Robotica (1998) volume 16, pp. 251–264. Printed in the United Kingdom © 1998 Cambridge University Press

Statistical representation and computation of tolerance and clearance for assemblability evaluation

Sukhan Lee*† and Chunsik Yi*

SUMMARY

The design of tolerances should take the functionality, manufacturability, assemblability and cost effectiveness of a product into consideration. This paper presents a method of analyzing and evaluating the assemblability of a product under given tolerance specifications in order to provide a tool to aide for tolerance design. Introduced is the range of feasible displacement of a part of an assembly as a result of the accumulation of clearances between mating parts along the serial as well as parallel chain of assembly, which can play a role of compensating for the pose error of the corresponding part due to the accumulation of tolerances. Assemblability is determined by the probability that all the parallel as well as serial chains of the assembly can be formed successfully under given statistical properties of tolerances. This requires to numerically compute the probability that a pair of mating parts forming a parallel chain can align themselves within the error margin defined by their mating clearance, accounting for the possible compensation of the ranges of pose errors associated with the mating parts due to tolerance accumulations by the ranges of feasible displacements of the mating parts due to clearance accumulations. Simulation results are shown.

KEYWORDS: Assemblability evaluation; Tolerance and clearance; Statistical representation.

1. INTRODUCTION

Manufacturing processes are inherently imprecise such that they produce parts that vary. The variability that can be accepted is the responsibility of the designer. To control the variation, the designer specifies tolerances which describe the maximum acceptable deviation from nominal. Consequently, this tolerance specification affects the manufacturability, assemblability, functionality, and the cost effectiveness of a product.

With the trend towards computer integrated manufacturing, there have been research efforts to define tolerances mathematically,[1–3] to represent them in solid modeling,[4–6] and to automatically synthesize and analyze them.[7–9] Mathematical definitions of tolerances avoid ambiguous interpretations and provide a sound basis for assessing the correct analysis techniques and algorithms. Representing tolerances in solid modeling is necessary for planning automatically the manufacture, inspection, and assembly of mechanical products. Current solid modeling systems lack the facilities for representing and manipulating tolerances. A tolerance analysis attempts to predict the effect of tolerances on the design functions. A tolerance synthesis attempts to compute tolerance limits that satisfy constraints on the design functions and that result in the lowest cost design.

Most tolerance analysis methods are based on predicting the effect of tolerances on the design functions. For example, a tolerance analysis may predict whether or not the clearance between two parts (or two surfaces) in an assembly is within the design constraint.[7, 10] However, the assemblability of the product design can be affected by tolerances also, since they can propagate and accumulate in the assembly. For example, Shalon et al.[11] described an assembly problem encountered in one of the Boeing airplane subassemblies. The mechanic was unable to install the hydraulic tube in the 767 Rolls Royce engine strut subassembly due to the varying distance between the hydraulic reservoir and the tube-interface fitting.

For the assemblability evaluation, a concept of adjustable displacement provided by the clearances between mating features is used. The deviations or pose errors of parts in the assembly caused by given tolerances may be compensated using the adjustable displacements during assembly. That is, some part poses may be adjusted to compensate the deviations by using the free space allowed by clearances. For example, Figure 1(a) shows that when assembled, nominal P1 and P2 permit the peg feature of P2 to be at its nominal pose, with respect to the coordinate frame of P1. However, Figure 1(b) shows that when assembled, toleranced P1 and P2 permit the peg feature of P2 to be deviated from its nominal pose. This deviation may cause the entire assembly to fail. Now, let's assume that nominal P1 and P2 have clearance as shown in Figure 1(c). The clearance exists between the peg feature of P1 and the hole feature of P2. The pose of P2 can be freely adjusted within the clearance to bring it to its desired pose, if possible, when P1 and P2 are toleranced. Figure 1(d) shows that the hole feature of P2 can be brought to its desired (nominal) pose by translating P2 in $+x$ direction by the amount the clearance permits, although P1 is toleranced. In this paper, we use the terminologies "clearance" and "adjustable displacement" interchangeably.

In this paper, we propose statistical representations for tolerances and clearances, and a systematic approach to an assemblability evaluation based on the propagation of tolerances and clearances. Tolerance is represented by an ellipsoid, which is assumed to be distributed normally. It is a statistical approximation of real tolerance. Clearance is

* Department of Computer Science, University of Southern California, Los Angeles, CA 90089-0781 (USA).
† Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109 (USA).

https://doi.org/10.1017/S0263574798000344 Published online by Cambridge University Press

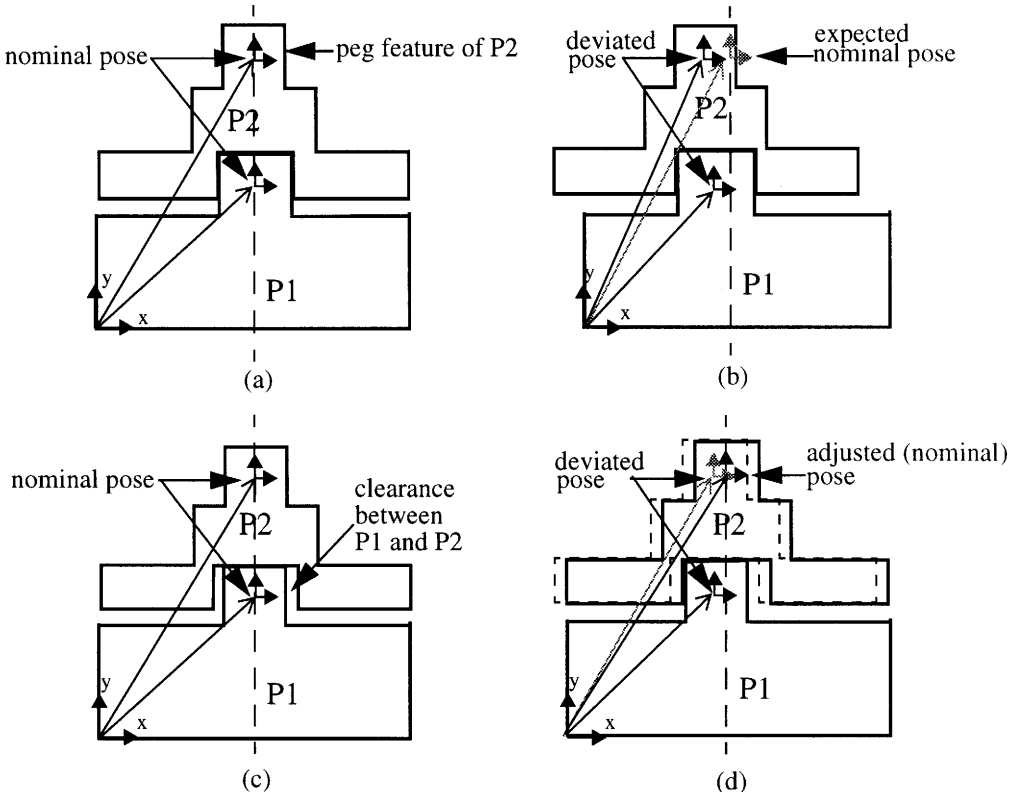


Fig. 1. Assembly of toleranced parts with and without clearance.

represented by a nominal ellipsoid and a range, which is assumed to be distributed normally. The same representations are used in the propagation of tolerances and clearances. Monte-Carlo method is used to approximate tolerances and clearances.

In this work, we assume the following: (1) Parts are rigid. Flexible parts, such as a thin metal plate, increase the uncertainty of the pose of a feature; (2) Features are rigidly attached to the part. Features that can rotate about an axis of the part increase the uncertainty of the feature pose; (3) Clearance cannot be negative. Tight-fit or press-fit with negative clearance is not permitted; (4) All parts satisfy tolerance specifications.

2. RELATED WORKS

Foster[12] and ANSI Y14.5M[13] describe geometric dimensioning and tolerancing for designing and manufacturing components of mechanical products. The main intended use is in manufacturing and inspections. However, if specified correctly by a designer, the tolerances can guarantee the assemblability of two parts with multiple mating features. However, these representations can be understood only by humans, and they are not computationally effective.

Whitney and Gilbert[14] proposed transform matrices to represent tolerances defined in ANSI standard. Using Chi-square error scheme, they computed the value for an optimizing factor which gave the optimal parameter values. A possible representation for clearance between two mating features was described briefly in terms of conditional variance. Clearance was treated the same as tolerance. However, neither a representation nor the computation for clearance was provided. We have adopted their approach for approximating the tolerances. However, our approach uses Monte-Carlo method to compute the tolerance propagation. In addition, we propose a clearance representation and an approach to clearance propagation. Moreover, the concept of clearance used in our approach is completely different from Whitney's work.[14]

Gilford and Turner[7] showed that a tolerance specification can be expressed as an in-tolerance region of a normed vector space. They developed the methods for tolerance analysis based on finding the relationship between in-tolerance region and in-design region. However, the dimensionality of the vector space can become large, since the dimensionality is proportional to the number of model variables. A model variable is a measure of extent to which some geometric property of a given part varies from nominal.

Su and Lee[15] modeled tolerance using differential transforms, and characterized it using mean and covariance. They proposed an analytical method to compute the pose uncertainty of an object during an assembly task. However, their method did not consider clearances.

Bjorke[10] proposed several methods for a tolerance analysis based on functionality. His goal was to derive a tolerance chain equation for a sum dimension. Then, this sum dimension was checked for the functionality criteria; for example, the clearances between two surfaces must be larger than zero and less than some limit. However, the analysis is limited to one dimensional space. Treacy et al.,[8] Wang and Ozsoy,[16] and Soderberg[17] implemented a data structure, an algorithm for generating the tolerance chain, and a computer inteface of Bjorke's approach, respectively.

https://doi.org/10.1017/S0263574798000344 Published online by Cambridge University Press

3. REPRESENTATION

We propose to represent the assembly by a graph called Feature Graph (*FG*). *FG* consists of a set of nodes and a set of edges. There are three types of nodes in *FG*: F-node, M-node, and P-node. F-node represents a mating feature with its tolerance. A mating feature is a feature of a part (e.g. cylindrical peg) which is used to interconnect with other parts. M-node represents a pair of mated features possibly with clearance. P-node represents the coordinate frame of a part. An edge in *FG* represents a relation between P-node and F-node with its transformation matrix describing the nominal pose of F-node from P-node. In *FG* figures, F-node is denoted by either a white circle or a gray circle, M-node is denoted by a gray circle inside a white circle, and P-node is denoted by a black circle.

A subset of nodes in a *FG* can be classified into a serial chain (S-chain), a parallel chain (P-chain), or a multi chain (M-chain). S-chain is defined as a chain, $\langle n_0, n_1, \ldots, n_i, n_{i+1}, \ldots, n_k \rangle$ such that (n_i, n_{i+1}) is an edge in *FG*, without repeated nodes. P-chain is a chain without repeated nodes, except $n_0 = n_k$. M-chain is a subgraph of *FG* with two or more P-chains.

3.1 Tolerance representation

ANSI Y14.5M defines varous types of geometric tolerances, e.g. location, orientation, form, profile, and runout, to represent tolerances of a feature. We classify these tolerances into two types: pose tolerance and dimension tolerance. Pose tolerance, such as location, orientation, and runout tolerances, allows position and orientation deviations of a feature. Dimension tolerance, such as form tolerance, does not affect the pose, but the clearance between the mating features.

Pose tolerance (or tolerance, for short) is represented by an ellipsoid with Gaussian distribution. Although manufacturing processes can have distributions such as rectangular and beta, Gaussian distribution is assumed. It is because the specific manufacturing processes are not known at the design stage, many manufacturing processes have Gaussian distribution,[17] and central limit theorem[18] may be applied to the tolerance propagation.

A tolerance ellipsoid is an approximation of real tolerance described in kinematic parameters. For example, the peg feature of P1 shown in Figure 2(a) has position tolerance of 0.02. Figure 2(b) shows that the position tolerance can have the tolerance zone representing the zone that the axis of the peg is allowed to reside. Figure 2(c) shows the tolerance volume of the tolerance zone. The volume shows that the peg axis can translate maximum of ± 0.01 in x-axis (indicated by the parameter x), and rotate maximum of ± 0.014 (indicated by the parameter θ). This tolerance volume shows that x and θ values are dependent. Finally, this tolerance volume is approximated by an ellipse shown in Figure 2(d). The ellipse approximates the density of the tolerance volume using Chi-square error reduction scheme. The optimization factor, α, can be obtained during the optimization process explained next.

We assume that the tolerance has Gaussian distribution, and simulate the possible feature poses using Monte-Carlo method.[19] For example, we randomly and normally generate a number (x_i) on the x-axis within the limit ± 0.01. Then, we randomly and normally generate a number (θ_i) on the θ-axis within the x_i limit. This random sample (x_i, θ_i) is an instance of one simulation result. By repeating this process *N* times,

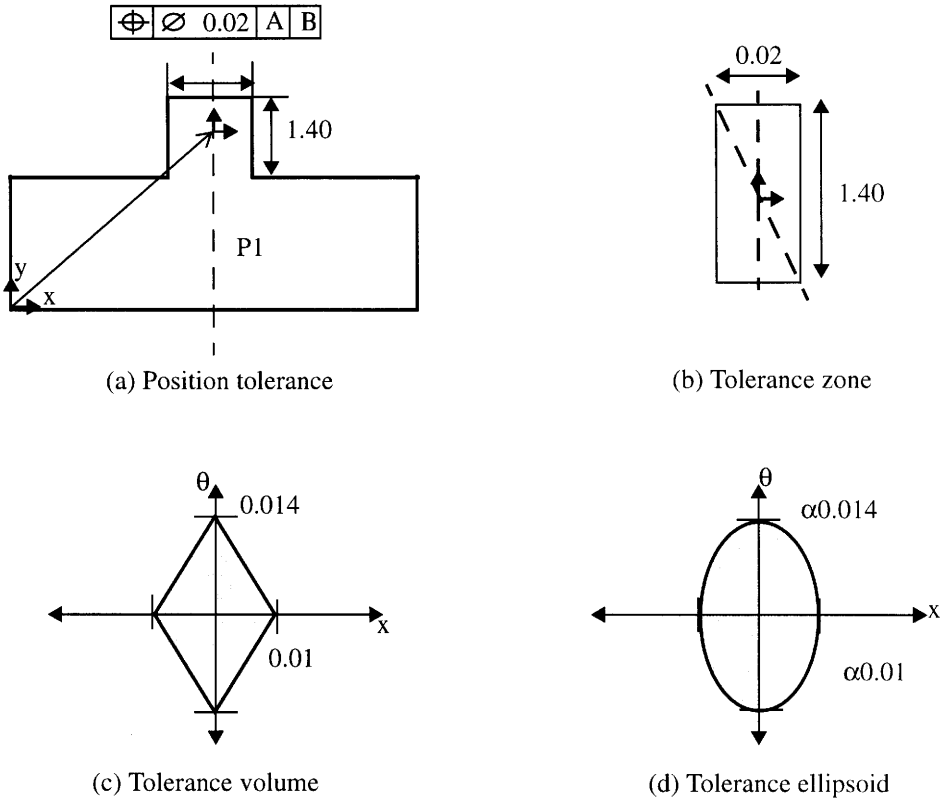


Fig. 2. Pose tolerance and representation.

https://doi.org/10.1017/S0263574798000344 Published online by Cambridge University Press

we have a simulated distribution of the tolerance. An analytic solution with Gaussian density function, where σ equals to one-third of the limit, is used to approximate the simulated solution. σ gives a good approximation, since the probability within the $\pm 3\sigma$ boundary of Gaussian distribution is 0.989 in 2-dimension.[20] This initial σ can be optimized using Chi-square error reduction scheme as described in Whitney's work.[14] For this, the simulated area and analytic area are partitioned into square grids. Then, for each grid, we add the square of the difference of two probabilities. This process is repeated iteratively by multiplying σ by an optimization factor, α, until the minimum sum is found for some α. The optimized ellipsoid has the axis lengths of $\alpha(3\sigma)$. This is shown in Figure 2(d). Whitney[14] provides detailed steps of this process.

3.2 Clearance representation

Clearance represents a possible free space between two mating features, where the free space allows one mated part to be adjusted in pose relative to the other part. This free space can be used to compensate the pose errors of a part in an assembly. Clearance comes from the combination of the nominal dimensions and the dimension tolerances of the two mating features, where the dimension tolerances are independent of datums or other features. That is, with the nominal dimensions, clearance has a nominal volume. However, with the dimension tolerances, clearance is permitted to vary from its nominal.

Some mating features may have a functionality constraint like an allowed clearance between the mating features. This functionality constraint is specified by the designer to ensure the functionality of the product. Therefore, the adjustable displacement between the mating features must take the specified functionality constraint under consideration. When the minimum limit of a functionality constraint is specified, it is subtracted from the nominal clearance of the mating features. The subtracted nominal becomes the adjustable displacement. However, when the maximum limit of a functionality constraint is specified, the adjustable displacement is the smaller of the two: the maximum limit of a functionality constraint and the nominal clearance of the mating features. When both the minimum and the maximum limits are specified, the nominal clearance is subtracted by the minimum limit, then the smaller of the two (the maximum limit and the nominal clearance after the subraction) is taken as the adjustable displacement. Throughut this paper, we use the terminologies "clearance" and "adjustable displacement" interchangeably to mean the adjustable displacement, which includes the functionality constraints between mating parts.

Clearance is represented by a nominal ellipsoid and a range which is assumed to have Gaussian distribution. A nominal ellipsoid approximates the nominal clearance. The range about the nominal ellipsoid represents the dimensional tolerances of the two mating features.

A clearance ellipsoid and a range are computed from the clearance zone of two mating features. A clearance zone is the real clearance that can be observed from the mating features. Similar to tolerance zone, this clearance zone can be transformed into a clearance volume in kinematic parameters. However, the variability of the clearance zone must be included. For example, the clearance zone and clearance volume between P1 and P2 in Figure 3(a) are shown in Figure 3(b) and (c), respectively.

The following algorithm shows how to compute the clearance ellipsoid and the range of a peg and hole:

Algorithm: Clearance Ellipsoid (Peg-Hole)

1. Randomly and normally generate the diameters of the peg and the hole, D_p and D_h, by using $3\sigma_p$ and $3\sigma_h$ equal to the diameter tolerances of the peg and hole, respectively. σ_p and σ_h are the standard deviations. The clearance zone is the rectangular area with its width equal to $D_h - D_p$, and the height equal to the depth of the hole. Then, the clearance volume is the clearance zone represented by kinematic parameters, similar to a tolerance volume.
2. Generate N clearance volumes from step 1.
3. Generate M uniformly distributed samples, S_c, inside a square area. This should cover the largest clearance volume.
4. For each clearance volume, V_i for i = 1, N, collect $S_i \subseteq S_c$ which is in V_i. This collection forms a flat-top bell shape distribution.
5. Approximate the collected sample distribution by Gaussian probability density function.
6. Using Gaussian-Sigmoid filtering method, calculate the

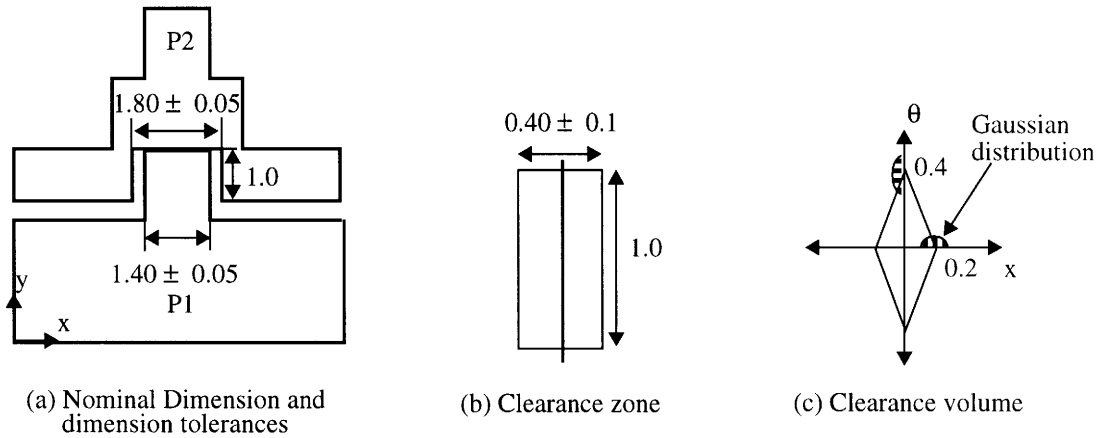


Fig. 3. Clearance of P1 and P2.

https://doi.org/10.1017/S0263574798000344 Published online by Cambridge University Press

optimized minimum and maximum clearance ellipsoids. The nominal ellipsoid is estimated to be the average ellipsoid of the minimum and maximum ellipsoids, and the range is calculated as half of the difference between the two ellipsoids' axis lengths.

Steps 5 and 6 are discussed in more detail in the next section. Note that steps 1 through 4 can be modified to collect samples of clearances for clearance sweep and clearance intersection operations for the propagation of clearances in an assembly.

4. COMPUTATION

In this section, we show the details of the computation process for calculating a clearance ellipsoid and a range from the clearance samples. The computation is based on Gaussian-Sigmoid filter method and Chi-square error reduction scheme.

4.1 Monte-Carlo method

The Monte-Carlo method is used to simulate the dimension tolerances of mating features to generate a possible set of diameters of a peg and hole. This simulation adds the randomly and normally generated dimension tolerances to the nominal diameters. Note that those random samples outside the 3σ limit are discarded to respect the tolerance specification. N simulated diameters are collected. Then, the clearance volumes are calculated using the randomly and normally generated diameters of the peg and hole with the depth of the hole. For example, the peg feature of P1 in Figure 3(a) has a diameter of 1.38, and the hole feature of P2 has a diameter of 1.81 as an outcome of the i_{th} trial of Monte-Carlo simulation. Then, the clearance zone and the clearance volume are computed, and they are shown in Figure 4(a) and (b), respectively. The value of θ is calculated using the small angle of approximation. Figure 4(a) shows that the clearance zone has a diameter of 0.43 and the height of 1.0. Figure 4(b) shows that the tolerance volume has a diamond shape with ± 0.215 in x-axis and ± 0.43 in θ-axis. The shape of the clearance volume denotes the dependency between the parameters x and θ.

This Monte-Carlo simulation results in N diamond areas. The distribution of these diamond areas is simulated by collecting uniformly distributed samples for each area, as described in steps 3 and 4 in the clearance ellipsoid algorithm. Note that similar clearance distributions are generated during the propagation of clearances as will be described in section 5. Next, we will show how this distribution can be approximated by an ellipsoid and a range.

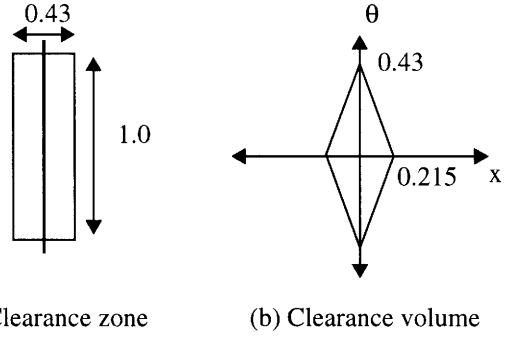
4.2 Gaussian-Sigmoid Filter

The Gaussian-Sigmoid filter method approximates the simulated solution by an analytical solution. This method approximates the simulated distribution of clearance by Gaussian-Sigmoid density function (3), with the parameter $Y = G(X)$, where $G(X)$ is a Gaussian probability density function (2). Two parameters, Σ and T, in (2) and (3), respectively, control the shape of Gaussian-Sigmoid distribution. Optimal values of these parameters, which give the minimum Chi-square errors beween the simulated density and analytical density, are used to calculate the clearance ellipsoid and the range.

Before beginning the optimization process, the initial approximation of the simulated distribution, D_s, is performed by computing the covariance matrix (Σ) of D_s, using (1). Note that Σ may not be a diagonal matrix, meaning that the matrix has rotational components. However, we can always compute the eigen values (Λ) and orthogonal eigen vectors (V) which correspond to the diagonal matrix of Σ and the rotation matrix of Σ, respectively. Throughout this paper, we assume that Σ is a diagonal matrix, unless stated otherwise. The rotated covariance matrix can always be recomputed by $\Sigma = V\Lambda V^T$.[21] Now, Σ is used in Gaussian probability density function (2). However, D_s is a flat-top bell shaped distribution. Therefore, this Σ is just an initial covariance, and must be optimized. This optimization process is explained in section 4.3.

Figure 5 illustrates the method. A simulated distribution of clearance samples is shown by a dark, broken line. Then, the covariace (Σ_i) of these samples is calculated using (1). The Gaussian distribution with Σ_i is shown by a light, solid line. This Σ_i is used in Gaussian-Sigmoid equation to generate the distribution shown by a dark, solid line.

The flat-top surface of Gaussian-Sigmoid distribution of Gaussian-Sigmoid distribution can be controlled by adjusting the parameter T. That is, the parameter T controls the slope of the Sigmoid function, whch in turn controls the area of the flat-top surface of Gaussian-Sigmoid distribution (3). Therefore, optimal T must be computed which gives the best fit of analytical flat-top surface to the simulated flat-top surface. Also, Σ must be optimized. This optimization process will be explained in the following subsection.



(a) Clearance zone (b) Clearance volume

Fig. 4. Clearance zone and volume for i_{th} trial.

https://doi.org/10.1017/S0263574798000344 Published online by Cambridge University Press

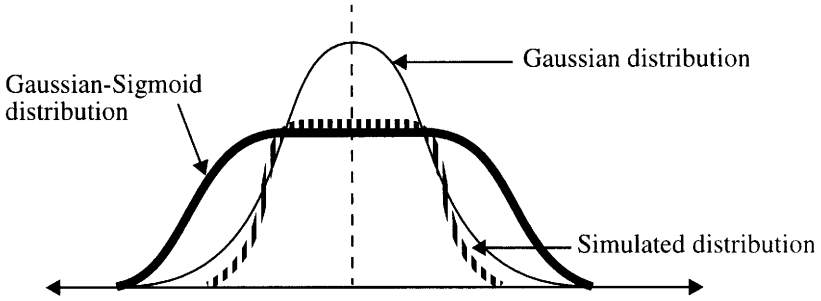


Fig. 5. Gaussian, Gaussian-Sigmoid, and simulated distributions.

$$\Sigma = \frac{\sum_{i=1}^{M} (pt_i^T \times pt_i)}{M} \tag{1}$$

$$G(X) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-(1/2)X\Sigma^{-1}X} \tag{2}$$

$$S(Y) = 2\left[\frac{1}{1+e^{Y/T}} - \frac{1}{2}\right] \tag{3}$$

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} (S(X_{i,j}) - P(X_{i,j}))^2 \tag{4}$$

4.3 Optimization process

A gradient-based method is used to optimize the parameter T and to find the optimizing factor for Σ, such that the analytic solution can best approximate the simulated solution. The initial Σ is the covariance matrix of the clearance samples collected during Monte-Carlo simulation. The initial T is calculated to start the gradient-based optimization. Then, two step optimization is performed: (1) optimize the flat-top surface area for the minimum ellipsoid, and (2) optimize the bottom area for the maximum ellipsoid.

The Chi-square (X^2) error equation is used in this problem, as shown in (4). In this equation, $S(X_{i,j})$ is the analytical density at $X_{i,j}$ and $P(X_{i,j})$ is the simulated density at $X_{i,j}$, where $X_{i,j}$ is the center position of the square grid(i, j). $P(X_{i,j})$ is normalized such that the number of samples in a square grid(i, j) is divided by the maximum expected density of the grid.

The normalization of $P(X_{i,j})$ may produce some grids with the density larger than 1. This is due to the random generation of samples generated during the Monte-Carlo simulation. To take this situation into account, the optimization of a flat-top surface begins by adding the areas of the grids that have the normalized density of 0.95 or larger. This sum (A_{FT}) of areas of the grids is the approximation of the area of the simulated flat-top surface. From this, initial T can be derived for the Gaussian-Sigmoid function.

For the initial value of T, first calculate the minimum ellipsoid (E_{min}) from A_{FT} and the ratio, $\rho = \sigma_1^2/\sigma_2^2$. This E_{min} has the equal area to A_{FT}. The axis lengths of E_{min} are a and b of $A_{FT} = \pi ab$ where $a = \rho b$ and $b = \sqrt{A_{FT}/(\pi\rho)}$. Then, T can be derived with X = (a, 0), or X = (0, b) as follows

$$k = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-(1/2)X\Sigma^{-1}X}$$

$$0.9985 = 2\left[\frac{1}{1+e^{-k/T}} - \frac{1}{2}\right]$$

$$T = \frac{-k}{\log[(2/(0.9985-1)) - 1]}$$

However, this T only approximates the area of the simulated flat-top surface. T can be optimized to better approximate the area with respect to the density of the flat-top area.

Optimal T should have the minimum error between the densities of simulated and analytic solutions, only in the area of the flat-top surface. The Chi-square error equation can be formulated such that the densities outside the flat-top area are zero for both densities. Since the only variable is T, the gradient δT can be formulated as

$$\delta T = -\alpha \frac{\partial \chi^2}{\partial T} = -\alpha \sum_{i=1}^{n}\sum_{j=1}^{m} \left\{ 2(S(X_{i,j}) - (PX_{i,j})) \frac{\partial S}{\partial T} \right\} \tag{5}$$

The coefficient α controls how fast the gradient-based method should converge. At each iteration. T is added by δT until T converges (or X^2 reaches its lower limit.) This T_{opt} is the optimal T for the given initial Σ Note that $S(X_{i,j})$ and $P(X_{i,j})$ are zeros for all $X_{i,j}$ not in A_{FT}.

Now, Σ can be optimized to best fit the flat-bottom of S(Y). When optimizing Σ, T_{opt} must also change accordingly to maintain the optimal flat-top area. There are two steps involved in this process: (1) calculation of the gradient δc for the optimization factor c for Σ, and (2) re-calculation of T to keep the flat-top area unchanged each time Σ is changed toward more optimal solution. The optimization factor c maintains the shape of the ellipsoid. The gradient δc can be formulated as

$$\delta c = -\beta \frac{\partial \chi^2}{\partial c} = -\beta \sum_{i=1}^{n}\sum_{j=1}^{m} \left\{ S(X_{i,j}, c\Sigma) - P(X_{i,j})) \frac{\partial S}{\partial c} \right\} \tag{6}$$

Finally, the clearance ellipsoid and the range can be

https://doi.org/10.1017/S0263574798000344 Published online by Cambridge University Press

computed from the minimum and maximum ellipsoids derived from (3) by letting S(Y)=Z, where Z is equal to 0.9985 and 0.0015, respectively. The ellipsoid equation derived from (3) is shown in (7) and (8). The nominal clearance ellipsoid is

$$1=\frac{1}{r}X^T\Sigma^{-1}X \tag{7}$$

$$r=-2Ln\left[\left(-TLn\left(\frac{1-Z}{1+Z}\right)\right)((2\pi)^{n/2}|\Sigma|^{1/2})\right] \tag{8}$$

the average ellipsoid of the two ellipsoids, i.e., the axis lengths of the nominal ellipsoid are the average lengths of the two ellipsoids. The range is the difference between the axis lengths of the nominal and the minimum ellipsoids (or maximum ellipsoid).

5. PROPAGATIONS

A tolerance propagation refers to the effect of the tolerance of a feature on other features in the assembly. Similarly, a clearance propagation refers to the effect of the clearance of two mating features on other features in the assembly. This section describes a systematic approach to calculating the propagation of tolerances and clearances. Three types of propagation are described for different types of chains: serial, parallel, and multi chains.

5.1 Serial chain

In a serial chain, the tolerances and clearances propagate from the begin node (B-node) to the end node (E-node) through the intermediate nodes (I-nodes). B-node is the first node in a serial chain. E-node is the last node of a serial chain. All other nodes in a serial chain are I-nodes. The tolerance propagation and the clearance propagation are calculated independently. For the propagation calculation, a sequence of sweep operations is applied from B-node to E-node. Note that a rotation component of i_{th} node creates a small translation component at $i+1_{th}$ node in a serial chain. This effect can be included in the sweep operations.

A sweep operation is defined as the addition of position vectors of two volumes for all possible combinations of vectors. The general idea of a sweep operation for two tolerance ellipsoids is to randomly and normally generate a sample point inside one tolerance ellipsoid. The sample point corresponds to an instance of a simulated feature pose. Then, place the other ellipsoid on this sample, and randomly and normally generate a point inside this ellipsoid. When these sample points are collected, they form a statistical volume. This volume is approximated by an ellipsoid. The algorithm for a tolerance sweep operation is as follows:

Algorithm: Tolerance Sweep Operation

1. Place the center of the first tolerance ellipsoid, E_1, at the origin of a coordinate frame.
2. Randomly and normally generate a sample point, p_i, within E_1.
3. Place the center of the second ellipsoid, E_2, at p_i.
4. Randomly and normally generate a point, p_j, within E_2.
5. Put p_j into a set T.
6. Repeat steps (2)–(5) N times.
7. Approximate these N samples in T by an ellipsoid. First, compute the covariace matrix, Σ, of N sample points. Then, compute the eigen vectors and eigen values of Σ, which give a good estimate of the ellipsoid orientation and axis lengths, respectively. The axis lengths of the ellipsoid are estimated to be 3σ, where each σ^2 is a diagonal covariance corresponding to eigen values.
8. The ellipsoid obtained from the above step can be optimized by using Chi-square error reduction scheme as described in section 3.1.

The general method of a clearance sweep operation for a serial chain is to repeatedly collect the samples from a maximum boundary that belong to the swept area of the two randomly created clearance ellopsoids. The maxiumum boundary should be at least the maximum possible sweep area of the two clearance ellipsoids. The samples in this maximum boundary are uniformly distributed. The collection of the samples belonging to the sweep areas forms a flat-top bell shaped distribution. The flat-top area corresponds to the minimum clearance ellipsoid of the sweep area. The key to the algorithm lies in the method of approximating the minimum and maximum clearance ellipsoids from this simulated clearance sweep distribution by using Gaussian-Sigmoid filter method discussed in section 4. The algorithm for a clearance sweep operation is as follows:

Algorithm: Clearance Sweep Operation

1. Create a maximum boundary, B. B is a rectangular box that can contain the swept volume of two maximum clearance ellipsoids.
2. For i=1 to R (R is the number of ellipsoids to be generated randomly and normally within the clearance range)
   - From the first clearance ellipsoid, randomly and normally generate an ellipsoid E_1 inside its clearance range.
   - From the second clearance ellipsoid, randomly and normally generate an ellipsoid E_2 inside its clearance range.
   - Randomly and uniformly generate G samples within B.
   - For each sample b_i in B, move E_1 to b_i. E_2 is on the origin. Check if E_1 and E_2 intersect. If they intersect, collect b_i into a set CLE. Note that the intersection check can be done through simulation. That is, randomly and uniformly generate M samples inside E_1. If there exists at least one sample in E_1 which intersects E_2, then E_1 and E_2 intersect. Otherwise, E_1 may not intersect E_2. If E_1 and E_2 intersect, it implies that b_i is in the clearance sweep area. Otherwise, it is not.
3. Calculate the clearance ellipsoid and the range from the CLE using Gaussian-Sigmoid filter method discussed in section 4.

Parallel chain

A parallel chain refers to a closed loop in a feature graph. In graph theory, a parallel chain corresponds to a simple loop,[22]

https://doi.org/10.1017/S0263574798000344 Published online by Cambridge University Press

which is defined as a path with all distinct nodes except the first and the last. Unlike a serial chain, a parallel chain may not always be achievable. In addition, the propagation of clearances and tolerances in a parallel chain can have a different result for the different choice of goal features. A goal feature is defined as a feature in a chain that is chosen to end the propagation. In other words, the solution is sought for that feature. A goal feature of a serial chain is the feature at the end of the chain. For a parallel chain, however, it can be any one of the features in a parallel chain since there is no end in a parallel chain. Usually, there must be a goal feature defined before the computation can start.

To compute the propagation of tolerances and clearances of a parallel chain, we hypothetically cut the parallel chain to form two serial chains, SC_1 and SC_2. The cut occurs at a goal feature. This cut will introduce two goal nodes for two serial chains sharing the same base node. Note that the clearance between the two goal node is assigned to the serial chain which does not have the goal feature of the parallel chain.

The propagation of tolerances and clearances of a parallel chain can be computed from the results of the propagations of SC_1 and SC_2. The algorithm for computing the serial chain propagation was already discussed in the previous section. Now, let's assume that SC_1 and SC_2 have tolerance ellipsoids, E_{t1} and E_{t2}, and clearance ellipsoids, E_{c1} and E_{c2}, respectively. E_{ci}, for $i=1, 2$, is a 2-tuple (E_c, R_c) where E_c is the clearance ellipsoid and R_c is the range. Then, the clearance ellipsoid, E_{c3}, and the tolerance ellipsoid, E_{t3}, of the parallel chain are computed using the following intersection operation algorithm:

Algorithm: Intersection Operation

1. Place the center of E_{t1} and E_{t2} at the origin of a coordinate frame.
2. Uniformly and randomly generate N samples, P_N, inside $\max(E_{c1})$, the maximum clearance ellipsoid of E_{c1}. $\max(E_{c1})$ can be simply calculated by adding the ellipsoid ranges to the ellipsoid axis lengths.
3. Normally and randomly generate a point, P_{t1}, inside E_{t1}, and a point, P_{t2}, inside E_{t2}.
4. Repeat R times:
   - Normally and randomly generate ellipsoids E_1 and E_2 from E_{c1} and E_{c2}, respectively.
   - Collect $P_M \subseteq P_N$ that are inside E_1, and translate P_M by P_{t1}. This spreads the clearance samples around the tolerance sample, P_{t1}.
   - Place the origin of E_2 at P_{t2}.
   - Collect samples $P_I \subseteq P_M$ that intersect with E_2.
   - Compute the mean, M_I, of P_I.
   - Move P_I to the origin such that the mean position of P_I is at the origin. Then, collect this P_I in C. Collect M_I in T.
5. Repeat step 3–4 K times.
6. Approximate the tolerance samples in T by an ellipsoid, E_{t3}.
7. Approximate the clearance samples in C by an ellipsoid, E_{c3}, using Gaussian-Sigmoid filter method.

The probability of successful assembly (assemblability measure) of a parallel chain is calculated by dividing the total number of failures by the total number of trials during the simulation. The failure is indicated by an empty P_I in step 4 of the algorithm.

5.3 Multi chain

A multi chain, M-chain, is composed of two or more parallel chains, P-chains. A recursive algorithm is proposed to compute the propagation of tolerances and clearances, given a base node (B-node) and a goal node (G-node). The method is based on recursively decomposing a M-chain into two simpler M-chains. Each simpler M-chain is formed by hypothetically cutting the P-chain, with G-node, of the original M-chain. Therefore, each simpler M-chain has one less P-chain than the original M-chain. When the solutions of the simpler M-chains are computed, the original M-chain can be solved by using the intersection operation on the P-chain that was hypothetically cut.

Algorithm: Multi Chain

1. Let P_b and P_g be the base and goal parts, respectively, of a M-chain, ChM. G-node can be either a F-node or a P-node. If G-node is a feature, then it is not in the M-chain, but it is attached to the M-chain, and P_g is a P-node that has G-node. Otherwise, Pg is a G-node. Select a feature node of P_g, $F_g \in P_g$, such that it is a member of a P-chain, ChP_g, where F_g, $P_g \in ChP_g$. P_b is a B-node.
2. Hypothetically disassemble the F_g from its mating feature, F_{mg}, in ChP_g.
3. Starting from F_g, follow the S-chain, ChS_g, until ChS_g reaches either P_b or P_i. P_i belongs to another P-chain, ChP_i, for $ChP_i \in ChM$. Note that all other chains that are not part of ChM are ignored, since they do not affect ChM.
   - If F_g reaches P_b, compute the tolerance and clearance propagations of ChS_g.
   - If F_g reaches P_i, compute the tolerance and clearance propagations at P_i from ChP_i and ChM′. $ChM' = ChM$ with two S-chains at F_g and F_{mg}, instead of ChP_g. The algorithm is recursively applied with $P_g = P_i$, $P_b = P_b$, $ChM = ChM'$, and $ChP_g = ChP_i$. Then, the result of tolerance and clearance propagations at P_i can be propagated to F_g through ChS_g.
4. Repeat the step (3) with F_{mg} and ChS_{mg}.
5. Perform an intersection operation with tolerances and clearances, (T_g, C_g) and (T_{mg}, C_{mg}), at the mating features F_g and F_{mg}, to compute the tolerance and clearance of the M-chain, (T_t, C_t), at F_g.
6. Propagate (T_t, C_t) to G-node.

Figure 6 illustrates the algorithm. For step (1), let $P_g = P_4$, $P_b = P_1$, and $ChP_g = ChP_1$ with ChM equal to the graph in the figure (a). In step (2), F_3 is selected as a hypothetically disassembled feature, F_g, $F_g = F_3$. This step produces two serial chains, $ChS_1 = (P_3, F_4, P_4, F_3^1)$ and $ChS_2 = (P_2, F_3^2)$. This is illustrated in the figure (b). Step (3) needs to solve for the tolerance and clearance propagations of two chains, ChS_1 and ChS_2. ChS_1 starts from P_3, where $P_3 \in ChP_2$. Therefore, the tolerance and clearance propagations at P_3 must be solved by recursively applying the above algorithm. Since ChS_2 is a parallel chain, it can be solved by applying the parallel chain method discussed before. The result of the parallel chain propagation is propagated to F_3^1 through ChS_1 from P_3. Similarly, the tolerance and clearance propagations

https://doi.org/10.1017/S0263574798000344 Published online by Cambridge University Press

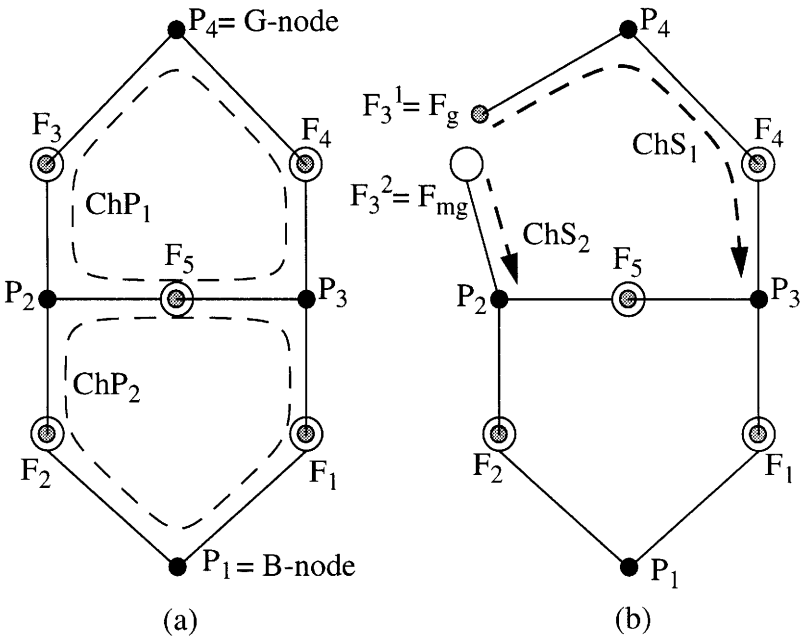


Fig. 6. Feature graph (Multi chain).

to F_3^2 can be computed, for step (4). In step (5), the intersection operation is applied to the results of steps (3) and (4). Step (6) propagates the result obtained from the step (5) to F_4.

6. ASSEMBLABILITY EVALUATION

The assemblability of a product design under the given tolerance specification is measured statistically by solving the tolerance and clearance propagations of parallel chains in the feature graph of the product design. That is, the assemblability measure is the product of the probabilities of successfully assembling all parallel chains.

The probability of adjusting a feature to its nominal pose can be numerically computed by computing the possible pose errors that can be compensated by clearances. This probability can be computed by the following algorithm:

Algorithm: Assemblability Measure

1. Place both tolerance and clearance ellipsoids at the origin of a coordinate frame.
2. Randomly and normally generate N tolerance samples within the tolerance ellipsoid.
3. Randomly and normally generate a clearance ellipsoid, E_i, within its clearance range.
4. Count the number of tolerance samples, N_i, which intersect E_i.
5. Repeat the steps (3) and (4) M times.
6. Calculate the assemblability measure, M_{asm}, using the following equation:

$$M_{asm} = \left(\sum_{i=1}^{M} N \right) / (\mathrm{M} \times \mathrm{N})$$

This algorithm counts M times those samples in the tolerance ellipsoid that can be always compensated by the clearance ellipsoid and the range. The area where these samples occupy is indicated by a light colored area in Figure 7. The other samples, which may be selected and counted probabilistically using the clearance range, lie in the area indicated by a dark color. Other samples which cannot be compensated even by the largest clearance are never counted.

7. SIMULATIONS

In this section we show three simulation results: (1) clearance ellipsoid and range computation between the peg

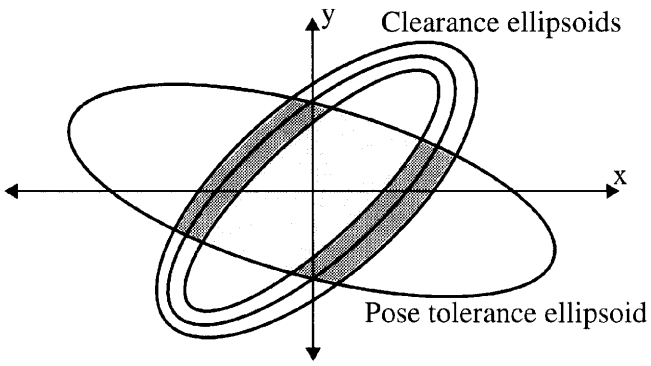


Fig. 7. Assemblability measure calculation.

https://doi.org/10.1017/S0263574798000344 Published online by Cambridge University Press

and hole mating features, (2) the propagations of tolerances and clearances of a serial chain, and (3) the propagations of tolerances and clearances of a parallel chain. The simulation was done based on a two dimensional assembly product shown in Figure 8 and Figure 9. Figure 8 shows coordinate frames of parts and mating features, and their relationships, whereas, Figure 9 shows dimensions and tolerances of parts. The simulation was performed using two kinematic parameters, one translation, x, and one rotation, θ, for simplicity and clarity.

Example 1: Clearance computation for mating features

In this example, we show the computation process and a result of a simulation for computing the clearance ellipsoid and the range of the peg and hole features shown in Figure 3(a). The clearance zone has a rectuangular shape with the width of 0.40 ± 0.1 and the height of 1.0. We performed Monte-Carlo simulation with $N = 50$. This results in 50 possible clearance zones and volumes. The uniform distribution of clearance was simulated by generating $S = 2500$ random and uniform samples within the maximum boundary. Then, samples $S_i \subseteq S$ that are in the clearance volumes are collected, for all N clearance volumes. The corresponding normalized density mesh of the simulated clearance is shown in Figure 10(a). The clearance area was partitioned into square grids with longer side of the area partitioned into 20 grids.

The flat-top surface area is computed by adding the area of grids which density is 0.95 or larger. This allows 5% error margin during the simulation. The density mesh figure is shown in Figure 10(b). The flat-top surface area is computed to be equal to 0.0873. It closely approximates the real minimum clearance of diamond-shape clearance volume, which is 0.0692. Using this area and covariance matrix (Σ) of the simulated samples, T parameter value is calculated. The corresponding analytical Gaussian-Sigmoid density, with initial Σ, is shown in Figure 10(c). Finally, optimization is performed on Σ, and the corresponding Gaussian-Sigmoid density is shown in Figure 10(d). From T and Σ, the axis lengths of [0.1171 0.2374] and [0.2254 0.4773] (where [x θ]) are calculated for the minimum and maximum ellipsoids, respectively. From this, the nominal ellipsoid is calculated to have the axis lengths of [0.1762 0.3573], and the range of [0.0592 0.1199].

Example 2: Tolerance and clearance propagation of a serial chain

This example shows the propagation of tolerances and clearances of a serial chain (P1, f13, f31, P3, f33, f41, P4,

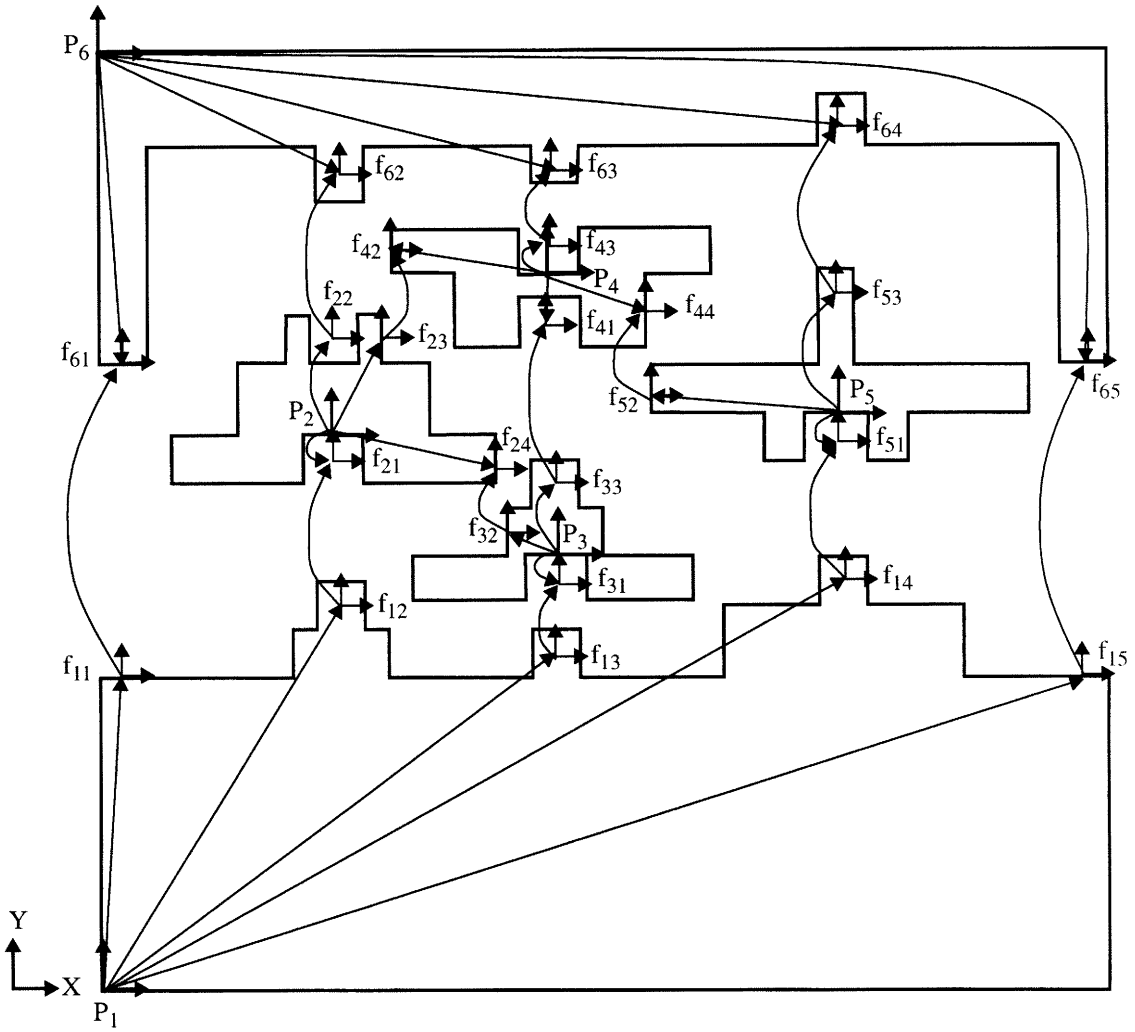


Fig. 8. Coordinate frames of a product.

https://doi.org/10.1017/S0263574798000344 Published online by Cambridge University Press

f43). That is, the tolerance and clearance at upper hole (f43) of P4 is computed for a subassembly consisting of parts, P1, P3, and P4. (See Figure 8.) In this simulation, the clearance was computed twice for two different values of diameters of the mating features: once with the diameters given in Figure 9, 1.40 ± 0.05 and 1.80 ± 0.05, and once with the diameters of 1.63 ± 0.001 and 1.67 ± 0.001 for pegs and holes, respectively. With the first values, the results of the simulations are reported. The tolerance ellipsoid is [0.123 0.028], with the rotation matrix,

$$\begin{bmatrix} 0.951 & -0.307 \\ 0.307 & 0.951 \end{bmatrix},$$

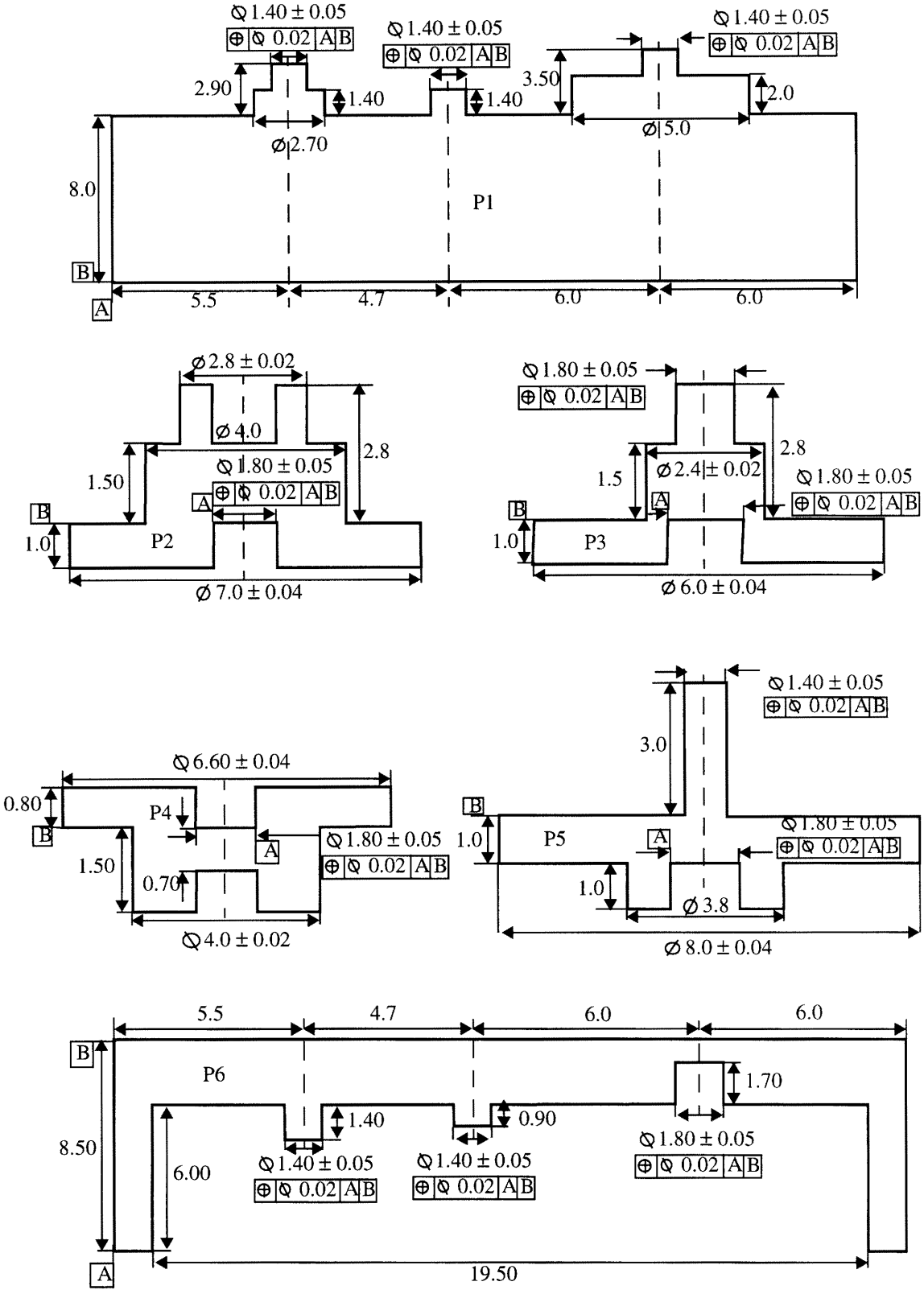


Fig. 9. 2-D assembly dimension and tolerance specifications.

https://doi.org/10.1017/S0263574798000344 Published online by Cambridge University Press

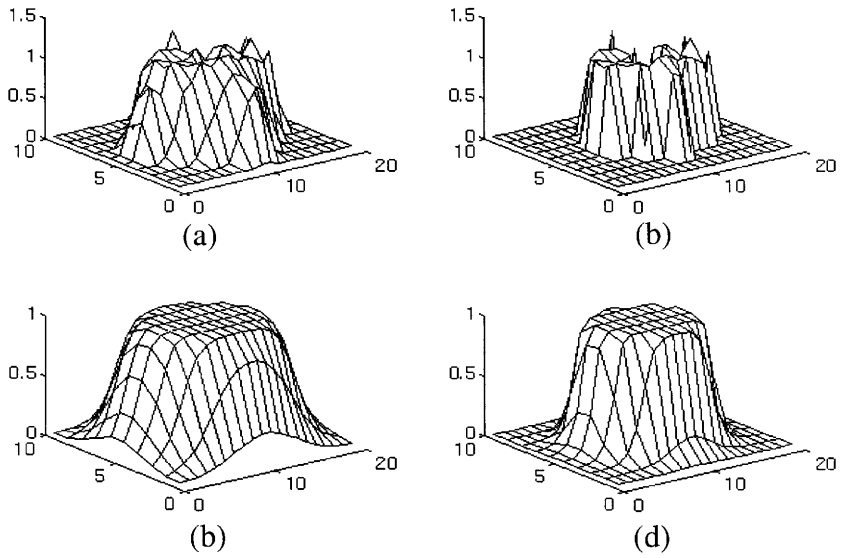


Fig. 10. Clearance representation simulation results.

and the clearance ellipsoid is [1.523 1.124], with the rotation matrix,

$$\begin{bmatrix} 0.999 & 0.007 \\ -0.007 & 0.999 \end{bmatrix},$$

and the range, [0.300 0.221]. With the second values, the clearance ellipsoid is [0.154 0.092], with the rotation matris,

$$\begin{bmatrix} 0.943 & -0.331 \\ 0.331 & 0.943 \end{bmatrix},$$

and the range, [0.066 0.039]. The results show that the first simulation has much larger nominal clearance at f43 than tolerance. However, the second simulation has the nominal clearance that is closer to tolerance. Note that a rotation matrix denotes the orientation of an ellipse.

Example 3: Tolerance and clearance propagation of a parallel chain

The parallel chain formed by the subassembly with P1, P2, and P3, is used in this simulation. Two serial chains, ChS1 = (P1, F12, f21, P2, f24) and ChS2 = (P1, f13, f31, P3, f32), were used by hypothetically cutting the parallel chain at the mating features, f24 and f32. Propagations of tolerance and clearances were computed for ChS1 and ChS2 using the serial chain propagation similar to example 2. Then, an intersection operation was applied to tolerances and clearances of ChS1 and ChS2. The final results are reported for two simulations using two different diameters for pegs and holes as in example 2. The simulation with larger clearances had the assemblability measure of 1. This implies that the clearances can compensate tolerances 100%. The resulting tolerance after the intersection operation [0.078 0.121] with the rotation matrix of

$$\begin{bmatrix} -0.743 & -0.668 \\ 0.668 & -0.743 \end{bmatrix}.$$

The clearance ellipsoid is [0.203 0.320], with the rotation matrix of

$$\begin{bmatrix} 0.712 & 0.701 \\ -0.701 & 0.712 \end{bmatrix},$$

and the range is [0.095 0.150]. This result is plotted in Figure 11, where dots, x's, and *'s indicate maximum, nominal, and minimum clerances, and +'s indicate tolerance. It shows that the area covered by +'s is smaller than the area covered by *'s. The second simulation was performed on the same parallel chain with smaller clearances, as inidcated in example 2. The assemblability measure was 0.41. The resulting tolerance after the intersection operation is [0.019 0.033] with the rotation matrix of

$$\begin{bmatrix} -0.740 & -0.671 \\ 0.671 & -0.740 \end{bmatrix}.$$

The clearance ellipsoid is [0.017 0.031], the rotation matrix is

$$\begin{bmatrix} 0.715 & 0.699 \\ -0.699 & 0.715 \end{bmatrix},$$

and the range is [0.008 0.015]. This result is plotted in Figure 12.

8. CONCLUSION

In this paper, a numerical method of analyzing and evaluating the assemblability of a product under given

https://doi.org/10.1017/S0263574798000344 Published online by Cambridge University Press

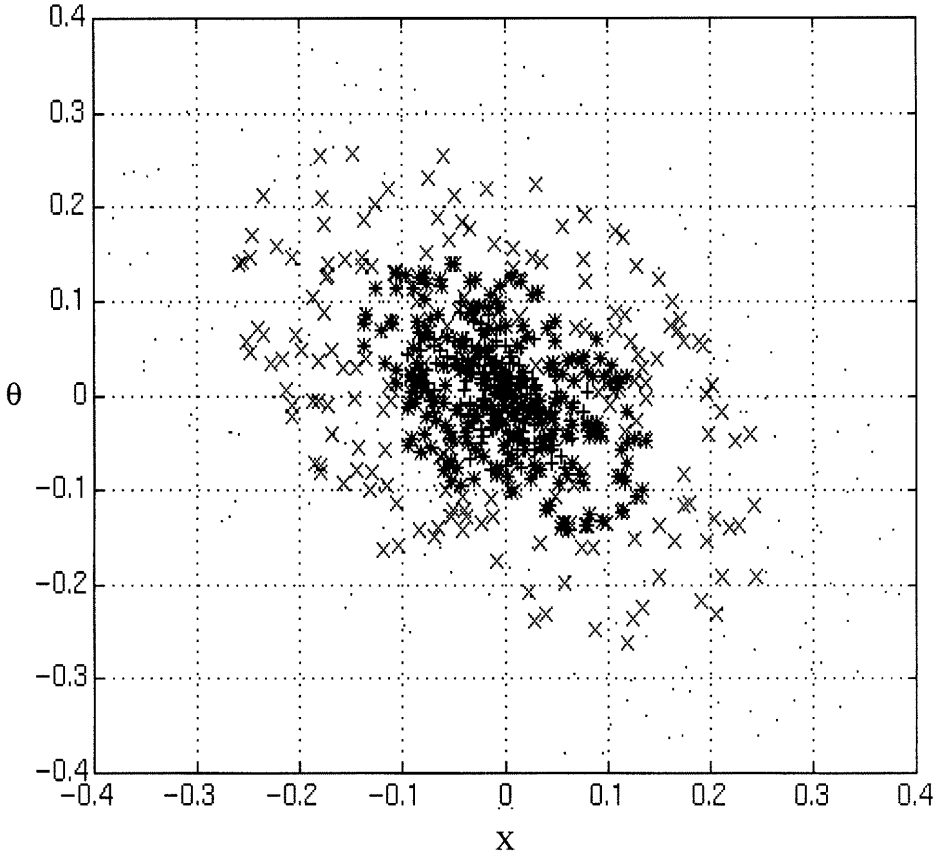


Fig. 11. Tolerance and clearance ellipses of with large clearances.

statistical specification of tolerances is presented. The method is based on a new concept that the assemblability of a product, defined as the probability that mating parts forming parallel chains can be aligned within the error

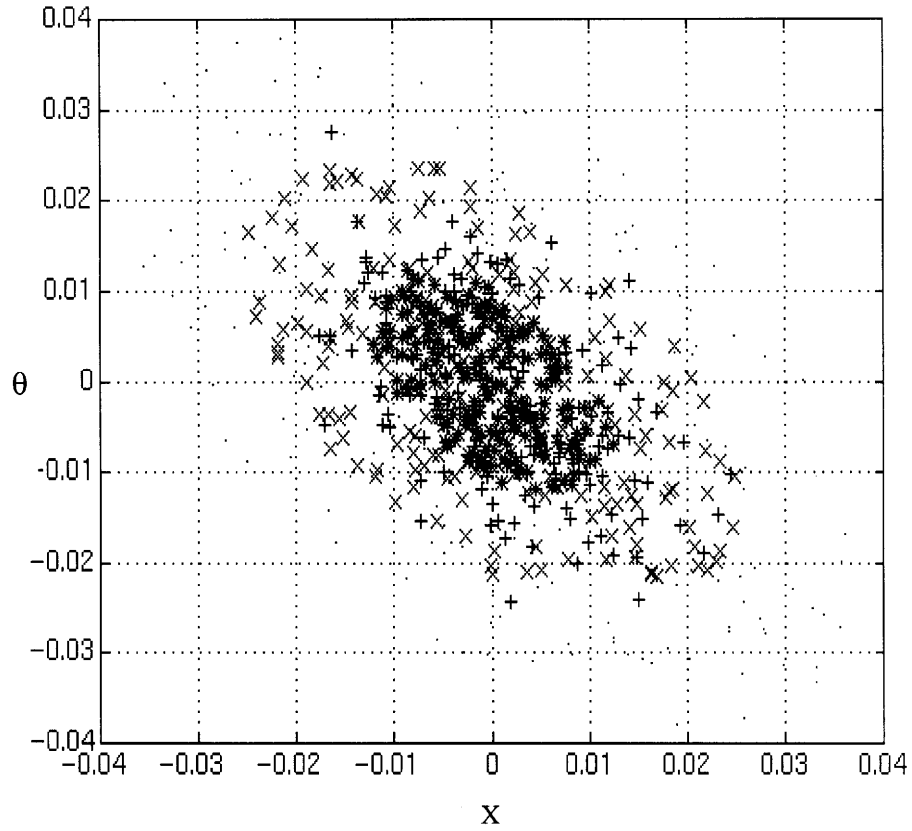


Fig. 12. Tolerance and clearance ellipses of with small clearances.

https://doi.org/10.1017/S0263574798000344 Published online by Cambridge University Press

margins allowed by their mating clearances, is a function not only of the pose errors of mating parts due to tolerance accumulations, but also of the feasible displacements of mating parts due to the accumulation of clearances along the chains of assembly that can compensate for the pose error of mating parts. A novel method of accumulating clearances numerically is presented for computing and approximating the feasible displacement ranges of parts statistically in terms of the reference coordinate frame based on, so called, Gaussian-Sigmoid filtering as well as Monte-Carlo simulation. The results obtained by this paper can be directly used as a computer-aided tolerance analysis and design tool, thus enhancing the capability of current geometric modeling and concurrent engineering systems.

References

1. A.A.G. Requicha, "Mathematical Definition of Tolerance Specifications" *Manufacturing Review* **6,** No. 4, 269–274 (1993).
2. F. Etesami, "A Mathematical Model for Geometric Tolerances" *J. Mechanical Design* **115,** 81–86 (1993).
3. P. Hoffmann, "Analysis of Tolerances and Process Inaccuracies in Discrete Part Manufacturing" *Computer-Aided Design* **14,** No. 2, 83–88 (1982).
4. A.A.G. Requicha, "Representation of Tolerances in Solid Modeling: Issues and Alternative Approaches" In: *Solid Modeling by Computers: From Theory to Applications* (J.W. Boyse and M.S. Pickett, Eds) (Plenum, New York, 1984) pp. 3–22.
5. G. Allen, "Tolerances and Assemblies in CAD/CAM Systems" *Manufacturing Review* **6,** No. 4, 320–328 (1993).
6. U. Roy and C.R. Liu, "Feature-based Representational Scheme of a Solid Modeler for Providing Dimensioning and Toleranciang Information" *Robotics and Computer-Integrated Manufacturing* **4,** No. 3/4, 335–345 (1988).
7. J. Guilford and J. Turner, "Advanced Analysis and Synthesis for Geometric Tolerances", *Manufacturing Review* **6**, No. 4, 305–313 (1993).
8. P. Treacy et al., "Automated Tolerance Analysis for Mechanical Assemblies Modeled with Geometric Features and Relational Data Structure" *Computer-Aided Design* **23**, No. 6, 444–453 (1991).
9. R.C. Hillyard and I. C. Braid, "Analysis of Dimensions and Tolerances in Computer-Aided Mechanical Design" *Computer-Aided Design* **10**, No. 3, 161–166 (1978).
10. O. Bjorke, *Computer-Aided Tolerancing* (ASME Press, New York, 1989).
11. D. Shalon, D. Gossard, K. Ulrich and D. Fitzpatrick, "Representing Geometric Variations in Complex Structural Assemblies on CAD Systems" *DE-Vol 44-2, Advances in Design Automation, ASME* **2,** 121–132 (1992).
12. L.W. Foster, *GEO-METRIC III: The Application of Geometric Dimensioning and Tolerancing Techniques (Using the Customary Inch System)* (Addison-Wesley Pub., Reading, Mass., 1994).
13. *Dimensioning and Tolerancing* (ANSI Y14.5M, American Society of Mechanical Engineering, New York City, 1982).
14. D. E. Whitney and O.L. Gilbert, "Representation of Geometric Variations Using Matrix Transformations for Statistical Tolerance Analysis in Assemblies" *Proceedings of the 1993 IEEE International Conference on Robotics and Automation* (1993) pp. 314–321.
15. S.F. Su and C.S.G. Lee "Uncertainty Manipulation and Propagation and Verification of Applicability of Actions in Assembly Tasks", *Proceedings of the 1991 IEEE International Conference on Robotics and Automation* (1991) pp. 2471–2476).
16. N. Wang and T.M. Ozsoy, "Automatic Generation of Tolerance Chains from Mating Relations Represented in Assembly Models" *Mechanical Design* **115**, 757–761 (1993).
17. R. Soderberg, "CATI: A computer Aided Tolerancing Interface" *DE-Vol. 44–2, ASME Advances in Design Automation* **2**, 157–164 (1993).
18. J. Stark and J.W. Woods, *Probability, Random Process, and Estimation Theory for Engineers* (Prentice Hall, Englewood Cliffs, N.J., 1986).
19. I.M. Sobol, *A Primer for the Monte Carlo Method* (CRC Press Inc., Boca Raton, 1994).
20. A. E. Bryson and Y-C. Ho, *Applied Optimal Control* (John Wiley and Sons, New York, 1975).
21. G. Strang, *Linear Algebra and Its Applications* (Harcourt Brace Jovanovich, Academic Press, New York, 1986).
22. T.H. Cormen, C.E. Leiserson and R.L. Rivest, *Introduction to Algorithms* (McGraw-Hill, New York, 1990).

https://doi.org/10.1017/S0263574798000344 Published online by Cambridge University Press