Abstract—Tolerancing decisions can profoundly impact the quality and cost of electro-mechanical assemblies. Existing approaches to tolerance analysis and synthesis in design entail detailed knowledge of geometry of the assemblies and are mostly applicable during advanced stages of design, leading to a less than optimal design process. During the design process of assemblies, both the assembly structure and associated tolerance information evolve continuously. Therefore, significant gains can be achieved by effectively using this information to influence the design of the assembly. Motivated by this, we identify and explore two goals for future research that we believe can enhance the scope of tolerancing for the entire design process. The first goal is to advance tolerancing decisions to the earliest possible stages of design. This issue raises the need for effective representation of tolerancing information during different stages of design and for effective assembly modeling. The second goal addresses the appropriate, synergistic use of available methods and best practices for tolerance analysis and synthesis, at successive stages of design. Pursuit of these goals leads to the definition of a multilevel design for tolerance process that enables tolerancing to be addressed at successive stages of design. This issue raises the need for effective representation of tolerancing information during different stages of design and for effective assembly modeling. The resulting design process, which we call the design for tolerance process, integrates three important domains:

1) design activities at successive stages of design;
2) assembly models that evolve continuously through the design process;
3) methods and best practices for tolerance analysis and synthesis.

We demonstrate major steps of our proposed approach through a simple, yet illustrative, example.

Index Terms—Assembly design process, assembly modeling, design tolerancing, statistical tolerancing, system level tolerancing, tolerance analysis, tolerance representation, tolerance synthesis.

I. INTRODUCTION

TOLERANCING is a critical issue in the design of electro-mechanical assemblies. In a 1997 workshop at the National Institute of Standards and Technology (NIST) [1], several leading researchers from industry, academia, and government emphasized the need for investigating assembly level tolerancing issues and for developing tolerancing standards related to assembly. Tolerancing is a major component in the open assembly design environment (OpenADE) architecture being developed and implemented at NIST [2]. Tolerancing includes both tolerance analysis and tolerance synthesis. In the context of electro-mechanical assembly design, tolerance analysis refers to evaluating the effect of variations of individual part or subassembly dimensions on designated dimensions or functions of the resulting assembly. Tolerance synthesis refers to allocation of tolerances to individual parts or subassemblies based on tolerance or functional requirements on the assembly. In this paper, we use the phrase design tolerancing to refer to tolerance analysis and synthesis during design.

A. Current Status of Design Tolerancing

Existing approaches to design tolerancing in electro-mechanical assemblies generally require detailed knowledge of the geometry of the assemblies and are mostly applicable during advanced stages of design. The current industry practice is to assign tolerances only during late stages of design, after nominal dimensions have been fixed by designers. Many firms use Monte Carlo simulation to conduct tolerance analysis on a detailed geometric model of the product. There are some important recent efforts, albeit preliminary, that attempt tolerancing decisions during early stages of design. These include the work based on key characteristics [3], [4]; and assembly-oriented design using assembly representations such as datum flow chains [5], [6]. Though some important design related decisions can potentially be enabled by these approaches during early stages of design, the actual tolerance analysis would require at least a rough geometric description of the assembled product.
Both worst-case tolerancing and to a less extent, statistical tolerancing, are currently practiced in industry [7]. Worst-case tolerancing involves establishing the dimensions and tolerances such that any possible combination will produce a functional assembly, i.e., the probability of nonassembly is equal to zero. Consequently, worst-case tolerancing can lead to excessively tight part tolerances and hence high production costs. Statistical tolerancing is a more practical and economical way of looking at tolerances and works on setting the tolerances so as to assure a desired yield, accepting a small percent of nonconformance.

There is now a vast body of literature on tolerance analysis and synthesis, with several survey papers available on important topics [8]–[19]. There are several software packages available exclusively for tolerance analysis and synthesis [17]. These packages are mostly simulation-based; simple analytical or probabilistic techniques are also provided. Industry best practices in design tolerancing include the well-known Motorola six sigma program [20]. Quality engineering techniques such as Taguchi Methods [21] are popular among some industries. There are also proprietary methods and software such as holistic probabilistic design (HPD) from Xerox [22], [23]. Monte Carlo simulation is the most popular technique used by industries and commercial packages.

Dimensional tolerancing has evolved mostly as an industrial practice without strong theoretical foundations [16]. The best tolerancing practices were collected and made available through an evolving series of tolerancing standards [24]–[27]. All international and most national standards have codified only worst-case tolerancing [7]. There are a few company specific internal standards for statistical tolerancing, such as in IBM [7] and Motorola [20]. The latest ANSI Y14.5M-1994 standard on dimensioning and tolerances [26], [29] provides a provision for including statistical tolerances. Currently, mathematically sound definitions of the syntax and semantics of statistical tolerancing are under development for inclusion into standards [27]. An ISO standard for statistical tolerancing is evolving [7].

B. Motivation

Tolerances must be considered early in the design cycle to develop product specifications for quality assemblies that can be produced cost-effectively. However, as described above, existing approaches to design tolerancing entail detailed knowledge of geometry of the assemblies and are applicable mostly during advanced stages of design, thus leading to a less than optimal design process. During the design process of assemblies, both the assembly structure and associated tolerance information evolve continuously. Therefore, significant gains can be achieved by effectively using this information to influence the design of the assembly. The success of Design for X concepts has established beyond doubt the efficacy of providing feedback on downstream manufacturing concerns. Motivated by this, we identify and explore two goals for future research that we believe can enhance the scope of tolerancing to the entire design process. The first goal is to advance tolerancing decisions to the earliest possible stages of design. This issue raises the need for effective representation of tolerancing information during early stages of design and for effective assembly modeling. These assembly models and tolerance representations should enable the designer to incrementally understand the buildup or propagation of tolerances and optimize the layout, features, or assembly realizations so as to ensure ease of tolerance delivery. The second goal addresses the appropriate, synergistic use of available methods and best practices for tolerance analysis and synthesis, at successive stages of design. Pursuit of these goals leads to the definition of a multilevel approach that enables tolerancing to be addressed at successive stages of design in an incremental, continuous, ongoing fashion.

C. Contributions and Outline

The primary contribution of this paper is to propose a multilevel approach to design tolerancing, which we call design for tolerance, to enable tolerancing to be addressed at successive stages of design in an incremental, continuous fashion. The proposed approach integrates three design-related domains.

1) Design activities at successive stages of design.
2) Assembly models for tolerancing that evolve continuously during the design process.
3) Methods and best practices for tolerance synthesis and analysis.

Fig. 1 shows a preview of the three major threads in the proposed methodology. A detailed description of this exhibit appears in the rest of this paper.

Though the investigations here emphasize electro-mechanical assemblies, much of the discussion is relevant for more general assemblies as well.

The paper is organized as follows. Section II presents an example of a chassis-like mechanical assembly and helps motivate the objectives of this paper. Different stages of its design process are delineated, from a tolerancing perspective, to bring out the need for and potential of an integrated, incremental approach to design tolerancing.

In Section III, we look into existing and emerging assembly modeling approaches that are appropriate to use during different stages of design. The leftmost part of Fig. 1 summarizes the assembly models for tolerancing. First, we survey assembly representations based on solid models, relational models, hierarchical models, and datum flow chains. Next, we investigate how these assembly models are useful for design tolerancing at different stages of design. We then identify the requirements and capabilities of an ideal model of assembly for tolerancing that can be used through successive stages of the design process.

Section IV is devoted to a brief survey of methods and best practices for design tolerancing. See Fig. 1, rightmost part, for a preview of the methods and best practices.

In Section V, we present a four-level, integrated approach for incremental and continuous tolerancing through successive stages of design. First, we establish a broad framework for assembly design process by looking into several candidate viewpoints in the literature. The middle part of Fig. 1 shows
this multilevel design for tolerance process. Next, we describe the four levels of the design for tolerance process and establish the coupling between these levels, the assembly models, and the tolerancing methods and best practices.

In Section VI, we consider a simple, representative example and delineate the major steps of our approach. Section VII concludes the paper with a statement of what lies ahead and the potential implications of this work.

II. MOTIVATING EXAMPLE

In this section, we present an example of a mechanical assembly. This example is chosen to illustrate the significant potential of using tolerancing considerations at successive stages of design. There are several examples in the literature that describe various ways in which tolerancing considerations can be used during early stages of design. For example, several case study articles in [30] describe tolerance related decisions at different stages of design. In [31], Altschul and Scholz discuss the tolerancing issues that arise when assembling a cargo door to an airplane body. When the cargo door, fitted with several hinges, is assembled to an airplane body, tolerancing problems could result, necessitating a careful tolerance analysis to be done. Problems such as how many hinges to use and how many gaps and lugs to have in a hinge also have tolerancing implications and represent decisions during early stages of design. More recently, Whitney [6], has provided several examples of illustrative assembly scenarios where tolerancing comes to play a decisive role in early stages of design.

Here, we present a simple and illustrative assembly example, give a rough sketch of its design process, and bring out the important role tolerancing considerations can play in successive stages of its design.

A. Chassis-Like Mechanical Assembly

We consider a simple chassis-like mechanical assembly comprising two major subassemblies—a lower subassembly (main body) and an upper subassembly (cover). (See Fig. 2). The lower subassembly comprises an envelope E and three parts A, B, and C to be assembled into the envelope. The upper cover is the subassembly D, which is designed to fit into the lower subassembly. Fig. 2 is intended to depict only a conceptual view of this assembly; the form shown is not to be viewed as implying any geometry or shape.

1) Assembly Response Functions: Let \( l_a, l_b, \) and \( l_c \) be the lengths of the parts A, B, and C, respectively; \( l_e \), the length of the inner boundary of the envelope E; and \( l_1, l_2 \), the lengths of the left arm and the right arm, respectively, of the cover.
The exact expressions for the gaps get decided as appropriate design decisions are taken. The order in which the gaps are established is also decided by the design process. For example, if part A is assembled to the envelope E first, the gap \( g_{ea} \) is established. If part B is assembled next, followed by part C, then gaps \( g_{ab} \) and \( g_{bc} \) are realized in that order. The gap \( g_{ce} \) is then automatically decided by the expression

\[
g_{ce} = l_c - l_a - g_{ea} - l_b - g_{ab} - l_c - g_{bc}.
\]

Thus, for the considered sequence of assembly, gaps \( g_{ea}, g_{ab}, \) and \( g_{bc} \) are decided in that order and the gap \( g_{ce} \) is dependent on the first three gaps. The first three gaps are decided essentially by the accuracy and process capability of the involved assembly steps or fixturing processes. The order of appearance of the terms in the above equation is important since it reflects the assembly sequence.

Another important issue is the level of detail of an assembly response function. For example, consider the function \( Y_1 = g_{ab} - l_1 \). At an early stage of design, this requirement may be adequate enough. Later in the design process, however, one may be interested in more details. For example, a left clearance and a right clearance may be specified while assembling the left arm into the gap \( g_{ab} \). Thus an assembly response function can evolve through the design process. Tolerance decisions during early design are based on aggregate or approximate versions of the response function. Another related issue is the progression from linear dimensions to complex 3D geometries as design matures. For instance, during early design, we may deal with \( l_1 \) and \( g_{ab} \) as linear dimensions, but as the design process unfolds, these variables can assume a nonlinear or 3D form. This again is caused by the evolution in the assembly response function.

2) Design of the Assembly Process: We now give a rough sketch of how the above assembly may be designed from an early conceptual stage and bring out the relevance and potential of tolerance-related decisions at different stages of the design process. A more generic description of the design process for electro-mechanical assemblies appears in Section V. The design will start with planning of the product, conceptualization, and generating the engineering specifications for the parts and the assembly. Since the lower subassembly and the upper subassembly are separate units, their design can proceed separately and in parallel. There is no need for designers to commit to any geometry during these early stages of design. The expressions for the assembly response functions \( Y_1 \) and \( Y_2 \) can be formulated very early in the design process, whereas the expressions for \( Y_3 \) and \( Y_4 \) can only be formulated later, as explained already. However, the assembly criteria \( Y_i \geq 0 \) for \( i = 1, 2, 3, 4 \) are known during early design itself. Note that \( Y_3 \) and \( Y_4 \) are related to the lower subassembly, while \( Y_1 \) and \( Y_2 \) are concerned with the interface between the two subassemblies.

Let us focus on the lower subassembly. We present four levels of decisions with respect to this subassembly, each more downstream than the previous one in the design process.

Selecting a Configuration Call the lower subassembly P. Fig. 3 shows three possible ways of configuring the four parts.
A, B, C, and E into P—there could be other configurations as well. In Configuration 1, all four parts are treated as individual parts and the assembly takes place in stages. In Configuration 2, P is composed of E and a subassembly that consists of parts A, B, and C. The motivation for considering this latter configuration could be that the subassembly is available off-the-shelf from a known supplier. Likewise, Configuration 3 is another candidate. In this case, the subassembly composed of A and B might be available from a different supplier. It is clear that the process capabilities and the associated parametric variations of the parts and subassemblies will influence the choice of configuration. The selection of one of the above three configurations could be based on how well the configuration enables proper fitting of the parts inside the envelope. Such decisions certainly need not wait until late in the design process.

Selecting Location Logic In this stage of design, our interest is in deciding the manner in which parts are located with respect to one another (location logic). Fig. 4 shows three candidate location logics. Candidate 1 corresponds to a location scheme where A and C are first located (in some order) with respect to a datum on the envelope E and B is next located relative to A and C. This scheme can be realized through the assembly sequence E → A → C → B or the sequence E → C → A → B. In general, a given location logic can be translated into several assembly sequences, thus location logic can be decided earlier than the assembly sequence. Both assembly sequences here are such that component B is assembled last. In Candidate 2 logic, B is the first one to be assembled into the envelope, whereas Candidate 3 logic corresponds to those sequences in which B is assembled in the middle between A and C (these two in any order). The directed acyclic graphs in Fig. 4 are called datum flow chains [5], [6]. They are described in more detail in Section IV. From the conceptual diagram of Fig. 2, it is clear that Candidate 1 may necessitate A and C to have two mating features; Candidate 2 may entail just one assembly feature each on A and C; and Candidate 3 may require either A or C to have two features while the other may have just one feature. One can evaluate, using simple probabilistic arguments and appropriate process capability data, these candidates based on ease of tolerance delivery. For example, Candidate 3 is likely to be better if there is high uncertainty in the dimension of B. The computation here would involve finding the probability that the assembly response functions \(Y_i\) lie in the desired tolerance zones. But once a candidate logic is selected, only those assembly sequences that satisfy that logic need to be pursued further, thus making the design process efficient.

Selection of Assembly Sequence Let us say Candidate 3 was chosen for location logic in the previous step. Then there are two possible assembly sequences: E → A → B → C or E → C → B → A. These two sequences could differ with respect to ease of tolerance achievement. Using the data available about the nominals, tolerances, and process capabilities for the individual parts, one can compute the probability that \(Y_i \geq 0\) for \(i = 1, 2, 3, 4\), and decide which sequence is better. For example, if A has more variability than C, then the second sequence is likely to be better than the first, since the higher variation of A can be transferred to where it is not important. In this case, this is intuitively clear but in complex assemblies, one necessarily needs to carry out such analysis.
Detailed Analysis and Synthesis When the design process reaches advanced stages, tolerance analysis and synthesis can be done in a detailed way. For example, given the assembly sequence; detailed specification of nominals and tolerances for A, B, C, and E; and detailed process capability data, one can compute the probabilities associated with each of the four conformance criteria. Also, detailed synthesis or design can be done. This could take one of three forms: optimize nominal dimensions; optimize tolerances; and establish a variance pool that can be distributed across the individual parts.

B. Need and Potential for an Integrated Approach

The discussion above has brought out the following issues.
1) Continuous evolution of assembly structure and tolerancing information during the design process.
2) Close coupling between the design process and tolerancing decisions.
3) Availability of a variety of assembly modeling methods at different levels of abstraction and relevant for different stages of the design process.
4) Applicability of methods and best practices of design tolerancing to successive stages of the design process.

This motivates the need for and the potential of an integrated approach to design tolerancing that enables tolerancing to be done in a continuous and incremental way.

III. ASSEMBLY MODELS FOR TOLERANCING

We first survey relevant assembly models and next look into how some of the assembly models have been used for tolerancing.

A. Relevant Assembly Models

There are a variety of assembly models available that capture assembly information at different levels of abstraction during the design process and are useful in specific ways. Assembly representations popularly discussed in the literature and applied in practice are based on solid models, relational models, and hierarchical models [6], [32], [33].

The solid models represent part positions in terms of their spatial coordinates. They provide sufficient information for graphic display of the assembly but are not convenient for purposes of tolerancing. For example, changes to the positions or dimensions of individual parts are not always propagated to neighboring parts in the assembly. According to Mantyla [34] and Whitney [6], geometric models of the type used in most solid models have some limitations.

1) They can represent the product structure at only a single level of abstraction and consequently do not support different kinds of analysis at successive stages of the design process.
2) They lack the capability to record the progression of the design process during various phases and thus cannot capture aspects of design intent.
3) Often, they cannot capture the distinction between essential and nonessential information. For example, they do not distinguish between mates and contacts. Mates are connections that pass dimensional and locational constraints from one part to another. Contacts on the other hand are all other connections made to provide strength or reinforcement, but not involved in providing locational constraint [5]. Both mates and contacts are important for tolerancing. Mates represent the interfaces to be controlled whereas contacts represent the sources where variation is transferred during assembly.
4) Changes in shape, geometry, and relative positioning to an individual part are not fully propagated to other parts of the model.
5) Geometric data is only one of several attributes of assembly/product data and does not become available until late in the design process. Many fundamental issues in design can be effectively addressed without having to use geometric data.

Relational models represent geometric relations in the form of mating features between individual parts or subassemblies. They are often called liaison diagrams or connective models of assembly [6]. The assemblies are usually modeled as undirected graphs where the nodes represent the parts and the arcs represent the geometric relations between them. The arcs can have annotations such as P (Part of); A (Attachment); C (Constraint); AS (Assembly), etc. [35]. The actual part or subassembly position can be represented by a coordinate transformation matrix, which is the result of a set of six rigid motions—three translational and three rotational. Fig. 5 provides a relational representation of the assembly of Fig. 2. It contains five nodes and six arcs in the model. Each arc represents a relation or a liaison between the parts or subassemblies at the two ends of the arc. Relational models cannot capture the order in which the geometrical relationships are established. They have been used in analysis applications such as robot path planning, generation of feasible assembly sequences, and robot assembly planning. [32]. Relational models, by themselves, are not adequate for tolerancing.

In a hierarchical model, an assembly is represented as a collection of subassemblies, which in turn are decomposed into individual parts or next level subassemblies. The actual part or subassembly position can be represented by a coordinate transformation matrix, as in the relational model. Though a hierarchical model captures assembly decomposition and aggregate-level precedence relationships in terms of its different levels, it does not assign any hierarchy on the order of establishment of liaisons between individual parts within a particular subassembly. Also, such a hierarchy is yet undeveloped during early design. A tree structure is most appropriate...
4) matrix transforms to

for representing a hierarchical model. Several variants of the hierarchical model have been employed [36], [37], [32], [38], [39]. Fig. 6 shows a simple hierarchical representation of the mechanical assembly of Fig. 2. The hierarchical model has been used in assembly sequence analysis, kinematics analysis, and tolerance analysis (during advanced stages of design).

A recent proposal for assembly modeling with emphasis on early design representation is that of datum flow chains (DFC) [5], [6]. A DFC is a directed acyclic graph that defines the hierarchy of dimensional relationships between parts in an assembly. Each node of a DFC is a part or a fixture or a defined feature on the part or fixture. A directed arc from Node A to Node B indicates that a designated datum corresponding to part A determines the dimensional location of the part B. Dotted lines, if used, (say between nodes B and E) indicate a contact between B and E. Fig. 7 shows a datum flow chain representing a particular way of locating the datums in the mechanical assembly of Fig. 2. Assume that each of the five parts, A, B, C, D, and E have well-defined datums on them. The location scheme in Fig. 7 implies that A and C are first located with respect to E; B is then located in reference to A and C; and the location of D is decided with reference to the locations of A, B, and C.

A DFC abstractly captures the underlying location logic of an assembly and often enables a visualization of the way in which tolerance may propagate. DFC’s can be used early in the design process to represent evolving assembly configurations. They have been shown to be useful in a variety of ways. For example, they can be used to identify important assembly sequence relationships. Also, when sufficient feature-related information is available, they can be used for deriving tolerance chains of assemblies. If a rough geometrical description (so called skeletal geometry) of the assembly is known, these tolerance chains can be used to conduct tolerance analysis [5].

B. Use of Assembly Models for Tolerancing

The models discussed above can potentially be used in many ways, such as assembly sequence analysis, kinematics analysis, and tolerance analysis. Since tolerancing is the main focus of this work, we now look into the use of these assembly models for tolerancing.

Representation of assemblies for automatic generation of tolerance chains has been described by Wang and Ozsoy [38]. Their model combines relational and hierarchical representations; the assembly is represented in an elaborate data structure with information on assembly decomposition; (4 × 4) transformation matrix for each instance of a component/subassembly; mating features; mating conditions (against, parallel, fit); dimensions and tolerances of the mating features; etc. The above information is used to algorithmically generate a tolerance chain for any given assembly. The chain can be used in tolerance analysis. This representation does not need geometric data but cannot be used in early stages of design due to the nature of information required to complete the data structure.

With tolerance analysis as the main objective, Whitney, Gilbert, and Jastrzebski [40] proposed a model of assembly that contains the following information: Mating features that build up the assembly; a graph representation of mating of parts (liaison diagram); underlying coordinate structure of the assembly; and homogeneous (4 × 4) matrix transforms to represent dimensions and tolerances of each part (in accordance with the ASME Y14.5M-1982 geometrical tolerancing standard). The transforms represent both the nominal relations between parts and variations caused by geometric deviations allowed by the tolerances. These transforms can be used to propagate tolerances through an assembly, which allows the location of any designated part to be captured starting from a reference part, taking into account variations in the locations, sizes, and shapes. The above representation can potentially be used in early stages of design.

Datum flow chains have been used to generate tolerance chains for assemblies during early design [5]. The method uses the location logic embedded in DFC’s with skeletal geometry of the assembly known, these tolerance chains can be used to conduct tolerance analysis [5].
assemblies, the knowledge of DFC is sufficient to perform a tolerance analysis. This is because all assembly sequences in a family have identical tolerance chains. Hence, if one assembly sequence fails (succeeds) to deliver the tolerance, so will all others corresponding to that family. In Type 2 assemblies, there is scope for in-process adjustments. So each assembly sequence within a family can produce different results. This would mean that additional information is required to do tolerance analysis.

There are several other papers that have dealt with the problem of assembly modeling in general and assembly modeling for tolerancing in particular. The reader is referred to [32], [37], [41]–[44].

C. Assembly and Tolerance Representation through the Design Process

The following are some important observations about the models for assembly and tolerancing discussed earlier.

1) Different models become available and are relevant, at possibly different stages of the design process. For example, a relational model becomes available earlier in the design process than a hierarchical model. The models discussed (liaison diagrams, trees, datum flow chains, solid models, etc.), when collectively used, cover a broad spectrum of the design process and therefore are useful for tolerancing at different stages of the design process. See the far-left portion of Fig. 1 for a preview of the various assembly models.

2) Different models capture the assembly at different levels of abstraction. For example, datum flow chains model design intent related to location logic at a fairly early stage of design. If suitable positioning information is available, DFC models enable tolerance analysis to be done at that (early) stage of design, leading to the viewpoint of tolerance achievement.

3) Both the assembly artifact and the tolerancing information evolve during the design process through successive refinement. Consequently, an assembly model continuously evolves through some or all stages of the design process. For example, during early design, not all geometric relations or mating features may be known, so a liaison diagram captures only a subset of all ultimate relations. As the artifact undergoes continuous transformation, existing relations may disappear and new relations can appear, leading to more detailed liaison diagram. Whitney [6] gives an example of how a datum flow chain model evolves as the design function progresses. The key to enabling effective tolerancing to be done at successive stages of the design process lies in a robust assembly model that gets modified and refined in a continuous way throughout the design process.

In our view, an ideal assembly model for tolerancing should

1) have a close coupling with the design process;
2) be mutable through successive stages of the design process;
3) be capable of representing the assembly and tolerance information at any level of abstraction.

Other important attributes of an ideal model would be: capture of design intent; embedding of different views (relational view, location logic view, etc.) in a unifying framework; and enabling all assembly information other than tolerancing, also to be captured in the model. This raises the issue of effective, integrated representations of assembly through the design process. Object oriented models are appealing since they enable such integrated representations of assemblies. There are some recent efforts in this direction. The first is the SHARED model [45]–[47], which is essentially an information model for cooperative product design. This is an object-oriented representation that captures both an evolving artifact and its associated design process. To represent artifacts as they evolve, the SHARED model defines objects recursively without any pre-defined granularity on the recursive decomposition, thus enabling the model to be used at any desired level of abstraction. The SHARED model, by virtue of using a single framework to couple the artifact with its design process, provides an attractive paradigm for assembly modeling for continuous tolerancing through the design process. Another effort [48] looks at an object oriented assembly representation that provides a general assembly model that can support both conceptual design at high levels of abstraction and feature modeling at low levels. This is achieved by incorporating functional knowledge and design intent as part of the assembly representation.

The far-left portion of Fig. 1 summarizes the assembly models for tolerancing. It presents certain selected, representative modeling formalisms only. When supplemented with appropriate information, these models are useful for making tolerance related decisions at different stages of design and constitute an important element of the design for tolerance methodology proposed in this paper.

IV. DESIGN TOLERANCING: METHODS AND BEST PRACTICES

Design tolerancing methods and best practices have an important role to play in enabling tolerance-related decisions to be made at successive stages of the design process. As stated earlier, tolerancing includes both tolerance analysis and tolerance synthesis [49]. In the context of electro-mechanical assembly design, tolerance analysis involves evaluating the effect of variations of individual part or subassembly dimensions on designated dimensions or assembly characteristics of the resulting assembly. Tolerance synthesis involves allocation of tolerances to individual parts or subassemblies of an assembly based on the tolerance requirements on the assembly. The far-right portion of Fig. 1 shows a listing of important methods for tolerance analysis and synthesis, and major best practices.

A. Methods for Tolerance Analysis

Tolerance analysis can be either worst-case or statistical. In worst-case tolerance analysis (also called deterministic or high-low tolerance analysis), the analysis considers the worst possible combinations of individual tolerances and examines the assemblability of the parts, so as to achieve 100%
interchangeability of parts in an assembly. This results in unnecessarily tight part tolerances and hence high production costs. Statistical tolerancing is a more practical and economical way of looking at tolerances and works on setting the tolerances so as to assure a desired yield. Here, the designer abandons the notion of 100% interchangeability and accepts some small percent of nonconformance.

Statistical tolerance analysis uses a relationship of the form

$$Y = f(X_1, \ldots, X_n)$$

where $Y$ is the response (a measurable characteristic such as assembly gap) of the assembly and $X_1, \ldots, X_n$ are the values of some characteristics (such as dimensions) of the individual parts or subassemblies making up the assembly. We call $f$ the assembly response function (ARF). The relationship can exist in any form for which it is possible to compute a value for $Y$ given values of $X_1, \ldots, X_n$. It could be an explicit analytic expression or an implicit analytic expression, or could involve complex engineering calculations or conducting experiments or running simulations. The input variables $X_1, \ldots, X_n$ are continuous random variables. In general, they could be mutually dependent. The function $f$ is a deterministic relationship; $Y$ is easily seen to be a continuous random variable. The general problem of tolerance analysis is to compute the probability distribution of $Y$ given the distributions of $X_1, \ldots, X_n$. However, more often we are usually interested in computing the first few moments (or mean, standard deviation, skewness, and kurtosis), given the distributions or first few moments of the input variables. Once the moments of $Y$ are determined, one can compute a tolerance range for $Y$ that would envelope a given fraction of the assembly yield.

There are a variety of methods and techniques available for the above computational problem. Essentially, the methods can be categorized into four classes [13].

1. Stack tolerancing or linear propagation (root sum of squares).
4. Monte Carlo simulation

For more details on these methodologies, we would like to refer the reader to the articles by Evans [13], Chase and Parkinson [15], Nigam and Turner [18], and Narahari et al. [50].

B. Methods for Tolerance Synthesis

In the context of electro-mechanical assembly design, tolerance synthesis usually refers to the allocation of specified assembly tolerances among the constituent parts and subassemblies, so as to ensure a specified yield or minimize a proper cost function. More generally, if $Y = f(X_1, \ldots, X_n)$ is an assembly response function, then the synthesis problem involves finding the best nominals and tolerances for $X_1, \ldots, X_n$, given nominal and tolerance specifications for $Y$. Synthesis is naturally an optimization problem; one can formulate an objective function that captures yield requirements or production cost requirements and pose an optimization problem by including tolerance related constraints.

There are several views and variants of the synthesis problem, depending on the objective function and the constraints. One view is to minimize the total manufacturing cost by choosing the individual part tolerances and the manufacturing processes for making the individual parts. This requires cost versus tolerance relationships for each individual dimension. Another view is to find robust nominals for individual dimensions, i.e., nominal values at which the effect of variations on the assembly response function is minimum. This is the problem addressed by Taguchi’s robust design methodology and HPD. Also, depending on the nature of the objective function and the constraints, the synthesis problem can be deterministic or stochastic.

Major approaches to tolerance synthesis include

1) iterative methods based on tolerance analysis [13], [50];
2) optimization methods which formulate tolerance synthesis as an optimization problem, leading to use of mathematical programming techniques such as linear programming, nonlinear programming, and integer programming, and heuristic techniques for optimization such as simulated annealing, genetic algorithms, Lagrangian relaxation, and Tabu search [13], [19], [50];
3) design of experiments, which uses systematic exploration of the design parameter space using statistical techniques. Taguchi’s robust design methodology [51], [52], [21], which has emerged as a best practice in many companies, uses design of experiments in a novel way.

C. Best Practices

In the last decade, many companies have established comprehensive programs in total quality management. These efforts include those of Motorola (six-sigma program) [20], [53], [54], Xerox (holistic probabilistic design) [55], [56], IBM, AT&T Bell Laboratories, and several others which have initiated formal, corporate programs for improved tolerance specification, monitoring, and control. For example, tolerance analysis and synthesis in the Motorola six sigma program [20], [53], [54] are based on

1) six sigma characterization of products and processes; the process capability indices $C_p$ and $C_{pk}$ are used as the vehicles to characterize the product-process quality;
2) simple, intuitive extensions to the RSS method, to enable tolerance analysis and synthesis in the presence of shifts and drifts of the process mean;
3) a well-defined, systematic program for design for quality, taking into account both the product perspective and the process perspective.

V. DESIGN FOR TOLERANCE PROCESS

We now propose an integrated approach, which we call Design for Tolerance, for enabling tolerancing decisions in an incremental and continuous ongoing fashion during the design of assemblies.
A. Design Process for Assemblies

In the literature, several researchers have presented their viewpoint of what the assembly design process should be. We provide a brief outline of some viewpoints that emphasize tolerancing. Whitney [6] advocates top-down design of assemblies, which uses the method of key characteristics (KC’s) [3]–[5]. Customer requirements or customer tolerances are specified in terms of product key characteristics (PKC’s), which are permanent properties of the design. These flow down to the subassembly and part levels resulting in a set of assembly key characteristics (AKC’s) and a set of manufacturing key characteristics (MKC’s). AKC’s define important dimensional datums, assembly mating features, and fixturing features on parts and assemblies [6]. These are defined in the context of a specific assembly process that is intended to deliver the PKC’s. MKC’s are basically parameters of manufacturing processes that are intended to deliver the AKC’s. Design of the assembly is driven by the KC’s and implemented using datum flow chains. Tolerance analysis forms an important part of verifying whether or not the key characteristics are delivered by the chosen location logic or assembly sequences.

Tolerancing best practices in major companies also advocate their own design processes for assemblies. The Motorola six sigma program [53] has a five-step process.

1) Perform preliminary design: Starting from customer specifications, establish a baseline design and develop a basic configuration. This will involve choosing baseline nominals for important dimensions.
2) Identify process variabilities.
3) Assign tolerances to related dimensions.
4) Compute the probability of conformance for each assembly gap and assembly response measure.
5) Optimize the probability of conformance for each assembly gap and assembly response measure. This may involve finding optimal nominals, determining best tolerances, and distributing the overall assembly variation among individual parts of the assembly. Confirm six sigma quality with respect to all the assembly gaps and assembly response measures.

The Xerox HPD methodology recommends the use of critical parameters that are derived from customer specifications and customer tolerances [56]. The critical parameters are similar to key characteristics. The critical parameters are systematically related to piece-part variabilities through flow-of-variance chains. Tolerance analysis and synthesis involve choosing the piece part variabilities so as to yield the customer desired tolerances for all the assembly response measures.

Taguchi’s robust design process follows a three-step approach [21]: system design, parameter design, and tolerance design. In system design, a basic functional prototype is designed after understanding the customer’s needs and the manufacturing environment. In parameter design, settings of product or process parameters that minimize the sensitivity of designs to the sources of variation are obtained. These settings are called robust nominals. In tolerance design, tolerances around the robust nominal settings are determined.

B. Design Tolerancing: An Incremental Process

The SIMA (Systems Integration for Manufacturing Applications) reference architecture formulated at the National Institute of Standards and Technology [57] provides a generic specification of design related activities for electromechanical products. Fig. 8 shows the various design stages and activities in the SIMA reference architecture. Stage A11 (Plan Products) involves developing the idea for the assembly depending on market needs and customer requirements and characterizing it in terms of function, target price range, and relationship to existing product lines. In Stage A12 (Generate Product Specifications), an engineering specification for the assembly is formulated. This involves mapping the customer requirements into engineering requirements and refining these in consideration of the relevant laws, regulations, patents, and product standards, etc. In Stage A13 (Perform Preliminary Design), the assembly design problem is decomposed into a set of component/subassembly design problems and specifications are developed for each component/subassembly problem. Interface specifications between the components/subassemblies are developed and a preliminary assembly layout is created. Finally, in Stage A14 (Produce Detailed Designs), all specifications needed to completely describe each subassembly or component are produced. This includes drawings and geometry, materials, finish requirements, assembly drawings, and fit and tolerance requirements.

There are several commonalities in the SIMA reference architecture and the assembly design processes outlined earlier. The design for tolerance process proposed in this paper embodies many of these ideas in the broad framework of the SIMA architecture, with emphasis on tolerancing.

![Fig. 8. Design stages and activities in the SIMA reference architecture.](Image)
detail. Also, the decisions taken at a particular stage influence and can simplify the decisions taken in the downstream stages. Like other attributes of a product design, tolerance information changes over time, through successive stages from product planning to detailed design through on-going production. Hence a robust tolerance representation would be mutable and directly related to the design process representation. The incremental refinement of processes and tolerance representations proceeds in symbiotic fashion. Consider, for example, a tooling design/build process. Both lead time and cost for tooling is often highly dependent on the tightness of a tolerance requirement. Scheduling of rough cutting for a die or mold can typically proceed prior to a final tolerance specification, but the finish cut, polishing, etc. must proceed afterward. Conversely, tolerance specification for a snap-fit in a high-precision injection-molded part must be preceded by a decision about assembly process (e.g., manual or robotic). For complex assemblies with many parts, the timing and precedence of tolerancing decisions can profoundly affect scheduling and total lead time. Analysis and synthesis for critical tolerance stack-ups is clearly related to process plan refinements. There are opportunities to compress cycle time by improved modeling prior to detailed design, but compatible, incrementally-refined representations of tolerances and processes are the key.

The incremental and continuous, ongoing nature of the process of tolerance decision making enables a natural aggregation/decomposition of tolerancing activities as the design matures. Another way of viewing this is in terms of the pruning that this causes at successive stages in the space of feasible solutions to the design problem. Early on in the design process, the solution space has a staggering cardinality and the tolerancing decisions, if taken in a continuous ongoing fashion, can lead to substantial early reduction in the space of possible solutions thus making the design process efficient. Another alternative view is in terms of marked reduction in design iterations or design rework. In this sense, design for tolerance is similar in spirit to design for manufacturing/assembly [58] that also has the effect of dramatically shrinking the space of solutions and reducing iterations. Furthermore, DFA, DFM, or such other design related strategies may have close coupling with tolerance related decisions and may both influence and be influenced by tolerancing at various stages.

C. Design for Tolerance: A Multilevel Approach

The first two stages A11 and A12 of the SIMA reference architecture and also the early stages of other assembly design processes (top-down design, Motorola process, Xerox HPD process, and the robust design process) essentially involve mapping customer requirements into product ideas and product specifications. Tolerancing is not directly involved in these early stages, except in very abstract terms; however, these stages provide critical inputs to the tolerancing decisions in the rest of the design process. See Fig. 9.

Thus we focus on Stage A13 (Perform Preliminary Design) and Stage A14 (Produce Detailed Designs) of the SIMA reference architecture. We divide these stages into the following four tolerance-related levels (TR Level) and develop a four-level approach to design tolerancing. Note the difference between SIMA stages and tolerance-related levels here.

1) SIMA Stage A13: Perform Preliminary Design:
   a) TR Level 1: Assembly Layout and Configuration;
   b) TR Level 2: Location Logic and Assembly Features;
   c) TR Level 3: Assembly Planning and Sequencing.

2) SIMA Stage A14: Produce Detailed Designs:
   a) TR Level 4: Detailed Tolerance Analysis and Synthesis.

These levels are fairly representative and generic for electromechanical assemblies. Neither the number of levels nor the description of the individual levels is to be viewed as being definitive. Fig. 9 captures the essence of this architecture for DFT.

1) TR Level 1: Assembly Layout and Configuration: Once the product concept is known and engineering specifications are generated based on the key characteristics, TR Level 1 of the proposed process can commence. TR Level 1 involves decisions regarding the preliminary assembly layout/configuration. Such decisions may include: rough allocation of space, number of subassemblies, the configuration of critical subassemblies, grouping of components into subassemblies, and rough layout of the assembly. The information available at this level can be described in the form of a liaison diagram (relations between parts or subassemblies), a tree (assembly decomposition), and a partial DFC (to capture whatever location logic is known at this point). Candidate layouts or configurations can be identified and represented using these models. These layouts or configurations and related manufacturing process selection typically might differ in terms of ease of tolerancing. The tolerancing considerations here are at a coarse level and may be directly influenced by customer specifications. To effect such high level tolerancing decisions, aggregate level manufacturing process capability data will be required and is often available at this point. Simple statistical assumptions and probabilistic calculations can be used at this stage. Also, for problems such as manufacturing process selection, optimization formulations can be used.

2) TR Level 2: Location Logic and Assembly Features: At the next level (TR Level 2), the following information is assumed to be available: assembly response functions (approximate); tolerance requirements at interfaces between major subassemblies and parts; and relevant process capability data. The decisions here are concerned with the location logic (how to locate subassemblies and components with respect to one another) and with choosing the appropriate assembly features to go with the location logic. The choice of features itself might depend on the assembly sequence (not the detailed sequence but a precedence specification among major assembly steps). The DFC model is suitable to capture the available/evolving assembly information here. There is close coupling among selection of features, selection of assembly sequence, and creation of DFC. Assembly models such as liaison diagrams are also relevant here. If the assembly is of Type 1, then the assembly features are predominantly decided by the functional requirements; if the assembly is of type 2, then the choice of assembly features is an important problem by itself. In the
latter case, the DFC alone will not be adequate to conduct a tolerance analysis. A more detailed model that captures the tolerance flow at this level will be required. Tolerance analysis here can tell us which location logic is better from a tolerancing viewpoint and which set of assembly features would best accomplish tolerance achievement. This stage might also help us to find preliminary target values and tolerances for individual parts.

Statistical tolerance analysis methods listed in Fig. 1 are all relevant here. Determining robust nominal values and preliminary settings of tolerances can be accomplished using Taguchi methods or Xerox Holistic Probabilistic Design methodology [56].

3) TR Level 3: Assembly Planning and Sequencing: We proceed next to TR Level 3 where the detailed assembly response function, detailed process capability data, skeletal geometry of the assembly, assembly features, and, specification of parametric or geometric tolerances of individual parts and features are assumed to be known. From the tolerance specification, one may derive \((4 \times 4)\) matrix transforms for the nominals and variabilities associated with the parts [40]. The decisions here could be with respect to the selection of the detailed assembly sequence that achieves the required tolerance specifications in the best possible way. The models that we employed in the previous stage, like DFC and liaison diagrams, can again be used here. In fact, they are now updated with richer and more detailed information. This kind of representation and analysis is presented in [38], where several data structures to capture tolerance related information are presented. With the information available here, one can also carry out tolerance synthesis.

4) TR Level 4: Detailed Tolerance Analysis and Synthesis: TR Level 4 corresponds to the detailed assembly design stage. Here, the complete assembly sequence is known; geometric data about the parts and features is available; detailed part level tolerance requirements are known; the assembly response function is available in complete form; and low level process capability data is accessible. Detailed tolerance analysis and synthesis can be carried out here. Most tolerancing studies and tolerancing tools available support this level of design.

D. Design for Tolerance: An Integrated Approach

The multilevel approach to design tolerancing integrates the design process, the assembly models for tolerancing, and the
tolerancing methods and best practices. This is captured by Fig. 9.

1) Design Process: The proposed design process follows the SIMA framework and has four stages: Plan product, generate specifications, perform preliminary design, and produce detailed designs. We have focused on tolerancing decisions during preliminary design and detailed design stages and proposed a four-level approach. It is to be noted that each level above is iterative both internally (feedback within a level) and across (feedback from a given level to a previous level).

The design process delineated here is focused on tolerancing. There are many other subprocesses of the design process that address important issues such as design for assembly, design for manufacturability, design for reliability, etc. All these processes are concurrent, cooperative, and often competitive. A thorough discussion of this is beyond the scope of this paper.

2) Assembly Models: As described in Section III, there are many assembly modeling approaches that capture the assembly at different stages and at different levels of abstraction. Successive levels of the design for tolerance process will need one or more of these models. The design process evolution is accompanied by a continuous refinement of the assembly models and the tolerancing information.

3) Tolerance Analysis and Synthesis: At successive levels of the DFT process, different kinds of tolerancing decisions need to be taken. These could vary in complexity from simple probabilistic calculations to complex and elaborate computations. As described already, there are a variety of methods and best practices for tolerancing. Which method or best practice to employ at a given level of the DFT process needs careful thought and can depend on a variety of factors such as the product domain, nature of the assembly response function, number of variables involved, and completeness of information.

VI. EXAMPLE

Recall the mechanical assembly example of Fig. 2. As stated in Section II, the diagram is conceptual and is not to be viewed as implying any geometry or shape. The conformance or functionality of the assembly is decided by the following criteria:

C1. \( Y_1 = g_{ab} - l_1 \geq 0 \).
C2. \( Y_2 = g_{bc} - l_2 \geq 0 \).
C3. \( Y_3 = g_{ea} \geq 0 \).
C4. \( Y_4 = g_{ec} \geq 0 \).

In the above expressions, the tolerance constraints are expressed in terms of linear dimensions. This is because, the gaps and the lengths are 1-D quantities. Therefore the tolerance zone in each case is an interval around the nominal length or nominal gap. More generally, if \( g_{ab} \) and \( g_{bc} \) represent complex geometrical gap elements, and \( R_{ab} \) and \( R_{bc} \) represent the tolerance zones for \( g_{ab} \) and \( g_{bc} \) respectively, the criteria C1 and C2 above can be expressed as:

\[
\begin{align*}
g_{ab} & \in R_{ab} \\
g_{bc} & \in R_{bc}.
\end{align*}
\]

The tolerance zones \( R_{ab} \) and \( R_{bc} \) will have the lengths \( l_1 \) and \( l_2 \), respectively, among their parameters. For the sake of simplicity, we shall consider here only parametric tolerances. Consequently, the tolerance zones become intervals. The discussion is similar for geometric tolerances, with appropriate extensions and reinterpretation.

We now discuss how tolerance related decisions can be taken at the four levels of the design for tolerance process (Fig. 9).

A. Selecting A Configuration

Fig. 10 shows three possible ways of configuring the five parts A, B, C, D, and E as product P; there could be other configurations as well. In Configuration 1, all five parts are treated as individual components and the assembly takes place in stages. In Configuration 2, P comprises E, D, and a subassembly that consists of components A, B, and C. The motivation for considering this configuration might be that the subassembly is available off-the-shelf from a known vendor. Likewise, Configuration 3 is another candidate. In this case, the subassembly comprising A and B might be available from a different vendor. It is clear that the process capabilities and the associated parametric variations of the components and subassemblies will influence the choice of the configuration.

To decide which of the above three configurations is best from a tolerancing viewpoint, we need to determine how well the criteria C1, C2, C3, and C4 are met by the configurations.
A natural way of doing this is to compute the probabilities

\[ p_1 = \Pr\{g_{ab} \geq l_1\} \]
\[ p_2 = \Pr\{g_{ec} \geq l_2\} \]
\[ p_3 = \Pr\{g_{ca} \geq 0\} \]
\[ p_4 = \Pr\{g_{ce} \geq 0\} \]

The following data is known about these configurations.

1) In the case of Configuration 1, the random variables \( l_a, l_b, l_c, l_e \), and \( l_2 \) are known (available from the vendors supplying these components or from local factory data). This means we know either the probability distribution or at least the first few moments of each random variable. The gaps \( g_{ab} \) and \( g_{ce} \) are not known since they depend on the assembly process. Similarly, the gaps \( g_{ca} \) and \( g_{ce} \) are also not known since they also depend on the assembly process. In fact, \( g_{ca} \) and \( g_{ce} \) are related by the following equation:

\[ l_e = g_{ca} + l_a + g_{ab} + l_b + g_{ce} + l_c + g_{ce} \]

If the assembly sequence is such that \( g_{ca} \) is decided first (that is, \( A \) is assembled to \( E \) earlier than \( C \)), the above equation can be used to determine \( g_{ce} \) (provided \( g_{ab} \) and \( g_{ca} \) are known). On the other hand, if \( g_{ce} \) is decided first, \( g_{ca} \) can be determined using the above equation.

2) In respect of Configuration 2, the following are known: The lengths \( l_a, l_b, l_c, l_e, l_1, l_2 \), and the gaps \( g_{ab} \) and \( g_{ce} \). The gaps \( g_{ca} \) and \( g_{ce} \) depend on the assembly process. Since \( l_1 \) and \( g_{ab} \) are known, the probability \( p_1 \) can be computed. Similarly, the probability \( p_2 \) can be computed since \( l_2 \) and \( g_{ce} \) are known.

3) In the case of Configuration 3, all the length-related random variables are known, while among the gaps only \( g_{ab} \) is known. Thus we can compute \( p_1 \) but not \( p_2 \).

The key to selecting the best among these configurations lies in choosing an important subset of criteria (probabilities) on which to base the decision, and then the ability to compute the chosen probabilities without bringing in assembly sequence or other downstream concerns. In the present case, it is reasonable to base the decision on \( p_1 \) and \( p_2 \), ignoring \( p_3 \) and \( p_4 \). To compute \( p_1 \) and \( p_2 \) for the above three configurations, we proceed as follows. It is straightforward in the case of Configuration 2, as already explained. In respect of Configuration 3, \( p_1 \) can be computed easily as explained above. To compute \( p_2 \), \( g_{ce} \) can be assumed to be the same as for Configuration 2 (this makes the comparison fair). As for Configuration 1, \( g_{ce} \) can be assumed to be the same as in Configurations 2 and 3: \( g_{ab} \) can be assumed to be either the minimum of the values of this gap for Configurations 2 and 3 (optimistic) or maximum of the values (pessimistic).

Having chosen a particular configuration (say configuration 1), another important decision needs to be taken. This concerns the supplier selection or manufacturing process selection. If the components \( A, B, C, D, \) and \( E \) are being supplied by two separate vendors and the components have differing specifications and costs, then which supplier to choose is an important question that can again be partially resolved by computing the probabilities above. Here, cost considerations also become important. If there is a wider choice of suppliers and each supplier has multiple offerings, the problem can be resolved by design of experiments or Taguchi methods, with a carefully chosen cost function. Another important decision concerns the manufacturing process selection. Here, the problem is to choose the best combination of manufacturing or assembly processes to make the components and assemble them, so as to satisfy tolerance requirements and minimize manufacturing/assembly cost. This can be solved as an optimization problem (see, for example, the integer programming formulation in [19]).

B. Selecting Location Logic and Assembly Features

In this stage of design, our interest is in fixing the location logic, which often allows the choice of assembly features. Fig. 11 shows four candidate DFCs; there could be other candidates as well. In Candidate 1, \( A \) and \( C \) in some order are first assembled into \( E \) and then \( B \) is located with respect to \( A \) and \( C \). Next, \( D \) is assembled with respect to \( A, B, \) and \( C \) to yield the proper gaps. In Candidate 2 logic, \( B \) is the first one to be assembled into the envelope, followed by \( A \) and \( C \) in some order and thereafter, \( D \) is assembled. Candidate 3 assemblies correspond to those sequences in which \( B \) is assembled in the middle between \( A \) and \( C \) (these two in any order). Note that \( D \) is assembled last in candidate logics 1, 2, and 3. In candidate 4, a fixture \( F \) can possibly be used to hold \( D \) and then \( A, B, \) and \( C \) are properly located with reference to the position of \( D \). Finally assembled to hold \( A, B, C, \) and \( D \). The use of a fixture is motivated by higher positioning accuracies that can possibly be achieved with well-designed fixtures. From the conceptual diagram of Fig. 2, one can also visualize how a particular location logic can influence the nature and choice of mating features.

To compare the above four candidates, we need to compute the probabilities \( p_1, p_2, p_3, \) and \( p_4 \). Recall that we know the distributions of \( l_a, l_b, l_c, l_1, l_2, \) and \( l_e \). The distributions or moments of the gaps are now to be computed knowing the location logic and relevant process capability data. For instance, consider candidate 1.

1) Since \( A \) and \( C \) are first assembled into \( E \), the distribution or moments of \( g_{ca} \) and \( g_{ce} \) can be computed first (assumed to be assembled first). The probabilities \( p_2 \) and \( p_4 \) can then be computed. These computations will need process capability data about the assembly operations.

2) Next, \( B \) is placed inside the envelope. Knowing the process capability data for this operation, we can compute the distributions or moments of \( g_{ab} \) and \( g_{ce} \).

3) Finally, knowing the process capability of assembling \( D \), the probabilities \( p_1 \) and \( p_2 \) can be computed.

We may remark that Candidate 2 is likely to be the best since it enables variation to be transferred to where it is not important. On the other hand, if there is high variability in the dimension of \( B \), then Candidate 3 may turn out to be a better choice. Also, note that design for assembly considerations may negate the choice of Candidate 1 for the reason that assembly may be difficult to achieve since component \( B \) is to be juxtaposed between \( A \) and \( C \), providing for the desired gaps.
Similar Statistical computations can be carried out using the
tolerance analysis methods of Section IV-A. Best practices,
such as the Motorola six sigma program and the Xerox Holistic
probabilistic Design, are also suitable for such computations.

If the parts are 3-D, then instead of linear dimensions
as above, more general methods will have to be used. If
(4 × 4) matrix representation is available for the parts and
their tolerances, then the matrix transforms can be used in the
computations [40].

C. Selection of Assembly Sequence

Each candidate logic can correspond to multiple assembly
sequences, thus selecting a sequence occurs later than selecting
a location logic. We observed in the previous subsection that
Candidate 3 is likely to be better if there is high uncertainty
involved in the dimension of B. If this candidate is chosen,
then there are two possible sequences: E → A → B → C
→ D or E → C → B → A → D. If this assembly were of
Type I, then, as observed previously, (in Section 2), both these
sequences will result in the same tolerance chain and hence the
same values for the probabilities

However, if we regard this as a Type 2 assembly (that is, features are
formed during the assembly process with the use of fixtures),
then the two sequences could differ with respect to tolerance
achievement. Using the data available about the distribution of
the lengths, and process capabilities for the individual parts,
one can compute the probabilities

and decide which sequence is better. For example, if \( l_\alpha \) has more
variability than \( l_\gamma \), then the second sequence is likely to be
better than the first sequence, since the higher variation of \( l_\alpha \)
can be transferred to where it is not important. In this case, this
is intuitively clear but in complex assemblies, one necessarily
needs to carry out such analysis.

D. Detailed Analysis and Synthesis

When the design process reaches advanced stages, tolerance
analysis and synthesis can be done in a comprehensive way
since we have access to to detailed data.

1) Tolerance Analysis: For example, the following information
may be known.

1) Assembly sequence: Say, E → B → A → C → D.
2) Distributions of \( l_\alpha, l_\beta, l_\gamma, l_\delta, l_\epsilon, l_\zeta, l_\eta, l_\theta \), and \( l_\iota \), or alternatively
their nominals \( N_\alpha, N_\beta, N_\gamma, N_\delta, N_\epsilon, N_\zeta, N_\eta, N_\theta \), and \( N_\iota \), and corresponding tolerances \( T_\alpha, T_\beta, T_\gamma, T_\delta, T_\epsilon, T_\zeta, T_\eta, T_\theta \), and \( T_\iota \).
3) Process capabilities of different assembly steps in the
assembly sequence.

First, B and A are located on the envelope, leaving the right
amount of gap \( g_{dk} \). Knowing the \( C_p \) and \( C_{pk} \) of this step,
the probability \( p_1 \) can be computed. Also, it is easy to see
that \( p_2 = 1 \). The next operation is to locate and place the
component C so as to get the correct gap between B and C
and also avoid interference between C and E. One can then
compute the probabilities \( p_2 \) and \( p_{13} \), knowing the appropriate
process capability data.

Here again, either statistical tolerance analysis methods
could be used. Very detailed analysis can be done using Monte
Carlo simulation.

The discussion above has again assumed linear dimensions
and tolerances. If the geometry of the individual components
and the assembly are known, then one can specify the data in
terms of the ANSI standard on geometric tolerances and use
(4 × 4) matrix transforms and repeat the above computations.

2) Design: Design or synthesis can assume several forms,
see for example, Harry and Stewart [53]. The possibilities
include: optimization of nominal dimensions; optimization
of tolerances; and optimal allocation of overall assembly
variation across individual parts.

Let us say the desired probability of nonconformance is
3.4 ppm, as in the Motorola six sigma program. If A, B, C,
and D are from external vendors and all appropriate data is
known (nominals and either tolerances or standard deviations)
for those, then for a given tolerance \( T_e \) of the envelope,
one can determine the nominal value \( N_e \) so as to assure the
required probability of conformance. This can be done both
optimistically (no shifts in the process mean) and realistically
(in the presence of shifts in the process mean).

Fig. 11. Datum flow chains representing different location logics.
Using Taguchi methods or Xerox Holistic probabilistic Design, one can determine robust nominals for all the parts involved, that is, the combination of nominals of the individual parts for which the effect of variations is minimized.

On the other hand, if all relevant data for A, B, C, and D is known, and the nominal $N_c$ of the envelope is fixed, one can determine the tolerance $T_c$ of the envelope so as to achieve a probability of nonconformance of say, 3.4 ppm. Here we are determining the capability of the process that fabricates the envelope.

VII. DISCUSSION AND FUTURE WORK

In this paper, we have outlined a continuous, multi-level approach to design tolerancing of electro-mechanical assemblies. The architecture integrates three main elements: Assembly models for tolerancing; methods and best practices for tolerancing; and the evolving design process. We have delineated a four level approach for incremental design tolerancing and illustrated the methodology for a simple, representative, mechanical assembly. The discussion has centered on parametric or linear tolerances. Extension to functional tolerancing is straightforward since the analysis and synthesis methods can handle arbitrary, possibly nonlinear, functional relationship between the individual piece-part characteristics and the assembly response. Also, by suitably defining tolerance zones, extension to geometrical tolerances is possible. Since the ultimate test of any such methodology is in successful application to industry-level products, a logical next step would be to look into industry-level implementation of the proposed approach. There are two important directions for further work on this topic. These are: implementation of a DFT environment and facilitation of standards development.

A. Implementation of a DFT Environment

As Fig. 9 suggests, computer implementation of an automated design for tolerance environment will involve integrating together the assembly models and the tolerancing techniques with the design process. Tolerance analysis and tolerance synthesis during the assembly design stage affect the design process in an influential way and therefore lead to a better understanding and formulation of the design process. Since improvements to the design process require effective modeling of the process, the proposed work will offer valuable insights into process modeling. For example, as tolerance-related information becomes available in increasing detail during the design process, both the process and product representations undergo successive refinements. This needs to be captured by the model. A number of methods have been proposed over the years to model design processes. However, these methods have several inadequacies. An ideal process model should enable faithful modeling of precedence relations, constraints, iterations, side effects, dependencies, abstraction, cost factors, and time-to-market determinants [59], [60]. The proposed work will help understand the process modeling requirements for assembly design. The work also raises interesting issues such as finding an integrated representation formalism for assembly modeling and also design process modeling. As already stated in Section III-C, object oriented models can form the foundation of such integrated product-process models.

B. Standards Development

It is expected that the proposed work on assembly modeling and assembly representation will provide preliminary specifications that can serve as the basis for assembly standards. The current standard (AP203) only allows the representation of an assembly as a collection of 3-D objects positioned and oriented in space. It does not make any provision for the capture of logical relationships between parts, mating feature relationships, part functionality, kinematic degrees of freedom, and tolerance information. The work here will provide useful inputs to the development of such a standard.

Dimensional tolerancing has mostly evolved as an industrial practice without strong theoretical foundations [16]. The best tolerancing practices were collected and made available through an evolving series of tolerancing standards [24]–[27]. All international, and most national, standards have codified only classical tolerancing [7]. The Deutsches Institut für Normung-German Institute for Standardization (DIN) standard issued in Germany [61] was a serious attempt at standardizing statistical tolerancing. The latest ASME Y14.5M-1994 standard on dimensioning and tolerances [26] provides a provision for including statistical tolerances. Currently, mathematical definitions of the syntax and semantics of statistical tolerancing are under development for inclusion into standards. An ISO standard for statistical tolerancing is evolving [7]. Improved understanding of the assembly design process from a tolerancing viewpoint and integration of various best practices at various stages of this design process will no doubt provide a critical input to the formulation of tolerancing representation and standards.

ACKNOWLEDGMENT

The authors would like thank Dr. J. Parks, Xerox Corporation, and with Dr. R. Kacker, Statistical Engineering Division, NIST, for many valuable discussions.

REFERENCES


Yadati Narahari (M’99) received the M.E. degree in computer science and the Ph.D. degree from the Indian Institute of Science, Bangalore, in 1984 and 1988, respectively.

He is an Associate Professor with the Department of Computer Science and Automation, Indian Institute of Science. During 1997, he spent a sabbatical at the National Institute of Standards and Technology, Gaithersburg, MD, where he was instrumental in developing an integrated approach for design for quality of electromechanical products.

He co-authored the textbook Performance Modeling of Manufacturing Systems (Englewood Cliffs, NJ: Prentice-Hall, 1992). In the same year, he was awarded an Indo-US Science and Technology Fellowship, and visited the Laboratory for Information and Decision systems, Massachusetts Institute of Technology, to work on factory modeling and dynamic scheduling. During 1996, he completed “Lecture Notes on Data Structures and Algorithms,” which is currently being developed as a web-based resource. His research interests are broadly in the areas of stochastic modeling and scheduling methodologies for future factories, supply chain management, electronic commerce, and object oriented modeling.

Dr. Narahari received the Sir C.V. Raman Young Scientist Award from the Government of Karnataka in 1998. He was a Guest Editor for two special issues of the journal SADHANA, on the topic of advanced manufacturing systems in 1997 and is an Associate Editor of the IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION and the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS.

Ram D. Sriram received the B.Tech. degree from the Indian Institute of Technology, Madras, India, and the M.S. and a Ph.D. degrees from Carnegie-Mellon University, Pittsburgh, PA.

He is currently leading the Engineering Design Technologies Group, Manufacturing Systems Integration Division, National Institute of Science and Technology (NIST), Gaithersburg, MD. Prior to that he was on the engineering faculty at the Massachusetts Institute of Technology (MIT), Cambridge, and was instrumental in setting up the Intelligent Engineering Systems Laboratory. At MIT, he initiated the MIT-DICE project, which was one of the pioneering projects in collaborative engineering. He has co-authored or authored more than 120 papers, books, and reports in computer-aided engineering, including 12 books. His most recent book is Intelligent Systems for Engineering: A Knowledge-based Approach (New York: Springer-Verlag).

Dr. Sriram received the Presidential Young Investigators Award from the National Science Foundation in 1989 and was a founding Co-Editor of the International Journal for AI in Engineering.

Kevin Lyons is a Manager with the Engineering Design Technologies Group, Manufacturing Systems Integration Division, National Institute of Standards and Technology, Gaithersburg, MD. He just completed an assignment as a Program Manager with the Defense Advanced Research Projects Agency (DARPA), where he was responsible for the conceptualization, development, and execution of advanced research and development programs in design and manufacturing. He supervised programs such as Rapid Design Exploration and Optimization (RaDEO), Agile Manufacturing, and Solid Freeform Fabrication and Design Programs. Prior to his DARPA assignment, served as project manager for a research effort at NIST, directed at developing a framework to facilitate the integration, interoperability, and sharing/exchange of engineering information. This work was scoped to address assembly relevant issues for electromechanical systems and evolved into the project titled “Open Assembly Design Environment (OpenADE).” Prior to this, Lyons worked in industry for 15 years dealing in product design and analysis, factory automation, and quality engineering. Other areas of his interest are design methodology and design knowledge representations in the domain of mechanical and electromechanical components, assemblies, and systems.