t}$.
3. The KCCs indexed from 10 to 12 and 16 to 18 are for angular mating features with three of them being located on a root part or a subassembly and the other three located on a mating part. The feasible region of the angular mating features is ±0.10–±0.25°. The cost function is $C(t) = 3{+}20/e^{3t}$.

The summation of all cost functions is defined as the objective function. The SOVA model in Equation (9) is established using the techniques in Huang, Lin, Bezdecny, Kong and Ceglarek (2007) and Huang, Lin, Kong and Ceglarek (2007). Threshold yield is set at 98%. The KCC design space is a 63 dimensional cube; ($n = 63$). The tolerance synthesis can be formulated as

$$\min\left\{\sum_{i=1}^{63} C_{\mathrm{i}}(t_{\mathrm{i}})\right\}, \text{subject to yield} \geq 98\% \text{ and } t_i^{\mathrm{L}} \leq t_i \leq t_i^{\mathrm{U}}, \tag{21}$$

where t_i is the tolerance of the ith KCC variable, and t_i^{L} and t_i^{U} are its lower and upper tolerance bounds.

6.1.1. *SM-based tolerance synthesis*

Step 1. Define the feasible tolerance region for the 63 independent variables as mentioned above.

Steps 2 and 3. Partition the s-cube into $k = 3$ sub-cubes. The feasible tolerance space (s-cube) is divided into 3 sub-cubes; C^1, C^2 and C^3 as shown in Fig. 4. The diagonal cubes C_1, C_2 and C_3 are defined accordingly. The range of tolerances of sub-cubes C^1, C^2 and C^3, and diagonal cubes are given in Tables 1 and 2. Then, $n_0 = 30$ samples are generated by NT-net in diagonal cubes C_1, C_2 and C_3 sequentially. The respective (n_0) yields of the samples $t_1, t_2, \ldots, t_{30}$ are estimated by the MDT-MC technique; sample size $N = 20\,000$ is used for yield calculation. Equation (9) is used as the assembly response function for KPCs and yield simulation. The maximum and average yields of each diagonal cube are listed in Table 3; the desired yield ≥ 98% is seen to be more likely to fall in diagonal cube C_2. Therefore, sub-cube C^2 is selected as the candidate space for the yield SM model fitting.

Step 4. NT-net generation: The candidate space C^2 is defined in Table 4. Then, $n_1 = 150$ samples ($t_1, t_2, \ldots, t_{150}$) are generated in C^2 by

Table 1. Tolerance ranges of variables in sub-cubes (mm)

	C^1	C^2	C^3
Fixture	0.06–0.15	0.06–0.12	0.06–0.09
Mating feature	0.10–0.25	0.10–0.20	0.10–0.15
Angular MF (°)	0.10–0.25	0.10–0.20	0.10–0.15

Table 2. Tolerance ranges of variables in diagonal cubes

	C_1	C_2	C_3
Fixture	0.12–0.15	0.09–0.12	0.06–0.09
Mating feature	0.20–0.25	0.15–0.20	0.10–0.15
Angular MF (°)	0.20–0.25	0.15–0.20	0.10–0.15

Table 3. Maximum and average yield (in percent) in each diagonal cube

	C_1	C_2	C_3
Average	91.83	98.62	99.97
Maximum	93.13	99.17	100

Table 4. Tolerance ranges in the candidate design space

	C^2
Fixture (mm)	0.06–0.12
Mating feature (mm)	0.10–0.20
Angular MF (°)	0.10–0.20

NT-net. Each sample consists of tolerances of 63 KCCs.

Step 5. Yield associated with each t_i is calculated by using MDT-MC simulation on 20000 samples.

Step 6. Model fitting and variable screening are conducted.

Step 7. Model check and updating with R^2_{adjusted} improvement from 85% to 89.4%.

Step 8. The bended model is obtained:

If predicted yield of the model ≤ 1.00: yield $= 0.0467 - 0.0173t_2 - 0.0258t_3 - 0.207t_5 - 2.86t_{10} - 0.545t_{11} - 1.82t_{16} - 0.449t_{17} - 0.0521t_{19} - 0.0177t_{20} + 0.205t_{21} + 0.00655t_{35} - 0.297t_{43} - 0.0435t_{45} - 0.0082t_{46} + 0.021t_{62} + 1.84t_5t_{43} + 39.0t_{10}t_{19} - 46.5\mathrm{t}_{10}\mathrm{t}_{21} - 1.26\mathrm{t}_{19}\mathrm{t}_{21}$

If response > 1.00: yield $= 1.00$

The regression model can predict the simulated response accurately in the candidate design space with an error of less than 4%. Direct explanation of the main effects of the significant tolerances should be avoided because the interaction effects and modeling noise may complicate the interpretation.

6.1.2. *Tolerance synthesis/optimization*

Optimization is conducted using the SQP method. The results are listed in Table 5. The predicted yield at the optimum value is also checked with MC simulation (98.21%). The optimal solution consists of the tightened tolerances of 15 significant KCCs (most of them are at or close to their lower bounds) and 48 insignificant tolerances taking their upper bounds. The optimal tolerance design represents the balance of the cost of the yield requirement. Either cost functions or the accuracy of the SM model may affect the optimal solution. The cost function adopted in the case study is designed for illustration. The accuracy of the SM, on the other hand, can be improved through the increased sample size in both model fitting and yield calculation.

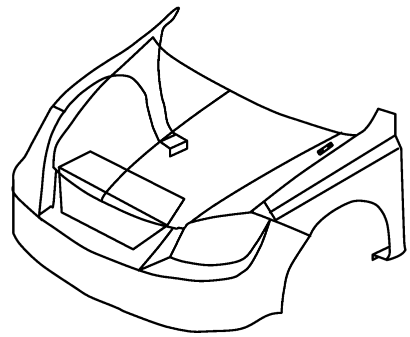

Fig. 7. Front-end assembly model.

6.2. *Case II*

An automobile front-end assembly model provided by one of the major domestic automobile companies is illustrated in Fig. 7. It consists of 215 variables (KCCs) including all the pins, locating holes and part–part mating surface dimensions. There are 61 measurement points (KPCs) to ensure the production functions. For illustration purposes, all the initial design tolerances were assigned as +/−1.00 mm. 3DCS Analyst was used for automatic SOVA model generation (Huang, Lin, Kong and Ceglarek, 2007). The specifications of all 61 measurements were assigned as +/−0.75 mm. The initial analysis showed a very low yield (≈0), because all the individual KPCs yield less than 0.8; about 60% of them less than 0.6, and the average is only 45%. It is obvious that, if all KPCs are independent with the individual yield equal to 0.8, the simultaneous conformity rate or the yield will be $(0.8)^{61} \approx 0$.

A yield SM model was created which includes 74 significant KCCs and a similar formulation and steps were followed for optimization. To keep a low cost, the rest of the 141 insignificant variables were set to their maximum allowable values (±1.00 mm). However, since the yield was very low we set the constraint for the yield to 0.45. Figure 8 compares the results before and after the optimization. In Table 6, the average yields increased by 80% with an average tolerance change from 100% to 72.2% of 215 KCCs.

6.3. *Comparative studies*

The floor pan case was used for a comparison study. The optimal tolerances obtained from the SM-based method were compared with current designs in industry and the optimization designs obtained by various algorithms in the literature. The MC-based global exhaustive search and GA method were used.

Table 5. The optimal tolerances of significant KCCs (cost = 899.84; * unit in rad)

KCC tolerance	t_2	t_3		t_5	$*t_{10}$	$*t_{11}$	$*t_{16}$	$*t_{17}$
Value (±mm)	0.105	0.105		0.105	0.00174	0.00174	0.00174	0.00174
KCC tolerance	t_{19}	t_{20}	t_{21}	t_{35}	t_{43}	t_{45}	t_{46}	t_{62}
Value (±mm)	0.105	0.105	0.105	0.175	0.105	0.105	0.105	0.105

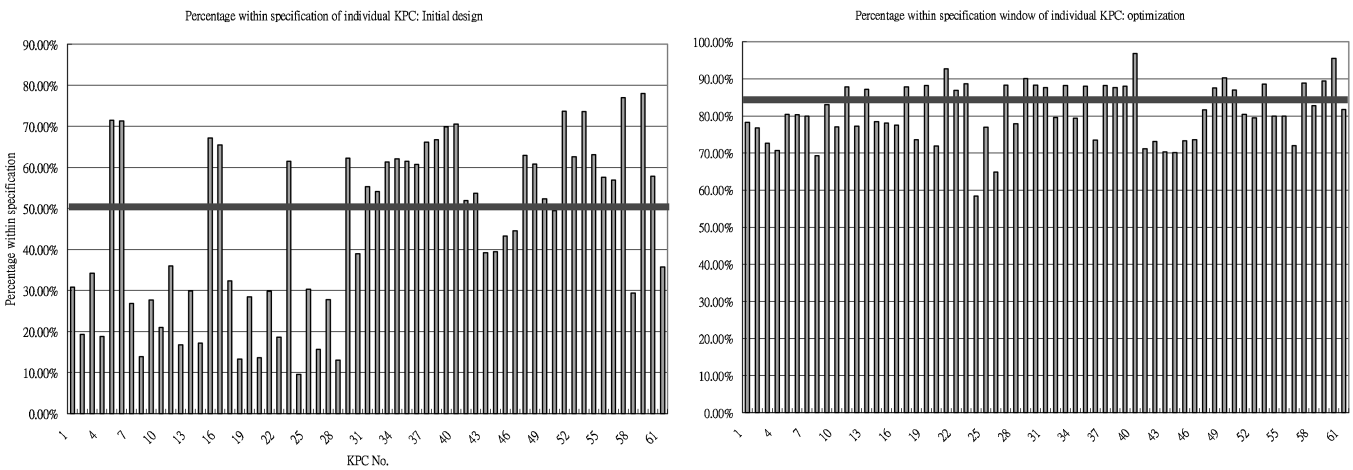

Fig. 8. Sixty one individual KPC's yields: initial design (left) and optimal design (right).

6.3.1. *Comparison with initial tolerance design*

The industrial tolerance design (initial design) of the floorpan assembly was provided by Dimensional Control Systems Inc. based on engineers' design experiences. The optimal tolerance design obtained by the SM-based method shows an average shrinkage from the current design tolerances by 6.55%. However, the yield is increased significantly by 12.02% from 86.19% to 98.21% as shown in Table 7.

6.3.2. *Comparison with MC simulation*

MC-based direct search is a straightforward approach for tolerance synthesis. Wherein large samples (trial design points) have to be generated and the yield on each of these points has to be assessed using the same MC technique. To avoid prohibitive computation in yield assessment, we tried the following strategy: 200 000 samples (trial design points) are initially generated; then the costs associated with these samples are calculated; to avoid tedious yield assessments, the yield check is designed to be done only for the points whose costs are less than or equal to 900 (the minimum in Table 5). However, among 200 000 samples, none is found satisfying this criterion. Clearly, MC is not as effective and efficient as the SM method in this tolerance synthesis problem.

6.3.3. *Comparison with GA method*

The same case study is used for comparison with the GA-based synthesis algorithm. Three different scenarios are considered. Computation times of SM and GA-based methods are listed in Table 8.

SM showed better performance over the GA method in computation time and optimal results in the case study, as shown in Tables 8 and 9. At the same level of yield (98%) optimal KCC tolerances obtained by the SM are bigger than that of the GA by 26.05% on average (Table 9). SM results in lower cost (899.84 versus 918).

Table 6. Comparison of SM versus initial design

	Initial design	*Optimization with significant KCCs*
Average change of tolerances (%)	100	72.20
Average KPC's individual yield (%)	45.15	81.03
Range of individual yield (min-max) (%)	9.50–77.96	58.39–96.84

Table 7. Comparison of SM versus initial design

	Yield (%)	*Shrinkage of tolerances (%)*
SM	98.21	−6.55
Initial design	86.19	0

Table 8. Comparison of SM and GA-based methods (Intel Pentium 4, CPU 2.80 GHz, 1GB RAM)

	SM	*GA*
Run time (minutes)	34	65 (1200 generations) 175 (1400 generations) 195 (1800 generations)

Table 9. Relative widths of tolerance windows (SM versus GA)

	Yield (%)	*Relative tolerance window width (%)*
SM	98.21	26.05
GA	98.37	0

7. Conclusions

The tolerance synthesis is defined as a probabilistic optimization problem with a general constraint on process capability. It provides a unified formulation for synthesis with generic multivariate statistical models and tolerance regions. Thus, it ensures the generality and applicability of the proposed model and methodology in application. A systematic yield SM-based method is developed for the tolerance synthesis. Case studies of the floor-pan and front-end assemblies in automotive body manufacturing with the MIMO SOVA model were conducted. Comparisons were made between an initial design from industry and the designs out of the developed SM as well as other optimization algorithms such as MC and GA algorithms. The proposed SM approach demonstrated advantages over these approaches in both computation efficiency and quality of the results.

The developed techniques can be applied in dimensional tolerance designs where assemblies are complex such as in the automotive, aerospace and shipbuilding industries. Examples of such assemblies include: automotive body, aircraft fuselage, aerospace structure, turbo-machinery etc. With recent advances in modeling the SOVA in multi-station manufacturing systems, the developed techniques also allow tolerance synthesis for product/process design integration. We envision that the proposed method can open up new research opportunities by shifting traditional *dimensional tolerancing* to *function-oriented tolerancing*. The process capability will be directly defined and analyzed on a product functional basis. The *function-oriented KPCs* (Section 4) are usually expressed as complex functions of *dimensional* KPCs, and thus may have complex distributions, irregular tolerance regions and similar computation challenges. The proposed unified formulation and SM method in design synthesis may play a critical role in this aspect.

Acknowledgements

The authors gratefully acknowledge the financial support of the Advanced Technology Program (ATP) award provided by National Institute of Standards and Technology (NIST) UK EPSRC Star Award EP/E044506/1 and US NSF-CAREER Award DMII-0239244. The authors would also like to thank the staff of Dimensional Control Systems, Inc. for their enthusiastic support and insightful discussions.

References

Camelio, J., Hu, S.J. and Ceglarek, D. (2003) Modeling Variation Propagation of Multi-station Assembly Systems with Compliant Parts, *Trans. of ASME, Journal of Mechanical Design*, **125**(4), 673–681.

Ceglarek, D., Huang, W., Zhou, S., Ding, Y., Ramesh, K. and Zhou, Y. (2004) Time-based competition in manufacturing: stream-of-variation analysis (SOVA) methodology—review. *International Journal of Flexible Manufacturing Systems*, **16**(1), 11–44.

Chang, M. and Gossard, D.C. (1997) Modeling the assembly of compliant, non-ideal parts. *Computer-Aided Design*, **29**(10), 701–708.

Chase, K.W., Greenwood, W.H., Loosli, B.G. and Hauglund, L.F. (1990) Least cost tolerance allocation for mechanical assemblies with automated process selection. *Manufacturing Review*, **3**, 49–59.

Chen, Y., Ding, Y., Jin, J. and Ceglarek, D. (2006) Integration of Process-Oriented Tolerancing and Maintenance Planning in Design of Multi-station Manufacturing Processes, *IEEE Trans. on Automation Science and Engineering*, **3**(4), 440–453.

Ding, Y., Ceglarek, D. and Shi, J. (2000) Modeling and Diagnosis of Multistage Manufacturing Process: Part I State Space Model, *Proceedings of 2000 Japan-USA Symposium of Flexible Automation*, July 23–26, Ann Arbor, MI, Paper No. 2000JUSFA-13146.

Ding, Y., Ceglarek, D. and Shi, J. (2002) Design evaluation of multi-station manufacturing processes by using state space approach. *Transactions of the ASME*, *Journal of Mechanical Design*, **124**(4), 408–418.

Ding, Y., Jin, J., Ceglarek, D. and Shi, J. (2005) Process-oriented tolerancing for multi-station assembly systems. *IIE Transactions*, **37**(6), 493–508.

Dupinet, E., Balazinski, M. and Czogala, E. (1996) Tolerance allocation based fuzzy logic and simulated annealing. *Journal of Intelligent Manufacturing*, **7**, 487–497.

Fang, K.T. and Wang Y. (1996) *Number-Theoretical Methods in Statistics*, Scientific Publishing Inc., Beijing, China.

Gadallah, M.H. and El Maraghy, H.A. (1994b) The tolerance optimization problem using a system experimental design in *Advances in Design Automation*, 69, ASME Publisher, New York, pp. 251–265.

Hong, Y.S. and Chang, T.C. (2002) A comprehensive review of tolerancing research. *International Journal of Production Research*, **40**(11), 2425–2459.

Huang, W., Ceglarek, D. and Zhou, Z. (2004) Tolerance analysis for design of multistage manufacturing processes using number-theoretical net method (NT-net). *International Journal of Flexible Manufacturing Systems*, **16**, 65–90.

Huang, W., Lin, J., Bezdecny, M.R., Kong, Z. and Ceglarek, D. (2007) Stream-of-variation (SOVA) modeling I: a generic 3D variation model for rigid body assembly in single station assembly processes. *Transactions of the ASME, Journal of Manufacturing Science & Engineering*, **129**(4), 821–831.

Huang, W., Lin, J., Kong, Z. and Ceglarek, D. (2007) Stream-of-variation (SOVA) modeling II: a generic 3D variation model for rigid body assembly in multi station assembly processes. *Transactions of the ASME, Journal of Manufacturing Science & Engineering*, **129**(4), 832–842.

Jin, J. and Shi, J. (1999) State space modeling of sheet metal assembly for dimensional control. *Transactions of the ASME*, *Journal of Manufacturing Science & Engineering*, **121**(4), 756–762.

Jordaan, J.P. and Ungerer, C.P. (2002) Optimization of design tolerances through response surface approximations.*Transactions of the ASME*, *Journal of Manufacturing Science & Engineering*, **124**, 762–767.

Kapur, K.C. (1993) Quality engineering and tolerance design, in *Concurrent Engineering: Automation, Tools, and Techniques*, Wiley, pp. 287–306.

Kopardekar, P. and Anand, S. (1995) Tolerance allocation using neural networks. *International Journal of Advanced Manufacturing Technology*, **10**, 269–276.

Kotz, S. and Johnson, N.L. (2002) Process capability indices—a review, 1992–2000. *Journal of Quality Technology*, **34**(1), 2–19.

Kusiak, A. and Feng, C.-X. (1996) Robust tolerance design for quality. *Transactions of the ASME, Journal of Engineering for Industry*, **118**, 166–169.

Lee, J. and Johnson, G.E. (1993) Optimal tolerance allotment using a genetic algorithm and truncated Monte Carlo simulation. *Computer-Aided Design*, **25**(9), 601–611.

Lee, W.J. and Woo, T.C. (1989) Optimal selection of discrete tolerances. *Transactions of the ASME, Journal of Mechanism, Transmission, & Automation in Design*, **111**, 243–251.

Lee, W.J. and Woo, T.C. (1990) Tolerances: their analysis and synthesis. *Transactions of the ASME*, *Journal of Engineering for Industry*, **112**, 113–121.

Lee, W.J., Woo, T.C. and Chou, S.Y. (1993) Tolerance synthesis for nonlinear systems based on nonlinear programming. *IIE Transactions*, **25**(1), 51–61.

Li, W., Bai, G., Zhang, C. and Wang, B. (2000) Optimization of machining datum selection and machining tolerance allocation with genetic algorithms. *International Journal of Production Research*, **38**, 1407–1424.

Liu, S.C., Lee, H.-W. and Hu, S.J. (1995) Variation simulation for deformable sheet metal assemblies using mechanistic models. *Transactions of NAMRI/SME*, **23**, 235–241.

Loose, J.-P., Zhou, S. and Ceglarek D. (2007) Kinematic Analysis of Dimensional Variation Propagation for Multistage Machining Processes with General Fixture Layouts, *IEEE Transactions on Automation Science and Engineering*, **4**(2), 141–152.

Ngoi, B.K.A. and Kuan, Y.C. (1995) Tolerance charting: the state-of-the-art review. *International Journal of Computer Applications in Technology*, **8**, 229–242.

Ngoi, B.K.A. and Ong, T.C. (1998) Product and process dimensioning and tolerancing techniques: a state-of-the-art review. *International Journal of Advanced Manufacturing Technology*, **14**, 910–917.

Ostwald, P.F. and Huang, J. (1977) A method for optimal tolerance selection. *Transactions of the ASME*, *Journal of Engineering for Industry*, **99**, 558–565.

Prabhaharam, G., Asokan, P., Ramesh, P. and Rajendran, S. (2004) Genetic-algorithm-based optimal tolerance allocation using least-cost-model. *International Journal of Advanced Manufacturing Technology*, **24**, 647–660.

Rong, Q., Ceglarek, D. and Shi, J. (2000) Dimensional Fault Diagnosis for Compliant Beam Structure Assemblies, *Trans. of ASME, Journal of Manufacturing Science and Engineering*, **122**(4), 773–780.

Roy, U. and Feng, Y.-C. (1997) Optimal tolerance re-allocation for the generative process sequence. *IIE Transactions*, **29**, 37–44.

Santner, T.J., Williams, B.J. and Notz, W.I. (2002) *The Design and Analysis of Computer Experiments*, Springer-Verlag, New York, NY.

Shi, J. (2006) *Stream of Variation Modeling and Analysis for Multistage Manufacturing Processes*, CRC Press: Boca Raton, FL.

Shiu, B.W., Apley, D.W., Ceglarek, D. and Shi, J. (2003) Tolerance allocation for compliant beam structure assemblies. *IIE Transactions*, **35**(4), 329–342.

Speckhart, F.H. (1972) Calculation of tolerance based on a minimum cost approach. *Transactions of the ASME*, *Journal of Engineering for Industry*, **94**, 447–453.

Spence, R. and Soin, R.S. (1997) *Tolerance Design Electronic Circuits*, Addison-Wesley.

Wu, Z., Elmaraghy, W.H. and Elmaraghy, H.A. (1988) Evaluation of cost-tolerance algorithms for design tolerance analysis and synthesis. *Manufacturing Review*, **1**, 168–179.

Zhang, C. and Wang, H.P. (1993a) Integrated tolerance optimization with simulated annealing. *International Journal of Advanced Manufacturing Technology*, **8**, 167–174.

Zhang, C. and Wang, H.P. (1993b) Optimal process sequence selection and manufacturing tolerance allocation. *Journal of Design & Manufacturing*, **3**, 135–146.

Zhou, Z., Huang, W. and Zhang, L. (2001) Sequential algorithm based on number theoretic method for statistical tolerance analysis and synthesis. *Transactions of the ASME, Journal of Manufacturing Science & Engineering*, **123**, 490–493.

Biographies

Wenzhen Huang is an Assistant Professor at the Mechanical Engineering Department, University of Massachusetts Dartmouth. He received his Ph.D. in Industrial Engineering from the University of Wisconsin–Madison in 2004. His current research includes multistage manufacturing process modeling and diagnostics, and stream of variation (SOVA) methodology for compliant/rigid assemblies. His research was supported by Advanced Technology Program (ATP/NIST) and in collaboration with Dimensional Control Systems (DCS) Inc. He has received Best Graduate Student Paper Awards by INFORMS Quality, Statistics and Reliability Section in 2002. His research interests include product and manufacturing process modeling, design analysis and optimization, and diagnosis for variation reduction and quality improvement, probabilistic design and optimization, metrology and sensor systems. The application area covers automotive, fuel cell, aerospace, shipbuilding, and general mechanical product design and manufacturing.

Tirawat Oak Phoomboplab received a B.Eng. in Mechanical Engineering from Chulalongkorn University, Bangkok, Thailand, in 2002; M.S. in Manufacturing Systems Engineering from the University of Wisconsin–Madison in 2006. He was a Project Engineer with Dow Chemical Company, the Bangkok, Thailand, from 2002 to 2003. Currently, he is a Graduate Research Assistant in the Manufacturing Systems Realization and Synthesis (MARS) Lab at UW-Madison, and Ph.D. Research Student in Digital Product Life-cycle Management at Warwick Manufacturing Group, University of Warwick, UK. His research interests include methodologies for the design synthesis in multi-stage assembly systems based on process-oriented design, and dimensional management for a new product development. This research is being performed in collaborations with General Motors and DCS Inc.

Dariusz (Darek) Ceglarek received his Ph.D. degree in Mechanical Engineering from the University of Michigan-Ann Arbor in 1994. He was on the research faculty at the University of Michigan-Ann Arbor from 1995 to 2000. In 2000, he accepted an appointment at the University of Wisconsin–Madison as Assistant Professor in the Department of Industrial and Systems Engineering where he rose to the ranks of Associate Professor and Professor in 2003 and 2005, respectively. Since 2006 he is Professor and EPSRC Research Chair at the University of Warwick, UK. Professor Ceglarek is a Fellow of CIRP (International Academy of Production Research), based in Paris, France. Professor Ceglarek served as Chair of the Quality, Statistics and Reliability Section of the Institute of Operations Research and Management Sciences (INFORMS); Program Chair for the American Society of Mechanical Engineers (ASME) Design-for-Manufacturing Life Cycle (DFMLC) Conferences, Associate Editor of the IEEE Trans. (TASE) and Associate Editor of the ASME Transactions, Journal of Manufacturing Science and Engineering. His research focuses on product life cycle modelling and analysis with emphasis on production and healthcare systems convertibility, scalability, and 6-sigma quality and root cause analysis (RCA) during design, manufacturing and service phases. He has developed a number of methodologies for 6-sigma variation reduction in manufacturing (SOVA methodology implemented by DCS, Inc.; "BIW Data Analyzer" implemented by GM;) and in service (FRL methodology for diagnosis of No-Defect-Found warranty faults implemented by Motorola; Remote diagnostics methodology for MR/CT field failures implemented by GE Healthcare). He has published widely in his research areas with more than 60 papers in leading journals. Two of his paper received Best Paper Awards by ASME Manufacturing Engineering Division and ASME Design Engineering Division. He has received numerous awards including in 2007 an UK EPSRC STAR Award (granted to an "exceptional senior faculty, recognised international leader in his/her research field"), US National Science Foundation (NSF) 2003 CAREER Award (NSF's "most prestigious awards in support of junior faculty who exemplify the role of teacher-scholars through outstanding research, excellent education and the integration of education and research."); 1999 Outstanding Research Scientist Award from the College of Engineering at University of Michigan; the 1998 Dell K. Allen Outstanding Young Manufacturing Engineer of the Year Award from the Society of Manufacturing Engineers (SME).