

Tolerancing for an Apple Pie: A Fundamental Theory of Tolerances

Joshua Roland Campbell

Department of Mechanical Engineering,
George Mason University,
Fairfax, VA 22030
e-mail: joshuarcampbell1998@gmail.com

George A. Hazelrigg¹

Department of Mechanical Engineering,
George Mason University,
Fairfax, VA 22030
e-mail: ghazelri@gmu.edu

Tolerancing began with the notion of limits imposed on the dimensions of realized parts both to maintain functional geometric dimensionality and to enable cost-effective part fabrication and inspection. Increasingly, however, component fabrication depends on more than part geometry as many parts are fabricated as a result of a “recipe” rather than dimensional instructions for material addition or removal. Referred to as process tolerancing, this is the case, for example, with IC chips. In the case of tolerance optimization, a typical objective is cost minimization while achieving required functionality or “quality.” This article takes a different look at tolerances, suggesting that rather than ensuring merely that parts achieve a desired functionality at minimum cost, a typical underlying goal of the product design is to make money, more is better, and tolerances comprise additional design variables amenable to optimization in a decision theoretic framework. We further recognize that tolerances introduce additional product attributes that relate to product characteristics such as consistency, quality, reliability, and durability. These important attributes complicate the computation of the expected utility of candidate designs, requiring additional computational steps for their determination. The resulting theory of tolerancing illuminates the assumptions and limitations inherent to Taguchi’s loss function. We illustrate the theory using the example of tolerancing for an apple pie, which conveniently demands consideration of tolerances on both quantities and processes, and the interaction among these tolerances. [DOI: 10.1115/1.4057040]

Keywords: design methodology, design optimization, design theory, product design, tolerance analysis and design

1 Introduction

It is likely that the modern concepts of tolerancing have their origins in the notion of interchangeability of parts [1,2]. Such concepts date back over half a millennium as Gutenberg’s press (1450s) relied on interchangeable letters. Over the ensuing years, it became clear that making parts interchangeable is not as easy as one might expect. Nonetheless, the emergence of steam power in the 1780s demanded that parts be made with challenging accuracy. Benjamin Franklin reported, in 1785, of a French gunsmith making muskets with interchangeable parts. And, 100 years later, with emergence of the Industrial Age, mass manufacturing on an assembly line required part interchangeability.

Parker [3,4], working at the Royal Torpedo Factory in Scotland, is credited by Liggett [5] with being the first to formally address “position tolerance theory.” Since that time, tolerance theory has emerged as a major subdiscipline of engineering design and manufacturing. In the earlier years, tolerances were mainly associated with part geometry resulting in the discipline of geometric tolerancing. The need to properly interpret part specifications led to standards for dimensional tolerancing [6] and, with the emergence of computers, Requicha and Voelcker [7–10] developed a theory of geometric modeling that enabled computer-aided design.

A key problem in the setting of tolerances is referred to as the problem of “stack-up” [11]. This problem occurs when a series of parts must fit or work together within an overall tolerance. Problems of this sort led to the notion of optimizing the allocation of the individual part tolerances to achieve the overall desired tolerance at the minimum cost [12–14].

A major contributor to a theory of tolerancing is Taguchi [15]. His philosophy may be summarized in four statements: “It is better to be precise and inaccurate than being accurate and imprecise; Quality should be designed into the product and not inspected; Quality is achieved by minimizing the deviation from the target; [and] The cost of quality should be measured as a function of the deviation from target.” [16] This philosophy resulted in the concept of the *Taguchi loss function*.

More recently, it is noted that several products are described not so much by their dimensions as by a “recipe” according to which they are manufactured. This is the case of integrated circuit chips and food products such as an apple pie. In these cases, tolerances largely determine the quality, lifetime, or reliability of the product. These are important attributes not often captured by product descriptions as they can significantly impact the proclivity of consumers to purchase a product. Again, recognizing that demanding narrower tolerances results in higher costs, several researchers have sought to meet a set of performance requirements at minimum cost [17,18].

The problem with minimizing manufacturing cost is that this objective results in the trivial solution of manufacturing none of the product. If the manufacturer manufactures no product, the manufacturing cost is \$0.00. This, obviously, is not a helpful solution. To render the solution helpful, it is then necessary either to impose constraints on the optimization problem or to change the objective. Constraints typically take the form of a set of product requirements, whereas an alternative objective may seek minimum cost per item produced. Hazelrigg and Saari [19] note that constraints only remove alternatives from the allowable set of design choices and, if they remove the optimal point, that is, if the constraints are active, they always penalize performance. Thus, for optimal design, constraints should be avoided to the extent possible. A way to avoid constraints is to change the objective function to one that more accurately reflects the preference of the responsible decision maker. Noting that the underlying

¹Corresponding author.

Contributed by the Design Theory and Methodology Committee of ASME for publication in the JOURNAL OF MECHANICAL DESIGN. Manuscript received January 14, 2022; final manuscript received February 26, 2023; published online March 29, 2023. Assoc. Editor: Paul Witherell.

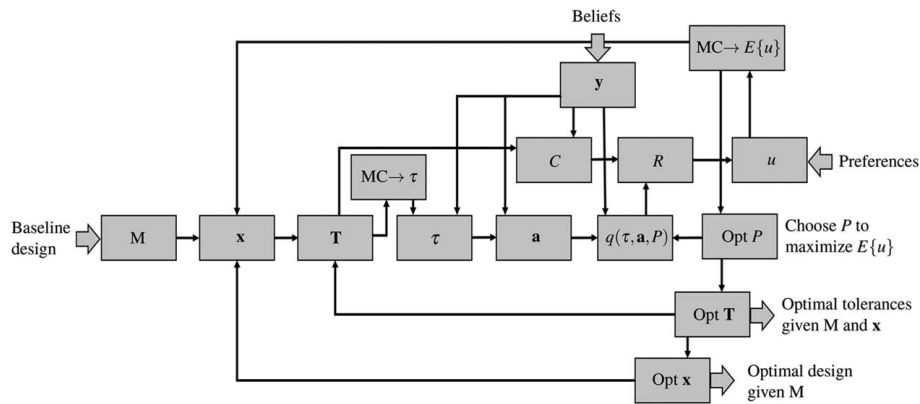


Fig. 1 A framework for optimal tolerance design

objective of a profit-making organization is to make money, Hazelrigg [20] presents a framework for product design optimization with this objective that also accounts for uncertainty.¹

Tolerances introduce the opportunity to intentionally allow variability in a product, the incentive being to reduce the cost of manufacture. As noted, this variability results in attributes of concern to customers of the product that obtain directly from the product-to-product variation. Product variability introduces risk into a purchase decision that is not present in a deterministic product. Optimization of tolerances must take this risk into account. Thus, the purpose of this article is to show that the basic logic of Hazelrigg's framework, with minor modification, can be applied to the optimization of both geometric and process tolerances separately or concurrently with the product design, with the objective being the maximization of a measure of net revenue or profit. The medium used to illustrate this application is the tolerancing of an apple pie.² Although the optimization framework is designed to an objective of profit maximization, it is conveniently adaptable to other valid preferences.

2 A General Framework for Tolerancing

Hazelrigg and Saari [19] show that optimal system design, including tolerances, demands that all design decisions be made using an overall system preference. Thus, the underlying tenet of this article is that the purpose of tolerancing is to increase the value, measured as expected utility, $E\{u\}$, of a product to the producer of the product. This is a sensible tenet for a number of reasons. First, it is the producer of the product who decides what the tolerances should be and, for rationality, this choice must be based on a preference of the decision maker. Second, for a product that has multiple consumers, it is, in general, not possible to express a joint consumer preference that would enable rational choice of tolerances [21,22]. Third, for most products, the consumers are too far removed from the technical aspects of a product to care about tolerances or even understand them. Fourth, vendors or parts suppliers have conflicting interests with the producer and, for this and other reasons, cannot be left to select the tolerances on the parts they produce.

With this in mind, the product design optimization framework that shall be used here is a modification of Hazelrigg's framework as shown in Fig. 1. The purpose of this framework is to enable

computation of an objective function, namely, expected utility, $E\{u\}$, which is based on a logical and defensible preference of the relevant decision maker and that enables product design optimization under uncertainty. There are three entry points in this framework: description of a baseline design, specification of a set of beliefs defined as "exogenous" variables that define the extant uncertainty, and the expression of a preference from which we will be able to determine a utility measure. A design is described fully by its configuration, M , its dimensions, x , and tolerances, T , on the variables x . These are deterministic design decision variables subject to optimization. Typically, M consists of a set of statements that describe the system in detail,³ and the x are continuous real numbers that may include weights, voltages, volumes, and other such variables in addition to dimensions. The values of the variables, x , are tolerated, whereas the statements that comprise M are not. While instantiations of the product or system achieve the design descriptors, M , achieved values of x , $\tilde{x}$, will vary from the design whenever $T \neq 0$. The variables T comprise the tolerances applied to x ,⁴ while $\tau(M, x, T)$ are the attributes of the system that are a function of the variability in $\tilde{x}$ as determined by T . The values of τ would typically, but not necessarily, be determined by a Monte Carlo simulation, taking into account the degree to which the tolerance is not held with precision. τ is an aggregate parameter accounting for the variability in the achieved values $\tilde{x}$. $a(M, x)$ are the as-designed system attributes taking M and x to have their nominal values, that is, with $T = 0$. We might refer to the attributes a as performance attributes, such as maximum speed, acceleration, and gas mileage for a car, and the attributes τ as quality attributes, such as reliability and lifetime. $C(M, x, T)$ is the cost of producing the system, and $q(\tau, a, P, t)$ is the demand for the system as a function of its attributes, τ and a , price, P , and possibly time, t .⁵ P is a design variable chosen to maximize $E\{u\}$. a and C are differentiable functions of x . R is the gross revenue derived from the system. y are exogenous variables that specify all uncertainties related to the system performance, cost, demand, and other variables such as the weather. u , utility, is a risk-adjusted measure of system performance in a specific simulation case obtained by optimizing price. u is determined by the overall system preference, for example, to make money, more is better. The final measure of the system performance is expected utility, $E\{u\}$, again typically obtained via a Monte Carlo simulation. Optimization loops are used to optimize the tolerances, T , and

¹The fact that this example demands the imposition of a self-imposed constraint highlights the fact that minimization of cost is not a valid preference. It is important that a theory of tolerancing enable the use of a valid preference that does not require self-imposed constraints.

²Approximately 50 million apple pies are manufactured annually for sale in grocery stores, generating an annual revenue of about a quarter of a billion dollars in 2020 (data from Information Resources, Inc.). The tolerancing of an apple pie is very much an engineering problem taken quite seriously by this industry, and it illustrates issues of tolerancing not obvious in more complex examples.

³For example, such a statement might be, "The car has four doors." These statements may include descriptions of manufacturing processes, operations, and maintenance, and even distribution, sales, and disposal/recycling.

⁴Tolerances may be expressed in any convenient form that describes the variability in the achieved variables, $\tilde{x}$, of x . For example, this may take the form of probability distributions on $\tilde{x}$.

⁵This framework enables C , R , and q to be functions of time, t .

system design variables $\mathbf{x}$. With this simple overview, we will now look at the elements of this framework in more detail.

Without denying the possibility of a producer having alternative preferences, the following is based on the notion that the underlying preference of a producer is to make money, and more is better. A full preference consists of three parts [23], the fundamental preference—taken here to be for money—a time preference, and a risk preference. A time preference is generally expressed through discounting, and the risk preference is expressed through the curve of utility versus money. The net value derived from a product, corrected for its time value (discounted), is given by

$$V = \sum_{\text{Today}}^{\infty} \frac{R - C}{(1 + r)^t} \quad (1)$$

where V is the net present value of profits, t is the time per period, and r is the discount rate per time interval ($r=0$ infers that equal sums of money have equal value independent of when they are received). Revenues are generated by selling things, and costs are generated by buying things. Normally, one sells the product produced, and the revenue generated at time t by the sale of the product is $R(t) = q_s(t)P_s(t)$, where $q_s(t)$ is the quantity sold at time t and $P_s(t)$ is the price at which it is sold. It is possible, however, that the production of the product produces other salable items (things that may appear to be “waste”), and the revenue generated by their sale should be included.

Risk preferences derive from a decision maker’s willingness to wager on an uncertain return. Generally this is calculated as a function of utility and presented as an Arrow-Pratt [24,25] measure of absolute risk aversion (ARA).

$$\rho(V) = -\frac{d^2 u(V)/dV^2}{du(V)/dV} \quad (2)$$

where $u(V)$ denotes the “utility” of V and $\rho(V)$ is a measure of ARA. Thus, if $\rho(V)$ is positive, the individual is risk averse; if it is negative, the individual is risk proverse; and if it is zero, the individual is risk neutral. Utility is a cardinal measure commonly determined via a decision maker’s response to a von Neumann–Morgenstern lottery [26]. Utility is typically a random variable, whereas expected utility is a deterministic, ordinal variable. Thus, the use of expected utility as an objective for optimization converts the nondeterministic function u into a deterministic objective function, $E\{u\}$, as is mathematically required for its existence [27].

The quantity of product sold at each point in time depends on the demand for the product, which is a function of its attributes and its price. The attributes of the product are a result of its design and its tolerances. Variability in products is the result of nonzero tolerances. The more nearly identical that each individual product is to a nominal product, the more predictable it will be, and predictability of a product may itself be an attribute of concern to customers, frequently referred to as the product’s “quality” [28–30]. For example, customers are often concerned about getting a “lemon,” particularly in the purchase of a car, and they show this preference by paying more for cars that have good reliability reports.

3 The Mathematics of Tolerances

Referring again to Fig. 1, the elements of $\mathbf{M}$ describe the configuration of a product or a system. These elements typically are not continuous variables nor are tolerances applied to them. The variables, $\mathbf{x}$, on the other hand, are continuous real numbers and typically are assigned tolerances. This differentiation between $\mathbf{M}$ and $\mathbf{x}$, together with the notion that variations in $\mathbf{x}$ are small because of small $\mathbf{T}$, enables us to write the expected utility of a design in the form of a Taylor series in a region near a reference design, $(\mathbf{x}_0, \mathbf{T}_0)$.

$$E\{u(\mathbf{x}, \mathbf{T})\} - E\{u(\mathbf{x}_0, \mathbf{T}_0)\} = \frac{\partial E\{u\}}{\partial(\mathbf{x}, \mathbf{T})} \bigg|_{(\mathbf{x}_0, \mathbf{T}_0)} \begin{pmatrix} \delta\mathbf{x} \\ \delta\mathbf{T} \end{pmatrix} + \frac{1}{2} (\delta\mathbf{x}^T, \delta\mathbf{T}^T) \frac{\partial^2 E\{u\}}{\partial(\mathbf{x}, \mathbf{T})^2} \bigg|_{(\mathbf{x}_0, \mathbf{T}_0)} \begin{pmatrix} \delta\mathbf{x} \\ \delta\mathbf{T} \end{pmatrix} + \dots \quad (3)$$

where $\delta\mathbf{x} = (\mathbf{x} - \mathbf{x}_0)$ and $\delta\mathbf{T} = (\mathbf{T} - \mathbf{T}_0)$. If we take the reference point $(\mathbf{x}_0, \mathbf{T}_0)$ to be a maximizing point of $E\{u(\mathbf{x}_0, \mathbf{T}_0)\}$, then the elements of the first-order term on the right-hand side of Eq. (3) are zero. In addition, the elements of the second-order cross derivatives, $\partial^2 E\{u\}/(\partial\mathbf{x}^T \partial\mathbf{T}^T) = \partial^2 E\{u\}/(\partial\mathbf{T}^T \partial\mathbf{x}^T)$, are zero, leaving, to second order,

$$E\{u(\mathbf{x}, \mathbf{T})\} - E\{u(\mathbf{x}_0, \mathbf{T}_0)\} = \frac{1}{2} \delta\mathbf{x}^T \frac{\partial^2 E\{u\}}{\partial\mathbf{x}^2} \bigg|_{\mathbf{x}_0} \delta\mathbf{x} + \frac{1}{2} \delta\mathbf{T}^T \frac{\partial^2 E\{u\}}{\partial\mathbf{T}^2} \bigg|_{\mathbf{T}_0} \delta\mathbf{T} + \dots \quad (4)$$

Note that we write the Taylor series in terms of the performance variable $E\{u\}$. This differs from the work of Zhang et al. [31] and Tarcolea and Paris [28], for example, who write the series in terms of the Taguchi loss function. While these authors then assume that the first-order term is zero since the loss function achieves a minimum at the target value, when viewed as written in Eq. (3), this would not appear to be the case in general. In order that the first-order terms in Eq. (3) be zero, it is necessary that the *expected utility* of the design be maximized with respect to $\mathbf{x}$ and $\mathbf{T}$, as the condition of design optimality for the variables $\mathbf{x}$ is $\partial E\{u\}/\partial\mathbf{x} = \mathbf{0}^T$, and for $\mathbf{T}$ is $\partial E\{u\}/\partial\mathbf{T} = \mathbf{0}^T$. This assures that the second-order terms in Eq. (3), which are, in fact, negative loss terms (i.e., benefit terms), dominate and that the loss related to $\mathbf{T}$ is axisymmetric and quadratic about the reference design point $(\mathbf{x}_0, \mathbf{T}_0)$ only if $(\mathbf{x}_0, \mathbf{T}_0)$ maximizes $E\{u\}$. Thus, to be precise, the

optimization of tolerances must be done concurrent with the optimization of the design variables, $\mathbf{x}$.

Unfortunately, concurrent optimization of both $\mathbf{x}$ and $\mathbf{T}$ can be a bit onerous, so we may be inclined to simplify the procedure with some approximations. To begin, it is reasonable to assume that the tolerances do not allow variations in the achieved values $\tilde{\mathbf{x}}$ that are large enough to significantly alter the attributes $\mathbf{a}$. This enables us to optimize the design with respect to $\mathbf{x}$ while holding $\mathbf{T} = \mathbf{0}$. Next, assuming that $\partial E\{u\}/\partial\mathbf{T} \approx \mathbf{0}$ in the vicinity of $\mathbf{T}_0 = \mathbf{0}$ and that typical tolerance values are such that $\delta\mathbf{T}$ is small, we can examine the rightmost term of Eq. (4) assuming that the system is optimized only with respect to $\mathbf{x}$, and we will take $\mathbf{T}_0 = \mathbf{0}$. Under these conditions, the negative of this term approximates the Taguchi loss function.

$$L \approx \frac{1}{2} \delta\mathbf{T}^T \left\{ \frac{\partial^2 [-E\{u\}]}{\partial\mathbf{T}^2} \bigg|_{\mathbf{T}_0} \right\} \delta\mathbf{T} \quad (5)$$

Note that $\partial^2 [-E\{u\}]/\partial\mathbf{T}^2 = \mathbf{H}$ is a positive definite Hessian matrix. Therefore, to second order, a surface of constant loss forms an n -dimensional hyperellipsoid, with the value of the loss dependent on the eigenvalues and eigenvectors of the Hessian matrix. Furthermore, although typical formulations of loss functions in the case of tolerances on multiple elements of $\mathbf{x}$ tend to treat the losses as independent of each other, this formulation shows that, in general, they are not independent. Indeed, the Taguchi loss functions are not linearly additive, that is, the total loss is not equal to the sum of the individual loss functions, and it would be a mistake to optimize tolerances individually.

We now see that the Taguchi loss function relies on several underlying assumptions that were not obvious in the absence of the aforementioned derivation. First, it requires that the basic system design, namely, the variables $\mathbf{x}$, be chosen to maximize $E\{u\}$, thus taking uncertainty and risk into account. Second, it requires that $\delta\mathbf{T}$ be small. But, just because $\delta\mathbf{T}$ is small does not mean that nonzero tolerances necessarily have a small impact on $E\{u\}$. Indeed, tolerance variables can have associated attributes, τ , such as reliability or safety, that have profoundly large impact on $E\{u\}$. In these cases, it is important that the decision maker's risk preference be taken into account. The Taguchi loss function does not do this.

So far, we have recognized that nonzero tolerances, $\mathbf{T}$, impart losses to the value of a system or a product. This alone would prompt a selection of $\mathbf{T}=\mathbf{0}$. However, countering this, tolerances come with an associated cost, and the smaller the tolerance, the higher the cost. Accordingly, the total loss is the sum of the loss function, Eq. (5), and the cost of the tolerance $\mathbf{C}_T(\mathbf{T})$,⁶

$$L_{tot} \approx \frac{1}{2} \delta\mathbf{T}^T \left\{ \frac{\partial^2 [-E\{u\}]}{\partial \mathbf{T}^2} \right\}_{\mathbf{T}_0} \delta\mathbf{T} + \mathbf{i} \mathbf{C}_T(\mathbf{T}) \quad (6)$$

where the row vector $\mathbf{i} = [1, 1, \dots, 1]$. As noted in Sec. 5, it should be expected that the distribution of $\tilde{\mathbf{x}}$ would be correlated with $\mathbf{C}_T(\mathbf{T})$. At this point, one might be inclined to minimize L_{tot} . But this would fail to account for the risk preference of the decision maker. Instead, one should maximize $E\{u\}$. This is a simple task only if the decision maker is risk neutral, that is, if utility equals profit, or if the variation in u because of the nonzero tolerances is sufficiently small that this is a reasonable approximation. Unfortunately, as noted earlier, the latter is not always the case as the quality attributes of a product have the potential to significantly affect demand. Hence, the selection of optimal tolerances is not as easy as the Taguchi method would have us believe.

Before proceeding further, it is appropriate to consider the terms that comprise the Hessian matrix in Eq. (6). This matrix provides an estimate of the expected utility loss because of a degradation (real or perceived) in product quality, reflected as a shift in the demand curve. Taking the variation in demand to be a continuous, differentiable function of the quality attributes, we can write the demand function in the form of a Taylor series,

$$q(\tau) = q_0 + \frac{\partial q}{\partial \tau} \delta\tau - e \frac{q_0}{P_0} \delta P + \dots \quad (7)$$

where $e = -(P/q)(\partial q/\partial P)$ is the price elasticity. To minimize the loss from reduced demand, depicted as the shaded area in Fig. 2, we must re-optimize the price.

Let $\Delta q = (\partial q/\partial \tau) \delta\tau$ denote the shift in the demand curve resulting from a nonzero tolerance specification, α_s be the increase in marginal production cost per unit, that is, the slope of the marginal supply cost curve, and S_0 be the marginal supply cost at a production rate of q_0 . Then, the loss is given by

$$L = -q_0 \delta P - (P_0 - S_0 + \delta P) \delta q + \frac{\alpha_s}{2} \delta q^2 \quad (8)$$

where

$$\delta q = -e \frac{q_0}{P_0} \delta P - \Delta q \quad (9)$$

Thus, L is a quadratic function of δP . Solving for the minimum loss, the optimum value of δP is given by

$$\delta P = -P_0 \frac{\Delta q(P_0 + \alpha_s e q_0) - P_0 q_0(1 - e) - e q_0 S_0}{e q_0(2P_0 + \alpha_s e q_0)} \quad (10)$$

⁶ L_{tot} may contain additional terms relating to expenses such as warranty costs and liability costs. We include these in the loss function to emphasize that they are associated with tolerances.

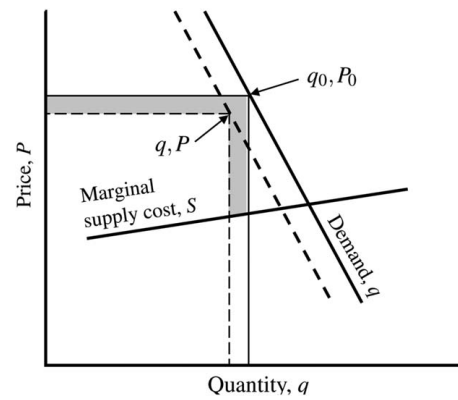


Fig. 2 Profit loss resulting from a quality loss

This yields the interesting, albeit intuitive, result that, for high demand elasticity (somewhat greater than unity), increasing tolerances results in lower optimal product prices, whereas for inelastic demand, increasing tolerances results in higher optimal prices. Examples show that optimal adjustment of product price to match the selected tolerances can significantly improve profitability.

4 Computational Procedure

The computational framework follows the logic flow shown in Fig. 1, which outlines a procedure for the optimization of the product design, including tolerances, as a unified process. Unfortunately, for most products, this can lead to a highly complex and time-consuming set of computations. The complexity of the problem makes it desirable to resort to Monte Carlo methods, which sacrifice computational efficiency to achieve a more simplified and less-prone-to-error mathematical formulation. However, even this may leave the problem intractable. As a result, it is desirable to separate the dimensional optimization of the product from the tolerance optimization. The assumptions leading to Eq. (5) enable this separation. Thus, in practice, it is convenient to apply the framework in two steps, first the optimization of the “dimensions” (target values of $\mathbf{x}$) of the product and then, based on these optimized values, the optimization of the tolerances placed on the target values.

We shall begin our outline of the computational procedure under the assumption that the basic product design has already been optimized. Keep in mind that the validity of Eq. (5) depends on this being the case. Tolerances place “constraints” on the variability of the outcomes, $\tilde{\mathbf{x}}$, of the decisions, $\mathbf{x}$, with a concomitant cost. The goal of the computational procedure is to enable a selection of these constraints such that they maximize the value, measured as the expected utility, of the product to the producer. Under the condition that the basic product design, assuming all $\mathbf{x}$ values achieve their nominal value, is optimized to achieve maximum expected utility, Eqs. (8)–(10) afford some degree of independence from the basic design in the consideration of tolerances. Indeed, in the case that the decision maker is risk neutral, that is, for whom utility equals profit, minimizing the expected loss is a solution. However, minimization of the loss does not assure maximization of expected utility for decision makers who are not risk neutral. Because of this, we are forced to compute a utility difference in the context of the expected value of the basic design. This requires evaluation of the expected utility of the basic design and evaluation of the total loss function as a deviation from the expected utility of the basic design.

The first issue, which would appear to be overlooked in many applications of the Taguchi loss function, is the need to take product price into account as a variable of choice to the manufacturer that also must be optimized. The thing that makes the determination of the optimal price shift tricky is that the quality loss is not realized on a product-by-product basis one by one as products come off the production line, but rather on consumer perceptions based on

a history of many products produced under the design variables and tolerances of the product. Thus, in order to simulate the demand shift, for each product outcome (achieved values of $\mathbf{x}$ on a product-by-product basis), we must compute the product loss function. This requires the inner Monte Carlo simulation shown in Fig. 1 between the selection of tolerances, $\mathbf{T}$, and their resulting quality measure, τ . This nesting of Monte Carlo loops can result in substantial increases in computational time. One approach to this problem is to assume that there is no significant variability in the outcome of $\mathbf{x}$ on a product-by-product basis, that is, a change large enough to alter the attributes, $\mathbf{a}$, and analyze only the impact of tolerance on one particular product instantiation. This provides an approximate result that can be later checked against a limited number of full simulations around the optimal tolerance design point.

The next problem we encounter is the appropriate expression for the cost of achieving a specific tolerance level. One way to achieve a given tolerance is to test to assure that all tolerances are met and to discard any parts of or products that fail to meet the tolerance. This results in a cost of wastage. The wastage costs result from the costs of manufacturing unsalable product. It is obviously desirable to keep wastage small. But this means maintaining tolerances with a high per-product probability, and this often demands more sophisticated and concurrently more expensive manufacturing equipment. Ergo, as a tolerance is reduced, the manufacturer must consider the purchase of more expensive manufacturing equipment. The tolerance cost model must reflect these costs.

While tolerances denote the limits of acceptable outcomes of $\mathbf{x}$, they do not describe distribution of these outcomes. Yet, this distribution is needed in order to compute the loss function. While it might seem natural to describe the distributions of outcomes as Gaussian with a mean and standard deviation, the Gaussian distribution has the property that it extends infinitely in both positive and negative directions. This causes problems for two key reasons. First, actual parts do not have negative dimensions, and second, actual parts do not get infinitely large. One might think that, for a Gaussian distribution, these are extremely rare occurrences that can be neglected. The problem is, however, that a Monte Carlo analysis will interrogate the distribution hundreds of millions, perhaps even billions, of times, and rare events that will cause errors are bound to occur. Thus, we have chosen to represent tolerances for the analyses presented here as beta distributions, although other distributions can be used within the context of the theory presented here. Beta distributions have finite limits, can be skewed, and can be shaped based on the distribution parameters α and β .

Lastly, we need to discuss the formulation and determination of the loss function. As noted earlier, the mathematics of the loss function confirm that it should take the form of an n -dimensional hyperellipsoid. Let the orthogonal unit vectors $\hat{\mathbf{x}}$ correspond to the elements of $\mathbf{x}$ to form a Cartesian coordinate system. Note that the critical point or center of a hyperellipsoid denoting a surface of constant loss is located at the design point $\mathbf{x}_0$, with its axes aligned with the orthogonal eigenvectors, $\mathbf{v}$, of the Hessian matrix $\mathbf{H}$. Denote by $\tilde{\mathbf{x}}$ the coordinates of a point $\mathbf{x}$ in the rotated and translated coordinate system defined by $\mathbf{v}$. The critical point of the hyperellipsoid is at the center of this coordinate system. Then, the loss function corresponding to Eq. (5) is given by the equation of a hyperellipsoid in the rotated and translated coordinate system,

$$L = \gamma \sum_i \nu_i \tilde{x}_i^2 \quad (11)$$

where ν_i are the eigenvalues of $\mathbf{H}$ and γ is a proportionality constant. The coordinates of $\mathbf{x}$ in the translated system are given by $\delta\mathbf{x} = (\mathbf{x} - \mathbf{x}_0)$. Rotating the axes to the eigenvector system gives the coordinates, $\tilde{\mathbf{x}}$, in the eigenvector system.

$$\tilde{\mathbf{x}} = \delta\mathbf{x}^T \hat{\mathbf{v}} \quad (12)$$

where

$$\hat{\mathbf{v}}_i = \frac{\mathbf{v}_i}{\sqrt{\mathbf{v}_i^T \mathbf{v}_i}} \quad (13)$$

are unit vectors comprising a coordinate system where the loss function hyperellipsoids are centered on the coordinate system with their axes aligned with the eigenvectors of the Hessian matrix.

5 Apple Pie

As an illustration of the decision theoretic formulation of the tolerancing problem, we have chosen the tolerancing of an apple pie. The detailed geometric tolerancing of an apple pie would be a formidable task and, in the end, rather futile. No two pies are exactly alike nor would anyone want that they would be. So, geometric tolerancing is not appropriate for a pie, save for, perhaps, the diameter of the pie as it has to fit in a box for marketing purposes. Instead of describing an apple pie by its detailed dimensions, which would involve volumes of numbers, we describe and tolerance an apple pie by its *recipe*. Accordingly, tolerances are placed on the measurable parameters of the recipe, including amounts of ingredients and processing parameters such as baking time and temperature. The apple pie recipe used in this example is given below.

Apple Pie

Ingredients

$\frac{1}{2}$ cup sugar, more to taste
 $\frac{1}{2}$ cup packed brown sugar
 3 tablespoons all-purpose flour
 1 teaspoon ground cinnamon
 $\frac{1}{4}$ teaspoon ground ginger
 $\frac{1}{4}$ teaspoon ground nutmeg
 6–7 cups peeled and sliced tart apples
 1 tablespoon lemon juice
 dough for double-crust pie
 1 tablespoon butter
 1 large egg white

Process

Preheat oven, 375 deg.
 Toss apples with lemon juice, add sugar, toss to coat
 Combine sugars, flour and spices
 Roll half of dough to $\frac{1}{8}$ -in.-thick circle,
 transfer to 9-in. pie plate,
 trim even with rim
 Add filling, dot with butter
 Roll remaining dough to $\frac{1}{8}$ -in.-thick circle
 Place over filling, trim even with rim, seal, and flute edge
 Cut slits in top
 Beat egg white until foamy, brush over crust
 Sprinkle with sugar
 Cover with foil, bake 25 min
 Remove foil and bake another 25 min
 Cool on wire rack

This recipe is divided into two parts, the first part specifying amounts of each ingredient, the variables of which are measurable, continuous real numbers as required by Eq. (6). Tolerances would typically be placed on these variables. The second part specifies the processing steps. Some of these steps are amenable to tolerancing, such as the baking temperature and time. But others defy tolerancing or even a clear definition. Indeed, we become rather

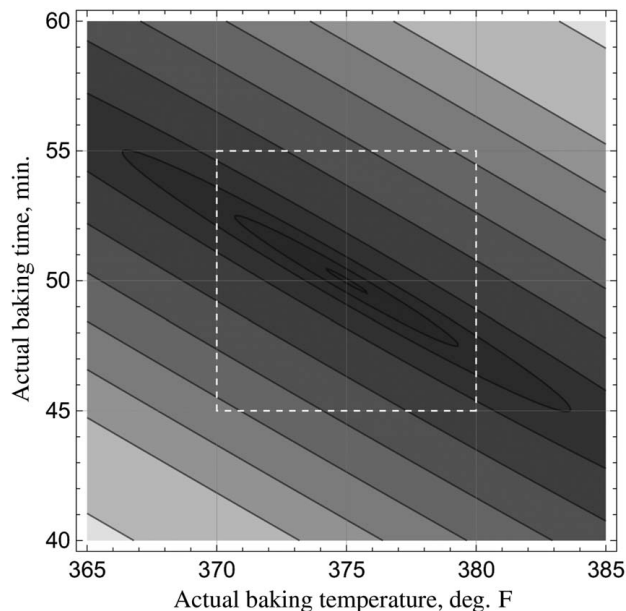


Fig. 3 Loss as a function of actual baking temperature and time, with target values of 375 deg and 50 min

philosophical at this point, invoking Gödel's theorem. Gödel's theorem deals with the limits of rationality in reflexive systems.⁷ Language is a reflexive system, that is, we define words with words. Hence, all definitions rely on knowing the definitions of other words, which are only known by knowing the definitions of other words, and so on. Thus, words describing the aforementioned process steps such as "toss," "combine," "beat," "roll," and "sprinkle" can never be defined with clear and precise precision. What this means is that the process steps of a typical recipe can never be transferred assuring no loss of clarity and, as a result, there will always be some element of art in the manufacture of any product that involves process steps.

Nonetheless, we assume that these ingredient amounts and process variables have been duly optimized ($\partial E\{u\}/\partial \mathbf{x}_0 = \mathbf{0}^T$, Eq. (3)) and will now examine tolerances on the baking time and temperature. Note that these variables are measurable, continuous real numbers. To begin, we construct an elliptical penalty function, taking into account the correlation between these variables. Obviously, if the oven temperature is a bit low, a longer cooking time will compensate at least partly for this deviation. Figure 3 shows an elliptical loss function with zero loss occurring at a baking temperature of 375 °F and a baking time of 50 min.⁸ The figure indicates that an increase in baking time of approximately 6 min will compensate optimally for a decrease in baking temperature of 10 deg. The dashed-line box in Fig. 3 denotes example tolerance limits of ± 5 deg on temperature and ± 5 min on baking time. If these tolerances were held, pies baked in conditions that exceed these tolerances would be discarded as a loss. However, we see that pies baked in conditions just outside the upper left and lower right corners of this box are classified as waste, while they are considerably more acceptable than those being sold that are baked in conditions corresponding to the lower left and upper right corners. Intuitively, a multiparameter tolerance criteria could enable the tolerances to be relaxed while reducing waste and maintaining or even improving quality. Multitolerance criteria can be easily implemented in the context of this

⁷A reflexive statement that illustrates the limit of rationality in language is, *This statement is a lie*. If the statement is a true, it must be a lie. And, if it is a lie, it must be true.

⁸The derivatives that determine the loss function (Eq. (6)) may be obtained by measuring demand as a function of the allowed variability of the product. For the example presented, we chose an illustrative loss function.

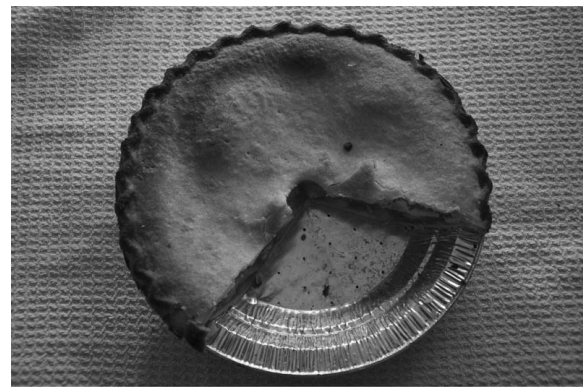


Fig. 4 A burnt apple pie. It did not taste very good.

tolerance-evaluation framework; however, our example problem sets tolerances on the variables independently.

Figure 4 shows a commercially manufactured apple pie that was baked with time and temperature parameters that exceeded appropriate tolerance limits. The crust is rather burnt and bears the taste of burnt pastry. Clearly, were this the norm, demand for this producer's pie would be significantly reduced. One might be inclined to think that we have been a bit facetious in choosing to go to so much detail to analyze production tolerances on an apple pie. Be assured, however, that this is taken quite seriously in the apple pie baking industry [32–35]. Indeed, detailed studies have been conducted to identify the attributes of importance to apple pie customers and to estimate how variations in these attributes might affect demand for the product. However, we did not choose variables for our example from the literature. Rather we chose them, while not entirely unreasonable, to emphasize aspects of tolerance optimization that one might encounter in typical manufacturing situations.

For our example, we assumed that the producer expected to be able to sell 10,000 pies per production period, with a demand elasticity of $e = 2.0$, a fixed investment cost (amortization of manufacturing equipment, rent, insurance, etc.) of \$5,000 per production period, and a marginal cost of production per pie of \$2.50 at a production rate of 10,000 pies per production period and increasing at a rate of \$0.001 per 1,000 pies. Tolerances around the target values of baking time and temperature were considered from near zero to ± 7.8 min and ± 29 deg, respectively. The distributions of times and temperatures were modeled as symmetrical beta distributions with parameters $\alpha = 4$ and $\beta = 4$, and with minimum and maximum limits of 25% below and above the tolerance limits. This leads to a rejection rate or wastage of about 12 pies per 1000. These pies are discarded, incurring production costs but producing no revenue. Given this description of the tolerances and with time and temperature distributions dependent on the tolerance limits, it seemed reasonable to model a fixed, per-period cost of maintaining the specified tolerance as a hyperbolic function of the tolerance.

$$C_T = \zeta T^\xi \quad (14)$$

where T is the tolerance ($\pm$ minutes or $\pm$ degrees), and ζ and ξ are constants. The functions used are shown in Figs. 5 and 6. As time is an easier tolerance to maintain than temperature, we modeled its cost as significantly lower than that for temperature, as the latter would likely demand more expensive ovens.

Finally, we took producer utility to be the log of the net revenue per production period. A simulation was coded that has the ability to take into account uncertainties in all major variables associated with the determination of performance (profit) as a function of tolerances. However, simulations of enough cases to map out expected utilities for even two tolerances, including all uncertainties, can be quite time consuming, as much as days of run time or longer. Thus, for the example provided here, we chose to assume that the demand, demand elasticity, and production costs are known

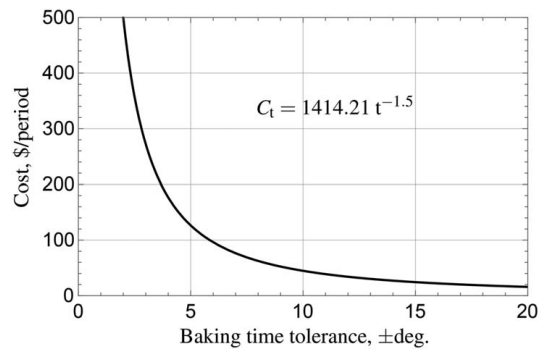


Fig. 5 Per production period cost, C_t , of maintaining a time tolerance, t

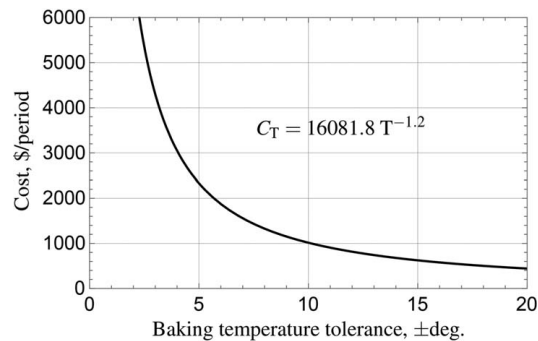


Fig. 6 Per production period cost, C_T , of maintaining a temperature tolerance, T

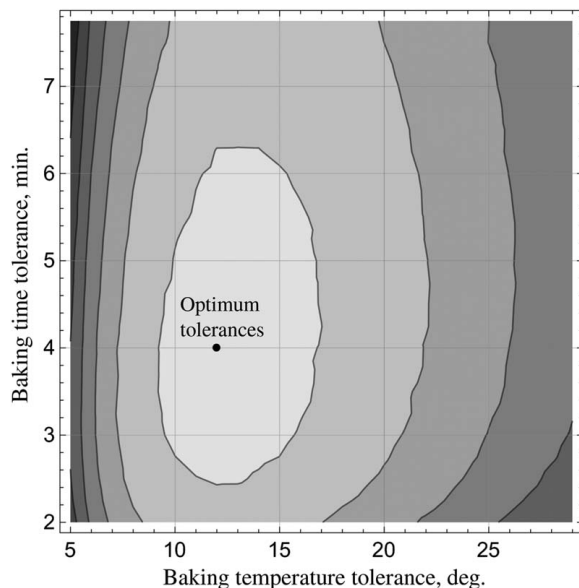


Fig. 7 Expected utility as a function of baking temperature and time tolerances showing the optimum tolerances

deterministically. With these assumptions, Fig. 7 is produced by computing results for combinations of every combination of baking time and temperature corresponding to the tick marks of this plot. This comprised a total of 600 time–temperature tolerance cases, with 1 million simulations per case. The run time for this was about 4 h.

Clearly, computer run times for cases that seek to optimize multiple tolerances with full consideration of uncertainties can be an impediment to application of this approach. Nonetheless, the approach can be used in a “deterministic” mode to locate the

regions of optimal solutions, and these can be verified with limited computations in the vicinity of the optimal solutions taking uncertainties into account. The key factor driving high computing time is the need for the solution of a nested Monte Carlo simulation, which could require a total of a billion or more simulations to achieve adequate accuracy.

6 Conclusions

The objective of this research is to cast the problem of tolerancing in the framework of decision theory. It was found that Hazelrigg’s design framework [20] could provide a mathematically rigorous basis for a theory of tolerancing with modification to enable the analysis of the so-called quality attributes emerging from product-to-product variability. The resulting analysis provides insights into the validity of the Taguchi loss function.

Taguchi defines a loss function that can be derived from a Taylor series around “target values” of design variables, with arguments that the first-order term of the series is zero because the loss is minimized at the target value. But this argument holds only if the design target values themselves are optimized with respect to an overall system or product value function, and only in the case where all derivatives of this value function with respect to the design variables x and T exist and are finite in the vicinity of these optima, thus validating the Taylor series. Otherwise, the first-order terms do not vanish and, in fact, diminish the concept of a tolerance by allowing larger tolerances to have the potential to improve certain samples of the product. Although Taguchi recognizes the need for an optimization criterion, it does not appear that this requirement is clearly recognized in applications of the loss function for tolerance optimization.

Second, the Taguchi method is most commonly applied assuming that the tolerances themselves are independent of each other. The decision theoretic formulation makes clear that this is not the case. Namely, the total value loss resulting from nonzero tolerances is not the sum of the Taguchi losses for each tolerance as determined independently.

Third, while the Taguchi loss function treats the cost of tolerancing to the manufacturer and the loss of value to the customers, the decision theoretic formulation makes clear that the important factor is profit or net benefit to the designer/manufacturer. It is this entity that decides what the tolerances should be, reaps the benefits of production, and owns the loss. This entity would likely prefer to maintain a profitable level of demand for the product, whereas nonzero tolerances reduce demand. Through consideration of demand, the decision theoretic approach takes consumer preferences into account, without the need to assess a group preference [21].

Fourth, the loss attributed to diminished demand resulting from nonzero tolerances can be mitigated by re-optimization of the price at which the product is sold. Although this is required for the optimization of tolerances, we see no evidence that it has been considered in applications of the Taguchi loss function.

Fifth, the determination of tolerances in a decision theoretic framework enables consideration of uncertainties affecting the optimal design of the entire product or system, and it accounts for the risk preference of the design decision maker. In this regard, it should be noted that, although the variation in the product resulting from nonzero tolerances may be small, it still has the potential to result in large losses in product value, thus invalidating the approximation of risk neutrality.

Lastly, we believe that the decision theoretic formulation of the tolerancing problem provides significant new insight into the mathematics of tolerancing and appears to encompass a range of tolerancing problems that span geometric dimensional tolerancing through process tolerancing.

Funding Data

- The National Science Foundation (Award No. CMMI-1923164).

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The authors attest that all data for this study are included in the paper.

Nomenclature

- q = demand for a product
 r = discount rate
 t = time
 u = utility
 $\mathbf{a}$ = a set of attributes that determine the demand for a product
 $\mathbf{v}$ = eigenvectors of the Hessian matrix
 $\mathbf{x}$ = a set of statements such as dimensions describing the measurable and continuous real variables that determine a basic design
 $\mathbf{y}$ = a set of statements that describe uncertainties on other variables
 H = Hessian matrix
 M = a set of statements describing a particular design configuration
 C = costs associated with the production of a product
 L = loss incurred because variables $\mathbf{x}$ do not achieve their target values
 P = price at which a product is sold
 R = revenue generated by the sale of a product
 V = net present value of profit
 $\mathbf{T}$ = a set of real numbers describing tolerances on the variables $\mathbf{x}$
 $\mathbf{T}^T$ = superscript denotes transpose
 $E\{u\}$ = expected utility
 τ = a set of attributes related to tolerances that affect demand for a product
 ν = eigenvalues of the Hessian matrix

References

- [1] Lienhard, J. H., "Engines of Our Ingenuity, No. 101: Interchangeability," Technical Report, <https://uh.edu/engines/epi101.htm>
- [2] Hounshell, D. A., 1984, *From the American System to Mass Production, 1800–1822: The Development of Manufacturing Technology in the United States*, Johns Hopkins University Press, Baltimore, MD.
- [3] Parker, S., 1940, "Notes on Design and Inspection of Mass Production Engineering Work," Gauge Design Drawing Office, Naval Ordnance Gauge Factory, Sheffield, UK.
- [4] Parker, S., 1956, *Drawings and Dimensions*, Sir Isaac Pitmen & Sons, Ltd., London.
- [5] Liggett, J. V., 1970, "Fundamentals of Position Tolerance," Society of Manufacturing Engineers, New York, Technical Report.
- [6] American Society of Mechanical Engineers, 1973, "Dimensioning and Tolerancing, ANSI Standard Y14.5-1973," American National Standards Institute, New York.
- [7] Requicha, A. A. G., and Voelcker, H. B., 1982, "Solid Modeling: A Historical Summary and Contemporary Assessment," *IEEE Computer Graphics Appl.*, **2**(2), pp. 9–24.
- [8] Requicha, A. A. G., 1977, "Mathematical Models of Rigid Solid Objects, Tech. Memo. No. 28, Production Automation Project," University of Rochester, Rochester, NY.
- [9] Requicha, A. A. G., 1980, "Representations for Rigid Solids: Theory, Methods, and Systems," *ACM Comput. Surveys*, **12**(4), pp. 437–464.
- [10] Requicha, A. A. G., 1983, "Toward a Theory of Geometric Tolerancing," *Int. J. Robot. Res.*, **2**(4), pp. 45–60.
- [11] Otsuka, A., and Nagata, F., 2015, "Stack-Up Analysis of Statistical Tolerance Indices for Linear Function Model Using Monte Carlo Simulation," *Proceedings of the 20th International Conference on Engineering Design (ICED15)*, Vol. 6: Design Methods and Tools – Part 2, Milan, Italy, pp. 143–152.
- [12] Song, Y., and Mao, J., 2015, "Tolerance Optimization Design Based on the Manufacturing-Costs of Assembly Quality," *Proceedings of the 13th CIRP Conference on Computer-Aided Tolerancing*, Hangzhou, China, May 11–14, Elsevier B.V.
- [13] Iannuzzi, M. P., and Sandgren, E., 1996, "Tolerance Optimization Using Genetic Algorithms: Benchmarking With Manual Analysis," *Computer-Aided Tolerancing*, Kimura, F., ed., Springer, Dordrecht, pp. 219–234.
- [14] Bryan, K., Ngoi, A., and Teck, O. C., 1997, "A Tolerancing Optimisation Method for Product Design," *Int. J. Adv. Manuf. Technol.*, **13**, pp. 290–299.
- [15] Taguchi, G., 1978, "Quality Engineering Through Design Optimization," *Proceedings of International Conference on Quality Control*, Tokyo, Japan, Oct. 17–20.
- [16] Kiran, D. R., 2017, *Total Quality Management*, Elsevier Inc., Cambridge, MA.
- [17] Kumar, R. S., Alagumurthi, N., and Ramesh, R., 2009, "Optimization of Design Tolerance and Asymmetric Quality Loss Cost Using Pattern Search Algorithm," *Int. J. Phys. Sci.*, **4**(11), pp. 629–637.
- [18] Jeong, S. H., Kongsuwan, P., Truong, N. K. V., and Shin, S., 2013, "Optimal Tolerance Design and Optimization for a Pharmaceutical Quality Characteristic," *Math. Probl. Eng.*, **2013**.
- [19] Hazelrigg, G. A., and Saari, D. G., 2022, "Toward a Theory of Systems Engineering," *ASME J. Mech. Des.*, **144**(1), p. 011402.
- [20] Hazelrigg, G. A., 1998, "A Framework for Decision-Based Engineering Design," *ASME J. Mech. Des.*, **120**(4), pp. 653–658.
- [21] Arrow, K. J., 1963, *Social Choice and Individual Values*, 2nd ed., John Wiley & Sons, New York.
- [22] Hazelrigg, G. A., 1996, "The Implications of Arrow's Impossibility Theorem on Approaches to Optimal Engineering Design," *ASME J. Mech. Des.*, **118**(2), pp. 161–164.
- [23] Howard, R. A., and Abbas, A. E., 2016, *Foundations of Decision Analysis*, Pearson, Essex, England.
- [24] Arrow, K., 1965, *Aspects of the Theory of Risk Bearing*, Yrjö Jahnssonin Saatio, Helsinki. Reprinted in: "Essays in the Theory of Risk Bearing," Markham Publ. Co., Chicago, 1971, pp. 90–109.
- [25] Pratt, J. W., 1964, "Risk Aversion in the Small and in the Large," *Econometrica*, **32**(1–2), pp. 122–136.
- [26] Hazelrigg, G. A., 2012, *Fundamentals of Decision Making for Engineering Design and Systems Engineering*, Vienna, VA.
- [27] Hazelrigg, G. A., 2014, "Guest Editorial: Not so Subtle Subtleties Regarding Preferences," *ASME J. Mech. Des.*, **136**(12), p. 120301.
- [28] Tarcolea, C., and Paris, A. D., 2011, "Loss Functions Used in the Quality Theory," *U. P. B. Sci. Bull. Ser. A*, **73**(1), pp. 45–54.
- [29] Vacarescu, C. F., and Vacarescu, V., 2010, "The Taguchi's Quality Loss Function in Development and Design for Bulk Goods in the Automotive Field," *Proceedings of the 21st International DAAAM Symposium*, Vol. 21.
- [30] Vasseur, H., Kurfess, T. R., and Cagan, J., 1997, "Use of a Quality Loss Function to Select Statistical Tolerances," *Trans. ASME*, **119**(3), pp. 410–416.
- [31] Zhang, Y., Lixiang, L., Song, M., and Yi, R., 2019, "Optimal Tolerance Design of Hierarchical Products Based on Quality Loss Function," *J. Intell. Manuf.*, **30**(1), pp. 185–192.
- [32] Moskowitz, H., 2007, "Interrelations Among Liking Attributes for Apple Pie: Research Approaches and Pragmatic Viewpoints," *J. Sensory Stud.*, **16**(4), pp. 373–391.
- [33] Sanford, K. A., McRae, K. B., and Deslauriers, M. L. C., 2001, "Sensory Descriptive Analysis and Correspondence Analysis Aids in the Choosing of Apple Genotypes for Processed Products," *J. Food Qual.*, **24**(4), pp. 301–325.
- [34] Stable Micro Systems Ltd, 2019, "How to Measure Food Texture and Why?" <https://www.azom.com/article.aspx?ArticleID=18254>, ArticleID=18254
- [35] P. Singham, P. B., and Yadav, B. K., 2015, "Importance of Objective and Subjective Measurement of Food Quality and Their Inter-Relationship," *J. Food Processing Technol.*, **6**(9), pp. 1–7.