Metrics for Evaluating the Barrier and Time to Reverse Engineer a Product

Stephen P. Harston
e-mail: sharston@gmail.com

Christopher A. Mattson
Assistant Professor
e-mail: mattson@byu.edu

Department of Mechanical Engineering,
Brigham Young University,
Provo, UT 84602

Reverse engineering, defined as extracting information about a product from the product itself, is a common industry practice for gaining insight into innovative products. Both the original designer and those reverse engineering the original design can benefit from estimating the time and barrier to reverse engineer a product. This paper presents a set of metrics and parameters that can be used to calculate the barrier to reverse engineer any product, as well as the time required to do so. To the original designer, these numerical representations of the barrier and time can be used to strategically identify and improve product characteristics so as to increase the difficulty and time to reverse engineer them. As the metrics and parameters developed in this paper are quantitative in nature, they can also be used in conjunction with numerical optimization techniques, thereby enabling products to be developed with a maximum reverse engineering barrier and time—at a minimum development cost. On the other hand, these quantitative measures enable competitors who reverse engineer original designs to focus their efforts on products that will result in the greatest return on investment. [DOI: 10.1115/1.4001347]

Keywords: reverse engineering, barrier to reverse engineering, product imitation, hardware imitation, time estimation, Ohm's law

1 Introduction

The introduction of innovative products into the marketplace is often accompanied by an interesting engineering and design dichotomy; on one hand, the original designer intends on maintaining his/her competitive advantage, gained through innovation, by offering the product to the masses without easily disclosing its enabling technology [1]. On the other hand, however, the competitor is determined to reverse engineer the innovative product so as to uncover the enabling technology and potentially earn a portion of the market by capitalizing on it [2]. Although seen from different perspectives, the notion of *barriers to reverse engineering* is critical in both cases. Ideally, to the original designer, all efforts are made to increase the barrier and time required to reverse engineer his/her design. To those reverse engineering the original designs, minimal time and barrier is desired so as to enter the market before it is saturated. In either case, these designers could benefit from general metrics and parameters for quantifying the time and barrier to reverse engineer a product [3,4]. This paper develops these metrics and parameters, provides practical insight into their use, and demonstrates their effectiveness with three examples.

Reverse engineering was defined in a variety of similar, yet unique, ways by the disciplines that have approached the topic in the literature [2,5,6]. Among the various disciplines that have addressed the topic of reverse engineering, the following three areas are predominant: (i) reverse engineering of software [5,7–10]; (ii) reverse engineering of hardware [2,3,11–13]; and (iii) reverse engineering of biological systems [6,14–16]. The reverse engineering of software is pervasive in the literature and is of particular interest as it relates to reverse engineering because software is being delivered to end-users with more mobile code in architecture-independent formats—thereby facilitating the reproduction of the original code with less effort. Strategies to prevent reverse engineering of software include tamper proofing, obfuscation, and watermarking [17].

The reverse engineering of hardware is generally addressed in the literature from within three areas of research: (i) performance benchmarking [2,11,13], which is the evaluation of competitive products in order to specify performance criteria and generate concepts for new products; (ii) geometric surface and shape recovery [18,19], which is the automated extraction of geometry from an existing product and the construction of 3D CAD models from the data; and (iii) empirical parameter estimation and surrogate model building by statistical sampling of hardware [20,21], which is simply the estimation of performance measures through testing an existing product and fitting a mathematical model to the test data, thereby developing an approximate parametric model of the product's performance.

Ingle [2] provides a basic four-stage methodology for the reverse engineering of hardware. Of the four stages presented by Ingle, the first two stages are of particular interest in the context of the present paper: Stage 1 is the evaluation and verification of a product or system, and Stage 2 is the documentation of the findings, usually in the form of technical data. As a note, Stage 3 is prototype verification, and Stage 4 is project implementation.

Finally, research in the reverse engineering of biological systems has gained more and more momentum as scientists and engineers seek to discover the building blocks of nature [14,15] and successful ways in which natural systems accomplish complex tasks [16].

Although related and useful to the design and reverse engineering of software, hardware, and biological systems, the developments presented in this paper focus on an articulated, yet unmet need in the literature—comparative metrics for barrier and time to reverse engineer a product or system. For clarity of scope, we provide the following three important definitions in the context of the present paper.

- Reverse engineering is the process of extracting information about a product from the product itself.
- Time to reverse engineer is the total required man-time to

Contributed by the Mechanisms and Robotics Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received June 1, 2009; final manuscript received February 4, 2010; published online April 22, 2010. Assoc. Editor: Mary Frecker.

reverse engineer a product without consideration to parallel activities.
- Barrier to reverse engineering is anything that impedes reverse engineering.

Barriers to reverse engineering include, for example, critical complex surfaces that are difficult to recreate, localized heat treating that creates difficult-to-discover heterogeneous material characteristics, and hidden in situ sensors that monitor performance. Importantly, we note that there are distinct differences between time and barrier, and that a large time to reverse engineer a product does not necessarily imply that there is also a large barrier to reverse engineering. For example, the barrier to extract geometric information from keys on a keyboard is relatively small as it only requires simple measurements that are easy to obtain. However, the time to reverse engineer the keys on a keyboard may not be small, due to the quantity of keys requiring analysis.

Various researchers have expressed the need to estimate the time and barrier to reverse engineer a product. The various perspectives in the literature range from those of the original designer [3,12,13], to those who reverse engineer [2,4,11], and to market analysts [3,22,23]. While these perspectives are insightful and suggest the need for quantitative measures, unfortunately none of them provide it.

Macmillan et al. [3] stated that it is critical to estimate the competitor's response lag (or time to reverse engineer and imitate a product) in order to understand the potential financial risks and profits. Pahl et al. [13] stated that effective product planning includes understanding the life cycle of the proposed product, as well as understanding the competitor's products. Therefore, effective product planning and definition of product life cycle is likely to (i) consider the time required for competitors to conduct reverse engineering activities and (ii) require a full understanding of competitive products through reverse engineering activities.

Shapiro [24] and Nelson and Winter [25] emphasized that the harder a product is to reverse engineer—dependent upon the competitor and their resources and skills available—the less incentive a competitor has to imitate the technology. On the other hand, there is little incentive for original designers to develop innovative products if competitors can imitate the products at a significantly reduced development cost with a larger return on investment [4].

While others previously presented the idea of barriers to reverse engineering [12,26,27], we pursue the concept of barriers to reverse engineering further by developing relationships that define quantitative representations of the barrier and time to reverse engineer any product.

In the present paper, we develop a set of metrics and parameters to quantify two things; (i) a measure of how difficult a product is to reverse engineer, and (ii) how much time is required to reverse engineer the product. While there are many applications of the metrics and relationships presented in this paper, they have been developed with the intention of using them in conjunction with numerical optimization approaches to maximize the barrier and time to reverse engineer a product.

The relationships developed in the present paper originate from Ohm's law and enable us to estimate the time to reverse engineer a product with an average error of 12.2%. We start by presenting a brief overview of the pertinent relationships used from Ohm's law in Sec. 2. The adaption of Ohm's law to product development is then presented in Sec. 3, followed by a discussion of model limitations and sensitivity analysis of the presented metrics in Sec. 4. Empirical validation of the developed relationships is presented in Sec. 5, with concluding remarks provided in Sec. 6.

2 A Foundation in Ohm's Law

The metrics that are developed in this paper have a foundation in Ohm's law. Ohm's law serves as an appropriate foundation because of an interesting phenomenon, which will be described later in this section. The observed phenomenon was sufficient to motivate the investigation into the application of Ohm's law to reverse engineering. The results documented herein, indicate that the developed relationships are appropriate for nearly all products and are accurate to the degree of an average error of 12.2%.

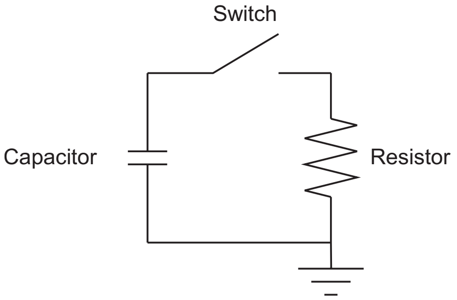

Fig. 1 Simple resistor-capacitor circuit. The capacitor is initially fully charged and begins to discharge the instant the switch is closed at $t=0$.

The history of Ohm's law is rich; Ohm [28] first presented Ohm's law in an 1827 publication. Since then, it was adapted and used to meaningfully characterize the behavior of many systems, including fluid systems [29], mechanical systems [30], thermal systems [31], and electrical systems [32,33].

To facilitate the ensuing developments, we consider the analysis of the simple resistor-capacitor circuit shown in Fig. 1, and outline mathematical relationships that enable the evaluation of a circuit's resistance R, capacitance C, and the time T to drain an initially charged capacitor. We present the following fundamental principles of Ohm's law because it is the foundation for the reverse engineering metrics and parameters presented in Sec. 3. Ohm's law characterizes the relationship between resistance, current, and voltage in a circuit as

$$R = \frac{V(t)}{I(t)} \tag{1}$$

while the capacitance C can be expressed as [32]

$$C = \frac{Q(t)}{V(t)} \tag{2}$$

where $V(t)$ represents the voltage difference across the resistor at current $I(t)$, and $Q(t)$ represents the charge stored in the capacitor. Notice that while V, I, and Q are time dependent, R and C are not. This important principle is used later in the paper to assist the designer in specifying reverse engineering parameters.

The resistance and capacitance of the circuit can be expressed in a way that is convenient to our discussion of reverse engineering. The convenience of this form is made evident in Sec. 3. When Q, I, and P are known and the following well-accepted [33] relationships are considered:

$$V(t) = \frac{W(t)}{Q(t)} \tag{3}$$

$$I(t) = \frac{Q(t)}{t} \tag{4}$$

and

$$W(t) = P(t)t \tag{5}$$

it follows that

$$R = \frac{P(t)}{I(t)^2} \tag{6}$$

and

$$C = \frac{Q(t)I(t)}{P(t)} \tag{7}$$

We reiterate that this form of the resistance and capacitance relationships is particularly useful in the context of information extraction during reverse engineering.

When R and C are known for a given system, the time to discharge a capacitor can be quantified as a function of the charge remaining in the capacitor by

$$T = -RC \ln\left(\frac{Q}{Q_0}\right) \tag{8}$$

where it is assumed that the capacitor begins to discharge at $t=0$, and T represents the time when the specified charge Q is remaining in the capacitor. An interesting characteristic of a discharging capacitor is that the discharge rate is dependent upon the voltage difference across the resistor shown in the resistor-capacitor circuit of Fig. 1. When the difference is large, the capacitor discharges quickly. When the difference is small, the capacitor discharges slowly. This behavior is exponential in nature.

This phenomenon is also observable in the reverse engineering of products. That is, the rate at which information can be extracted from a product is dependent upon the difference between the unextracted information that exists in a product, and how much of that information is known by the individual reverse engineering the product—we hereafter refer to this difference as *information difference*. For this reason, Ohm's law is the foundation for the metrics developed in this paper.

As Eq. (8) is an exponential relationship, the time to fully discharge the capacitor is infinite. For this reason, Q is often selected to be a positive nonzero value with the bounds

$$0 < Q \leq Q_0 \tag{9}$$

which results in a finite quantity of time.

Therefore, by these relationships, any resistor-capacitor circuit can be analyzed and, importantly, a prediction of time to discharge the circuit's capacitor can be made. Additionally, by using Ohm's law as a basic building block, circuits of any complexity can be analyzed using well structured, well-known approaches such as Kirchhoff's current and voltage laws [33]. As presented in Sec. 3, we use this same basic relationship to estimate the time required to discharge information about a product, from the product itself.

3 Development of Metrics and Parameters for Reverse Engineering

In this section, we present metrics and parameters for characterizing the barrier and time to reverse engineer any product. The presentation of the metrics and parameters is divided into three main parts in this section. Section 3.1 presents the general relationship for barrier and time to reverse engineer any product, with a brief description of the supporting parameters and metrics. Section 3.2 provides practical insight into specifying the needed parameters, and quantifying barriers and time for small subsets of a larger problem. Section 3.3 shows how the solutions to these small subsets can be reintegrated to solve the large problem.

3.1 General Metrics for Reverse Engineering. The barrier B to reverse engineer a product can be expressed as

$$B = \frac{P}{F^2} \tag{10}$$

where P is the power—the work per time to extract information—and F is the rate at which information can be extracted from a product. The time T to reverse engineer a product is

$$T = -BS \ln\left(\frac{K}{K_0}\right) \tag{11}$$

where K is the information contained by a product at a specific moment in time and K_0 is the information initially contained by a product. For simplicity, K is often defined as a fraction of K_0 (i.e., $K=0.05K_0$). Specifically, the quantity K is constrained to

$$0 < K \leq K_0 \tag{12}$$

which ensures that Eq. (11) yields a finite quantity of time. The quantity S in Eq. (11) is evaluated as

$$S = \frac{KF}{P} \tag{13}$$

where S is termed *information storage ability* of a product, which is analogous to electrical capacitance. As a note, while the general form of the equations presented in this section are true for K, F, and P at any time, it is worth noting that K_0, F_0, and P_0 are typically the simplest to specify.

Similar to the electrical relationships, the reverse engineering metrics can be rearranged to solve for any variable that is known or easily determined. In this paper, the metrics have been presented in a form that utilizes the variables K, F, and P as they are more readily determined than S, B, or T.

3.2 Decomposition of a Product for Barrier and Time Analysis. This section discusses how to determine the values of K, F, and P for the computation of the metrics as presented in this paper. In a realistic setting, it can be difficult to accurately determine the values of K, F, and P for the product as a whole. However, a product can be decomposed into disparate information components, allowing for a more simple quantification of K, F, and P for each component. In this section, we present an approach for decomposing a product based on information components, and analyzing each component to determine B and T. In Sec. 3.3, we discuss how the quantities B and T for each component can be systematically combined to determine the total barrier B^* and the total time T^* to reverse the product as a whole.

We start by discussing the parameter K, and the categorization of it. Recall that K is the estimated or actual information contained by a product, and that the purpose of reverse engineering is to extract information contained by a product from the product itself. Some examples of information contained by a product include material, geometry, electrical conductivity, and color. While there are many different ways where a product can be decomposed, we present a process by which products are decomposed according to categories of information contained by the product. For the purposes of the present paper, information contained by a product K is categorized according to the taxonomy chart in Fig. 2.

As seen in the taxonomy chart, the general information contained by a product can be separated into three basic levels. At the highest level, information is categorized into information types such as geometric information, material information, and function information. The second level of categorization separates each information type into information classes. For geometry, information classes include linear dimensions and radial dimensions, among others. When applicable, another categorization of the geometric information class can include microdimensions, mesodimensions, and macrodimensions. The final level of categorization on the taxonomy chart is the information subclass, which only has two categories—information that is pertinent to product performance and information that is superfluous. Generally speaking, the product should be decomposed into the minimum number of levels needed to easily specify the parameters K, F, and P for all the information contained by the product. As values for K, F, P, S, B, and T are specified or calculated for each information type, a subscript $[\ ]_i$ is used to distinguish information types or informa-

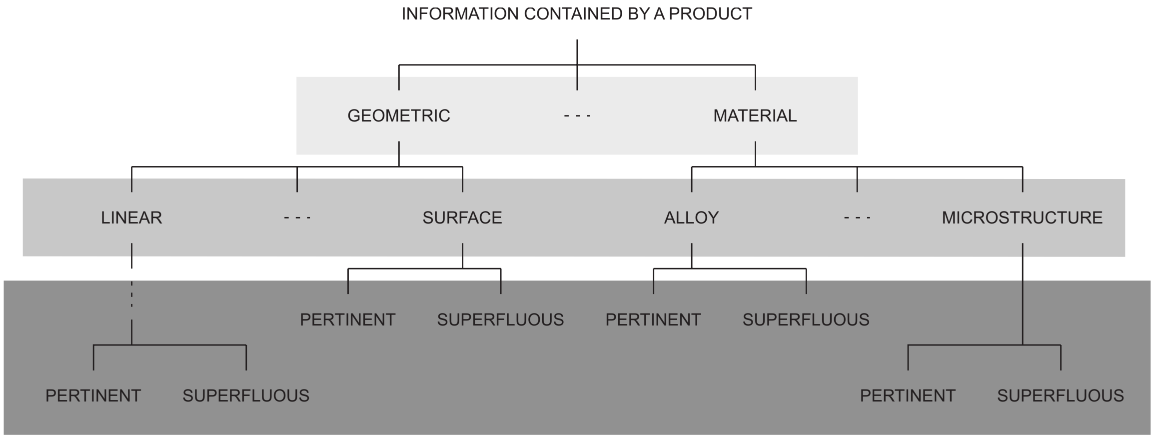

Fig. 2 A basic taxonomy of information contained by a product

tion classes—depending on the level for which K, F, and P are being analyzed—while the superscript $[\ ]^*$ represents the values of $[\ ]$ that pertain to the product as a whole.

With the different information types defined, K is more fully defined as the *estimated*, *unextracted*, *pertinent* information contained by a product *at a specific time*. The quantity of information contained by a product is therefore a function of information type i and time t. The quantity of pertinent information contained by a product is determined as the number of relevant units of information that is critical to the performance of the product.

For convenience in specifying the parameters K, F, and P, we define two reference time frames. Time in the t domain is the traditional representation of time, which captures any moment during the reverse engineering process. As it may be difficult to determine, the quantity of pertinent information contained by a product, and the rate at which it is extracted, at any time t when the product contains both pertinent and superfluous information, a second reference time frame is used. This second reference time frame, in the domain τ, is a theoretical time frame when all the values of K, F, and P are known, and all information is deemed pertinent. In the τ time frame, the time-independent quantities of B and S are more easily calculated. Since these quantities are time independent, they can also be used directly in the t time frame where there exists many unknown factors.

We pause now to make a clear distinction between $K(\tau)$ and $K(t)$. The parameter $K(t)$ represents only the pertinent information contained by a product, while the parameter $K(\tau)$ represents the *total* information contained by a product, be it pertinent or superfluous. In general, the most conservative value of $K(\tau)$ is when $K(\tau)$ is set equal to $K(t)$, implying that competitors know exactly what information is pertinent and what is superfluous. The quantity $K(\tau)$ is principally used for calculating S. A similar process of using two different reference frames is often used to determine the capacitance and resistance of electrical elements. If a resistor value is unknown, one can apply a known voltage and measure the current and determine the resistance of the system using Ohm's law. The resistance of a resistor is not dependent upon the electrical current, voltage, or time. Therefore, the resistance may be known for all times t, once it is known for a single time τ where a known voltage and current has been applied.

When a product is reverse engineered, no amount of superfluous information will benefit those extracting the information. For this reason, we are only interested in the rate F at which *pertinent* information can be extracted. For a product that contains both pertinent and superfluous information, it may be difficult to determine the flow rate of pertinent information when both pertinent and superfluous information is being extracted. For this reason, the flow rate of information is determined in the τ reference frame where all information is assumed pertinent. The quantity $F(\tau)$ is principally used for calculating S and B for individual information types.

Typically when extracting information contained by a product, the information that is quickly and easily extracted is extracted at a high flow rate. At times in the information extraction process, information becomes more difficult to extract, resulting in a lower flow rate. It is also apparent that the flow rate of one information type such as geometric linear dimensions may not be the same flow rate as another information type such as material grain orientations. The flow rate of information in the τ reference frame can be determined experimentally by measuring the time to extract information of particular information classes, such as geometric linear dimensions.

The measure of work per time to extract information contained by a product is characterized as power P. It is important to note that while an individual may put forth a consistent effort, the quantity of work achieved per unit of time does not remain constant during the reverse engineering process since some information requires little work to extract, while other information require significantly more work. Not only was this obvious from the empirical validations, but also the equations that define P—both in the electrical engineering perspective and in the metrics presented in this paper—show that P decays exponentially as a function of time. The quantity P is also determined in the τ reference frame and is used in calculating both S and B. The value of P is constrained by

$$0 < P \leq 1 \tag{14}$$

where zero represents no work being accomplished and 1 represents that maximum work is accomplished per unit of time while reverse engineering a product. The value of power should be selected to accurately represent the competitor's actual performance. Often it is simplest to specify P when $t=0$, therefore, we have specified P_0 to be a value of 1 for this study—which is the most conservative value of P_0. With the values of K, F, and P defined, B and S can be calculated according to Eqs. (10) and (13) for each information type i or for the product as a whole, if it can be evaluated as a whole. When the product cannot be evaluated as a whole, the developments of Sec. 3.3 become important.

3.3 Integration of Analyses for Overall Product Evaluation.

In this section, the total time to reverse engineer, and the total barrier to reverse engineering, are calculated by strategically combining the barrier and time to reverse engineer each information component, as discussed previously. In Sec. 3.2, we discussed how a product can be decomposed into various information types to facilitate the selection of K, F, and P, resulting in a B and T for each information type. Importantly, when multiple barriers exist for the same information type, those barriers may be added together in the same way electrical resistors in parallel and in series may be added together.

A different approach is required for calculating the total barrier and time to reverse engineer a product when it contains *multiple types* of information. Under this approach, each information type, including the respective barrier and storage ability, may be considered as an independent resistor-capacitor circuit. Calculating the total time to reverse engineer a product is analogous to quantifying the total time required to discharge multiple independent resistor-capacitor circuits where the number of circuits is equivalent to the number of information types contained by the product. Knowing the length of time required to discharge the independent circuits, the combined quantity of charge initially stored by the circuits, and the capacitance of the capacitors enables us to create a pseudo resistor-capacitor circuit that will result in the same quantity of time to discharge as the summed time of the independent circuits—when the pseudo circuit has the same capacitance and charge as the sum of the individual circuits. With the capacitance, charge, and time to discharge known for the pseudo circuit, the resistance of the pseudo circuit can be calculated.

To estimate the total barrier and time to reverse engineer the product as a whole, we perform a similar analysis on a pseudo product that has the same performance as one that has the considered information types combined, enabling an estimation of B and T for the entire product.

The total time T^* to reverse engineer a product, the total information K^* contained by a product, and the total storage ability S^* of a product can be determined by

$$T^* = \sum_{i=1}^{N} T_i \tag{15}$$

$$K^* = \sum_{i=1}^{N} K_i \tag{16}$$

and

$$S^* = \sum_{i=1}^{N} S_i \tag{17}$$

where N is the quantity of information types that the product has been decomposed into.

When individual information types are analyzed, the known values include F and P. With the pseudo product, however, the flow rate is calculated by

$$F^* = \frac{K^*}{T^*} \tag{18}$$

which enables P^* to be calculated as

$$P^* = \frac{K^* F^*}{S^*} \tag{19}$$

Note that Eq. (19) is obtained by rearranging Eq. (13) and solving for P.

Only now that the effective rate at which information can be extracted from the pseudo product and the power required to extract information are known, the effective barrier for the entire product can be determined by using Eq. (10). It is important to note that the barrier and time to reverse engineer a product are dependent upon skills and resources available (both affecting the flow rate of information). Therefore, the barrier to reverse engineer a product may vary depending upon the group performing the reverse engineering activities [34]. In general, the metrics presented in this paper will be more accurate if the individual reverse engineering is familiar with the reverse engineering process, the tools to be used while extracting information, and has a general understanding of the product being reverse engineered since the rate of information extraction often changes rapidly for those learning new processes or tools.

4 Model Limitations and Sensitivity Analysis

In this section we present the limitations for the reverse engineering metrics presented in this paper, as well as a sensitivity analysis of the input parameters.

The accuracy of the time and barrier to reverse engineer a product is dependent upon accurate selection of the parameters K, F, and P. Depending upon the reverse engineering perspective taken, some parameters may be more accurate, lending to a better estimation of the time and barrier to reverse engineer a product. Recall that there are at least two practical reverse engineering perspectives: that of the original designer who seeks to determine, and even maximize, the difficultly to reverse engineer their product; and that of the competitor who seeks to reverse engineer the innovative product.

When the original designer uses the relationships presented in this paper, he/she is able to accurately determine the actual quantity of pertinent information K contained by the product, but will only be able to estimate the rate at which the competitor can extract information F. The competitors, on the other hand, will be able to accurately determine the rate at which they (the competitors) can extract information F, but will be forced to estimate the initial quantity of pertinent information contained by the product. Additionally, it may not be obvious to the competitor what information is pertinent and what is superfluous—especially if the designers developed the product to be difficult to reverse engineer. It is likely that the original designers can estimate the information extraction rate for the competitors, more accurately than the competitors can estimate the quantity of pertinent information contained by a product. A simple approach would be for the original designer to specify a flow rate of information extraction, based on their own skill and motivation, as it is likely that their competitors have similar skills and motivation. Also, as discussed in Sec. 5, products must be of a sufficient complexity to ensure accurate estimations of B and T.

There are also limitations regarding the input parameter P. In this paper, we present $P_0=1$ for all of the presented examples. This is because the conditions defining $P_0=1$ can be understood in terms of product development, which involves a maximum effort being put forth with the maximum work achieved. Unfortunately, the conditions defining P less than 1 are not yet understood, and are the focus of a separate study by the authors. Fortunately, the conditions defining $P_0=1$ are also the most conservative. It is additionally beneficial that the term P does not affect the time estimation. The parameter P cancels out in the T equations, therefore negating any error that may be introduced due to a poor selection of P, however, the barrier estimation is still affected.

The first order sensitivity analysis confirms this notion and shows the sensitivity of B and T to the input parameters, which may be seen in Figs. 3 and 4, which present the percent error of the calculated B and T, respectively, with respect to the error of the input parameters. As can be seen from the figures, the flow rate of information extraction generally has the largest impact on the accuracy of the barrier and time estimations. Therefore, it is likely most beneficial to ensure that the flow rate is accurate. In our studies, we have found that a typical F error has been found to be $\pm 5\%$ when F is determined by the methods outlined in this paper.

5 Empirical Validation of Developed Metrics

In this section, we present an empirical study with the purpose of showing that the time and barrier to reverse engineering can be estimated by the relationships presented in this paper for products

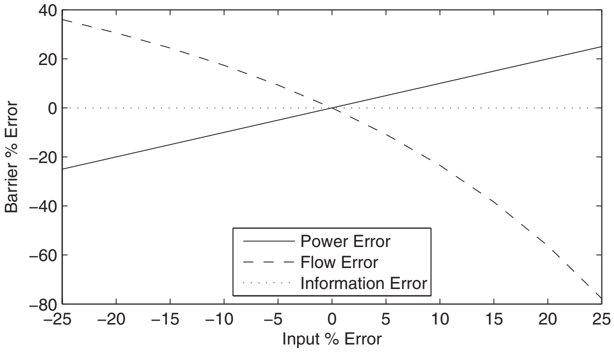

Fig. 3 First order sensitivity analysis of the reverse engineering barrier

Fig. 5 Part 127 as presented in Sec. 5

of sufficient complexity. For the empirical studies presented here, only geometric information is considered and K is assumed to be $0.05K_0$ (see Eq. (11)).

For any information type, the time to extract a unit of information varies from unit of information to unit of information (within a product) and from product to product. An effective and efficient way to handle the differing times is to determine an individual's general rate of information extraction; by general we mean valid for all products of sufficient complexity. As a note, this extraction rate is the rate of information extraction $F(\tau)$, as described in Sec. 3.2. We obtain $F(\tau)$ for geometric information experimentally by issuing a uniform dimension extraction test. The test is set up to allow the individual to familiarize themselves with the dimension to be extracted, then instructed to extract that dimension with a measurement tool while the time is recorded. This process is repeated multiple times for different dimensions to obtain an average dimension extraction rate of the individual using the measurement tool—a rate that is independent of the time spent developing the dimension extraction sequence or checking to ensure all dimensions have been extracted. The dimension extraction rate (F) is then used in Eq. (10), enabling calculation of T by Eq. (11) to estimate the time to reverse engineer any product of sufficient complexity, as discussed in this section. The accuracy of the exponential time estimations are dependent upon accurate measurement of the information extraction rate. When the actual information extraction rate is known, the estimated time is the same as the actual time to reverse engineer a product. The test we use has been found to be an adequate measure of information flow rates resulting in time estimations with an average error of 12.2%.

To illustrate, four individuals were asked to reverse engineer Part 127 and Part 128, as seen in Figs. 5 and 6, respectively. Before beginning the reverse engineering process, the information extraction rate F was determined for each individual by the process outlined above, K was determined by counting the dimensions required to fully describe each part, and the initial power was selected to be $P_0=1$, assuming that individuals put forth a maximum effort with a maximum work achieved. The individuals, without knowing the values of K, F, and P, were then instructed to extract and record the dimensions with enough detail that the product could be recreated if needed.

The plots seen in Figs. 7 and 8 are the results of a single individual reverse engineering each product and compared with the linear and exponential time approximations. The linear relationship is defined as

$$T=\frac{K}{F} \tag{20}$$

where the information extraction rate (F) of the individual is the slope and the number of dimensions (K) to be extracted is the y-intercept on a plot of dimensions versus time. While the plots are for a single individual, they are representative of all the individuals that reverse engineered the products and are consistent with other tests we have performed. The data in the plots has been rearranged according to the time to extract each dimension—with the shortest times plotted first—and are not plotted in the order of dimension extraction. Tables 1 and 2 present the predicted time to reverse engineer each product, for each individual, as well as the calculated barrier to reverse engineering. From Tables 1 and 2, we see that the barrier to reverse engineering is the same for both parts for each individual. This is due to the fact that the barrier is only dependent upon the individual and the type of information being extracted, and not dependent upon the quantity of information extracted.

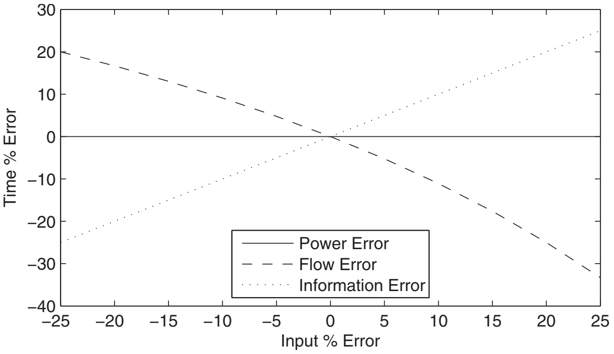

Fig. 4 First order sensitivity analysis of the time to reverse engineer a product

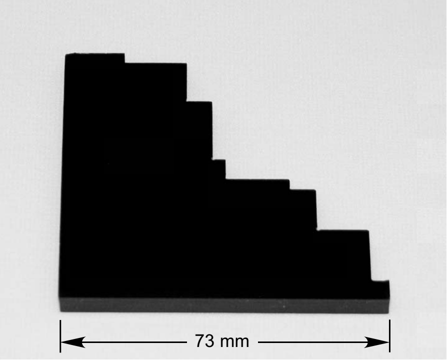

Fig. 6 Part 128 as presented in Sec. 5

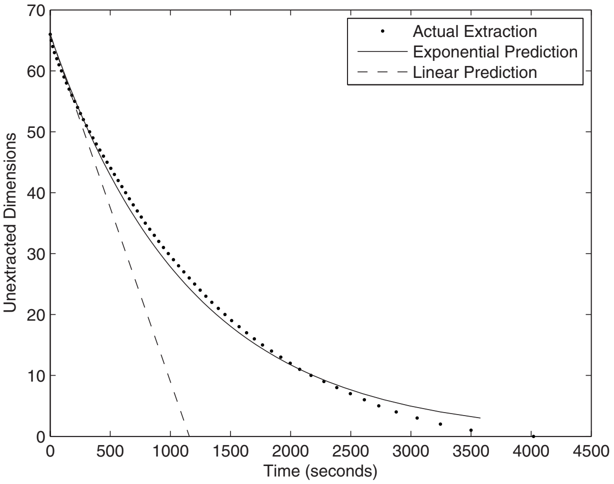

Fig. 7 Plot of unextracted dimensions remaining in Part 127 versus time as compared with the linear and exponential time predictions for Individual 1

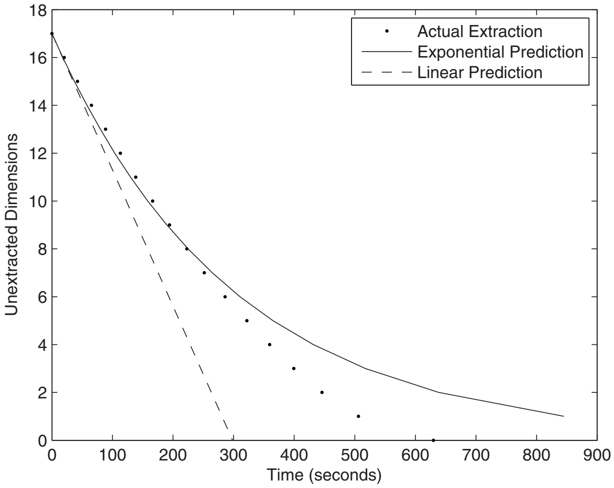

Fig. 8 Plot of unextracted dimensions remaining in Part 128 versus time as compared with the linear and exponential time predictions for Individual 1

Fig. 9 Figure of keyboard before disassembly

To determine the validity of the relationships presented, multiple individuals have reverse engineered multiple products, resulting in over 50 sets of data for geometric information extraction. By observation and data analysis, we have verified that the time to reverse engineer the geometry of a product can be approximated by an exponential relationship. We have also observed that simple products tend to be less accurately estimated by the exponential relationship. Part 128 was specifically selected to test the exponential relationship near the limits of application, and it may be seen that a linear approximation may be more accurate for the simplest of parts. However, Part 127, while still relatively simple, has been found to be sufficiently complex to be accurately estimated by the exponential relationship.

Products of higher degrees of complexity have also been analyzed and have also been found to be accurately represented by the exponential relationship. To illustrate this, we briefly discuss the reverse engineering of Apple Inc.'s recently released computer keyboard, as seen in Figs. 9 and 10. As with the previous examples, we will only reverse engineer geometry and do not reverse engineer the material properties or the keyboard electronics. However, if the flow rate of information extraction is determined for extracting material properties and analysis of electronics, the same relationships used for estimating the time and barrier to extract geometric information can also be used to estimate the time and barrier to extract information about material properties and the electronics of a system.

We reverse engineered the keyboard to the degree that we could recreate keyboard parts that would be interchangeable with the current product. In order to fully extract the geometric information contained by the keyboard, some disassembly was required. While disassembly time may be important to quantify [35,36], it was not the focus of this study or of the developed metrics. Therefore, the keyboard was considered disassembled when reverse engineering

Table 1 Table of predicted and actual times to extract geometric information from Part 127. Time is in seconds.

Individual	Actual time (s)	Linear prediction	Linear % error	Exponential prediction	Exponential % error	Barrier
1	4020	1158	−71.20	3579	−10.97	307.8
2	3473	847	−75.60	2656	−23.51	339.2
3	1367	517	−62.19	1433	4.84	260.9
4	2201	826	−62.48	2323	5.55	272.8

Table 2 Table of predicted and actual times to extract geometric information from Part 128. Time is in seconds.

Individual	Actual time (s)	Linear prediction	Linear % error	Exponential prediction	Exponential % error	Barrier
1	629	298	−52.65	845	34.16	307.8
2	568	298	−44.48	887	56.20	339.2
3	595	242	−59.28	656	10.27	260.9
4	522	264	−49.33	733	40.49	272.8

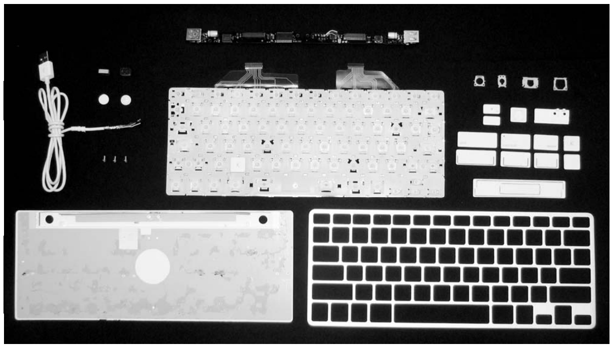

Fig. 10 Figure of keyboard disassembled

began. Utilizing the relationships presented in this paper to estimate the time and barrier to reverse engineer the keyboard resulted in a barrier of 307.8 and an estimated time of 25,649 s. In actuality, it took 23,667 s to reverse engineer the keyboard—an 8.38% error when compared with the predicted time of 25,649 s. The estimated and measured times were determined independently so that neither influenced the other. Figure 11 compares the actual time to reverse engineer the keyboard with the exponential and linear predictions.

6 Concluding Remarks

In this paper, we have presented general metrics for evaluating the barrier and time to reverse engineer a product. We have also defined supporting metrics and parameters for evaluating the barrier and time. The metrics and parameters presented are adapted from Ohm's law and are based on resistor-capacitor circuits and capacitor discharge time estimates. The effectiveness of the metrics outlined in this paper has also been demonstrated with an empirical study.

The presented relationships enable a systematic and consistent comparison of products—pre- or post-production. This brings the designer a distinct ability to quantify the amount of time to reverse engineer different variations of a product while in the early design stage. Such quantification can readily support trade-off studies of production costs with market strategies. The ability to quantify the barrier and time to reverse engineer a product early in the design process enables designers to strategically implement product features that will increase the difficulty of reverse engineering the product, while minimizing implementation cost. For those reverse engineering, the systematic estimation of the reverse engineering time facilitates management decisions such as reverse engineering costs, project timelines, and market strategies.

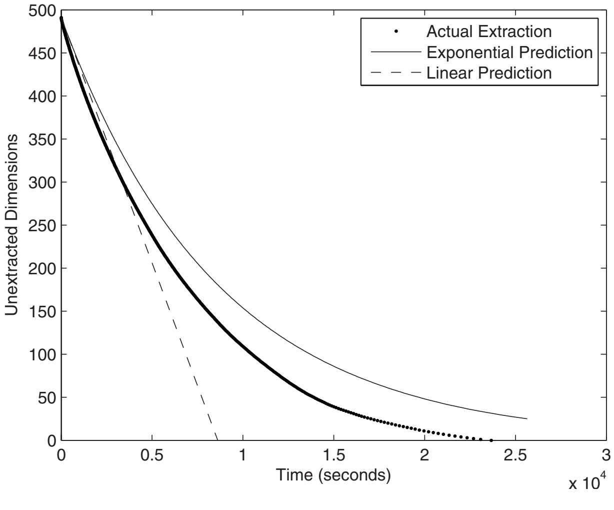

Fig. 11 Plot of unextracted dimensions remaining in keyboard versus time as compared with the linear and exponential time predictions

Ongoing and future developments include the use of these metrics in the design and optimization of products. Specifically, the metrics developed herein bring a new and additional objective of increasing the time and/or barrier to reverse engineer a product to the decision making process. We reiterate here that the total time T and total barrier B is the time and barrier only to extract information from a product. Other developments by the authors include the time and barrier to fabricate the reverse engineered product [37], and the influence of repetitive measurements on the reverse engineering process [38].

Acknowledgment

This research was supported by National Science Foundation under Grant No. CMMI-0800904 for Christopher A. Mattson and Brent L. Adams.

Nomenclature

B = barrier to extract information about a product from the product itself
F = estimated rate at which information is extracted from a product
K = estimated or actual information contained by a product
P = estimated power exerted to extract information contained by a product
S = a measure of a product's ability to contain information
T = estimated time to extract information K
t = reference time frame for reverse engineering a product
τ = reference time frame when all parameters are known

Subscripts, Superscripts, and Other Indicators

$[\ ]^*$ = indicates total measure of $[\ ]$
$[\ ](t)$ = indicates $[\ ]$ is a function of time, in the t domain
$[\ ](\tau)$ = indicates $[\ ]$ is a function of time, in the τ domain
$[\ ]_0$ = indicates $[\ ]$ is evaluated at time t or τ equal to zero
$[\ ]_i$ = indicates $[\ ]$ is of information type i

References

[1] Harston, S. P., Mattson, C. A., Adams, B. L., and Ahmadi, S., 2008, "Capitalizing on Heterogeneity and Anisotropy to Find Designs With Unexpected Performance," *Proceedings of the 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Victoria, British Columbia, Canada, Paper No. AIAA 2008-6015.
[2] Ingle, K. A., 1994, *Reverse Engineering*, McGraw-Hill, New York.
[3] MacMillan, I., McCaffery, M. L., and van Wijk, G., 1985, "Competitors' Responses to Easily Imitated New Products-Exploring Commercial Banking Product Introductions," Strategic Manage. J., **6**(1), pp. 75–86.
[4] Mansfield, E., Schwartz, M., and Wagner, S., 1981, "Imitation Costs and Patents: An Empirical Study," Econom. J., **91**(364), pp. 907–918.
[5] Chikofsky, E. J., and Cross, J. H., 1990, "Reverse Engineering and Design Recovery: A Taxonomy," IEEE Software, **7**, pp. 13–17.
[6] Csete, M., and Doyle, J., 2002, "Reverse Engineering of Biological Complexity," Science, **295**, pp. 1664–1669.
[7] Waters, R., and Chikofsky, E., 1994, "Reverse Engineering: Progress Along Many Dimensions," Commun. ACM, **37**, pp. 22–25.
[8] Buss, E., De, R., Gentleman, M. M., Henshaw, J., Johnson, H., Kontogiannis, K., Merlo, E., Muller, H., Mylopoulos, J., Paul, S., Prakash, A., Stanley, M., Tilley, S., Troster, J., and Wong, K., 1994, "Investigating Reverse Engineering Technologies for the CAS Program Understanding Project," IBM Syst. J., **33**, pp. 477–500.
[9] Nelson, M. L., 1996, "A Survey of Reverse Engineering and Program Comprehension," ODU CS 551—Software Engineering Survey.

[10] Muller, H. A., Jahnke, J. H., Smith, D. B., Storey, M.-A. D., Tilley, S. R., and Wong, K., 2000, "Reverse Engineering: A Roadmap," *Proceedings of the International Conference on Software Engineering (ICSE)*.
[11] Otto, K., and Wood, K., 2001, *Product Design*, Prentice-Hall, Upper Saddle River, NJ.
[12] Reed, R., and DeFillippi, R. J., 1990, "Casual Ambiguity, Barriers to Imitation, and Sustainable Competitive Advantage," Acad. Manage. Rev., **15**, pp. 88–102.
[13] Pahl, G., Beitz, W., Feldhusen, J., and Grote, K.-H., 2007, *Engineering Design: A Systematic Approach*, Springer-Verlag, London.
[14] Moles, C., Mendes, P., and Banga, J., 2003, "Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods," Genome Res., **13**, pp. 2467–2474.
[15] Kremling, A., Fischer, S., Gadkar, K., Doyle, F. J., Sauter, T., Bullinger, E., Allgöwer, F., and Gilles, E. D., 2004, "A Benchmark for Methods in Reverse Engineering and Model Discrimination: Problem Formulation and Solutions," Genome Res., **14**, pp. 1773–1785.
[16] Benyus, J. M., 2002, *Biomimicry: Innovation Inspired by Nature*, Harper, New York.
[17] Naumovich, G., and Memon, N., 2003, "Preventing Piracy, Reverse Engineering, and Tampering," Computer, **36**, pp. 64–71.
[18] Besl, P. J., and Mckay, N. D., 1992, "A Method for Registration of 3-D Shapes," IEEE Trans. Pattern Anal. Mach. Intell., **14**(2), pp. 239–256.
[19] Várady, T., Martin, R. R., and Cox, J., 1997, "Reverse Engineering of Geometric Models: An Introduction," Comput.-Aided Des., **29**, pp. 255–268.
[20] Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M., 2009, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments (Wiley Series in Probability and Statistics)*, 3rd ed., Wiley, New York.
[21] Simpson, T., Poplinski, J., Koch, P., and Allen, J., 2001, "Metamodels for Computer-Based Engineering Design: Survey and Recommendations," Eng. Comput., **17**, pp. 129–150.
[22] Scherer, F. M., 1967, "Research and Development Resource Allocation Under Rivalry," Q. J. Econ., **81**, pp. 359–94.
[23] Mansfield, E., Rapoport, J., Romeo, A., Villani, E., Wagner, S., and Husic, F., 1977, *The Production and Application of New Industrial Technology*, Norton, New York.
[24] Shapiro, C., 1985, "Patent Licensing and R & D Rivalry," Am. Econ. Rev., **75**, pp. 25–30.
[25] Nelson, R., and Winter, S., 1982, *An Evolutionary Theory of Economic Change*, Belknap Press, Cambridge, MS.
[26] Hill, C. W. L., 1992, "Strategies for Exploring Technological Innovations: When and When Not to License," Org. Sci., **3**, pp. 428–441.
[27] McEvily, S. K., and Chakaravarthy, B., 2002, "The Persistence of Knowledge-Based Advantage: An Empirical Test for Product Performance and Technological Knowledge," Strategic Manage. J., **23**, pp. 285–305.
[28] Ohm, G. S., 1827, *Die galvanische Kette, mathematisch bearbeitet*, T. H. Rieman, Berlin, Germany.
[29] Akers, A., 2006, *Hydraulic Power System Analysis*, Taylor & Francis, Boca Raton, FL.
[30] Chernov, I. N., Eyink, L., Lebowitz, J. L., and Sinai, Y. G., 1993, "Derivation of Ohm's Law in a Deterministic Mechanical Model," Phys. Rev. Lett., **70**, pp. 2209–2212.
[31] Jiles, D., 2001, *Introduction to the Electronic Properties of Materials*, CRC, Boca Raton, FL.
[32] Dorf, R. C., 1993, *The Electrical Engineering Handbook*, CRC, Boca Raton, FL.
[33] Rizzoni, G., 2004, *Principles and Applications of Electrical Engineering*, McGraw-Hill, New York.
[34] von Hippel, E., 1998, "Economics of Product Development by Users: The Impact of "Sticky" Local Information," Manage. Sci., **44**, pp. 629–644.
[35] Collado-Ruiz, D., and Capuz-Rizo, S., 2010, "Modularity and Ease of Disassembly: Study of Electrical and Electronic Equipment," J. Mech. Des., **132**(1), pp. 014502.
[36] Thevenot, H., and Simpson, T. W., 2009, "A Product Dissection-Based Methodology to Benchmark Product Family Design Alternatives," J. Mech. Des., **131**(4), p. 041002.
[37] Knight, D. C., Mattson, C. A., and Adams, B. L., 2009, "Maximizing Return on Investment by Constructing Optimal Barriers Against Competitors' Market Entry," *Proceedings of the 50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Proceedings of the 17th AIAA/ASME/AHS Adaptive Structures Conference, and Proceedings of the 11th AIAA*, Palm Springs, CA, Paper No. AIAA-2009-2224.
[38] Curtis, S. K., Harston, S. P., and Mattson, C. A., 2009, "A Generic Formulaic Characterization of the Time to Reverse Engineer Tolerances of a Product," ASME IMEC Paper No. 2009-12123.