Comparison of screwdriver tips to the resultant toolmarks

by

Julie Ann Kidd

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Materials Science and Engineering

Program of Study Committee:
Scott Chumbley (Major Professor)
Lawrence Genalo
Max Morris

Iowa State University
Ames, Iowa
2007

Copyright © Julie Ann Kidd, 2007. All rights reserved.
# TABLE OF CONTENTS

LIST OF FIGURES ................................................................................................................... iii

LIST OF TABLES ....................................................................................................................... iv

ABSTRACT ................................................................................................................................ v

CHAPTER 1. BACKGROUND .................................................................................................... 1
  1.1 Tools and toolmarks ......................................................................................................... 2
    1.1.1 Toolmark characteristics .......................................................................................... 3
    1.1.2 Toolmarks as they relate to firearms ....................................................................... 5
    1.1.3 Toolmarks as they relate to tools ............................................................................ 6
    1.1.4 How toolmark comparisons are made ...................................................................... 7
    1.1.4.1 Consecutive Matching Striae ............................................................................. 9
  1.2 A brief history of toolmarks ............................................................................................. 10
    1.2.1 Uniqueness with respect to firearms ...................................................................... 15
    1.2.2 Uniqueness with respect to toolmarks ................................................................... 15
  1.3 Rules of evidence and the Daubert criteria ..................................................................... 17

CHAPTER 2. STATISTICAL ANALYSIS OF TOOLMARKS .................................................... 20

CHAPTER 3. METHODS AND PROCEDURES ..................................................................... 27
  3.1 Scanning Electron Microscopy ....................................................................................... 28
  3.2 Profilometry .................................................................................................................. 29
  3.3 Stereomaging and reconstruction .................................................................................. 30

CHAPTER 4. RESULTS AND DISCUSSION ......................................................................... 36
  4.1 Matching of profilometer data from toolmarks .............................................................. 37
  4.1 Evaluation of marks at varying angles .......................................................................... 39
  4.2 Linking a tool to a mark ................................................................................................ 42
    4.2.1 SEM stereopair analysis ......................................................................................... 43
    4.2.2 Other methods ....................................................................................................... 52
      4.2.2.1 Profilometry on a tool tip ................................................................................. 53
      4.2.2.2 Confocal florescence and x-ray tomography .................................................. 54
      4.2.2.3 Laser profilometry ......................................................................................... 55
      4.2.2.4 Dye penetrant ............................................................................................... 56
      4.2.2.5 InfiniteFocus ................................................................................................. 57

CHAPTER 5. SUMMARY AND CONCLUSIONS .................................................................... 61

ACKNOWLEDGEMENTS ......................................................................................................... 63
LIST OF FIGURES

Figure 1. Striated toolmark ................................................................. 3
Figure 2. Optical comparison of a mark highlighting the match region ............ 7
Figure 3. Illogical match ..................................................................... 21
Figure 4. Matches selected along the steep edges of the profile .................... 22
Figure 5. Diagram of correlation and validation windows on comparative profiles 24
Figure 6. Poor separation (left), good separation (right) ........................... 25
Figure 7. Sections of a tool ................................................................ 27
Figure 8. Overlapping images of the tool ............................................. 28
Figure 9. Jig to make toolmarks ......................................................... 29
Figure 10. Detail of jig making marks ................................................... 29
Figure 11. Marked lead sample with red lines representing profilometer traces 30
Figure 12. Signal-noise ratio ............................................................... 32
Figure 13. Sufficient structure ............................................................. 33
Figure 14. Depth of focus (DoF) .......................................................... 33
Figure 15. Tilting around the y axis ..................................................... 34
Figure 16. Disparity .......................................................................... 35
Figure 17. Comparisons of profilometer mark match (left) and non-match (right) 38
Figure 18. Comparison of matches and non-matches ............................... 39
Figure 19. Comparison of known matches .......................................... 42
Figure 20. Comparison of non-matches .............................................. 42
Figure 21. Comparison of sides A and B ............................................. 42
Figure 22. 95% confidence intervals .................................................... 42
Figure 23. Profile drawn on image surface ......................................... 43
Figure 24. Multiple profiles drawn adjacent on a single image .................. 44
Figure 25. Relative noise of SEM and profilometry profiles ....................... 44
Figure 26. Diagram showing sources of individual profile ....................... 45
Figure 27. Comparison of stereopair profiles to profilometer data ............... 46
Figure 28. Comparison of SEM profiles at high 500x, 750x, 1000x respectively 47
Figure 29. SEM results using AutoCalibration and a digitally controlled stage 48
Figure 30. Comparisons of adjacent profiles of SEM tip 39b to profilometer 39b 49
Figure 31. Comparison of correlation values in profilometer v. SEM matches 50
Figure 32. Comparisons between same sections of toolmark (top graph) and the tool (bottom graph) ................................................................. 52
Figure 33. A match (left) and a non-match (right) comparing profilometry of the tool tip and the toolmark ................................................ 54
Figure 34. Comparison of matches and non-matches between profilometry of tip and mark ........................................................................ 54
Figure 35. Imaged edge of tool in confocal florescence system .................. 55
Figure 36. Surface measurements with the laser ...................................... 56
Figure 37. InfiniteFocus comparison of primary profiles of tool 38 and toolmark 38 58
Figure 38. Validation output from IFM profiles ...................................... 59
LIST OF TABLES

Table 1. T Statistic comparison of IFM profiles ................................................................. 59
ABSTRACT

The subjective nature of correlating tools and toolmarks has been called into question since the 1993 Florida Supreme Court ruling in *Daubert v Merrell Dow Pharmaceuticals, Inc.* This has led to law enforcement agencies and officials placing an emphasis on developing objective techniques with known error rates to replace the traditional subjective comparisons. Additionally, if such objective techniques could be automated the heavy workloads currently faced by forensic examiners could be reduced. Development of a semi-automatic process that utilizes a three dimensional profilometer shows potential as a technique that may yield statistically verifiable results, removing the subjective nature currently inherent to toolmark evaluation, and be automated.

This work involves characterizing a number of consecutively manufactured tools with a scanning electron microscope (SEM) and comparing that tool to the resultant mark. By using software to analyze both the roughness of a tool and a toolmark—evaluated by SEM and profilometry—the two surfaces can be statistically compared and a correlation determined in a region of best fit. The project has sought to answer two distinct questions: Can a toolmark be related to a particular tool (and only that tool) on a statistical basis? Can a series of toolmarks be obtained and compared in an automated manner to yield a statistically valid match? Providing answers to these questions based upon quantitative techniques rather than subjective analysis removes the uncertainties raised by the Daubert decision. The methods employed have the potential for automation, thereby offering a means for decreasing examiner workload. Thus, successful completion of this project could lead to development of an automated system that produces statistically valid and verifiable results.
CHAPTER 1. BACKGROUND

Optical characterization between tools and toolmarks through a method of comparative matching is a technique that has been utilized for nearly a century. Experience has shown that tools, generally accepted to possess unique surface characteristics, can be accurately paired with toolmarks, i.e., marks made on softer surfaces by the tool. Marks are often left on metal when a tensile, shear, or compressive force is applied. Comparative identifications of tools and corresponding toolmarks have been used to prove that a particular tool was responsible for a mark in criminal investigations\(^1\), with the assumption that each mark represents unique characteristics of the tool that created it. A similar assumption of uniqueness has been held true of fingerprints, although scientific studies to prove this assertion have not been conducted.

In 1993, the case of Daubert v. Merrell Dow Pharmaceuticals created a higher standard for federal courts to accept expert witness testimony; the new standard calls for scientific knowledge with a basis in the scientific method to be the foundation for testimony of expert witnesses (in this field, toolmark examiners). The field of toolmark examination has therefore been forced to examine the validity of the basic assumption that toolmarks are unique. Development of a method of analysis that reduces the subjective nature of comparative evaluation and provides statistical confirmation of a match, with known error rates and confidence intervals, is desirable.

This study will utilize fifty sequentially produced screwdriver tips obtained from Omega Company; sequential production makes them as similar as possible. The tips were examined using the SEM, profiles of their surfaces generated, and toolmarks made
by an expert examiner. The toolmarks were then examined with the SEM and surface profilometry. SEM images, reconstructed into stereopairs, and data collected from the profilometer provide quantitative information about the three dimensional surface.

This study seeks to establish a quantifiable match between a tool tip and the resultant toolmark using SEM and profilometry. Subsequent description of the match will be in statistical terms. Additionally, an automated method for comparing tools and toolmarks that assigns comparisons a ranking representative of the match quality, based in statistics, will be designed and tested.

### 1.1 Tools and toolmarks

The Association of Firearm and Toolmark Examiners (AFTE) defines a tool as “an object used to gain mechanical advantage; also thought of as the harder of two objects which when brought into contact with each other, results in the softer one being marked.”\(^2\) This definition allows for a broad range of objects to be classified as tools. The area of the tool that comes into contact with the softer material to leave behind a mark is known as the working surface of the tool. Toolmarks, which can be made by virtually any object, are created when the tool’s working surface comes into contact with a softer material, and leaves a representation of its surface.

Comparisons of tools and toolmarks fall into two key categories according to Biasotti and Murdock: pattern fit and pattern transfer. Pattern fit, also described as a physical match or a fracture match, is a term describing the unique features of surfaces fitting together uniquely; the more contours a surface possesses the higher the probability of a unique match. For example, if a piece of glass was fractured into two pieces and the
pieces were fit perfectly back together, a pattern fit would have been made. Pattern transfer is not as simple as pattern fit; it involves the impressions and striations of two and three dimensional marks. Toolmarks are considered pattern transfer, Figure 1. Impressions are created when force and motion applied to the tool are perpendicular to the surface being marked. For example, a hammer impact is an impression. Contours created when force and motion are applied parallel to the surface being marked are known as striations. Scraping a surface with a pry bar creates a striated toolmark.

![Figure 1. Striated toolmark](image)

### 1.1.1 Toolmark characteristics

Individual characteristics are a completely unique series of features on a surface “produced by the random imperfections or irregularities of tool surfaces. These random imperfections or irregularities are produced incidental to manufacture or are caused by use, corrosion or damage,” according to F. Taroni, author of “Statistics: A Future in the Toolmarks Comparison?” Individual characteristics are unique and distinguish it from all other tools of similar type.

Class characteristics are indicative of the source of the tool; they are marks characteristic of the class of tools, often resulting from the tool design. Class characteristics are typically more macroscopic in nature. For example, in the area of
firearms class characteristics are related to the matching of caliber of the firearm and cartridge or bullet, and the rifling pattern contained in the barrel of the firearm as it is transferred to a bullet.  

Subclass characteristics are more distinctly defined—they are related to manufacture, have, a narrow source of origin, and are ever changing. An example of a subclass characteristic would be a tool produced from a common master that shares characteristics present only in other tools produced by the same master. In 1949, Churchman observed subclass characteristics in a series of bullets fired from consecutively made rifle barrels. Twenty-six years later, in 1975, Skolrood’s observations were similar to Churchman’s; he detected subclass characteristics when he examined three similar rifle barrels. Nichols explains what qualifies a characteristic as a subclass characteristic:

If one were to examine a cast of the bore of a firearm, such characteristics would have to exist for the entire length of the cut surface. If a certain characteristic appeared after the cut surface had already started, then it would be an imperfection caused by the current process. If it disappeared before the end of the cut surface, then it is gone and by definition of its absence cannot be passed onto the next cut surface. Therefore, the only characteristics capable of being defined a subclass would be those that persist for the entire length of the cut surface.

Examiners also have found class and subclass characteristics in toolmarks. In 1968, Burd and Kirk’s study of screwdrivers that had not experienced finishing work had the potential to show subclass characteristics. Miller documents research into subclass characteristics present in tongue and groove pliers, nails, metal punches, metal snips, and screwdrivers. In each instance, subclass characteristics were present and yet experienced
toolmark examiners were able to distinguish between different tools used to create the marks. 11

Subclass characteristics are partly defined by their ability to evolve over time. The evolution of subclass characteristics in firearms is attributed to use—this may include cleaning, handling, or dismantling. The barrel interior is affected primarily by erosion, corrosion, and deposition of particles. Bonfanti’s review of literature explores the lifetime of a subclass characteristic; she emphasizes differences in subclass characteristics from weapon to weapon and the need for the subjective interpretation of photographic evidence by a toolmark examiner. 12 Even in consecutively made toolmarks from the same tool, differences in individual surfaces may be present. However, the slow change of tool surfaces does not prohibit identification criteria to be established and positive identifications to be made. 13 Experienced examiners, those who understand the differences between class characteristics and individual characteristics, are crucial to distinguishing true matches. 14

1.1.2 Toolmarks as they relate to firearms

Toolmarks created by firearms have been extensively studied with the purpose of determining whether a specific weapon fired a specific bullet or cartridge. The bore of a firearm consistently creates unique toolmarks on bullets as a result of compressive and tensile forces. The bores of firearms are rifled to allow the bullet to spin in a more controlled pattern, increasing stability and accuracy, and it is the markings created by this rifling that are transferred to the bullet when it is fired. Similarly, markings on the firing pin, breech, and ejector mechanism can be transferred consistently to the cartridge of
each bullet as it is fired. Thus, a large number of markings exist in a firearm investigation.

If a weapon is suspected of being a match with a piece of evidence and the firearm is in the examiner’s possession, marks are relatively easy to produce due to a standard method of operating a firearm. Characteristic marks from firing either will or will not be transferred.

1.1.3 Toolmarks as they relate to tools

As described earlier, toolmarks are primarily classified as impressions and striations. To prepare to make a comparison from a specific tool, test marks must be created. Toolmark identification, in comparison to firearm identifications, faces different challenges—no standard shape or size can be expected from a toolmark. Toolmarks will vary as a function of how they were made—pressure applied, angle of the tool and twisting all introduce variations to the mark. Tools are more often subjected to abuse (using the tool for something other than its intended purpose) than firearms, creating the possibility of significantly altering the original surface. It is important to try to replicate the conditions that made the evidence mark to make the comparison as similar as possible.

When a tool and a toolmark are suspected of having a correlation, the tool is used to make a series of marks in an attempt to produce a mark similar to the evidence mark. The test marks are generally made in lead or a similar soft material. Because tools lack a standard method of use, many marks may need to be made with varying angles and pressures to replicate the found toolmark. According to Biasotti and Murdock, preparing
a comprehensive set of test marks yields the best chance of positive identification, if one does indeed exist, and reduces the probability of a false inclusion. The series of preliminary marks would be compared to one another to predict if the tool did make the mark. Biasotti and Murdock expect the following four items to be true if the tool did create the toolmark in question: the tool was used to make the evidence mark; the working surface had not been altered since the creation of the evidence toolmark; the evidence toolmark is characterized by unique features; and the surface is not simply a subclass surface or a surface that simply possesses characteristics of other similarly created tools.  

1.1.4 How toolmark comparisons are made

Toolmark comparisons, for both firearms and toolmark identifications, are typically made using an optical system that utilizes a camera mounted on a microscope. The angle of illumination used produces regions of high and low reflectance due to the uneven surface, yielding an effectively two-dimensional method of comparison. A comparative microscope employs dual stages that allow two samples to be compared simultaneously. The comparative microscope image, Figure 2, compares a section of an ‘evidence’ sample on left to a ‘standard’ sample on right—the region of match is highlighted by the box.

**Figure 2. Optical comparison of a mark highlighting the match region**
It should be noted that the two-dimensional optical striations have intrinsic three dimensional roughness. While evaluation by light microscopy is easily accomplished, it is time intensive and is dependent upon the light source and view point. The image can change substantially as the lighting is changed, while the three dimensional roughness remains constant.

The AFTE Theory of Identification as it Relates to Toolmarks\textsuperscript{17} remains qualitative, describing a match as a “sufficient agreement.” “Sufficient agreement” refers to the duplication of the surface contours of the tool in the toolmark. Two or more sets of contours are compared utilizing features of heights of peaks, widths of peaks and valleys, and curvatures. Agreement is considered “significant” when it exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools. The agreement must also be self-consistent, i.e. toolmarks known to have been produced by the same tool and be identifiable and consistent with each other. The statement ‘sufficient agreement’ exists between two toolmarks means that the optical agreement between the two patterns displayed by the marks is so exact that the likelihood another tool could have made the mark is considered a remote possibility.\textsuperscript{18}

A mark examined with a comparison microscope is classified as: identification, inconclusive, eliminated or unsuitable. Identification is defined as characteristics whose agreement exceeds those of known non-matches and is consistent with the expectations the toolmark examiner has. Inconclusive matches may demonstrate some agreement of all characteristics but an insufficient amount to declare a match, agreement of all class characteristics without agreement in individual characteristics, or agreement of class characteristics with or without agreement of individual characteristics. Elimination
occurs when there is significant disagreement between class and/or individual characteristics. A classification of unsuitable indicates that the sample is not appropriate for comparison on the comparative microscope.  

Biasotti and Murdock recommend the following be considered until a method of matching is developed: studies of consecutively manufactured tools; importance of method of manufacture on working surfaces, mechanical and mathematical models of consecutive marks, quantity and quality of match agreements among known non-matches.  

1.1.4.1 Consecutive Matching Striae

Historically matches have been made using a method described as pattern match. Consecutive matching striae (CMS) introduce a more statistical approach to identifying matches. While pattern matching techniques seem qualitative, patterns assessed often possess features that can be quantified. The following are examples of features sought to make a match that can be quantified: positions of striae relative to a reference point; height and width of striations; and consecutive series of known height and width striae. CMS may be best described as a method to determine the best non-match observed; the understanding that matching striations occur in known non-matches creates a standard comparison within the examiner’s experience for the minimum number of matching striae for an examiner to confidently declare a match.

Pattern match and CMS represent the same science but use different methods to describe it. Additionally Nichols believes an examiner who utilizes CMS (described as a line counter) may appear more impartial because the method used to describe the work is
more likely to be understood by a lay person and they are able to utilize the best non-match from someone else’s observation to supplement their own training and experience.\(^{21}\)

A variety of CMS of known matches are reviewed by Ronald Nichols. Different firearms and toolmarks are examined and each author puts forth the minimum number of CMS they determine through their examinations. Examination of these works allows examiners to call upon other examiners’ experiences to help them evaluate specific cases of evidence before them. As a point of reference, Biasotti and Murdock suggest that for three dimensional toolmarks (only ridges are counted), at least two separate groups of three CMS appear relative to one another or six CMS in a single group when compared to a test mark; for two dimensional toolmarks (striae that match exactly with respect to width and relative position), two groups of five CMS appear relative to one another, or a single group of eight CMS when compared to the test mark. Biasotti and Murdock’s conservative recommendation for CMS identification has been supported by approximately 4,800 known non-match comparisons; these tests reported include no false inclusions based on their criteria.\(^{22}\)

1.2 A brief history of toolmarks

Early firearms identification began in London, England in 1835. Henry Goddard, considered to be the father of modern firearm identification, identified the mold used to manufacture the ball fired from a firearm used in conjunction with a murder. He was also able to establish a link with paper used to seal the ball from the gunpowder in the black powder firearm to paper found in the guilty man’s residence.\(^{23}\) Dr. R. Kockel of the
University of Leipzig was one of the first to publish his findings with regard to striations around 1900. Kockel investigated knife cuts in wood at various angles, examining them with “powerful magnification, oblique lighting and measure of relative spacing by vernier calipers.”²⁴ The Charlie Stielow trial of 1915 was key to assembling a group of prominent forensic researchers to make their science better understood. C. E. Waite, an investigator from New York, determined that Stielow’s firearm was not used to commit the murder for which he had been convicted. Waite, seeking more information about forensic toolmarks, assembled a group of individuals to research the topic of forensic identification. The group included himself; Phillip O. Gravelle, a microbiologist and photographer; John H. Fisher, a tool designer; and Calvin Goddard of the Bureau of Forensic Ballistics. This group was influential in creating the comparative microscope for firearm examination, and in 1925 the group became known as the Bureau of Forensic Ballistics which went on to found the Scientific Crime Laboratory at Northwestern University in 1930.

Early cases focused primarily on ballistic forms of toolmarks. Dr. Thomas, of the University of Ghent, released a paper in 1948 on toolmarks left by an axe in a victim’s skull; the article illustrated the technique used to determine the match. Thomas’ paper was an early example of non-firearm toolmark study.

In 1958, John Davis developed an instrument known as the striagraph—an instrument to map surface roughness in lines—that can be considered a forerunner of the profilometer. This instrument provided a new way to map toolmarks. Unfortunately, data storage was the limiting factor on the amount and quality of analysis that could be
performed. This allowed optical examination techniques to persist into the twenty first century.\textsuperscript{25}

In 1969, the Association of Firearms and Toolmark Examiners (AFTE) was established in the United States; this group remains the premier organization overseeing the development of toolmark identification. The formation of this group allowed people interested in toolmarks to come together to share ideas and educate each other about available techniques.

The 1970s saw research published that utilized SEMs to capture images of striated surfaces.\textsuperscript{26,27,28,29} The nature of image collection in the SEM eliminated the challenge of directional light. Investigations that employ SEMs for toolmark analysis emphasize three advantages inherent to the SEM: depth of field; magnification; and imaging. Large depth of field is advantageous for deep impressions. Examples given include ballistic examples of firing pins and an entire bullet. Magnification ranges from approximately fifteen times to thousands of times; this allows examination of a large variety of surface features. SEMs are designed to produce high quality images, useful for reference and demonstrations in court. SEM analysis also proves advantageous in the case of damaged bullets or casings, when the marked area is smaller than usual, or the striae detail is shallow. The magnification capabilities provide these advantages. Mann’s study of a 7 mm copper jacket sample demonstrated how a limited area available for comparison that could not be decisively matched with an optical system resulted in two consecutive matches in the sample utilizing high magnification in the SEM.\textsuperscript{30}

Utilizing SEMs for forensic toolmark identification evolved in three phases. The first phase, presented in 1972,\textsuperscript{31} placed a sample in the chamber of the SEM for imaging.
After imaging the first sample, it was exchanged for the second sample. After the second sample was imaged the resulting micrographs could then be compared. The second phase required two SEMs operating in tandem with software that allowed for side by side analysis. Most recently, a firearms comparison stage was placed in an SEM so that samples could be adjusted and viewed live, mimicking the comparative optical microscope.

The 1972 publication by Grove, “Examination of Firing Pin Impressions by Scanning Electron Microscopy” examined firing pin impressions from a series of 16 semi-automatic weapons with a SEM at a magnification of fifty times. They found images gathered with the SEM to be of superior resolution and to have greater consistency among comparisons in contrast to images produced with an optical microscopy system. Grove also sites depth of field as a characteristic of SEM images that is superior to optical units. In a paper released supplementing the 1972 publication, he examined firing pins from shotguns and rifles. He found that fifty percent of the shotgun impressions and 75 percent of the rifle impressions could be positively identified. His successes with identification indicate the value that this technique possesses.

Firearms examiners are typically limited by time constraints to compare tools and toolmarks one by one in a manual process. The advancement of computing and the development of large data storage devices in the late twentieth century have enabled more research to be done that utilizes three dimensional mapping of surfaces.

The National Integrated Ballistic Information Network (NIBIN) allows federal, state and local law enforcement agencies to quickly acquire and compare, across the
nation, markings made by firearms on bullets and cartridge casings. A semi-automated imaging process allows information to be uploaded into a database for comparison with other markings, providing an automated approach to comparisons among a database consisting of 228 sites, which utilize the automated integrated ballistic imaging system (IBIS). The NIBIN makes images available which previously would have been virtually inaccessible. IBIS merges its forbearers BULLETPROOF and BRASSCATCHER, automated projectile and cartridge casing comparative systems respectively, to provide firearms examiners a database for comparison. The NIBIN now logs information on more than 26,000 crime scene evidences and has provided more than 12,500 positive matches.

While the NIBIN currently employs two-dimensional technology, advances in measuring surface topology, including laser techniques, are being introduced to provide three dimensional comparisons. The use of a laser eliminates variations from light intensity, material of the sample, angle of the light, and the type of light used. However, three dimensional characterization requires more instruments to use and increased documentation making it an expensive technique.

For more than 170 years toolmark identification has been on the minds of people, many with formal training in scientific fields. The years of their combined work has created a body of knowledge that today’s technology can build upon to create a more accurate and faster method of evaluation of toolmarks. However, in all the years in which comparative identification has been used, the basic assumption underlying firearm and toolmark analysis, namely, that the mark in question could come only from a single gun or tool, has rarely been questioned. Indeed, studies that relate a particular mark to a
particular tool, and analyze the quality of that comparison on a statistical basis, are nonexistent. Without such studies the reliability of evidence submitted in legal proceedings can be (and has been) called into question.

### 1.2.1 Uniqueness with respect to firearms

The forensic study of markings of bullets and shell casings is an area of toolmark examination that has been widely studied, with standards in place for evaluation. These methods form the basis of a standard method of examination for other types of toolmarks. Bullets carry unique marks inherent from their manufacture and from the path they experience during firing. Optical methods are used to compare striations in bullets and corresponding marks.\(^{37}\)

### 1.2.2 Uniqueness with respect to toolmarks

The theory that tools have and make unique marks has been the premise for many investigations. Over the last century many studies have been conducted in an attempt to prove this theory; the majority find that toolmarks are indeed unique. The preponderance of this research, as well as more thorough investigations into toolmark evaluations, has been performed with firearms, making these ballistics findings a valuable resource when investigating other types of toolmarks.

After a critical toolmark match was deemed inadmissible by the State of Florida in Ramierz v. State of Florida due to the fact that the uniqueness of toolmarks had not been specifically evaluated, evidence was gathered to support the fact that toolmarks are indeed unique. As early as 1926, Calvin Goddard reported that every piston barrel (a ballistic tool), even those newly manufactured, has unique characteristics on the barrel’s
surface. Cassidy in his examination into tongue and groove pliers concluded that the wear and damaged areas on workings surfaces of tools over the course of years of use make them unique. Due to this, a tool will leave marks that only that tool can produce. Butcher and Pugh examined successively manufactured bolt cutters and Watson consecutively manufactured knife blades and crimping dies—from which it was hypothesized that toolmarks are indeed unique, their uniqueness rising from defects in materials and manufacturing. Biasotti and Murdock agree with the theory of uniqueness stating that “it is possible to individualize toolmarks because there are practical probability limits to: (1) the number of randomly distributed consecutive matching striae, and (2) the number of randomly distributed matching individual characteristics in impression toolmarks in known non-match positions.” Thus, they suggest a quantitative match criterion based on two groups of three consecutive striae in three dimensional toolmarks in positions relative to each other, or a single group of six consecutive striae. In two-dimensional toolmarks, they suggest two groups of five striae or a single group of eight striae when compared to the test toolmark.

Not all reports support the uniqueness of all tools. Extensive ballistic testing reviewed by Bonafanti and DeKinder call attention to the few cases tested where positive identification could not be made solely from bullets fired from a specific gun. In the majority of the cases positive identification is possible—however, that does not diminish the significance of the few for which identification was not possible. It should be noted, however, that these studies really did not address whether marks were unique. Rather, they addressed the question concerning the ability to distinguish between marks, which is a more practical question for a forensic examiner. Assuming uniqueness, Bonafanti and
DeKinder reviewed the reproducibility of marks made by firearms after extensive firing. Some tests were able to correctly identify the $5000^{th}$ bullet to the first bullet while others were only able to match the $50^{th}$ bullet to the first. A trend also existed chronologically; tests performed more recently had higher success rates than their forbearers. This may be attributable to the advances in imaging technology.  

Testing the uniqueness of firearms is simple compared to the effort needed to replicate toolmarks made with other sorts of tools. To discharge a firearm, a predictable, controlled series of events occurs. This is not the case with many simple tools. A screwdriver, for instance, may not simply be used to drive screws, it may also be used to pry or scrape, making the wear pattern difficult to predict. For example, a large number of people could use a firearm and the markings produced on the bullet and cartridge could be exactly the same. However, if two different people were asked to use a screwdriver to pry open a locked door, the number and variety of marks could vary widely due to applied pressure, twist of the tool, angle at which the person held the tool, etc. While it is true that marks made at the same angle, twist, and pressure should be similar, the sheer number of potential marks greatly increases the complexity of the comparison. The conclusion one can draw from examination of these studies is that uniqueness, while supported by the literature, has not been proven.

### 1.3 Rules of evidence and the Daubert criteria

In 1993, Daubert v. Merrell Dow Pharmaceuticals changed the standard for expert witness testimony that had been in place since the 1923 Frye v. U.S. case created the standard accepted in federal courts and most states. Expert witness testimony defined by
Frye was accepted if it had gained general acceptance in the particular field in which it belongs. It was believed that if the scientific field from which the testimony originated thought it valid, it had been thoroughly tested. The Federal Rules of Evidence, passed by congress in 1975, set a new standard to define what was acceptable as expert witness testimony. The Frye standard was so firmly rooted that it continues to be the accepted method in many states. Under Daubert, trial judges determine whether expert witness testimony is admissible based on four criteria: testability of scientific principle, known or potential error rate, peer review and publication, and general acceptance in a particular scientific community (the Frye criterion is included as a requirement under Daubert).

Daubert controls all federal cases, but not all states accepted the Daubert criteria as their standard. Many states have modified or expanded the Fry criteria and use their individual standard for expert witness testimony, creating a disparity between standards for state and federal courts regarding acceptability.

Forensic examiners face a series of challenges in order to meet the Daubert criteria with regard to toolmarks. Matches are based upon theories believed and supported by much evidence but disagreement within the community still exists about factors like uniqueness and what is required to declare a match. Matching is still defined as a qualitative technique and while CMS does introduce a quantitative method of evaluation there is no clear cut standard because each case is unique. M.J. Saks, author of “Implications of the Daubert Test for Forensic Identification of Science,” suggests that forensic identification can meet the standards Daubert imposes when three premises are fulfilled:
That many kinds of physical entities exist in unique, one-of-a-kind form...That they leave correspondingly unique traces of themselves...That the techniques of observation, measurement, and inference employed by forensic identification science are adequate to link these traces (toolmarks) [sic] back to the one and only object that produced them. 46

Manufacturing processes inherent to tool production create classes and subclass characteristics which can be separated from individual characteristics. If a tool can be used to create reproducible toolmarks that contain the individual characteristics, then it can be said it leaves unique traces of itself. By adopting a universal quantitative system of description, toolmarks will become less questionable under governing rules of Daubert.

Quantitatively determining a positive match is the challenge set by Daubert. The AFTE Theory of Identification47 remains qualitative—an evaluation made by a trained examiner. This study seeks to make identifications between tools and toolmarks utilizing an automated, statistical approach. Relative position of surface features will remain key to evaluation but this method does not require individual counts of striae. Instead, toolmarks will be made and compared to known matches and non-matches. The comparisons of the two data sets, matches and non-matches, will be a statistical indication of the likelihood of a tool being related to an individual mark.
CHAPTER 2. STATISTICAL ANALYSIS OF TOOLMARKS

An algorithm used to compare profiles of tool and toolmarks has been the ongoing work at Iowa State University of Professor M. Morris, David Faden and Jeremy Craft. It is a project occurring in parallel to the toolmark study with a partnership in place for the benefit of both groups. Morris’ group focuses on developing a successful match criteria based on the statistical analysis of supplied data rather than the current subjective method of visual matching. Supplying data for testing of their algorithm has been one of the objectives of the current project. In order to understand the validity of the data it is first important to understand how the algorithm works.

The computer program compares and matches data in a rapid and objective manner. It implements the afore mentioned algorithm that sequentially compares user specified segments (windows) from each trace, to imitate the work of a forensic examiner, and determines the best match between the two profiles. It is important to bear in mind that a match will always be found.

The program has experienced many stages of development. Problems with early versions of the program included orientation of the comparisons, placement of the matches, tendencies to match straight lines, values consistently higher than 0.9 which lead to difficulties distinguishing matches from non-matches; and determining appropriate window size. Figure 3 is an example of an illogically placed match; Figure 3 contains four graphs: the top graph is a plot of the first profile of interest, the bottom graph the second profile of interest. The highlighted red window of the first profile (first graph) is the area of the profile most similar to the red window in the second profile.
(fourth graph). The second and third profiles are simply expansions of the red windows of the first profile and fourth profile respectively. Figure 3 illustrates the tendency to match straighter line segments, most commonly found along the edge of the profile, because they are the most similar.

Figure 3. Illogical match
Figure 4. Matches selected along the steep edges of the profile

Another complicating factor in this type of comparison is the fact that all profiles generated from SEM profiles are overlapping profiles. This complicates the comparison procedure because the overlapping region is selected by visual estimation, yielding inconsistent overlap regions. This is compounded by the fact that the marks may contain the slightest twist defects and the depth of each mark, and thus the quality of the striations, also vary from mark to mark. These variations require that each match be visually inspected to make sure the match is a logical choice.

The algorithm relies on a validation step to ensure matches are appropriately placed. This is done with two windows known as the correlation and validation windows. The correlation window is what the program designates the most similar area of a user specified size; the validation window, also user defined, may be a different size than the correlation window.
Validation windows are utilized to increase quality of the match area. The user first defines the size of the correlation window, then the size of the validation window, and finally the number of pairs of validation windows to be used. For the two toolmarks being compared, to implement the validation step, a random number of pixels are chosen such that the windows of the specified validation window size which are an equal distance from the correlation window are compared for the two datasets by calculating the Pearson’s correlation coefficient of these windows. This procedure of equal distance comparisons is repeated the number of times as specified by the number of pairs of validation. If the two tools come from a match, the correlation would be expected to be relatively large. If the two tools do not match, it would be expected that the correlation coefficient would tend to not be better than arbitrarily selecting two differing distances from the correlation windows from the various toolmarks and calculating the correlation coefficient from these windows. To allow for this comparison, a procedure is performed in the validation step where arbitrarily selected differing distances for the two toolsets are compared and the Pearson’s correlation coefficient is calculated. This is also repeated the number of times as specified by the number of pairs of validation. Special care is taken to ensure that all chosen validation windows do not overlap the correlation window that was determined in the first step. Figure 5 shows the correlation window in red and a validation window in blue. Many validation windows would be checked before it is assigned a ranking comparing it to other matches. This is especially important as a quality check of the distributions between known matches and non-matches.
Next, all correlation coefficients are ranked and the ranks obtained from equal distance steps from the correlation window are summed to give the statistic \( W = \sum_{j=1}^{NumPairs} S_j \), where \( S_j \) are the ranks mentioned above. The test statistic, \( T \), then takes the form

\[
T = \frac{W - \{n(n+n+1)/2\}}{\{nn(n+n+1)/12\}^{1/2}},
\]

which is the Wilcoxon two-sample rank sum statistic.

Although the \( T \)-statistic won’t have asymptotic normality properties due to independence violations, one expects matching tools to have a distribution centered at a larger value than that of non-matching tools. Further research is being done to investigate such distributional properties.

An example of a comparison is shown in Figure 6. The relative separation of the data is indication of the comparative quality of the data. If a series of data is gathered and compared to known non-matches as well as suspect matches or known matches, relative separation of the two plots can be expected. The plot at left shows no separation. In fact, the window where most of the correlation values fall for non-matches is within the window of correlation for matches. This plot would indicate that the believed match
was of similar quality to known non-matches and could be reevaluated accordingly. Conversely, the plot at right shows a clear separation between the two box plots. There is no overlapping in either the boxes or outliers, indicated by the vertical lines capped with horizontal lines. In this case the conclusion could be drawn that the quality of the suspected matches is far superior to known non-matches and is far more likely to be a match than the plot at left. Before a positive match could be determined however, a study of similar plots would be required to determine what parameters yielded the highest quality non-match.

**Figure 6. Poor separation (left), good separation (right)**

In the present study it is important to keep in mind when using an automated system the facts presented concerning the nature of toolmarks, specifically, the presence of class and subclass characteristics. These features can and do exist, and could possibly fool an automated matching system and produce false positives unless the system is
designed in such a way as to either avoid or overlook such characteristic markings.

Considering the scale of the observations used in this study, class and subclass characteristics should have little bearing on the measurements made using the profilometer and the SEM for analysis by the matching algorithm. These methods consider large regions of the tool at high resolution, thus subclass characteristics should not influence the quality of results. Similarly, class characteristics are so coarse in size that they are not considered.
CHAPTER 3. METHODS AND PROCEDURES

Fifty identical screwdrivers produced by a single manufacturer, Omega Company, were obtained. The screwdrivers were certified by Omega to have been produced sequentially although the exact order was not known. Specifics of the manufacturing were not provided by the manufacturer. Mr. Jim Kreiser, former head toolmark examiner for the State of Illinois, believes this set of tools was formed from a hexagonal rod; turned on a lathe to the appropriate shape and size; and cut off, likely with a screw machine. Examination with a stereoscope reveals circular cutting marks on the end of the tip, along the shank of the tip, and in the notches cut into the hexagonal shaft.

The sides of the tip were likely formed by grinding. The edges of the tip that create striations and are the features of interest for the SEM study are at the intersection of the cut circular surface and the ground surface. The sequential production of the screwdrivers should ensure that they are as practically identical as is possible. Though the population is limited, repetition in the comparison of these tools will provide an
adequate sample base, enough to form a small database that could be used to generate blind test matching.

3.1 Scanning Electron Microscopy
The screwdrivers were imaged with a JEOL 6060 LV SEM to obtain an initial data set. A series of images were gathered over each edge of a screwdriver tip to guarantee the entire surface was imaged; the number of images was dependent upon the magnification. Figure 8 shows overlapping images; each colored box represents an imaged section. Preliminary SEM images were gathered at 30º, 60 º, and 85 º angles. The 30º angle is preferred because it is difficult to tilt from vertical to high angles (60º and 85º) with the tilt limitations of this stage. A sample holder that positions the tool at an angle before it is tilted on the stage brought the tip of the tool into dangerous proximity of the detector. For these reasons, remaining samples were imaged around a 30º tilt.

![Figure 8. Overlapping images of the tool](image)

After imaging each tool, Mr. Kreiser used the screwdrivers to mark lead samples at angles of 30º, 60º, and 85º using a jig to maintain the selected angle for the toolmark, as illustrated in Figure 9 and Figure 10. Both sides of the screwdriver were used. Four replicates were made of each toolmark. After the marks were made, the toolmarks themselves were imaged with an SEM, as described above.
3.2 Profilometry

A Hommelwerk Surface Stylus Profilometer was used to measure surface roughness on all toolmarks and selected tools. The advantage of a stylus profilometer is that it mechanically measures the surface, eliminating possible distortions generated from reflected light. The disadvantage is that the surface is affected by the passage of the stylus, albeit only slightly. Scans on each toolmark were performed in a region where the mark was found to be most complete by visual examination. Each scan consisted of ten separate traces run perpendicular to the toolmark striations, illustrated in Figure 11 by the red lines on the right trace in the image. Each trace sampled 9600 points along a line approximately 7 mm in length. The vertical resolution of this device is 0.005 microns. Scans on the selected tools were obtained by mounting the tool in a holder at a fixed angle of 30 degrees, to simulate the angle at which the corresponding toolmarks were produced.
3.3 Stereoimaging and reconstruction

A further series of images was collected in the SEM to obtain stereo images for three dimensional analysis using software obtained from Alicona, Inc. MeX software is designed to allow for image reconstruction and analysis. MeX is able to reconstruct the surface topography of an observed object in three dimensions based on stereoscopic images that are captured via simple tilting of the stage. MeX was particularly appealing to this application because it is a non-tactile technique. By merging images that have been tilted, typically at angles between 3° and 7°, MeX is able to create a three dimensional representation upon which a profile, area, or volume analysis can be performed. An additional benefit to this reconstruction method is the ease of identification between regions appearing in the profile and their respective area in the image. The profile is user defined with a line drawn on the image. This path can follow a straight line or a contour of interest on the surface. The manufacturers of the software state that the profile can be analyzed in various ways such as direct depth measurements,
measurements of overall average roughness, and waviness determination according to standards now in place such as EN ISO 4287.⁵²

Stereoscopic reconstruction relies on parallax, the apparent position of an object when two views of it are obtained from varying distance from the viewer.⁵³ This works in human eyes by the nature of their spacing which allows for two viewpoints and enables the brain to create depth perception. A two dimensional image results in the loss of information, whereas tilting and reconstruction of two separate images allows the use of parallax to measure depth. Compiling stereopairs from images which are too similar (created from too small a tilt angle or too smooth of a surface) will not create a discernable height difference.

To calculate the coordinates of a three dimensional system, X, Y and Z, knowledge of x and y in both images comprising the stereopair are necessary. Accurate measurement of the magnification of the images, as well as the tilt, between the two images is necessary to calculate coordinates in the three dimensional system. L and R will designate left and right images. Using the calculations below, the parallax will be positive.⁵⁴

\[ Z = \frac{P}{2M \sin(\alpha / 2)} \]
\[ X = \frac{x_L - (P / 2)}{M} = \frac{x_R + (P / 2)}{M} \]
\[ Y = \frac{y_L}{M} = \frac{y_R}{M} \]
\[ P = x_L - x_R \]
\[ \alpha = \Theta_2 - \Theta_1 \]

The software MeX and subsequent routine to MeX termed AutoCalibration rely on mathematical procedures similar to these to create stereopairs.
When obtaining stereopairs from SEM images, a feature at the center of the image is selected around which the stage is tilted. Images used for reconstruction in MeX must meet the following criteria: good signal to noise ratio, Figure 12; sufficient structure in the images, Figure 13; quality depth of focus throughout the image, Figure 14; well exposed image areas; appropriate ratio of height vs. field of view; and geometrically correct tilting.

![Good SNR vs. Poor SNR](image)

Figure 12. Signal-noise ratio
For reconstruction MeX requires information about conditions while the image was captured. Pixel size is needed for both the x and y directions. The working distance needs to be the same for all images contributing to a reconstruction. Alicona recommends
a working distance of around 10 mm. Images can be captured over a vertical or horizontal tilt axis. For reconstruction, images must be rotated so that they are tilted about the y axis, see Figure 15.

![Image of tilted image setup](image)

**Figure 15. Tilting around the y axis**

Finally, the tilt angle (angle of the left image minus angle of the right image) is necessary for input. While small deviations can be tolerated and compensated for, large deviations present too great of a parallax for accurate measurement. A specific tolerance is not given. Disparity (the apparent change in location as related to the reference point) is the method used to calculate depth within the image. Depth resolution is best when it is limited to three times the field of view. The minimum ratio of height to diagonal is 1:70. Disparity can be increased by increasing magnification and/or by increasing tilt angle.

Figure 16 shows two merged images; the left image is blue, the right image is red, the disparity is highlighted by the white arrows. It is necessary to adjust magnification and tilt so that height changes are distinguishable.
Figure 16. Disparity
CHAPTER 4. RESULTS AND DISCUSSION

The project has sought to answer two distinct questions: Can a toolmark be related to a particular tool (and only that tool) on a statistical basis? Can a series of toolmarks be obtained and compared in an automated manner to yield a statistically valid match? These questions have been addressed by first obtaining surface measurements (using a profilometer and SEM) from tools and toolmarks, and then comparing these measurements in a statistical manner. Comparison of the obtained data has been carried out in conjunction with Dr. Max Morris and Mr. Jeremy Craft.

Since the performance of the match algorithm was critical to answering these questions a series of experiments were conducted to test the validity of the algorithm and define the best operational parameters, and a brief summary of these tests was outlined in Chapter Two. Presented below is a more complete discussion of the results of select studies run to determine if the algorithm could statistically support known facts as well as assertions experienced examiners routinely hold to be true due to their experiential knowledge.

Comparisons of toolmarks obtained from the 50 screwdriver tips involved in this study were made in a number of ways. For example, four replicates were made for each side of every screwdriver; therefore, comparisons between profilometer data obtained from replicas could easily be made where a high match correlation should exist. As a corollary to this, a large database of toolmarks was created so that a particular toolmark can be compared to marks made using other tools. In this case a low correlation should exist. Comparisons could also be made between both sides of any given tool, between marks made using a given tool at various angles, etc. If the assertions held by toolmark
examiners are true and the match algorithm works in an acceptable manner, then comparisons made from the same tool at the same angle should show high correlations. Every other comparison should have low correlation values.

Since the algorithm only requires a data file consisting of height measurement versus distance, files obtained either from the profilometer or the SEM using stereoimaging could be compared and analyzed. The routine could be utilized to evaluate files and obtain comparisons in three different ways: profilometer versus profilometer, profilometer versus SEM stereo, and SEM stereo versus SEM stereo.

4.1 Matching of profilometer data from toolmarks

Profilometer profiles of the toolmarks created the most complete data set of any measuring device utilized; therefore, extensive testing of the algorithm was conducted using this data. Known matches and non-matches of the profilometer profiles were compared to examine similarities between marks and test the software on a large dataset. Known matches are those comparisons where it was known that the same tool was used to create a series of toolmarks. For example, tool 43 side A was used to make 4 marks at 30°. Comparisons of these marks to each other should yield known matches. Comparison of a mark made by tool 43 side A to a mark made using side B, or any mark made by any other tool regardless of which side was used, should result in a known non-match. Results obtained from tests of this nature—based on a correlation window of 750, a validation window of 300, and 100 pairs—are shown in Figure 17; the left side shows a known match while the right shows a non-match. Visual inspection of the large scale plots show that the overall shape of the measured surface is similar for the known match
and vastly different for the non-match, even though a “match” is made. This example shows that one must bear in mind that the program will always find a match, no matter how poor that match may be, just as a web search always turns up a match regardless of its quality. What is important is the comparison of the T-statistic, which will indicate the quality of the match. Figure 18 compares known matches from a mark made from tool 43 side B to non-matches from tool 43 side A versus side B. The opposite sides, treated as non-matches, center around the 0 T-statistic, while matches average 10; the clear separation of the two box plots indicates that statistically the known matches are distinctly different from the non-matches.

Figure 17. Comparisons of profilometer mark match (left) and non-match (right)
Comparisons of this type conducted in conjunction with Mr. Craft determined the minimum window sizes necessary to distinguish matches from non-matches. It was found that comparisons of known matches exceeded values of known non-matches when the correlation window was larger than 100, the validation window was larger than 100, and number of pairs was also larger than 100. Similar studies would need to be performed on different types of comparisons to identify the minimum match and maximum non-match criteria.

4.1 Evaluation of marks at varying angles

Experienced examiners routinely make a number of marks at varying angles using both sides of a screwdriver when they are conducting a toolmark investigation. This is done because, experientially, two assertions have been found to be true:
Assertion #1. Toolmarks must be made at nearly the same angle in order for a match to be made. Toolmarks whose angles vary by more than 10-15 degrees cannot be matched.

Assertion #2. For a screwdriver tip, the opposing sides of the screwdriver result in significantly different marks. Comparisons between opposite sides of a screwdriver show no higher likelihood of matching than marks from two totally different screwdrivers.

If the above assertions could be supported by the results of the algorithm, then for the first time statistical validation of these assertions would be available. A successful result from this study would also mean that it is reasonable to assume that further use of the algorithm to match a tool to a toolmark should show the same degree of validity as matching one toolmark to another.

Figure 19 through Figure 22 are based on comparisons made between 44 tools with a correlation window of 100 pixels, a validation window of 500 pixels, and 100 pairs. Figure 19 shows the high ranking of comparisons performed at the same angles, 30-30, 60-60, 85-85, versus those performed at differing angles, 30-60, 30-85, 60-85 of known matches. Known matches means that all the replicates made by a tool at the three angles were compared to each other—e.g. for tool 1, side A, comparisons were made between all angles; for tool 2, side A, comparisons were made between all angles; etc. The results of Figure 19 clearly show a high correlation in T-statistics when toolmarks are compared at the same angle; a low correlation corresponds to marks made using the
same tool but at angles differing by more than 10°. These results are especially informative when compared to those of Figure 20, where marks were compared from varying tools. In this case no matches should be possible; low T-statistics with values roughly symmetric around zero are observed. The observation that the known matches compared at angles greater than 10° exhibit T-statistic values similar to those seen at all angles for known non-matches clearly supports Assertion #1 (above) held by toolmark examiners. Especially interesting are the results of Figure 21, which show comparisons of side A of any tool to side B of the same tool. The T-statistic values again are equivalent to values for known non-matches, indicating enough difference exists between the two sides of a single tool to make them analogous to non-matches. This observation supports Assertion #2, again in agreement with examiner observations. All of the comparison data for the angular study are summarized in Figure 22, which illustrates the range of T-statistics covered in the 95% confidence intervals of matches, non-matches, and different sides. A clear separation exists between a match and a non-match, indicating that the algorithm should be suitable for the next phase of the project: to establish a definite link between a tool and the resultant mark using a statistical analysis.
4.2 Linking a tool to a mark

With the baseline was established by the profilometry data, work progressed to establishing a link between the working surface of the tool and toolmark.
4.2.1 SEM stereopair analysis

The first stereopairs were created from images gathered on a manually controlled stage plus or minus 3° around a 30° angle; the first sets of stereopairs were gathered at a magnification of 75x. Because the stage was not motorized, the accuracy of the tilt was within one degree. A typical output of the program is shown in Figure 23 which shows results obtained using this routine. The profile drawn on the image surface and the corresponding profile plotted below the image. Figure 24 shows how multiple profiles are gathered from a single stereopair. Adjacent parallel lines are expected to have similar profiles.

Figure 23. Profile drawn on image surface
These initial stereopairs produced unsatisfactory profiles due to a large amount of noise. In Figure 25, a SEM profile is compared to the profile generated by the profilometer, illustrating this problem.

A typical large scale comparison between the profilometer and SEM profiles is shown in Figure 26. The surface profiles are shown on the left, while the tool and the
resultant toolmark are shown imaged on the right. Note the length of the profilometer scan is much longer than the SEM stereo image, 5000 pixels to 500 pixels. This is due to the limited region that can be imaged in the SEM at one time.

Figure 26. Diagram showing sources of individual profile

Figure 27 shows three comparisons between SEM profiles and profilometer data. An extra profile is included (second from the top) that scales the profile to the same length as the SEM profile for better visual inspection. The blue box shows the expected region of match and the red the actual region of match. The low quality of data is evident; best matches are found in areas known to be outside the region of potential match in all cases. The far left comparison shows a match inside the blue box, but the subsequent location inside that box is incorrect.
Alicona was contacted for suggestions to improve the results. They immediately suggested a eucentric stage with motorized tilt and the use of higher magnifications for acquiring an image. A eucentric stage was not available but a stage with motorized tilt was used to produce the second set of stereopairs. A variety of magnifications (100x, 250x, 500x, 750x, 1000x) were tested but profiles still contained excessive amounts of noise. Figure 28 shows adjacent traces of the same reconstructed image at three increasing magnifications, 500x, 750x, and 1000x. In all cases it is obvious that the profiles are significantly different, even though the data was obtained along nearly identical profilometer traces, indicating that inaccuracies persist. It is unknown whether the inaccuracies are due to the lack of a eucentric stage or the magnification still being too low. Even at relatively high magnifications (e.g. 1000x) the data is still poor. At magnifications higher than 1000x features considered to be responsible for creating
toolmark striations begins to be obscured by smaller scale features unrelated to the surface profile of the toolmark.

![Graphs showing comparison of SEM profiles at high 500x, 750x, 1000x respectively.]

Figure 28. Comparison of SEM profiles at high 500x, 750x, 1000x respectively

Alicona was contacted again; they responded that the low signal to noise ratio, the smooth surface, and the height change to image dimension ratio were preventing quality reconstructions. They recommended their software AutoCalibration in place of StereoCreator as a solution.63

AutoCalibration is a newer software package that utilizes three images for more accurate three dimensional profile reconstructions. It was hoped that by using this software the problem of a non-eucentric stage could be overcome. Initial examination of the new plots appeared encouraging in that the regions visually appeared to correspond more closely to the profilometer scans. The noise was still higher in the stereoimage than the profilometer data obtained from the same region. The reconstruction of the three
dimensional information from stereoimages was expected to contain a higher noise level than direct measurement of the surface.

Several tests were run to determine whether the data obtained using AutoCalibration was reliable. Unfortunately, results from these comparisons were no better than those obtained from the previous stereopairs. As shown in Figure 29, matches are still found outside of the expected regions; red boxes identify regions of match while the blue boxes indicate where matches should occur.

**Figure 29. SEM results using AutoCalibration and a digitally controlled stage**

Another test involved drawing three adjacent profiles on a single stereoimage—the profiles were drawn as closely as possible. These profiles should be very similar due to their close proximity to one another. The subsequent matches to the profilometer profile obtained from a toolmark made by this tool are shown in Figure 30. The stereoimage profiles, shown at the bottom of the three match comparison boxes, are visually similar to each other as one would expect. However, the matches made to the
profilometer profile are wildly inaccurate. The red box highlights the program’s selected region of match while the blue box shows the expected region of match. All three of these sections resulted in a match region located outside the region the stereoimage originated from. These results were consistent for all comparisons generated between SEM stereoimages and profilometer data. A comparison of the T-statistic correlation values for profilometer vs. SEM stereoimages is shown in Figure 31. The lack of separation and the average value near zero indicates the matches are no better in quality than known non-matches. Although a variety of window sizes were tried, all led to the same conclusion.

**Figure 30. Comparisons of adjacent profiles of SEM tip 39b to profilometer 39b**
Figure 31. Comparison of correlation values in profilometer v. SEM matches

At this point it was apparent that two possibilities existed. Either, 1) there was no relationship between a tool and the resultant toolmark. The long history of comparative matching between tools and toolmarks—as well as the results of this study—showed toolmarks can be conclusively identified using the developed algorithm. One could argue that the likelihood of the first possibility being true is essentially zero. Or 2) the software was still inadequately measuring the surface roughness of the tool. To test whether user error was the source of the observed problems, a sample was sent to Alicona’s test lab where another attempt to obtain suitable data from the tool tip was made. Alicona was also unable to produce a reasonable profile using their SEM, responding that they could not achieve correlation with the SEM images due to limitation of stereoscopic analysis.64 (N.B. While this is Alicona’s opinion, theoretically stereoscopic analysis should be successful. It seems more likely that the problem is a limitation with the way their software implements stereoscopic analysis.)

Analysis of the problem suggests that creating stereopairs from the SEM images was unsuccessful for a variety of reasons. Firstly, the small field of view with respect to
the vertical resolution—the difference between horizontal and vertical was orders of magnitude different (millimeters long versus micrometers high)—was outside the range that could be easily accommodated by the technique. Alicona officials explained that the large difference between horizontal and vertical dimensions was problematic. The result: too small of a height change to perceive a height difference and a lack of information about the surface. In other words, by the time the magnification was high enough to achieve good resolution, it was also so high that it was difficult to discern surface characteristics. Secondly, stitching images to obtain a large image from which to work was impractical. While stitching images produced a large image, it was impractical to use this composite for stereoscopic analysis because it led to multiple image centers around which the images were tilted. This resulted in an image that could not be reconstructed. One possible solution to this problem could be to create overlapping three dimensional reconstructions. However, overlaying consecutive reconstructions to create a complete profile was found to be impossible in this case because of the noise contained in the profiles. It was never clear where overlapping sections should begin and end.

As a final attempt to obtain results using the SEM and MeX, a notch was cut with a dremmel in the center of the region of interest of tool 1. The notch in the center provided a region of known match near the center of the tool tip. Toolmarks were then made with the notched tool tip by Mr. Kreiser for examination. This allowed areas on each side of the notch to be easily identified, and very specific images were obtained at high magnifications in this area. Unfortunately, the profilometer was undergoing repairs and unavailable for use to analyze the toolmark. Instead, the toolmark was examined
with the SEM and a profile related to the same area of the tip was generated. In other words, MeX was used to create a three-dimensional reconstruction of both the tip and the resultant mark, and these data files were compared. Figure 32 shows the results of parallel sections from SEM profiles of the mark and the tip. It is clear that the results are consistent with previous results that indicate inconsistencies with SEM profiles generated from stereopairs.

Figure 32. Comparisons between same sections of toolmark (top graph) and the tool (bottom graph)

4.2.2 Other methods

Given the failure of the SEM imaging to establish any link between a tool and the resultant tool tip, various other techniques were used to measure the tool surface for the comparison of the profilometer scan of the mark. The results of those experiments are detailed in the following sections.
4.2.2.1 Profilometery on a tool tip

Given the quality of the profilometer data, attempts were made to run profilometer scans on the very edge of the tool for comparison to the resultant toolmark. For these attempts, the tool was mounted vertically and three parallel profilometer traces were run as closely to the edge as possible. Three profiles were obtained of both side A and B of tool 35 before the profilometer experienced problems. Comparisons were made using the match algorithm; a command was added to re-orient comparisons between profilometer profiles of toolmarks and tips. The results of these experiments are shown in Figure 33; the left image shows a match between tool tip and toolmark 35, and the right image a non-match between tool 35 and toolmark 39. Figure 34 shows a comparison of the T-statistics of 29 comparisons for both matches and non-matches. The separation of the two box plots indicates that the matches are indeed statistically better than the non-matches, although more separation is desired. A larger data set should be examined to improve the quality of these comparisons and confirm this conclusion.
Figure 33. A match (left) and a non-match (right) comparing profilometry of the tool tip and the toolmark

Figure 34. Comparison of matches and non-matches between profilometry of tip and mark

4.2.2.2 Confocal florescence and x-ray tomography

An attempt was also made to resolve surface features on the tool tip using Confocal Florescence. Confocal Florescence depends on a laser source directed onto a sample surface. The laser stimulates the release of longer wavelengths which are collected by a detector. Bob Doyle, Associate Scientist in Genetics, Development and Cell Biology at Iowa State University, was contacted and volunteered to assist with this attempt. An example of the image obtained is shown in Figure 35. Unfortunately the
attempt was not able to resolve surface features as the image collected did not have enough data to achieve the desired resolution upon reconstruction.

Figure 35. Imaged edge of tool in confocal florescence system

X-ray tomography was also considered. X-ray tomography utilizes many x-rays profiles collected in stacks (vertically adjacent x-rays) to create a digital three dimensional reconstruction. Joe Gray, of the Center for Nondestructive Evaluation at Iowa State University, examined a tool as a candidate for x-ray tomography and determined that x-ray tomography was not a good option to gather a sample profile because the surface features that create striations are too small.

4.2.2.3 Laser profilometry

Given the success of the stylus profilometer results, an attempt was made to use a laser profilometer, located at the Center for Nondestructive Evaluation at Iowa State University, to measure the sample surface. This technique focuses a laser beam onto a sample surface and measures the length to the surface. The laser moves back and forth across a small distance in the x direction while simultaneously moving along the y axis and measuring the distance to the surface, z. This technique has an advantage over stylus profilometry in that nothing is physically in contact with the sample during
measurement. Because the area of interest was along the sample edge, the laser had some trouble remaining focused along the entire length of the tool. Since large areas of the sample did not stay in focus the profile created was quite noisy, as is seen in Figure 36. Mr. Craft examined the resulting profile to see if a match could be made using the match algorithm by mathematically removing the noise from the signal. However, the conclusion was reached that after removing all the static, a large amount of the profile would be lost and the profile would not be of sufficient quality to achieve a high caliber match with the current validation software.

Figure 36. Surface measurements with the laser

The data from the regions where the laser was in focus visually appear quite good; it seems reasonable that laser profilometry might be a suitable candidate for measurements of this type if the focus problem can be resolved.

4.2.2.4 Dye penetrant

An attempt to measure the surface of the sample toolmark using dye penetrant was also made. This method involves pouring a gelatinous dye solution into surface
crevices and measuring fluorescence of the dye with a microscope designed for this purpose. This again was appealing because although there was contact with the surface the gelatinous materials could be easily removed without damaging the sample. Jerry Sedgewick of the University of Minnesota assisted with this type of test. Initial results indicate that this test is also unsuitable for determining the surface roughness of the tool. Two different types of dyes were employed; neither dye was able to sufficiently penetrate into the striations of the toolmark. While the width of the toolmark could be determined the details of the striations considered to be characteristic of the tool were not discernable.

4.2.2.5 InfiniteFocus

As a final attempt, Alicona suggested their InfiniteFocus (IF) system as an alternative measurement system. InfiniteFocus depends on focus variation to create accurate depth of focus measurements for three dimensional reconstructions. White light is passed through a beam splitter to a series of objectives. A three dimensional reconstruction is created when multiple focal planes are superimposed. The varying Z distance is important to the reconstruction to bring all areas of the surface into focus.68

Alicona already possessed the sample that had been sent for their attempt using SEM imaging making it relatively easy for them to use the same samples to obtain data with the optical system. The results from this attempt are shown in Figure 37. Three curves are provided; measurement from tool sides A and B (green and blue respectively) and a measurement from the toolmark produced from side B (red). Note that the profiles have been flipped to correspond to the impression of the toolmark. Visually, the
similarities of the fine features between the tool and the mark (blue and red) are readily apparent. The differences between the corresponding sides of the tool and the plate (toolmark) profile probably arise due to specifics concerning how the mark was made and the variations in angle and pressure, which are expected to occur even using a jig to minimize variation. Equally apparent are the differences between green profile (side A) and the blue profile (side B), again illustrating the significant difference between two sides of the same tool.

![Comparison of primary profiles - Tool angle 60°](image)

**Figure 37. InfiniteFocus comparison of primary profiles of tool 38 and toolmark 38**

The match algorithm was used to test the validity of the visual impressions. A range of window sizes was employed for the correlation and validation portions of the algorithm since only one match and one non-match comparison is available for examination. In all cases the T-statistic correlations found for the match set are significantly higher than for the non-match set, as shown in Table I. The highest match ranking, a T-statistic of 11.58111, was achieved using a correlation window of 800 pixels, a validation of window 400 pixels and 90 matches, Figure 38.
Figure 38. Validation output from IFM profiles

Table 1. T-statistic comparison of IFM profiles

<table>
<thead>
<tr>
<th>Correlation Window</th>
<th>Validation Window</th>
<th>Number Pairs</th>
<th>ToolSideBvPlateSideB (T-statistic)</th>
<th>ToolSideAvPlateSideB (T-statistic)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>150.75, 90</td>
<td></td>
<td>8.31389</td>
<td>-2.60347</td>
<td>10.91736</td>
<td></td>
</tr>
<tr>
<td>500, 250, 90</td>
<td></td>
<td>11.72</td>
<td>1.23879</td>
<td>9.93321</td>
<td></td>
</tr>
<tr>
<td>600, 300, 90</td>
<td></td>
<td>11.03465</td>
<td>4.47916</td>
<td>6.55549</td>
<td></td>
</tr>
<tr>
<td>750, 375, 90</td>
<td></td>
<td>11.37513</td>
<td>4.24278</td>
<td>7.13235</td>
<td></td>
</tr>
<tr>
<td>800, 400, 90</td>
<td></td>
<td>11.58111</td>
<td>3.04119</td>
<td>8.53992</td>
<td></td>
</tr>
<tr>
<td>1000, 500, 90</td>
<td></td>
<td>11.40088</td>
<td>4.09402</td>
<td>7.30686</td>
<td></td>
</tr>
</tbody>
</table>

These initial results using the IF optical system indicate that this method holds a high degree of promise for matching of tool to toolmarks. While the T-statistics were high for all correlation and validation windows, most important is the separation between known matches and non-matches. The separation decreases a small amount as window size increases, suggesting that the smaller window may be best for analysis on this type
of profile. The high T-statistics that describe the match suggest matching is good over a large portion of the surface rather than simply over a small constrained area. This indicates that if conditions could be duplicated exactly an almost one-to-one relationship with a high degree of statistical certainty could be stated to exist between features on the tool and the resultant toolmark. To confirm the accuracy of these initial results, a larger series of tests needs to be performed under more tightly controlled conditions.
CHAPTER 5. SUMMARY AND CONCLUSIONS

Utilizing Alicona’s stereoimages for this application has proven unsuccessful. The very smooth surface is difficult to detect variations upon without increasing magnitude beyond the point of being able to see multiple individual characteristics. Even high magnification images contained a significant amount of noise preventing successful matching of similar locations between two profiles of the tip or the tip and the mark. The low ratio of height change to image size is another obstacle to using this method. If stereoimages are ever to work, the amount of noise in the profile has to be reduced and the software would have to be able to detect small surface changes over large areas. The conclusion, confirmed by Alicona, is that at this time it is impossible to create quality stereoimages, from SEM micrographs, with their software for this application. This conclusion should not detract from the concept of stereoimages and is specific to stereoimages reconstructed in this manner.

IF shows the most promise as a direct, non-tactile comparison method. It does not suffer from the problems associated with stereoimages using SEM because it reconstructs the three dimensional surface in a different manner. It is able to detect small changes over a long surface without losing focus. Laser profilometry could produce similar results if the technique was refined to eliminate the lack of data in some parts of the profile.

Mr. Kreiser believes that an experienced examiner has the ability to make a conclusion about whether a match is present based on a match length of 0.5 mm or less, depending on the condition of the mark. The algorithm can distinguish successfully
between matches and non-matches of profilometer profiles based on a minimum correlation window of 100 pixels or 0.073 mm. Validation windows support this correlation window checking at least 100 other windows of the same size in relative positions. In each case, window sizes will need to be validated by a comprehensive study to understand the maximum values of known non-matches before reaching a conclusion of minimum match criteria. If studies are performed with regard to individual applications this software should indeed produce statistically valid results.
ACKNOWLEDGEMENTS

Special thanks to Professors Scott Chumbley, Larry Genalo, and Max Morris for their contributions and guidance. Thanks also go to fellow students David Faden, Jeremy Craft, Stephen Davis, Charles Fisher, and Laura Barker for their contributions to this research. A final thank you to my family and friends for their input and support.


Kreiser, Jim. Email to the Author. 20 January 2007.

Besser, Matt. "Re: 0.5 Microns?" Email to the author. 16 August 2006.


“Requirements on Images for Good Results with MeX.” Advertising Brochure. Alicona Imaging.

“Requirements on Images for Good Results with MeX.” Advertising Brochure. Alicona Imaging.

“Requirements on Images for Good Results with MeX.” Advertising Brochure. Alicona Imaging.

“Requirements on Images for Good Results with MeX.” Advertising Brochure. Alicona Imaging.


“Requirements on Images for Good Results with MeX.” Advertising Brochure. Alicona Imaging.


Walter, Gernot. "Re: more problems with MeX." Email to the author. 22 September 2006.

Scherer, Stefan. "Re: screwdriver tip for Mex imaging." Email to the author. 19 December 2006.

Scherer, Stefan. Alicona "Re: solution?" Email to the author. 10 January 2007.

Scherer, Stefan. Alicona "Re: solution?" Email to the author. 10 January 2007.


Kreiser, Jim. Email to the Author. 2 Feb 2007.