PREDICTIVE PROCESS MODELING OF TOOL WEAR IN FRICTION STIR WELDING OF METAL MATRIX COMPOSITES

By

Tracie Joy Prater

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Mechanical Engineering

August, 2012

Nashville, Tennessee

Approved:

Professor Alvin M. Strauss

Professor George E. Cook

Professor Jimmy Davidson

Professor Robert Pitz

Professor Jason Valentine

To my parents, Teresa and Benjamin Prater,

and

To my grandmother, Joyce Jones.

ACKNOWLEDGEMENTS

This work was made possible by financial support from NASA Space Grant and the NASA Graduate Student Research Program (GSRP). I am especially grateful to Drs. Alvin Strauss and George Cook of Vanderbilt University as well as Dr. Art Nunes of NASA Marshall Space Flight Center (MSFC), who directed this research. I am indebted to Dr. Jim Davidson, also of Vanderbilt, for his insight and assistance with diamond coatings. Special thanks to Dr. Frank Six of NASA MSFC for his support of my academic endeavors.

I am grateful to all those I have worked with during my tenure in the Vanderbilt University Welding Automation Lab: Brian Gibson, Chase Cox, Jorrie Ballun, Kathy Dharmaraj, David Lammlein, Thomas Bloodworth, and Paul Fleming. Much thanks and love to my parents, Teresa and Benjamin Prater, and my grandmother, Joyce Jones, to whom this dissertation is dedicated. Additional gratitude to two very special dogs: Gustavus Adolphus von Weinerschnitzel and Daisy.

PREFACE

This dissertation presents several separate but related studies on the wear process in FSW of metal matrix composites (MMCs): modeling the variation of wear with process variables, developing a phenomenological model to describe the wear process, assessing the impact of material properties of both the tool and workpiece material on wear, and exploring the use of process signals to sense wear in-process. Together these original studies represent a significant contribution to the literature on materials joining, composite materials, and tribology.

The work presented in the dissertation is organized into ten chapters. Chapter I provides an overview of the FSW process, its advantages, and current industrial application. This broad discussion includes information on common FSW tool materials and geometries, joint configurations, workpiece materials, and process variables. The features of FSW joints (including defects and flaws) as well as techniques (both destructive and non-destructive) used to evaluate the quality of the finished joint are also discussed in this chapter.

Chapter II introduces metal matrix composites (MMCs). The first portion of the chapter focuses on the structure of these materials, their advantages, and the classification system used to succinctly identify their composition. The remainder of the chapter presents the problems encountered when joining MMCs to themselves or other materials using fusion techniques. MMC joints produced using FSW exhibit characteristics superior to their fusion-welded counterparts. Representative data for hardness, tensile strength, fatigue life, and ductility of FSW joints (taken from published parameterization

studies on FSW of MMCs) is presented. Wear experienced by the FSW tool when joining MMCs is of particular interest. While there are a few existing studies which examine the evolution of the FSW tool shape with wear, little to no work has been done to predict, combat, and sense wear in this process. These unstudied aspects of wear phenomena in FSW of MMCs are the principal focus of this dissertation.

Any study of wear usually begins with an identification of the wear mechanism. The bulk of chapter III profiles three of the major wear mechanisms commonly encountered in manufacturing scenarios: adhesion, abrasion, and surface fatigue. In addition to outlining the characteristics of worn surfaces produced by each mechanism, the discussion emphasizes the point that multiple wear mechanisms can occur simultaneously in a single system. Tribology and microscopy techniques are essential to the clear identification of the wear mechanism(s) which act on a surface. This chapter also assesses the methods commonly employed to quantify wear, including mechanical gauging, optical comparison, and radiotracers. Careful attention is paid to those techniques which are applicable to the measurement of tool wear in FSW of MMCs.

Chapter IV describes the majority of the research methods which are used to design and analyze the statistical models of FSW tool wear presented in subsequent chapters. Since FSW includes multiple process variables, experimental design is essential to maximize experimental efficiency. The first portion of the chapter enumerates the advantages of experimental design and explains several designs typically used to construct plans of experiments for manufacturing and quality control applications: factorial design, Taguchi method, and Response Surface Methodology. Statistical techniques used to analyze the data generated using these experimental

matrices are also presented in detail. The bulk of this discussion centers on multiple regression modeling (MRM) and performance metrics which can be used to evaluate MRM models (such as tests for significance). The latter half of chapter IV examines the literature on tool wear in machining of MMCs. The purpose of this literature review is twofold, as it 1) critically examines whether the methods outlined earlier in the chapter have been successfully applied to a similar wear problem and 2) can be used to draw parallels between variables which may influence wear in both machining and FSW of MMCs . The work summarized serves as a precursor for the studies of FSW tool wear which appear in chapters V-IX; the experimental techniques and variables selected for these investigations are often based upon those found in the literature on tool wear in machining.

The exploratory study which is the subject of chapter V seeks to quantify the impact of FSW process parameters (rotation speed ω, traverse rate ν, and length of the weld joint ℓ) on the amount of volumetric tool wear incurred in joining MMCs. The experiments for this investigation were designed using the Taguchi method. Volumetric wear was calculated by optically comparing pre and post-weld cross-sectional images of the tool probe in imaging software. Multiple regression was used to construct a predictive process model in the form of a linear equation which estimates tool wear based on the major process variables considered in the study. The model's predictive capability was assessed using standard statistical metrics in addition to cross-validation techniques. The concluding section of this chapter addresses concerns about scalability and expresses wear in terms of the dimensionless parameter $\frac{\omega \ell}{\nu}$.

Chapter VI attempts to explain the dependence of tool wear on process parameters observed in the previous study through physics-based modeling of the wear process. This model of wear in FSW of MMCs is based on the rotating plug model for friction stir welding developed by Dr. Arthur Nunes of NASA's Marshall Space Flight Center. In Nunes's formulation, the probe is surrounded by a plug of plasticized metal (the "rotating plug"). The width of this plug, δ, is determined by the process parameters ω and ν as well as the angular position θ in the x-y plane (the plug is symmetric about the y-axis and varies in thickness from δ=0 at θ = -π/2 to $\delta_{maximum}$ at θ = π/2). Nunes's model can be slightly modified to account for the presence of abrasive particles (such as the inclusions found in Al-MMCs). The model indicates that for a given particle diameter *D*, wear occurs only when the particle is able to span the width of the shear zone and press into the tool surface. This model is consistent with the results of Chapter V – parameters which correspond to thinner shear zones result in greater wear, while parameter sets associated with larger values of δ experience less wear. The rotating plug model leads us to postulate a piecewise criterion for wear: wear only occurs when the radius of the particle exceeds the width of the shear zone ($\frac{D}{2} > \delta$). This corollary to Nunes's work implies that wear is also affected by material properties, such as particle size and percentage reinforcement.

In chapter VII, tracer experiments based on the work of K. Colligan (who used steel shot tracers embedded along the joint line to study material flow in FSW) were used to test the effects of particle size. In each experiment, a tool is used to join Al 6061 plates containing a specific size of SiC particulate (the particles are contained in a milled slot which runs the length of the advancing face of the joint). The wear pattern on the

worn surface of the tool is then analyzed using contact profilometry and SEM microscopy. The former technique is used to track variations in the wear pattern (quantified in terms of the surface roughness) with increasing particle diameter.

Chapter VIII focuses on methods to combat wear. In chapter VII, the wear mechanism at work in FSW of MMCs was diagnosed as abrasion: it follows that the best method to combat wear is simply to use a tool material which is harder than the reinforcement particles. This strategy stems directly from Rabinowicz's principle of abrasive wear, which states that material A can scratch material B only when the hardness of B is less than that of A. This chapter explores the effect of the hardness ratio (defined as the ratio of the hardness of the tool to the hardness of the abrasive) on wear. Based on classical theory, a system for which the hardness ratio exceeds 1 should experience zero wear. The list of materials with hardness values greater than the ceramic reinforcements typically used in Al-MMCs is short. The best candidate for a wear-resistant tool material is diamond. Unfortunately, the inordinate expenses associated with monolithic diamond tools preclude their use in all but the most mission-critical manufacturing applications. Prevailing concerns about cost efficiency lead us to consider diamond coatings as a more economical alternative to tools fabricated from bulk diamond. The major obstacle to the implementation of coated tools is the coating's tendency to delaminate under the stresses imposed by the joining process. When delamination occurs, the (often superior) wear behavior of the coated tool regresses to levels associated with that of the substrate; a strong bond between substrate and coating is thus needed to prevent delamination and preserve the wear-resistant characteristics of the coated tool. The plan of experiments used to test the effect of the hardness ratio on wear

is a factorial design which considers four different tool materials (O1 tool steel, WC micrograin, WC submicrograin, and WC coated with diamond). Previous research has demonstrated that the brittle behavior of ceramic and refractory metals (which are good candidate substrates for diamond coatings) renders them virtually useless for FSW applications. It is generally agreed that coated tools are only cost efficient for rapid wear scenarios in which wear cannot be effectively controlled using other means (i.e. parameter selection).

Chapter IX focuses on research toward developing a method of sensing wear in-process using feedback signals. Work by Dr. Arthur Nunes and Mr. Brian Gibson has demonstrated that radial deterioration of the probe during joining of MMCs by FSW should correspond to a decrease in the magnitude of the torque experienced by the tool. Simulations by Gibson et al. showed that *in situ* estimates of wear could be made using an adaptive torque controller [34]. According to Nunes's equation for torque based on the rotating plug model, the torque signal is sensitive to flow stress, temperature, and plunge depth as well as geometry. These additional (and sometimes coupled) dependencies make it difficult to isolate changes in the torque signal that can be attributed solely to the radial loss of tool material. The experimental work presented in this chapter deals with the collection of process signals and an assessment as to whether they can provide information about the degree of wear. Force and torque signals (axial force F_z and spindle torque M_z) are recorded using a low-end dynamometer. Changes in spindle torque are shown to correlate with wear -- this work thus serves as proof of concept for an in-process wear sensor in FSW of high-temperature alloys. Such a sensor would be particularly useful for FSW of MMCs and steels, processes which are plagued

by rapid and severe tool wear. Currently, all of the techniques used to evaluate wear (such as those used to analyze the tools in chapters V-VII) are performed post-process and require that the tool be taken offline for examination. An in-process wear sensor such as the one proposed in this chapter would exploit changes in the torque signals to transmit real-time information about the degree of tool wear. The end-user can utilize this information to identify when an unacceptable level of material loss has occurred and the tool needs to be replaced. In-process sensing minimizes disruption of the process; the primary application is thus for longer welds of abrasive materials where an *in media res* wear analysis is not possible.

Chapter X considers the sum of the work presented on predictive process modeling and sensing of wear in friction stir welding of metal matrix composites. The discussion synthesizes these analyses and identifies possible directions for future research in this area. The suggested ancillary research based on the dissertation work includes: a formal analysis of temperature effects on wear rate, the design and implementation of an adaptive torque controller to potentially control the amount of wear, and modeling of the wear process using computational fluid dynamics (CFD) software. Throughout chapters V-IX, the results and trends observed for tool wear in FSW of MMCs are compared with those documented in the literature on tool wear in machining. Each of these chapters expands upon the concordance and discrepancies in wear behavior for FSW and machining and offers possible explanations for the differences (or similarities) between wear processes in these applications. The discussion situates the methodology used in the FSW tool wear studies within the broader field of tribology. While the optical imaging, microscopy, profilometry, and weighing methods used to quantify wear in Chapter V-

VIII are standard techniques for analysis of worn surfaces, the particle-tracer experiments represent a novel and inexpensive means to study the effects of reinforcement size and concentration on wear. Although the particle-tracer technique was necessitated by the limited availability and expense of off-the-shelf MMC materials with desired combinations of particle size, inclusion material, and percentage reinforcement, it is generally applicable to the study of wear in other applications (and may be more cost efficient than many current methods). The concluding remarks of this chapter emphasize the contribution of the original work presented in the dissertation to the fields of friction stir welding, metal matrix composites, and wear of materials.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CAD Computer Aided Design

CCD Central Composite Design

CFD Computational Fluid Dynamics

CVD Chemical Vapor Deposition

DOE Design of Experiments

EB Electron Beam

FGM Functionally Gradient Material

FSSW Friction Stir Spot Welding

FSW Friction Stir Welding

GTA Gas Tungsten Arc

HAZ Heat Affected Zone

HSS High Speed Steel

IPM Inches Per Minute

LB Laser Beam

LDA Linear Discriminant Analysis

MMC Metal Matrix Composite

MRM Multiple Regression Modeling

MSFC Marshall Space Flight Center

NDE Non-destructive evaluation

PCD Polycrystalline Diamond

RPM Rotations per minute

RSM Response Surface Methodology

SEM Scanning Electron Microscope

SNR Signal to noise ratio

TMAZ Thermo-Mechanically Affected Zone

TPMC Thermoplastic Matrix Composite

TWI The Welding Institute

UTS Ultimate Tensile Strength

VINSE Vanderbilt Institute of Nanoscience and Engineering

VPPA Variable Polarity Plasma Arc

VUWAL Vanderbilt University Welding Automation Laboratory

Al_2O_3 Aluminum (III) Oxide

B_4C Boron Carbide

CrN Chromium Nitride

SiC Silicon Carbide

WC Tungsten Carbide

$R_{adj}{}^2$ adjusted R-squared value

T_{init} or T_∞ ambient temperature

θ angular position

R_a average surface roughness

F_z axial force

z axial position along probe length

B Biot number

Δm change in mass of insert

δ clearance between rotating plug and tool surface

δ_0 clearance between rotating plug and tool surface at $\theta = 0$

D_{cr} critical particle diameter necessary for abrasion

W_c critical wear value

ΔC cutting arc

Df	degrees of freedom
ρ	density
D	diameter of reinforcement particle
N	dimensionless group
$\frac{\omega \ell}{\nu}$	dimensionless group for wear
R_z	distance between the average height of five highest and lowest peaks
R_p	distance from mean height of the surface to the highest peak
e_i	error
e	errors in column matrix form
Y	experimental outcomes in matrix form
$\boldsymbol{\beta}$	experimental weights in matrix form
F	F-statistic
H_r	hardness of reinforcement
H_t	hardness of tool
H	hardness ratio
x	indentation depth
m_i	initial mass of insert
ℓ	length of weld
T_{max}	maximum tool temperature
MS_M	mean sum of squares of regression model
MS_R	mean sum of squares of residuals
ε	mechanical dissipation
T_m	melting point of matrix
SS_M	model sum of squares
E	modulus of elasticity
y_i	observed value
P	original pixel count of probe cross section

R_t	peak to valley roughness
R	Pearson correlation coefficient
P	Peclet number
P'	pixel count of probe cross section after a specified weld distance
$\widehat{y_i}$	predicted value
r	radius of probe
R	radius of shoulder
R'_i	ratio of R_i in region 3 of tool surface to R_i in region 1
β_i	regression weight
SS_R	residual sum of squares
R_q	root mean square value of surface roughness
ω	rotation rate
Ω	rotation rate
τ_p	shear stress of tool
τ_M	shear stress of workpiece
p	signficance value
c_t	specific heat of tool
c_m	specific heat of workpiece
SA	surface area of probe
t	t statistic
k_t	thermal conductivity of tool
k_m	thermal conductivity of workpiece
α	tool tilt angle
SS_T	total sum of squares
ν	traverse speed
τ or T	torque

v_m	volume fraction of matrix
v_r	volume fraction of reinforcement
W	wear of FSW probe
$\frac{\omega}{v}$	weld pitch
T_w	weld temperature
Y_{sm}	yield strength of material
Y_{st}	yield strength of tool

CHAPTER I

AN INTRODUCTION TO FRICTION STIR WELDING

1.1 The Friction Stir Welding Process

Patented in 1991 by researchers at The Welding Institute of Cambridge, England, Friction Stir Welding (FSW) is a novel solid-state joining process used in applications worldwide [1]. The process, which occurs below the melting temperature of the joint material, represents a departure from traditional fusion welding methods. In conventional FSW, a rotating tool is plunged into the surface of adjoining metal plates. The rotation of the tool generates heat at the interface, resulting in local plasticization of the material due to shear stress. As the tool traverses along the joint line, the material behind the tool consolidates, forming a welded region with a width roughly corresponding to the diameter of the tool in contact with the surface [2]. The process is illustrated in Figure 1. Note that the advancing and retreating sides of the weld are defined relative to the direction of tool rotation. The advancing side is the side of the weld in which the tool rotation is in the same direction as the traverse; thus the retreating side is the side where tool rotation and welding direction are in opposition. During a weld, material is swept from the advancing side and deposited on the retreating side.

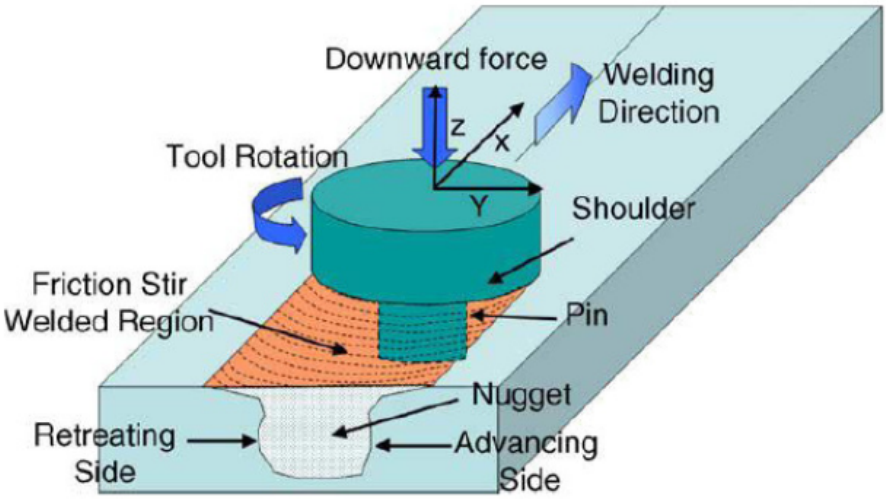


Figure 1 Schematic of FSW process [3]

As a relatively nascent technology, much research has devoted to the effect of process parameters on the quality of the finished joint. There are four major process parameters which can be varied in FSW: rotation speed, traverse speed, tilt angle, and plunge depth. Rotation speed (ω) is in units of rotations per minute (rpm) and designates the rate at which the tool rotates. Traverse speed (ν) indicates the speed at which the tool traverses the material. ν is usually specified in units of inches per minute (ipm). Depending on the system configuration, the tool may remain stationary while the material is advanced at traverse rate ν (or vice versa). The selection of ω and ν, often expressed in terms of the weld pitch ratio $\frac{\omega}{\nu}$, largely governs heat input. Excessive heat input can contribute to the formation of voids in the joint, while insufficient heat input can result in tool fracture. The establishment of an operating window – a set of parameters which produce acceptable welds – is essential for many applications.

The plunge depth, *d*, is defined as the distance the shoulder penetrates below the surface of the workpiece [3]. Plunging the shoulder of the tool helps ensure that the heat generated by the shoulder is adequate to plasticize the workpiece material. The plunge depth required to maintain sufficient shoulder contact is based on the tool tilt angle α (analogous to the travel angle in fusion welding) and the diameter of the tool shoulder. The plunge depth in turn determines the weld ligament, the distance from the backing anvil to the bottom of the tool pin. For a full penetration weld, the ligament should be less than .01" to prevent the formation of root defects caused by insufficient stirring.

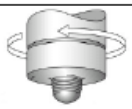
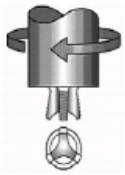
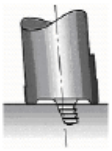
1.2 Friction Stir Welding Tools

Another important process parameter is tool design. As depicted in Figure 1, the FSW tool is comprised of two distinct cylindrical segments: the pin and shoulder. The pin refers to the portion of the tool which is plunged into the material, while the shoulder maintains contact with the workpiece surface. The pin's primary purpose is to facilitate stirring of the material. For full penetration welds, the pin length is usually marginally less than the thickness of the material with a diameter that is approximately 1/3 that of the shoulder. The shoulder serves a dual purpose, as it contains the expelled material while providing a majority of the heat input.

The shape of the pin has a significant impact on the flow of metal. In many instances, the formation of voids in the joint can be prevented by adding features to the pin which enhance vertical flow. As documented in the modeling work of Shercliff and Colegrove, a useful parameter for assessing the mixing capability of a given pin geometry

is the ratio of the swept volume to the volume of the pin [4]. This ratio is 1 for a smooth cylindrical probe (a geometry for which void formation is commonly observed), but can be as large as 2.6 for more complex geometries such as the MX Triflute pin developed by TWI. Additionally, machining scrolls into the shoulder of the tool The choice of a particular tool design is based on a number of factors: the joint configuration, thickness of the material, and the nature of the application, among others. Industrial applications where production of a defect free joint is of critical importance may necessitate the use of more complex, featured tools, while simpler tools may be sufficient for fundamental research studies in which production of defect-free joints is of less concern. A variety of tool designs, ranging from the common to the exotic, are detailed in Table 1.

Table 1 FSW Tool Geometries [5]

Tool	Cylindrical	Whorl™	MX triflute™	Flared triflute™	A-skew™	Re-stir™
Schematics						
Tool pin shape	Cylindrical with threads	Tapered with threads	Threaded, tapered with three flutes	Tri-flute with flute ends flared out	Inclined cylindrical with threads	Tapered with threads
Swept volume to pin volume ratio	1.1	1.8	2.6	2.6	depends on pin angle	1.8
Application	Butt welding; fails in lap welding	Butt welding with lower welding torque	Butt welding with further lower welding torque	Lap welding with lower thinning of upper plate	Lap welding with lower thinning of upper plate	When minimum asymmetry in weld property is desired

1.3 Materials

Friction Stir Welding tools are usually fabricated from tool steels; a cursory review of the literature suggests that AISI H13 is the most common tool material. H13 is a strong material (elastic modulus 30,500 ksi) with a melting point of 2600 F, approximately three times as great as temperatures likely to be experienced by the tool during the FSW process [6,7]. While FSW tools made from conventional materials have a semi-infinite life when used to join Aluminum alloys, they exhibit wear in the welding of harder materials such as metal composites and steel. These applications require the use of harder material tools such as Tungsten Carbide (WC).

The solid-state nature of the FSW process makes it ideal for joining Aluminum alloys which are difficult to weld using fusion methods. As such, there has been much research devoted to FSW of Aluminum 2XXX, 6XXX, and 7XXX series; FSW is generally regarded as a mature process for use with these alloys [7]. Research in recent years has focused on extending the application of FSW to include higher temperature materials such as steel, titanium, and magnesium [7]. Thermocouple measurements taken during FSW welds in Aluminum indicate that the temperature of the material remains below 900 F [7]. In welding of steel, titanium, and magenesium, the weld temperatures are significantly greater, necessitating the use of more advanced tool materials and designs. Additionally, there is interest in using FSW to weld metals such as copper, magnesium, steel, and composites to Aluminum. The joining of dissimilar materials using FSW has potential applications in the maritime and automobile industries.

1.4 Advantages of Friction Stir Welding

As indicated previously, the foremost advantage of FSW lies in its ability to weld materials which are difficult to join using fusion processes. Problems commonly encountered in welding of Aluminum alloys that are not observed in FSW of these materials include porosity, cracking, and distortion. Other advantages of FSW cited by Shercliff, Threadgill, and Withers are enumerated below:

- Since FSW is an autogenous process, the quality of the finished weld does not depend upon the skill of the operator. This contributes to enhanced repeatability.
- FSW is considered a "green process" in the sense that it does not use shielding gas or produce potentially harmful radiation.
- FSW is a robust process that is largely insensitive to small gaps along the jointline (known as fit-up errors) or slight variations in thickness.
- FSW is capable of producing welds with efficiencies, hardness values, and other mechanical properties comparable to (or in some cases exceeding) fusion-welded counterparts. From a microstructural perspective, FSW joints are generally superior to fusion joints.
- FSW can be applied to a variety of configurations, including butt, T, and lap joints. Recent advances in tooling have made it possible to weld contoured surfaces [8,9].

Though FSW typically compares favorably with other processes, there are some instances in which potential benefits may be overshadowed by cost. The purchase price of an FSW apparatus made by a commercial company such as Materials Testing Solutions (MTS)

can easily exceed tens of thousands of dollars. When cost is a concern, a more economical alternative may be to convert a milling machine for FSW through the addition of motor drives and computer control (an example of one such conversion is documented in reference 10).

1.5 Applications of Friction Stir Welding

Though some may argue that licensing fees associated with the international patent on the process have stifled the growth of the technology, there is little evidence of this in the aerospace, automotive, and maritime industries, which play host to a burgeoning set of FSW applications. In the aerospace industry, FSW has largely replaced fusion processes such as Variable Polarity Plasma Arc (VPPA) welding, the method previously used to weld the Space Shuttle's External Tank. Friction Stir Welding proved an ideal process for joining Aluminum Lithium 2195, the weight-saving alloy which constitutes the shuttle's only non-reusable component [11]. The elements which comprise NASA's next generation space architecture (Space Launch System and the Orion Crew Exploration Vehicle) will also make extensive use of FSW.

Since NASA first demonstrated the viability of the process for spaceflight applications in the late 1990s by integrating FSW into the external tank program, other companies have followed suit. The Delta II and Delta IV rockets, heavy-lift vehicles designed by Boeing and used primarily to launch military satellites, have friction stir welded fuel tanks [12]. The SpaceX corporation's Falcon series of rockets rely almost exclusively on FSW for joining components [13]. Eclipse Aviation, an aircraft company

which specializes in the manufacture of business jets, recently produced its first aircraft using FSW. In the Eclipse 500 model, 263 friction stir welds replaced approximately 7300 fasteners [14].

Friction Stir Welding has also expanded into the marine industry, where it is used to join the superstructures of large sea-faring vessels such as cruise ships. The Japanese ship "Ogasawara", a high-speed vessel for passenger and cargo transport, has Aluminum ship decks which were joined using FSW [15]. The auto industry makes extensive use of Friction Stir Spot Welding (FSSW), a spinoff of conventional FSW sometimes referred to as "fixed-point FSW", to join Aluminum panels. Mazda used FSSW for both its RX-7 and RX-8 sports car models, reporting savings in both cost and energy [16].

1.6 Microstructure of Friction Stir Welded Joints

The wide-ranging applications of FSW can be attributed to the superior mechanical properties of the finished joint. These properties are explained in part by an examination of the weld microstructure. The classification scheme for FSW microstructure is illustrated in Figure 2. The microstructure is essentially divided into three distinct zones: the unaffected (or parent) material, the heat affected zone (HAZ) and the thermomechnically affected zone (TMAZ).

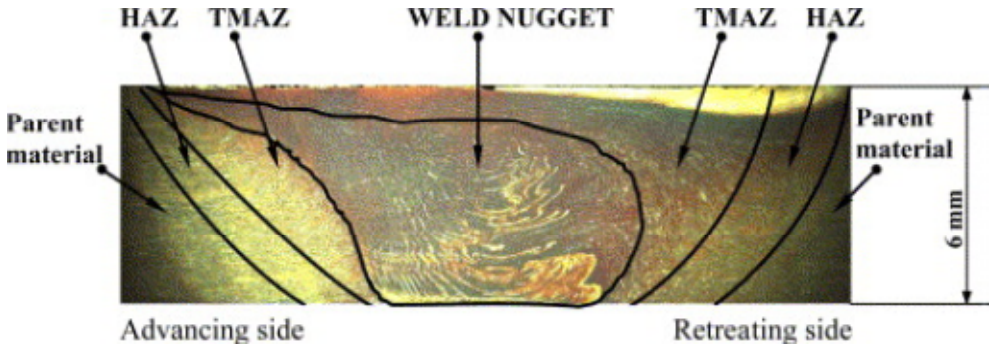


Figure 2 Microstructural zones in FSW [17].

The unaffected zone consists material distant from the weld that remains unaffected by heating. As such, it is able to retain the mechanical properties of the parent material. In the heat-affected zone, the material has been modified by heating but not plastically deformed. It is apparent from figure 2 that there is no obvious distinction between the grain structure of the HAZ and that of the parent material; rather the transition from unaffected material to the HAZ is signified by a drop in hardness [7]. Material in the thermomechanically affected zone experiences both heating and plastic deformation. This plastic deformation is most pronounced in the weld nugget, a region of the TMAZ that corresponds to a rise in hardness (a typical hardness profile for a FSW joint appears in Figure 3). The TMAZ is characterized by finely structured, recrystallized grains which enhance the material's resistance to fracture. Shercliff, Threadgill, and Withers report that the rate of crack growth in the TMAZ is significantly lower than that for the parent material. The extent of improvement of the aforementioned mechanical properties observed in the weld nugget depends upon the degree of recrystallization (and thus on the temperature attained during the weld).

An additional feature of the TMAZ is the circular markings, commonly termed "onion rings", which are formed by the revolution of the tool. There appears to be a connection between the number of onion rings formed per revolution and the symmetry of the tool geometry. For instance, a circular pin tool produces one ring per revolution, while a Trivex pin (which has three lines of symmetry) generates three rings with each rotation. Though some research in FSW has focused on correlating the size and shape of the rings with the presence of internal defects, there appears to be no relationship between the appearance of the rings and the resulting quality of the weld [7].

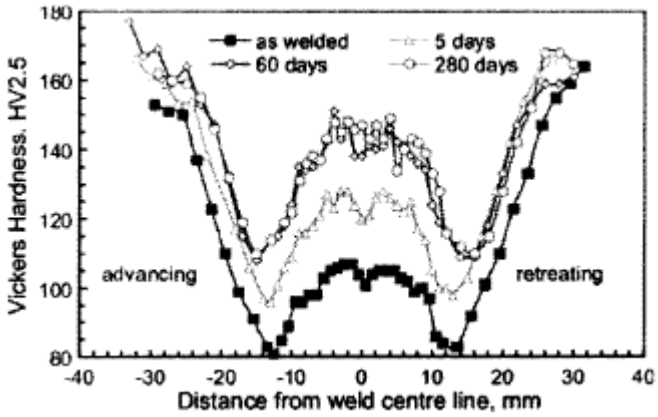


Figure 3 Hardness profile of FSW joint taken from Leonard [18]. Aging of the joint increases the hardness values.

1.7 Weld Defects

In addition to providing a glimpse into the features of the weld zones, macrographic analysis can be employed as a diagnostic tool to reveal internal defects. Before we address the various types of faults which can be detected, it is important to distinguish between weld defects and flaws. The Welding Institute defines a flaw as an imperfection in the weld; although undesirable, flaws are present in most welds [3]. A flaw which proves detrimental to the integrity of the weld is classified as a defect. Simply put, a defect is an unacceptable flaw. In an industrial setting, the criteria used to distinguish between a flaw and a defect depends on a number of factors: the location of the joint in the structure, the reduction in strength associated with the flaw, and the required safety factor.

The internal flaw most commonly encountered in Friction Stir Welding is a void in the weld nugget (Figure 4a). The formation of these voids (sometimes referred to as "wormholes") is usually attributed to insufficient stirring of the material. Thus a threaded pin geometry, which enhances material flow, can reduce the likelihood of wormhole development. Other factors which may contribute to void formation include inadequate (or excessive) heat input and deficient axial force. A void which is confined to the portion of the weld nugget closest to the base of the joint is termed a root flaw (Figure 5). A root flaw can be caused by either incorrect plunge depth or a pin which is substantially shorter than the thickness of the material, resulting in a large weld ligament; in both instances the material below the pin is insufficiently stirred. It is important to note that voids may extend the entire length of the joint line (creating a microchannel) or appear

only in a segment of the joint, the latter case being potentially less detrimental to the integrity of the weld than the former.

Figure 4 Wormhole on advancing side of weld nugget [18], a flaw common to unthreaded pin geometries [7].

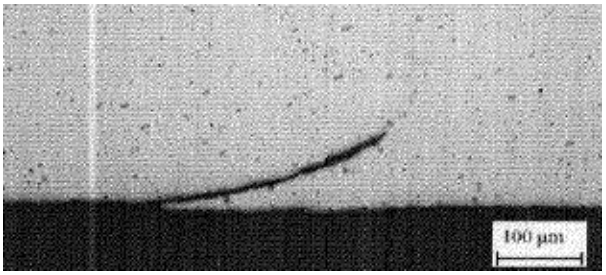


Figure 5 Root flaw produced as a result of insufficient plunge depth [7]

Though wormholes are common to all FSW joint types, there are some defects which are specific to joint configuration. The hooking defect, in which the unwelded boundary is rotated, is only observed in lap welds. On visual inspection, hooking appears as a notch feature on the advancing side of the weld (Figure 6). Since hooking is a defect which severely compromises the strength of the joint by thinning the plate, much research

has been devoted to eradicating it. Thomas et al. have linked the presence of the hooking defect to tool geometry and proposed a design to minimize its occurrence in lap joints [20].

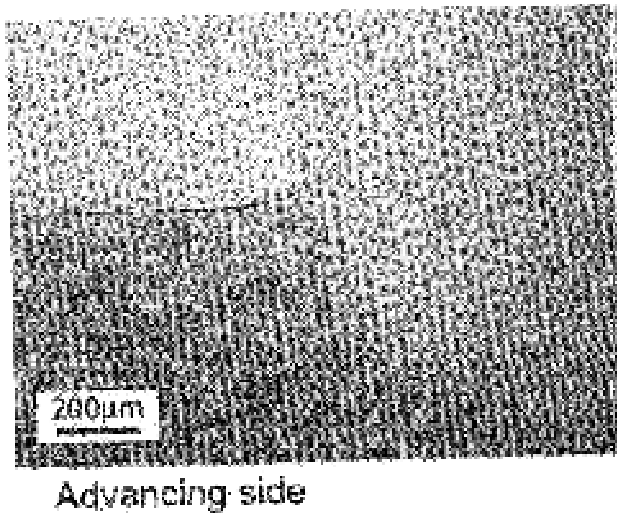


Figure 6 Hooking defect in FSW lap joint [7]

1.8 Weld Inspection and Evaluation

The flaws defined in section 1.7 may be insignificant. However, there are instances in which these flaws can seriously impact the integrity of the finished joint. Distinguishing between an inconsequential flaw and an unacceptable defect requires careful inspection of the weld. The easiest flaws to detect are those, such as flash or a lack of bonding, which are visible on the surface of the weld. Flash occurs when additional material is expelled from the joint and is principally attributed to excessive

plunge depths. Both flash and lack of bonding (indicated by the gap in the weld) are visible on the surface of the joint in Figure 7.

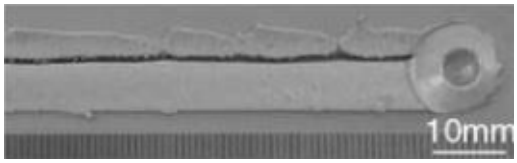


Figure 7 Joint with flaws (flash and lack of bonding) visible on surface [21]

Less observable flaws are those which are internal to the joint. Non-destructive evaluation (NDE) techniques such as ultrasonics and X-rays can detect the presence of voids on the order of micrometers. However, NDE remains expensive and there are some concerns regarding its reliability for FSW [7]. The most definitive method of flaw detection is destructive inspection. As its name implies, destructive evaluation requires cutting of the joint for microscopic evaluation and/or mechanical testing. For microscopy, a transverse cross-section of the joint is cut, polished, and etched to permit examination of the microstructure. Flaw detection using an optical microscope is limited to flaws which are visible under magnification (such as the relatively large wormhole pictured in Figure 4). The presence of more subtle flaws can be revealed through a series of mechanical tests: tensile tests, fatigue tests, and bend tests.

A tensile test provides a quantitative assessment of weld strength. Transverse or longitudinal segments of the weld are machined into specimens known as coupons (Figure 8) which are subjected to increasing tensile loads. The test is complete when the

specimen fractures; for FSW samples, failure usually occurs on the advancing or retreating side of the weld nugget [7]. The tensile strength of the joint is often expressed relative to the ultimate tensile strength of the parent material, a metric referred to as joint efficiency. For the proper parameters, FSW joint efficiencies compare favorably with those achieved in fusion welds: efficiencies exceeding 90 percent have been documented in FSW of Aluminum alloys [7].

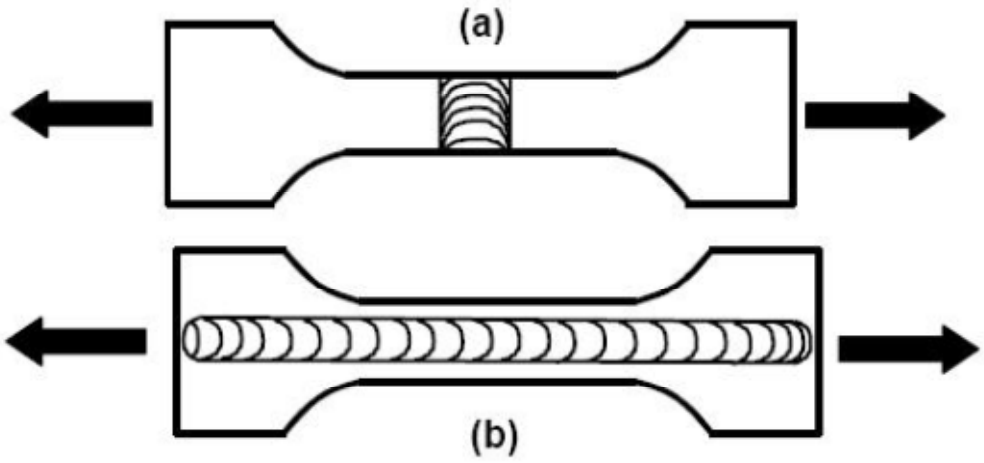


Figure 8 Two-dimensional rendering of tensile samples for FSW butt welds. (a) is tranverse segment and (b) is longitudinal. Arrows indicate the direction of the applied load [22].

In fatigue testing, the weld specimen is subjected to repeated application of loads which are below the yield strength of the parent material. The performance of the sample is plotted on an S-N curve (Figure 9) in which S represents the applied stress (the load normalized with respect to the cross sectional area of the specimen) and N denotes the number of loading/unloading cycles until failure. Although the fatigue performance of

FSW joints is considerably better than that observed in fusion welds, it is still markedly less than the parent material. Magnusson and Kallman have demonstrated that post-weld machining of the surface can further improve the fatigue properties [23].

In a bend test, the weld specimen is bent around a radius; the degree of the bend is increased until the sample cracks or fractures. While flaw free samples can be bent 180 degrees without cracking, welds containing any discontinuities or voids will fail the bend test [24]. As such, the bend test can be a powerful tool for identifying the presence of flaws which are too small to be detected by other methods. The bend test is particularly sensitive to flaws which occur at the root of the weld; Figure 9 shows a sample which failed the bend test as a result of a root flaw defect.

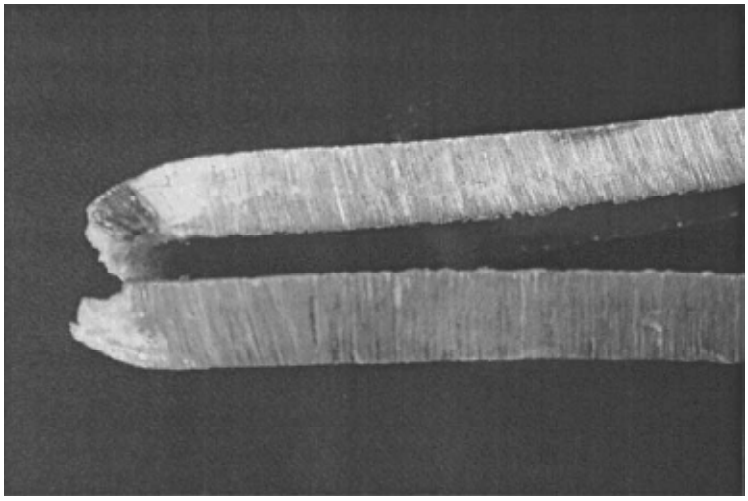


Figure 9 Bend test failure [25]

The importance of flaw detection is largely dependent on the FSW application. Monitoring and control of flaws is especially critical in aerospace structures which are subjected to extreme vibration during launch; a joint failure in these assemblies could be catastrophic. Flaw detection is an essential component of parameterization studies, which seek to establish an operating envelope (a set of parameters which produce defect-free welds) for a particular tool/workpiece combination. However, there are other instances, such as the wear studies presented in subsequent chapters, where flaw detection is a peripheral concern, superseded by the fundamental understanding and characterization of the wear process. Ultimately the criticality of flaw detection is dictated by the application and nature of the research.

CHAPTER II

METAL MATRIX COMPOSITES

2.1 Material Structure and Classification

Among the recent advances in the field of materials engineering is the development of composite materials, a class of materials particularly valued by the aerospace and defense industries for their high strength to weight ratios, resistance to temperature, and rigidity. Metal Matrix Composites (MMCs) are a class of composite materials which consist of two separate phases. The continuous phase, termed the matrix, is present in larger quantities. Embedded within the matrix is the reinforcement phase, which consists of abrasive material in the form of fibers or particles. MMCs are classified according to the materials which comprise the matrix and reinforcement phases, the shape of the reinforcement, and the concentration of the reinforcement Aluminum MMCs (metal composites in which the matrix is an Aluminum alloy) are categorized using a scheme developed by the Aluminum Association [26]. For instance, the classification Al 359/SiC/20p indicates that the material is Aluminum alloy 359 reinforced with 20 percent Silicon Carbide particulate by volume.

A microscopic view of Al 359/SiC/20p appears in Figure 10. The Silicon Carbide particles in this material display an irregular, angular geometry, although some particulate reinforcements may have a spherical or cubic structure. In most cases, the particles are randomly oriented and dispersed through the matrix, a condition which gives

rise to the term discontinuously reinforced Aluminum (DRA). MMCs can also have fibrous reinforcements consisting of long fibers of a harder material. If the aspect ratio, defined as the ratio of the length to the cross-sectional diameter, is large the reinforcement is classified as continuous; discontinuous fiber reinforcements are characterized by lengths and diameters on the same order [27]. While both fiber and particulate reinforced MMCs find applications in industry, particulate reinforcement is preferred due to its lower cost and ease of manufacture. Additionally, particulate reinforced MMCs are isotropic. Although the anisotropy of fiber reinforced MMCs can be exploited to enhance the strength of a structure in a particular direction, the directional dependence of material properties makes them more difficult to characterize and model.

Though the constituent materials of MMCs are relatively inexpensive, the price of composites is driven upward by costs associated with their manufacture. The most common forming methods for MMCs are liquid metallurgy and powder metallurgy. In a stir casting method described by Sahin, reinforcement particles or fibers are introduced into initially molten aluminum through a pipette [28]. Alternately, particulate reinforcement can be inserted into the solid matrix alloy through drill holes and mixed throughout once the metal has melted. The mixture is poured into a cast iron mold and allowed to solidify; hydraulic pressure is applied before billet removal to reduce the incidence of porosity [28]. A more uniform particle distribution can be attained through the powder metallurgy technique of hot isostatic pressing, which applies heat and pressure concurrently to the elemental matrix and reinforcement to induce consolidation [29], Both solid and liquid metallurgy techniques for MMC fabrication require the development of specialized equipment. The primary advantage of solid metallurgy is that

is can produce near-net shape parts. While liquid metallurgy yields a more uniform reinforcement distribution, significant machining is required to transform the part into the desired shape.

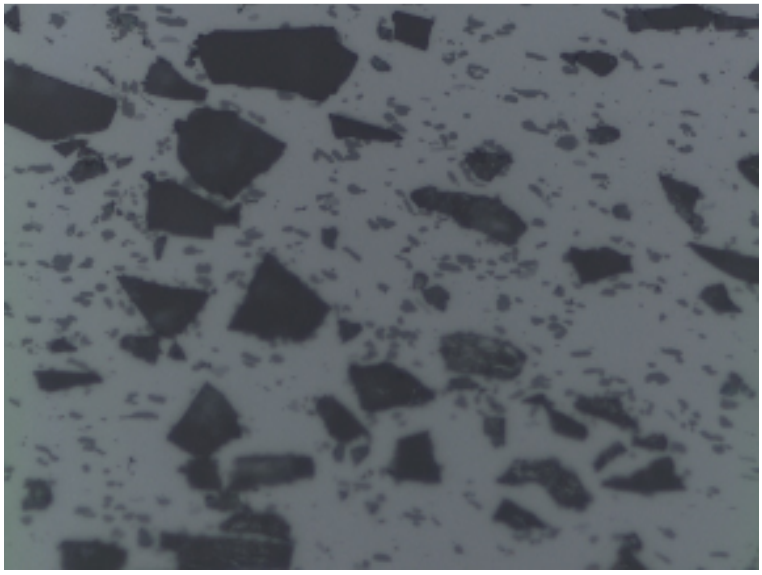


Figure 10 Microscope image of Al 359/SiC/20p at 100X magnification.

The degree of enhancement in mechanical properties attained with use of a metal composite (such as increased strength, temperature resistance, and hardness) is determined by the percentage reinforcement. Since the material property data available for MMCs is limited due to both their diversity and categorization as specialty materials, mechanical properties are estimated using the Law of Mixtures. Commonly used in calculations involving composite materials, the Law of Mixtures expresses a given

composite material property as a sum of weighted averages, where the weighting factors are the fraction of matrix and reinforcement present in the material:

$$X_c = x_m\, v_m + x_r\, v_r \quad \text{(equation 1-1)}$$

X_c corresponds to the mechanical property of the composite (such as density or elastic modulus); x_m and x_r denote the value of this property for the matrix and reinforcement materials, respectively; v_m and v_r represent the volume fractions for the matrix and reinforcement (for MMCs which contain only two phases, $v_m + v_r = 1$). The expression in (1) is also valid for weight fractions (recall that volume fractions can be converted to weight fractions using density) [27].

It is apparent from equation 1 that MMCs are in many respects "customizable": that is, we can design a composite with a desired property value by carefully selecting the matrix and reinforcement materials and controlling the proportions in which they are present. In the case of particulate reinforcement, which is usually added to the Aluminum alloy as a powder, the elastic modulus of the alloy can be significantly augmented with only a nominal increase in weight. Kunze and Bampton report that the elastic modulus of conventional Aluminum can be increased by 300 percent through the addition of SiC particulate reinforcement in a volume fraction of 70 percent [29]. However, while the increase in particle concentration substantially enhances the strength of the alloy, it makes machining more difficult, as cutting through the particulates results in rapid and severe wear of the cutting tool. As demonstrated by the work of Gugger, diamond tools have shown great potential in overcoming the problem of tool wear [30].

As a result of the challenges and costs incurred in machining MMCs, their use is currently relegated to applications where the added strength is critical to performance (and thus justifies the additional investment). Some "weight-saving" structures in which MMCs have been successfully implemented include the Space Shuttle Orbiter's structural tubing, the Hubble Space Telescope's antenna mast, control surfaces and propulsion systems for aircraft, tank armors, and braking systems for roller coasters [29, 31]. The potential applications of MMCs are equally varied, ranging from nuclear fuel containers to auxiliary components for satellites [29].

2.2 Fusion Welding of Metal Matrix Composites

An additional barrier to the widespread use of MMCs is the difficulty encountered when joining them to other MMCs or unreinforced materials in a larger structure. When MMCs are welded together using fusion methods, the quality of the joint is degraded by solidification and chemical reactions, effects documented by Stojohann et al. in reference [32]. The disparity in densities between the matrix and reinforcement materials results in particle segregation; this "clumping" of the reinforcement creates nonuniformity in the joint. Additionally, the higher viscosity of the composite material inhibits material flow during welding, potentially impacting the weld stress distribution and contributing to a reduction in strength [33].

The most problematic aspect of fusion welding Aluminum MMCs is the formation of deleterious theta phases, induced by the reaction of molten Aluminum with reinforcement particles. In the case of Silicon Carbide reinforcement, liquid Aluminum

reacts with SiC to produce Aluminum Carbide (Al_4C_3). Stojohann et al. indicate that the reaction, initiated when the weld temperature exceeds the melting point of Aluminum, is highly dependent on temperature. The amount of Al_4C_3 present in the completed joint (visible in the microscope image of Figure 11 as needle-like formations) is proportional to the difference $T_w - T_m$, where T_m is the melting point of the matrix and T_w signifies the weld temperature (the reaction proceeds only when $T_w \geq T_m$).

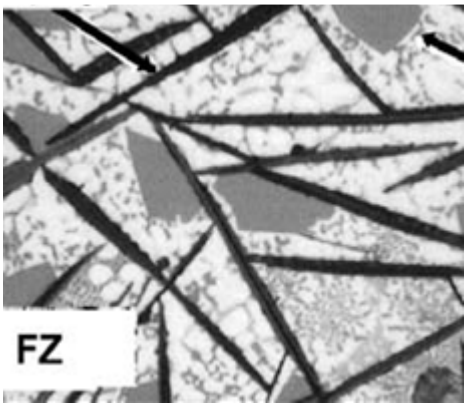


Figure 11 Aluminum Carbide in the fusion zone of a laser-beam weld of SiC reinforced Aluminum [32].

Stojohann et al. assessed the propensity for Al_4C_3 formation and particle degradation in three different fusion welding processes: gas tungsten arc (GTA), electron beam (EB), and laser beam (LB). Two composite materials were considered, one with Al_2O_3 particulate reinforcement and another with SiC fiber reinforcement. Optical evaluation of the microstructures in the welded Al_2O_3 reinforced composites indicated degradation of the reinforcing particles for each of the processes considered, a result of the decomposition of Al_2O_3 reinforcing particles to gas in the presence of molten

Aluminum [32]. The clumping of particulates, observed for GTA and EB welds, creates porosities in the joint. For the SiC reinforced composites, the SiC whiskers remain intact, but the Al_4C_3 phase is ubiquitous in all of the fusion welds. Though the amount of Al_4C_3 produced can be somewhat mitigated through careful control of heat input (since the reaction which governs the formation of Al_4C_3 is accelerated by an increase in temperature above T_m), Al_4C_3 is still prevalent in the EB welds, the fusion process with the lowest heat input per unit length. Thus Storjohann concludes that the formation of the Al_4C_3 phase is unavoidable in fusion welding of Al-SiC composites.

Due to the undesirable chemical reactions which take place in the composites when the matrix alloy is melted, a solid-state welding method may yield better results. Since the weld temperature in FSW is less than the melting point of the matrix alloy ($T_w < T_m$), the undesirable chemical reactions which lead to the dissolution of reinforcement in Al-Al_2O_3 and the formation of Al_3C_4 in Al-SiC are unable to proceed. Storjohann used FSW to produce defect-free welds in both the Al-Al_20_3 and Al-SiC composites. Although intermittent particle clustering was observed in the Al-Al_2O_3, the particle distribution was more uniform than that found in fusion welded joints. While the number of particles present in the parent material and the joint is on the same order (as the Al_2O_3 will not dissolve in the absence of liquid Aluminum), the particle distribution in the joint is skewed toward smaller diameter particles. This outcome is a sign of particle breakage, an effect Storjohann attributes to the brittleness of the reinforcement and its inability to "flow" with the surrounding Aluminum alloy.

In the samples of Al-SiC composite joined by FSW, none of the deleterious Al_3C_4 phase was detected. Additionally, no breakage of the SiC whiskers was observed, owing

to their small size and ability to reorient themselves in the flowfield of plasticized Aluminum. In the parent material, the whiskers are aligned along the x direction; in the joint their orientation is primarily in the z direction, the axis coincident with the direction of welding (Figure 12). Storjohann postulates that this change in orientation may prove favorable to the mechanical properties of the joint, though further research is required to substantiate this claim.

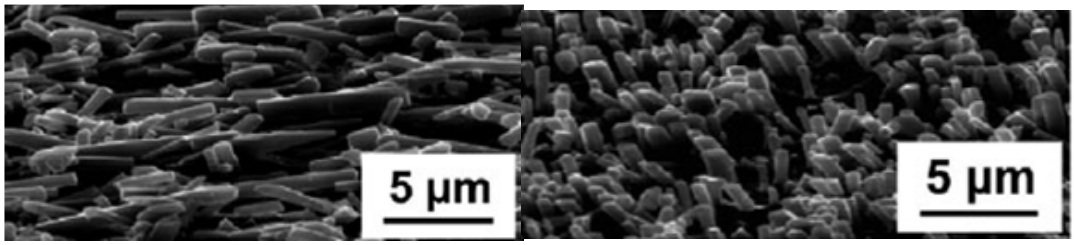


Figure 12 Re-orientation of SiC whiskers in FSW of Al-SiC composite [32]. The left SEM image is the whiskers as they appear in the base metal, while the right image captures the whiskers in their re-oriented state in the weld nugget.

2.3 Features of Friction Stir Welded MMCs

As documented by Storjohann et al., Friction Stir Welding is the preferred method for joining metal composites since it precludes formation of deleterious phases within the joint. There have been a number of published studies which assess the characteristics of MMC joints produced using FSW processes. These distinguishing features are enumerated briefly in the list which follows.

- Reduction of reinforcement particle size in weld zone. When compared with the shape of the particles in the parent material, particles in the weld region have

smaller diameters and are round, rather than angular, in shape [34]. This size reduction, which may be as large as 20%, suggests that the brittle particulate reinforcement fractures in the wake of the stresses which accompany plastic flow of the metal matrix [35]. These changes in size are particularly prominent in the weld nugget, the zone coincident with the largest amount of heating and plastic deformation [36]. The size reduction and dispersal of particles is evident in Figure 13, which compares SEM images of the particles in the parent material and joint for Al 2009/SiC/15p [34]. Fracture of the particles, apparent in the right image, is confined to areas with concentrations of large diameter particles; small particles debind from the matrix rather than fracture.

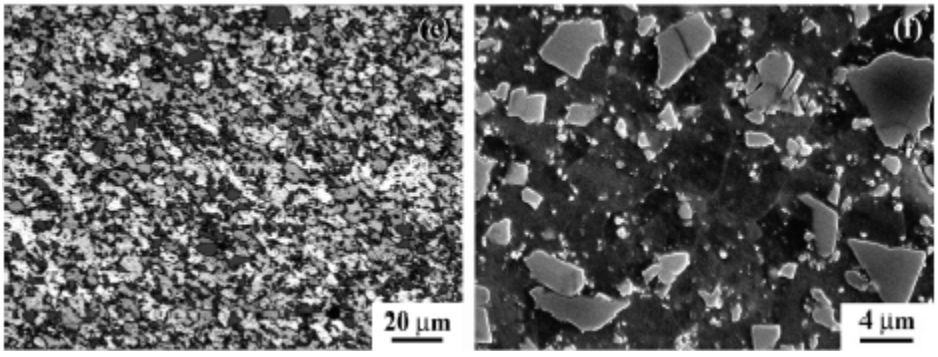


Figure 13 Comparison of SiC particles in base metal Al 200-/SiC/15p (left) and FSW joint (right) [34]

- Changes in particle distribution. In addition to rounding the reinforcement particles, stirring of the metal composite in FSW induces their rearrangement. In the base metal of a discontinuously reinforced metal composite, there exist aggregates of reinforcement which are significantly larger than the median-sized

particle; thus the particle size distribution (plotted as a histogram of frequency versus size) will be skewed toward larger particle sizes. These large particle conglomerates are transformed into a homogeneous distribution through the particulate rounding and scattering effects of FSW [37]. A comparison of the pre and post-weld particle size distributions for Al_2O_3 particulate reinforcement reveals that the Al_2O_3 particles are homogeneously distributed in the weld nugget by the stirring action of the tool. The abrasive particles in the nugget are of similar size and shape (aspect ratio), an attribute that is not observed in the base metal.

- Grain refinement. The phenomenon of grain refinement is well-documented in FSW of Aluminum alloys and it is even more pronounced in FSW of Al-MMCs. The precise mechanism by which grain refinement occurs is a topic of scholarly debate and is beyond the scope of our research concerns. The effect is closely related to the dynamic recrystallisation of the plasticized material which consolidates behind the tool to form a weld. In FSW of metal composites, grain refinement is enhanced by the reinforcing material, as the larger reinforcement particles provide additional nucleation sites to facilitate grain growth, subsequently reducing the size of the grains in the finished joint [35]. Humphrey has developed a model which predicts the recrystallized grain size, *D*, from the volume fraction of the composite (F_v) and the initial average diameter of the reinforcing particles (*d*) [38]. Based on Humphrey's expression $= dF_v^{\frac{-1}{3}}$, the grain size in the weld nugget decreases with increasing reinforcement percentage,

a prediction that is consistent with the nucleation site hypothesis proposed by Boromei in reference 35. Thus recrystallization efficiency is improved by the presence of reinforcing particles.

- Hardness. There is some disagreement in the published literature regarding the hardness of the weld zone. Marzoli et al. observe a hardness profile similar to that found in Al 6XXX alloys (Figure 3), with a minimum hardness value in the heat-affected zone [39]. The hardening of the HAZ which occurs in fusion welding, a consequence of reinforcement particle migration toward the center of the joint line, is not detected in FSW joints by Marzoli. Similar findings are reported in reference 36; Ceschini records an 11 to 25 percent drop in hardness in the weld zone as compared with the base material. However, other researchers have documented a hardening of the HAZ by as much as 30 percent, an increase evident in the hardness profiles of Figure 14 [40]. Though this hardening appears to be anomalous result, Shindo et al. attribute the discrepancy to their use of an as-cast Al-SiC composite, which has different characteristics than the Al-Al_2O_3 composites used in the other studies. Shindo claims that the rise in hardness of the weld zone is more pronounced for slower traverse rates. Additional variations in hardness coincide with wear of the tool; the relationship between hardness and wear will be discussed in detail in a subsequent section.

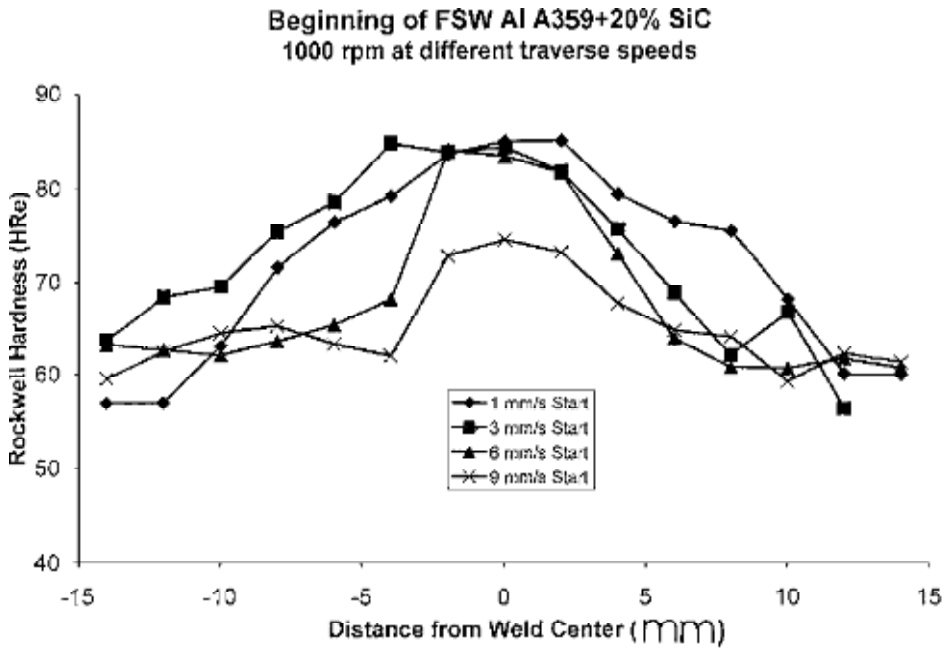


Figure 14 Comparison of hardness profiles for FSW of Al 359/SiC/20p for a range of travel speeds [40].

- Joint Efficiency and Weld Envelope. Joint efficiency, defined as the ratio of the yield strength of the weld specimen to that of the parent material, is highly dependent on process parameters. The highest joint efficiency documented in the literature for friction stir weld of a metal composite is 84.4%, recorded by Marzoli et al. in a sample of Al 6061/Al_2O_3/20p. Other parameterization studies report similar values, typically in the range of 70 to 80 percent [35, 36]. Although this range of maximum joint efficiency is smaller than that achieved in FSW of unreinforced alloys, it can be enhanced by heat-treating the as-welded specimen to a T4 condition [34]. Comparison of joint efficiencies for welds produced at various process parameters is used to define an operating window, a set of parameters which produce acceptable joints. The process envelope, taken from

Marzoli et al., for the frictions stir welding of Al 6061/Al_2O_3/20p is shown in Figure 15.

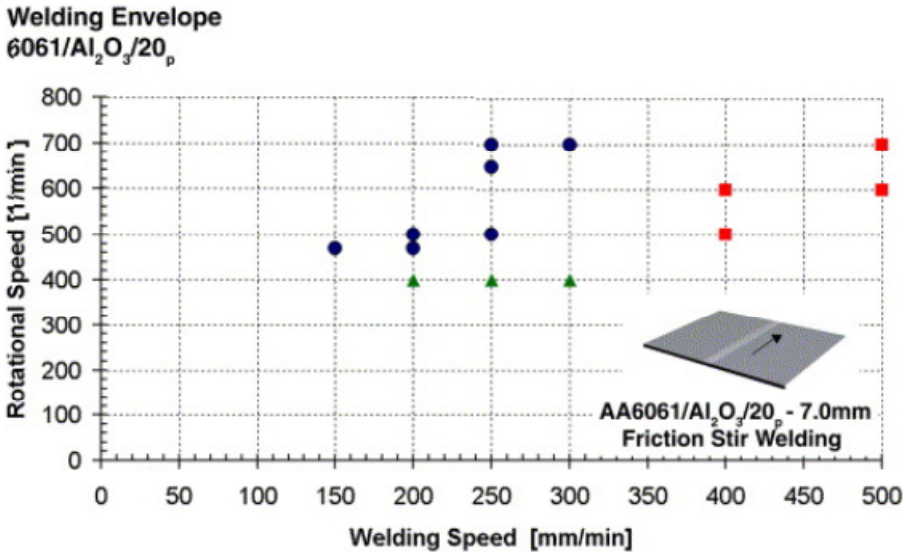


Figure 15 Weld envelope for FSW of Al 6061/Al2O3/20p. The triangles delineate a process limit, below which no acceptable welds were produced. The circles denote welds with satisfactory joint efficiencies. Squares indicate mechanical limits of the welding apparatus [39].

Compared with the weld envelope for unreinforced Aluminum alloys, the range of operating parameters available for FSW of metal composites is much narrower. The composite is in some ways less forgiving than its unreinforced counterpart, as the presence of the brittle reinforcing particles inhibits plasticized flow of the matrix alloy [34]. This decreased metal flow contributes to the formation of voids which can negatively impact the tensile strength. It may be possible to counteract the tendency for void formation in MMCs by implementing more complex tool geometries (Table 1) which enhance flowfield characteristics.

The rounding of the particles and the more uniform particle distribution which arises in the FSW process increase the ductility of the weld specimens, a change which translates into an improvement in tensile properties. This claim is supported by the computer simulations conducted by Marzoli et al.; particle distributions in which the particles are smaller and similar in size demonstrate increased ductility and are therefore able to withstand higher tensile stresses [39]. Thus the high joint efficiencies which comprise the process envelope correspond to parameters wherein the material is stirred sufficiently to generate favorable particle distributions.

- Temperature. MMC materials possess a high specific heat and resistance to temperature, properties which make the plasticization of the material required for FSW more difficult. To achieve plastic flow in MMCs, the tool must impart additional heat to the material at the start of the weld. The higher heat input necessitated by the temperature resistance of MMCs is evident in Figure 16, a plot of thermal cycles for six locations along the weld joint in FSW of Al 6061/Al_2O_3/20p [39]. Although this need for a "dwell" period in which the tool is stationary but rotates to generate heat for plasticization is present in conventional materials, it is of longer duration in MMCs since more heat is required to induce flow of the matrix alloy. Referring to the operating envelope developed by Marzoli (Figure 15), it seems that larger weld pitches (consistent with higher heat input) produce joints with improved properties.

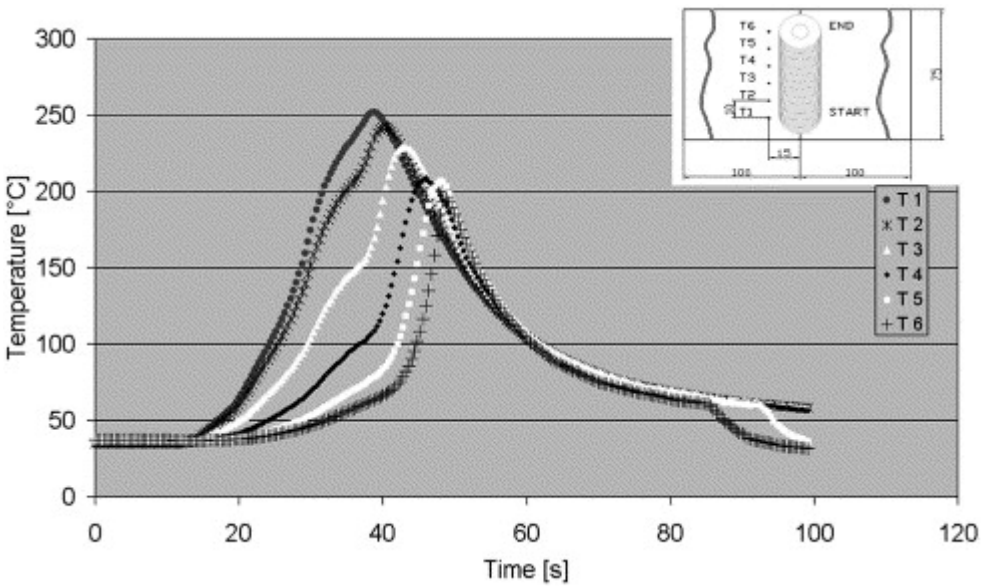


Figure 16 Temperature profiles recorded in FSW of Al 6061/Al_2O_3/20p. Each curve corresponds to a thermocouple location (shown in the inset) [39].

- Surface characteristics. The surface appearance of an Aluminum joint made with FSW is characterized by a polished, mirror-like finish. MMC joints have a coarser surface structure, presumed to be a consequence of decohesion between the matrix alloy and reinforcement at the surface of the joint [33]. Tools made of harder materials with a lower lubricity, such as Boron Carbide (B_4C), produce a smoother joint surface due to slippage at the interface between the workpiece and the tool [33].

- Fatigue. The fatigue life of the welded MMC specimen is lower than that observed in the parent material [35, 36]. This reduction in fatigue strength is most prevalent in samples with void defects: when subjected to high strain-amplitudes, these specimens fail prematurely, a consequence of the high stress concentration

present in the vicinity of the void [35]. The reduced fatigue life in specimens which do not contain voids can be attributed to the disturbances to the base material caused by the FSW process [36]. Boromei et al. report a 30 to 1 ratio for fatigue lives at high strain-amplitudes between the base composite material (Al 7005/Al_2O_3/10p) and the welded specimen. This ratio decreases to 10:1 for low strain-amplitudes, observations which are consistent with the claims made in references 35 and 36 concerning the impact of void defects on fatigue life.

- Ductility and failure mechanisms. As discussed in the section on joint efficiency, the ductility of metal composites is reduced by the presence of brittle reinforcement particles. This decreased ductility is evident in the mechanical response of the specimen to tensile stress; elongation in the weld is slightly less than that of the base material [41]. However, FSW welds of composites exhibit increased ductility compared to fusion welded specimens of the same material, owing to the rounding of the reinforcement particles induced by the FSW process. Tensile failure in FSW samples is consistent with ductile fracture originating in the weld's thermo-mechanically affected zone (TMAZ). This ductile fracture mechanism is indicative of the good plasticity properties of the joint and the refined, recrystallized grain structure created by FSW [41].

  According to Ceschini et al., failure of an FSW metal composite can generally be attributed to one of three causes:

1. Cracks, originating in large reinforcement particles that have not been broken up by the stirring action of the FSW tool, propagate and increase the specimen's succeptibility to fracture.
2. Decohesion of the matrix and the reinforcement particles, often apparent in the weld's surface structure (coarse surfaces are characteristic of decohesion).
3. Presence or growth of voids. As mentioned previously, voids introduce stress concentrations which can compromise the mechanical properties of the specimen. Voids which are initially present can increase in size as a result of decohesion [36].

The cause of failure, when discernable, can inform selection of process parameters which produce joints with more favorable properties. For instance, an increase in weld pitch ($\frac{\omega}{v}$) results in increased stirring of the material, an effect which may reduce the formation of the voids which are often responsible for specimen failure.

2.4 Wear of Friction Stir Welding Tools in the Joining of Metal Matrix Composites

2.4.1 Measurement of Wear

A final characteristic of MMC joints made by FSW is the presence of debris products in the weld specimen, deposited along the joint line as the tool wears. Wear of the tool in the FSW process is a phenomenon unique to high strength alloys, having been documented in the literature only for FSW of steels and metal composites [26]. Prado et al., who used FSW of Al 6061 as a baseline for comparison with tool consumption in MMCs, report no detectable wear for FSW of conventional Aluminum alloys [42]. The wear of tools in composite materials is rapid and severe enough to be discerned with the naked eye; Figure 17 shows the shape changes observed in a steel FSW probe for successive welds of Al 6061/Al_2O_3/20p. After only two welds (1.82 meters) the initially threaded tool resembles a smooth cylindrical probe with rounded tip.

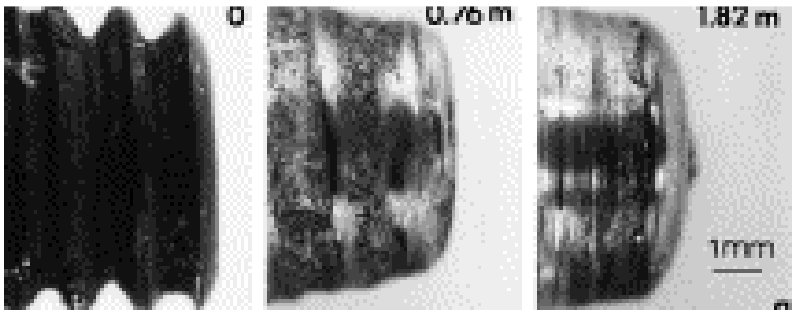


Figure 17 Images from Prado et al. showing shape changes in steel FSW tool for two successive welds of Al 6061/Al_2O_3/20p. Parameters are 500 RPM and 1 mm/s and each weld was 0.76 m in length [43].

Tool material eroded as a result of contact with hard reinforcing particles is deposited along the joint line. This abraded material can be detected and quantified by performing an energy dispersive X-ray (EDX) analysis on a transverse cross-section of the weld [34]. EDX can detect the presence of particles originating from the tool as well as reveal the regions in which these erosion products are concentrated. The results of Feng, Xiao, and Ma show an increased iron concentration in the weld nugget, an indication of a rise in wear in regions with greater concentrations of reinforcement [34]. Thus this observation corroborates the prevailing theory that contact between the tool and the reinforcement particles is the primary cause of wear in FSW of metal composites.

Although EDX provides useful metrology information, destructive methods (i.e. sectioning of the joint) are required to obtain a sample suitable for analysis. There are a number of alternative methods which analyze wear based on post-weld examination of the tool rather than the welded specimen. The most obvious and easiest to implement of these methods compares the weight of the tool before and after welding; the weight lost by the tool is approximately equal to the amount of material eroded. Unfortunately, sensing tool wear via mass comparison is complicated by the Aluminum which accumulates on the tool during welding. Although this Aluminum residue can be etched away using a base solution, weighing is regarded as a crude method for this application due to the concurrent acquisition and deposition of material inherent in FSW of high-strength alloys [26]. Mechanical gauging methods, such as the use of calipers to record changes in probe length and diameter, have similar limitations [26]. Additionally, gauging techniques are largely insensitive to the nuanced shape changes which occur as the tool wears (Figure 17). In contrast, measurement systems which rely on optical

comparisons of the tool are generally regarded as more reliable indicators of wear [26]. Stellwag and Lienert use a shadowgraph technique to quantify the percentage of the tool probe consumed during welding. In a novel hybrid of weighing and optical comparator methods, Fernandez and Murr measure wear by capturing close-up photos of the tool probe, cutting out the probe in these images, and comparing the weights of the cut-outs. The assumption that the weight of the two-dimensional image cutout is indicative of the tool's material loss is substantiated by a series of parallel experiments which calculate percent wear by comparing masses of the etched tool after each weld [44].

Most studies of wear in FSW of composite materials are exclusively concerned with wear of the probe. The succeptibility of the shoulder to wear is a subject of some debate. Liu et al. measured the shoulder diameter between subsequent welds, but found the differences to be negligible [45]. A conclusive indicator of shoulder wear would be an increase in the probe length after welding due to erosion of tool material at the probe/shoulder interface, an effect that is observed by Lee, et al. [33]. However, since wear of the shoulder is not observed in any other published research, it appears that the outcome documented in [33] is an anomalous result, attributable to an excessive plunge depth used to increase heat input and ensure plasticization of the temperature-resistant workpiece material.

2.4.2 Variation of Wear with Process Parameters

Once a robust method of wear measurement has been established, the variation of wear with process parameters can be investigated. The wear experiments detailed in references 40 and 42-45 relate the volumetric wear of the probe to the process parameters rotation speed (ω), travel speed (v), and distance welded (ℓ). Though the studies utilize different metal composite materials and tool geometries, there are some trends which seem to hold in general for FSW of metal composites. An obvious direct proportionality between wear and linear weld distance ℓ is reported in all of the aforementioned articles. In particular, Liu et al. note the dramatic reduction in the diameter of the probe with increasing weld length [45]. Lee et al. determine the average reduction in probe diameter to be 0.010 inches per foot of weldment, degradation substantial enough to necessitate replacement of the probe after approximately 5 feet of weld [33]. It is clear from the plots of percent tool wear versus weld distance displayed in Figures 18 and 19 that the variation of wear with distance is non-linear. In fact there is usually some critical distance, x_{cr} , beyond which no additional wear is observed. The significance of this plateau in the wear curve observed for values of ℓ greater than x_{cr} is discussed in section 2.4.3.

The dependence of wear on rotation speed and traverse rate is less palpable. To isolate the effect of rotation speed on wear, Fernandez and Murr performed welds at a fixed traverse rate while varying the spindle speed from 500 RPM to 1000 RPM. As evidenced by the wear curves plotted from their data (Figure 18), percent tool wear increases with increasing rotation speed. This trend is also observed by Prado et al., who find that wear increases with rotation speed up to 2000 RPM [43]. To characterize the

dependence of wear on traverse rate, Fernandez and Murr expanded their experimental matrix to include variations in traverse speeds [44]. The wear curves for the seven combinations of traverse and rotation speeds considered are plotted in Figure 19. By comparing wear data for parameter sets with the same rotation speed but different traverse rates, it is apparent that wear decreases with increasing traverse speed. Though the inverse relationship between wear and v is nonintuitive, it provides experimental evidence for the theory that tool wear in the FSW process is a shear, rather than drag, phenomenon. Shear versus drag phenomena will be discussed at length in chapter IV, which proposes a physics-based model for tool wear in FSW.

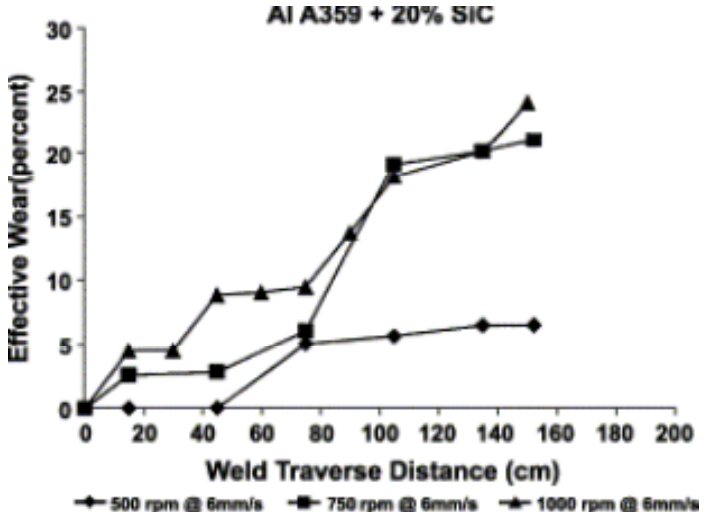


Figure 18 Plot of wear versus weld distance for 500, 750, and 1000 RPM [44].

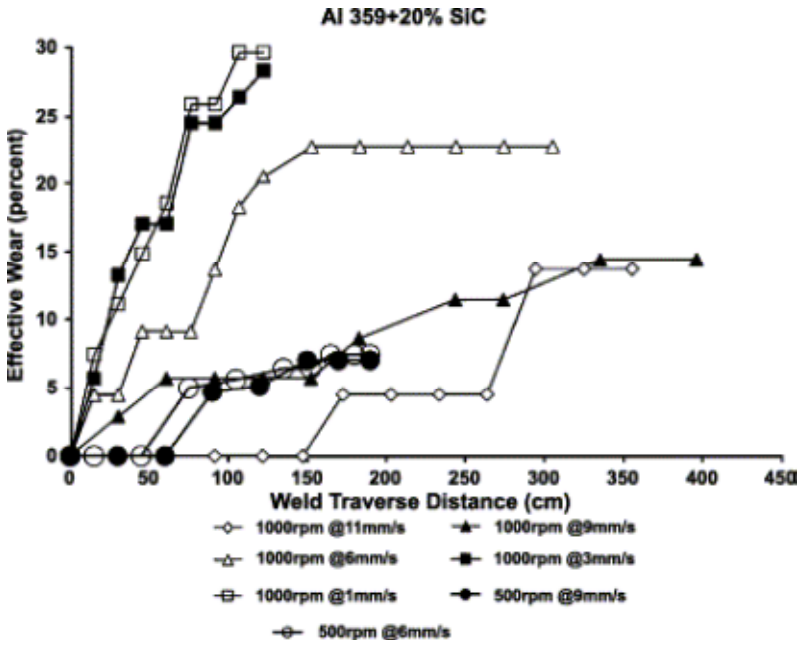


Figure 19 Plot of wear versus weld distance for a range of rotation and traverse rates [44].

While the data cited above is by no means a complete characterization of the variation of wear with process parameters, it indicates that material loss in FSW of metal composites is inversely proportional to traverse rate and directly proportion to the rate of rotation:

$$W \propto \frac{\omega}{v} \qquad \text{equation (1-2)}$$

The symbol W denotes percent wear, the complement of the ratio of the tool volume after welding to the original tool volume expressed as a percent (recall that for the Fernandez/Murr data, differences in the weights of the photo cut-outs are regarded as indicators of volumetric wear). Based on the functional dependence of wear on ω and v indicated in (2), slow rotation speeds and high traverse rates are paramount to the

minimization of wear. It is thus not surprising that of the parameters selected for the Fernandez/Murr study, the least amount of tool wear was observed in the 500 RPM/6 mm/s and 500 RPM/9 mm/s cases, as the values of the parameters in these cases correspond to the maximum traverse speed (11 mm/s) and minimum rotation speed (500 RPM) considered. Though Fernandez and Murr do not interpret their results in terms of the weld pitch ratio $\frac{\omega}{v}$ of equation 1, it is apparent from Figure 19 that lower weld-pitch conditions result in significantly reduced wear. Unfortunately, minimization of weld pitch is not a viable solution to tool wear when weld quality is a concern: Crawford, et al. argue that the formation of wormholes is more likely to occur at low weld pitches owing to the lack of adequate flow [46]. The competing effects of defect formation and rapid wear discussed in suggest that controlling wear through the selection of process parameters may be a difficult prospect.

The process parameters v, ω, and ℓ are not the only factors which influence wear. A study conducted at NASA Marshall Space Flight Center by R. Diwan concludes that the ease of welding a metal composite is in inverse proportion to its percentage reinforcement. Categorizing the materials based on reinforcement concentration, Diwan ascertains that category IV and V composites (which have SiC reinforcement percentages of 40 and 55 percent, respectively) are "unweldable" using conventional tool materials [47]. Diwan's findings are similar to those of Lee, et al., who determine that the production of satisfactory welds using steel tools is limited to materials with a reinforcement percentage no greater than twenty-five percent [33].

2.4.3 Rate of Wear and Evolution of the Self-Optimized Shape

It is evident from the wear curves of Figure 18 and 19 that the rate of wear (recorded in units of percent effective wear per centimeter) varies with process parameters v and ω as well as the distance welded. In general, the initial wear rate of the tool decreases with increasing traverse rate, a trend shown in the plot of Figure 20 [40]. In the Fernandez/Murr data (Figure 19), the wear curve for the maximum traverse speed considered (11 mm/s) has a slope of zero (indicating that no wear of the tool occurs) until the weld distance exceeds 150 cm [44]. For parameters with greater values of $\frac{\omega}{v}$, the wear rate is most rapid at the outset of welding, an effect that is particularly pronounced in Figure 19 for the 1000 RPM/1 mm/s and 1000 RPM/3mm/s cases. Prado et al. postulate that this increased initial wear rate is related to the temperature of the workpiece material: because the material is colder at the start of the weld, higher forces are required to move the tool through the workpiece [42]. Similar results appear in reference 45, wherein the maximum wear rate is always recorded during the early phases of welding. Stelwag and Lienert document an analogous increase in initial wear rates for friction stir welding of steel (approximately eighty percent of the total wear occurs during the initial plunging and dwelling stages), a rise they also attribute to greater material flow stress at reduced temperatures [26].

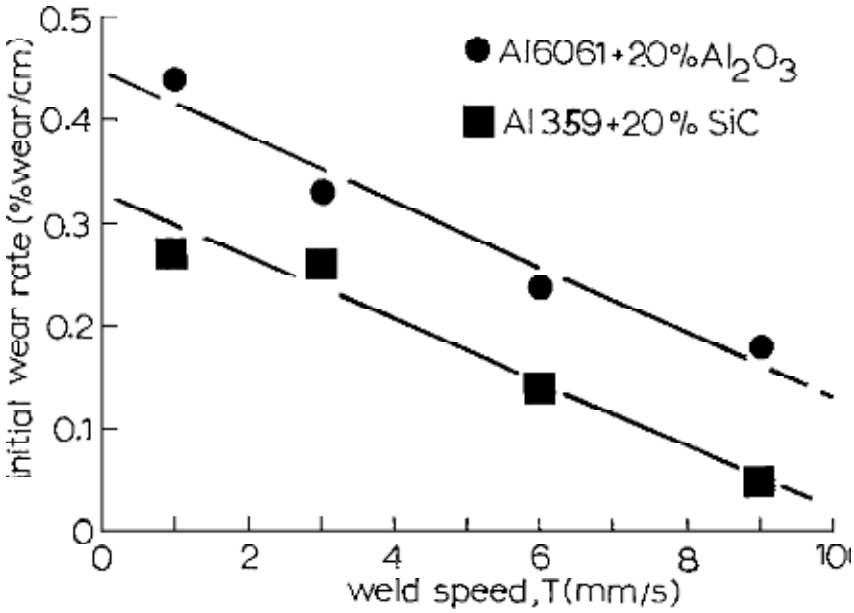


Figure 20 Plot of initial wear rate versus traverse rate (rotation speed is constant at 1000 RPM). The plot combines data from Prado et al. and Shindo et al. [44]

As mentioned in section 2.4.2, there are some parameters for which the wear curve plateaus with increasing weld distance. The termination of wear with some critical distance x_{cr} is clearly visible in the 1000 RPM/6 mm/s curve of Figure 19, where wear effectively ceases for distances greater than 150 cm. This critical distance is not observed for wear curves corresponding to parameters with comparatively large values of $\frac{\omega}{v}$ (such as 1000 RPM/1 mm/s and 1000 RPM/3 mm/s in Figure 19). Shindo et al. contend that the high initial wear rates consistent with these parameter sets preclude the existence of a critical weld distance, as material loss during the initial weld phase is substantial enough to make the steady-state plateau associated with zero wear unattainable.

The shape of the tool that is present at the cessation of wear is referred to as the self-optimized shape [40,44]. The sequence of images in Figure 20 shows the evolution of a probe from its original state to self-optimization [40]. During the initial stage of the weld, Fernandez and Murr claim that the workpiece material "fills" the threads, inhibiting the wear mechanism and delaying the onset of wear. However, the threads are eventually eroded by the abrasive reinforcement particles, producing the smooth frustum-like shape seen in the rightmost image of Figure 21. The figure also demonstrates that radial wear along the length of the probe is nonuniform. For the 1000 RPM/9 mm/s, maximum radial wear occurs at the root of the probe, while the diameter near the shoulder of the probe remains virtually unchanged (indicating that little to no wear occurs in this region). Liu et al. also studied the variation of radial wear with distance, ultimately finding that different regions of the probe experience different rates of wear. The amount of wear recorded at a particular location is additionally depends on the process variables, an observation that is apparent in Figure 22, which compares self-optimized shapes for two of the parameter sets considered by Shindo et al. These images, which correspond to the zero wear rate regimes in the wear curves of Figure 19, clearly illustrate that the optimized shape varies significantly with welding speed. The left image (1000 RPM/6mm/s) exhibits the pseudo-hourglass shape characterized by Liu et al., wherein the maximum wear occurs at a location 1/3 of the distance between the root and the end of the probe. Using the results of Liu et al. and defining x as the linear distance from the shoulder and ℓ as the length of the pin, the probe can be divided lengthwise into three distinct sections corresponding to low, moderate, and high wear:

i) $0 < x < \ell/3$. Little to no radial wear of the probe occurs in the vicinity of the shoulder ($x \approx 0$), but wear increases with increasing distance from the shoulder x.

ii) $x \approx \ell/3$. This is the region of maximum wear -- for the optimized shape, the probe diameter is smallest in this region.

iii) $\ell/3 < x < \ell/3$. The amount of wear decreases from $x \approx \ell/3$ to $x = \ell$. While the diameter of the probe in this region is greater than that in (ii), it is slightly reduced from the diameter in (i). Thus the wear regime for this range of x values is classified as moderate.

This hourglass shape described by Liu et al. is more pronounced for parameters which experience comparatively greater wear (such as the 1000 RPM/6 mm/s in Figure 22). Not surprisingly, parameters which result in less wear produce more subtle changes in shape. There is no existing literature which explains the observed shape variations in terms of a physical wear mechanism. Additionally, the current research on optimization of tools in FSW of metal composites offers little explanation as to why the self-optimized shape is attainable for some parameters but not others.

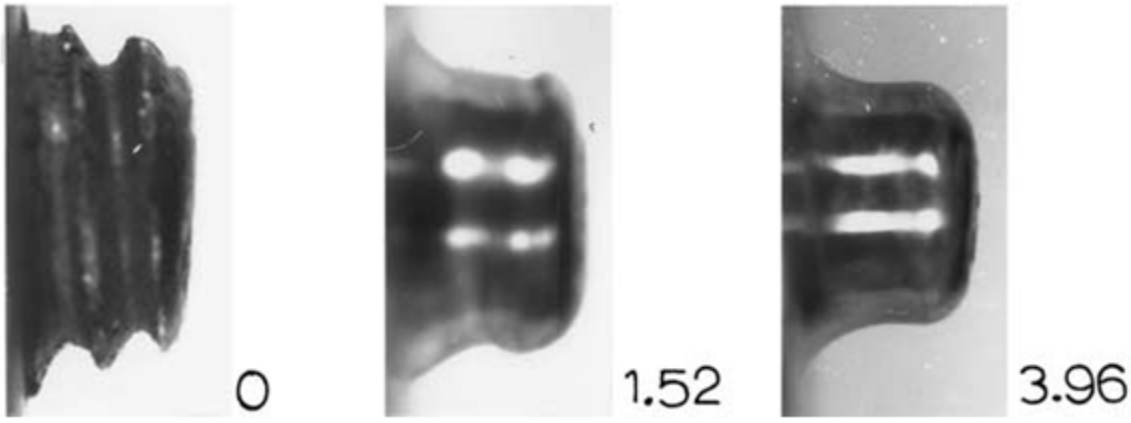


Figure 21 Evolution of self-optimized shape for 1000 RPM, 9 mm/s case. The leftmost image (0 m welded) is the original probe geometry; the middle image shows the probe after a weld distance of 1.52 m; the right image corresponds to the self-optimized shape (attained after 3.96 m of weld) [40].

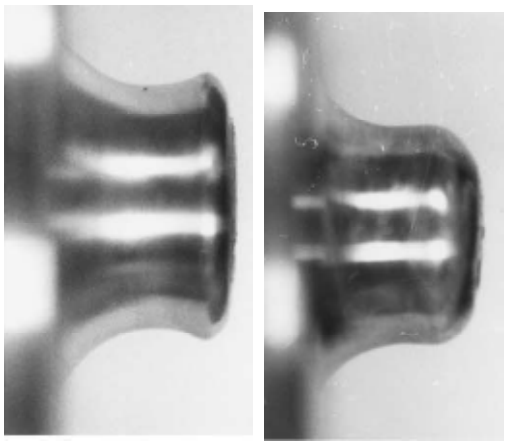


Figure 22 Comparison of images of self-optimized probe shapes taken from Shindo et al. [40]. 1000 RPM/6 mm/s (distance welded 3.1 m) is on the left; 1000 RPM/9 mm/s (distance welded 3.96 m) is on the right.

A research question which naturally arises from the studies presented in references 41 and 43-46 concerns what effect, if any, the variations in tool shape which

accompany wear impose on the quality of the weld. The results of hardness tests of welds in the wearless regime (the plateau regions of select wear curves in Figure 19) indicate a significant increase in the hardness of the HAZ [42]. This rise in strength associated with the use of a self-optimized probe is consistent with a homogeneous weld zone formed in the absence of a wake effect. The gradual disappearance of threads during the course of welding reduces the vertical flow of material around the tool; once the tool has reached a self-optimized shape, the threads have been completely removed and there is little to no circulating flow of material, a result manifested in the hardness measurements of Prado et al. [43]. Additionally, when the probe has reached the optimized condition, the onion structure characteristic of the surface of an FSW joint is less apparent [44]. These changes in the weld attributes can be ascribed to differences in the flowfields for the initially threaded geometry and the self-optimized shape, illustrated in Figure 23 [44]. Although the vortex flow is absent in the self-optimized shape, friction at the tool/workpiece interface generates sufficient heat to produce acceptable welds [42]. Shindo et al. contend that the ability of the self-optimized tool to yield satisfactory weld specimens renders the use of featured tools in FSW of metal unnecessary. Fernanzdez and Murr agree that the inclusion of features is inefficient, as they will eventually be eroded by the abrasive action of the particles in the workpiece. However, since some vortex flow is desirable to reduce the likelihood of void formation, Prado et al. suggest that the ideal tool design for this application is a smooth probe designed with a minimum amount of features to mitigate porosity [42]. This recommendation informs our selection of the Trivex geometry for the wear study presented in chapter V. Prado et al. also advise the use of tools machined to closely resemble the self-optimized shape.

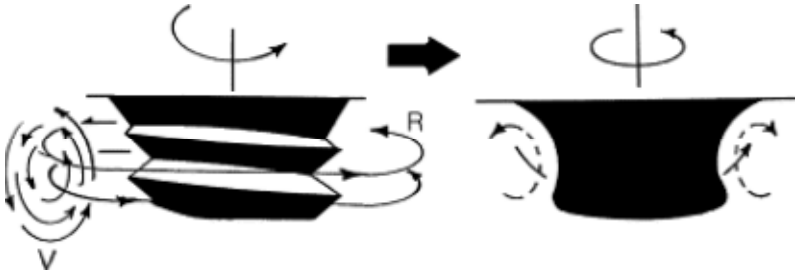


Figure 23 Comparison of flow regimes for threaded probe (left) and probe with self-optimized shape (right). [44]

Stellwag and Lienert contend that the the wear of the tool seems to have no adverse effect on the mechanical properties of the weld [26]. This claim is not corroborated by other research. Feng et al. conclude based on tensile tests that the presence of erosion products in the joint results in deterioration of mechanical properties. It is also possible that the reduced vertical flow associated with the self-optimized shape facilitates the growth of voids that can negatively impact joint strength. When wear of the tool becomes severe enough to compromise weld quality, the tool must be replaced. Replacement may prove a costly venture when the tool geometry is complex and/or there is a large amount of material to be welded. For these reasons, considerable research has focused on reduction of wear in frictions stir welding of metal matrix composites.

2.4.4 Combating Wear in FSW of Metal Composites

As discussed in section 2.4.3, preserving the original tool geometry eliminates costs associated with the manufacture of replacement tools and renders concerns about

reduction in tensile strength related to porosity obsolete. Assuming that the mechanism which underlies the wear in this process is abrasive, the most obvious approach to combating it, and thus preserving the original probe shape, is to select a tool material harder than the reinforcement particles which cause abrasion (this plan of attack is founded upon the principle of abrasive wear, discussed in chapter III). Polycrystalline Boron Nitride (PCBN), Boron Carbide (B_4C), and diamond are suitable tool materials for this application, having hardness values greater than the ceramics (Al_2O_3 and SiC) frequently used as the reinforcing phase in metal composites.

The selection of a wear-resistant tool material is fraught with challenges. While materials such as PCBN, WC, and diamond are harder than the reinforcing particles (and thus should guard against wear), they are both expensive and difficult to machine. Many of these materials are brittle and would thus fracture under the high forces associated with FSW [33]. In an effort to improve ductility, a brittle material can be alloyed with a less brittle material, resulting in a new material with mechanical properties that are more suitable for FSW applications. Two common alloys of this type are WC-Co and Mo-Rh; the addition of the comparatively more ductile Cobalt and Rhenium limits the propensity for fracture and improves machinability without significantly reducing material hardness.

Because it is expensive to fabricate the entire tool from harder, wear-resistant materials, researchers often rely on coatings of substrates made of materials which are cheaper and easier to manufacture. Lee et al. used a h13 steel tool with a B_4C coating to join Al 6092/SiC/17.5p. Unfortunately, the tool did not demonstrate increased wear resistance, as the B_4C coating delaminates from the h13 substrate after only a short weld distance. Once the coating has worn away, the coarse joint surface characteristic of

decohesion between matrix and reinforcing particles reappears. Welds made using the coated tool show a marginal improvement in mechanical performance; for the same process parameters, the ultimate tensile strength (UTS) associated with the coated tool is 61.9 ksi, a value approximately 13 percent greater than the UTS measured for the uncoated tool [33]. This slight improvement in UTS can be explained in terms of Feng's hypothesis, which argues that the presence of erosion products negatively impacts joint strength [34]. In this scenario, the hard B_4C coating functions as a barrier between the tool and abrasive particles, affording a delay in the onset of wear. This in turn prevents the accumulation of eroded tool material which can reduce the UTS. Lee et al. intimate that although the B_4C coating marginally improves joint strength, its tendency to delaminate under FSW conditions makes it ineffective from a cost perspective. While coatings are commonly employed in other machining processes to mitigate wear, their use in FSW of metal composites is largely unexplored. Lee et al. note that the use of diamond as a tool base material or coating may yield more promising results and suggest that further investigation on this subject is needed.

Another option for reducing wear involves alteration of the workpiece material. Lee et al. investigated replacing conventional metal composites, which have a uniform distribution of reinforcement throughout, with Functionally Gradient Materials (FGM), a type of composite which has a high percentage of reinforcement everywhere except near the jointline [33]. This reduced presence of reinforcement surrounding the weld joint limits the tool wear, while the large concentration of reinforcement present at all other locations in the material preserve the enhanced strength associated with MMCs (Figure 24). The gradient in the reinforcement is created via infiltration casting: a ceramic

perform material (Saffil) comprises the 0.25" edge sections of the plate. Lee et al. were able to successfully join Aluminum 2195 alloy to FGM composites with SiC reinforcement percentages of 5, 18.5, and 27 near the edge of the plates and 50 percent elsewhere.

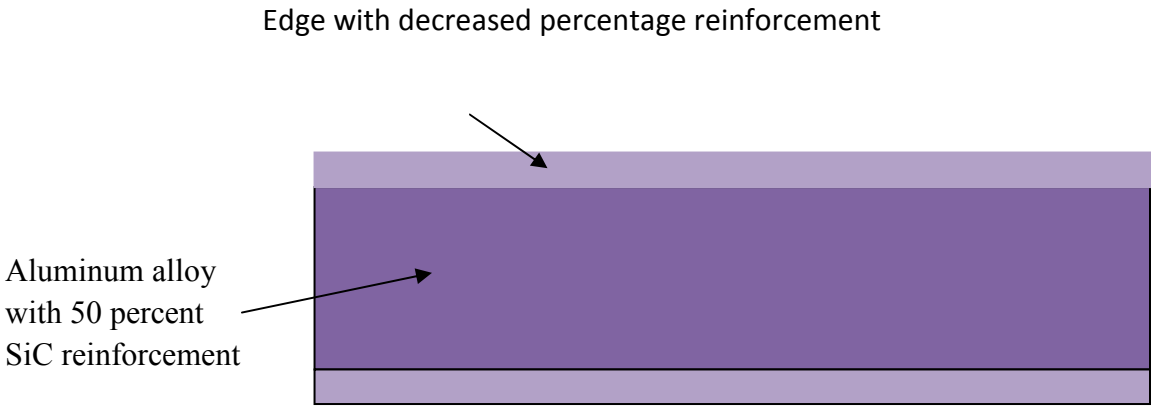


Figure 24 Functionally Gradient Material (FGM) used by Lee et al. [33]

J.A. Wert proposes the joining of metal composites to monolithic alloys as a possible means of wear mitigation. This method is advantageous because it does not require alteration of the composite (i.e. the reinforcement gradient which is created in FGM) and limits the use of the metal composite to regions where increased strength is most critical. There is thus the potential for a threefold reduction in cost: there are lower costs associated with the production of the MMC (as no special modification to the material is required), the amount of material which must be purchased (since the MMC is used in less of the structure), and replacement of tooling (if the joining of MMCs to

monolithic alloys is shown to reduce wear). While the method posited by Wert is attractive for economical reasons, it proves exceedingly difficult to implement owing to the dissimilarity in material properties between the metal composite and the unreinforced alloy. In order to soften the MMC (which has a substantially higher flow stress), the heat input to the weld must be increased. As a result, MMC to monolithic alloy joints are characterized by overheating and the formation of undesirable eutectic phases containing Mn, Si, Cu, and Al [48]. Additionally, the placement of the MMC on the advancing or retreating side of the weld has great effect on the surface morphology. Since stirring harder material into softer material is easier than stirring softer material into harder material, it is recommended that the comparatively harder MMC be placed on the advancing side. Positioning the composite on the retreating side with respect to the tool rotation produces joints with cross-sections characteristic of insufficient mixing (Figure 26). The effect of joining MMCs to conventional Aluminum alloys on tool wear has yet to be investigated; at the time of this writing, there are no studies which demonstrate that FSW of MMCs to Aluminum reduces the rapid consumption of tool material documented in welds of similar composite materials.

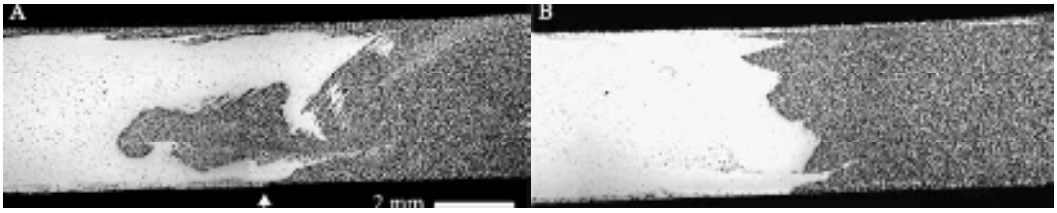


Figure 25 Comparison of transverse cross-sections for welds of dissimilar materials (MMC and Al 2024). In (A), the metal composite is on the advancing side; in (B), the composite is on the retreating side. The morphology in (B) indicates little to no mixing of the materials [48].

CHAPTER III

WEAR

3.1 Significance of Wear

In his seminal text Friction and Wear of Materials, Ernest Rabinowicz defines wear as "the removal of material from solid surfaces as a result of mechanical action" [49]. As discussed in Chapter II, wear is incredibly important from an economic perspective; in the manufacturing industry, cost is in direct proportion to the frequency with which a particular component must be replaced. Additionally, worn machine elements can result in catastrophic failure of a part or entire system This is the case for jet engines, in which the bearings are repeatedly subjected to loads in excess of 15,000 psi; surface fatigue wear (induced by the repeated cycles of loading and unloading) results in a loss of material known as "spalling" [50]. Once spalling is initiated, the bearing fractures due to the rough contact between the bearing and other engine components. Bearing fracture (and subsequent seizure of the engine) severely compromises the performance of multiple-engine aircraft and is devastating for craft with single engines. To complicate matters further, the wear responsible for the potential loss of the aircraft may be so minute that it is indistinguishable using conventional macroscopic inspection techniques. As a result, the aircraft industry devotes substantial resources to advanced non-destructive evaluation tools which can detect spalling at its

onset. A lubricant (usually oil) is circulated beneath the bearings to prevent and/or significantly delay the onset of spalling.

The example of fatigue wear in aircraft engine bearings illustrates why monitoring wear is important as well as the difficulties inherent in its detection. Wear as a phenomenon is underexplained, a deficiency Rabinowicz attributes to the pervading idea among researchers that wear is a chaotic process, so complex that it cannot be robustly characterized or predicted. Rabinowicz contends that in most applications wear is a gradual process typified by a "steady and continuous" loss of material. The results summarized in chapter II show that the amount of tool wear incurred in FSW of metal composites can potentially be predicted from process parameters. The dependence of wear on process parameters ω, ν, and ℓ (equation 2-2) is empirically observed for a wide range of materials and tool geometries [42-44, 51, 52]. Similar studies of tool wear in other machining processes such as turning and drilling have demonstrated that wear can be represented mathematically with some degree of accuracy [53-54]. Studies such as these go far in debunking the myth that predictive process models for wear are unattainable owing to the complexity of the mechanisms at work.

While wear is usually regarded as detrimental to system performance, there are scenarios in which the wear process can be beneficial. Metallurgists rely on abrasive wear to polish specimens: the outer layer of material is eroded by abrasive particles, exposing a smooth surface suitable for imaging and analysis. A pencil eraser is able to remove graphite from a paper surface through adhesive wear [49]. A more unusual instance of favorable wear comes from the animal kingdom, where the incisor teeth of carnivores are continuously sharpened by the contact between meat and bone with exposed dentine during

mastication [49]. In some cases, wear is a useful diagnostic tool, as it can provide a forensic record of the processes at work in the system.

3.2 Types of Wear

In some ways, we can approach the problem of wear as a physician approaches diagnosis and treatment of an illness. Suppose a patient enters a doctor's office complaining of pain in the abdomen. The first task of the physician is to determine the cause of the pain -- he or she usually relies on a series of diagnostic tests (X-Rays, MRI, cell cultures, etc.) to aid in this assessment. Based on the results of these analyses, the physician settles on a probable cause and treats the patient accordingly. The method of treatment hinges on determination of the cause. In a similar fashion, determination of the underlying wear mechanism in systems that exhibit wear is critical to understanding and combating the process. Just as a stomach ache can be symptomatic of a large number of illnesses ranging from mild to severe, wear can be caused by a number of mechanisms, each with its own characteristic indicators. EDX, surface inspection, mechanical gauging, and radiotracer techniques are diagnostic tests the engineer or metallurgist may use to reveal these indicators and pinpoint a particular mechanism as the primary mode of wear. Misclassification of the wear mechanism, akin to misdiagnosis in medicine, can result in a negative outcome. Much as we do not combat adhesive wear in the same manner as abrasive wear, the doctor in our previous example would not prescribe the same medicine for motion sickness as he would for a gastrointestinal infection. Once a mechanism (or mechanisms) has been identified, a course of action can be chosen to

mitigate the amount of wear which occurs during a given process. It is important to note that elimination of wear is a best-case scenario. Depending on the specifics of the wear mechanism, it may not be possible to eradicate wear altogether (just as there are some illnesses which cannot be cured).

Rabinowicz classifies wear into four major categories: surface fatigue, adhesion, abrasion, and corrosion. A system in which wear is observed may display characteristics of any or all of the mechanisms, an interrelatedness which is visually depicted in the Venn diagram of Figure 26. While overlap between wear regimes is common, a dominant mechanism (one that is responsible for the majority of the measured wear) usually emerges during analysis. Thus even though there may be many mechanisms at work, the dominant mechanism is cited as the primary cause of wear. We propose an addendum to Rabinowicz's classification scheme in which wear within any category (or categories), may be further classified according to severity. However, the boundaries delineating light wear from moderate wear and moderate from severe wear are not well-defined; it would thus be left to the judgment of the researcher to make such determinations.

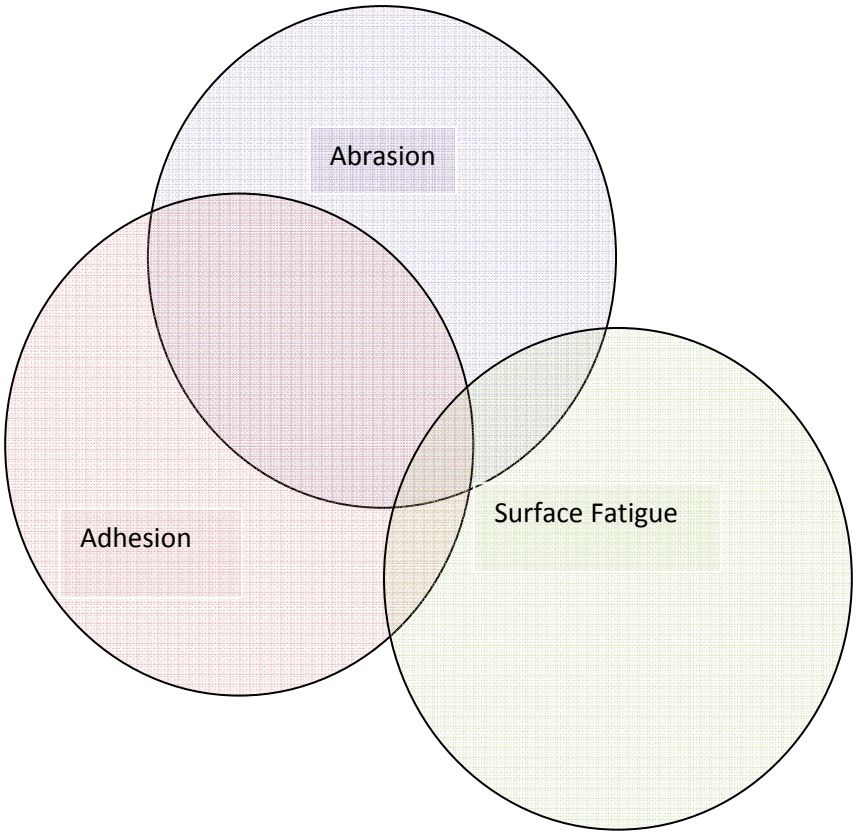


Figure 26 Interrelatedness of three major wear mechanisms. The mechanisms which most frequently overlap are adhesion and abrasion.

3.2.1 Adhesive Wear

Adhesive wear is by far the most common of the wear mechanisms, present in all systems where one surface slides over another. The adhesive wear process is characterized by the preferential adherence of one material to another material with which it is in contact; this adherence is a product of the atomic forces of attraction which exist between dissimilar materials. In adhesive wear, a small region of one surface (referred to as a junction) contacts a junction of similar size and shape on the opposite surface; it is possible that when this contact is broken the break will occur within one of

the materials rather than at the interface, resulting in expulsion of material in the form of a wear fragment. Formally stated, the condition for formation of an adhesive wear particle is as follows:

"If the force required to break through the interface of materials is larger than the force required to break through some continuous surface inside one of the materials, the break will occur along this latter surface, and a transferred wear particle will be produced." [49]

It is most often the case that the shear strength of the interface is much less than the strength of either of the bulk materials in contact; hence no wear fragment is produced. According to Rabinowicz, less than 5 percent of junctions shear in a manner that is conducive to formation of a wear particle. On the relatively rare occasion that a wear fragment emerges, it usually originates from the softer material. However, there are instances where adhesive wear particles exhibit material properties consistent with the harder material, an indication of the presence of local low-strength regions within the harder contacting surface.

The particle size distribution of adhesive wear fragments is typically skewed toward particles with smaller diameters [49]. This tendency to form small particles is explained by Archard's model of adhesive wear, which relates the size of the wear fragments formed to the cube of the junction diameter. This assumption, represented in the diagram of Figure 28, provides the foundation for Archard's equations which describe adhesive wear (detailed information on the development of the Archard model can be found in reference 55). Archard's fundamental law of adhesive wear, equation 3-1,

expresses the volume of wear particles (V) formed as a function of the applied load (L), the sliding distance (x), the flow pressure (p) for the softer material, and the coefficient of wear (k). The coefficient of wear is an empirically determined constant that is specific to the combination of materials in contact.

$$V = \frac{kLx}{3p} \qquad \text{(equation 3-1)}$$

Based on the relationships expressed in equation 3-1, Archard formulates three laws which govern the adhesive wear process:

(1) The total amount of wear varies directly with the applied load.

(2) Wear increasing with increasing sliding distance.

(3) Wear varies inversely with the hardness p of the softer material.

When stated in this qualitative manner, Archard's laws may seem like a statement of the obvious, as the laws agree with our intuition regarding wear processes. In particular, the observation in (3) that harder materials are less susceptible to wear is a concept that is well-understood even by the non-engineer. The simplicity of Archard's laws, however, belie their predictive power. Based only on Archard's fundamental equation, an engineer can "treat" a system that has been diagnosed with adhesive wear (returning to the medical analogy of the previous section) by prescribing shorter sliding distances, decreased contact pressures, or harder materials. Fortunately, adhesion is an infrequent occurrence which produces only small wear particles; it is thus associated with comparatively lower wear rates.

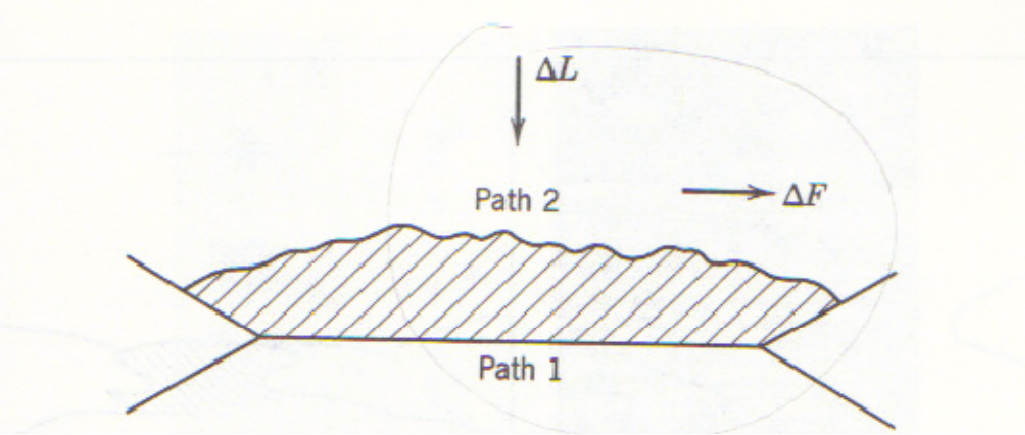


Figure 27 A junction in shear. The shaded region represents a wear fragment produced when the shear strength of the junction exceeds the bulk strength of the softer material. In this situation, shear occurs along path 2 [50].

3.2.2 Abrasive Wear

Systems which exhibit high wear rates are indicative of abrasive wear. Unlike adhesive wear, which is present in virtually all systems in which dissimilar materials are in contact, abrasive wear favors systems in which a harder surface is able to impinge on a softer surface and plow out grooves in the softer material (wear fragments consist of the material expelled from the groove). This type of abrasive wear is termed two-body abrasion. The alternative, three-body abrasion, arises when an abrasive grain comes between the two sliding surfaces. Because the grain is harder than one or both of the contacting materials, it is able to plow out grooves with widths approximately equal to the grain diameter.

Wear rates associated with both types of abrasion can be predicted using the abrasive wear model described by Rabinowicz in reference 49. In this simplified model, the abrasive material (represented by a conical region of width $2r$ and included angle θ)

impinges on the surface under load ΔL and penetrates the surface a distance x, travels along a horizontal path of length dl and removes a volume of material dV. This model is represented geometrically in Figure 28, excerpted from Rabinowicz. The resulting equation, which expresses the volume of material removed per unit distance as a function of load ΔL, cone angle θ, and hardness p, is given by equation 3-2:

$$\frac{dV}{dl} = \frac{\Delta L \, tan\theta}{\pi p} \qquad \text{(equation 3-2)}$$

Integrating equation 3-2 with respect to l results in an expression for the amount of material lost that is similar in form to the Archard equation in 3-1. Thus the qualitative claims made by Archard regarding adhesive wear also hold true for abrasion: that is, wear inversely proportional to the hardness of the softer material and directly proportional to the applied load and distance over which the load is applied. The wear constant for abrasion, k_{abr}, depends on the included angle Some experimentally determined values for k_{abr} are tabulated by Kruschev in reference 56. k_{abr} is much larger for two-body abrasion, a discrepancy Rabinowicz attributes to the tendency of abrasive particles in three-body abrasion to "roll." The ability of the grains to easily slide between the two materials in three-body abrasion impedes the formation of wear particles in the system.

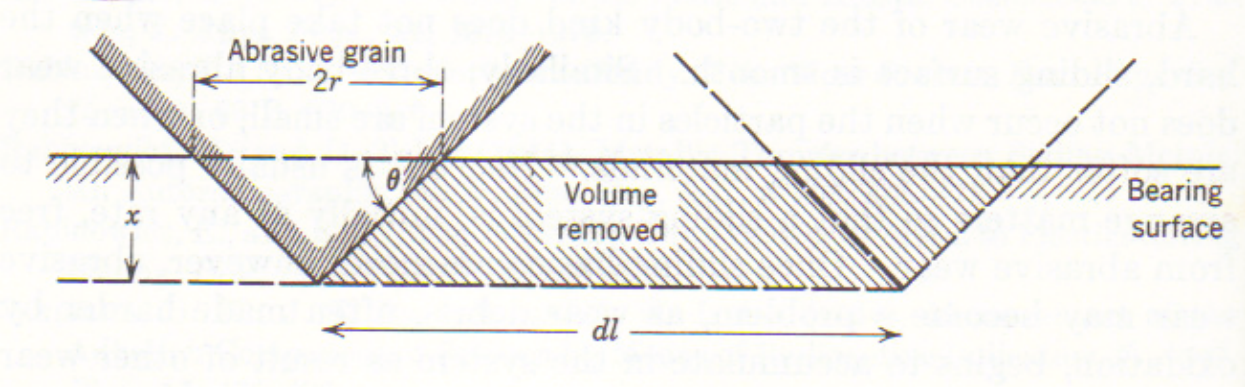


Figure 28. Geometric representation of abrasive wear model [49].

Abrasive wear has been shown to be dependent on other factors not represented in equation 3-2, namely the size of the abrasive grains and the hardness of the abrasive. Experiments in which the diameter of the abrasive is varied while other conditions are unchanged indicate that wear increases with abrasive particle size [49]. There exists a critical abrasive size above which the wear rate is largely insensitive to changes in grain size. For particle sizes below the critical size, the dependence of wear rate on size is more pronounced. This behavior is explained by Rabinowicz in terms of indenter geometries (Figure 28). For smaller particles, particle size influences the shape of the indenter formed an abrasive grain indents into the contacting material. However, indenter geometry is largely invariant for particles with volumes greater than the critical size. While the effect of particle size (and indenter geometry) does not explicity appear in equation 3-2, it may be included in the surface roughness term ($tan\theta$). Smaller particles would result in "smoother" surfaces, as the cone of removed material would have a smaller included angle. The selection of smaller particle sizes can thus help to reduce material loss due to abrasion, although the extent of this reduction depends on the

hardness of both the abrasive and the abraded material. The tracer experiments of chapter VII are an attempt to capture a similar dependence of wear on abrasive grain size in friction stir welding of metal composites. The physics-based model which accompanies these experiments theorizes that there is a critical particle size (D_{cr}) required to score the welding tool at a particular location.

A surprising relationship exists between abrasive wear and the moisture content of the ambient air. A high concentration of water vapor in the air can function as a lubricant in the system, effectively flushing away wear debris. Since the accumulation of wear particles can act as a "buffer" against additional wear, their removal increases the efficiency of the abrasive action [49]. Rabinowicz reports that abrasive wear can increase by as much as 10 to 20 percent under conditions of high humidity. Thus humidity should be carefully controlled in abrasion experiments to ensure that results are reproducible.

The rate of wear in systems undergoing abrasion is typically rapid. However, Mulhearn and Samuels [57] found that the wear rate for some systems exhibit asymptotic behavior: that is, the total volume of material lost (V) approaches a limit (V_∞) as the sliding distance (l) increases with time *t*. The upper limit of material loss is governed by expression 3-3, where V_∞ and β are experimentally determined constants specific to the combination of materials in contact. Samuel and Mulhearn explain this behavior with regard to the accumulation of wear fragments; at some point, the abrasive surface becomes clogged with adherent wear debris, thus impeding the ability of the abrasive grains to impinge on the contacting material. This is an instance in which one form of wear (adhesion) proves beneficial by interfering with the action of another, more severe form of wear (abrasion).

$$V = V_{\infty}(1\text{-}e^{-\beta l}) \qquad \text{(equation 3-3)}$$

The asymptotic dependence of abrasive wear on sliding distance (Figure 29) is similar to the critical distance x_{cr} reported by Prado et al. (in friction stir welding of metal matrix composites, x_{cr} is the length of weldment beyond which little to no additional tool wear occurs). The wear response of the systems studied by Mulhearn and Samuels are characterized by an initial, rapid erosion of material followed by a gradual decrease in the wear rate ($\frac{dV}{dl}$), behavior which is nearly identical to that documented in studies of tool wear in friction stir welding of metal composites. This similarity in response (the decay in wear with increasing distance l) can be viewed as evidence that the wear mechanism in both scenarios (steel sliding on silicon carbide paper and a steel tool joining silicon carbide reinforced aluminum) is abrasion.

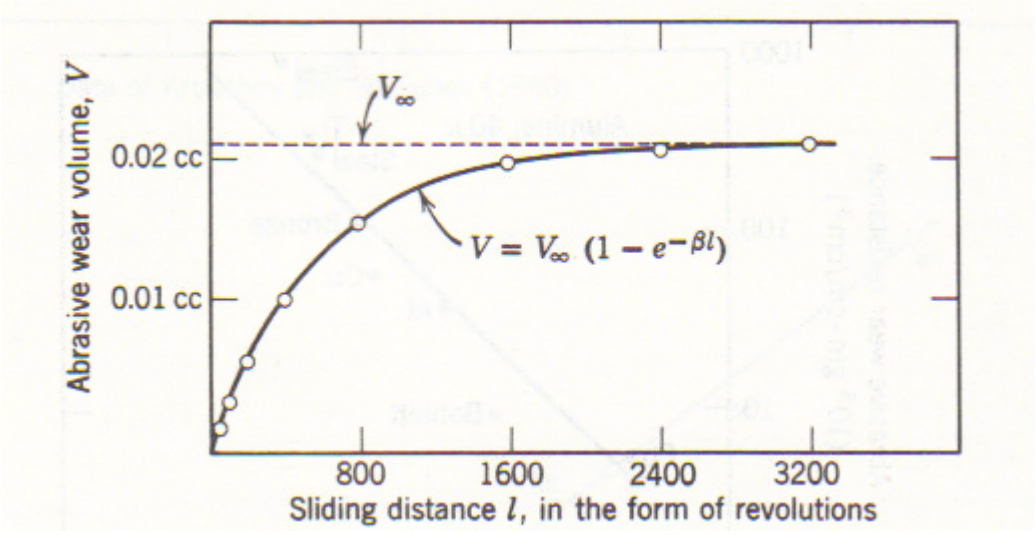


Figure 29 Plot of material removed (V) versus sliding distance (l) for steel sliding on silicon carbide [49].

A noticeable drop in the wear rate occurs when the hardness of the material being abraded (p in equation 3-3) approaches that of the abrasive grains. This simple observation serves as the basis for the principle of abrasion, which states that "if mineral A is harder than mineral B, A scratches (i.e. abrades) material B, but is not scratched in return" [49]. Thus if our wear doctor of section 3.2 were to advise a course of treatment for abrasive wear it would most likely call for the use of a contacting material which is harder than the abrasive. The precedent for using harder materials to combat abrasive wear is found in the work of Aleinikov, who documented a decay in the abrasive wear rate as the hardness of the material being abraded is increased [58].. This decay begins when the hardness of the material being abraded is approximately 70 percent as hard as the abrasive; once the hardness of the material being abraded supercedes the hardness of the abrasive grains, the wear rate drops off precipitously. Further quantification of this relationship suggests that the wear rate is inversely proportional to the square of the ratio R by which the hardness of the materials in contact differ [49]. R is defined for two materials a and b in equation 3-3a, where H_a denotes the hardness of the softer material and H_b is the hardness of the abrasive. The ratio can be used as a metric to evaluate the ability of a particular tool material to resist wear. For example, in a system with silicon carbide abrasive, the ratio R for a h13 steel tool (heat treated to RC-50) is approximately 0.40.. Replacing h13 with tungsten carbide increases R to 0.74 and thus should reduce the wear rate (equation 3-3b) by a factor of three. Unfortunately, selection of materials with higher hardness values is complicated by their propensity to fracture under the high stresses associated with many manufacturing processes. It is often the case that the hardest materials available to combat abrasive wear in a particular process may also be

the most brittle. As such, there are inevitable trade-offs between hardness and ductility which much be dealt with in the design process. In manufacturing systems prone to abrasive wear, care must be taken to minimize stress concentrations in elements fabricated from harder, brittle materials. There is an additional possibility that if the harder material is abraded, it will produce a wear particle that is larger than the diameter of the circular groove ploughed out by the abrasive. Thus while wear particles from harder materials are produced with less frequency, they tend to be larger than wear particles originating from softer material.

$$R = \frac{H_a}{H_b} \qquad \text{(equation 3-3a)}$$

$$Wear\ rate \propto \frac{1}{R^2} \qquad \text{(equation 3-3b)}$$

3.2.3 Surface Fatigue Wear

The third major category of wear, surface fatigue wear, occurs when a component in a system fractures due to stress cycling. .Unlike the other wear mechanisms, surface fatigue wear occurs suddenly. For many cycles, the element may appear to be unaffected by the application and removal of the loading; large fragments then begin to break off the element with little or no warning. The time it takes for this failure to occur is predicted by equation 3-4a, where L denotes the magnitude of the applied loading and c is an experimentally determined constant. For geometries where L can be defined as the cube of the maximum elastic stress (σ_m) , the time to failure t can be rewritten as equation 3-4b (this is the case for contacting materials with hemispherical areas, such as bearings).

The predictive capability of equations 3-4a and 3-4b is somewhat compromised by the experimental scatter which is common in fatigue studies. Values for the life of an element (determined by the results of identical, controlled experiments) may vary by factors as large as 200! S-N graphs, which plot the applied load against the number of cycles to failure, can also exhibit a high degree of variability.

$$t = \frac{c}{L^3} \qquad \text{(equation 3-4a)}$$

$$t = \frac{c}{{\sigma_m}^9} \qquad \text{(equation 3-4b)}$$

When it comes to reducing the likelihood of surface fatigue wear, designers essentially have two options:

1) Choose a more ductile material. Failure in surface fatigue wear is usually initiated by the formation of a subsurface crack. For this reason brittle materials, which are less able to yield under stress, are more succeptible to surface fatigue wear. S-N curves demonstrate that fewer stress cycles are required to induce surface fatigue wear in brittle materials. Fracture is most likely in materials whose strength in tension is less than 30 percent of the compressive strength [50].
2) Reduce the loading. Based on equation 3-4a, a reduction in the load by half increases the life of the component by a factor of 8.

Since the failure of the element in surface fatigue wear is caused by the repeated application of stresses that are usually well below the yield strength of the material, there

is no design modification which can guarantee that the possibility of failure is zero. A typical machine element is subjected to many loading and unloading cycles during its lifetime. While a change in material and/or a load reduction can prolong component life, it is not an assurance that surface fatigue wear will not occur. An example of how surface fatigue wear can severely compromise system performance appears in the chapter on diamond coatings for FSW of metal composites.

3.2.4 Other Types of Wear

There are a few others types of wear separate from the three major wear mechanisms identified in sections 3.2.1-3.2.3. First among these is corrosive wear, which occurs in sliding systems where one or both of the surfaces in contact react with the surrounding environment. Material loss is initiated when the products of the corrosion reaction leave the surface. The formation of a corrosive film inhibits the efficiency of the wear action [49]. Figure 30 compares the amount of corrosion over time for systems in which a corrosive film is formed with the progression of corrosive wear in systems where no film is formed. The film prevents the surface from reacting with its environment, thus guarding against further corrosive action. In the absence of the film, corrosive wear proceeds at a constant rate.

Corrosive wear is quantitatively characterized by equation 3-5. The depth worn of material worn away by corrosion (h) is given by the product of the sliding distance with a system-specific dimensionless constant (k). For most situations, k is assumed to be 10^{-4}

or 10^{-5}. This expression is only applicable to processes in which a film is formed on the surface.

$$h = \frac{kx}{3} \qquad \text{(equation 3-5)}$$

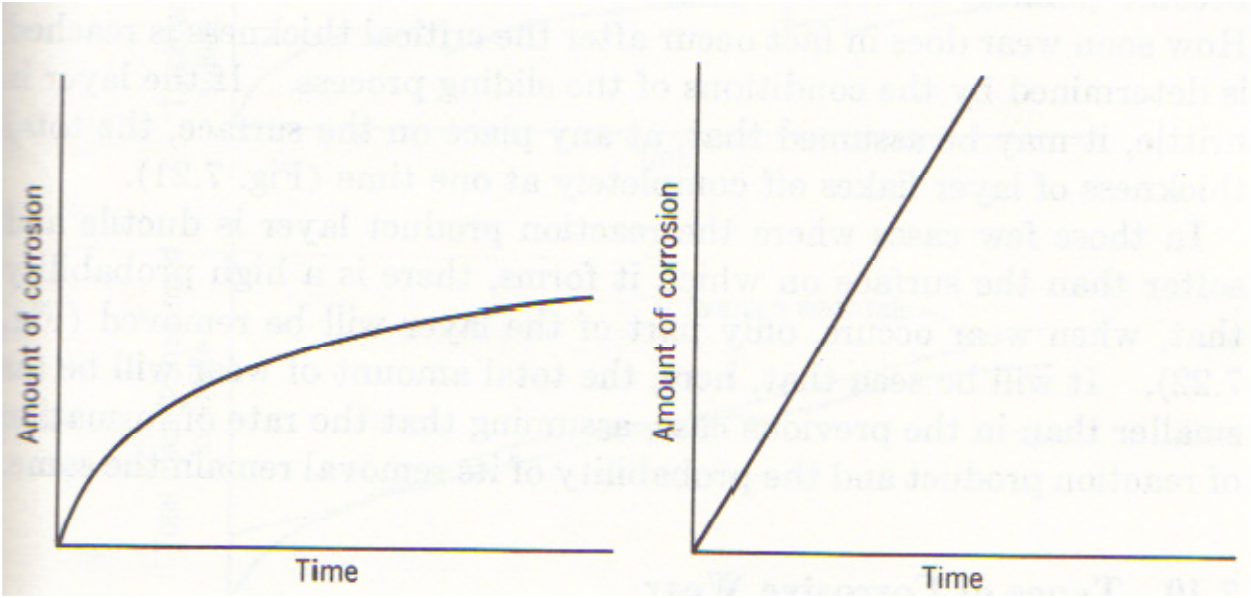


Figure 30 Progress of corrosive wear with time in sliding systems when a corrosive film is present (left image) and when no film is formed (right image).

Fretting is a type of wear which develops when one of the surfaces in contact undergoes oscillation. Although the amplitude of oscillation is typically very small, it results in the production of large wear particles. Oscillatory behavior leads to the formation of adhesive wear fragments which can then corrode to remove additional material from the surface. Since fretting is a combination of multiple wear mechanisms (adhesive, corrosive, and in some instances, abrasive), it cannot be easily classified. The incidence of fretting can be reduced by eliminating slippage at the material interface [49].

This usually involves coating one of the materials with a harder material such as CrN or PCBN. While the introduction of a coating guards against fretting, it increases the system's susceptibility to surface fatigue wear.

3.2.5 Comparison of Wear Mechanisms

Fortunately for researchers who seek to identify the wear mechanism in a particular system, each of the four major wear processes identified by Rabinowicz follows a set of unique laws. Hence all that is required to pinpoint a mechanism as the source of wear is to determine which wear process its characteristics are most consistent with. These characteristics are summarized in Table 2, which compares the governing equations, fragment size, microscopic appearance, and rate of wear for the wear mechanisms considered thus far. The qualitative assessments which appear in some cells of the table are defined relative to other wear mechanisms. For instance, the fragments produced by abrasive and surface fatigue wear are characterized as "large"; however, they are only large when compared with the fragments from corrosion and adhesion, which produce wear particles on the order of 30 μm.

Table 2 Comparison of Wear Mechanisms

Wear Mechanism	Governing Equation	Fragment Size	Microscopic Appearance	Rate of Wear
Adhesive*	$V = \frac{kLx}{3p}$	Small; usually semi-ellipsoidal in shape; may be in loose form	Fragments of material from surface A visible on surface B	Gradual
Abrasive°	$\frac{dV}{dl} = \frac{\Delta L \, tan\theta}{\pi p}$	Large	Grooves ploughed out by abrasive material	Rapid; asymptotic
Corrosion[a]	$h = \frac{kx}{3}$	Small	Worn surface is smooth and rounded	Gradual; linear if no film is formed, otherwise asymptotic
Surface Fatigue+	$t = \frac{constant}{L^3}$	Very large (on the order of 1000 μ)	Pitting of surface	Time prior to onset of wear is highly variable; once wear is initiated (via cracking in the surface or subsurface), progression to failure is rapid

* V = volume of wear particles/amount of material removed, k = proportionality constant specific to the materials in contact, x = sliding distance, p = hardness of softer material, L = applied load

° V=amount of material removed, ΔL = applied load, p = hardness of softer material, θ = included angle (in this model, abrasive material is represented by a conical region of width $2r$ and included angle θ), l = length of contact between abrasive grain and surface of opposing material

[a] h = depth of material worn away, x = sliding distance, k = proportionality constant (values on the order of 10^{-5} are typical)

+ t = time to failure, L = applied load

3.3 Measuring Wear

The measurement of wear is often specific to both the wear mechanism acting on the system and the experimental conditions used to study it. Adachi and Hutchings emphasize that because wear is not an intrinsic property of a system, it can vary greatly with test conditions [60]. Some of this variation is desirable, as it allows researchers to extract the effect of a particular process parameter on the amount of wear. Undesirable variations which originate from variables which may not fall under the direct control of the researcher (such as humidity) must be minimized to make certain that the effects induced by the experiments are both "predictable and reproducible" [60]. The detection and measurement of wear requires carefully designed and controlled experiments to ensure that the change in the system response is a result of the variations imposed on the system by the researcher rather than uncontrolled factors.

3.3.1 Microscopy

Surface inspection is the best method to readily identify the wear mechanism. An examination of the worn surface will usually reveal the dominant wear mechanism at work in a system. As summarized in Table , each wear mechanism has a distinct microscopic signature. For instance, abrasion manifests itself as a series of parallel grooves on the worn surface. This grooving is clearly visible in Figure 31, which shows the surface of a steel bearing after sliding contact with Silicon Carbide slurry (volume fraction 0.01). The parallelism is characteristic of two-body abrasion; three-body abrasion produces grooves with no discernable directionality [59]. The width of the

scratches is an indication of the size of the junction, while the number of scratches can be assumed to roughly correspond to the number of junctions [49]. An analog of this optical method is used in Chapter VII to study the abrasive wear process in friction stir welding of metal composites. Features of the tool surface are analyzed using optical microscopy; these measurements are then used to evaluate the predictions of a physics-based model for wear in FSW of MMCs.

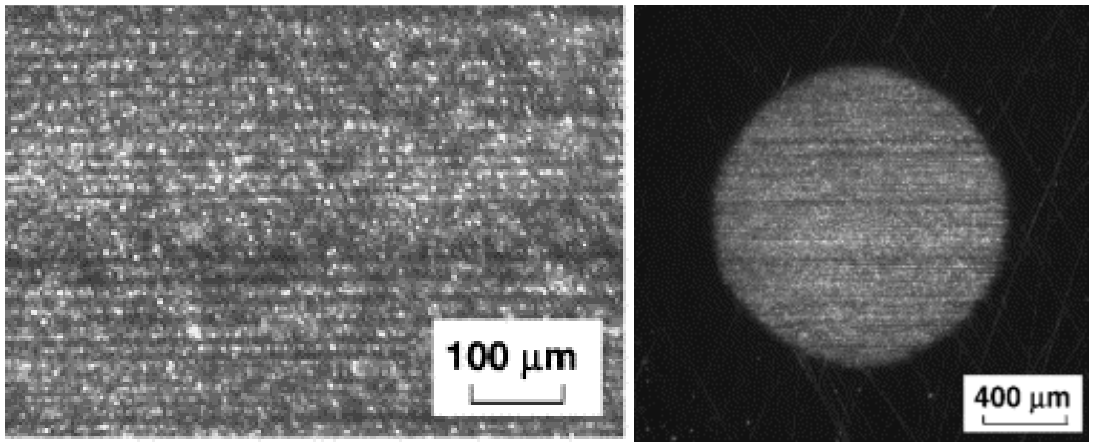


Figure 31 SEM image of worn surface of steel bearing. Bearing was in sliding contact with Silicon Carbide slurry (volume fraction 0.01). The normal load for this experiment is 0.2 N [59].

There are instances in which post-process examination of the worn material can be misleading. The most common misclassification occurs when adhesive wear is mistaken for abrasion [50]. This is the case for copper sliding on steel, where the scratches on the worn surface point to abrasive action. However, the root cause of the wear is not abrasion, as the transfer of the hard particles which produce the scoring on the worn surface is facilitated by adhesion. To prevent incorrect classification of the wear

process, it is best to use microscopy in conjunction with other methods. Additional methods of characterizing wear include mechanical gauging, weighing, and the use of radiotracers.

3.3.2 Weighing

The simplest method of detecting wear is to compare the weights of the materials in contact before and after sliding. As material is lost due to wear, the weight of the worn component is reduced; the reduction in weight is indicative of the amount of volumetric wear which has occurred. The percent wear is given by equation 3-6a, where m is the mass of the unworn specimen and m' is the mass of the specimen after sliding. The mass of material removed can be converted to volume V by multiplying the change in mass by the density of the worn material ρ (equation 3-6b). The resolution of this measurement technique hinges on the equipment used to measure the changes in weight. A typical mass balance has a resolution of at least 10^{-3} g, making this method better suited for wear processes that remove large amounts of material, such as abrasion. The weighing method is problematic for surfaces on which material accumulates during sliding. This is the case in FSW of metal composites, where the aluminum from the metal matrix collects on the tool during welding. It is critical to clean the surface thoroughly before weighing to remove the additional mass associated with residues or films.

$$Percent\ wear = \frac{(m-m')}{m} x100 \qquad \text{(equation 3-6a)}$$

$$V = (m - m')\rho \qquad \text{(equation 3-6b)}$$

3.3.3. Mechanical Gauging

Wear measurements can also be obtained through mechanical gauging methods, in which a micrometer is used to measure the dimensions of the specimen before and after sliding. The change in the dimensions as a consequence of the wear process is taken as representative of the amount of material loss. While mechanical gaging is an easily implemented method of wear detection, the accuracy of the method is severely limited by the resolution of the micrometer used to make the measurements. The resolution of a typical micrometer is 10^{-2} mm; wear measurements of specimens with initial dimensions less than the device resolution are thus associated with a large degree of uncertainty. As such, mechanical gauging is best suited for larger specimens with regular geometries. Wear of irregular geometries is difficult to quantify with a micrometer owing to the many measurements which must be made to accurately characterize the profile. Profilometry is a more sensitive type of mechanical gauging and the best alternative for specimens which have substantial variations in shape and/or are too small to be characterized by micrometry. A contact profilometer uses a diamond-tipped stylus to construct a contour plot of a specimen surface. The stylus is lowered vertically until it contacts the surface and translated in the x-y plane, recording the variations in the height of the surface as it is moved across. This height data is then used to produce an image such as that in Figure 32, where the height of the surface is indicated by a range of colors [60]. Recent advances in optical profilometry allow for the use of laser-scanning techniques to construct three-dimensional profiles without contacting the specimen.

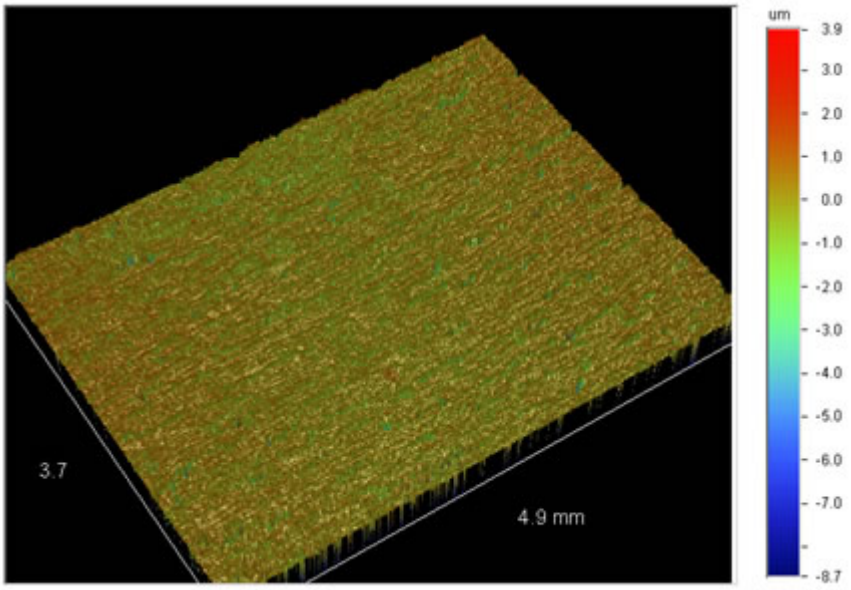


Figure 32 Three-dimensional surface plot of Aluminum alloy 7075 after exposure to 0.1M NaCl solution. Image from Fontana Corrosion Center [60].

3.3.4 Radiotracers

There are some scenarios where the amount of wear is so minute that it cannot be detected by mechanical gaging or weighing. The resolution can be increased to a sufficient level by incorporating radioactive tracer elements into the wear experiments. An additional advantage of radiotracer wear measurement is that post-experiment X-rays of the specimens can provide insight into how and where the wear occurs [49]. There are a number of variations on the radiotracer method for studying wear in a sliding system:

1. Detection of transferred radioactive material. In this setup, one of the surfaces in contact is made radioactive. Thus any material that is transferred from this surface can be detected on the other surface using a Geiger counter or X-ray. The mass of transferred material is equal to the material lost by the opposing surface.

2. Circulation of nonadherent wear debris. A lubricant flows between two contacting surfaces, one of which is made radioactive. If the radioactive surface wears, debris will accumulate in the lubricant. The amount of wear particles captured in the lubricant can be detected using a Geiger counter. This method is most suitable for systems with nonadherent wear debris, such as abrasion. However, as discussed in section 3.2.2, the introduction of lubricant exacerbates the wear mechanism in systems exhibiting abrasive wear (flushing away of the wear debris exposes a fresh surface and improves the efficiency of the abrasion). Thus radiotracers should only be applied in this manner to study abrasive wear in systems where circulating lubricant is already present (such as in engines or pumps) [49].

3. Selective radioactivity. Due to the size of the surface or the geometry, it may not be feasible to make an entire specimen radioactive. In this instance, only a portion of a specimen is exposed to radiation (in most cases, the radioactive portion coincides with what is believed to be the region of highest wear). Abowd used selective radioactivity to study the variation of wear with location in a piston ring [61]. By using different radiotracer particles for different regions, Abowd was able to identify the regions in which wear was the most severe. If the wear is localized, then the use of harder materials to combat wear can be confined to the areas where it is most crucial. Exploratory studies using selective radioactivity are thus be a cost-effective means of wear diagnosis. The tracing properties of the radioactive isotopes also provides a window into the kinematics of the process and is a useful tool for identifying the wear mechanisms at work in a system.

While radiotracer techniques are most frequently employed in studies of abrasive wear, they can be used as effectively to examine adhesive or corrosive wear processes. Radiotracers are particularly useful in measuring the very small particles and gradual wear associated with adhesion; the smallest change in mass that can be detected using this technique is 10^{-12} g, a resolution that is much greater than that of other wear measurement methods. Due to the increased cost and preparation associated with radiotracer experiments this procedure is best suited to systems in which the amount of wear is too minute to be detected by other methods, investigations concerned with identifying regions of high-wear, and/or exploratory studies which seek to identify the wear mechanism,

3.3.5 Comparison of Techniques to Measure Wear

Table 3 briefly summarizes and compares the techniques considered in sections 3.3.1-3.3.4. The advantages and disadvantages of each method are identified as well as which scenarios the detection technique is best-suited for. The resolutions specified in the table are order of magnitude estimates; the actual resolution of a particular technique is largely determined by the equipment used to measure material loss (for instance, some profilometry equipment can measure changes in depth as small as tens of nanometers). The resolution values presented are approximations intended for the purpose of comparison.

As we will see in section 3.4, these techniques are not mutually exclusive. The wear study presented in chapter VI uses a modified radiotracer technique in conjunction

with microscopy and profilometry to assess the influence of particle size, location, and process parameters on the mechanics of wear in FSW of metal composites. In a related study of wear for this process, Fernandez and Murr use a combination of weighing and optical comparator methods [44]. Wear studies are generally enhanced by incorporating multiple methods to gauge wear. It is often the case that a simpler, cruder method of wear detection (for example, weighing or mechanical gauging) is paired with a more sensitive technique (such as radiotracing or microscopy).

Table 3 Comparison of Wear Measurement Techniques

Measurement Technique	Resolution	Advantages	Disadvantages
Microscopy (SEM)	$10^{-6}\ m$	Wear mechanism can be identified by shape and orientation of surface markings; can be used for any wear process which results in features distinct from the original surface appearance	Best for qualitative analysis; depth of field may not be sufficient to measure small changes in height of specimen surface; sample chamber limits specimen size
Weighing	$10^{-3}\ g$	Equipment is relatively inexpensive compared to other techniques; best for large specimens which experience high rates of wear	Crude; low resolution; surface must be thoroughly cleaned before weighing
Mechanical Gauging	$10^{-5}\ m$	Best for large samples	May not be able to account for small variations in shape; difficult to use with irregular geometries
Profilometry	$10^{-7}\ m$	Best for irregularly shaped and/or small specimens	May not be applicable to large specimens
Radiotracers	$10^{-12}\ g$	Very high resolution; best for samples which experience gradual wear; can provide insight into dynamics of wear process	Samples require additional preparation; technical expertise and specialized equipment is needed for analysis

3.4 Identifying and Measuring Wear in Machining of Metal Matrix Composites

There exists only a modicum of published papers which concern wear in friction stir welding of metal composites. Previous work in this area is summarized in chapter II, with emphasis on the studies which appear in references 40 and 42-44. Because so little has been written about wear in this process, researchers must look to tool wear studies involving metal composites in the field of machining. The abundance of work in this area can provide insight into identifying and measuring wear in FSW of MMCs.

Most of the papers on FSW of metal composites identify abrasion as the primary wear mechanism. Though the assumption of abrasive wear is a reasonable one owing to the presence of Silicon Carbide particles which are much harder than the tool material (SiC is approximately 1.5 times harder than h13 tool steel), the conclusion of abrasive wear is premature. The authors of references 40 and 42-44 do not perform any tribological assessments of the tool surface to verify the assumption of abrasion or rule out other forms of wear (notably adhesion) which may act on the tool. In papers on machining of metal composites using steel tools, care is taken to identify the wear mechanism prior to studying the effect of process parameters on the amount of wear. Since each of the major wear mechanisms has a unique microscopic signature (Table), the acting mechanism can be easily identified by examining the tool surface under a microscope. Sahin and Sur inspected the worn surface of tools used in cutting Metal Matrix Composites under an optical microscope; a representative image from this work appears in Figure 33. The grooving across the flank of the cutting tool is parallel to the direction of cutting, an indication of two-body abrasion [62]. Though the experimental

conditions and tool geometries differ, nearly identical micrographs were produced by Kishaway et al. in their study on cutting tools in machining of MMCs [63]. If abrasive wear (and specifically two-body abrasive wear) occurs in friction stir welding of metal composites as postulated in 40 and 42-44, microscopic evaluation of worn FSW tool surfaces should produce images similar to those of Sahin and Sur. While Davim and Baptista find some evidence of adhesion in machining of MMCs (SEM analysis indicates that a small amount of composite material is deposited on the flank of the tool during cutting), they conclude that two-body abrasion is the dominant wear mechanism [64].

Three-body abrasion is evidenced by indentations on the tool surface (Figure 34). The steel surface in this image was subjected to micro-abrasive wear testing, wherein abrasive slurry is deposited between a steel plate and a rolling ball of the same material which contacts the surface under a specified load [59]. In contrast to the grooving in Figure 33, the notches in Figure 34 have no directionality. This is not, however, to say that three-body abrasion cannot occur in cutting of MMCs. While Kishaway et al. observed two-body abrasion patterns along the flank of the cutting tool nearly identical to those seen by Sahin and Sur, they discovered that three-body abrasion dominates in the presence of interfacial cracking. Cracking occurs when the abrasive particles debond from the matrix and are free to "roll" between the surfaces in contact, a change in system dynamics which constitutes the scenario simulated by Adachi and Hutchings in their micro-abrasion experiments. Because three-body abrasion in cutting of MMCs occurs only in this special case, two-body abrasion is regarded as the dominant mechanism for this process. The prevalence of two-body abrasion in machining of MMCs indicates that the abrasive particles are able to adhere temporarily to the surface which exhibits wear

(plowing out a groove in the process) and/or becomes provisionally embedded in it [49]. Correct identification of the wear mechanism is critical to subsequent modeling of the phenomena.

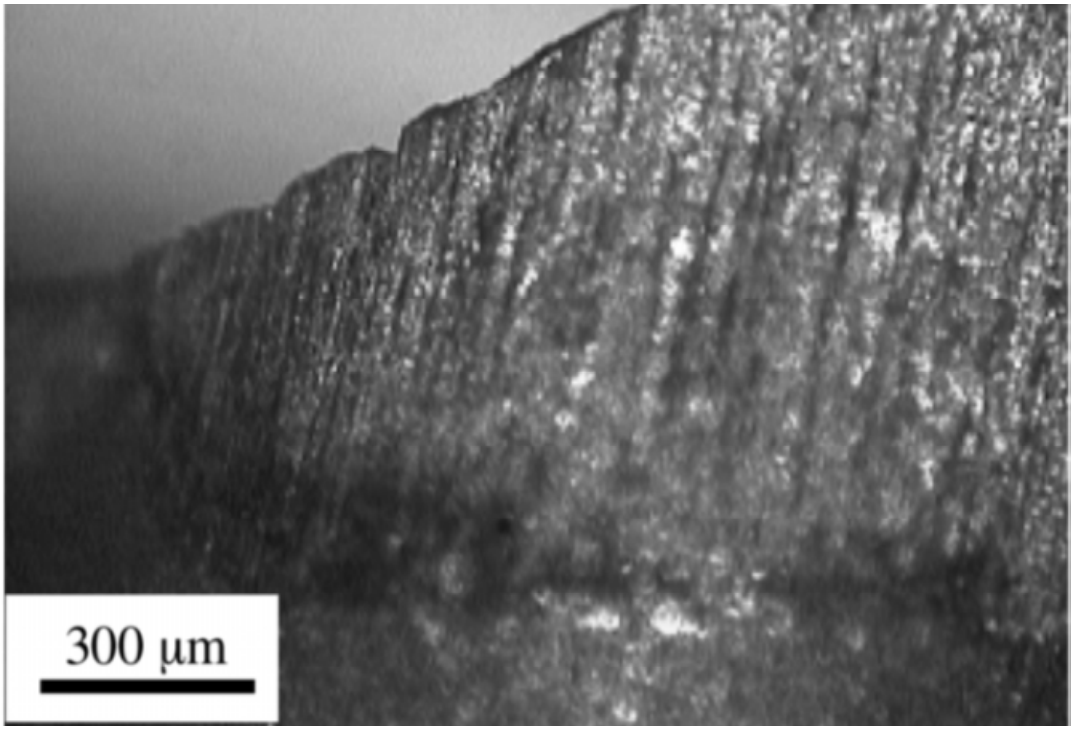


Figure 33 Optical micrograph of tool used in cutting of an Aluminum alloy with 10 percent Silicon Carbide reinforcement. Parallel grooving is synonymous with two-body abrasion [62].

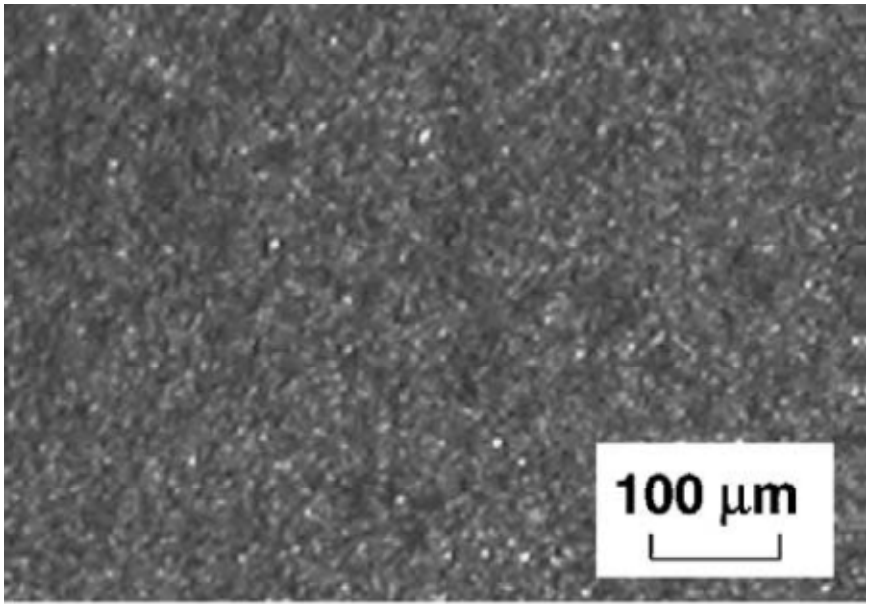


Figure 34 Three-body abrasive wear on steel surface abraded with ceramic slurry [59].

The studies of wear mechanics presented by Davim, Adachi/Hutchings, Kishaway et al., and Sahin and Sur rely on post-test examination of the tool surface to discern the wear mechanism [53-54, 59, 62-64]. As an alternative to this method of analysis, Quigley et al. devise a "quick-stop" apparatus for cutting which enables them to obtain specimens in which the chip is intact (i.e. has not yet been sheared from the surface by the cutting tool). Details of device construction and functionality can be found in reference 65. These *in media res* samples, which capture the process of chip formation in cutting of MMCs, are then examined in an SEM (Figure 35). This composite chip exhibits feature distinct from those encountered in chips formed from conventional aluminum alloys. The most prominent of these is the extensive cracking along the chip-workpiece interface, an effect which is exacerbated when a propagating crack contacts an abrasive reinforcement particle. Other SEM images show abrasive particles that have been pulled from the surface of the MMC and are held between the workpiece and the

cutting tool (the three-body abrasive action documented by Kishaway et al. and precipitated by interfacial cracking). The quick-stop technique shifts the focus from the tool to the sample undergoing machining and is one precedent for the "emergency stop" procedure employed in the tracer study presented in chapter VII.

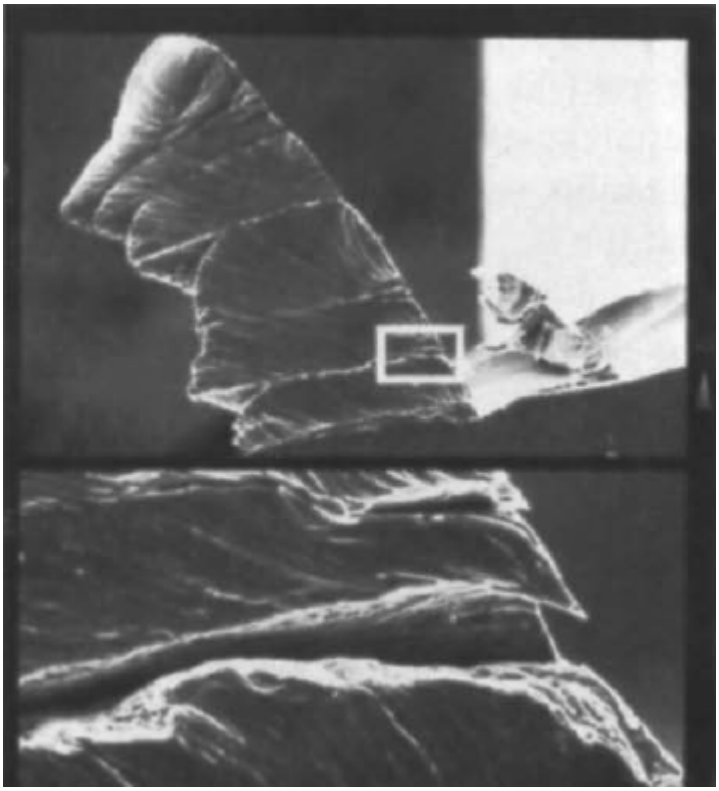


Figure 35 SEM image of chip formation in cutting of Al/SiC composite [66].

Once the wear mechanism has been identified, the question becomes how to quantify the amount of wear which occurs under specific conditions. Davim tracks the progression of tool wear in turning of metal composites using a Mitutoyo optical microscope [53]. Under a magnification of 30X, the dimensions of the tool can be measured to within 1 μm. Sahin and Sur also use an optical microscope to quantify wear,

comparing the dimensions of the cutting tool prior to and after testing [62]. Kishaway et al. regard the width of the tool flank (denoted by VB) as a critical dimension for gauging wear incurred in cutting metal composites [64]. For drilling of metal composites, tool wear is expressed in terms of VB_{max}, the average of the maximum wear observed at the edges [54]. Width of the flank (VB for cutting and VB_{max} in drilling) is the industry standard for measuring wear in machine tools, where wear of the flank is accepted as a robust indicator of volumetric wear. Marinov amends this procedure to measure wear in terms of the cross-sectional area lost by the cutting tool [66]. The area of the tool is computed from a microscope image using a graphical integration.technique. This versatile method of gauging wear can be applied to machine components with geometries disparate from those commonly associated with cutting, turning, or drilling tools. Marinov's process of quantifying tool wear using cross-sectional area is used in the wear study of FSW tools presented in chapter V.

One method that is noticeably absent from recent work on tool wear in machining MMCs is that of radiotracer techniques. This can be attributed to several factors:

1) The wear mechanism encountered in the machining of MMCs is predominantly abrasion. Abrasion can be diagnosed using a variety of other methods, such as microscopy, which are easier to implement than techniques which rely on radioactivity.
2) The rates of wear which accompany abrasion are large enough to render the increased sensitivity associated with radiotracer techniques unnecessary.
3) The logistics and additional preparation required to incorporate radiotracing into wear studies may limit its application. For mechanisms such as abrasion,

where wear occurs rapidly and leaves behind distinct tool markings visible under moderate magnification, it is often the case that little additional information may be gained by the use of radiotracers. However, the usefulness of radiotracers in abrasion should not be discounted: if the focus of a study is to trace the path of wear particles and/or the abrasive rather than simply measure the effects of wear, radiotracers (or a derivative technique) present the best option.

The methods commonly used to pinpoint the mechanism responsible for wear and measure the resultant material loss are not specific to machining (i.e. cutting, turning, and drilling). Chapters V through IX demonstrate that these existing methods, with some modifications, can be applied to the study of wear in FSW of metal composites. Once the dominant wear mechanism has been identified and a robust method of wear measurement established, variables which influence the wear process can be investigated. Experimental designs and statistical techniques which can be used to characterize the variation of wear with process parameters are the focus of the next chapter.

CHAPTER IV

VARIATION OF TOOL WEAR WITH PROCESS PARAMETERS AND MATERIAL PROPERTIES: STATISTICAL METHODOLOGIES AND RESULTS FROM MACHINING

Once a wear mechanism has been identified and a robust method of wear measurement established, researchers can proceed with characterization of the wear phenomena. This typically entails a series of experiments designed to evaluate how the amount of wear is impacted by variations in the process parameters. These parameters, referred to in statistical nomenclature as factors, are variables which may influence the outcome variable (in this case, the wear incurred by the component). The number and type of factors are specific to the machining or manufacturing process being studied. For instance, tool wear in cutting of metal composites may be influenced by any or all of the following: spindle speed, feed rate, depth of cut, reinforcement particle size, abundance of reinforcement, particle hardness, cutting tool geometry, tool material, and temperature of the tool during cutting. To characterize the dependence of wear on these factors, we might perform a series of tests in which one parameter is varied while the others are held constant. Varying one factor at a time creates "stacks" of data from which researchers can extract the effect of changing a particular process parameter on the outcome variable. For the nine factors in the cutting example, there would be one set of experiments where spindle speed is varied while the other eight factors are fixed, another set of experiments where feed rate is varied, and so on until variation in all of the factors has been accounted for. A crude (albeit often effective) experimental design, the results generated by these tests can be plotted to elucidate the relationship between a particular factor and the

outcome variable. This technique appears in the wear studies of Prado et al. and Fernandez/Murr detailed in Chapter II [42-44]. A cursory review of the data plotted in Figure 18, for instance, reveals that for a constant weld distance and traverse speed, wear varies directly with the tool's rotation speed. While there are important observations which can be gleaned from this style of experimentation, it does not account for possible interactions among the factors. Returning to the example of tool wear in cutting, the temperature of the cutting tool is believed to have some impact on wear. However, this temperature is a latent variable, a factor which itself depends on other factors. According to El-Wardany et al., the temperature profile of the cutting tool is a function of feed rate, depth of cut, geometry (width of tool and nose radius), and the rake angle [67]. In addition to neglecting the possibility of factor interdependency, the experimental methods employed in references and are difficult to implement when the outcome variable (wear) is believed to be influenced by more than three factors. In these situations, the number of experiments grows rapidly – in the case of cutting composites, at least nine "stacks" of data are needed to completely characterize the wear process (although we have not given these data stacks a formal designation, the experiments used to generate them most closely resemble factorial design). This first part of this chapter summarizes techniques which can be used to generate and analyze wear data. We begin by examining these methodologies and progress to consider the results obtained using these methods.

An efficient experimental design is one which "derives the required information at the least expenditure of resources" [68]. By using design, we avoid the danger of simply accumulating data in a haphazard fashion. Experimental design takes us away from arbitrary variation of parameters and directs us toward a robust course of

experiments whose results can be analyzed using established statistical methods. Designs commonly employed in wear studies include factorial, Taguchi, and Central Composites Design (CCD). Each has its own advantages; the choice of one particular method over another is generally determined by the number of factors, whether relationships may exist between the factors, the mathematical techniques which will be used to analyze the data generated, and/or the ultimate objective of the study.

Before proceeding with a discussion of experimental design, it is necessary to define some basic terminology used in the field. The *outcome variable*, generally denoted by Y, is the characteristic of a system that researchers wish to study and in most cases predict, A *factor* is any variable that is hypothesized to influence the outcome variable. For instance, if the objective of a study is to predict an individual's likelihood of developing heart disease, the factors in the experimental matrix would consist of variables which are believed to influence heart health. These factors can be either quantitative (such as age, body mass index, waist-to-hip ratio, calories consumed per day, cholesterol levels) or qualitative (for instance, gender, family history of heart disease, genetic markers). Factors which fall into the latter category can be converted to a numeric metric: in the case of gender, 0 can be chosen to represent men and 1 to represent women. 0 and 1 are referred to as levels for the factor "gender": a *level* is simply the value that a factor can assume. Levels may be binary, as in the case of gender, or they may encompass a wide range of values. In some studies, it may be possible to assign levels to each factor. In other cases, such as the heart disease example, the levels associated with each factor are determined by the subjects in the study (in general, researchers do not have the freedom to choose levels when the population is selected

randomly). Together, the factors and levels comprise an experimental matrix; data from this matrix can be represented in the form of a statistical model. An *additive statistical model* proposes that the data points of the outcome variable Y can be expressed as a sum of the factors x_i where i ranges from 1 up to the number of factors n. The mathematics which go into the construction of such a model are discussed in section 4.4.

4.1 Factorial Designs

The simplest method of experimental design, known as factorial design, considers all possible combinations of levels and factors. The best way to illustrate the construction of a factorial design is through an example. Suppose an experiment has three factors (*a, b, c*) and that each of these factors has two levels (*0, 1*). The first experiment in the matrix is denoted by $a_0\, b_0\, c_0$, where the subscript *0* indicates that each factor is at level 0. Similarly, $a_1 b_0\, c_0$ means that factor a is at level *1*, while factors *b* and *c* remain at level *0*. Thus the first factor can be at one of two levels, the second factor can be at one of two levels, and the third level can also assume one of two levels. Using simple combinatorics, the number of experiments needed is thus 2 x 2 x 2 = 2^3. The eight experiments correspond to the eight combinations of factors and levels:
$a_0\, b_0\, c_0$, $a_0\, b_0\, c_1$, $a_0\, b_1\, c_0$, $a_0\, b_1\, c_1$, $a_1\, b_0\, c_0$, $a_1\, b_0\, c_1$, $a_1\, b_1\, c_0$, $a_1\, b_1\, c_1$. In general, the number of experiments in a factorial design is given by the expression m^n, where *n* is the number of factors and m is the number of levels.

The factorial design is ideal for studies in which the factors can assume two levels. Since the number of experiments in this instance is a function of 2^n, the

inclusion of each additional factor *n* doubles the number of experiments in the design matrix. Provided the number of levels remains fixed at 2, researchers can expand the design to study more factors while keeping the number of experiments within a manageable range. An increase in the number of levels m has a more dramatic effect on the number of experiments. For $n = 4$ factors at $m = 2$ levels, 16 experiments are needed. Doubling the number of levels to $m = 4$ increases the number of experiments to 256 (a 16-fold increase)! It is apparent that running every combination of factors and levels is likely impractical for studies with more than 4 factors at 3 or more levels. When the number of experiments in a factorial design becomes so large that experimental efficiency is compromised, we are compelled to consider alternative experimental designs which reduce the number of tests required.

4.2 Taguchi Method and Factor Interactions

The Taguchi method of experimental design was developed by Dr. Genichi Taguchi, a Japanese engineer, in the mid-twentieth century [69]. Although Taguchi's techniques were designed specifically for use in manufacturing, their application has grown to include economics and the social sciences. More than simply a technique for constructing design matrices, Taguchi represents a paradigm shift in how system responses are classified. Taguchi methodologies avoid binary classifications schemes, in which the value of a response variable is characterized as either "acceptable" or "unacceptable." Prior to the advent of the Taguchi system, manufacturing engineers would identify a target value or range for a response variable. For instance, consider the

conditional statement that "a weld is only acceptable if its ultimate tensile strength (UTS) is greater than or equal to 72% of the strength of the parent material." The target UTS of the joint (σ_{UTS}) in this example is defined by the inequality $\sigma_{UTS} \geq 0.72\sigma'_{UTS}$, where σ'_{UTS} is the UTS of the parent material. Any response which has a value outside this range is deemed unacceptable. In this philosophy, any joint with a UTS less than $0.72\sigma'_{UTS}$ is classified as unacceptable with little regard for its distance from the target value. Joint with UTS values of $0.15\sigma'_{UTS}$ are discarded alongside joints with ultimate tensile strengths of $0.71\sigma'_{UTS}$. Berger and Maurer refer to this manufacturing philosophy as the "goal-post" approach: engineers are only concerned with whether the measured value falls within the goal posts (i.e. the target value or range). In contrast, the Taguchi method considers not only whether the response value lies within the accepted range, but if it should fall outside the "unacceptable zone" how far beyond this range it lies. While any departure from the optimum region represents a loss in time, money, and/or quality, in the Taguchi philosophy values which are outside the "goal posts" yet close to the desired value represent less of a loss than those which are farther away. Returning to the example of joint strength, while welds with $0.15\sigma'_{UTS}$ and $0.71\sigma'_{UTS}$ are both unacceptable, the latter represents less of a loss because it is closer to the target value. The dependence of loss on distance from the accepted value is expressed using the Taguchi loss function L (equation4-1). L represents the loss in dollars per component, K is a constant related to the cost associated with the loss in quality/time/money, Y is the value measured in the experiment, and T is the target value.

$$L = K(Y - T)^2 \qquad \text{(equation 4-1)}$$

The differences between the Taguchi and "goal-post" philosophies are illustrated in Figure 36, which compares the loss functions of both systems. In the goal-post loss function, a data point coincident with the target represents no loss, while points outside the acceptable range represent a loss of equal magnitude, regardless of distance from the desired value. The Taguchi loss function is 0 at the target and increases in both directions with increasing deviation from the desired value.

Figure 36 Comparison of goal post loss function and Taguchi loss function [69]. The y-axis represents financial loss– the target value is 0. The x-axis is the signed difference between the observed value and the target value.

The continuum approach of the Taguchi method makes it well-suited for optimization problems. The quadratic dependence of the loss incurred L in equation 4-1 indicates the existence of an absolute minimum at $T = Y$. Problems which are studied using Taguchi method are thus recast in terms of minimization. In a study where the response of a system depends on a single variable, the question becomes "which value of variable A minimizes loss for this process?" This question is more versatile than its goal-post analog, as the 'solution" to the Taguchi loss function (the value of A at which the target value is equal to the measured value) points toward a nuanced representation of the

spectrum of responses. Not only do we know the location at which the loss is minimized, but also the manner in which the loss varies with distance from the target. The Taguchi loss function can thus be interpreted in terms of the mean-square error (the square of the distance between the measured and target values), a concept which serves as the mathematical basis for the regression analysis techniques discussed in section 4.4.

Another interesting feature of Taguchi techniques is the use of implicit control to minimize variability. In explicit control, the change in system response *f(x)* is minimized by controlling the change in the input *x* [69]. Taguchi promotes implicit control as a means to reduce the system's sensitivity to variability in the input factor, making the process more robust. For the same change in input Δx, a system under implicit control results in a smaller change in output $\Delta f(x)$. Implicit control is accomplished through careful selection of the levels which the factors will assume in the plan of experiments. Berger and Maurer refer to this as "designing quality in rather than weeding quality out", the latter being the approach associated with explicit control [69]. Implicit control is incorporated into the Taguchi method to facilitate construction of a regression model of the system which can predict the response based on inputs.

The preceding discussion establishes the philosophy behind the Taguchi techniques as well as features of the design which are advantageous to manufacturing applications. But how is a plan of experiments based on the Taguchi ideas constructed? Critics of the method often characterize Taguchi designs as a "cookbook approach": for each experimental combination of *n* factors at *m* levels, there is a Taguchi table which lists the experiments to be performed. Researchers employing Taguchi techniques simply "look up" the appropriate table to be used in their study. This simplified design process

was intentional on the part of Taguchi, who believed that engineers would be unlikely to utilize design of experiments if it required specialized knowledge of statistics. The Taguchi tables, more formally termed *arrays* and denoted by the symbol L, are organized in such a way that the row names correspond to the experiment number while the column headings denote the name of the factors under study. The entry of a cell located at the intersection of a row and column indicates the level of the factor in a particular experiment. The levels are denoted by integers 1 through m, where m is the number of levels in the experiment. For instance, if there are 3 levels, "1" is assigned to the level with the lowest value, "2" represents the value of the intermediate level, and "3" is the highest value in the level range. The format of a Taguchi table in its simplest form is illustrated in Table 4: the L_8 array uses 8 experiments to characterize a process with up to 7 factors at 2 levels. Each experiment consists of a combination of factors and levels referred to as a treatment.

Table 4 Taguchi's L_8 array

Experiment	**Factor A**	**Factor B**	**Factor C**	**Factor D**	**Factor E**	**Factor F**	**Factor G**
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

Selection of an appropriate Taguchi matrix is determined not only by the number of factors and levels in an experiment, but also whether these factors interact. Interaction is a broad term, but in statistics it means simply that the effect of a treatment combination is not additive. If there is no opportunity for interaction between factors, the overall effect of a treatment combination on the response variable can be expressed as the sum of the effect of A on the response variable added to the effect of B on the response variable and so on until all the factors have been accounted for. Regardless of interaction, the first experiment in a Taguchi matrix (the "1" treatment combination) is performed with all of the factors set to the lowest level. The remainder of the matrix for the no interaction case is constructed by listing the possible main effects and assigning each of them to a row/experiment in the array. A L_n array can account for up to $n-1$ factors. To use the array for an experiment with factors numbering less than $n-1$, the matrix is partitioned by column. The number of experiments (rows) remains the same, but the treatment combination for each experiment is reduced to include only the factors of interest. For example, an L_8 array for an experiment with four factors at two levels consists of the first four columns of the L_8 array in Table 4 (which can accommodate a maximum of seven factors).

When interactions are considered, orthogonal arrays can still be used but factors can no longer be assigned to columns in a random fashion. Because each interaction occupies a column in the array, the inclusion of interactions reduces the number of factors that can be incorporated into a design matrix. If we use an L_8 array with 4 factors (A, B, C, D), there are 3 (7-4) columns available for interactions. The possible two-level interactions for this example are AB, AC, AD, BC, BD, and CD. The three-level

interactions may be ABC, ABD, ACD, and/or BCD. The only four-level interaction possible is ABCD. Based on our knowledge of the process under study and an idea of how these factors interact, 3 interactions are selected to occupy the 3 available columns in the L_8 array. The assignation of interactions to columns is done in such a way that an interaction follows the factors which comprise it: for instance, the AB interaction column is placed after the A and B columns. Entries in the interaction columns are left blank since they do not represent true factors to which a level can be assigned. An example of an L_8 Taguchi array for a 3 factor, 2 level experiment with interactions, adapted from an example problem in Berger and Maurer, is displayed in Table 5. A comparison of Table 4 (the L_8 array with seven factors and no interaction) and Table 5 (an L_8 array with four factors and three two-way interactions) reveals that although the column labels are different, the position and value of the cell entries remain unchanged (the entries corresponding to interaction columns have been omitted).

Table 5 Example of a L_8 Taguchi array with interactions

Experiment	Factor A	Factor B	Interaction AB	Factor C	Interaction AC	Interaction BC	Factor D
1	1	1		1			1
2	1	1		2			2
3	1	2		1			2
4	1	2		2			1
5	2	1		1			2
6	2	1		2			1
7	2	2		1			1
8	2	2		2			2

Sometimes it is helpful to represent factors and interactions as linear graphs. This visual method presents a simple way of denoting the relationships which may exist between factors For a particular geometry, effects can be assigned to both sides and vertices. Two-way interactions are placed along the line connecting the interacting factors. The placement of a three-way interaction on a linear graph is more complex. Since it must represent the modular multiplication of the effects it comprises, it can either be placed outside the plane of a closed geometry or at the end of a segment which has a single effect at the opposite vertex and an interaction term along the line. To illustrate this point, the linear graph of an L_8 Taguchi matrix with five factors and two interactions (of both the two and three factor variety) appears in Figure 37. Effects which do not appear in the graphs are assumed to be zero. The major drawback of this graphical representation of Taguchi arrays is that it provides no obvious information regarding the levels of the factors in the experiment. When factors and levels are summarized in the tree diagram (familiar to students of combinatorics), treatment combinations can be extracted at a glance.

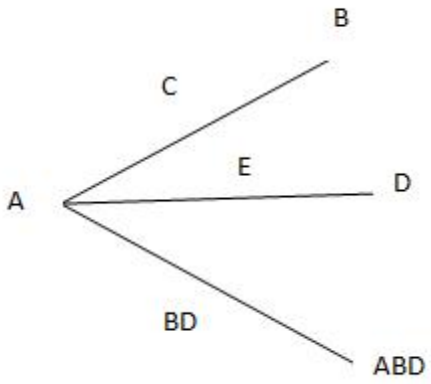


Figure 37 Example of linear graphs for a Taguchi L_8 array. There are five factors, *A, B, C, D,* and *E*, with one three-factor interaction *ABD* and one two-factor interaction *BD* [69]..

There are instances in which all or part of a Taguchi experimental matrix is identical to a design constructed using factorial techniques. For instance, a Taguchi L_{27} matrix which considers three factors at three levels consists of the same treatment combinations as a factorial design with the same specifications. Though Taguchi arrays can be created using other techniques, the major advantage Taguchi methods offer over factorial and fractional-factorial methods is the ease with which these arrays are generated. A criticism commonly leveled at Taguchi methodology is its catalog approach to experimental design, yet in many engineering applications this is considered an advantage. The purpose of the preceding, somewhat detailed information on experimental design is to put these techniques in perspective in hopes that the experimental designs for the studies which follow will not be interpreted as merely excerpts from a "cookbook."

An additional benefit associated with use of Taguchi methods is that the design of experiments (and subsequent analysis of experimental results) is done with process optimization in mind. The significant reduction in the number of experiments possible with Taguchi arrays makes data replication more feasible. A ratio of the response (signal) to the variation in the response (noise) is calculated by dividing the average value of the response variable at a particular treatment combination by the standard deviation in the response over repeated experiments. The signal to noise (SNR) ratio is a performance metric The reduction in the original number of experiments, however, can come at a cost. In selecting the factors (columns) in a Taguchi array, it is necessary to assume that many of the interactions are zero. This may be a reasonable assumption for some processes, but ignoring interactions when they are present can impair the quality of the results (and the accuracy of any models based on them).

4.3 Response Surface Methodology (RSM)

Response surface methodology (RSM) is an experimental design technique which traces its origins to regression analysis (section 4.4). In RSM, the factor levels are assumed to be continuous. This precludes the use of categorical factors (which typically take on discrete values) [70]. The term *factor* in RSM is thus synonymous with *independent variable*, since the response variable Y varies continuously with factors X_1 through X_i. The course of experiments dictated by RSM generates a set of points (X_1, X_2, …,X_i, Y) which define the response surface f. The height of the response surface is a function of the independent variables/factors and can be written as $Y = f(X_1, X_2, \dots, X_n)$, where r $f(X_1, X_2, \dots, X_n)$ is a continuous function. The continuous nature of the response

surface poses a problem for the experimenter: to completely reproduce the contours of the response surface would require thousands (or perhaps even millions) of experiments. Fortunately, RSM does not seek to generate every point which lies on the response surface, as this would be both costly and time-consuming. Rather, response surface methodology generates enough data points to capture the manner in which the system responds to changes in the independent variables. Interpolation between these data points is made possible by multivariable regression, a technique which fits the points to a continuous function (referred to as the response function).

The differences between response surface methodology and the methods of experimental design discussed in sections 4.1 and 4.2 are subtle. Both Taguchi and RSM seek to characterize the response of a system using the smallest number of experiments. Both also emphasize optimization of the response: RSM searches for "peaks" or "valleys" in the continuous response surface which indicate a maximum or minimum. The major difference lies in the objective of the designs. Factorial and Taguchi designs are primarily exploratory. These methods consider a number of factors which the researcher hypothesizes may have some effect on the response variable. Results of experiments in which the levels of the factors are varied are used to elucidate two primary pieces of information: which factors the outcome variable depends on and of these, which factors which have the greatest influence on the response. In RSM, researchers already know which factors govern the response of the system. The results of RSM experiments give a more precise, focused characterization of the process (which has presumably been studied previously using a more general method). Suppose we examine a slide containing bacteria under a microscope, Taguchi and factorial designs are

analogous to viewing the slide at a low magnification: we can count the number of bacteria on the slide and classify the bacteria by their shape. However, if we need to know the characteristic length of the streptococcus in the sample, we must reduce the distance between the lens and the slide to zoom in on a single bacterium of this variety. In experimental design, RSM is the mechanism researchers use to "zoom in" on a response surface.

In the RSM philosophy, results of Taguchi and/or factorial experiments are intended to serve as a starting point. They pinpoint the factors which are significant. Based on how the response in these studies varies with changes in the factors, researchers can select values for levels which might be located near a maximum or minimum on the response surface. This is the essence of RSM: based on the results of a cruder study, experiments are designed to locally explore the response surface. RSM techniques are used in Chapter VII and VIII to explore the relationship between wear in FSW of MMCs and two key variables: particle size and hardness of the tool material. The range of the variables considered in these studies was narrowed based on the results of investigations of wear in both machining and FSW of MMCs.

4.4 Statistical Modeling: Multivariate Regression

Sections 4.1 and 4.3 detail methods which can be used to design controlled experiments. A plan of experiments constructed using these techniques permits researchers to obtain desired information while minimizing the expenditure of resources. After the experiments have been performed, researchers need to analyze the data generated in a meaningful way. While graphical representations of data are of immense

value to researchers, the ultimate goal of experimentation is to construct a predictive equation-based model of the process being studied. This is usually accomplished through regression modeling.

A regression model relates the value of the outcome variable y to the predictor variables (synonymous with the term "factors" in experimental design). The resulting equation represents a one-to-one mapping of the set of predictors $\{x_1, x_2, \ldots., x_i\}$"onto" the corresponding value of the outcome variable y_j. To transform a set of predictor variables into y_i, each predictor is multiplied by a set of weights, denoted as $\{\beta_1, \beta_2, \ldots., \beta_i\}$. The number of weights, p, is equivalent to the number of predictors. In the regression equation, the outcome y_i is expressed as the sum of the products $\beta_i x_i$ added to e_i (equation 4-2) e_i is defined as the distance between y_i (the value obtained from the experiment) and the sum of the $\beta_i x_i$ terms. While e_i is often thought of as the "intercept term", it is technically a residual -- the signed difference between the the observation y_i and the predicted value $\widehat{y_i}$. Equation 4-2 represents a linear regression model, where $\widehat{y_i}$ is the line which best approximates the shape of the data. Most regression analyses begin with the construction of a linear model. If this model is judged to be insufficient, higher-order models can be explored (a comprehensive treatment of the mathematics of non-linear regression can be found in reference 72).

$$\widehat{y_i} = \sum_{i=1}^{p} \beta_i \, x_i + e_i \qquad \text{(equation 4-2)}$$

Sometimes it is helpful to represent the terms in the linear regression equation 4-2 as matrices (equation 4-3). X is a matrix with dimensions $N \; x \; p$. The number of rows N is equal to the number of test cases; the number of columns p is the same as the number

of predictors. Each entry in the matrix, x_{ij}, corresponds to the value of the jth predictor in the ith experiment. B is a column vector of dimensionality $p\ x\ 1$ containing the set of linear weights $\{\beta_1, \beta_2, \ldots., \beta_i\}$. The result of matrix multiplication of X and B is equivalent to the summation of products, $\sum_{i=1}^{p} \beta_i\, x_i$, in equation 4-2. The term E in equation 4-3 is a $N\ x\ 1$ column matrix whose entries consist of the residual terms $(\widehat{y_i} - y_i)$.

$$Y = XB + E \qquad \text{(equation 4-3)}$$

The set of weights $\{\beta_1, \beta_2, \ldots., \beta_i\}$ which "carry" the predictors into the outcome variable are not unique. In fact, there are many combinations of weights which produce regression lines whose shape approximates that of the data. The key to regression analysis lies in finding the optimal function (among many possible functions) which most closely approximates the observations in Y. This optimization hinges on determination of the weights $\{\beta_1, \beta_2, \ldots., \beta_i\}$ which minimize the distance between the observed values and those predicted by the regression model. The weights which correspond to the line of best fit can thus be obtained by minimizing the sum of the squared errors $({e_i}^2)$. The set of weights satisfying the least-squares criterion is given by equation 4-4. X' denotes the transpose of X, $(X'X)^{-1}$ represents the inverse of the matrix product $X'X$, and y is the $N\ x\ 1$ matrix of observations. The dimensionality of the matrix B given by this equation is consistent with that of equation 4-3, which indicates X is post-multiplied by a $p\ x\ 1$ matrix of linear weights. Although the expression used to generate B is stated here without proof, the derivation of B appears in reference 70.

$$B = (X'X)^{-1}X'y \qquad \text{(equation 4-4)}$$

Substituting equation 4-4 into 4-3 and writing E in terms of the residuals yields equation 4-5. This expression summarizes the mathematical procedure for determining the line of best fit. A word of caution: the modifier "best" in this context is not necessarily synonymous with "accurate." While the line Y represents the equation with the smallest sum of squared errors, this does not indicate that it is a good representation of the data (metrics used to assess the fit of a regression model are discussed section 4.5). Rather, the line is the best fit possible given the constraints placed on the model's construction (for instance, linear and satisfies least squares criterion).

$$Y = X(X'X)^{-1}X'y + (X(X'X)^{-1}X'y - y) \quad \text{(equation 4-5)}$$

It is clear from equation 4-5 that construction of a regression model by hand can become tedious, particularly when multiple predictor variables are considered. Commercial software packages such as SPSS (Statistical Package for the Social Sciences) are powerful tools which can be used to generate regression models ranging from linear to logarithmic. The features of these packages are also useful in assessing the predictive power of a regression model. A description of the various statistical tests which are used to assess the accuracy of a regression model is the subject of the next section.

4.5 Evaluating the Regression Model

Once the regression model is constructed, the next task is to assess how well the model reflects the shape of the data. A visual inspection of the regression model superimposed on the data can provide some idea of the goodness of fit. If the observed

data lies close to the regression line, then differences between the actual values and those predicted by the model are small, an indication that the model is good representation of the process under study. While valuable as an initial estimator, visual comparison of the model and data leaves much to interpretation. Fortunately, there are a number of quantitative tests and metrics which can be used to evaluate the "goodness" of a particular model. These metrics are briefly defined and summarized in the section which follows.

4.5.1 Quantitative Assessment of the Regression Model

The simplest model of any data set is the mean of the observations, $\bar{y}$ (equation 4-6). When plotted, this model consists of a straight line defined by the equation $Y = \bar{y}$. The total sum of squares (SS_T), is the squared sum of the distances between the mean and each data point (equation 4-7). SS_T becomes a useful metric when it is compared to the residual sum of squares (SS_R) (equation 4-8), the squared sum of the differences between the observations and the values predicted by the best-fitting regression model. Residual sum of squares is the quantitative analog of the visual inspection method discussed previously (a smaller SS_R indicates a better overall fit). The difference between the residual sum of squares and the total sum of squares (the model sum of squares SS_M) indicates the degree of improvement associated with use of the regression model as a predictor (equation 4-9). A large SS_M value means that the regression model represents a significant enhancement in our ability to predict the outcome variable; smaller values of SS_M signify that the regression model offers little improvement over a model which uses

the mean as a predictor. The ratio of SS_M to SS_T is denoted as R^2 (equation 4-10). The Pearson correlation coefficient, R, is simply the square root of this ratio. R^2 is the metric most frequently used to evaluate the fit of a regression model. The closer the value of R^2 lies to unity, the better the regression model fits the data. R^2 can be interpreted as the percentage of the variation in the observations that can be explained by the regression model [70].

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad \text{(equation 4-6)}$$

$$SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad \text{(equation 4-7)}$$

$$SS_R = \sum_{i=1}^{n} (y_i - Y_i)^2 \qquad \text{(equation 4-8)}$$

$$SS_M = SS_R - SS_T \qquad \text{(equation 4-9)}$$

$$R^2 = \frac{SS_M}{SS_T} \qquad \text{(equation 4-10)}$$

While the metrics discussed thus far arise from the sum of squares, the mean sums of squares (MS) can also be used to extract information regarding the fit of the model. The MS of the regression model (MS_M) and the residuals (MS_R) are calculated using equations 4-11 and 4-12, respectively. n corresponds to the number of observations in the sample and p is the number of predictors in the regression model. The ratio of MS_M to MS_R forms the F statistic (equation 4-13).

$$MS_M = \frac{\sum_{i=1}^{n} (Y_i - \bar{y})^2}{p-1} \qquad \text{(equation 4-11)}$$

$$MS_R = \frac{\sum_{i=1}^{n}(y_i - Y_i)^2}{n-p} \qquad \text{(equation 4-12)}$$

$$F = \frac{MS_M}{MS_R} \qquad \text{(equation 4-13)}$$

The *F* statistic represents a point on the x-axis of the F distribution (Figure 38). The graph in Figure 38 shows only the general shape of the F curve: each combination of $p-1$ and $n-p$ (the degrees of freedom of the model and the residuals, respectively) has a unique probability distribution. In hypothesis testing, the value of *F* determines whether the null hypothesis is accepted or rejected in favor of the alternative hypothesis. For regression modeling, the null hypothesis H_0 states that the value of the predictor variables in the model $\{x_1, x_2, \ldots, x_i\}$ have no measurable effect on the outcome variable Y. The alternative hypothesis H_a asserts that the values assumed by $\{x_1, x_2, \ldots, x_i\}$ effect the value of Y. A critical value c is set such that the region to the right of *c* consists of an area which is a fraction α of the area under the entire distribution. If *F* is greater than *c*, the null hypothesis is rejected and the model is said to be "significant at the α level" [71]. The significance value of F corresponds to the area underneath the F distribution curve to the right of the F statistic. An F statistic which exceeds the critical value indicates that the probability that the null hypothesis is true (that the predictors have no effect on the outcome variable) is less than α. This is sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis. The value of α is thus equal to the probability of a type I error, which occurs when the null hypothesis is rejected but is actually true. Since larger values of F coincide with smaller values of α, the probability of a type I error decreases as F increases. For larger F statistics, we can

conclude with greater confidence that the predictors considered in the model have a measurable effect on the outcome variable.

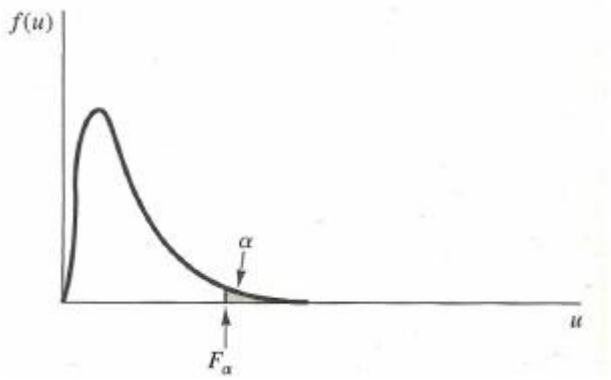


Figure 38 Shape of F distribution [71].

The F-statistic provides important information concerning the strength of the relationship between the set of predictors $\{x_1, x_2, \ldots, x_i\}$ and Y. To extract the importance of an individual predictor on the outcome requires the t-test. In a multiple regression model, each predictor x_i has an associated coefficient β_i. The t-test compares the value of the coefficient β_i with 0 (the null hypothesis states that $\beta_I = 0$, in which case the parameter x_i has no impact on the output of the system). The t-statistic for a predictor (equation 4-14) is the ratio of the predictor's coefficient in the multiple regression model to the standard error (SE_β). In this formulation, the standard error is simply the standard deviation of the data. The t-test is executed and interpreted in a manner similar to the F-test. If the coefficient is different from zero, the value of t will lie in the rejection region

of the t-distribution (Figure 39). As in the F-distribution, the shape of the curve is determined by the degrees of freedom. In a two-tailed t-test, the rejection region is defined by the inequality $\{t \leq \frac{-\alpha}{2} \cup t \geq \frac{\alpha}{2}\}$. If the t-statistic of a coefficient lies within this region, it is said to be significant at the $\frac{\alpha}{2}$ level. As in the F-test, the significance value (sometimes referred to as the p value) is synonymous with the probability of a type I error. For sufficiently small values of p ($p < .05$), the argument can be made that the parameter significantly contributes to the predictive capability of the model.

$$t = \frac{\beta_i}{SE_{\beta_i}} \qquad \text{(equation 4-14)}$$

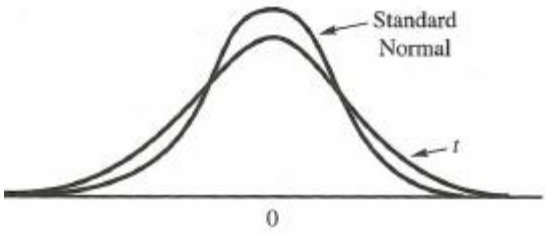


Figure 39 Comparison of t-distribution with standard normal [73].

When statistical software such as SPSS is used to construct a multivariate regression model, the analysis outputs a great deal of information. The output includes the common metrics for evaluating goodness of fit and significance (the F and t-tests discussed above) as well as less familiar measures. Other tests which can used to evaluate a regression model include residual statistics and the adjusted R^2 value.

- Residual statistics are used to gauge the influence of an individual case on the model. An important residual statistic is Cook's distance, which assesses the effect of deleting an observation. A case with a Cook's distance greater than 1 indicates that the case may bias the model. If the value of Cook's distance flags a case as a potential outlier, Field suggests rerunning the analysis with the data from the case(s) in question omitted [70]. The larger the change in the model coefficients, the greater influence the omitted case has on the model. Since an outlier represents a substantial deviation from the overall trend of the data, it is likely to negatively affect the predictive capability of the regression model. How to treat outliers in a statistical analysis is left to the discretion of the researcher. The experimental case which produced the outlier can be repeated in an effort to produce a non-anomalous result. If repeated experimentation yields the same result, the outlier may be omitted from the analysis altogether or treated as a missing data point.
- The adjusted R^2 value predicts how effectively the model can be used to generalize a population beyond the sample used to construct it. The adjusted R^2 value can be calculated using a variety of methods; SPSS uses the simpler formulation of Wherry (equation 4-15) [70]. The data shrinkage (defined as the difference between the R^2 and adjusted R^2 for the model) represents the reduction in the predictive power of the model when it is applied to the population. A small difference predicts that the model will generalize well. Data shrinkages of less 5 percent enhance confidence in the applicability of the model and the conclusions

which may be drawn from it. Using the adjusted R^2 value to estimate the accuracy of the model across the population is discussed further in section 4.5.4.

$$adjusted\ R^2 = 1 - \frac{(1-R^2)(n-1)}{(n-k-1)} \qquad \text{(equation 4-15)}$$

4.5.2 Assumptions and Limitations of Regression Modeling

While regression modeling is a powerful tool which can enable researchers to characterize and predict complex processes, its usefulness in some instances may be limited by the assumptions which underlie the mathematics (section 4.4) used to construct the model. According to Field, the major assumptions inherent in a regression model include:

Restrictions on predictors. Values of predictor variables must be quantitative. In the case of a categorical/quantitative predictor, a numeric value (such as 0 or 1 when a predictor can be classified as one of two categories) is assigned. Additionally, the variance of the predictor values must be non-zero. Regression modeling also precludes multicollinearity among predictors. A data set exhibits multicollinearity when the value of a predictor can be represented as a linear combination of two or more of the other predictors in the set. The value which a predictor assumes must be "uncorrelated with external variables" [70]. This statement implies that researchers must be able to exercise control in selecting the value of the predictor variable; the level of the predictor should not be influenced by other variables which are not a part of the study.

Restrictions on errors. The residuals in the sample (the difference between the observed outcome and that predicted by the regression model) should be uncorrelated. In addition to violating a critically important assumption of regression modeling, highly-correlated residuals reduce the standard error, thereby increasing the value of the t-score for a predictor (equation 4-14). Autocorrelation of errors can thus steer researchers toward a false positive result when assessing the significance of predictors. The metric which gauges the amount of correlation among the residuals is the Durbin-Watson statistic [70]. While values of Durbin-Watson range from 0 to 4, in general any value smaller than 1 or exceeding 3 is cause for alarm.

Another assumption of regression is that errors are normally distributed. The histogram of the range of the values of the residuals plotted against the frequency with which each value occurs should approximate a Gaussian distribution. The majority of the residuals have values near 0 (the center of the distribution). Larger residuals are observed less frequently.

Restrictions on outcome variables. The values of the outcome variable should be independent. "Independent" in this context means that each result originates from a separate case [70]. A first-order regression model additionally assumes that the relationship between the outcome variable and the predictors over the increments considered is linear. Formally stated, the mean values of the outcome variable calculated over a fixed increment of the predictor range (a plotting procedure referred to in signal processing as a moving window average) form a straight line.

Before proceeding with regression modeling as a means of analyzing a process, it is important to verify that the above assumptions are valid. Checking these assumptions can serve as an initial test of the model's robustness. In most models, all of the assumptions may not hold true. The criticality of the assumptions varies: while violation of some assumptions has little effect on the model, violation of others (such as uncorrelated residuals) can skew significance values and lead to incorrect conclusions. The appropriateness of regression modeling for an application is ultimately left to the judgment of the researcher. In general, processes which do not meet a majority of the criteria outlined should not be represented in terms of a regression model.

As a final caveat, researchers should exercise caution when evaluating regression models constructed using statistical software packages. These models are influenced not only by the values of the predictors, but also by the manner in which the predictors are entered. The most common method of entry (and the default setting for most software packages) is *forced*. In forced entry, all predictors are entered simultaneously; the researcher is not called upon to make any judgments prior to the analysis regarding which predictors are the most (or least) significant. In *hierarchal* entry, researchers rank the predictors based on their perceived importance. The factor judged to be the most significant is entered into the model first -- entry of the remaining factors proceeds in descending order. The third form of entry, *Stepwise*, is usually employed only when there are a very large number of predictors. Predictors are added to the model incrementally and tested for significance. Predictors which do not meet the specified significance criteria (for instance, $p < .005$) are discarded and omitted from the regression model. Of these three approaches identified by Field, forced entry is regarded

as the most "honest" as it does not permit the researcher to make decisions regarding the importance of a predictor until after the analysis is complete. In academic writing, researchers often do not specify the manner in which predictors were entered. Readers should exercise caution in drawing conclusions based on the results of stepwise models, since the fit of model can be substantially improved by selective inclusion (or omission) of factors.

4.5.3 Testing the Regression Model

The major problem with regression modeling is that of generalization. While a model may fit the sample data set well, it cannot be assumed that this predictive capability is maintained over the entire range of the process variables. To validate the regression model, it is necessary to assess how accurately the model predicts outcomes for sets of process parameters which lie outside the sample used to construct it. This testing is analogous to certain aspects of machine learning: a machine's predictive power is tested using data sets separate from those used to train the machine and build the classification algorithm, a process referred to as cross-validation. If the predictions of the model agree with experimental observations, the model is said to generalize well to the population. A generalizable model enables researchers to extend the model's predictions beyond the initial data set. A successfully validated model is essential for establishing confidence that the data collected in the sample accurately characterizes the behavior of the larger population.

As discussed in section 4.5.3, the adjusted R-squared value ($R_{adj}{}^2$) is one metric which can be used to assess the model's applicability to different data sets. A model with a small shrinkage (the difference in the R-squared and adjusted R-squared values) is believed to generalize well. Although there are several different formulas which can be used to compute the adjusted R-squared for a regression model, the meaning of the quantity is the same regardless of the formula used to calculate it. $R_{adj}{}^2$ indicates the proportion of variance in the population accounted for by the model. The closer $R_{adj}{}^2$ lies to R^2, the smaller the shrinkage. A small difference between these values means that the model should lose little of its predictive capability when applied to the population [70].

The adjusted R-squared value provides one estimate of the model's generalization. Another technique which can be used to gauge how accurately the model represents the population is data splitting. In this method, the data collected is divided into two groups, each containing half the data points. One of the groups is used to construct the regression model; the cases which comprise the other group are used for testing. The cases in the half of the data matrix set aside for validation are substituted into the regression model and the value predicted by the model is compared with the value observed. More accurate models have smaller discrepancies between observed and predicted values. Data splitting is best-suited for large data sets, where omission of half the cases leaves enough remaining data points to construct a reliable model. While sample size is dictated in many cases by the choice of experimental design, Miles and Shelvin's plot of sample size versus number of predictors (Figure 40) can be used to estimate the number of experimental cases needed to reduce the probability of a type II

error [70]. To use data splitting as a means of model validation, the size of the original sample must be approximately twice as large as that indicated in Figure 40.

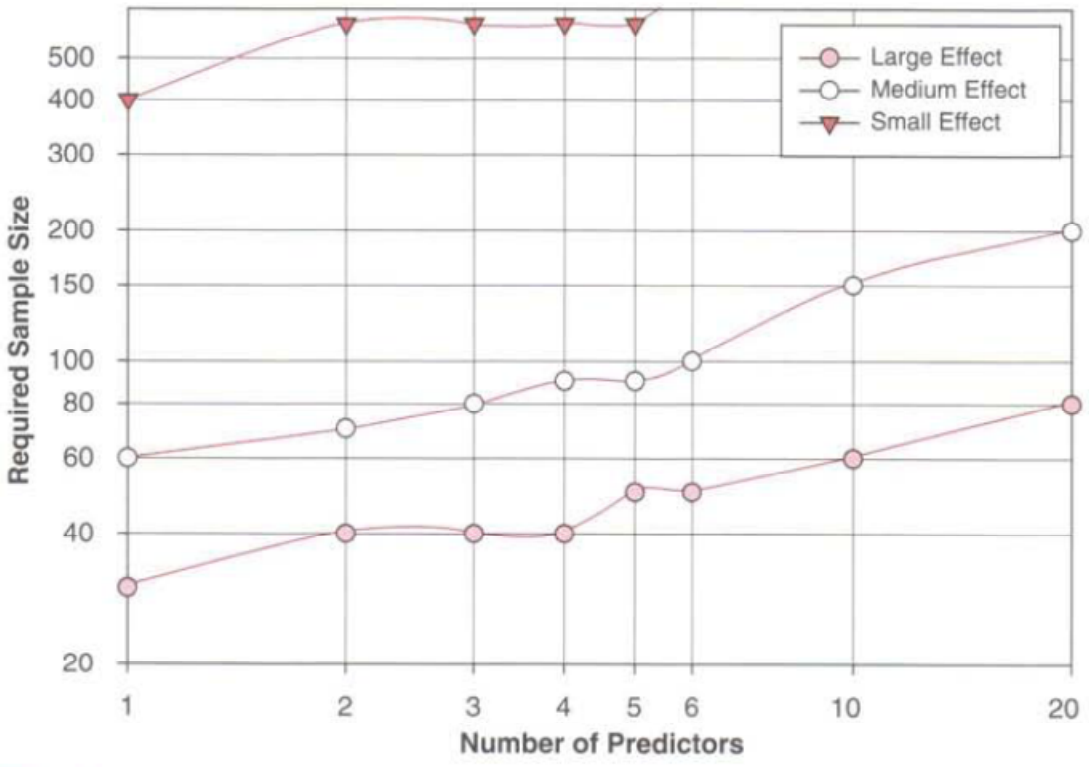


Figure 40 Miles and Shelvin's recommended sample size for regression based on number of predictors and size of effect [70].

The most reliable test of a model's capacity for generalization is cross-validation. Since cross-validation does not use any of the cases in the original sample, it can be applied to a study of any size. The experimental array for cross-validation consists of cases with levels separate from those used to construct the model. Cross-validation thus requires generation of additional data points. However, a large number of test cases is unnecessary in most instances. A few cases are sufficient, provided they are

representative of the range of values the process variables can assume. The model generalizes well if the residuals for the test cases (the difference between the experimentally observed value and the value predicted by the model) are small. It is common to cast residuals in terms of percent error; observed and predicted values which differ by greater than 10 percent are cause for concern. In some ways, cross-validation is itself a test of the efficacy of the adjusted R-squared metric. It is neither impossible nor unprecedented for a model with roughly equivalent values of R^2 and ${R_{adj}}^2$ to fail cross-validation. The adjusted R-squared value is merely a prediction of how the model will translate to the population, not a concrete test of applicability If a model does not cross-validate well, conclusions drawn from it are restricted to the sample from which the model originates. Depending on the levels chosen in the experimental design stage, this may be a small range. Not surprisingly, models with a high degree of specificity are less useful than those which can be generalized to the population at large.

4.6 Results from Literature: Characterization and Prediction of Wear in Machining of Metal Matrix Composites

The preceding discussion focused on the efficient design of experiments and subsequent assessment of models derived from experimental results. We will now examine how the methods discussed in sections 4.1 through 4.5 have been applied to study wear in the machining of metal matrix composites. The studies summarized here serve as both a precedent and guide for our investigations into wear in frictions stir welding of these materials.

4.6.1 Variation of Wear with Process Parameters in Machining of Metal Matrix Composites

A number of published studies consider the effect of processing parameters on wear of the tool in machining (drilling, turning, etc.) of composites. The majority of these studies use Taguchi orthogonal arrays (section 4.2) in conjunction with multiple regression modeling (section 4.4). In this manner, the effect of several parameters on the outcome variable can be determined using a single experimental treatment [53]. Although several of the process variables in machining are distinct from those encountered in FSW, these studies demonstrate that tool wear in these processes is deterministic and repeatable. This summary does not encompass all published machinability studies on MMCs, but focuses on examples with methodologies similar to those used in the research presented in chapters V-VII.

Cutting of MMCs. The term machining is usually synonymous with cutting, a method of material removal in which the workpiece (or part) is rotated on a lathe while a cutting tool traverses parallel to the axis of rotation [72]. A cutting tool should be significantly harder than the material being cut; most tools are made from high speed steel or carbide to facilitate easy removal of material from the workpiece. In the case of cutting MMCs, the abrasive reinforcement is in many cases significantly harder than the cutting tool, resulting in wear of the tool bit. The effect of process parameters on tool wear in cutting of MMCs has been investigated extensively by J. Paolo Davim and colleagues. Using Taguchi matrices in conjunction with multiple regression analysis, Davim was able to quantify the contribution of cutting velocity V (the difference between the cutting tool and the rotating workpiece), feed rate f (the speed at which the cutting tool advances along the workpiece), and cutting time T to tool wear [54]. Wear was monitored by

comparing the dimensions of the tool flank before and after machining. The regression model derived from Davim's experiments indicates that wear of the cutting tool is directly proportional to cutting velocity and cutting time but varies inversely with feed rate [53]. There are several other studies which report similar results. Sahin and Sur, Pramanik et al., Quigley and Monaghan, and Kilickap et al. all found that greater cutting speeds produce an increase in wear [62,65,73-74]. The relationship between tool wear and cutting speed is the most statistically (and physically) significant of the process parameters considered. Kilickap et al. found that doubling the cutting speed produces a three-fold increase in flank wear [74]. Sahin and Sur argue that this effect can be attributed to thermal softening of the workpiece (since heat input increases with cutting speed), while Pramanik et al. suggest it is a result of increased contact between the abrasive reinforcement and the cutting tool [73].

The inverse dependence between wear and feed rate in cutting of MMCs documented by Davim and Baptista is similar to that observed in FSW of these materials, where the amount of volumetric tool wear behaves oppositely to the speed at which the tool traverses the workpiece [64]. There is, however, some disagreement as to the nature of this effect in cutting of MMCs. Kilickap et al. found that the influence of feed on wear varies depending on cutting speed – at higher cutting speeds, higher feed rates result in a slight increase in tool wear [74]. Davim and Baptista suggest that the increased tool life they observed for higher feed rates is characteristic of an abrasive wear mechanism [64]. Consequently, they recommend that operators machining these materials run at the highest acceptable feed rate to minimize wear of the cutting tool.

The classical analysis of tool wear in cutting is Taylor's tool life equation (equation 4-16), which expresses tool life T as a function of cutting speed V and two constants, n and C [72]. n is an index which accounts for the compatibility of the tool and the workpiece material; material interactions which result in increased wear have higher values of n. Values of n for more common material/workpiece combinations can be found in a machinist's handbook. C is a constant calculated based on measurements of tool life. However, the percent change in tool life which results from a change in cutting speed can be calculated without explicit knowledge of C. The form of Taylor's equation can be slightly modified to include the additional process parameters feed rate f and depth of cut D (equation 4-17) [72]. The inclusion of f and D adds a level of complexity to the Taylor equation, as the constants associated with these variables generally do not appear in published data and thus must be determined experimentally.

$$VT^n = C \quad \text{(equation 4-16)}$$

$$VT^n D^x f^y = C \quad \text{(equation 4-17)}$$

The primary usefulness of Taylor's equation lies in its ability to provide a crude approximation of tool life for various cutting scenarios. Conclusions drawn based on the Taylor equation are generally in agreement with the observations reported in studies regarding tool wear in cutting of metal composites. As cutting speed increases, tool life decreases as a result of increased wear. The variation of tool life with feed rate is less clear. Based on equation 4-17, whether tool life increases or decreases with feed depends on the relative magnitude of the f^y term. The Taylor equation thus offers a potential explanation for the discrepancies in the results of Davim.(who observed that higher feed

rates decreased tool wear) and Ozben et al. (who found that wear is directly proportional to feed) [75-76]. Since there is some overlap in the range of feeds examined in each study -- feed rates considered by Ozben et al. (0.10 mm/rev to 0.30 mm/rev) were only slightly faster than those studied by Davim (0.05 mm/rev to 0.20 mm/rev) – disparities in the observed dependence of wear on feed rate may instead be related to material properties. The sets of researchers used different workpiece materials (Davim studied a Al 359/SiC/20p MMC, while Ozben et al. used a unspecified Aluminum alloy with 5 to 15 percent SiC reinforcement by weight) as well as different tool materials (the tools used by Ozben were K10 grade, but several tools were coated with TiN; those used by Davim were fabricated from polycrystalline diamond). Thus changes in the values of the constants which occur when the workpiece and/or tool material is modified may account for incongruities in the relationship between tool life and feed rate. The constants in the Taylor equation serve as a reminder of the important effect that material properties exert on wear processes, a subject discussed further in section 4.6.2.

Drilling of MMCs. Drilling is a material removal process in which a cutting tool (referred to as a drill bit) is used to form a cylindrical hole in the workpiece. The process variables encountered in drilling are nearly identical to those of cutting. The feed rate f indicates the relative velocity between the tool and the workpiece; for drilling, the velocity vector has only a vertical component, as the motion of the drill bit is almost always in a direction perpendicular to the workpiece. The cutting speed V represents the speed at which the drill bit rotates. As with cutting, wear of the tool in drilling is observed only for high-temperature alloys and abrasive materials. In these scenarios, drill bits are fabricated from harder materials and/or coated to improve wear resistance.

Published data on tool wear in drilling of MMCs is scarce. The most comprehensive study on this subject is Davim's characterization of the effect of process variables on wear in the drilling of Al 356/20p/SiC [54]. Davim employed the same methodology used in his previous work on wear in cutting of MMCs (a Taguchi design of experiments coupled with the statistical analysis technique ANOVA) to gauge the contribution of the process variables f (feed rate), V (cutting speed), and T (cutting time) to the flank wear of the drill bit. Of these parameters, Davim found that cutting time has the strongest influence on wear. The contribution of T to flank wear is estimated at 50 percent, a value which is substantially greater than either the effects of feed rate (24 percent) or cutting speed (7 percent). The relationship between wear and process parameters in drilling MMCs is distinct from that observed by Davim in his studies on cutting: the peculiar inverse dependence of wear on feed rate documented in [53] is absent in drilling.

Davim's study is an extension of earlier work by Coelho, Yamada, et al., who measured tool wear in drilling of Al 2618/SiC/15p for four different feed rates and two cutting speeds [77]. For a constant cutting speed, wear increases with increasing feed rate (and vice versa). Not surprisingly, the amount of wear also increases increases linearly with the number of holes drilled n. From the plots which appear in reference [77], it appears that n is the strongest predictor of wear; this agrees with Davim's conclusions in [54], which pinpointed machining time T as the major influence on tool life.

Additional considerations: Scalability, thermal effects, and cutting time. Although the findings discussed in this section provide valuable information regarding the effect of

process parameters on wear in machining of MMCs, there are some points which require additional clarification. Based on the relationships established for flank wear in cutting ($VB \propto V, T, 1/f$) and drilling ($VB \propto V, T, f$), it would seem that wear could be easily minimized for either process by choosing parameters which correspond to the least amount of wear. In real manufacturing scenarios, however, the choice of parameters is restricted by both the apparatus limits and the operating window; the latter is bounded by the sets of parameters which result in a satisfactory product. Operators must thus negotiate a compromise between wear and quality by selecting parameters which mitigate wear yet still produce a cut with an acceptable surface finish.

Davim found the interactions between process parameters to be negligible in both cutting and drilling. While these interactions have F values that are statistically significant, their percentage contribution P to the outcome variable wear is small enough (in most cases less than 5 percent) that they can be omitted from the regression model. It may be tempting to think of the predictor variables in these machinability studies (V, f, T) as entities which can be set independently from one another. This is certainly true for Davim's experiments, in which the values of the parameters were carefully selected to fit an orthogonal array. In actual production, however, parameter values are interrelated and their selection is governed largely by the quality of the resulting cut. In cutting, for example, the feed rate is determined by the depth of cut D. For faster feed rates, the depth of cut must be made shallower or the cutting speed must be increased to facilitate removal of the desired amount of material. Additionally, the cutting time T depends on both the feed rate f and the dimensions of the part being machined. Because these variables are coupled (to an extent) in actual manufacturing scenarios, it may be possible

to consolidate the predictor variables into a single factor related to the rate of material removal. In terms of wear prediction, a single grouping of several variables (referred to as a dimensionless group when the units of these variables cancel one another) permits scalability. While multivariate regression is a powerful tool, the disadvantage of analyses such as those used by Davim is that they are essentially case studies, specific to the experimental conditions chosen by the researcher. While the models cross-validate well, it is uncertain to what degree their predictive capability is affected by changes in apparatus, tool geometry, and/or the selection of process variables which lie outside the range of the parameters used to construct the model. Scalability is a major concern in all manufacturing processes, including friction stir welding. In general, experimentally derived models are very sensitive to scaling. The use of dimensional analysis to improve the scalability of predictive process models of tool wear in machining (and by extension, FSW) is discussed in chapter V.

As a final note, it is important to exercise caution when interpreting relationships between the predictor variables and an outcome. In references -, variations in the amount of wear encountered in machining of MMCs are attributed to changes in process parameters; the relative strength of the correlation between a particular process parameter and wear is indicated by the results of statistical tests. In accordance with the familiar caution that "correlation does not imply causality", the results of the studies discussed in this section should not be interpreted to mean that the selection of process variables in machining MMCs causes wear of the cutting tool. The relationship between wear and process parameters may be classified as correlation without causality: the classic example of this type of relationship is the link between the damage from a household fire in

dollars (Y) and the number of fire trucks deployed to handle the situation (X). While there is a strong and obvious correlation between these two variables, it would be erroneous to claim that the amount of damage done by the fire is caused by the fire department's dispatch of additional trucks. This type of relationship is known as a spurious correlation: although the variables are highly correlated, changes in the outcome variable cannot be attributed to variation in the predictor variable(s). Instead, variability in the outcome is related to a third variable (referred to as a latent variable since it is not directly observed). In the previous example, the latent variable is the size of the fire, which determines both the predictor variable (number of trucks) as well as the outcome variable (damage).

There are several candidates for a latent variable in the wear in machining of MMCs which may require additional investigation. Chief among these are the duration of the machining process and the temperature of the tool and/or workpiece. The temperature experienced by either of the contacting materials during machining is itself a function of the process parameters. For instance, an increase in cutting speed tends to increase heat input; increasing the feed rate (or making the cut shallower) often has the opposite effect. The idea that changes in the materials brought about by temperature effects have the potential to exacerbate (or inhibit) wear is posited by Sahin and Sur in reference [62]. When a substantial amount of heat is transferred from the workpiece to the tool, the tool material becomes softer. The effect of thermal softening is a decrease in the yield strength of the tool, which in turn reduces the force required for hard particles to impinge on the tool surface and remove material. Scenarios where the majority of the heat flows from the cutting tool to the workpiece may experience less wear since the

tool's material properties are not compromised by excessive heat input. It may be possible to estimate the degree of thermal softening which can occur by comparing the product of the volumetric heat capacity and the thermal conductivity for the materials in contact [78]. Large discrepancies in these values may indicate that thermal effects are significant and must be considered in analysis of the wear process. Temperature effects can also be monitored using a thermal camera or thermocouples embedded in the material or fixed to the tool shank. If the tool experiences very high temperatures, then there is a possibility that wear is intensified by thermal softening of the tool. If the observed tool temperature remains within material limits, then the thermal component of wear may be negligible.

As we have noted, hardening or softening of one of the contacting materials in relation to the other may either facilitate wear or guard against it. The temperature variations within the system are influenced by the process parameters, making temperature a potential latent variable in predicting wear. The time for which the materials are in contact,t, is similarly determined by the process parameters. Although it may not explicitly appear in the data collected, the duration of the machining process is in most instances the reciprocal of the feed rate f multiplied by the length of the cut. The effect of time of contact on wear can provide insight into the wear mechanism (section 3.2). For wear mechanisms which behave in accordance with principles of drag, wear becomes more pronounced as the feed rate is increased. An inverse relationship between wear and feed rate, such as that observed by Davim for cutting, suggests the wear mechanism is a shear phenomenon since faster feed rates reduce the machining time [53]. Reducing the time which the materials are in contact (by increasing the feed rate) limits

the peak temperature of the machining process. As indicated by Sahin and Sur, wear in cutting of MMCs may be exacerbated by high temperatures which induce thermal softening of the tool [62]. The puzzling relationship between feed rate and wear in cutting of MMCs can thus potentially be explained in terms of interaction between the latent variables temperature and time of contact. The lower cutting temperatures which accompany increased feed rates may preclude thermally-driven changes in the tool's material properties which accelerate wear. Whether the fluctuations in tool temperature encountered at lower feed rates and higher cutting speeds are considerable enough to influence the wear process requires additional study. In-process temperature data acquired using thermocouples and/or thermal imaging can be used to assess the degree to which temperature may affect wear.

The existence of latent variables such as temperature and time of contact may explain some of the variation in the amount of wear which was unaccounted for by the process parameters considered by Davim, Ozben et al. and others. The predictive model for tool wear in turning of MMCs presented by Davim has an R^2 value of 0.6084, which can be interpreted to mean that approximately 40 percent of the variation in tool wear (the complement of the R^2 metric) can be attributed to error and/or variables which are not part of the regression analysis [53]. Factors other than process parameters which may influence tool wear in machining of MMCs are discussed in the next section.

4.6.2 Variation of Wear with Material Properties

The material-specific constants which appear in the most basic form of the Taylor equation (equation 4-16) demonstrate that the amount of wear depends not only on process parameters, but is also substantially influenced by the properties of the materials in contact. There are a number of additional factors which may affect tool life in machining of metal composites: the diameter of the abrasive particles which reinforce the material, the percentage in which these particles are present, and the hardness of the reinforcement relative to the tool material. Unfortunately, the end user may have less control over these variables than the process parameters. While a major advantage of MMCs is their ability to be customized by varying the matrix alloy, particle size, and/or percentage reinforcement, those purchased "off the shelf" from a materials manufacturer typically contain smaller particles (FEPA grade F500) of SiC, Al_2O_3, or B_4C, embedded in Al 359, 6061, or 7005. The choice of alloy is dictated by material compatibility: for a given reinforcement material, there are a limited number of conventional Aluminum alloys with which it can be combined easily. The material properties of the composite are additionally affected by the process used to produce it. For instance, particulate-reinforced composites produced using powder metallurgy (specifically hot isostatic pressing) exhibit a higher degree of isotropy than those manufactured with liquid techniques [29]. Characteristic reinforcement percentages for MMCs range from 10-40 percent by volume; as discussed in section 2.1, the constraints on percentage reinforcement are dictated by machining capabilities. Metal composites with 40 percent reinforcement represent the upper limit of what can be machined using standard methods: materials with reinforcement percentages above this threshold require exotic (and more

expensive) techniques to produce a net shape part due to rapid wear of the tool. The selection of the tool material for machining of MMCs is limited by budget as well as the nature of the machining operation: while harder materials such as ceramics are typically used to combat wear, they are brittle and have a propensity to fracture under the high shear stresses associated with some machining processes. A change in any of the aforementioned properties (particle size, matrix alloy, reinforcement material, percentage reinforcement, tool material, or tool geometry) has the potential to impact the wear process. This section synthesizes the existing research into the variation of wear in machining with properties of the tool and/or workpiece.

4.6.2.1 Variation of wear with percentage reinforcement

There are only a few studies which characterize machine tool wear in cutting MMCs as a sole function of percentage reinforcement. This may be a consequence of the limited control researchers have over percentage reinforcement when materials are purchased from a manufacturer. Li and Seah were able to circumvent this constraint by manufacturing their own composites using liquid phase casting (in which preheated particulate is stirred into a molten matrix alloy and solidified in a steel mold to form a plate) [79]. Using this inexpensive method, Li and Seah created an array of Aluminum composites with reinforcement percentages ranging from 2.5 to 15 percent by weight [79]. Perhaps not surprisingly, the results of their subsequent experiments demonstrate that the rate of tool wear in cutting of MMCs is directly proportional to percentage reinforcement. The explanation for this dependence is simple and intuitive. Increasing

the reinforcement percentage (without changing the matrix alloy or the dimensions of the formed composite) increases the population of abrasive particles embedded in the material. Materials with large reinforcement percentages contain more particles per unit volume; the distance between particles is thus significantly reduced. This increased particle density enhances the likelihood that the tool will come in contact with the abrasive, an effect which Li and Seah term particle interference.

Another important facet of Li and Seah's work is their determination of a critical reinforcement percentage. The plot which appears in Figure 41 shows flank wear for various cutting speeds as a function of the percentage reinforcement. Tool wear is minimal when the reinforcement percentage is below 5 percent. Above 5 percent reinforcement, however, the wear rate begins to sharply increase. The same trend is observed across a range of feed rates. The wear data from Li/Seah illustrates the compromise between material strength and tool wear that must be negotiated in machining metal composites. While the addition of powder abrasives in large quantities can double or even triple the elastic modulus of the base alloy without substantially increasing its weight, the high particle densities which correspond to these levels of reinforcement render the material virtually unmachineable using conventional tool materials. From a material design standpoint, this means that the amount of reinforcement which can be included in net shape parts is limited by the amount of flank wear associated with various reinforcement percentages. For instance, if an operator determines that flank wear must be maintained below 1.6 mm to ensure the quality of the cut, this limits the selection of workpiece materials to composites with less than 7.5 percent reinforcement (assuming the cutting conditions lie within the window considered

by Li and Seah). A plot or chart which indicates the class of materials and/or process parameters which meet the wear criteria for a particular application is known as a machinability map. The usefulness of machinability maps is discussed further in section 2.6.2.3, which addresses how wear varies jointly with both percentage reinforcement and particle size.

The outcomes of investigations by other researchers corroborate the relationship between percentage reinforcement and the amount of tool wear observed by Li and Seah. Kishaway et al. estimate an approximately 1 to 1 correspondence between flank wear and percent reinforcement: for identical cutting conditions, if the volume fraction is doubled from 10 to 20 percent, the flank wear also increases twofold [63]. Marinov found that volumetric wear of the cutting tool increases with reinforcement percentage over the range of 5 to 30 percent, a relationship illustrated by the plot in Figure 42 [66]. The solid line of best fit shows that the response of the system to variations in percentage reinforcement is nonlinear, a result consistent with Li and Seah and contrary to Kishawy et al. The effect of the amount of reinforcement on tool wear in FSW of MMCs was documented in the technical report by Diwan discussed in section 2.4.2. The author's own investigation into percent reinforcement and wear in FSW is incorporated into the wear study of Chapter VIII.

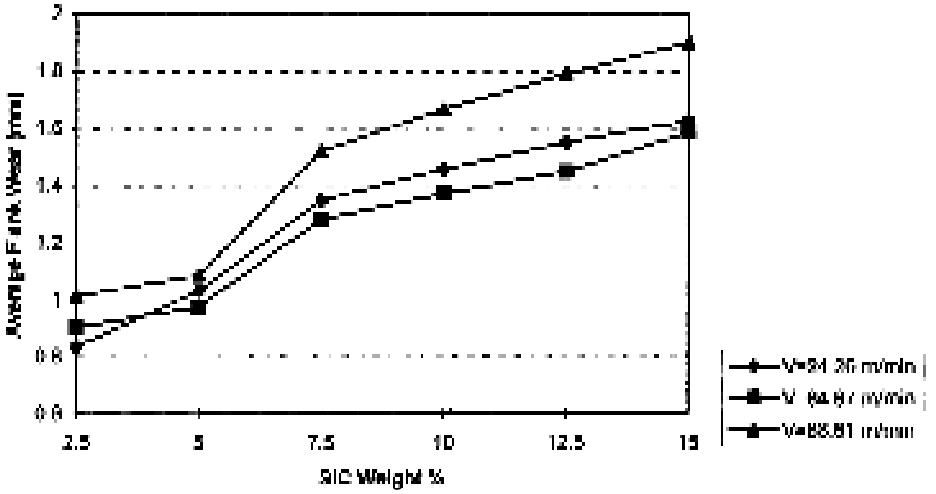


Figure 41 Average flank wear as a function of reinforcement percentage [79].

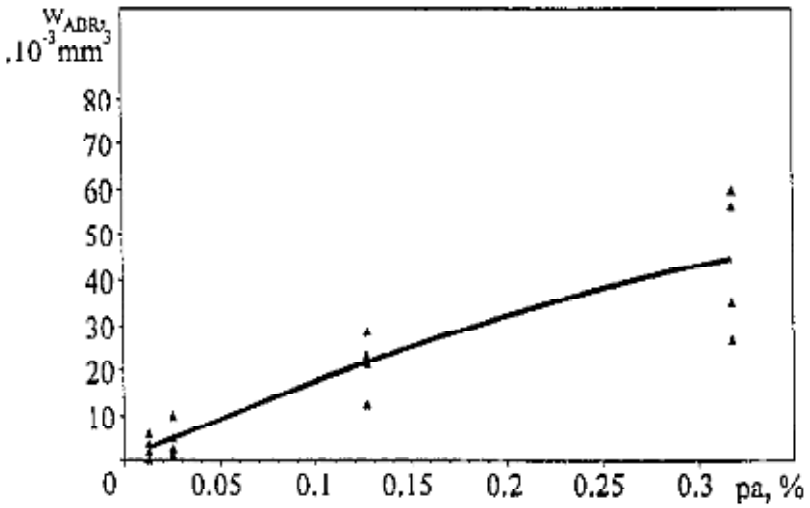


Figure 42 Plot of volumetric wear versus percentage reinforcement [66]. The amount of reinforcement (expressed as a proportion of the volume) is plotted on the x-axis. The y-axis is the material removed in cubic millimeters.

4.6.2.2 Variation of wear with reinforcement particle size

Another material property which may influence wear in machining of MMCs is the diameter of the reinforcement particles embedded in the base alloy. Marinov found that wear of the cutting tool is slightly accelerated for larger particle sizes, an effect attributed to the ability of larger particles to remove grain conglomerates from the tool surface [66]. A similar particle size effect is documented by Kishawy et al., who report that the presence of coarser particles impedes machinability [63]. Sahin and Sur considered two sizes of SiC particulate (29 μm and 110 μm): flank wear is more pronounced in tools used to cut the alloy containing larger diameter particles, a relationship which seems to be independent of process parameters [62]. Sahin and Sur suggest that the link between particle size and wear can be explained in terms of dislocation theory. In the case of fine SiC particles, the abrasive reinforcement is small enough to deform with the surrounding matrix. However, the matrix is unable to incorporate coarse particles into its deformation due to dislocations which occur in the vicinity of the abrasive. These dislocations are a product of work hardening which occurs when the matrix alloy is plastically deformed. When the matrix is hardened, the tool is unable to press the particles into the surrounding material. Sahin and Sur believe that the diversion of particles into the matrix that is possible when the abrasive is small has the effect of slightly reducing wear. Dislocations and work hardening associated with coarser reinforcements inhibit this wear reducing mechanism.

Not all researchers are in agreement regarding the effect of particle size on tool wear in machining of metal composites. Marinov reports that a few previous studies have found that particle size acts opposite to tool wear or that the contribution of particle

size is negligible [66,80,81]. The former result, in which larger diameter particles result in less wear, seems to be specific to the titanium stabilized workpiece material investigated in work by Hrustchov and Babitchev [81], However, the negligible correlation between wear and particle size observed by Faulring may be corroborated by Marinov, whose work does not preclude an effectively zero correlation between these variables [66,80]. The plot of abrasive wear versus particle size constructed by Marinov, shown in Figure 43, indicates that the relationship between the variables exhibits some degree of local linearity over a small portion of the y-axis range, though there is very little difference in the average wear mesaured for each particle size. Additionally, the non-trivial variation in wear for identical cutting conditions and particle sizes raises concerns about the repeatability of experiments in which the diameter of the abrasive is varied (for the 40 μm case, there is a difference in material loss of approximately 0.015 mm^3 between experiments performed at the same cutting parameters) [66]. It is thus essential to perform multiple experiments for each treatment to gauge the effect this variability may have on the perceived relationship between wear and particle size.

The range of particle sizes considered in the Marinov study (8 to 90 μm) is narrow in comparison to the spectrum of abrasive grain sizes used commercially (12 to 1400 μm) [66]. It may seem that the small window of particle sizes used in Marinov's work is insufficient to capture the effect of abrasive diameter on wear. However, the results of Kannan et al. indicate that wear is extremely sensitive to particle size. Kannan found that varying the diameter of the reinforcement by as little as 5 micrometers produced variations in tool flank wear on the order of 10^{-1} mm (Figure 44) [82]. The results of this study provide evidence that the variation in wear with particle size is a

detectable effect -- tools in contact with particle diameters corresponding to the extrema of grain sizes used in manufacturing applications (12 and 1400 μm) should exhibit large variations not only in the amount of wear but also the pattern of abrasive tracks left behind at the particle/tool interface. The importance of elucidating the effect of particle size in modeling wear should not be underestimated, as the relationship between these variables provides valuable information about both the wear mechanism and the mechanics of the particle/tool interaction which causes wear.

Based on the plot from Kannan et al. in Figure 44, the magnitude of the difference in flank wear between composites with 20 and 25 micron diameter abrasive appears to decrease with increasing cutting time. The change in the amount of wear with another process parameter (T) calls into question the magnitude of the influence of particle size on wear relative to other variables. At a cutting time of approximately 1700 seconds (28 minutes), the effect of particle diameter on wear becomes less pronounced. While further study is needed, the Kannan data indicates that the impact of particle size is greater at the onset of cutting. As the process continues, the time of contact T may overtake particle size as the primary factor which governs the amount of wear. The work by Kannan et al. suggests that particle size has significantly less influence on material machinability than percent reinforcement [82].

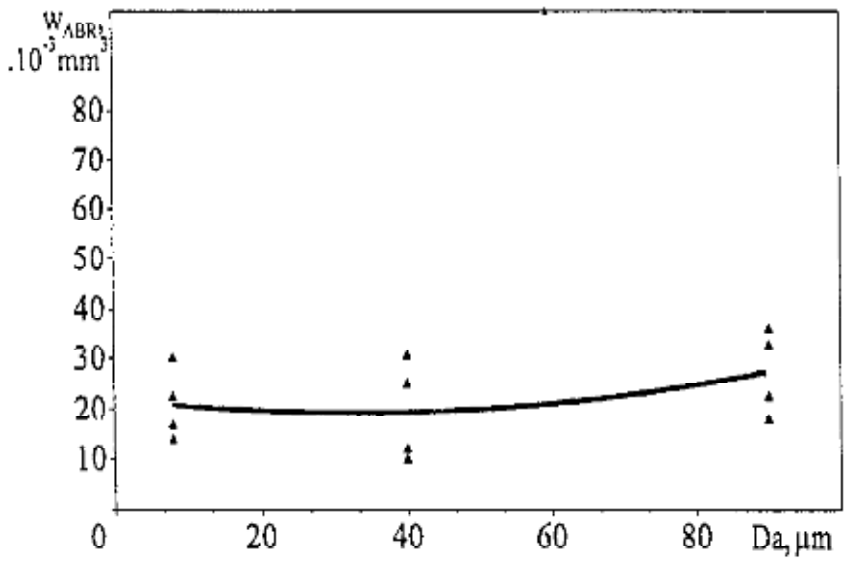


Figure 43 Plot of flank wear versus particle diameter from Marinov. The unique measurement associated with each experiment is denoted by a triangle and the solid trendline connects the mean value of the volumetric tool wear recorded at each particle size [66].

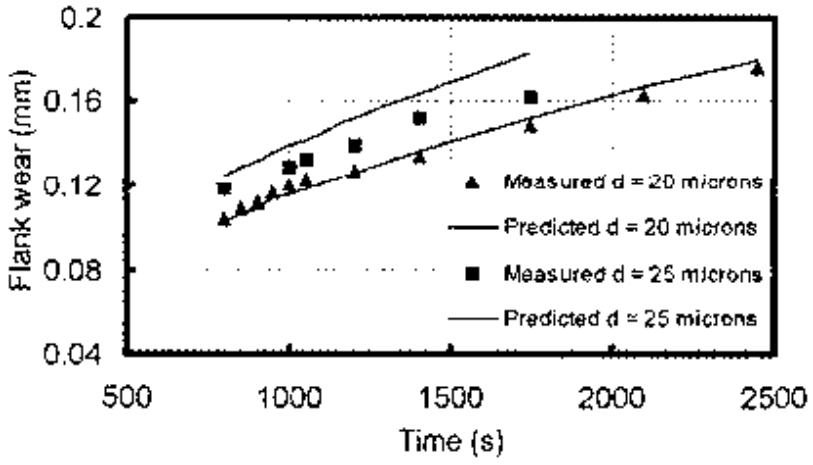


Figure 44 Plot of maximum flank wear versus cutting time for composites containing abrasive with diameters of 20 microns (denoted by the triangle) and 25 microns (square). The solid lines represent the values for flank wear predicted by the mathematical model developed by Kannan et al. [82].

4.6.2.3 Combined effects of percent reinforcement and particle size on tool wear

Published studies which attempt to isolate the effect of particle size and/or percentage reinforcement on wear are somewhat scarce. This may be because the effects of these variables on wear are generally believed to be coupled. The investigations of Marinov and Kannan et al. demonstrate that particle size and percentage reinforcement are in fact related and have a single, combined effect on the amount of flank wear incurred in cutting MMCs. Marinov found that wear in this process is directly proportional to n, the number of abrasive particles contained in the specimen being cut [66]. n is in turn related to particle concentration n^* (defined as the percentage reinforcement pa divided by the volume of a single abrasive particle V_p) by the relation $n = C_1(n^*)^{\frac{2}{3}}$, where C_1 is a constant specific to the geometry of the cutting tool [66]. The units for n^* are the number of particles per unit volume, or m^{-3}. The implicit presence of the n^* term in the experimentally determined wear relationship $w_{abr} \sim n$ indicates that abrasive wear is influenced by both particle diameter and particle size. An increase in particle concentration (which can be accomplished by either decreasing the size of the particles or increasing the reinforcement percentage) thus contributes to an increase in flank wear.

Li and Seah varied percentage reinforcement and particle size simultaneously to elucidate the degree to which changes in these variables affect the critical reinforcement percentage. Previously discussed in section 6.4.2.1, the critical reinforcement is defined as the percentage reinforcement which corresponds to the sometimes abrupt transition from low to high levels of wear. Li and Seah found that this critical reinforcement

percentage depends on particle size: for larger particles, the transition from the low to high wear regime occurs at a lower value of percentage reinforcement (Figure 45). This relationship can be explained in terms of particle interference. During cutting of MMCs, abrasive particles are able to remove tool material by carving out grooves along the face of the cutting tool. If the impinging particle (particle A) should encounter another particle (particle B) along this path, the ability of A to be dislodged by the motion of the tool (thus terminating the abrasive action) is impeded. A drawing of the particle's interaction with the cutting tool and a neighboring particle is shown in the right hand image of Figure 46 [79]. The effect of particle interference is an acceleration in the wear rate of the tool (the critical reinforcement percentage can be thought of as the percentage reinforcement at which particle interference effects begin to dominate). Particle interference is more likely to occur in the presence of larger particles, which require longer paths to dislodge. Since large particles also occupy a greater volume, a larger-diameter particle is also more likely to encounter a neighboring particle on its path along the tool face. For composites containing large particles, the particle interference phenomenon emerges at reinforcement percentages which are lower than those observed for composites reinforced with smaller particles (Figure 45). This explains why critical reinforcement percentage decreases with increasing particle size. Particle interference may be one reason why metal composites purchased from a manufacturer typically contain smaller particle sizes. According to Li and Seah, the selection of smaller particles permits an increase in the reinforcement percentage without an accompanying acceleration in wear of the cutting tools. In manufacturing composites which must be cut, Figure 45 (from Li/Seah) can serve as a machinability map, providing guidelines for

choosing particle sizes and reinforcement percentages which will limit material loss to the low-wear region.

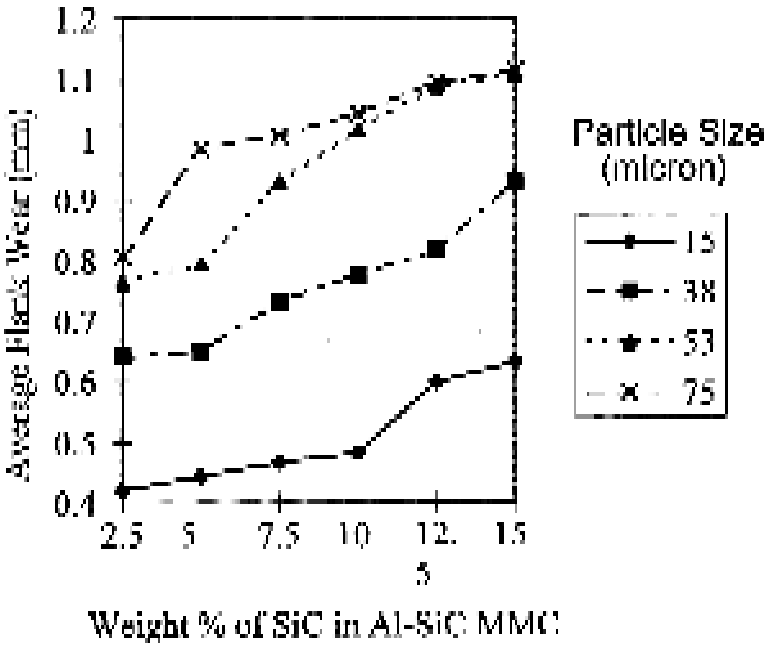


Figure 45 Plot illustrating effect of particle size and percentage reinforcement on flank wear and critical reinforcement percentage [79].

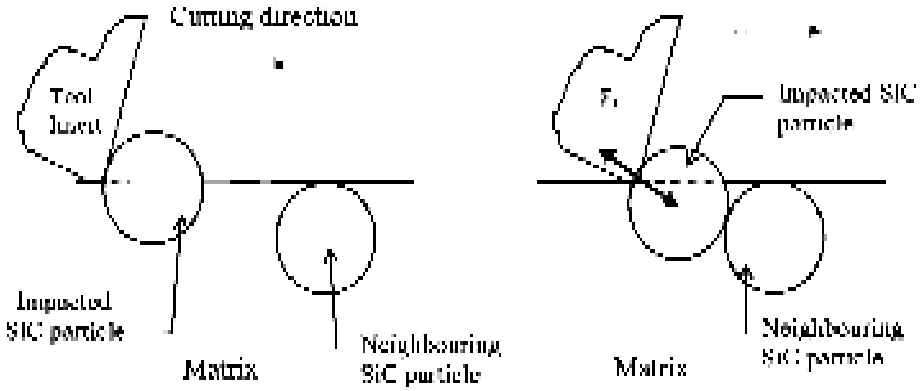


Figure 46 Drawing of particle interference. The drawing on the left represents the scenario in which particle interference does not occur, while the drawing on the right shows the interaction of a particle with both a neighboring particle and the tool face [81].

4.6.3 Hardness of tool material relative to hardness of reinforcement

As noted in sections 4.6.1 and 4.6.2, wear can be reduced through careful selection of process parameters and material properties (percentage reinforcement and particle size), but cannot be eliminated completely as long as there is a discrepancy in the hardness of the tool and the reinforcement. According to the classical principle of abrasive wear, abrasive action is only able to proceed when one of the two materials in contact is significantly harder than the other. Rabinowicz claims that the rate at which this wear occurs is proportional to the difference between the hardness of the reinforcement and the hardness of the abrasive [49]. Similarly, Marinov argues that the wear rate is governed by the ratio of the hardness of the abrasive reinforcement (H_z) to the hardness of the tool material (H_m). Marinov uses this metric to define the "abrasive capability" C_a (the probability that abrasive wear takes place) and set forth the conditions under which abrasive wear is possible in machining of MMCs [66]. Wear of the machine tool is only observed for combinations of tool and workpiece materials where $H_z > H_m$

(C_a = 1 for these cases since there is a 100 percent probability that some degree of abrasion will occur). Circumstances in which the hardness ratio is less than 1 forbid tool wear since the hardness of the tool is greater than that of the reinforcing material (hence $C_a = 0$). Marinov's hypothesis is that wear varies linearly with the hardness ratio when $H_z > H_m$. For a given temperature, an increase in the hardness ratio (accomplished by changing the reinforcement material) produces a proportional increase in the amount of volumetric wear [66]. Coelho et al. determined that the dependence of wear on this hardness ratio holds true across a range of tool materials (high-speed steel, tungsten carbide, diamond-coated tungsten carbide): the closer the ratio of the hardness of the abrasive reinforcement to the hardness of the tool lies to 1, the less wear is observed [77]. Research by Weinert and Konig likewise emphasizes the importance of the properties of the reinforcement material relative to those of the tool, declaring it to be *the* determining factor in the abrasive wear process [83]. While the amount of wear can be dramatically affected by changing one or both of these materials, it is generally easier to limit wear by modifying the tool, as the ceramics most commonly used as reinforcements all possess similar hardness values.

Silicon Carbide (the most common reinforcement material used in commercially manufactured metal composites) is harder than all materials used for machine tools with the exception of diamond and cubic boron nitride (cBN). Because it is expensive to fabricate entire tools from diamond or cBN, Tungsten Carbide (WC) is often chosen as a more economical alternative. While it is to be expected that a machine tool made of Tungsten Carbide (WC) used to cut a SiC reinforced composite would exhibit some wear (WC is only 75 percent as hard as SiC at room temperature), it shows significantly better

wear resistance than steel tools which are an order of magnitude softer. The relatively low expense of WC along with its improved wear resistance and ductility have made the use of WC drill bits and cutting tools for machining abrasive materials common practice.

The use of harder tool materials in machining of MMCs prolongs tool life and enhances performance. While WC represents a significant improvement over conventional steel tools in this regard, a study by Weinert and Konig concluded that WC was only effective at eliminating wear in the machining of Aluminum reinforced with Al_2O_3, a ceramic which is slightly less hard than WC [84]. As documented by Weinert/Konig and Coelho et al., WC tools exhibit substantial wear in machining of metal composites reinforced with SiC or B_4C [84,77]. For this reason, most researchers recommend that even harder tool materials, namely diamond or polycrystalline diamond (PCD), are essential for this application. Although these materials are significantly more expensive than WC, they currently represent the best means of eliminating tool wear in machining of ceramics or materials with a ceramic constituent. Harder tools possess the added benefit of lowering cutting forces; reduced cutting forces have the potential to improve surface finish as well as enable better overall control of the machining process.

For WC and polycrystalline boron nitride (PCBN) tools, wear additionally depends on the grain size of the tool material. In SiC-reinforced composites, WC tools with carbides composed of coarse grains fare better than those made with fine grains; the opposite is true for composites containing Al_2O_3 [83]. The reversal of the trend in wear with grain size which results from a change in reinforcement can be explained in terms of the cobalt binder phase (the inclusion of Cobalt is common in WC machine tools, as its presence improves WC's ductility and reduces the material's propensity for fracture

under mechanical stresses). Finer WC grains can be easily stripped from the tool by SiC of B_4C particles; the removal of these grains makes the tool more susceptible to abrasion. If wear occurs in machining an Al_2O_3 reinforced composite using a WC tool, it is because the abrasive particles are able to remove the cobalt binder phase through adhesion, thereby liberating the WC particles. Thus in the case of Al_2O_3, a finer WC grain structure actually inhibits wear. The observed relationship between wear and grain size for WC tools is also true for PCBN: when the reinforcement material of the MMC is harder than PCBN, coarse grains reduce wear [83]. Ding et al. have demonstrated that binderless PCBN tools exhibit better wear resistance than those which are alloyed with another metal [84]. While omission of the binder phase improves wear resistance for the reasons discussed in Wienert et al., its absence decreases the ductility of the tool (depending on the machining process and the magnitude of the stresses the tool will be subjected to, a significant reduction in ductility may be unacceptable). Based on these findings, grain structure may be an important consideration in adapting WC tools for FSW of MMCs.

Wear can be reduced by using harder tool materials which are binderless or have very fine grain sizes. Of course, the best means of eliminating tool wear in any process where abrasion is the dominant mechanism is to fabricate the tool from the hardest material available. Diamond, which occupies the topmost position on every hardness scale, embodies the best candidate for a tool material which will guard against abrasive wear. Concerns about wear, however, are nearly always tempered by economics: the inordinate expense of fabricating a solid tool from diamond drives the use of diamond coatings (apart from expense, the brittleness of single crystal diamond deters its use as a bulk part). The logic which underlies the inclusion of coatings is obvious: wear

properties of any candidate tool material (many of which are relatively inexpensive) can be improved by depositing a layer of diamond, the hardest known material, on its surface. Diamond coatings are usually deposited using chemical vapour deposition (CVD). CVD is a blanket term used to refer to a range of processes in which source gases (or in the case of direct liquid injection CVD, source liquids) are pumped into a chamber containing the substrate. A reaction is initiated which decomposes these material precursors to form diamond, silicon dioxide, or polysilicon, which then diffuses onto the exposed surface of the substrate (as in photolithography, portions of the substrate can be masked to create patterns) [85]. The rate of this diffusion can be accelerated by modifying the environment of the coating chamber; for example, a plasma enhanced CVD can reduce coating times and temperatures substantially, enabling faster coatings of larger areas and coating of materials with lower temperature resistances. Though still expensive, the economic efficiency of deposited diamond coatings has improved in recent years as a result of increased demand for these methods in industry (and the subsequent proliferation of configurations where large batches of parts can be coated simultaneously). Unfortunately, the successful implementation of coatings is complicated by mismatches between the coating and the substrate (the base material upon which the coating is deposited).

The best materials from a coating perspective, which possess a high degree of temperature resistance and can withstand the intense temperatures of the coating chamber, are often very brittle. For instance, diamond has a high affinity for both Molybdenum (Mo) and Silicon Carbide (although differences in the lattice structure of diamond and SiC can make coating the latter more challenging). While diamond adheres

readily to Mo and SiC, the inability of these materials to deform under stress causes them to fracture when subjected to loads associated with machining. Thus even though they are harder, these tools have significantly shorter lives than softer tools made of more ductile materials such as steel. Less brittle materials which are ideal for machining applications often do not meet the temperature resistance criteria for coating using deposition techniques. Depending on the specifics of the coating process, the temperatures the tool is subjected to in the coating chamber may be sufficient to induce melting of a substrate with low temperature resistance. Additionally, the formation of a strong bond between the coating and the substrate is critical in order to prevent delamination (the separation of the coating from its substrate which can occur during use of the tool). Ductile materials which have a high carbon content (such as steel) can only be coated with diamond by first depositing an intermediate layer, usually Chromium Nitride (CrN). While steel is cheaper and much less brittle than other candidate substrates, research has indicated that substrates with double and triple coatings are the most susceptible to delamination [62]. Microscopic examination of the tool post-use can diagnose this type of failure: for tools in which the coating has delaminated, chipping and/or peeling of the coating will be apparent. Delamination can be prevented by good adhesion between the coating and substrate as well as the careful selection of "speeds and feeds."

Negotiating a compromise between ductility, hardness, and temperature resistance in the selection of a substrate was previously discussed in Chapter II with regard to coatings of FSW tools; as noted in this section, many of the same concerns are present in selecting substrates for coated machine tools. Tungsten Carbide represents a good compromise between the desired properties of the base materials, the coating, and

the coating/substrate interface. WC possesses a fracture toughness that is more than adequate for the majority of machining applications and its resistance to fracture is superior to that of harder materials such as PCBN, PCD, Molybdenum, and Silicon Carbide [86]. The Cobalt binder that is alloyed with WC improves the ductility of the tool but ultimately impedes the deposition of diamond on its surface. This binder, typically present in amounts under 10 percent, can be etched from the surface of the substrate using ethanol (C_2H_5OH) [87]. Sibuki, Sasaki, et al. have shown that the adhesive strength of the coating/substrate bond is improved by preferential removal of the Co phase [88]. The performance of the coated cutting tool strongly depends on the chemistry of the adhesive bond; a strong bond translates into improved performance. Even after removal of the Co binder, additional surface treatments of WC prior to coating may be necessary due to microstructural mismatches between diamond and WC. Shibuki, Sasaki, et al. were able to obtain a quality diamond surface film through a two step surface pre-treatment: the substrate was first stripped of carbon via immersion in an H_2-O_2 plasma environment, then sprayed with diamond powder to increase the nucleation density and facilitate diffusion of diamond vapor onto the surface [88]. In commercial settings, the specifics of the surface treatment regime which the tool undergoes may be considered proprietary. For WC, such surface treatments are essential to ensure a favorable chemistry at the coating/substrate interface -- good adhesion at this location guards against delamination and lengthens tool life. Unfortunately, achieving a good diamond coating on WC substrates is often an iterative process, as the grain size and thickness of the resultant coating depend on the surface treatment, the conditions present in the coating chamber, and the duration of the coating process, among other factors. In

scenarios where an acceptable diamond coating cannot be obtained or proves too costly, another coating, such as Titanium Nitride (TiN) or Chromium Nitride (CrN) may be substituted in place of diamond.

There are several studies in the body of literature on tool wear in machining MMCs which assess performance of coated tools relative to uncoated tools. Coelho et al. compared the wear performance of WC and WC coated with TiN in the drilling of an Al/SiC composite [77]. Comparison of flank wear data for the tools reveals that the tool life of WC and WC/TiN (expressed as number of holes drilled before failure) is virtually identical for a given set of process parameters, a result which prompts Coelho et al. to suggest that the margin of improvement obtained was not great enough to justify the additional expense associated with coating. Though the quality of the hole was a peripheral concern, holes produced using the WC/TiN tool exhibited higher than acceptable values of surface roughness. The ineffectiveness of the coated tool in this application may be attributed to the selection of TiN as a coating for WC: coatings from the Nitride family, while cheaper, are typically less effective than diamond in preventing wear. In a similar study, Davim compared the performance of CVD diamond coated inserts and brazed PCD tools in cutting Al 356/SiC/20p. The CVD inserts had tool lives approximately 1/10 as long as those of the PCD inserts. The more rapid wear of the CVD tools in this instance cannot be attributed to delamination caused by abrasion, as optical microscope images of the tool show that the coating is still present along the majority of the clearance and rake faces of the tool even after the onset of wear [76]. Andrews et al. suggest that wear of the CVD tools is instead the result of adhesion, since surface irregularities in the coating make it easier for aluminum to adhere to diamond. When the

strength of the adhesive bond between diamond and aluminum exceeds that between diamond and the WC substrate, a portion of the diamond film is stripped away from the tool surface [89]. Figure 47 shows the erosion of the coating along the flank face caused by adhesive wear.

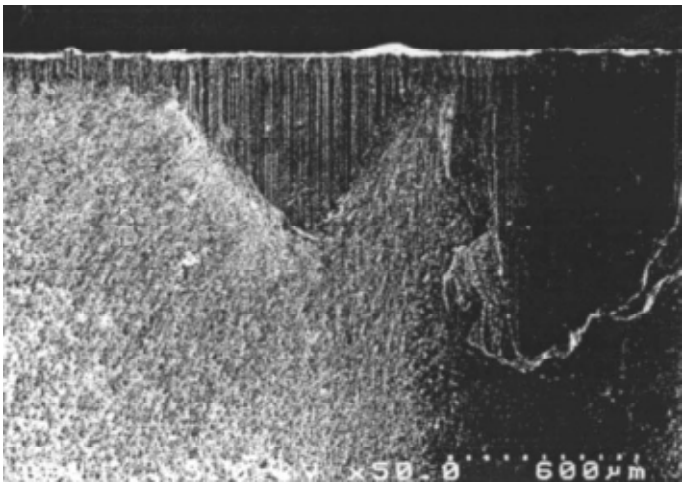


Figure 47 SEM image of flank face of CVD diamond coated WC tool after machining Al 380/SiC/20p. Portions of the coating have been eroded by adhesion of diamond to the aluminum matrix [89]

PCD tools are associated with the best overall performance in machining MMCs because the tool is monolithic – that is, the entire tool is made of a material which is harder than the abrasive reinforcement. The uniform composition of the tool means that delamination is not a concern. For diamond coated tools, the coating/substrate interface is the portion of the tool that is most vulnerable to wear. Once the coating begins to delaminate, the wear resistance of the tool degrades to the level associated with the substrate. The major advantage of coated tools over monolithic PCD is economical:

coatings present a cheaper means to improve wear resistance in scenarios where cost is prohibitive. The results of Andrews and Davim, however, emphasize the need for the development of coatings which guard against abrasion as well as adhesion [89,76]. To be effective at preventing wear in machining of MMCs, a coating must possess two characteristics: 1) the coating must be harder than the reinforcement to counteract abrasive wear and 2) the coating must strongly adhere to the base tool material, as a strong bond at this interface reduces the likelihood that the coating will be removed by adhesive wear. Oles et al. claim that high levels of adhesion in diamond-coated WC cutting tools can be obtained through treatment of the substrate surface prior to coating [90]. The pretreatment regime recommended by Oles et al. calls for processes which induce evaporation of the cobalt binder from the tool surface and increase the surface roughness (R_a) of the substrate [90]. Maintaining lower substrate temperatures during coating and accelerating the rate of deposition aid in creating a thin film with a morphology free of the surface irregularities which promote adhesive wear. Oles et al. found that WC tools coated using this method exhibited wear performance superior to PCD tools at 4 of the 6 turning conditions considered in their study [90]. The histogram in Figure 48 compares the tool life (in minutes) for PCD and diamond-coated WC tools over a range of speeds and feeds. Post-process micrographs of the tools diagnosed the failure mode of the coated WC tools as abrasion: the wear zone of the diamond tool shows no shows no signs of flaking, an indication that the strong adhesive bond between the coating and substrate was successful in delaying the onset of delamination.

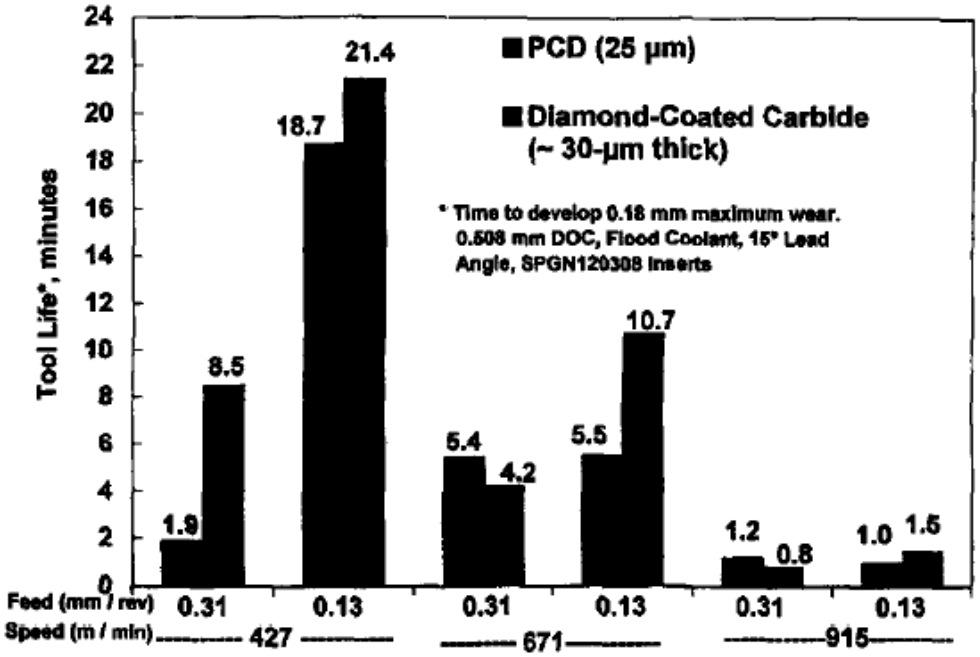


Figure 48 Histogram comparing tool life of PCD (grain size 25 μm and diamond-coated WC tools in turning of Al 359/SiC/20p for six process parameters. For each case, the PCD life is the left bar while WC + diamond is represented by the right bar. Tool life is expressed as the number of minutes required to reach the failure criterion (defined as 0.18 mm of flank wear) [90]

An auxiliary concern in the use of coated tools is their effect on the surface roughness of the finished part. In most cases, coated tools produce an inferior surface finish; additional machining is required to bring the surface roughness of the finished part into compliance with standards dictated by the end-user and/or the engineering application. Oles et al. maintain that the surface finish of the part is largely determined by the surface finish of the tool used to machine it [90]. Buffing the surface of the coated tool can decrease the roughness value at the tool surface to a level that is close to the R_a characteristic of PCD tools (according to Oles et al., buffing reduces the R_a of the coated tool by approximately 0.5 μm). Buffing has a dramatic effect on the quality of the

finished part: cuts made using coated tools which have undergone the buffing operation have R_a values which are equivalent or superior to cuts made with PCD tools [90]. The research of Oles et al. demonstrates that it is possible to produce durable coatings which are resistant to adhesive wear, prolong tool life, and produce parts with acceptable surface finishes [90].

The research summarized in this section emphasizes the use of harder tools in reducing or eliminating tool wear incurred in machining of MMCs. Though halting the progress of wear is critical in machining applications where tooling carries a high capital cost, minimizing tool wear has several ancillary benefits. Foremost among these is that the elimination of tool wear improves the quality of the machined surface. The workpiece surface typically becomes rougher as the tool features which facilitate material flow (such as threads or flutes) are eroded by abrasive wear processes; preserving the original tool geometry thus enhances the surface texture characteristics of the finished part; Davim found that the surface roughness of the workpiece is lowest ($R_a < 0.8\ \mu m$) for cutting conditions which correspond to reduced wear [91].

An important aspect of the work documented in this section is that the previously observed relationships between wear and process parameters hold true for harder tool materials. For instance, Coelho et al.'s findings that wear increases with cutting speed and decreases with feed rate agrees with much of the previous research discussed in section 4.6.1 with regard to wear of steel tools [77]. Similarly, Sahin and Sur observed that tool life for tools with multiple coating layers decreased with increasing cutting speed [62]. A common concern raised about the efficacy of statistical models is that their applicability may be limited to the specific experimental conditions used in their

construction. Observations of the same trends even when the workpiece and/or tool material is altered suggest that the development of a comprehensive predictive process model of wear in machining MMCs is feasible.

Perhaps a surprising result of the collected research on the wear process in machining MMCs is that the hierarchy of wear performance for various materials is not based solely on hardness. The study of wear in drilling of Al-SiC alloys by Coelho et al., previously referenced in the discussion of coatings, compared the wear resistance of high-speed steel (HSS), WC, WC coated with diamond, PCD, and diamond plated high-speed steel [77]. Using flank wear (VB_{max}) as a metric, Coelho ranked the materials as follows (from most wear to least wear): Diamond-plated HSS, HSS, WC, diamond-coated WC, and PCD. The diamond-plated HSS tool, which ranks among the hardest materials considered in the study, experienced the earliest failure. Discrepancies between the rankings of the tool materials in terms of hardness and wear resistance indicate that hardness is only one of many material properties which must be considered when selecting a material to combat wear: brittleness, fatigue life, and the strength of the adhesive bond between coating and substrate are equally as important. Though the use of harder tool materials (and coatings) is replete with complications, it is the most effective means of combatting wear in this application. The tool material may in fact be the most important factor in predicting the wear behavior of machine tools. Results of the Coelho et al. study, which considered the effect of cutting speed, feed, rate, depth of cut, and time on tool wear in addition to tool material, suggest that of these, tool material is the most statistically significant parameter which influences tool life [77].

4.6.4 Determination of the underlying wear mechanism

Based on the results of the studies discussed in section 4.6.1 – 4.6.3, a list of factors which have been shown to significantly influence machinability of MMCs can be compiled. These factors include process parameters (such as cutting speed, feed rate, depth of cut, machining time) as well as material properties of both the composite (percentage reinforcement, size of reinforcement particles) and machine tool (hardness, brittleness, adhesion of coating and substrate). The degree to which each of these variables influences the wear process can provide important clues to the wear mechanism(s) that are responsible for the loss of tool material. In studies related to wear in machining of MMCs, researchers often assume that wear is abrasive. This is a reasonable assumption, since the tool is in most cases much softer than any of the ceramic materials commonly used as reinforcement. The observed relationships between volumetric wear and several of the aforementioned factors are also consistent with abrasion. Though the wear mechanism in this instance may seem obvious, clear identification of a root causal agent for wear is essential to understand and model the fundamental mechanics of the wear process. This especially important aspect of the study of wear is often neglected – even in scenarios where abrasive wear is the likely cause, it is necessary to verify the mechanism before attempting to model the process. Fortunately, identification of the wear mechanism can be accomplished through a simple post-process evaluation of the tool surface.

As discussed in section 3.2, each type of wear leaves behind a distinct microscopic signature. Since multiple wear mechanisms can occur simultaneously, tribology is necessary to decouple the effects of each wear mechanism in situations where

more than one process acts on the tool. Figure 49, taken from Kishawy et al., is a scanning electron micrograph of a worn tool used to cut Al 356/SiC/20p [63]. The grooves shown in the close-up image of the tool surface (inset) are clear indicators of abrasive wear. The directionality of the grooves is used to further classify the wear mechanism as two or three body abrasion. Since the grooves are parallel to the cutting direction, two body abrasion is determined to be the dominant cause of wear. A similar diagnosis of the wear mechanism in machining of metal composites was made by Sahin/Sur, Davim, and Weinert/Konig in their respective research [62,77,83]. There is some indication in Davim's study that the wear process may have an adhesive component, particularly in the case of tools with coatings (discussed in section 4.6.3) [76]. The micrographs taken by these researchers indicate the prominent presence of adhesion in wear of PCD inserts [76]. Based on the body of literature on wear in this application, it can be reasonably argued that adhesion is only observed for tools which are fabricated from diamond or have a diamond coating. In the case of tool materials which are softer than the reinforcement, adhesion is a relatively small contributor to wear in comparison to abrasion; in these instances, the proportion of wear which can be attributed to adhesion is small enough that the wear mechanism can be considered purely abrasive. Additionally, coated tools and/or those fabricated from brittle materials are vulnerable to fracture (a form of fatigue wear). The sum of these observations is the caveat that the wear mechanism in many cases depends on the tool material as well as the material properties of the workpiece. Researchers should exercise caution in evaluating wear processes. As the studies referenced above indicate, the presence of abrasive particles in the workpiece material does not guarantee that abrasion is the primary mode

of wear. Rather, the effectiveness with which a particular wear mechanism acts on the tool is dependent on many factors.

In some cases, it may not be possible to investigate the wear mechanism using microscopy due to the size and/or shape of the worn part. While methods which do not employ microscopy are less definitive at identifying the wear process(es), they become necessary when the part being examined is too large (even when trimmed to a representative section) to fit in a microscope chamber or has features which make analysis exceedingly difficult (for instance, characterizing the wear pattern inside a hole or crevice). An alternative means of assessing the wear process in these instances is to gauge the response of the system when specific conditions are imposed. For example, if abrasive wear is present, the presence of cutting fluid will increase the wear rate (since the presence of larger wear particles in sliding systems actually impedes further abrasion, removal of this material renders the abrasive wear mechanism more effective). Abrasive wear is also very sensitive to fluctuations in humidity. The wear rate increases dramatically with increasing humidity, an effect which can be attributed to the moisture in the air acting as a lubricant [49]. These are just two examples of system responses which are uniquely associated with a particular wear mechanism. The ability of a specific condition to elicit (or fail to elicit) a change in the system behavior can be used to identify (or eliminate if no response is observed) one or more of the possible wear processes. This method of classifying the wear mechanism is indirect and should be used as a substitute for tribology only in rare cases where microscopic examination is not possible. It can, however, provide confirmation of a dominant wear mechanism; analysis

of system response in this manner may prove especially useful in situations where there is microscopic evidence of multiple wear mechanisms.

Clear identification of the wear mechanism facilitates modeling of the wear process. For instance, if the wear mechanism is known to be abrasive, then it is safe to assume that the wear process is independent of temperature (thus permitting researchers to disregard thermal effects). Knowledge of the wear mechanism can also provide a preliminary idea of the relative importance of variables which may impact the amount and/or rate of wear. For abrasive wear, the hardness of the tool relative to the reinforcement (the subject of section 4.6.3) is typically the most important factor in predicting the amount of wear which will occur in the machining process. Further classification of the mechanism (for example, as two-body abrasion or three-body abrasion) conveys additional information regarding the expected magnitude of the wear constant (represented as k in many of the classical wear equations presented in chapter III) as well as the anticipated relationships between wear and properties of the workpiece material. The wear mechanism often provides a clue to the best means of combatting wear in a particular process. In the case of abrasion, the most effective means of reducing tool wear is the use of materials which are harder than the substance causing the abrasion (a strategy in accordance with Rabinowicz's classic principle of abrasive wear). There is a wealth of information which can be discerned in large part just by identifying the wear mechanism; this simple analysis thus has the potential to substantially decrease experimental time.

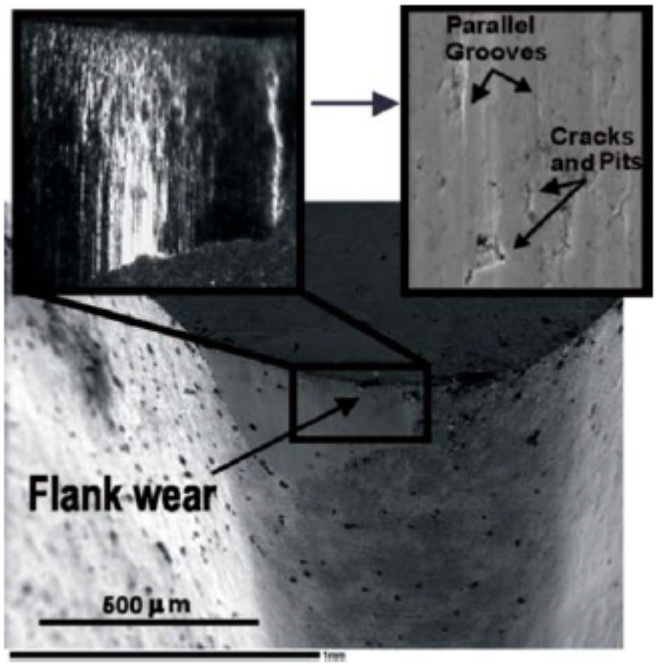


Figure 49 SEM image of flank surface of cutting tool used to machine Al 359/SiC/20p. The parallel grooving pictured in the inset is characteristic of two-body abrasion [64].

For most research into machining of MMCs (such as studies which seek to characterize the variation in tool wear with process parameters), identification of the wear mechanism is adequate; as such, further characterization of the wear pattern is typically not pursued. However, qualitative assessments are insufficient if we hope to accurately capture the physics of the wear mechanism. An investigation of residual wear tracks on the tool surface which goes beyond simple identification could provide potentially valuable insights into the mechanics of the wear process. According to Sahin and Sur, the abrasive wear pattern observed in tools made of softer, less expensive materials (such as steel of WC) is at least partially governed by the ability of the particles to plastically deform with the surrounding matrix [62]. When conditions are such that the matrix cannot accommodate the deformation of the reinforcement, the abrasive presses into the

tool and carves out a groove along the tool's surface. The wear pattern which results from the interaction of the tool with the reinforcing particles in this manner is defined by the number, spacing, and depth of the parallel grooves left by the abrasive particle. Measurements of variations in the height of the flank surface (obtained using contact profilometry or a similar method) can be used to calculate an average value for surface roughness, R_a. The surface roughness value of the tool in some cases have been shown to correlate with the amount of volumetric wear – this is investigated by Davim in references 53 and 91 . The distance from the surface of the tool to the valley of the profile over a specified sampling length (denoted as R_v) as well as the mean depth of the valleys (R_{vk}) can be used to estimate the depth of the grooves. If machining conditions which produce deeper grooves and/or grooves which are spaced closer together (spacing is given by the surface texture parameter R which measures the distance between successive valleys) also give rise to greater volumetric wear, then it may be possible to model the wear process in terms of the surface asperities (not unlike the conical simplication of abrasive wear presented by Rabinowicz in reference 49). The tribology-based approach to the study of tool wear in MMCs recommended here is used to collect wear data in the experiments of Chapter VII, which are designed to isolate the effects of reinforcement particle diameter and weld pitch on the wear pattern in FSW of metal composites. Though investigations of this nature require specialized equipment and detailed analysis, they pay dividends in the form of insight gained into the fundamental physics which underlie the wear process.

4.7 Summary

Tool wear is one of the most extensively studied aspects of machining MMCs. Approximately thirty percent of journal articles on machining MMCs published between 1991 and 2010 concern predicting, modeling, and combatting tool wear. The plethora of research available on the wear of machine tools in MMCs stands in stark contrast to the scarcity of literature devoted to tool wear incurred in frictions stir welding of these materials. Apart from the few preliminary studies which appear in references 32-45, welding of MMCs represents largely uncharted research territory. As discussed in chapter II. welding remains a prohibitive concern for the use of MMCs in industrial applications: a solid-state joining process has the potential to solve the problems of weldability which are inevitably encountered when integrating MMCs into larger structures. Results from investigations into the variation of wear with material properties and process parameters in machining of MMCs are examined in this section as a precursor to related work on FSW of MMCs presented in chapters V-IX. The material properties which are pinpointed in the literature on machining as variables which influence wear (percentage reinforcement, particle size, hardness of machine tool relative to hardness of reinforcement) can easily be incorporated into experiments designed to study wear FSW of MMCs. In most cases, the process parameters considered in machining do not have a direct analog in FSW. However, the argument can be made that the resemblance between certain machining parameters and FSW process variables is strong enough to enable comparisons. For instance, cutting speed is analogous to rotation speed in FSW, the feed rate in machining is roughly equivalent to traverse rate, and the depth of cut is akin to the plunge depth of the FSW tool. Thus the factors examined in

the research on machining MMCs may also prove useful for predicting wear in FSW of these materials.

Perhaps not surprisingly, many of the same relationships between wear and material properties are observed for both processes (FSW and machining operations such as cutting, drilling, or turning). The results documented in chapter VI illustrate that the amount of wear in FSW of composites increases with increasing percentage reinforcement and larger particle sizes, a result nearly identical to those summarized in sections 4.6.2 for machining. Chapter VII demonstrates that wear is reduced when harder FSW tool materials are used, an effect also observed in investigation of how machine tool wear in composites varies with the hardness ratio of tool to reinforcement (section 4.6.3). Wear in FSW of MMCs increases with rotation speed (Chapter V), a trend similar to the increase in wear with cutting speed noted in section 4.6.1. The similarities in these results lend credence to the argument that methods used to model wear in machining can also be applied to study and predict wear in FSW.

The broader literature survey presented in this chapter is not only intended to pinpoint several important factors which influence wear in machining of MMCs (which serve as a guidepost for our similar investigation of wear incurred in welding these materials), but also to introduce the methodologies commonly used to analyze wear processes in industrial settings. Apart from a few preliminary studies by Prado et al. on the variation of wear within a narrow range of process parameters, there are no existing studies which use experimental design and/or statistical analysis to efficiently determine the conditions which minimize wear in FSW of MMCs. Chapters V and VI (which focus on the variation of wear in FSW with process parameters and material properties,

respectively) apply the design of experiments (DOE) methods discussed in this chapter (factorial design, Taguchi, and response surface methodologies) to the study of tool wear in FSW of composites. These experimental designs have been proven as effective tools in the characterization of wear processes, as evidenced by the many references cited in section 4.6.1-4.6.3 which utilize them. Taking a cue from research into wear in machining, we begin our own investigation of wear by studying the dependence on wear on the three major FSW process variables (traverse speed, rotation speed, and length of weld joint), using Taguchi method in conjunction with multivariate regression to create a machinability map which can be used to predict the volumetric wear of the FSW tool over a range of process parameters. As was the case for machining, results of this research can subsequently inform physics-based modeling of the wear process (Chapter VI-VII) as well as aid in the design of wear resistant geometries and/or selection of materials which reduce wear (Chapter VIII).

CHAPTER V

VARIATION OF WEAR WITH PROCESS PARAMETERS IN FRICTION STIR WELDING OF METAL MATRIX COMPOSITES

In chapter III, the investigation of wear processes was explained using the metaphor of a doctor treating a patient who is ill. This metaphor functions as an explanatory schema, describing the key concepts which underlie the study of wear in a way that is familiar and accessible to the common reader. We will return to this metaphor throughout the results and analysis chapters of the dissertation (chapters V-VIII), as it fits with our perspective on wear and introduces important ideas succinctly and understandably. The medical metaphor is an apt one since the techniques used by tribologists and material scientists to diagnose wear are analogous to those employed by medical professionals when a patient initially presents with a disease.

As regular viewers of the television show *House* can attest, the curmudgeonly title doctor's preferred protocol for diagnosis often centers on the study of causal phenomena through performance tests [92]. Treadmill ambulation (also known as a stress test, in which the patient walks or runs on a treadmill while sensors monitor heart rate, blood pressure, and respiration) and tilt table testing (in which the patient lies on a table that is tilted from the flat position to vertical, a technique used to identify the presence of conditions which are triggered or exacerbated by changes in position) are two common examples of performance testing [92]. The purpose of these tests is to gauge the response of the patient's body systems to stimuli – the manner in which the patient reacts can

provide valuable clues to diagnosis and/or indicate the most effective course of treatment. The wear experiments detailed in this chapter are direct analogs to the performance testing used in medical diagnosis. Just as performance tests are used to identify the conditions which aggravate or minimize a patient's symptoms, these wear experiments are designed to characterize how wear varies with process parameters and identify the range of process variables which correspond to increased or decreased wear. While it may be tempting to assume that wear is abrasive (perhaps based on the adage "when you see hoofprints, look for horses, not zebras"), the material interactions in FSW of MMCs may be different from those documented in machining. Although an abrasive wear mechanism is the most likely causal agent since it fits with our intuitive understanding of the interaction between the weld tool and the harder reinforcement particles, a full investigation is needed to 1) identify the wear mechanism(s) (which may be abrasive, adhesive, fatigue-related or some combination thereof) and 2) determine which factors (either process variables or material properties) contribute to the degree of wear.

As discussed in chapter IV, there is a substantial body of literature on tool wear in machining of MMCs. There are comparatively fewer published studies on FSW of MMCs: of these, most focus on identifying characteristics of MMC joints and/or parameterization of the joining process. Hardly any studies focus specifically on tool wear in FSW of MMCs. The scarcity of published work on this topic represents a critical gap in the literature: understanding the wear problem in FSW of MMCs is key to expanding the usability of these materials to applications (such as aerospace structures) where they would be of maximum benefit.

Storjohann et al. found that fusion-welded MMC joints were characterized by porosity in the heat-affected zone, the dissolution of reinforcing particles, and the formation of a deleterious theta phase (Al_4C_3) caused by localized melting [32]. These precipitates, which appear on a microscope image as needle-like formations (Figure 11), tend to segregate into the molten matrix, leaving a phase-depleted, low-strength region at the joint interface in their wake. Storjohann et al. found the decomposition reaction which results in the formation of Al_4C_3 ($4SiC + 3Al \rightarrow 4Si + Al_3C_4$) to be highly temperature-dependent; the proportion of this precipitate present in the finished joint is directly proportional to heat input. A lower-heat, solid-state process such as friction stir welding is the preferred method for joining MMCs primarily because it precludes formation of precipitates and the accompanying low-strength region which results from particle segregation; FSW is able to preserve the properties of the parent material to an extent that is not possible with fusion welding.

FSW of MMCs, however, is not without its own set of unique problems: chief among these is rapid and severe wear of the welding tool, an undesirable consequence of contact between the weld tool and the comparatively harder reinforcement particles. Although the physical mechanism underlying tool wear in FSW of MMCs is not well-understood, previous experimental studies have attempted to characterize the effect of process parameters on wear. An investigation published by Prado et al. provides a preliminary assessment of the wear of cylindrical threaded tools in the butt welding of Al 6061/Al_2O_3/20p [42-43]. Prado et al. observed that the most dramatic wear coincided with higher rotation speeds and lower traverse rates. Data reported in an analogous study by Fernandez and Murr, which considers butt welds of Al 359/SiC/20p supports a similar

conclusion (a more detailed discussion of this research can be found in the literature review in chapter II) [44]. The research presented here is not an attempt to reproduce or verify the results of previous work, but instead to establish a predictive model of the wear process in FSW of MMCs (which, to the author's best knowledge, does not exist in the literature). While there are no existing models (statistically based or otherwise) for wear in FSW, these types of analyses abound in studies of tool wear in machining. The previous chapter summarized results of studies on wear processes in the latter; the methods used to construct predictive models for machine tool wear serve as both a precedent and guide for our own research into wear phenomena in FSW.

The aforementioned work by Prado et al. and Fernandez/Murr investigated the dependence of wear on rotation speed and traverse rate, widely regarded as two of the most influential process parameters in determining weld quality. While there are some inconsistencies among their data, overall the observed trend seems to be that wear increases with rotation rate, but may decrease slightly (depending on the parameter values) with traverse rate. The applicability of these studies is somewhat limited by the small range of process parameters considered in the experiments; all welds were performed at rotation speeds equal to or below 1000 RPM. Additionally, neither study provides an assessment of the relative strengths of the correlations between wear and each process variable. Although the authors classify the relationship between wear and each parameter as either direct or inverse, they do not speculate on the physical phenomena which may explain these dependencies. Together, the limitations of these studies, the scarcity of published literature on wear in FSW of MMCs, and the potential of this research topic to extend the applicability of metal composites highlight the need

for a robust study of wear phenomena in this process. The research in this dissertation builds upon the preliminary investigations of Prado et al. and Fernandez/Murr, going beyond their work to construct models of wear behavior and explain the physics of the underlying wear mechanism(s). Collectively, our research provides insight into wear which will allow researchers to understand, predict, and more effectively combat the wear processes which are detrimental to performance of FSW joints.

The first phase of this work (and the subject of this chapter) lies in characterizing the variation of wear with process parameters. The focal point of this study is a predictive model (derived using multiple regression techniques) that estimates the volume loss the tool will experience during a weld based on the values of three major process parameters. This model is similar to the machinability maps used by machinists to identify parameters for machining of abrasive materials which minimize wear yet still produce a cut, hole or other feature with acceptable quality. The regression equation and the associated dimensionless group (which can be used to scale results) are intended to inform parameter selection in FSW of MMCs, enabling operators to select parameters which yield optimal results with regard to both wear and quality. The material presented here is the first systematic study of wear in the FSW literature.

5.1 Experimental Procedures

5.1.1 Materials

The material selected for this study was Al 359/SiC/20p, an off-the-shelf, commercially available composite purchased from mc21, a metal composites manufacturer and supplier based in Reno, Nevada. The designation Al 359/SiC/20*p* indicates that the matrix alloy is Aluminum 359, the reinforcement material is Silicon Carbide, and that the SiC inclusions are in the form of particles (*p*) rather than fibers (*f*). Al 359 is a cast alloy (Aluminum alloys with the four digit designation are wrought). Cast alloys typically have a higher silicon content than their wrought counterparts as well as superior corrosion resistance. The reinforcement percentage is 20% by volume. Most commercial composites are manufactured using the stir casting technique discussed in chapter II, wherein particles are funneled into the molten base alloy through drill holes and mechanically stirred into the matrix by an impeller housed inside the crucible. Once mixing is complete, the contents of the mixing chamber are transferred to a mold and solidified [93]. mc21 uses a variation of this technique known as the rapid mixing concept (depicted in Figure 50). In rapid mixing, reinforcement particles are introduced into the matrix through a hollow shaft which extends below the surface of the molten metal. The mixing head positioned underneath the shaft is specially designed so that its motion does not induce vertical flow; this permits operation at faster rotation speeds which promote shearing of the material, reducing mixing time. According to mc21, composites produced using this technique have less porosity and a more uniform reinforcement distribution than composites made using conventional stir casting methods. Since rapid mixing occurs in ambient air (stir mixing takes place in a vacuum chamber),

its associated capital costs are lower: the at-cost price of Al 359/SiC/20p is listed as $2.20/kg. The majority of commercial MMCs produced are the form of flat plates. The material used in our work on FSW of MMCs comes from a surplus of ¼" thick Al 359/SiC/20p fabricated for an aerospace industry client. These plates were sheared into custom strips measuring 1 ½" in width and 8" in length for our application. For joining, two of these strips are positioned adjacent to one another along their length in a typical butt weld configuration.

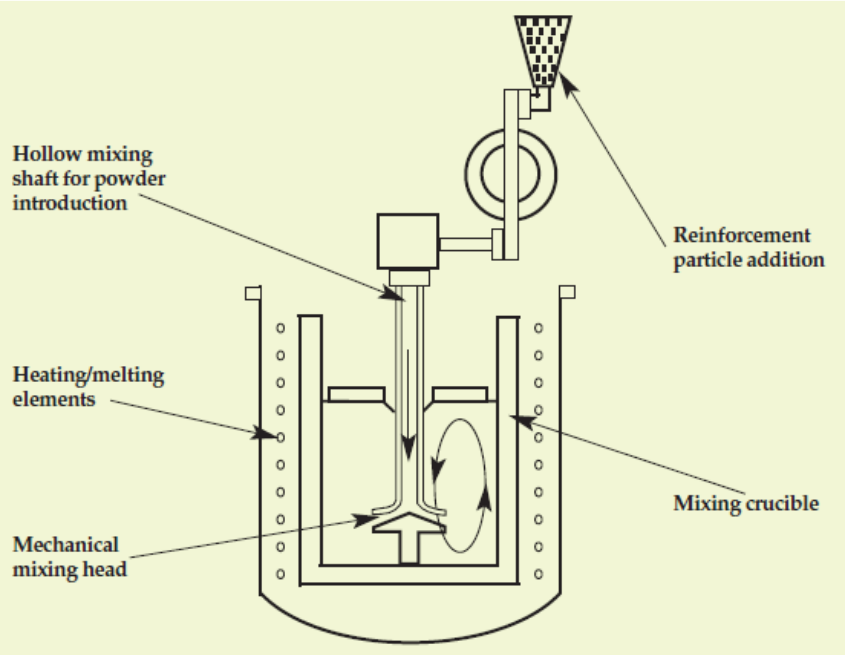


Figure 50 Schematic of rapid mixing technique developed and patented by mc21, Inc. [93]

The reinforcement particles in mc21's Al-SiC class of composites are classified as F500 on the FEPA scale (a detailed explanation of the FEPA scale can be found in chapter VI). A F500 grade means that 95% of the particles in a batch have diameters

which do not exceed 12 microns ($4.7x10^{-4}$ inches). An optical microscope image of a transverse cross-section of the unwelded MMC (referred to as the "parent" material) reveals the irregular shape of the SiC inclusions (the dark areas visible in Figure 51). The particles are approximated as spherical to simplify their characterization -- the FEPA scale classifies abrasive according to the distribution of particle diameters. Given the hardness and angular features of the particles (which resemble microcutters), it is not surprising that they are responsible for high rates of wear.

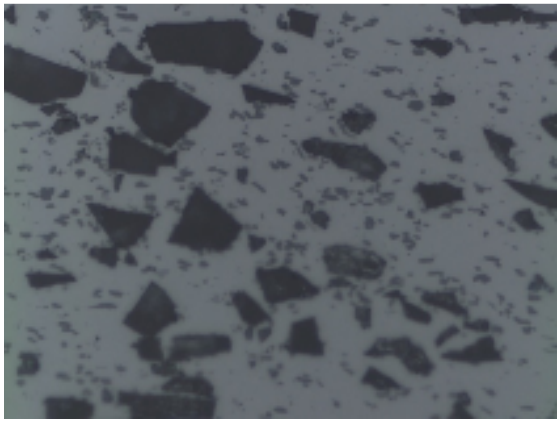


Figure 51 Optical microscope image of etched transverse cross-section of Al 359/SiC/20p. Magnification: 100X. Image taken using laboratory facilities at NASA Marshall Space Flight Center.

The inclusion of the reinforcement significantly increases the strength of the Aluminum alloy. Since particulate reinforcement is introduced into the molten matrix as a powder, this enhancement in strength is unaccompanied by a proportional increase in weight. Aluminum-MMCs thus possess strength to weight ratios which are much greater than conventional alloys. Table 6 (adapted from a study by mc21) compares properties for alloys traditionally used in tank armor (Al 5083, Titanium, and rolled homogeneous

armor – RHA) with an Al-MMC reinforced with 30 and 45 percent ceramic, respectively. Properties reported for Titanium, RHA, and MMCs are typical rather than specific. The properties listed for MMCs are average values for the range of metal composites produced by mc21. While it is clear that MMCs cannot compete with titanium and the rolled homogeneous armor used for tanks in terms of strength, the specific modulus (Young's modulus normalized with respect to density) for either category of MMCs is at least 1.5 times greater than any other material in the table.

Table 6 Comparison of properties of typical MMCs with other alloys []

	MMC 30	MMC 45	Al 5083	Titanium	RHA
Density (g/cc)	2.8	2.9	2.66	4.51	7.5
Young's modulus (GPa)	115	175	75	110	200
Tensile strength (MPa)	325	180	95	730	690
Specific modulus	40	60	28	24	27

Addition of reinforcement to an aluminum alloy produces an effect similar to doping in electronics. When impurities (dopants) are introduced into a semiconductor, the presence of these additional electron acceptors (or donors) alters the material properties (the change in properties is proportional to the amount and type of dopant added). Just as electronics can be doped to make them more conductive, the addition of hard reinforcing particles to Aluminum strengthens the base material. Inclusion of the reinforcement material has little effect on weight (which explains the high values of

specific modulus and specific strength observed for metal composites). At the extreme end of the doped circuit spectrum are semiconductors which behave as conductors, materials referred to as degenerate. The type and percentage reinforcement in MMCs can be similarly manipulated to create materials which display properties (such as fracture toughness, hardness, and temperature resistance) far exceeding those of the matrix material: Kunze and Bampton report that Aluminum alloys reinforced with 70 percent SiC possess an elastic modulus comparable to that of steel [29].

When selecting a metal composite for an application, it is important to be mindful of the tool wear which may occur during machining and/or welding. Studies in the machining literature on MMCs consistently report that wear increases with percentage reinforcement. According to Diwan, this trend also holds true for friction stir welding of MMCs [47]. Diwan argues that the weldability of a metal composite is directly proportional to its percentage reinforcement. Diwan's research classifies materials into one of five categories based on the amount of reinforcement, ranging from category I (less than 10% reinforcement, characterized by comparatively mild and gradual wear of the tool volume) to V (greater than 50% reinforcement, results in tool fracture). Diwan characterizes any material with a reinforcement percentage exceeding 40% as unweldable. Concerns about wear thus impose an upper limit on the amount of ceramic reinforcement which can be introduced into a metal alloy (which in turn limits the degree of strength enhancement which can be attained). In MMC material selection, care must be taken to negotiate a compromise between strength and machinability/weldability. The objective of composite material design is thus to increase the strength of the base alloy yet ensure that the addition of ceramics which enhance the properties do not render the

resulting material unweldable (or unable to machined). For the wear study presented in this chapter, an MMC with 20% Silicon Carbide reinforcement was selected; this level of reinforcement results in properties which represent a significant improvement over the unreinforced material. Table 7 compares the material properties of an Al 359 unreinforced alloy with its MMC counterpart (properties of the MMC are theoretical and were calculated using the law of mixtures). The reinforcement percentage is small enough to categorize the material as class II using Diwan's classification scheme; while class II MMC materials cause rapid wear of steel tools, they are still considered weldable.

Table 7 Properties of Al 359 and Al 359/SiC/20p [6]

	Al 359	Al 359 w/ 20% SiC reinforcement*
Density (g/cc)	2.67	2.8
Modulus of Elasticity (GPa)	72.4	140
Specific modulus	27.1	50
Specific Heat Capacity (J/g-C)	0.963	0.9
Thermal Conductivity (W/M-K)**	138	125.9
Hardness (Knoop)	114	691.2

*Properties listed for Al 359/SiC/20p are theoretical(estimated based on the law of mixtures).

**Thermal conductivity represents value at 400 C.

5.1.2 Experimental Apparatus

The experiments in this study were performed using a Milwaukee #2K milling machine modified for friction stir welding. A photograph of the entire apparatus with labeled components appears in Figure 52. The milling machine was retrofitted with three motors (traverse, lateral, and vertical) which control movement of the stage in the x, y, and z directions, respectively; a separate motor (the 20 HP motor located above the vertical head) controls spindle rotation. The vertical head attachment functions as a reinforcing mechanism which prevents vertical motion (axial forces during welding are typically on the order of thousands of pounds). The motors are collectively controlled through a general user interface (GUI) called *Weld Controller*. This program communicates with the motor controllers, enabling active user control of three major process variables: traverse speed, rotation rate, and plunge depth.

Prior to welding, the sample is clamped to ensure that it remains stationary under the forces associated with the FSW process. For butt welds, the sample consists of two pieces, each measuring 8" long, 1.5" wide, and 0.25" in thickness. The pieces are aligned along their length and clamped into place using bolts (tightened to 50 N-m) on either side of the joint. The tool is inserted into the spindle and fastened in place by means of a double set screw (the screw contacts a 1/4" square flat machined into the tool shank). The desired position of the tool is set prior to welding via an automatic zeroing routine. The zero location refers to the vertical position of the tool when the surface of the shoulder just contacts the top of the workpiece material (this location coincides with a spike in the axial force that can be detected by the sensing apparatus). The plunge depth specifies the tool's position when it is fully engaged in the workpiece (for instance, a

plunge depth of 0.007" means that the tool is advanced this distance beyond the zero location). The weld procedure consists of three stages which occur sequentially:

1) Plunge/dwell. In the plunge/dwell step, the tool rotates in a stationary position as the vertical motor raises the stage. Vertical movement of the stage ceases when sufficient contact between the shoulder and the surface of the workpiece material has been established. Once the tool is fully engaged in the material (i.e. has reached its plunge depth), it remains stationary for a few seconds. Referred as the "dwell time", this period of stasis ensures that the material is plastically deformed before welding begins. If the material at the beginning of the joint is plasticized prior to welding, the tool is less likely to fracture when traverse movement of the stage is initiated.
2) Welding. The welding phase is defined by advancement of the tool along the jointline. The traverse motor propels the stage at a constant feed rate; traversal continues until the tool reaches the user-specified end location. With each rotation, the tool sweeps material from the advancing side of the joint and deposits on the retreating side. As the tool advances, the material behind the tool recrystallizes and consolidates to form a weld.
3) Extraction. The end of the weld typically coincides with the end of the jointline. At this point, the stage is gradually lowered until the tool has completely disengaged from the workpiece (the tool continues rotating during extraction). Spindle rotation ceases once the stage has reached the change height (the height at which there is sufficient distance between the tool and the workpiece to permit changing of the sample and/or tool).

While the automatic welding routine outlined above is specific to the apparatus used in the Vanderbilt University Welding Automation Laboratory, the general procedure is very similar to that used for industrial frictions stir welding robots (such as the iStir PDS used for process development and testing at NASA Marshall Space Flight Center) [95]. In the iStir PDS and the VUWAL robot, the tool remains stationary while the workpiece traverses. This is by far the most common configuration. However, there are applications (such as FSW of ships, automobiles, and space vehicles) for which translation of the workpiece is impractical. These applications rely on robots, such as the ESAB Rosio, which have an articulated arm for FSW [96]. In assembly, NASA uses articulated arm robots custom-built by Materials Testing Services (MTS) for FSW of the Ares I upper stage. These robots have a much larger workspace and greater versatility than the iStir PDS.

Table 8 summarizes the range of apparatus limits for the VUWAL machine. Though the minimum rotation speed is listed as 100 RPM, most experiments are performed at higher rotation speeds (typically 1000 to 2000 RPM). The range of parameters and workspace volume is somewhat limited compared to what is possible with commercial robots. The VUWAL apparatus is intended to serve as a test bed for process development. For the results of this research to be applied with confidence to an industry application, it would be necessary to either repeat the experiments so that results are specific to the apparatus which will be used for joining or scale the results using dimensionless groups, a concept which is discussed in section 5.5 of this chapter.

Table 8 Capabilities of VUWAL FSW Apparatus

Traverse speed (IPM)	1 to 14
Rotation speed (RPM)	100 to 2200
Length of joint (inches)	3 to 18
Width of joint (inches)	up to 4”
Thickness of joint (inches)	up to ¼”
Tilt angle (degrees)	-4 to 4
Possible joint configurations	Bead on plate, butt, T-joint, lap

During welding, axial force and torque are recorded using a low-end dynamometer developed by Brian Gibson, a graduate student in Mechanical Engineering. This instrument is indicated in Figure 52. Axial force is detected using a series of strain gauges mounted on the vertical head. Application of axial force causes deflection of the vertical head, which produces a proportional change in resistance of the gauges. The voltage output of the strain gauge circuit is converted to force using a calibration curve; data can be plotted and analyzed post-process in Microsoft Excel or MATLAB. Torque measurements are derived from a voltage signal transmitted wirelessly via strain gauges housed inside the low-end dynamometer. At the time of this writing, the VUWAL apparatus lacks sensing capabilities for traverse and longitudinal forces (F_x and F_y).

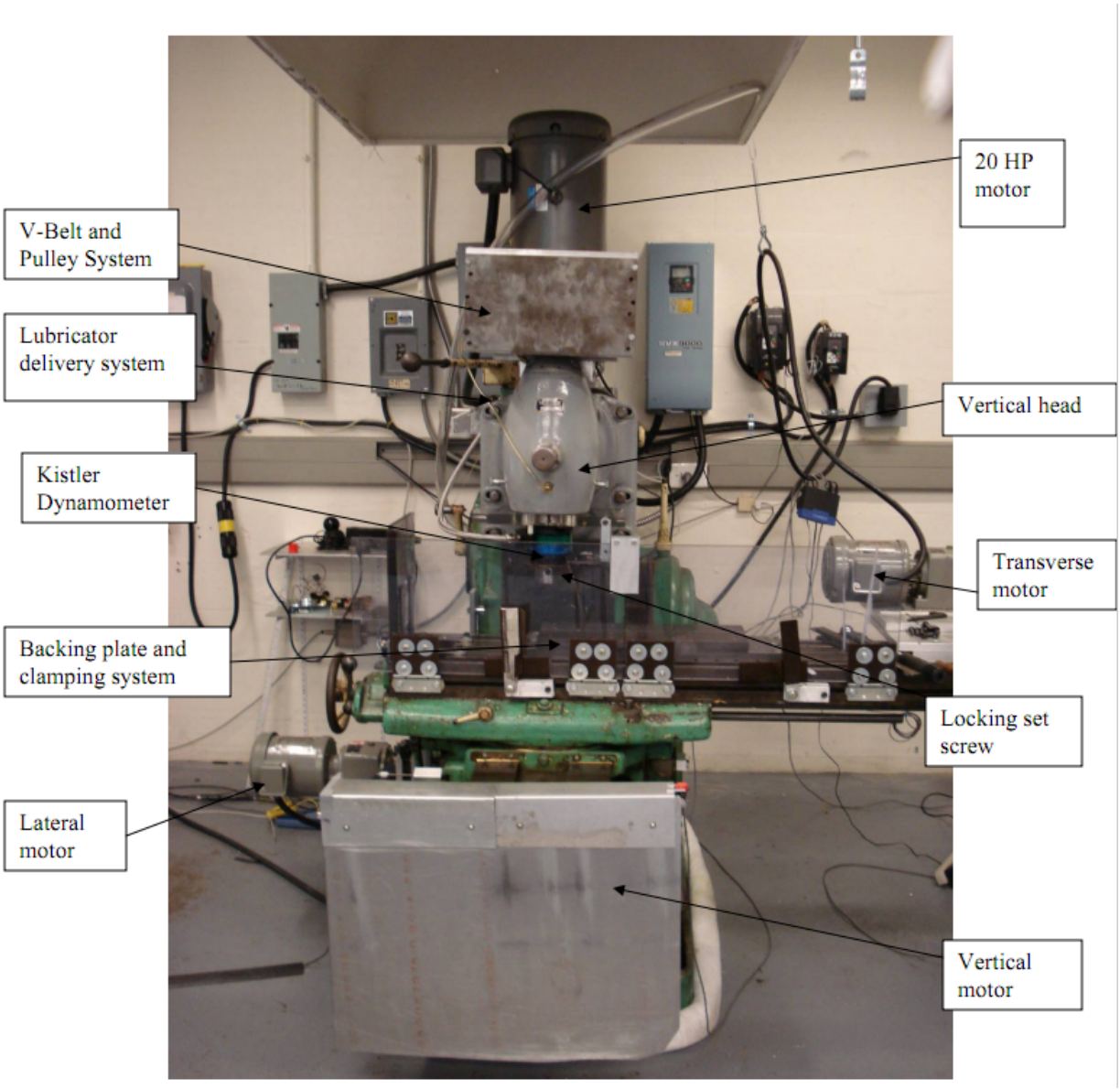


Figure 52 Overview of FSW apparatus

5.1.3 Tool Design

Though tools used in FSW range from simple (a smooth cylinder) to incredibly complex (the threaded and fluted cylinder introduced by TWI), there are some basic features which are common to all tool designs. As discussed in section 1.2, all tools consist of a shoulder (a larger cylinder which sits on the surface on the weld, generating heat as the tool rotates and containing material which would otherwise be expelled from the joint) and the pin (a smaller cylinder which extrudes from the shoulder and plunges into the workpiece and stirs the material). Enhanced pin features, such as threads or flutes, are included to enhance material mixing, which reduces the likelihood of void formation. Research by TWI has demonstrated that the inclusion of threads and flutes improves joint quality and widen the operating window (the range of parameters which produce acceptable joints) [4]. Scrolling on the shoulder surface improves surface quality and suppresses flash which may be extruded during the welding process.

The choice of a tool design depends on a number of factors ranging from the workpiece material, joint configuration, and shape of the workpiece to the desired quality of the joint. There are several published studies which develop processing windows for FSW of MMCs [34-36]. These investigations used tools with threaded probes to enhance material mixing and thereby improve joint efficiency. While this design strategy works for FSW of conventional alloys, the use of featured probes in FSW of MMCs is complicated by the presence of abrasive particles, which rapidly erode tool features. In Figure 17, adapted from [43], the once-threaded tool evolves to resemble a smooth cylinder after approximately 30 inches of weld. Considering these results, it seems impractical to use featured tools for this application, as any threads will be lost after only

a short weld distance. It makes sense to choose a smooth cylindrical probe for this application, since this is the shape a featured probe will eventually assume when subjected to abrasive wear. Unfortunately, the simplicity of the probe design results in decreased material flow; smooth probes typically produce lower quality joints with defects for all but a few parameter combinations. The processing window for FSW is made even narrower by the high temperature resistance of the MMC material. Section 2.3 ("Characteristics of Friction Stir Welded MMCs") discusses the differences in operating windows for Al alloys and Al-MMCs.

Although parameterization is not part of this work (research in this area is documented in references -), it is important to make some effort to ensure that the flow of material along the jointline is sufficient to preclude formation of large voids. The best option for this application is an intermediate tool design: a tool which is not threaded (since threads are quickly removed by wear processes) but not smooth (since a smooth probe combined with a harder material is a recipe for defect formation). The design chosen for the study in this chapter is the Trivex, a tool developed and patented by TWI. Shown in Figure 53, the Trivex geometry arose from the CFD modeling work of Shercliff and Colegrove [4]. The edges of the probe are convex and the three vertices, when connected, form an equilateral triangle. The ratio of swept area to probe area (a metric known as the dynamic volume) for the Trivex design is greater than that of a cylindrical probe (which has a dynamic volume of 1.0). The additional dynamic volume associated with the Trivex increases material mixing, which helps to reduce the occurrence of weld defects caused by insufficient flow.

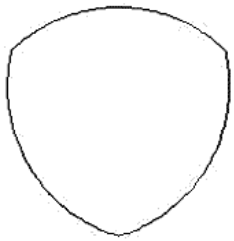


Figure 53 Top view of Trivex probe

An ancillary benefit of the Trivex geometry is that it reduces traversing forces by 18 to 25 percent and axial forces by as much as 12 percent [4]. Since forces are normally higher in FSW of MMCs (metal composites are harder and more temperature resistant than unreinforced Aluminum alloys), this force reduction offers an advantage over other tool geometries in this application [97]. In preliminary experiments which used a smooth probe to join MMCs, there were a few combinations of parameters for which the steady-state axial force exceeded the limit set for the VUWAL apparatus. No such force limits were encountered when using the Trivex tool at the same parameter sets. It is easy to imagine a similar situation could occur in an industrial setting, where the higher forces associated with joining MMCs might impose additional requirements on the stiffness of the FSW robot (unless the increase in force was counteracted by the use of a tool such as the Trivex).

Figure 54 is a CAD rendering of the specific tools used for these experiments. The shoulder is 0.75" in diameter and the pin has a swept diameter of 0.25." The probe length is 0.185", which is substantially less than the 0.25" thickness of the material stated by the manufacturer. There is a high degree of variability in thickness between samples in the MMCs provided by mc21 -- the probe length was chosen for compatibility with

the thinnest pieces, which measure approximately 0.20” in thickness. Because of this accommodation, welds of ¼” thick samples made using these tools are partial penetration. The double-sided tool design was chosen to reduce machining costs. The square flats milled into the tool shank interface with the locking set screw indicated in Figure 52.

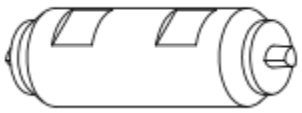


Figure 54 CAD rendering of Trivex tool used in wear study.

5.1.4 Design of Experiments

The objective of this phase of the research is to develop a predictive model for wear on process parameters. The published studies of tool wear in machining of MMCs (summarized in chapter IV) consider the effect of the parameters cutting velocity, cutting time, feed rate, and depth of cut. Most, if not all, of the parameters included in the machining studies have direct analogs in FSW: feed rate is analogous to traverse speed in FSW, cutting velocity parallels rotation speed, and the length of the cut (or depth of hole in drilling) is similar to joint length. Rotation speed, traverse rate, and distance welded also happen to be the parameters selected for the few published studies on wear in FSW. However, the literature is devoid of any models for wear in FSW which 1) quantify the contribution of each of these factors and 2) develop a predictive process model to inform selection of parameters which limit wear. Together, these objectives describe the goal of the work presented in this chapter. The methodology used to develop and test the wear

model for FSW of MMCs (Taguchi design of experiments, linear regression modeling, and cross-validation) was used by Davim et al. to construct predictive models for wear in cutting and drilling of MMCs [53-54].

In Friction Stir Welding of MMCs, there are many variables which could potentially influence wear. These range from process variables (rotation speed, traverse rate, tilt angle, plunge depth, distance welded, tool geometry) to material properties (reinforcement material, tool material, reinforcement size). A factorial design in which all of these factors are varied simultaneously is unwieldly – the alternative is to study these variables in a piecewise manner. This chapter considers only the effect of process parameters on tool wear in FSW of MMCs; chapters VI and VII are devoted to modeling and characterizing the impact of various material properties on wear. Discretizing the study in this way focuses the investigation and significantly reduces experimental time without compromising the quality of the analyses.

5.1.4.1 Variable Selection

The experimental results of Prado et al. pinpoint three common FSW process variables which impact wear: rotation speed ω, traverse rate ν, and distance welded ℓ. Since this work is intended (at least in this initial study) to build upon the work of these researchers, the variables considered in this phase of our investigation are identical to those studied in references [42-43]. There are some process parameters which we have neglected, notably tilt angle and plunge depth. Tilt angle has been shown to have little, if any, impact on the properties of the FSW joint [5]. The plunge depth is a constant

(except in the case of hemispheres or materials with highly varying thickness) dictated by the thickness of the workpiece material, the probe length, and the tilt angle. Since tilt angle and plunge depth are coupled, their variation is joint (i.e. a change in tilt angle must be accompanied by a compensatory change in plunge depth to ensure sufficient shoulder contact). To reduce complexity, the variables tilt angle and plunge depth are omitted from this study. An assessment will be made (based on the regression model) as to whether the percentage of the variation of wear that is unaccounted for by the variables ω, ν, and ℓ can be attributed to the effect of tilt angle and/or plunge. The effect of tool geometry is also not considered. As discussed in section 5.1. Trivex tool geometry was chosen as a compromise between a threaded probe and a smooth cylinder. The Trivex design reduces forces and stirs the material more than a cylindrical probe. While the material flow associated with the Trivex is less than the threaded probe, the former geometry is much easier to image (see section 5.1.5 on measuring wear). As demonstrated by Shindo et al., the threaded geometry is of little benefit in FSW of MMCs since its features are quickly removed by abrasive action [40].

5.1.4.2 Taguchi Matrix

Since out study involves multiple variables, it is important to develop a systematic means of varying parameters in order to minimize the number of experiments which must be performed. Plans of experiments based on factorial designs typically become impractical when multiple variables (and interactions between variables) are considered. Experimental design techniques provide researchers with a valuable alternative: an

experimental plan which minimizes the number of experiments yet is still large enough to accurately gauge the effect of each input variable on the outcome. The goal in using design of experiments (DOE) methods is the same as that of writing a plot summary in literature – to succinctly capture the "big picture." While there is the possibility that some of the nuances which might be evident in a larger experimental matrix will be overlooked (just as a summary of a long novel may omit some subplots), DOE offers the greatest experimental efficiency. To illustrate this point, consider a factorial design for 4 variables at 5 levels ("level" refers to the discrete value the variable can assume). A factorial design would require 625 (5^4) experiments; the same information could be extracted using Central Composites Design (CCD) in only 32 experimental cases.

A common criticism leveled at experimental design techniques is that they represent a cookbook approach and require little, if any understanding, on the part of the researcher about the process being studied. The objections expressed by some toward experimental design, however, are far outweighed by its methodological merits. DOE represents a valuable tool for researchers, as it allows us to consider multiple variables using a concise and manageable plan of experiments. In this work, DOE is merely a means to an end. We employ experimental design to reduce the number of experiments (conserving material is an important consideration due to the expense of composites) as well as direct our investigation. The design matrix allows us to systematically elucidate the strength of the relationships which exist between wear and process parameters as well as assign quantitative measures to these dependencies. Additionally, the organization inherent in DOE makes the technique a good basis for regression modeling.

The experimental design technique chosen for this investigation is the Taguchi method. Discussed in detail in chapter IV, the Taguchi method is a plan of experiments developed specifically for manufacturing applications. Taguchi is a continuum approach which seeks to systematically measure the change in the output variable by controlling for changes in the levels of the inputs. Our investigation into the variation of wear with process parameters considers three variables: rotation speed ω, traverse rate ν, and distance welded ℓ. A Taguchi L_{27} array allows us to study the effect of these three factors on wear simultaneously, including interactions. The L_{27} matrix consists of 27 experiments, which are listed in coded form in Table 1. The first column is assigned to factor A (rotation speed), the second column to factor B (traverse rate), and the third column to Factor C (distance welded). Additionally, there are three columns (not shown) which are assigned to the interactions between the factors (AxB, AxC, and BxC). The entry of each cell in the coded array is 1, 2, or 3. These numbers correspond to the three levels each factor in the array can assume. Summarized in Table 10, the levels for a factor are chosen such that they are equidistant from one another; the levels are intended to represent the low (1), medium (2), and high (3) value for each parameter. The factors and levels indicated in Table 10 can be used to "translate" the array in Table 9 into an uncoded form (the plan of experiments in Table 11). The numerical value of each factor is clearly indicated in Table 3 for each of the 27 experimental cases.

Table 9 Taguchi L_{27} array

	Factor A	Factor B	Factor C
1	1	1	1
2	1	1	2
3	1	1	3
4	1	2	1
5	1	2	2
6	1	2	3
7	1	3	1
8	1	3	2
9	1	3	3
10	2	1	1
11	2	1	2
12	2	1	3
13	2	2	1
14	2	2	2
15	2	2	3
16	2	3	1
17	2	3	2
18	2	3	3
19	3	1	1
20	3	1	2
21	3	1	3
22	3	2	1
23	3	2	2
24	3	2	3
25	3	3	1
26	3	3	2
27	3	3	3

Table 10 Factors and Levels

Factor	Parameter	Symbol	Units	Level 1	Level 2	Level 3
A	rotation speed	ω	RPM	1000	1500	2000
B	traverse rate	ν	IPM	5	7	9
C	length of weld	ℓ	Inches	8	16	24

Table 11 Plan of experiments

	Ω	ν	ℓ
1	1000	5	8
2	1000	5	16
3	1000	5	24
4	1000	7	8
5	1000	7	16
6	1000	7	24
7	1000	9	8
8	1000	9	16
9	1000	9	24
10	1500	5	8
11	1500	5	16
12	1500	5	24
13	1500	7	8
14	1500	7	16
15	1500	7	24
16	1500	9	8
17	1500	9	16
18	1500	9	24
19	2000	5	8
20	2000	5	16
21	2000	5	24
22	2000	7	8
23	2000	7	16
24	2000	7	24
25	2000	9	8
26	2000	9	16
27	2000	9	24

The levels chosen for these experiments were dictated by several constraints:

1) The apparatus limits of the VUWAL machine. The values of the levels must fall within the apparatus limits summarized in Table 12.

Table 12 Apparatus Limits for VUWAL

	Range
Rotation speed (RPM)	750-2250 RPM
Traverse rate (IPM)	0.5-14 IPM
Tool tilt (degrees)	-5 to 5 degrees
Length of joint (inches)	up to 18”
Width of joint (inches)	3”-4”
Thickness of joint (inches)	1/8” – 3/8”
Joint configurations	Bead on plate, butt joint, lap joint, T-joint, spot welds

2) The nature of the FSW process imposes additional restrictions on possible combinations of traverse rate and rotation speed. A lower rotation speed combined with a comparatively high rate of travel results in a colder weld temperature, which contributes to the formation of voids and increases the likelihood of tool fracture. On the other side of the coin, higher rotation speeds accompanied by a slow traversal rate can produce extreme heating in the workpiece material. While excess heat can also be correlated to weld defects, welds that are “too hot” are preferable because they do not run the risk of tool breakage (a failure which may befall their “too cold” counterparts). It is thus important to choose levels which ensure parameter compatibility (i.e. select parameters such that the lowest rotation speed is compatible with the highest traverse rate and vice versa) The rotation speeds considered are faster than those commonly encountered in the literature on FSW (in most academic studies, the rotation speeds reported do not exceed 1000 RPM). Higher rotation speeds have been shown to reduce forces [98]; force reduction is important in welding MMCs

since forces are normally higher than those encountered in FSW of conventional aluminum alloys [97]. Extrapolating the dependence of wear on rotation rate observed by Prado et al., it is anticipated that higher rotation speeds will accelerate deterioration of the tool (exacerbated wear may make it easier to detect and measure changes in volume) [42-43]. Travel speeds fall within the low and mid-range of the traverse motor's capabilities; this is because harder materials such as MMCs require slower travel speeds to ensure that the heat input is sufficient to plasticize the material.

The parameter levels span a wide range of values. It is clear that some parameter combinations will maintain weld temperatures that are colder, while others will result in welds which have too much heat input. A few parameters will be in the "Goldilocks zone", producing steady-state temperatures and material flow conditions which do not adversely impact weld quality. While it is important that at least a few of the welds are free from voids, the focus of the study is on wear rather than parameterization. It is necessary to operate at the extremes of the envelope (i.e. consider parameter combinations which are anticipated to produce a low-quality joint) in order to fully characterize the wear process. If we confined our study to only the parameters associated with acceptable welds, the data set would be very small (parameterization studies have demonstrated that MMCs have a very narrow process window) and the predictive capability of the resulting model would probably suffer as a consequence.

3) The levels for the factor "distance welded" are limited by the dimensions of the workpiece material. While the clamping system can accommodate a sample as

long as 18", the surplus material purchased from mc21 was sheared into strips (each strip constitutes ½ of a butt weld) measuring 8 inches in length. Hence each of the three levels for ℓ is a multiple of 8.

Although there are 27 experimental cases listed in Table 3, only 9 tools are required to execute the plan of experiments. Each tool is associated with a unique combination of rotation speed and traverse rate: a given tool is used to perform three welds (each 8 inches in length) at the specified parameter. The nine parameter combinations were randomized prior to experimentation; hence experiments were not performed in the same order as they appear in Table 11. Wear measurements were taken prior to and at the conclusion of each experiment. The next section discusses the techniques used to gauge wear.

5.1.5 Measuring Wear

The outcome variable for the experiments in Table 11 is wear. Wear in this instance is defined as the volume that is lost by the tool while joining an MMC sample. Chapter III summarizes the various techniques which can be used to quantify the amount of wear. The simplest method is to compare the pre and post-weld weights of the specimen – any mass which is lost is attributed to wear. This method is not well-suited for the wear study in this chapter, as the initial mass of the tool far exceeds the mass of the material which is lost. Though minute changes in tool mass can be detected using a high resolution mass balance, the tool's mass typically exceeds the instrument's mass rating. Precision balances with the resolution desired for this application typically have

an upper weight limit of 250 grams, a value which is slightly less than the characteristic initial weight of the tool pictured in Figure 11. Mechanical gauging can be used as an alternative to weighing. The dimensions of the tool (specifically the diameter and the length of the probe) are measured after each experiment and compared with the dimensions taken prior to welding. The implementation of this method is complicated by the variation of the probe's diameter with distance from the shoulder. To maximize accuracy, great care must be taken to ensure that the diameter is always measured at the same axial position (the radial measurement is an approximation since the Trivex geometry is not precisely circular). Additionally, if there is significant shoulder wear, the length of the probe will appear to increase over time. While shoulder wear is a concern for all methods of wear measurement, it is particularly problematic for mechanical gauging techniques.

Radiotracing, in which one of the elements in contact is made radioactive, has the highest resolution of any method of wear measurement. If the tool is irradiated prior to welding, the amount of radioactive material detected in the weld post-process precisely indicates the mass of material lost by the tool during FSW. In addition to being incredibly difficult to implement, the resolution of this technique ($10^{-12}g$) far exceeds what is required for this application [49]. The major advantage of radiotracer techniques is that they can provide insight into the dynamics of the wear process which are often difficult (or impossible) to directly observe. A modified tracer technique (which does not rely on radioactive material) is used in chapter VI to test the effect of particle size on wear and assess the validity of the rotating plug model.

It is the author's opinion that the best method to quantify tool wear in FSW of MMCs is via an optical measurement technique. Optical methods have a distinct advantage over mechanical gauging in that they can be used to gauge changes in two dimensions (x and y) simultaneously. The most common optical method is the shadowgraph (also known as a profile projector), an apparatus which uses a Fresnel lens to focus light from a source aimed at the specimen. The result is a shadow image of the part (Figure 55) projected on a screen directly behind the specimen. An optical microscope image of the part provides a similar visual reference; as with the shadowgraph, comparison of successive images can visually indicate any changes in shape which may occur as a result of wear. The problem with these methods are two-fold: 1) images obtained are generally low-magnification (in the case of the shadowgraph, the Fresnel lens results in blurring at the edges of the part) and 2) the images, in their raw form, provide only qualitative information about the volumetric deterioration of the tool due to wear. Both of these issues can be resolved using the optical technique (a custom-made optics bench used in conjunction with imaging software) proposed in this section.

Figure 55 Shadowgraph image of Trivex tool prior to joining MMCs [51]. Image taken at maximum magnifaction possible with profile projector.

Figure 56 is a schematic of an optics bench which can be used to quantify wear based on cross-sectional images of the examined specimen. The shank (the widest cylindrical section of the tool) is inserted into the tool mount and aligned with the grid (squares measure 1 cm x 1 cm) affixed to the back of a thin Aluminum plate positioned behind the tool mount. A set screw holds the sample in place and prevents vertical translation of the tool during imaging. An image of the specimen is captured using a high-resolution camera rigidly mounted in front of the specimen holder as indicated in the drawing. The vertical position of the camera lens is slightly higher than the top of the mount stage (the probe and shoulder are the only tool features which extrude from the tool holder -- the tool shank lies below the camera's field of view). A halogen lamp can be used if further illumination of the tool is required to obtain a clearer image. Holes in the optics bench permit horizontal movement of the tool fixturing relative to the camera. For these experiments, the horizontal distance from camera to tool is fixed at 6 inches. Images were taken using a Canon Powershot A620 camera with 7.1 Megapixels and 4X

digital zoom. The result is a sharp, low distortion image for which the edges of the tool surface can be easily identified (Figure 57).

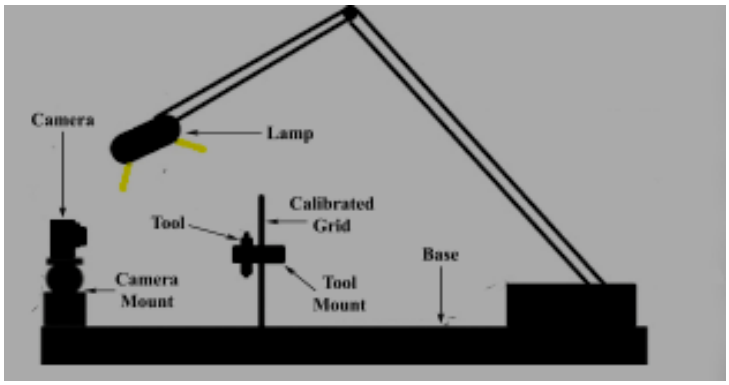


Figure 56 Schematic of optics bench used to quantify wear.

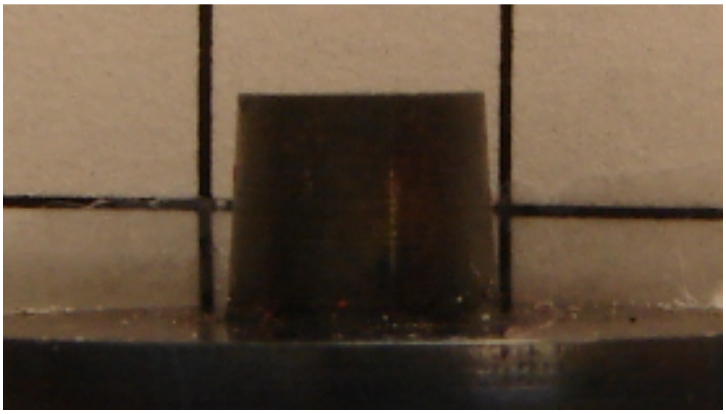


Figure 57 Image of tool probe taken using optics bench.

The images taken using the optics bench are imported into image processing software for analysis. Within Photoshop, the outline of the probe can be traced using the

magnetic lasso tool. With this tool, the area is not defined by free-hand selection using a mouse (a technique that relies on user dexterity and, as such, is particularly difficult when the shape being traced has an irregular or curved geometry). Rather, the magnetic lasso enables the user to click points which lie along the edges of the geometry to be analyzed -- once these points are selected via clicking and dragging, the lasso "snaps" to the boundaries of the surface. The user identifies "base points" on the left and right of the geometry where the tool probe extrudes from the shoulder. The magnetic lasso tool draws a straight line between these points, forming an enclosed area that is coincident with the shape of the tool probe The lasso tool is best suited for outlining geometries when there is a high degree of contrast between the foreground and background of the picture, conditions which are characteristic of the probe images in these experiments. The settings/configuration of the lasso can be adjusted within the software: for instance, the user can define the width (the number of pixels the lasso uses to find contrast) as well as the frequency (the number 1-100 of intermediate points that are placed between the selected points).

The quantitative measure of the area of the surface defined using the lasso is given in pixels. By similarly defining a 1 cm x 1 cm square on the grid visible behind the tool (the image in Figure 9 has been cropped to remove the portions of the image that were extraneous to the analysis), we can derive a conversion factor (with units of $\frac{pixels}{cm^2}$) which permits conversion between pixel counts and cross-sectional area. There are a few important assumptions which underlie this method of wear measurement:

1) **Wear is symmetric with respect to the longitudinal axis of the tool.** The rotation rates considered are fast enough to equalize the effects of any wear process which may act on the tool. Reconciling the longitudinal asymmetry of the rotating plug model with the symmetry of volumetric wear (and the conditions under which this symmetry exists) is addressed in the next chapter. The volumetric wear percentages reported in this study are based on the symmetry assumption (i.e. that revolution of the cross-section 360 degrees about the longitudinal axis represents the volume of the probe). The cross-section is defined as one-half of the probe in Figure 9 (the one-half factor is a consequence of the longitudinal symmetry of the probe with respect to the axis of revolution). Stated mathematically, $V = \pi \int_a^b [f(x)]^2 \, dx$, where $f(x)$ is the function that defines the shape of the probe cross section (for a right circular cylinder $f(x) = r$).
2) **For successive images of the same tool (i.e. the initial image and the images taken after the tool has been used to weld some distance), there is a proportional relationship between the decline in the pixel count of the probe cross-section and the amount of material lost by the tool.** Pixels can easily be converted to units of area using the conversion factor discussed above. It is not necessary that the relationship between the reduction in pixels and volumetric wear be a 1 to 1 correspondence.
3) **The surface of the tool is clean and free from residue.** During the course of welding, Aluminum accumulates on the tool probe. This Aluminum can be removed post-process by immersing the tool in a solution of NaOH and water.

The compound exothermic reactions $2\ Al + 6\ H_2O \rightarrow 2\ Al(OH)_3 + 3\ H_2$ and $Al(OH)_3 + NaOH \rightarrow Na^+ + [Al(OH)_4]^-$ dissolve excess Aluminum without otherwise disturbing the tool surface. If the surface is not cleaned prior to analysis, the presence of extra material will impede accurate measurement of the cross sectional area. Excess material "stretches" the dimensions of the tool image in both x and y directions, leading to an overestimation of the area (and a corresponding underestimation of any wear which has occurred). A clean surface is also essential for the analysis techniques which appear in subsequent chapters (SEM, profilometry, and weighing).

4) **Wear for this process is cumulative.** Wear of the tool is assessed prior to and after each experimental case. It is assumed that the effects of wear are additive. For a given tool (each tool corresponds to a single set of process parameters), the wear value reported for a specific experiment represents the sum total of wear for that experiment as well as all the experiments which preceded it. Wear reported for a specific tool can thus be expressed in terms of a recursive sequence: $W_{n,T} = W_{n-1,T} + W_{n-2,T} + W_{n-3,T} + \cdots + W_{n-i.T}$. Wear observed for the nth experimental case of the *T*th tool, $W_{n,T}$, is equal to the inclusive sum total of the wear which occurred in *each* of the previous *i* experiments. Based on this cumulative property of wear, the wear associated with an individual experimental case can be obtained by simply subtracting the wear value for the previous case (which is also cumulative) from it. The final point on cumulative wear is obvious but is stated for completeness: any volumetric deterioration observed for an experimental case is attributed only

to the wear processes which occurred during that specific weld. It is assumed that no material is lost during preparation of the tool surface for examination (or, more precisely, that any material loss which takes place during etching is miniscule compared to the material loss associated with wear). The wear data for a tool is not expected to fit an arithmetic sequence. The difference between wear terms may not be constant since the shape of the tool changes with volume loss. It is anticipated that changes in tool shape which accompany wear have some influence on the wear rate, a topic which is further discussed in chapter VI.

Before presenting the results of this study, we will address some questions/objections which may arise concerning experimental procedure. In a few of the previously published studies, welds were performed on a continuous, long sheet of metal (length on the order of meters). The tool is used to weld a specified distance, extracted and examined, then plunged back into the material as welding continues along the same jointline. The experiments thus consist of a single, long-pass weld with regularly spaced interruptions to permit examination of the tool In contrast, each experiment in the matrix in Table 11 comprises a separate weld (and thus requires a new workpiece), a modification necessitated by the apparatus limits discussed in section 5.1.4.2. We maintain that these two configurations are equivalent since every experiment (regardless of whether it is a new weld/workpiece or simply a continuation of the weld along the same jointline) consists of the same four stages: plunge, dwell, welding, and extraction. The question has previously been raised as to how dwell time (the initial period of the weld for which the traverse rate is slower than the weld speed in order to ensure proper

plasticization of the material) may impact wear [33]. It is likely that if wear does depend strongly on process parameters, the dwell time will have some influence on the amount of wear (since process parameters for the dwell period are typically distinct from those used in steady-state welding). Unfortunately, the dwell time cannot be eliminated without risking probe fracture, particularly in the case of parameter combinations with faster traverse rates which produce welds with relatively colder temperatures. As a compromise between these competing concerns, the dwell time is carefully controlled via an automatic weld routine. The rotation speed during dwell is synonymous with the rotation speed during welding; however, the dwell traverse rate of 0.5 inches/minute is significantly slower. The dwell period lasts until the tool has traversed 0.25 inches along the joint line (a dwell time of approximately 30 seconds), at which point the traverse speed is abruptly increased to the welding speed. To further minimize the effect of the initial stages of the weld on wear, the plunge and dwell periods are consolidated by advancing the tool into the material from the side of the joint. The usual method of plunge, in which the tool rotates in place while the stage is raised to the weld height, may result in additional wear which could potentially skew the wear measurements and/or the correlation between wear and process parameters. For these reasons, it is critical that we be able to carefully control (i.e. minimize) the amount of wear which originates from plunge and/or dwell. In all the experiments which comprise the dissertation, this control is accomplished using the WeldController, a software developed specifically for VUWAL by graduate student Paul Fleming.

5.2 Results and Analysis

5.2.1 Wear Data

As emphasized in the previous section, the quality of the results of the experiments in Table 3 hinges largely on our ability to measure wear. It is thus critical that the method we have established to quantitatively gauge wear based on optical techniques be both robust and reliable. Care has been taken to ensure that both the experimental setup and the wear process satisfy each of the assumptions detailed in bullets (1)-(4) of section 5.1.4. While the wear measurement technique proposed here represents a unique method for gauging tool wear in FSW, it is aligned with methods which have been used in similar studies on wear in machining of MMCs.

The way in which we choose to express material loss (the outcome variable in these experiments) depends on the measurement technique. Material lost by the tool during welding can be expressed in terms of grams, length, cross sectional area, or volume. Since these metrics are related to one another by geometry and/or density, if one value is known it is possible to calculate all the others using geometric formulas or conversion factors. For example, the reduction in the probe mass (grams) can be converted to volume (cm^3) by simply dividing the former value by the density of the probe material (g/cm^3). Similarly, the cross sectional area can be converted to volume by rotating the cross section 360 degrees about the transverse axis of the tool (or equivalently, rotating one-half of the cross section 360 degrees with respect to the longitudinal axis). To avoid confusion about units and conversion, it is better to cast wear in terms of a pure (unitless) number. The fraction of a particular metric (whether

weight, area, or volume) that is lost due to wear is computed using the formula $\frac{M-M\prime}{M}$, where $M$is the original value of the metric and M' represents the reduced value associated with wear. For our study, wear W is calculated in terms of the original pixel count of the probe cross section (P) and the pixel count of the same cross-section after a specified weld distance (P'). If wear has occurred, then the data will satisfy the inequality $P > P'$. Otherwise, $P = P'$. In no case should $P < P'$; this observation is physically possible only in instances where the shoulder has experienced substantial wear. To convert fractional wear to a percent, $W$is multiplied by 100.

The results of the experiments which comprise the Taguchi L_{27} array are listed in Table 13. The wear values in the final column were obtained using the optical measurement technique explained in section are calculated using the formula $W = \frac{P-P\prime}{P}$) and represent the cumulative fraction of material lost from the tool probe during welding. A table containing pixel counts and the corresponding cross sectional area measurements for each case (as well as images of each tool probe) can be found in Appendix A.

Table 13 Experimental Matrix and Results for FSW of Tool Wear in MMCs

	ω	N	ℓ	Wear
1	1000	5	8	0.036928
2	1000	5	16	0.059150
3	1000	5	24	0.075490
4	1000	7	8	0.012418
5	1000	7	16	0.027451
6	1000	7	24	0.037908
7	1000	9	8	0.014243
8	1000	9	16	0.046704
9	1000	9	24	0.049685
10	1500	5	8	0.058262
11	1500	5	16	0.156320
12	1500	5	24	0.160458
13	1500	7	8	0.029690
14	1500	7	16	0.123002
15	1500	7	24	0.171615
16	1500	9	8	0.040288
17	1500	9	16	0.065509
18	1500	9	24	0.155912
19	2000	5	8	0.035404
20	2000	5	16	0.078156
21	2000	5	24	0.128591
22	2000	7	8	0.071523
23	2000	7	16	0.117219
24	2000	7	24	0.175828
25	2000	9	8	0.040482
26	2000	9	16	0.100703
27	2000	9	24	0.149548

5.2.2 Analysis of Wear Data

The goal of these experiments is to relate wear of the tool to process parameters. The data in Table 13 can be used to measure the correlation between wear and each of the three process parameters considered. To assess the influence of each factor on wear, the signal to noise ratio (SNR) is computed for each case (Table 14). The SNR ratio is defined for this application by equation 5-1, where Y_i denotes the observation for the i_{th}

experiment and n corresponds to the number of cases in the experimental matrix ($n = 27$). The SNR value is an indicator of how the response value (the fraction of tool material lost as a result of wear) compares to some target value. For these experiments, the target is 0, a value which represents the amount of tool wear observed in an ideal FSW scenario. The formula used to calculate SNR depends on both the data and the target value. Since the wear data is non-negative and has a target value of zero, the "smaller is better" criteria for SNR (equation 5-1) is applied. Higher values of SNR are thus associated with the tests which minimize system response.

$$SNR = -10log_{10}(\frac{{Y_i}^2}{n}) \qquad \text{(equation 5-1)}$$

Table 14 Wear Data with Tabulated SNR

Test number	Wear	SNR
1	0.03693	42.96652
2	0.05915	38.87454
3	0.07549	36.75585
4	0.01242	52.43260
5	0.02745	45.54247
6	0.03791	42.73902
7	0.01424	51.24161
8	0.04670	40.92656
9	0.04969	40.38913
10	0.05826	39.00593
11	0.15632	30.43335
12	0.16046	30.20641
13	0.02969	44.86143
14	0.12300	32.51539
15	0.17162	29.62253
16	0.04029	42.21012
17	0.06551	37.98762
18	0.15591	30.45605
19	0.03540	43.33259
20	0.07816	36.45439
21	0.12859	32.12943
22	0.07152	37.22472
23	0.11722	32.93368
24	0.17583	29.41188
25	0.04048	42.16840
26	0.10070	34.25279
27	0.14955	30.81803

The SNR is not very useful as a stand-alone statistical metric. It can, however, be used to rank the influence of each factor on the outcome variable. Table 15 organizes the SNR data into a response table. The cell entries are the average value of SNR for all experiments where the factor in the column is fixed at the level indicated by the corresponding row. Δ represents the magnitude of the difference between the average SNRs for a particular factor when the level changes from 1 to 3. Based on the relative

magnitude of this difference, each factor is assigned a ranking concordant with its significance.

Table 15 SNR Response Table for Wear in FSW of MMCs

Level	Rotation speed	Traverse rate	Length of weld
1	43.54	36.68	43.94
2	35.26	38.59	36.66
3	35.41	38.94	33.61
Δ	8.13	2.26	10.33
Rank	2	3	1

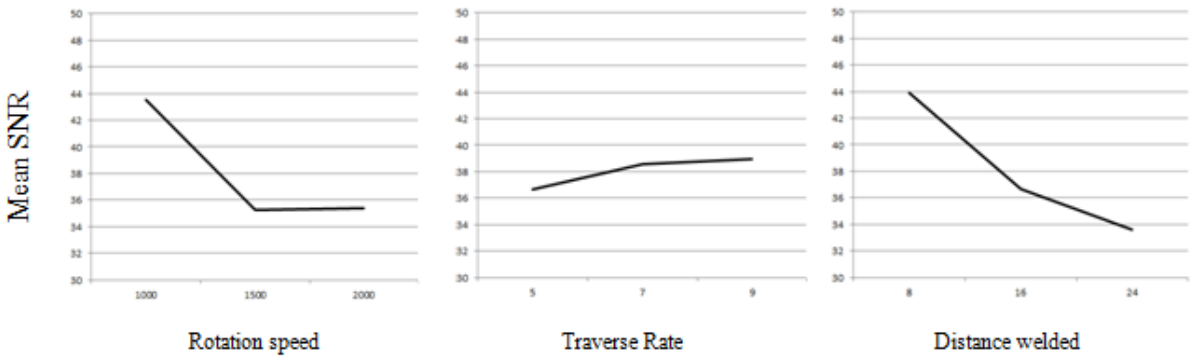


Figure 58 Main Effects Plots for SNR

Figure 58 displays the effects plots for each factor based on the SNR means. In evaluating the plots, it is important to remember that the value of SNR is inversely proportional to wear. Hence a lower mean SNR for a particular factor level means that the value for that level promotes wear, while a higher SNR indicates just the opposite. The larger the difference in SNR between factor levels, the greater influence that factor has on the outcome variable. The plots of the mean value of wear versus levels for each

factor convey a similar story. The average wear increases with rotation speed and length of weld, but decreases slightly with the traverse rate. Wear is expressed as the proportion of material lost, a value obtained by calculating the fractional change in cross sectional area of the probe recorded for each experiment. As with SNR, the magnitude of the maximum change in the outcome variable as the level is increased indicates the importance of the factor. Based on the mean plots, the significance of the factors are ranked (from most to least important) as 1) length of weld, 2) rotation speed, and 3) traverse rate. This ordering agrees with the SNR response rankings in Table 15. The optimum levels (i.e. the level settings which minimize wear for each factor) indicated by the plots are: level 1 rotation speed, level 3 traverse rate, and level 1 length of weld. Comparing this prediction with the actual data, we find that the parameter 1000 RPM/7IPM, rather than the 1000 RPM/9IPM combination predicted by the main effects plots, is actually associated with the least material loss (the difference in wear between the predicted and actual minimal wear parameter, however, is small). While the main effects plots can give an indication as to the relative importance of factors and general trends in the data, a multivariate linear regression model (section 5.2.2) serves as a more robust tool for predicting process behavior.

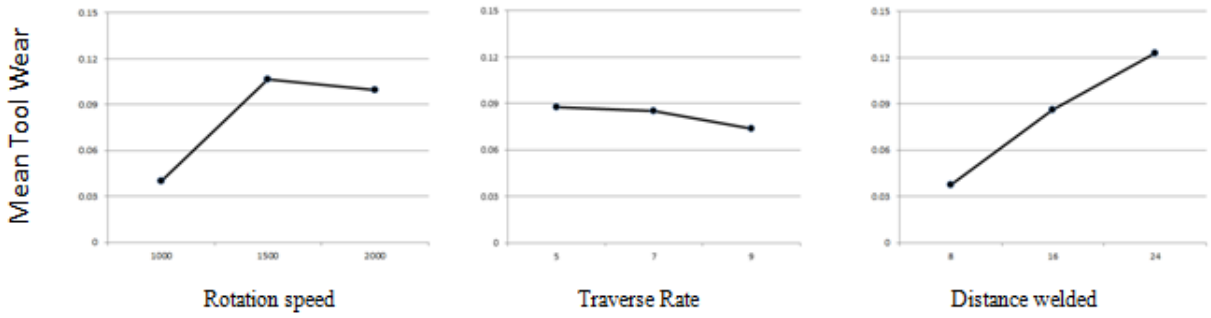


Figure 59 Main Effects Plots for Mean Tool Wear

The L_{27} Taguchi array chosen for this study also allows us to investigate the possibility of interactions between factors. The main effects plots for interactions are summarized in Figure 60 (SNR interactions) and Figure 61 (mean wear interactions). Interactions are possible in any experiment where there are multiple independent variables. An interaction is formally defined by as an effect which occurs "when a relation between (at least two) variables is modified by (at least one) other variable." An interaction simply means that an effect is not the same across all observations. Consider the top center plot in Figure 61 (the main effects plot of rotation speed vs. traverse rate). For the 1500 RPM group, increasing traverse rate decreases wear, while the opposite is true (at least initially) for the 2000 RPM group. The reversal of this trend for these different data groups is presumably due to the effect of a third variable (each data point represents an average of the wear data for that parameter over three weld lengths). The idea behind statistical interaction is somewhat analogous to the concept of conditional probability. Written as $P(A|B)$, a conditional probability simply means that the probability of an event A depends on the condition B (for independent events $P(A|B) = P(A)$). For example, the probability of drawing a Jack from a fresh deck of cards (event

A) is different than the probability of drawing a Jack given that two Jacks and two Aces have already been drawn without replacement (condition B). Similarly, the probability of observing a particular value of wear at any parameter combination depends on the distance welded. This is the essence of an interaction: the wear value for any two variables is affected by the presence of a third variable. Interactions can be additionally classified as either one-way or two-way. A one-way interaction is typified by directionality (i.e. change in the level of a factor has the same effect across data sets). A two-way interaction is bidirectional – the previously mentioned interaction between traverse rate and rotation speed in Figure 61 fits this criterion since the effect on the outcome variable is not uniform across the data series.

The strength of a potential interaction between variables can be gauged by the distance between the data points for series plotted on the same axes. Data series which lie close to one another (such as the approximately parallel lines in the center right plot of Figure 61) indicate that the interaction between the variables in question is weak. A stronger interaction manifests itself in wider spacing between data series (such as the "distance welded x traverse rate" graph in the bottom center of Figure 60). The strength of the interaction between variables is thus proportional to the distance between the data series which represent the group means. Every two variables interact to some degree; the attribute "no interactions" does not mean that interactions are not present, but rather that they are weak enough to be neglected. A determination regarding the magnitude of the interactions and whether they are significant (the term "significant" is applied based on p-values and is generally only assigned to relatively strong interactions) is made as part of the ANOVA analysis presented in the next section. A major benefit of the Taguchi

method is that it generates a balanced dataset which makes it easy to discern interactions and extract the effect of each factor on the outcome variable.

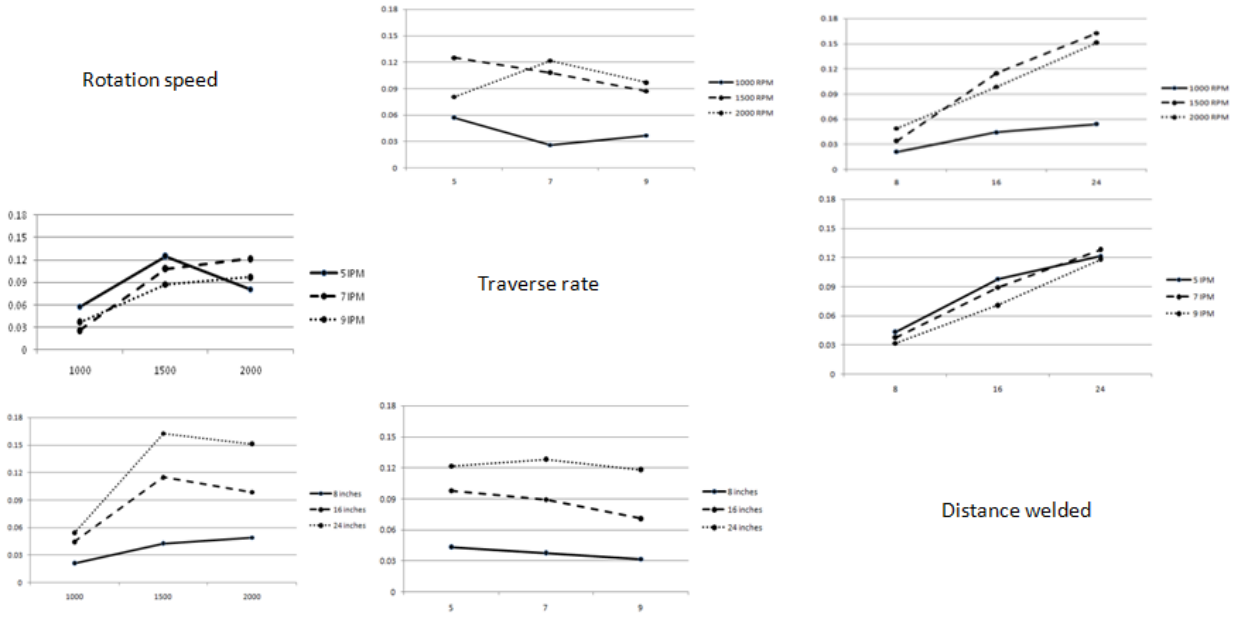


Figure 60 Interaction plots for mean fractional volume loss. Clockwise from top: rotation speed x traverse rate, rotation speed x distance welded, traverse rate x rotation speed, traverse rate x length of weld, length of weld x rotation speed, length of weld x traverse rate.

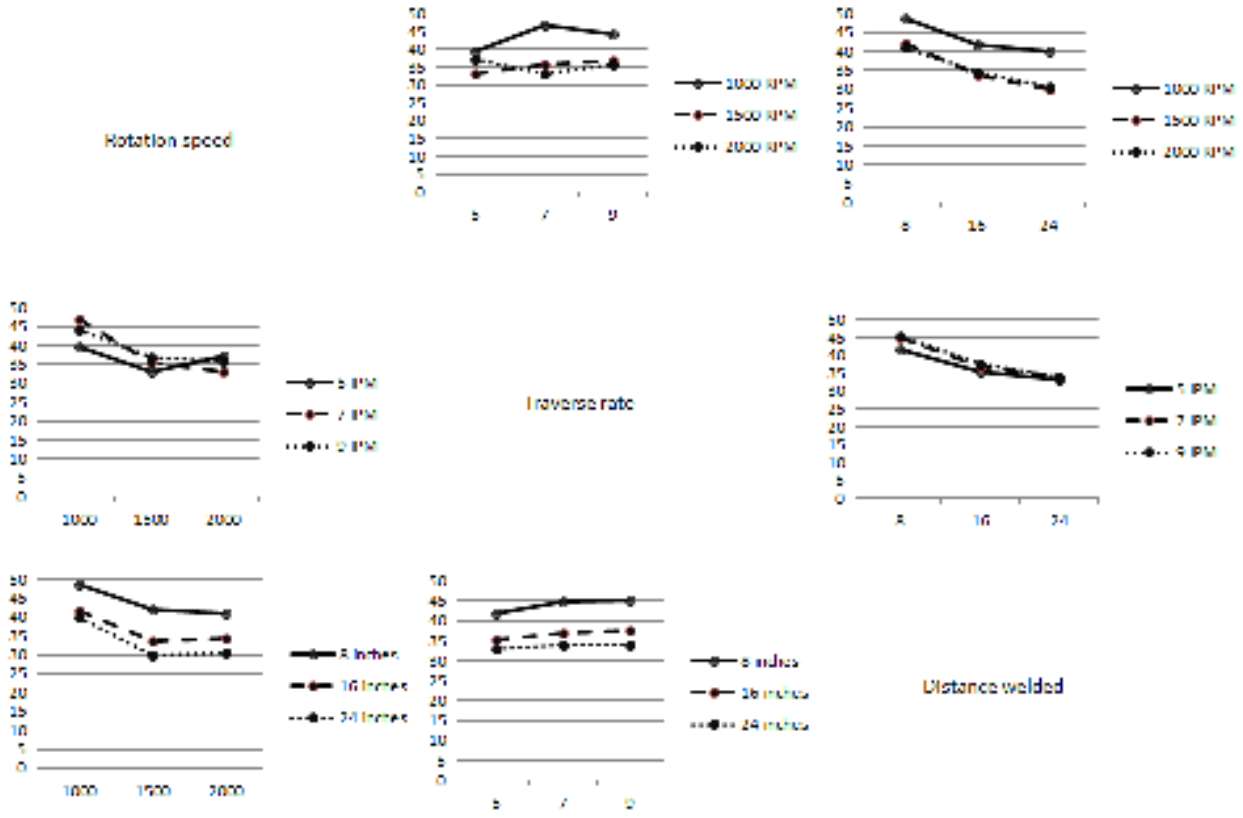


Figure 61 Interaction plots for signal to noise ratio. Clockwise from top: rotation speed x traverse rate, rotation speed x distance welded, traverse rate x rotation speed, traverse rate x length of weld, length of weld x rotation speed, length of weld x traverse rate.

Some general trends begin to emerge based on the response analysis. The amount of wear increases with rotation speed and distance welded but decreases slightly with traverse rate. The relationships between the factors and wear is mirrored in the signal to noise ratio, which decreases with distance welded and rotation speed, but increases with traverse rate. Since SNR is inversely proportional to material loss, the direct relationship between SNR and traverse reflects the inverse proportionality between the amount of wear and travel speed ν (similarly, the inverse dependence of SNR on rotation and distance translates into the direct relationship observed between wear, spindle speed, and length of joint). While response analysis can help us to understand the broader vestiges of the wear process and provide information as to which factors have the greatest effect on the outcome variable, it offers little in the way of wear prediction beyond identifying the experimental conditions (selected from among the conditions explicitly considered in the experiments) which minimize it. To fully understand the dependence of wear on process variables, we need to know the relative strength of the correlations between the factors and wear, the percentage contribution of each factor to wear, and what percentage of variation in the wear of the tool is unaccounted for by the factors considered in the experiment. These questions can be answered using regression analysis.

5.2.3 Constructing the Regression Model

The ultimate goal of these experiments is to construct a predictive model for wear of the tool in FSW of MMCs using process parameters as inputs. This model is constructed using multiple regression analysis, a statistical method commonly used for predictive process modeling of multivariate systems. The mathematics which underlie

this technique are discussed in detail in chapter IV. To make the Taguchi array in Table compatible with regression analysis, it is partitioned into two matrices. The input side of the Taguchi array (which has the experimental factors as columns and the levels of the factors as the numeric elements) forms the matrix **X**. The dimensionality of **X** is 27 (the number of experimental cases) x 3 (the number of factors). The second matrix, **Y**, consists of the wear data associated with each experimental case; **Y** is thus a column matrix with dimensionality 27 x 1. The objective of multiple regression analysis is to find a set of linear weights β_i (the column matrix **β**) which "map" the factors into the outcome variable. Regression analysis seeks to find the column matrix **β** which satisfies the equation $\mathbf{Y}=\boldsymbol{X\beta}+\mathbf{e}$. The **e** term represents the column matrix of errors. Error is defined tautologically as the difference between the observed value y_i and the predicted value given by $\boldsymbol{X\beta}$. While there are many sets of weights which meet this criterion, the transformation $(\boldsymbol{XX'})^{-1}(\boldsymbol{X'Y})$ (where **X'** is the transpose of the data matrix **X**) yields the unique matrix **β** which minimizes the sum of the mean squared errors.

Expanding the $\boldsymbol{X\beta}$ term using matrix multiplication, the regression model **Y** yields a linear equation of the form $y = x_1\beta_1 + x_2\beta_2 + \cdots + x_n\beta_n + e_1 + e_2 + \cdots + e_n$. For our specific experiments, n=3 since there are three factors which are varied. The expression for wear thus takes the form $W = \omega\beta_1 + \nu\beta_2 + \ell\beta_3 + C$, where β_1, β_2, and β_3 denote the coefficients associated with rotation speed, traverse rate, and length of weld, respectively. As discussed in chapter IV, these coefficients function as weighting factors which "map" the factors into the outcome variable. C represents the sum of the error terms. The multivariate regression model derived for our particular data set appears in equation 5-2:

$$W = 0.584\ell - 1.038v + 0.009\omega - 6.028 \qquad \text{(equation 5-2)}$$

The relationship between wear and process parameters indicated by this equation is consistent with the results of the response analysis detailed in section 5.2.1. Wear increases with rotation rate and length of weld, but decreases with traverse rate. Figure 62 plots the regression model as a series of planar surfaces in 3-dimensions. Percent wear is plotted on the z-axis; traverse rate and rotation speed lie along the x and y axes, respectively. Each of the three planes represents a specific length of weld – from bottom to top, the parallel "slices" correspond to 8 inches, 16 inches, and 24 inches of weld. It is clear from this visualization that the parameters which minimize wear are characterized by slow rotation speeds, fast traverse rates, and shorter weld lengths. Unfortunately, slow rotation speeds coupled with faster traverse rates are often in opposition to the parameters which will produce acceptable welds. The lack of heating which accompanies this parameter combination typically results in voids, or in the most extreme cases, failure of the tool due to fracture. When selecting parameters which mitigate wear, weld operators must negotiate a compromise between reducing wear yet preserving an acceptable level of quality; a very similar dilemma is faced by machinists when using a machinability map to choose parameters for machining materials which will result in wear of the machine tool.

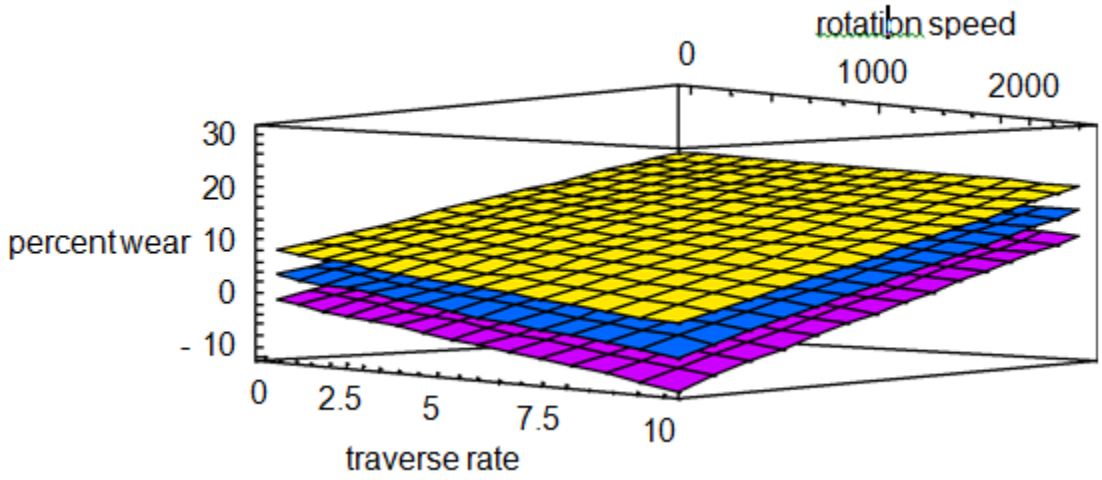


Figure 62 Plot of regression model in 3-space. Traverse rate, rotation speed, and length of weld are plotted on the x, y, and z axes, respectively.

A glaring peculiarity of the plot in Figure 15a is that there are some parameters which correspond to negative values of wear. A negative wear value indicates that the volume of the probe has increased as a result of welding, a physical impossibility in our system. These negative values stem from the choice to plot the regression model as a continuum (i.e. there are no restrictions on the values of the parameters represented on the x and y axes). The prediction of values which lie outside the range of reality is common in multiple regression analyses. The classic example is a regression model which predicts the IQ of a child based on the IQ of its parents. Although there is a strong linear correlation between the factors (IQ of father and mother) and outcome variable (IQ of child) in this instance, there are some values of the factors for which the child's IQ is predicted to be 0 or negative. In these situations, the regression plot can be modified to better reflect the dynamics of the system by restricting the range of the x and y axes such that the value of z is always nonnegative. It is generally expected that empirical based

statistical models will have an upper and/or lower limit on their range of validity. Although there may be a portion of the model for which the predicted values lie outside the range of what is physically possible, this may does not compromise the model's usefulness. For the wear model, it is nearly certain that the parameters which correspond to negative wear would not be selected as a weld schedule since they typically coincide with void formation and/or tool fracture. In the IQ example, it is similarly unlikely that both parents would have IQs which are below the threshold values which correspond to child's IQ=0.

The usefulness of the regression model is that we can use it to pinpoint which factors will generate a desired response. When applied in this way, the model can answer common questions about wear which may arise in joining of MMCs. Such questions usually take the form, "if welding a __ inch sample at a spindle speed of __ RPM and the total wear of the tool must remain under 10 percent, what is the maximum possible rate of traverse"; higher traversal rates are generally desired in manufacturing scenarios since they minimize the amount of time spent on each part (this is especially true for machining and/or welding on a mass production scale). The particular β coefficients associated with each of the three variables can be interpreted in terms of the change in system response (wear) that accompanies a step change of 1 unit in the input factor:

-wear increases 0.009% per unit increase in spindle speed (RPM)

-wear increases 0.584% per unit increase in length (inch)

-wear decreases -1.038% per unit increase in traverse rate (inches/minute)

This interpretation of the model can be used to estimate the effect of increasing or decreasing the value of a particular process parameter on total wear – the coefficients thus have a very practical application for weld operators. For example, if the operator decreases the rotation speed by 500 RPM while keeping all other parameters equal, the amount of wear (compared to the wear that the tool would experience if the rotation speed were unchanged) should decrease over the length by 4.5 percent. While more work needs to be done to characterize the effect of wear on weld quality, the relationships implied by the coefficients can be used to make back of the envelope calculations regarding tool life based on the current, measured state of the tool. For example, if 20% is the threshold at which tool wear becomes severe enough to negatively impact the weld and the current value of tool wear is estimated as 10%, the tool will last for another 17 inches of weld ($\frac{10\%}{\frac{0.584\%}{in}}$) before it needs to be replaced (assuming all other process parameters remain constant). We can make these estimates because the wear process for this system is continuous (as compared to more spontaneous wear processes, such as a fracture due to fatigue wear, which are less predictable). The coefficient interpretation can be exploited for operator control: if tool wear is measured near the critical value (either using an off-line evaluation technique or an in-process sensor of the type proposed in chapter VIII), the weld operator can make a compensatory adjustment in either the rotation speed or traverse rate to prolong tool life. Extending tool life is particularly important for more exotic (and expensive) tool materials. Operators should, however, keep in mind that any change in process variables which reduces wear could potentially have a negative impact on weld quality.

Two questions that always arise in system modeling problems are those of uniqueness and linearity. In choosing to fit the data to a multiple linear regression model, we have assumed that there is a 1 to 1 correspondence between the input factors and the outcome variable: each combination of process parameters is assumed to be associated with a unique value of percent wear (uniqueness). Furthermore, any change in the factors is accompanied by a proportional change in the system response (the condition for linearity). The "fit" of the model is used to assess the validity of these assumptions. If the "fit" is poor, it means that the function is nonlinear or does not satisfy the 1 to 1 criterion. Although the topology of the three dimensional plots in Figure 15 represent the variation between wear and process parameters as linear, these plots were constructed based solely on the regression equation and tell us nothing about how well the model represents the actual data. While it is understood that the transformation $(\boldsymbol{XX'})^{-1}(\boldsymbol{X'Y})$ selects the model with the best linear fit, it is important to maintain perspective that "best" in this case is defined relative to the other possible mappings (i.e. the mapping for the data set which is linear and minimizes mean square error). If all the possible mappings are poor representations of the data, then the regression model can at best be described as the "least bad." The next section discusses quantitative metrics which are used to assess the "goodness" of a multiple linear regression model. These metrics are used to assess whether 1) the data is linear and 2) the constructed model accurately captures process dynamics.

5.2.3 Evaluating the Regression Model

Researchers who employ statistical modeling in their work rely on statistical metrics to assess how well the model represents the observed data. The best measure for goodness-of-fit is the R^2 value, defined in Chapter IV as the ratio of the model sum of squares SS_M (the sum of the squared distances from the mean of the data set to the regression line) to the total sum of squares SS_T (the sum of the squared distances from the mean value of the data set and each observed value). SS_M, SS_T, and SS_R (the sum of the squared distances between the observed data and the regression line) are tabulated as part of the analysis of variance in Table 16. R^2 (based on the sum of squares interpretation) is thus 0.813, a value which represents the proportion of the variation in the outcome variable which is accounted for by the variations in the three predictor variables (ω, ν, and l). The Pearson correlation coefficient R is the square root of this value ($R = 0.902$ for the three-predictor model).

Table 16 ANOVA table for tool wear model with ω, ν, and l as predictors

	Sum of	**Df**	**Mean**	**F**	**Sig**
Regression	812.74	3*	270.91	33.33	1.5E-
Residual	186.90	23	8.12		
Total (SS_T)	999.65	26°			

*coincides with number of predictor variables

†*Df = number of test cases – (number of predictors + 1)*

° *Df = number of test cases – 1*

Another metric used to determine goodness-of-fit is the F value, the ratio between the mean sum of squares associated with the model MS_M ($MS_M = \frac{SS_M}{Df}$) and the mean sum of squares for the residuals MS_R (MS_R=$\frac{SS_R}{Df}$). The F value can be interpreted in practical terms as the ratio of the degree of improvement when the model (rather than the mean of the data set) is used to predict the outcome variable (MS_M) to the discrepancies between the model and the data (MS_R). Hence for a good predictive model, MS_M should be substantially greater than MS_R. The F ratio corresponds to a value on the x axis of the F-distribution. For an F-distribution with 26 degrees of freedom, an F ratio of 33.339 corresponds to a significance value (the area under the curve which is greater than or equal to F) of $1.5x10^{-8}$. This significance value represents the probability that we reject the null hypothesis (that the regression model is not an accurate representation of the data) in favor of the alternative hypothesis (that the regression model accurately reflects the process behavior manifested in the data set) when the opposite is true. The significance value is simply the probability of a type II error. For this particular model, the significance value is very close to zero, an indication that the likelihood of the data set occurring by chance (i.e. that the outcome variable is essentially uncorrelated to the factors) is very small.

The metrics which reflect something about the overall model (F, R, R^2) are promising. However, these metrics consider the sum total of the effect of the predictors on the outcome variable. It is important to consider additional metrics which assess the significance of the predictors individually, rather than collectively. For an individual assessment of significance, the t-test is applied to each of the β coefficients. Table 17

shows each coefficient, the associated t-score, and the significance value. The t statistic is the ratio of the coefficient to the standard deviation and corresponds to a point on the t-distribution for All coefficients are significant at the .01 level. This result means that there is a very small probability (less than 1 percent) that the null hypothesis ($\beta = 0$) has been incorrectly rejected in favor of the alternative hypothesis (that β associated with a factor is equal to the value given by the regression model).

Table 17 Factor coefficients and significance values

	$\boldsymbol{\beta_i}$	**Significan**
ℓ	0.584	0.000
N	-1.038	0.005
ω	0.009	0.000

The diagnostics discussed thus far indicate that the model is a good fit for the data set: R is close to 1, the F-statistic shows that the regression model is significant at the 0.01 level, and t-tests confirm the significance of each coefficient. It is clear from the results of these analyses that the multiple linear regression model accurately reflects trends in the data set. However, the question remains as to whether the model can predict wear for parameters which are different from those used in the original design matrix. The data used to define the regression model represents a sample taken from the larger population (the population for this study would include all possible combinations of traverse rate, rotation speed, and length of weld). If the model is said to "generalize well", this means it accurately and consistently predicts the outcome variable for values of factors different from those used to construct it. It is not safe to assume that any

model, even the best-fitting one, will generalize well. In fact, it is sometimes the case that a model is an exceptionally good fit for a sample, but sees its predictive capability dramatically reduced when applied to the larger population. If a model does not generalize well, its application is limited to the (perhaps narrow) data set that was included in the regression analysis

In most scenarios, an assessment of the model's scalability relies upon cross-validation techniques. Cross-validation entails performing additional experiments (at parameters different from those included in the original design), measuring the outcome variable, and comparing the value predicted by the regression model with the value observed. There are, however, some techniques which do not call for the generation of additional data points. The easiest of these is to simply calculate the adjusted R^2 value based on the number of test cases n, the number of factors k, and the correlation coefficient R^2:

$$R^2{}_{adjusted} = 1 - [(\frac{n-1}{n-k-1})(\frac{n-2}{n-k-2})(\frac{n+1}{n})](1\text{-}R^2) \qquad \text{(equation 5-3)}$$

$R^2{}_{adjusted}$ is an estimate of the proportion of the variation in the outcome variable that would be accounted for *if* the model were based on the general population rather than the sample. $R^2{}_{adjusted}$ is thus less than R^2; how close these values lie to one another is an indicator of the model's scalability. If the difference between $R^2{}_{adjusted}$ and R^2 (a value referred to as data shrinkage) is small, then the model is expected to generalize well. This is the case for the wear regression model, where $R^2 = 0.813$ and $R^2{}_{adjusted} = 0.751$. Conversely, a large discrepancy between these values means that the model's scope is likely limited to the original data set. As a statistical metric, $R^2{}_{adjusted}$ gives

some indication as to whether the model's predictive capabilities will be diminished when applied to the population.

A second validation technique which requires only existing data is data splitting. In this method, the data is divided exactly in half: one-half of the data is used to construct a multiple regression model, while the other half of the data is used to test the model. Values predicted by the regression equation are directly compared with observed data; the error between predicted and observed values gives some indication as to whether the model can be generalized to the population. The caveat for this method is that it should only be used when there are a very large number of data points (typically numbering in the hundreds). Since the experimental matrix for this study is small, losing half of the parameters would make it difficult to construct a model which captures the wear process for the entire range of process parameters present in manufacturing. To implement data splitting for this application, the size of the design matrix would need to increase at least three-fold. Such an increase is not feasible due to costs associated with materials and tooling.

An alternative cross-validation technique was applied to the wear regression model, a choice made based on the small size of the sample (which makes data splitting impractical). Furthermore, the $R^2{}_{adjusted}$ metric is an estimator based on the work of Stein and, as such, is not a definitive indicator of scalability. The most traditional cross-validation technique requires the investigation of a few additional test cases which are disparate from those used to construct the regression model. This technique is similar to machine learning, in which a machine's decision-making/predictive ability is assessed by testing it on data that is not part of the data set used for training. For testing of the wear

regression model, three parameters were selected as test cases. While these parameters (summarized in Table 18) were not included in the original study, they do lie within the range of the factor values originally considered. Conditions for these welds were identical to those used in the prior experiments: materials, tool geometry, plunge depth, tilt angle, dwell time, and joint configuration remained unchanged. After each experiment, the wear of the tool was measured using the same imaging technique outlined in section 5.1.5. The predicted value of wear (calculated by substituting the values of the factors for each case into equation 5-1) can then be compared with the observed value (Table 19). The residuals (defined as the difference between the actual and predicted values) are tabulated in this table. The residuals can be normalized in terms of percent error by simply dividing the number in the residual column by the observed value and converting to a percent. Overall, the residuals are very small and the percent error is below 10 percent for all but case A (1250 RPM/5 IPM/8 inches), for which the difference between the actual and predicted value is 13.4%.

Table 18. Parameters for cross-validation study

Test	**ω (rpm)**	**v (in/min)**	**ℓ**
A	12	5	8
B	17	9	8
C	17	9	16
D	14	6	8

Table 19 Comparison of observed and predicted percent wear values for test cases

Test	Percent Wear	Percent Wear	Residual	Percent Error
A	5.45	4.72	0.73	13.39
B	4.95	5.05	-0.1	-2.02
C	10.06	9.72	0.34	3.38
D	5.45	5.02	0.43	7.89

On average, the model tends to underpredict the amount of wear experienced by the tool. Uncertainties surrounding wear incurred during the initial plunge/dwell time are likely the primary source of error for the model (errors are consolidated into the constant term of the regression equation), an effect which is exacerbated for longer welds. Since samples purchased from the manufacturer have been sheared into 8 inch long segments, a length of 24 inches actually consists of three separate 8 inch welds. This is problematic but ultimately unavoidable: even for a continuous length of material, continuation of the experiment after each periodic offline evaluation of the tool requires a new dwell period before welding can commence. The plunge/dwell period, which in our experiments occurs at the indicated rotation speed but a slower traverse rate (0.5 inches per minute), is essential to prevent fracture of the tool. The conditions for this study were as follows: the tool was advanced into the material from the side (thus eliminating the "plunge" phase) and the dwell time was 60 seconds for each weld (the tool advances 0.5 inches along the weld seam during this time). After one minute, the traverse rate is abruptly increased to coincide with the welding velocity specified in the experiment. While the effects of dwell time on wear are inherent in the regression model, it is anticipated that more wear (per inch) occurs during the dwell period because the shear zone is narrower than during steady-state welding. The hypothesized effect of the shear zone on wear is discussed and

tested in the next chapter. For now, we will only mention that the shear zone is not fully developed in the early stages of the weld (the dwell period) due to a lack of heating (shear zone increases in width with temperature). A thinner shear zone is believed to promote abrasion of the tool by the reinforcement particles. The best way to assess wear during dwell time versus wear during steady-state welding is to perform a full-length weld at the dwell conditions and measure the amount of wear. It is expected that more wear will occur in the dwell period since the traverse rate is much slower (0.5 inches/minute for our experiments) than in steady-state welding. However, the length of material traversed at the dwell parameters (approximately 0.25-0.5") may be sufficiently small to render this phase's contribution to the overall wear of the tool negligible.

If the regression model were to be applied in an actual manufacturing scenario, operators should simply use it as a means to roughly estimate the amount of wear the tool will incur during the course of the welding routine. The cross-validation study shows that the observed wear value should lie close to the predicted value, but may not precisely match it due to the (in many cases uncontrollable) variations in the welding environment: the process dynamics during the dwell period, the evolution in temperature profiles for the backing anvil, workpiece, and tool, as well as the ambient temperature and humidity. While all of these factors probably have some effect on the amount of wear (and thus may be responsible for the slight discrepancies between the predicted and actual values in Table 13), the regression model is able to account for most of the variation in wear (81 percent) by considering only three major process parameters: rotation rate, welding speed, and length of joint.

5.2.5 Dimensional Analysis and Scaling

Even with cross-validation of the regression model, concerns about scalability persist. With any empirically based predictive model, there are lingering questions as to whether predictions are only valid under the specific experimental conditions used in the study from which the data originates. A similar dilemma is confronted by drug researchers in clinical trials with regard to drug dosage (i.e. will the drug have the same observed effect on patients who differ in some ways from those included in the initial study). In fluid dynamics, the problem of scalability manifests itself as an attempt to understand the effect of changes in length, viscosity, velocity, and density on flowfield characteristics without generating immense amounts of data which capture every forseeable combination of flow conditions. Both scenarios call for a scaling factor: a multiplier which can be used to easily gauge the quantitative effect of changes in the factor(s) on the outcome variable. In some cases, the relationship between the factor and outcome can be expressed succinctly as $y = cx$, where c is the scaling parameter (c may correspond to an actual physical quantity). In the drug dosage example, c most likely represents some variant of patient weight (for example, heavier patients will require a larger volume of a drug to achieve the concentration of medicine in the bloodstream that will produce the desired effect). The dosage for any patient would be calculated by multiplying the patients weight x, by c, a multiplier which adjusts the baseline dosage to account for how the patient differs in weight from the "average" patient.

In fluid dynamics, the most widely-used scaling factor is a dimensionless quantity known as the Reynolds number ($Re = \frac{\rho V L}{\mu}$ where ρ is the density of the fluid, V is the

flow velocity, L is a characteristic length, and μ is the fluid viscosity). The group is termed "dimensionless" because the units of the terms cancel, resulting in a pure number. For any dimensionless group, there are multiple combinations of values of the constituent variables that will yield the same dimensionless number, a property which can be exploited for the purpose of scaling. The usefulness of dimensionless groups hinges on the hypothesis that the system response is the same for equivalent values of the dimensionless number, a property known as similitude. The system behavior is thus determined by the value of this number rather than the values of the variables which comprise it. In fluid dynamics, similitude is exploited for the purposes of scaling: the object being analyzed can be made smaller as long as the other flow conditions are proportionately increased to maintain the same Reynolds number.

Since the process parameters associated with FSW of MMCs are highly specific to the test bed, it is important to develop a scaling factor for this application. The range of FSW process parameters is highly dependent on the apparatus and application. For instance, the PDS (Process Development System) at NASA MSFC has traverse capabilities which are comparable to the VUWAL apparatus, but its rotation speeds are much slower (typically in the 100 RPM to 1000 RPM range). The backing anvil and clamping system for the PDS can also accommodate a much larger weld sample (up to 36" in length, 8" in width, and 1" in thickness). Even though the cross-validation procedure documented in the previous section alleviates some apprehension about scalability, applying the regression model to the PDS with the expectation that it will accurately predict results for parameters well outside the range for which it was developed is a dubious proposition. A more viable option to extend the applicability of

the model is to develop a dimensionless group N, consisting of some combination of the factors ω, ℓ, and ν, which is strongly correlated with the outcome variable percent wear W. The relationship between wear and *N* is thus of the form $W = cN$, where c is a proportionality constant.

The derivation of the group N is rather straightforward. The step-by-step method of Ipsen was used to generate a list of all possible relationships between the variables ω, ℓ, and ν for which the units cancel [99]. These groupings are referred to as π groups, a name derived from the Buckingham π theorem which serves as the basis for this type of analysis. The regression model indicates that wear of the tool is a function of the three major process parameters: $W = f(\omega, \nu, \ell)$. Since we have four physical quantities, m in the Buckingham-π theorem is equal to 4. Wear is a pure number since the loss in cross-sectional area was normalized with respect to the original cross-sectional area of the tool and expressed as a percent in the prior analysis (section 5.1.5). Thus there are really only 3 physical quantities (*m=3* in the Buckingham π formulation. ω is an angular velocity measured in rotations/minute; ν is a linear velocity with units of inches/minute; ℓ is a length measured in inches. Table 20 summarizes these quantities and expresses them in terms of the fundamental units of time T and length L. Since there are only 2 fundamental units, *r* in the Buckingham-π theorem is set equal to 2. The theorem states that for *m* physical quantities with *r* unique fundamental units, there are *p=m-r* dimensionless groups. Based on this equation ($p = 3 - 2$), there is only one possible dimensionless group which can be formed from the quantities included in the regression model. This group is evident upon inspection. The center quantity, traverse rate, has the fundamental unit $\frac{L}{T}$. When this quantity is divided by rotation speed ($\frac{1}{T}$) the

fundamental unit of time T cancels. L, the fundamental unit of length, can be similarly eliminated by dividing the traverse rate by the distance welded. The dimensionless group associated with these variables is thus $\frac{v}{\omega \ell}$.

Table 20. Physical quantities, units, and fundamental units for dimensional analysis.

Physical Quantity	**Units**	**Fundamental Unit**
Rotation speed	$\frac{rotations}{minute}$	$\frac{1}{T}$
Traverse speed	$\frac{inches}{minute}$	$\frac{L}{T}$
Distance welded	$inches$	L

A more generalized dimensional analysis of process variables in FSW conducted by Hendricks, Cox, Gibson, and the author in the summer of 2009 arrived at a similar result. This comprehensive study considered sixteen variables which may be of interest in the FSW process: the thermal conductivities, specific heats, and yield strengths of both the tool and workpiece; the elastic modulus and density of the workpiece; the ambient temperature and the maximum weld temperature; the surface area of the pin; and the process parameters traverse rate, rotation speed, and length of joint. These physical quantities appear alongside their fundamental units in Table 21. There are 4 unique fundamental units associated with these quantities: mass M, length L, temperature θ, and time T. According to the Buckingham-π theorem, there are 12 possible dimensionless groups. These groups were generated using the dimensional analysis toolbox in MATLAB (since constructing an elimination table for this many variables would be time-

consuming and tedious). These groups are listed in Table 22 (variables are defined in the legend below the table).

Table 21 Physical quantities and fundamental units for variables in the FSW process.

Physical quantity	**Fundamental unit**
Thermal conductivity	$\frac{ML}{\theta T^3}$
Elastic modulus	$\frac{M}{LT^2}$
Yield strength	$\frac{M}{LT^2}$
Viscosity	$\frac{M}{LT}$
Density	$\frac{M}{L^3}$
Temperature	Θ
Specific heat	$\frac{L^2}{\theta T^2}$
Surface area	L^2
Traverse speed	$\frac{L}{T}$
Length	L
Force	$\frac{ML}{T^2}$
Torque	$\frac{ML^2}{T^2}$

Table 22 Dimensionless Groups (results of MATLAB analysis)

	Group
π_1	$\frac{k_m}{k_t}$
π_2	$\frac{E\ell^2}{F}$
π_3	$\frac{Y_{st}\ell^2}{F}$
π_4	$\frac{Y_{sm}\ell^2}{F}$
π_5	$\frac{\rho v^2 \ell^2}{F}$
π_6	$\frac{T_{max}k_t\ell}{vF}$
π_7	$\frac{T_{init}k_t\ell}{vF}$
π_8	$\frac{c_tF}{k_tv\ell}$
π_9	$\frac{c_mF}{k_tv\ell}$
π_{10}	$\frac{SA}{\ell^2}$
π_{11}	$\frac{\tau}{\ell F}$
π_{12}	$\frac{\omega\ell}{v}$

k_m ≡ thermal conductivity of workpiece
k_t ≡ thermal conductivity of tool
c_m ≡ specific heat of workpiece
c_t ≡ specific heat of tool
Y_{sm} ≡ yield strength of material
Y_{st} ≡ yield strength of tool
E ≡ modulus of elasticity
F ≡ axial force
τ ≡ torque
v ≡ traverse rate
ω ≡ rotation speed
ℓ ≡ length of joint
SA ≡ surface area of pin
ρ ≡ density
T_{max}≡ maximum tool temperature
T_{init} ≡ initial (ambient) temperature

The Buckingham-π theorem is used to generate all the possible dimensionless groups which can be constructed using the variables considered. However, the theorem gives no indication as to whether the groups which arise from the analysis have any physical significance. The best method of assessing significance is to plot each group against the outcome variable to be optimized (for instance, tensile strength, wear or fatigue life) and calculate the correlation coefficient R which exists between the group and the outcome. The distance between R and 1 is an indication of the ability of the group to predict the variable of interest. The closer R lies to 1, the greater the significance of the group and its constituent variables with respect to the outcome.

Ideally, the ability of each dimensionless group to predict wear would be determined on an individual basis using the R value. The correlation coefficient can then be used as a metric to comparatively rank the fit of the linear relationships between each group and the outcome from best-fitting (1) to worst-fitting (n, where n corresponds to the number of π groups). Unfortunately, such a comprehensive assessment is not feasible for most of the dimensionless groups listed in Table 22 using the existing data matrix, as the experiments in the study were designed to test only three specific variables (the weld process parameters) hypothesized to have the greatest influence on wear. While factors such as the yield strength, elastic modulus, thermal conductivity, and specific heat of the tool and workpiece were constant for these experiments, it may be the case that their variation also influences behavior of the wear system (the results of experiments designed to test the effects of material properties on wear appear in chapter VI and VII). Other variables in Table 21 which were not included in the initial wear study include: temperature, force, torque, tool geometry, and temperature. The omission of these

variables can be attributed to limitations in the sensing capabilities of the FSW apparatus as well as logistical concerns. It is essential (particularly when cost is a concern) that researchers limit the number of variables to ensure that the number of experiments falls within the range of what can be reasonably executed by the researcher (while it is tempting to consider the effect of all sixteen variables simultaneously, the associated design matrix would contain hundreds of experiments). With regard to the former point, torque and force were not recorded for many of these welds due to an equipment failure. A dyanamometer breakage resulted in the loss of sensing capabilities for torque along with forces in the x, y, and z directions. Although detection of axial load and torque was eventually recovered using a system of strain gauges and wireless force transducers, these sensors were still in the development phase at the time the Taguchi experiments were performed.

The decision not to collect temperature data for the wear experiments was related to several factors. Thermocouples, the simplest and most widely used means of temperature logging, are difficult to implement for the FSW process due to positioning. A thermocouple can be embedded in the workpiece material, but it will always lie at some distance from the tool (since placement of the sensor directly in the path of the tool will result in destruction of the sensing element). The distance between the sensor and the tool makes it difficult to gauge the variable of interest, tool temperature, with much accuracy. It is possible to roughly estimate the temperature at the tool surface using Fourier's law for steady-state conduction when the measured temperature, the distance from the thermocouple to the tool, and the thermal conductivity of the tool and the workpiece are known. A second option for temperature recording is a thermal camera.

Since the metals typically used for FSW are highly reflective, it may be necessary to coat the workpiece prior to welding with a spray of known emissivity to suppress reflection and improve sensing. Additionally, temperature was not included because it has been demonstrated that wear rates in abrasive wear are largely insensitive to temperature (wear is presumed to be abrasive since the tool is in contact with abrasive particles) [49]. Temperature has not been neglected completely, however, since the steady state temperatures of the tool and the workpiece are themselves governed by other variables which *are* measured in these experiments. According to Schmidt's model of heat generation in friction stir welding, the amount of heat input (and thus the weld temperature) is controlled by the process parameters rotation speed and traverse rate, the tool geometry (specifically the cubed difference between the shoulder diameter and pin diameter), and plunge depth [100]. Based on Schmidt's analysis, the effect of temperature on wear may be at least partially captured by variations in process parameters (for instance, higher weld pitches $\frac{\omega}{\nu}$ correspond to greater heat input). Neither have we permanently disregarded the potential for a relationship between wear and material properties (explored in this chapters VI-VIII) nor the possible link between force, torque, and wear (assessed in chapter IX). At each step of the investigation, dimensional analysis is included where appropriate (i.e. when all of the variables which comprise one of the dimensionless groups in Table 22 are accounted for).

Since it is not possible to investigate the relationship between all twelve dimensionless groups and wear using the existing data, we instead focus on the dimensionless group identified as π_{12}. The variables in this group (ω, ℓ, and ν) coincide with those included in the Taguchi study. Note that this group is the reciprocal of the π

group previously derived by hand using the Buckingham-π theorem. This dimensionless group $\frac{\omega \ell}{\nu}$ is plotted against percent wear in Figure 16. The value of the dimensionless group (calculated by substituting the levels for each specific case into the relationship $\frac{\omega \ell}{\nu}$; the numerator has been multiplied by the factor $\frac{2\pi}{rotation}$ to eliminate the unit associated with the ω term) is plotted along the horizontal axis; the corresponding range of wear values is represented on the vertical axis. The relationship fits the linear relationship of the form $W=cN$, where N is the dimensionless group and c is the proportionality constant (synonymous with the slope of the line of best fit). The Pearson correlation coefficient for this data is 0.82, a result which indicates that the wear data collected in the Taguchi experiments is dependent on $\frac{\omega \ell}{\nu}$. The equation for the line of best-fit is $W = 0.0004 \frac{\omega \ell}{\nu}$. Physically, this equation means that each unit increase in the group $\frac{\omega \ell}{\nu}$ is accompanied by a .0004% increase in percent wear. This constant, like the constants in the multiple regression equation (5-2), can be used to estimate relative changes in wear based on process parameters: for instance, a 10,000-fold increase in $\frac{\omega \ell}{\nu}$ would correspond to a 4 percent increase in percent wear. The plot in Figure 63 consists of only data from the original matrix The second plot (Figure 17) incorporates data collected as part of the cross-validation study. The inclusion of these additional data points does not improve the fit of the model, which remains constant at $R = 0.82$ or alter the proportionality constant (c=0.0004 for both models). Isolation of this data (in Figure 64, the cross-validation data are represented by the red diamond markers) demonstrates that the cross-validation study closely follows the same trend observed for the Taguchi matrix: that is, wear generally increases with increasing $\frac{\omega \ell}{\nu}$. The fact that the trend holds for the cross-

validation data as well is evidence that the applicability of this relationship extends beyond just the cases considered in the Taguchi matrix.

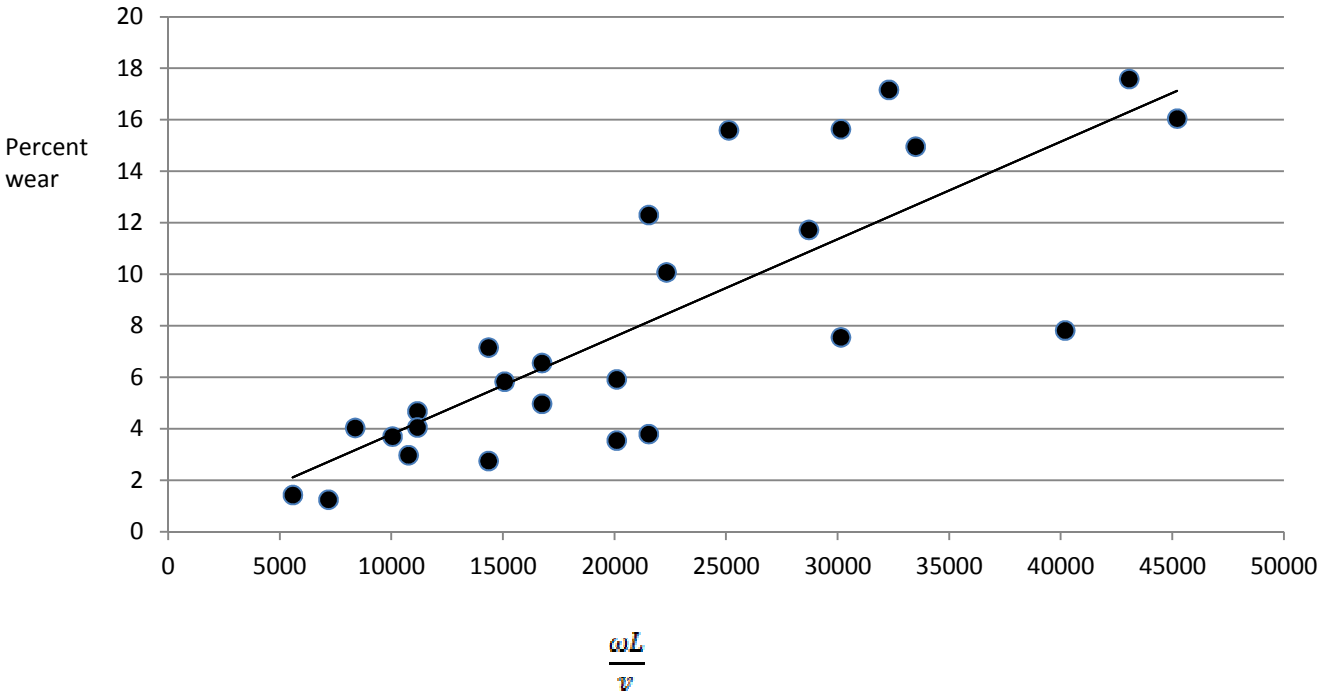


Figure 63 Plot of wear versus dimensionless group $\frac{\omega \ell}{\nu}$.

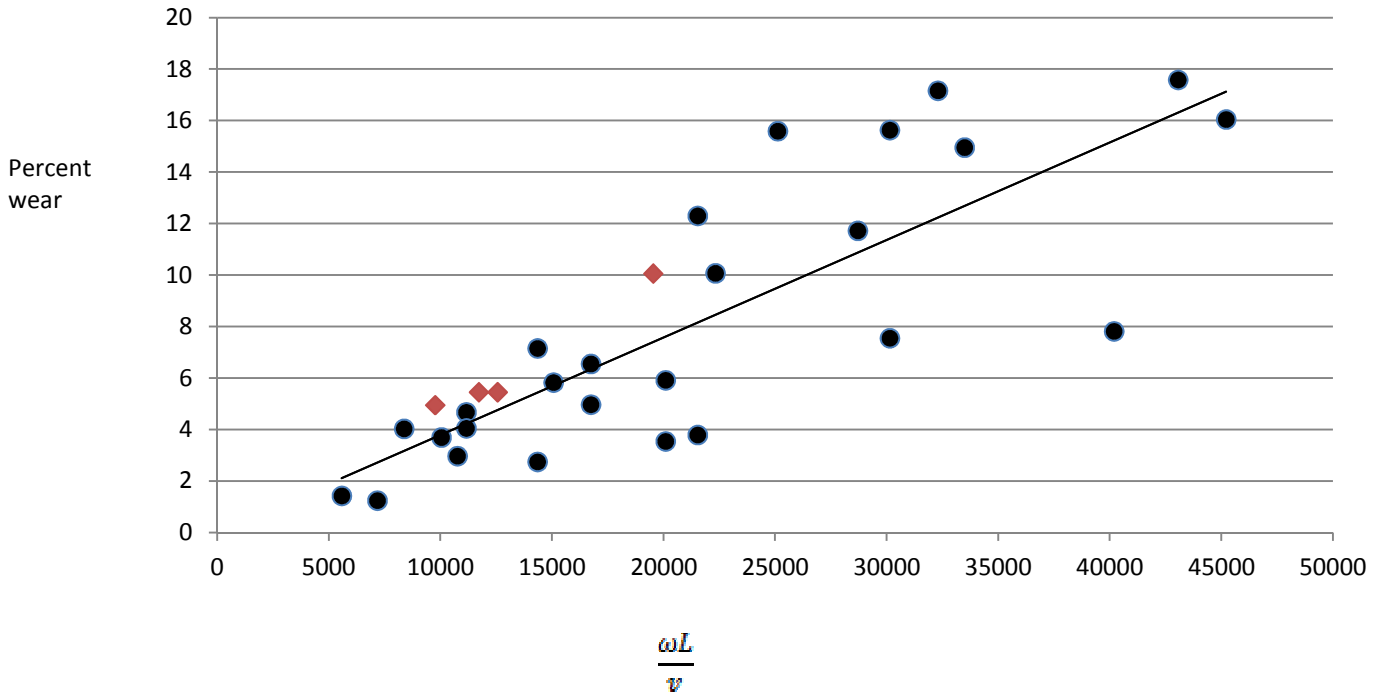


Figure 64 Plot of wear versus dimensionless group $\frac{\omega \ell}{v}$ (Wear from the original Taguchi study is represented by a black dot; data from cross-validation study is indicated by a red diamond).

The representation of wear as a dimensionless group is especially important when considering scalability. For example, the successful determination of a dimensionless group for wear in FSW of MMCs presents a useful means of solving the manufacturing dilemma posited in section 5.2.1. The scenario posed was that of a weld operator tasked with identifying sets of parameters corresponding to wear values which lie below some critical threshold (this threshold is specific to the application and material and approximately coincides with the point at which the changes in tool shape which accompany volume loss result in weakening of the joint to an unacceptable degree). It is assumed that the operator in this situation knows the length of the weld as well as the critical value of material degradation and wants to identify combinations of rotation rate

and traverse speed which will maintain wear below this level. Knowledge of a dimensionless number and how it varies with wear can be used to pinpoint the conditions which keep wear values below the defined threshold. From this list of possible parameters, the operator can then select those which are compatible with the process limits of the specific FSW apparatus yet maintain an acceptable level of weld quality (the data used to make assessments would be determined using a parameterization study similar to those documented in references 34-36,39).

The development of a dimensionless parameter is potentially a very powerful predictive tool, but its practical application is complicated somewhat by the question of uniqueness. It is apparent from the graphs in Figures 63 and 64 that there are some values of $\frac{\omega \ell}{\nu}$ which correspond to multiple values of percent wear. While many of the values associated with a particular value of $\frac{\omega \ell}{\nu}$ are clustered close together (for instance the two cases in the cross-validation study 1250 RPM/5 IPM/8 inches [$\frac{\omega \ell}{\nu}$=12,560] and 1400 RPM/6 IPM/8 inches [$\frac{\omega \ell}{\nu}$=11,723] produce identical values of wear), there are a few instances where nearly equivalent values of $\frac{\omega \ell}{\nu}$ correspond to divergent results (this is the case for 1000 RPM/5 IPM/24 inches and 1500 RPM/5 IPM/16 inches; these parameters have the same value of $\frac{\omega \ell}{\nu}$ yet the former weld results in 8 percent less wear than the latter). This may be a consequence of the asymmetry of effects: as detailed in section 5.2, the process variables do not carry equal weight (i.e. a change in one variable may have a greater impact on wear than a proportional change in another variable). Since the mapping between $\frac{\omega \ell}{\nu}$ and wear is not always one-to-one, a better option may be to divide

the data into classes and use $\frac{\omega \ell}{\nu}$ to predict the range of the expected wear, rather than an exact percentage of volume loss. For the bar graph in Figure 65, the wear data has been partitioned into two categories: unworn (<10 percent total wear) and worn (≥10 percent).* The proportion of data points which fall into each category is plotted for 5000 unit increments of the dimensionless number. It is evident from the bar graph that the low wear regime (<10 percent total wear) dominates for numbers which are less than 20,000. From 20,000 to 30,000 wear becomes more prevalent, as the occurrence of worn tools increases to account for 60 percent of the data in this range Beyond 30,000 the data is dominated by tools which meet the criteria (>10 percent wear) to classify them as "worn." For, the last half of the plot (30,000 to 50,000), the ratio of worn to unworn tools is 6 to 1. Stated probabilistically, this means that the value of $\frac{\omega \ell}{\nu}$ is greater than 30,000, the likelihood that it will be "worn" (exhibit greater than 10 percent wear over the course of the weld) is 86 percent.

*The designation of a tool as "worn" or "unworn" based on the amount of wear is somewhat arbitrary. "Worn" in this instance refers to volume losses in excess of the critical value at which defects begin to form due to degradation in tool shape. Classifying a defect as acceptable or unacceptable based on size (and linking the size of the defect to the amount of wear), a topic which is critical to quality control and weld qualification of friction stir welded joints, is beyond the scope of this research. Our choice of 10 percent to mark the boundary between "worn" and "unworn" is an educated assessment based on previous papers in this research area and a knowledge of how tool shape influences flow patterns. Typically, a flat cylindrical geometry can accommodate a small amount of wear without a reduction in flow characteristics which promote material consolidation. The value which we have assumed to mark the division between "worn" and "unworn" (10 percent) is the point at which noticeable wear has occurred. When wear is discernable with the naked eye, the amount of material loss is sufficient to cause problems (i.e. reduced flow, conditions which promote void formation, significant deposition of worn material along the jointline). Wear of this magnitude can be taken as an indication that the tool needs to be replaced. Similar plots can be constructed for any critical value of wear designated by the researcher/experimenter.

Figure 66 is an extension of this thinking – the curve represents the cumulative, conditional probability that a tool is worn given that the value of the dimensionless number is greater than or equal to the value on the x-axis (the probability is "conditional" since we have restricted the data from which the probability is calculated). As in the bar graph, it is clear that the probability of observing a worn tool increases as the value of the dimensionless number becomes greater. There is a segment of the plot (the range 30,000 to 45,000) for which the probability plateaus, a stagnation which is probably related to the comparatively fewer number of data points which fall in this range. The scarcity of data points in the upper ranges of the dimensionless number (a consequence of the parameter compatibility issues discussed previously) means that while the raw number of unworn tools observed is less than those concentrated at the lower end of the $\frac{\omega \ell}{\nu}$ spectrum, the proportion of unworn tools may increase simply because there are so few total data points which lie above a particular value. If we were to conduct multiple experiments at different parameters which correspond to values of 30,000-45,000 for the dimensionless number, we would expect (based on our hypothesis) that the proportion of worn tools would far exceed those of the unworn (a scenario analogous to flipping an unfair coin). The analysis presented here is limited since it was based solely on data from the Taguchi study of chapter V, which was designed to test the effect of process parameters on tool wear rather than the influence of $\frac{\omega \ell}{\nu}$ (a dimensionless number that was developed, in part, based on the results of the Taguchi study). Wear at higher values of the dimensionless number will be further evaluated in an ancillary study included in chapter IX (which discusses validity of the research, conclusions, and future work). The take-home message of the probability plot in Figure 66 is this: in general the likelihood of

observing a "worn" tool increases with the dimensionless number. Eventually a condition will be reached such that 100% of tools wear whenever the dimensionless number exceeds some critical x-value.

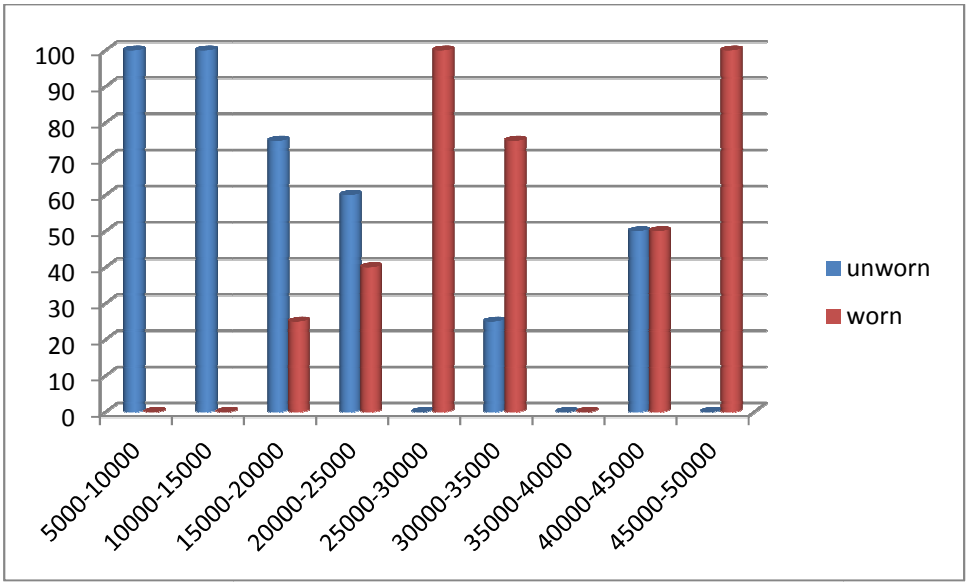


Figure 65 Proportion of tools (expressed as a percentage of the total data points within range of $\frac{\omega \ell}{\nu}$ specified on the x-axis) that classified as "worn" or "unworn." "Unworn" indicates the tool exhibits less than 10 percent wear. "Worn" tools are those which have lost greater than 10 percent of the probe volume. There are no data points which fall in the 35,000 to 40,000 range.

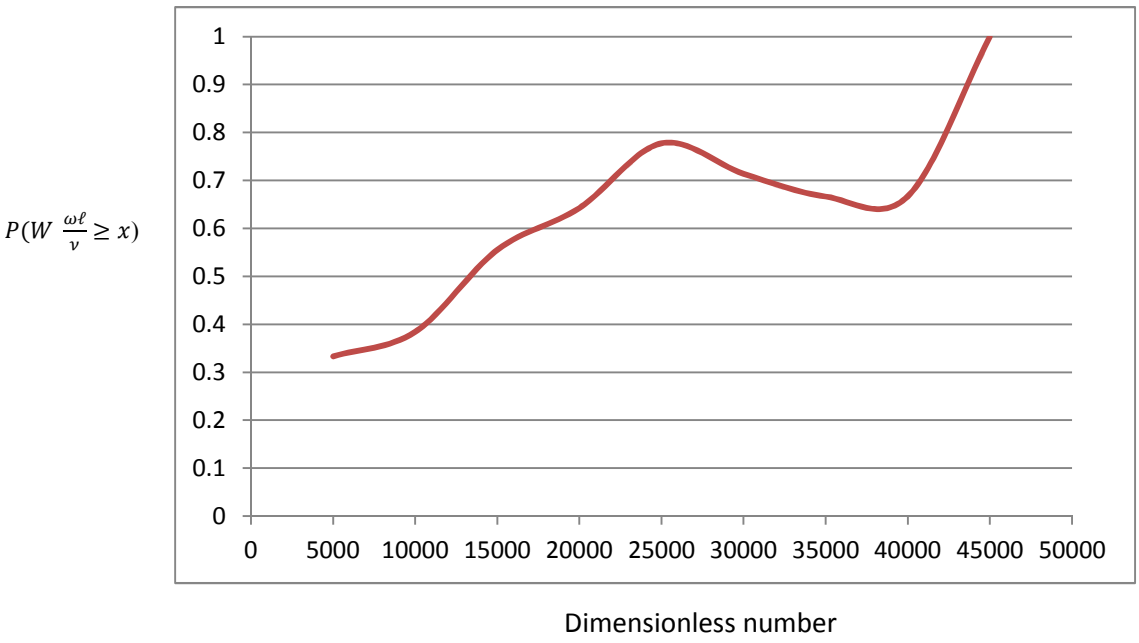


Figure 66 Plot of the cumulative conditional probability that a tool is worn $P(W)$ given that its dimensionless number is greater than or equal to the value indicated on the x-axis $P(W \frac{\omega\ell}{\nu} \geq x)$.

The graphs in Figures 65 and 66 can be used to visually identify the value of the dimensionless number which separates the "worn" tools from those which are "unworn" (in our example, "unworn" corresponds to tools which exhibit less than 10 percent wear). The concept of a transition point for wear is similar to the critical Reynolds number in fluid dynamics, which delineates the division between the laminar and turbulent flow regimes. The critical Reynolds number can be used to identify the location along the geometry where this transition occurs (the Reynolds number is length-dependent) as well as inform design changes to the geometry which delay the onset of turbulence. For flow over a flat plate, the critical Reynolds number Re_{cr} is approximately $5x10^5$. The flowfield exhibits laminar characteristics for conditions corresponding to Reynolds

numbers below this value; the turbulent flow regime is expected only for values which exceed Re_{cr}. Although the transition between regimes is modeled as an abrupt step change (wherein the flow changes from entirely laminar to turbulent at a specific point), in actuality the transition phenomena is continuous. As the Reynolds number approaches the transition point, the flow begins to exhibit less laminar characteristics. Flow with Reynolds numbers which only slightly exceed Re_{cr} are more turbulent than laminar, but may still have some laminar components. The division between laminar and turbulent flow is not a clear one – rather, the Reynolds number is an indication as to which flow regime (laminar or turbulent) is dominant at a particular location. In aerodynamics, the fundamental transition number is the Mach number (the ratio of freestream velocity to the speed of sound). For Mach numbers greater than 1, the flow is supersonic; subsonic flows have Mach numbers less than 1. Unlike the critical Reynolds number (which depends on a number of factors, including the geometry of the surface), the Mach number is a rigidly defined delineation for all flows. Like the Reynolds number, the Mach number is a discretization of a continuous phenomenon. For example, there may be portions of an airfoil immersed in a flow with a free-stream subsonic Mach number which exhibit characteristics of supersonic flow (as any commercial passenger who has experienced turbulence due to the formation of shock waves on the wing can attest). While binary classification schemes based on Reynolds number, Mach number, and our own dimensionless parameter may overlook nuances such as these, the value of binary classification to analysis and design cannot be overstated. We seek to implement a similar schema for the wear data of chapter V: first by graphical estimation and then by applying linear discriminant analysis (LDA) to the data.

For our data, it appears that low wear (below 10 percent) is predominant for smaller values of $\frac{\omega\ell}{\nu}$; higher wear (the "worn" regime) emerges when $\frac{\omega\ell}{\nu}$ is increased past some transition point, identified as W_c. Like the critical Reynolds number, W_c divides the spectrum of data into two regions. Values of $\frac{\omega\ell}{\nu}$ which exceed W_c correspond to the conditions under which a probe volume loss of greater than 10 percent is expected; values below W_c characterize the low wear regime (less than 10 percent volumetric deterioration). Visual identification of the critical wear number is complicated by the data's asymmetry with respect to the range of $\frac{\omega\ell}{\nu}$ (as mentioned previously, the plots in Figures 19 and 20 were constructed using data from the Taguchi matrix in chapter V, which was designed to test the effect of process parameters on wear but not specifically the single, combined effect of the dimensionless ratio $\frac{\omega\ell}{\nu}$); the data are clustered at the lower end of the $\frac{\omega\ell}{\nu}$ spectrum since these are the parameters which were compatible with our FSW application. From the graphs, it appears that values of $\frac{\omega\ell}{\nu}$ which exceed 20,000 generally coincide with a volume loss of greater than 10 percent (the criterion established to designate a tool as "worn"). This is the point at which we expect that more than a majority of tools will exhibit noticeable wear: based on Figure 20, the probability that the tool is "worn" given that its $\frac{\omega\ell}{\nu}$ value is greater than 20,000 is approximately 60 percent. The unworn regime is relegated primarily to π group values less than 20,000.

As with the Reynolds number, we favor a continuum interpretation of this transition number. Parameters corresponding to values of $\frac{\omega\ell}{\nu}$ greater than W_c are not a guarantee that the tool will exhibit substantial wear, but rather an indication that the latter

scenario (where $W > 10\%$) is a likely outcome. The critical value of $\frac{\omega \ell}{\nu}$ thus has important implications for weld operators in that the selection of process parameters satisfying the inequality $\frac{\omega \ell}{\nu} < W_c$ help to preserve the shape of the tool. Preservation of tool shape is often crucial to weld quality since the material flow pattern (and subsequent consolidation of material along the weld path) is facilitated by the probe's features. For a given length of weld, the critical dimensionless number can be exploited to determine the values of the other process parameters (ω and ν) which will maintain volume loss below 10 percent. Conversely, if the parameters have already been selected, the dimensionless number can give the operator an estimate as to when the tool needs to be replaced ($\ell = \frac{\nu W_{cr}}{2\pi\omega}$). The use of the critical dimensionless number in this manner is quite similar to problems in fluid dynamics which call for the identification of the distance x along a flat plate at which the flow transitions from laminar to turbulent ($x = \frac{Re_{cr}\mu}{\rho V}$).

To this point, we have estimated (based solely upon visualization of the data collected in chapter V) that the critical value of the dimensionless number for tool wear in FSW in our data is 20,000. This is a very crude approximation and in general, graphical techniques alone are insufficient for these type of calculations. A more robust determination of the transition between wear regimes employs linear discriminant analysis (LDA). Although the mathematics of LDA are glossed over here (a more thorough treatment of the underlying theory can be found in chapter IX, where LDA is applied to process signals), the thrust of linear discriminant analysis is to derive a linear function which maximally discriminates between two categories of data (in our case, "unworn" and "worn") based on an input factor(s). The result of LDA is a continuous

linear function of the form $f(x_1,x_2,\ldots,x_n)$ (where n is the number of input variables) which can be used to predict the group to which an experimental case belongs. This classification is based on the sign of the value obtained when the input variables are substituted into the discriminant function. In order to make the data from the Taguchi matrix compatible with LDA, the dependent variable (wear) must be converted into a form consistent with a binary classification scheme. In our analysis, a "1" indicates that the tool has not lost enough volume to be considered worn (the discriminant function is negative), whereas a categorization of "2" coincides with the worn condition (in which case the discriminant function is nonnegative). The amount of tool material loss which can be accommodated is highly dependent on the application: wear in excess of $x\%$ may be tolerated in one structure yet deemed unacceptable in welds for structurally similar but mission-critical structures such as flight hardware. The percent volume loss which necessitates tool replacement (and constitutes "wear") may depends on many factors: the criticality of the application, the degree of wear which can occur without an unacceptable reduction in desired joint properties, and the initial cost of the tooling, among others. Since these factors are specific to the manufacturing scenario, we cannot use the data from chapter V to construct a "catch-all" classifier for wear based on the dimensionless parameter $\frac{\omega\ell}{\nu}$. The purpose of this investigation, rather, is to assess the feasibility of the dimensionless parameter to discriminate between two tool conditions. Since the criteria for classification are application specific, we have performed a separate LDA analysis for each of the possible criteria for classification using the data from chapter V. An LDA-based classifier was constructed for each of the cutoff values ranging from 3 percent (which means that tools which exhibit greater than 3 percent wear are considered "worn",

while all others are “unworn”) to sixteen percent (in which case the worn condition applies only to tools which lose more than 16 percent of their total probe volume). If the value of wear associated with a case is less than the cutoff value, the case is identified a belonging to group I; cases for which wear equals or exceeds the cutoff value are classified in category II. Table 23 demonstrates the application of this criterion to the Taguchi and cross-validation data of Chapter V for the 5 percent, 10 percent, and 15 percent cutoffs.

Table 23 Wear data from Taguchi experiments and cross-validation study (in form compatible with linear discriminant analysis)

Experiment Number	$\frac{\omega \ell}{\nu}$	Classification (5 percent cutoff)	Classification (10 percent cutoff)	Classification (15 percent cutoff)
1	10,048	I	I	I
2	20,096	II	I	I
3	30,144	II	I	I
4	7,177	I	I	I
5	14,354	I	I	I
6	21,531	I	I	I
7	5,582	I	I	I
8	11,164	I	I	I
9	16,747	I	I	I
10	15,072	II	I	I
11	30,144	II	II	II
12	45,216	II	II	II
13	10,766	I	I	I
14	21,531	II	II	I
15	32,297	II	II	II
16	8,373	I	I	I
17	16,747	II	I	I
18	25,120	II	II	II
19	20,096	I	I	I
20	40,192	II	I	I
21	60,288	II	II	I
22	14,354	II	I	I
23	28,709	II	II	I
24	43,063	II	II	II
25	11,164	I	I	I
26	22,329	II	II	I
27	33,493	II	II	I
28	12,560	II	I	I
29	9,769	I	I	I
30	19,538	II	II	I
31	11,723	II	I	I

The output of each linear discriminant analysis is the classification function $f(x) = a_1 x + a_0$, where x represents the value of the dimensionless parameter $\frac{\omega \ell}{\nu}$. The

discriminant function is similar in form to a linear regression equation with a single predictor variable: the factor term is multiplied by a weighted coefficient a_1 and then summed with another constant to yield a single value. However, in LDA this value is corresponds to a category. In the LDA presented here, negative discriminant scores correspond to group I ("unworn"), while a nonnegative scores places the case in group II ("worn"). Mathematically, this piecewise classification criteria is defined as:

$$\begin{cases} f(x) < 0, & \in group\ I \\ f(x) \geq 0, & \in group\ II \end{cases}$$

The discriminant function $f(x)$ uses the dimensionless parameter $\frac{\omega \ell}{\nu}$ as a way to predict which group a tool will belong to if it is used to join a metal composite at the indicated process parameters. The numerical output of this equation expresses the likelihood that the tool exhibits a degree of wear that is consistent with a particular group. For example if the discriminant score $f(x)$ is negative, it is more likely that the tool condition is "unworn"; when $f(x)$ is nonnegative, the tool condition is more likely to be "worn." In some respects, LDA mimics the internal workings of the sorting hat in J.K. Rowling's *Harry Potter* series, which assigns students to one of four houses (Hufflepuff, Ravenclaw, Slytherin, or Gryffindor) based on their personal characteristics.*

*The sorting hat would actually use a specialized form of LDA known as logistic regression since the categorical variable can assume more than two states (there are four houses of Hogwarts).

The LDA function in this application is used to predictively discriminate between two tool conditions ("worn" and "unworn") based on three process parameters ($\omega, \ell,$ and ν, expressed in terms of the π_{12}dimensionless group).

To assess the efficiency of the classification functions at discriminating between tool conditions in FSW of MMCs, the value of $\frac{\omega\ell}{\nu}$ associated with each of the 31 cases in the Taguchi and cross-validation study is substituted into each of the discriminant function. The predicted group is determined entirely by the sign of the numerical value of $f(x)$. The application of the classification function to the cases in the Taguchi matrix (in addition to those considered in the cross-validation study) is summarized for the 10 percent cutoff value in Table 24. The third column shows the discriminant score. The fourth and fifth columns compare the prediction of the LDA classifier with the actual classification based on the measured wear. Shading indicates cases for which a discrepancy exists between predicted and actual group membership. In SPSS, the classification of a case is based on the comparison of posterior probabilities in addition to the raw discriminant score. The posterior probability, calculated using Bayes' Theorem, gives the conditional probability that a case is in group *i* given that the discriminant function is equal to *k*. In the SPSS printouts, P $(G = I|D = k)$ denotes the posterior probability for category I, while $P(G = II|D = k)$ is the posterior probability for category II. In most cases, the posterior probabilities and the raw discriminant score are in agreement (i.e. a negative discriminant score corresponds to a higher group I posterior probability and a nonnegative discriminant score is consistent with a higher group II posterior probability). There are, however, exceptions. For the analysis in Table 24, there are two cases (numbered 18 and 26) which have nonnegative discriminant scores

but are classified in category II because the group II posterior probability is higher than that for group I. Any apparent inconsistencies between the raw discriminant score and the resulting classification can usually be resolved by examining the posterior probabilities. For SPSS analyses, the discriminant score simply serves as a starting point from which the posterior probabilities can be calculated. The comparison of the posterior probabilities determines the group classification.

Table 24 Predictions of tool condition using classification functions from LDA analysis with 10% cutoff

Experiment Number	$\frac{\omega\ell}{\nu}$	Discriminant Score	Predicted group	Actual group
1	10,048	-1.178	I	I
2	20,096	-0.153	I	I
3	30,144	0.873	II	I
4	7,177	-1.472	I	I
5	14,354	-0.739	I	I
6	21,531	-0.006	I	I
7	5,582	-1.634	I	I
8	11,164	-1.065	I	I
9	16,747	-0.495	I	I
10	15,072	-0.666	I	I
11	30,144	0.873	II	II
12	45,216	2.411	II	II
13	10,766	-1.105	I	I
14	21,531	-0.006	I	II
15	32,297	1.093	II	II
16	8,373	-1.349	I	I
17	16,747	-0.495	I	I
18	25,120	0.360	I	II
19	20,096	-0.153	I	I
20	40,192	1.898	II	I
21	60,288	3.950	II	II
22	14,354	-0.739	I	I
23	28,709	0.726	II	II
24	43,063	2.192	II	II
25	11,164	-1.065	I	I
26	22,329	0.075	I	II
27	33,493	1.215	II	II
28	12,560	-0.922	I	I
29	9,769	-1.207	I	I
30	19,538	-0.210	I	II
31	11,723	-1.007	I	I

*Cases for which the predicted and actual group membership differ are shaded. Blue/darker shading identifies cases for which the raw discriminant score indicates the tool should be classified in category II, but the posterior probability for group I based on the discriminant score exceeds that of group II (SPSS classifies the case in the group with the higher posterior probability).

The usefulness of a linear discriminant function lies in its ability to correctly classify data based on the input variable(s). The accuracy of the classifier is expressed as the percentage of correctly classified cases (i.e. the percentage of cases for which the predicted and actual group membership are in agreement). For the LDA using a 10 percent cutoff value to distinguish between "worn" and "unworn", the misclassification rate is 19.4% (6 cases out of the 31which appear in Table 12 were incorrectly classified using the discriminant function). The LDA for this cutoff value is thus accurate 80.6% of the time. The classification accuracy for each LDA function is plotted against its associated cutoff value in Figure 67. The cutoff values (which delineate the boundary between "worn" and "unworn" tool conditions) vary from 3 percent to 16 percent and each cutoff value corresponds to a unique LDA function. The highest misclassification rate (and the lowest classification accuracy) is observed for the LDAs associated with the 4 and 5 percent cutoff values. The highest classification accuracies are synonymous with the discriminant functions which represent the cutoff values at the high end of the spectrum (87.1 percent and 90.1 percent accuracy for 15 and 16 percent volume loss, respectively). Thus the functions which are most successful at accurate classifying this data are those which are the least sensitive to subtle changes in the volume of the tool. The classification algorithm based on $\frac{\omega \ell}{\nu}$ is best-suited for identifying conditions which will likely result in comparatively larger volume losses (unfortunately, depending on the rigor of the application, this state may be well beyond the point at which the tool is, for all practical purposes, considered to be worn). Of the misclassified cases, the majority occur when a worn tool belonging to group 2 is incorrectly classified as an unworn (group 1). From a manufacturing perspective, this is probably a more severe error than

classifying an unworn tool as worn when its true condition corresponds to the former. In an incorrect group 1 classification, the volume of the tool may have degraded to a point that is detrimental to the weld structure before the operator is alerted. While an incorrect group 2 classification does result in an unnecessary tool replacement, the quality of the weld (at least in terms of voids which may result from tool wear) is preserved.

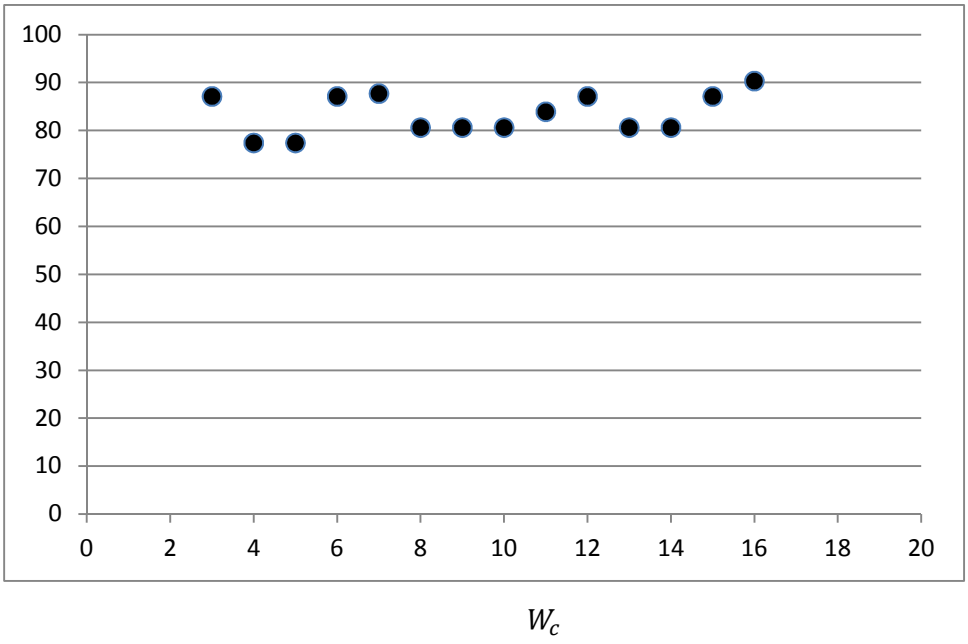


Figure 67 Plot of classification accuracy versus cutoff value of tool wear (W_c) used to classify the tool as "worn" (volume loss greater than or equal to W_c) or "unworn" (volume loss less than W_c)

Classification accuracy is only one of several metrics which can be used to evaluate discriminant functions derived from LDA. Others include Wilks' lambda, the F-statistic, and the *a priori* group probabilities:

1) Wilks' lambda, a multivariate extension of the F distribution, tests the statistical significance of the discriminant function. As discussed in chapter IV, significance can be interpreted as the likelihood that a particular outcome would be observed by chance [70]. If the significance value is sufficiently small, we reject the null hypothesis (that the discriminant function is not a reliable classifier) in favor of the alternative hypothesis (that the discriminant function can be used to predictively differentiate between groups based on one or several input variables). The value of the Wilks' lambda statistic corresponding to each discriminant function is tabulated in Table 25. All discriminant functions are significant at the 0.05 level and all but two (the discriminant functions for 3 and 4 percent) are significant at the 0.01 level.

2) For each discriminant function, the F-statistic is used to test for equality of group means. The group mean for each classification group is calculated by averaging the values of $\frac{\omega\ell}{\nu}$ for all the cases which correspond to volume losses consistent with that particular group. For example, the group mean for category I at the 3% cutoff value represents the average of the parameter $\frac{\omega\ell}{\nu}$ for all cases with less than 3 percent wear. Likewise, the group mean for category II is the average of $\frac{\omega\ell}{\nu}$ for all cases with wear equal to or exceeding 3 percent. To support the hypothesis that wear is proportional to $\frac{\omega\ell}{\nu}$ (and that $\frac{\omega\ell}{\nu}$ can be used to classify tool condition), the cases in category II (worn) should have a group mean exceeding that in category I (unworn) across all

discriminant functions. This is true – for each cutoff value, the mean value of $\frac{\omega\ell}{\nu}$ for the "worn" group is higher than that for the cases with comparatively smaller values of wear which place them in the "unworn" group. The significance values also reflect the inequality of the group means. All the F-values associated with the test for mean equivalency are significant at the .05 level and all but two (4.571 and 4.770, associated with the 3 and 4 percent cutoff values, respectively) are significant at the .01 level. These significance values indicate that there is sufficient evidence to reject the null hypothesis (that the group means are equivalent) in favor of the alternative hypothesis (that there is a significant difference between average $\frac{\omega\ell}{\nu}$ for groups of tools in the unworn and worn conditions, respectively).*

*As in chapters IV and V, the significance value represents the probability that we have rejected the null hypothesis in favor of the alternative hypothesis when the former is true (a type II error). The significance values associated with the F statistic are nearly identical to the significance values calculated based on the Wilks' lambda distribution in column 2. This equivalency phenomenon is attributable to the close mathematical relationship between the F statistic and Wilks' lambda: as mentioned in (1), the F-distribution is a subset of the more general multivariate Wilks' lambda distribution.

3) The final column in Table 25 lists the sum of the squares of the *a priori* probabilities for each discriminant function. The *a priori* probability is the proportion of the total cases which fall into a particular category. For instance, at the cutoff value of 3%, there are 3 (out of 31) cases in which the tool is classified as "unworn" and 28 cases for which the tool is "worn" (wear exceeds the cutoff value). The corresponding *a priori* probabilities are 9.7 percent and 90.3 percent, respectively. The value reported in column 7 is simply the square of these prior probabilities added together. This calculation is essentially a way to adjust/compensate for the disparities in group membership (scenarios in which the data is divided equally between groups are rare). The *a priori* sum of squares can be interpreted as the probability that the counts for the actual and predicted group membership are equal. For the discriminant function to be useful, the classification accuracy must be greater than the sum of the squares of the *a priori* probabilities. All of the discriminant functions satisfy this criterion. The functions associated with the mid-range cutoff values (which have lower classification accuracies than functions at the extreme ends of the spectrum) offer the greatest degree of improvement over the prior probabilities.

Table 25 Metrics used to assess statistical significance of discriminant functions

Cutoff Value	Wilks' Lambda	F statistic	Significance value	Classification accuracy	*A priori* sum of squares*
3%	0.864	4.571	0.041	87.1%	0.775
4%	0.859	4.770	0.037	77.4%	0.650
5%	0.654	15.356	0.000	77.4%	0.589
6%	0.495	29.632	0.000	87.1%	0.501
7%	0.454	34.838	0.000	87.1%	0.505
8%	0.561	22.649	0.000	80.6%	0.542
9%	0.561	22.648	0.000	80.6%	0.542
10%	0.561	22.649	0.000	80.6%	0.542
11%	0.502	28.728	0.000	83.9%	0.588
12%	0.523	26.397	0.000	87.1%	0.618
13%	0.735	10.473	0.003	80.6%	0.687
14%	0.735	10.473	0.003	80.6%	0.687
15%	0.778	8.257	0.008	87.1%	0.730
16%	0.768	8.749	0.006	90.3%	0.825

*same as proportional by chance accuracy

The take-home message of this analysis is that while all the discriminant functions in Table 13 satisfy significance criteria as well as the condition that classification accuracy exceed the *a priori* sum of squares, not all classifiers are created equal. The classifiers which offer an improvement of 25 percent or greater over the proportional by chance accuracy are highlighted in Table 25. A high classification accuracy alone is not an indication of the classifier's effectualness and can be misleading. There are several cases, particularly at the extremes of the spectrum, where classification accuracy is contraindicative – for instance, the highest classification accuracy (90.3 percent for the 16 percent cutoff) represents only a 9 percent improvement over the *a priori* sum of squares (the lowest degree of improvement observed among all the classifiers in Table 25). The second-lowest value of improvement (12 percent) occurs at the lowest cutoff value (which, perhaps surprisingly, also has the second highest value for classification accuracy

at 87.1 percent). Thus classification accuracy and statistical significance are not necessarily synonymous with the best classifiers. The best classifiers lie in the mid-range – the analyses where the volume loss which separates unworn and worn tools falls between 5 and 12 percent, inclusive. These classifiers show a substantial improvement (>25%) over the proportional by chance accuracy and satisfy the criteria for classification accuracy discussed previously.

An evaluation of the classifier based solely on classification accuracy and the propotional by chance accuracy is incomplete since (as with regression) the equation itself tells us little about how the classifier will generalize to the larger population. A "leave one out" analysis is used to assess the ability of the LDA classifier to correctly identify group membership for parameters which were not used in the construction of the model. Similar to data splitting in regression, "leave one out" is a cross-validation technique which uses a subset of the sample data (all but one case) to construct a classifier and then tests it on the remaining cases. This procedure is repeated until each case has been left out once: the cross-validation accuracy for the discriminant function represents the percentage of the sequentially omitted cases which were correctly classified. While we expect that the accuracy will shrink somewhat when applied to the general population, a robust discriminant function must have a cross-validation accuracy which is greater than the proportional by chance accuracy (a term synonymous with the *a priori* sum of squares calculated in Table 25). Table 26 compares the cross-validation accuracy with the classification accuracy and the proportional by chance accuracy for each of the LDA classification functions constructed. The analyses cross-validate well: in all cases the proportional by chance accuracy is less than both the classification

accuracy and cross-validation accuracy. For the LDA functions where the classification accuracy and the cross-validation accuracy are not equal, the function experiences only a slight (less than 5 percent) reduction in predictive capability. The difference between the classification accuracy and cross-validation accuracy (referred to as shrinkage) for each function is listed in the rightmost column of Table 26.

Table 26 Comparison of *a priori* sum of squares, classification accuracy, and cross-validation accuracy for discriminant functions

Cutoff Value	*A priori* sum of squares	Classification Accuracy	Cross-Validation Accuracy	Shrinkage
3%	0.775	87.1%	87.1%	0
4%	0.650	77.4%	77.4%	0
5%	0.589	77.4%	74.2%	0.2%
6%	0.501	87.1%	83.9%	3.4%
7%	0.505	87.1%	87.1%	0
8%	0.542	80.6%	80.6%	0
9%	0.542	80.6%	80.6%	0
10%	0.542	80.6%	80.6%	0
11%	0.588	83.9%	83.9%	0
12%	0.618	87.1%	87.1%	0
13%	0.687	80.6%	80.6%	0
14%	0.687	80.6%	80.6%	0
15%	0.730	87.1%	83.9%	3.4%
16%	0.825	90.3%	87.1%	3.2%

The LDA analyses presented in this chapter are not intended to be applied to FSW of MMCs in a manufacturing setting directly, since the data used to build and test the classifiers in Table 25 are specific to the joint configuration, tool and workpiece materials, apparatus limits, and tool geometry used in these experiments. However, the approach we have taken can be broadly applied in an industrial setting to predict wear in

FSW of MMCs for any combination of these variables. An operator can design a matrix of parameters (in this study we used a Taguchi L_{27} orthogonal array for three factors at three levels, but an alternative design may be more appropriate, depending on the number of process variables), the results of which are used to build a classifier which will discriminate between tool conditions (some ancillary research may be required to quantify the amount of wear that can occur without an unacceptable reduction in quality for the specific application). The significance values, classification accuracy, cross-validation accuracy, and the proportional by chance accuracy for the classification function(s) can be used as metrics to assess the goodness of a classifier. When a reliable classifier is found, it is applied to predictively classify tool condition based on the process variables. In application, a classifier works similar to the machinability maps included in machinist's handbooks for drilling, cutting, or turning of abrasive materials: operators can use the discriminant function to determine whether wear for a set of parameters will exceed the critical threshold identified as integral to quality or solve the inverse problem (determine parameters will maintain wear that is below some acceptable level). As with the regression model, information about the conditions which will maintain an unworn tool (and prolong tool life) are inherent in the equation derived from the LDA.

What separates our particular methodology from traditional LDA based on process variables is the use of the dimensionless parameter $\frac{\omega l}{v}$ as a basis for the discriminant function. Given the diversity of FSW applications and equipment, it is perhaps unreasonable to expect that the ranges of the process parameters for all manufacturing applications involving FSW of MMCs will be similar to those found in the VUWAL experimental test bed. The dimensionless number functions as a similitude

parameter (like the Reynolds number) that enables our (specific) results to be generalized and extended to predict wear for manufacturing scenarios with test conditions disparate from our own. The major advantage of the dimensionless parameter lies in its ability to simplify the relationship between wear and process variables. As indicated by the regression model, the variation is inverse for wear and traverse rate but direct for wear and both rotation rate and traverse velocity. The dimensionless parameter consolidates these dependencies into a single, unitless quantity, reducing the relationship between wear and process parameters to a linear equation of the form $W = kx$. Analysis of FSW of MMCs using this parameter thus hinges on determination of the proportionality constant k. The dimensionless parameter virtually eliminates the need for sophisticated experimental design techniques, since the problem now centers on gauging the effect of a single input variable (rather than 3 process variables considered independently) on a single output variable (percent wear). The dimensionless number thus has the potential to significantly reduce experimental time, as investigations which utilize it require fewer experiments to establish the relationship between wear and process parameters as well as to construct an LDA-based classifier (the representation of ω, ℓ, and ν as a single variable cuts the number of possible LDA functions to 1*) . Although we cannot expect similarities in parameter selection across applications/equipment, the correlation between wear and the dimensionless ratio $\frac{\omega l}{\nu}$ suggests that the system behavior (i.e. the variation of wear with the dimensionless parameter) should be similar.

*The number of maximum possible discriminant functions is equal to the number of independent variables. Thus independent consideration of $\omega, l,$ and ν would yield three unique discriminant functions. The best/most statistically significant of these would be used as a classifier.

Dimensional analysis is particularly useful for Friction Stir Welding (and FSW of MMCs) due to the multiple variables which can impact the process (depending on the data available to a researcher, any of the dimensionless parameters derived in table could serve as a basis for an LDA provided it was related to the outcome variable). Although dimensional analysis coupled with discriminant analysis demonstrates great potential for establishing standards/guidelines for wear in FSW of MMCs, the latter technique does have some limitations. For instance, LDA requires that the case to variable ratio be greater than 5 and preferably larger than 20 (for the experiments in Table 13, this ratio is 31). Additionally, the number of cases in the smaller classification group must be greater than the number of independent variables. Although there is only one true independent variable in our analysis (the parameter $\frac{\omega l}{v}$), the maximum and minimum cutoff values were chosen conservatively to ensure that there were a minimum of 3 cases in the smaller member group for each analysis. This restriction on sample size for the smaller group sets the minimum and maximum cutoff values for the LDA at 3 and 16 percent, respectively.

The analyses presented here may seem coarse; there are only two categories into which a tool can be classified and in some cases the group identifiers may not reflect the true condition of the tool (i.e. a tool may be classified as "unworn" but still exhibit some evidence of wear). To improve the sensitivity of the LDA analysis and more accurately capture subtle changes in tool volume due to wear, a multi-classifier approach may be adopted. Multi-class LDA is a generalization of binary LDA that can accommodate increased variability in the non-metric dependent variable. A multiclassifier works the same way as a binary classifier in that classification is based on the discriminant score

and/or a comparison of the posterior probabilities for each class – the difference is that there are now at least three categories into which a case can be sorted. Multi-class LDA is recommended for mission-critical applications where the operator needs to be alerted when the tool experiences a slight change in volume (instances where the ability to identify the tool as simply "unworn" or "worn" is insufficient and a knowledge of the intermediary stages of wear is desired). If an application calls for a more nuanced classification scheme, the methodology used to construct the binary classification function used in our analyses can be easily extended to include multiple groups.

A multi-class linear discriminant analysis was not applied to the data in this study for several reasons:

1) The multi-class form of LDA represents a departure from the conventional (i.e. binary) partitioning used in Fischer's technique. In the latter method, categorization can be easily determined based on the sign of the numerical output of the classifier and a comparison of the posterior group probabilities. Since the discriminant scores are standardized, a case's probable group membership can be ascertained from the sign (positive or negative) of the discriminant function's numerical output. If the posterior probability calculation is being done by hand, the probabilities associated with the binary LDA are less computationally intensive than those for the multi-class function (this makes sense since there are more possible groups to which a case can belong in the multi-class).
2) The demarcation between classes (what constitutes the low, medium, and high wear regime) is application-specific. While we can anticipate the cutoff values which may distinguish a worn tool from one that is considered unworn, it is more

difficult to forsee the divisions which will separate multiple classes (or even the number of classes) for a specific application.

3) In many cases, the binary classifier is more statistically significant than the classifier derived from multi-class LDA.
4) Multi-class LDA may require generation of additional data to better represent the range of the output variable (specifically the points between the extrema). If a large number of classes exist, the data set must be proportionately large in order to satisfy the LDA criterion (which also applies to multi-class systems) that the smallest group contains a number of cases greater than the number of input variables.

Perhaps the most useful result which emerges from the dimensional analysis and LDA is the representation of wear as a function of the unitless ratio $\frac{\omega l}{v}$. The functional relationship between wear and wl/v implies something similar to the regression model The practical significance of this dependency is that it may provide fundamental insight into the process dynamics of wear in FSW. While the causes of tool wear are well understood in machining of harder materials (and specifically MMCs), there understanding of wear in FSW of MMCs is basic The relationships indicated by the dimensional analysis (and previously by the regression model) have clear implications for modeling of the wear process in FSW of MMCs. The inverse relationship between wear and traverse rate suggests that wear for this process is a shear, rather than a drag, phenomenon. Furthermore, isolation of the $\frac{\omega}{v}$ term suggests that the amount of wear may be related to the width of the shear zone as formulated in the rotating plug model. Developed by Dr. Art Nunes, this model postulates that the probe is surrounded by a

region of plasticized material in the x-y plane (a "plug") which rotates with the tool as it advances along the jointline [101]. Extending the rotating plug model to explain wear phenomena in FSW of MMCs is the focus of the next section. The model also suggests that other, previously ignored factors (namely, material properties of the workpiece such as particle size and temperature) may be important contributors to the wear mechanism.*

5.2.6 Ranking Effects using Multiple Regression Modeling and Significance Tests

Section 5.2.1 ranked the effects of each process variable on wear based the calculated signal to noise ratio. Multiple regression analysis can also be used for this purpose. In addition to the model developed in section 5.2.2 (which includes all three predictors), there are six other models which can be constructed from the experimental data. These seven models, compared in Table 27, correspond to every possible combination of the three process variables (ω, ℓ, and ν) included in the Taguchi study. There are four possible multivariate linear regression models (with $\omega/\ell/\nu$, ω/ν, ω/ℓ, and ℓ/ν as predictors, respectively) and three univariate regression models (each of which uses ω, ℓ, or ν as a single predictor). Table 15 summarizes the relevant statistical metrics for the multiple regression model associated with each variable or group of variables. The chart indicates the variables included (and excluded from) each model as well as the corresponding significance value (p-value) and squared correlation coefficient. While the significance values have the same meaning as in previous discussions, the R^2 value in this instance takes on a slightly different interpretation.

*There is some evidence in the literature that the FSW process can be robustly modeled without knowledge of the specific material characteristics of the tool and/or workpiece.

Here, the R^2 value is taken to represent the percentage of the variation in the outcome variable (wear) which can be accounted for by the variables included the model (w, l, or v or some combination thereof). This interpretation of the correlation coefficient can give us some idea of the percentage contribution of each factor to the variation observed in the outcome variable.

Table 27 Possible regression models

Model ID	Predictors included	Predictors excluded	R	R^2	Adjusted R^2	F change	Sig. F change
1	ω	ν, ℓ	0.585	0.342	0.316	13.006	0.001
2	ν	ω, ℓ	0.279	0.078	0.041	2.102	0.160
3	ℓ	ω ,ν	0.627	0.393	0.369	16.204	0.000
4	ω, ν	ℓ	0.648	0.420	0.371	8.682	0.001
5	ω,ℓ	ν	0.858	0.735	0.713	33.862	0.000
6	ℓ, ν	ω	0.686	0.471	0.427	10.677	0.000
7	ω,ℓ, ν	none	0.902	0.813	0.789	33.339	0.000

Not surprisingly, the best-fitting model is model 7, which includes all three predictors (this is the model that was constructed and tested in section 5.2.3). While the other models (which include only one or at most, two of the variables in the Taguchi study) have a diminished predictive capability compared to the three-predictor model, they provide valuable information about the contribution of each factor to the outcome variable. The $\boldsymbol{R^2}$ value is synonymous with the percentage of the observed variation in the outcome which can be attributed to changes in the included predictors. For example, model 1, which considers only the effect of rotation speed, has an $\boldsymbol{R^2}$ value of 0.342 – this means that 34.2 percent of the variation in tool wear can be attributed to the selection

of ω. In this manner, the squared correlation coefficient of the three single predictor models can be used to rank the effects of each variable on wear. The most significant parameter in determining the amount of wear is thus length of weld (with a 39.3 percent contribution), followed by rotation speed (34.2 percent). Traverse rate comes in a distant third, accounting for only 7.8 percent of the variability in the wear data. The magnitude of effects indicated by this analysis are consistent with the rankings compiled using signal to noise ratio in section 5.2.1. However, the R^2 metric does more than simply confirm previous results: it offers a means to quantify the contribution of each variable to the system response (which is not as readily apparent with SNR ratios).

All models are highly significant ($p < 0.005$) with the exception of model 2, which excludes all variables except traverse rate ν. The comparatively smaller contribution of traverse rate does not mean that it is insignificant (and that the effects of this variable should be discarded), but rather that ν is less significant than the other parameters considered in the study. The significance values associated with the three predictor model (0.000) and the traverse rate coefficient (0.005) confirm that the selection of traverse rate has a significant impact on the amount of wear. A comparison of the unstandardized coefficients for from the multiple regression model (Table 5) indicate that a unit change (an increase or decrease of 1 inch/minute) in traverse rate results in a relatively larger change in the volume loss (1.038 percent) than a unit change in rotation speed (a change of 1 RPM corresponds to a 0.009 percent change in wear) or length of weld (0.584% per unit change in length).

Although all three of the included parameters are significant contributors to wear, they do not account for all of the variability in the wear data. The three-predictor model

has an R^2 value of 0.813, which means that 19.7 percent of the variation in wear cannot be explained by the process parameters (ω,ℓ, ν) included in the experimental study. Despite this substantial proportion of unexplained variation, the three-predictor regression model exhibits good predictive capability (as demonstrated by the results of the cross-validation study in section 5.2.3). However, the distance of the R^2 value from 1 suggests that there are other factors not included in this study which may impact the wear process. The influence of these additional factors, such as percentage reinforcement, particle size, and tool material, will be thoroughly explored in subsequent chapters.

5.2.7 Verifying Assumptions

Before we proceed with an investigation of the wear process based on the information derived from this exploratory study, it is important to verify that the formal assumptions which underlie the mathematics of regression modeling are satisfied. These assumptions, taken from the list compiled by Field et al., are enumerated in the list below [70]:

1) Predictor and outcome variable(s) are quantitative. Value of outcome variable is unconstrained.
2) Variance of predictors is non-zero.
3) Predictor variables are uncorrelated ("no perfect multicollinearity")
4) Predictors and "external variables" (those not included in study) are uncorrelated
5) Residual terms are uncorrelated (Durbin-Watson statistic between 1 and 3)

6) Errors are normally distributed

7) Independence of outcome variable

8) Linearity

Assumptions 1) and 2) are clearly satisfied by the experiments which comprise the Taguchi study. The predictors and the outcome variable are numerical (rather than categorical) and the outcome variable (percent wear) is free to assume any value. Assumption 3), that the predictor values be uncorrelated, is also met – the levels of each of the predictors were chosen independently and the choice of any value for a particular predictor has no influence on the value of the other two predictors. Assumption 4) asks us to evaluate whether there are other predictor variables not included in this analysis which could exert some influence on the predictors included in the analysis (i.e. are there any excluded predictors which can influence included predictors?) As discussed in section 5.2.4, the R^2 value of the three-predictor model suggests that there may be other variables which influence the amount of wear (tool material, reinforcement particle diameter, percentage reinforcement, hardness of reinforcement), but the degree of their influence is thus far unsubstantiated. These variables were not included in the initial study in an effort to keep the number of experiments at a manageable level. However, the selection of these factors is clearly independent of $\omega, \ell,$ and ν (for instance, the choice of particle size has no influence on rotation rate, etc.). Thus criterion 4 is also satisfied.

Testing assumption 5) requires calculation of the correlation coefficient ρ between each of the residual terms (the residual term is the difference between the amount of wear predicted by the model for a particular case and the observed wear). A faster, less computationally intensive method to gauge the strength of the overall

relationship between the residual terms is to use the Watson-Durbin statistic, a metric which is calculated by SPSS and output as part of the basic regression analysis. The Watson-Durbin statistic can range from 0 to 4: a value between 1 and 3 indicates that correlation among the residuals is minimal. For the multiple regression model, the Watson-Durbin statistic is 1.886 (a Watson-Durbin statistic of 2.0 corresponds to a correlation of 0).

Assumption 6) states that the distribution of residuals/errors should approximate a normal distribution. The normality of the residual data can be visually assessed by constructing a histogram with the signed distance of the error from 0 on the x-axis and the corresponding frequency with which it occurs plotted on the y-axis. The data is divided into several equally spaced classes which span the range of the residuals. If the distribution is Gaussian, the majority of the residuals will lie close to zero -- the categories at the lower and higher end of the spectrum should occur with significantly less frequency. The error distribution for our model (shown in Figure 68) exhibits characteristics consistent with a normal distribution. The plot is centered at zero and the data is divided into several classes, referred to as "bins." Of the 27 residual cases, 19 of them (70.4 percent) lie within 1 standard deviation (+/- .034) of the mean residual (μ=-0.0076). For a true normal distribution, 68 percent of the data should fall within one standard deviation of the mean. 26 of the 28 cases plotted in Figure 21 (93 percent) are within two standard deviations of the mean (between -0.0756 to 0.0604); for a perfect Gaussian distribution, this proportion would be 95 percent. Continuing with this comparison, the range bounded by 3 standard deviations on either side of the mean should encompass 99.7 percent of the data; all of the residuals in our data set fall between

0.0944 ($\mu + 3\sigma$) and -0.1096 ($\mu - 3\sigma$). Thus the distribution is roughly normal and the criterion specified in 6) is satisfied.

An alternate graphical representation which can be used to determine the shape of the residual distribution is to construct a P-P plot, which compares the expected cumulative probability versus the observed cumulative probability for two distributions. For a residual distribution that is truly Gaussian, the relationship between the expected and observed cumulative probabilities is linear and 1 to 1. The P-P plot of normally distributed residuals appears in Figure 69 alongside the expected vs. observed cumulative probabilities compiled from our test data (represented by circles). This plot of the two distributions on the same axes, included in the standard SPSS output for regression analyses, presents another method (other than the histogram in Figure 68) to visually assess the normality of the residuals. Comparing the two distributions, it is apparent our test data closely follows the P-P plot for a normal distribution. While there is some slight deviation from the normal line (which is probably a consequence of the model's slight tendency to overpredict wear), the overall linear trend in the cumulative probabilities for the residual data confirms the previous conclusion that this data is approximately normal. Normality is desirable in residuals because it indicates that large discrepancies between observed values and those predicted by the model are rare. For normally distributed residuals, the difference in actual and predicted values is most frequently zero.

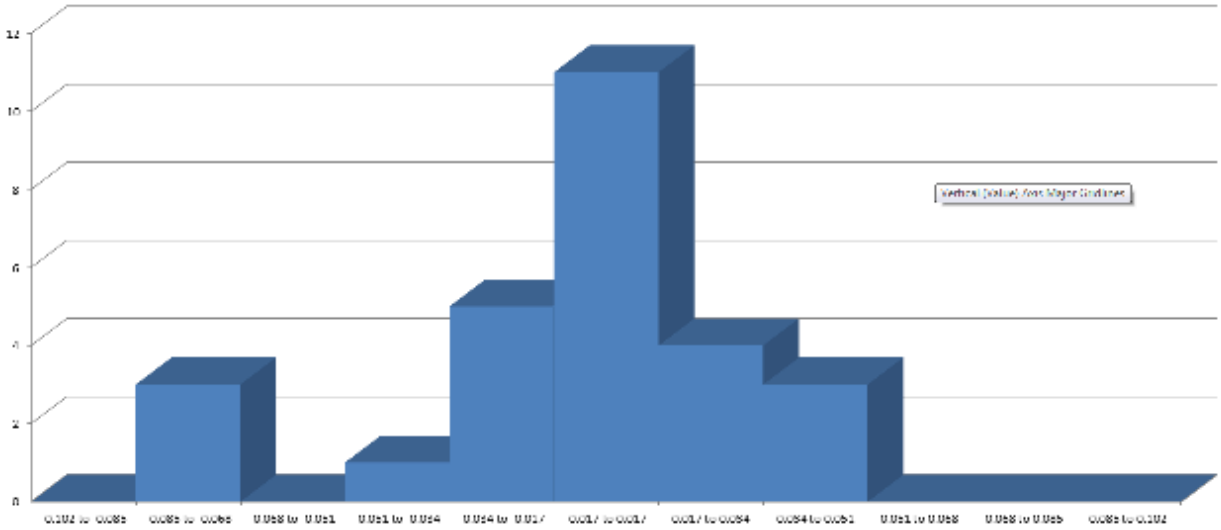


Figure 68 Histogram of frequency vs. range of residuals. X-axis range of residual plot is defined by three standard deviations on either side of the mean.

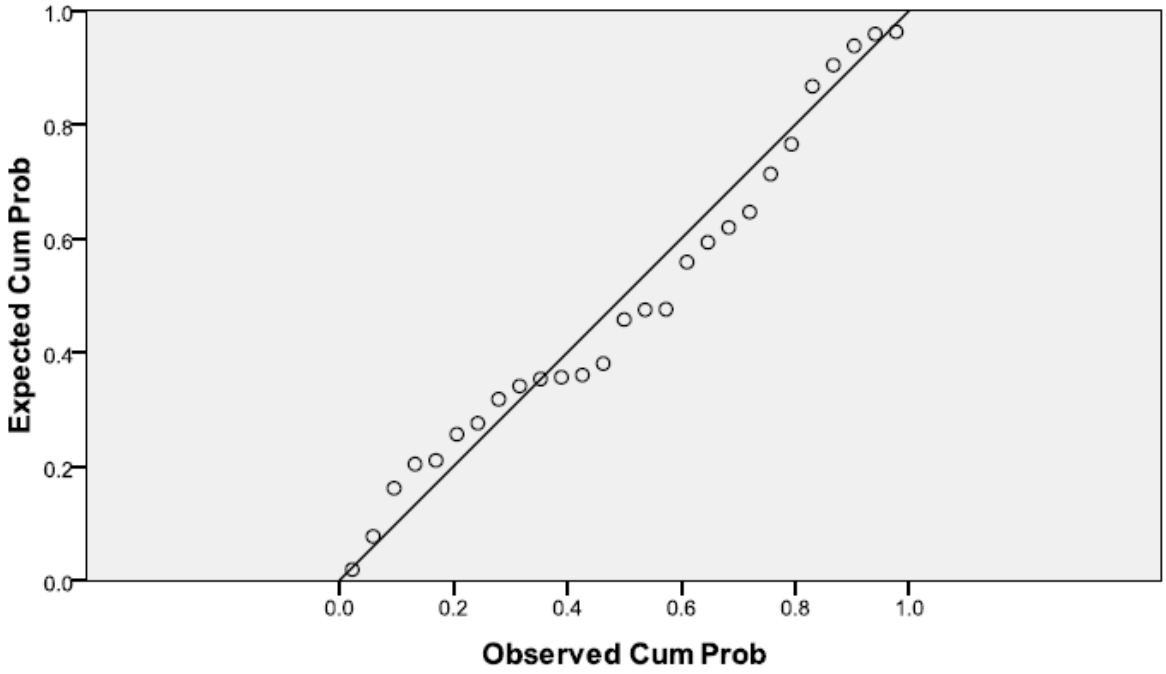


Figure 69 P-P plot of expected cumulative probability versus observed cumulative probability for a normal distribution (straight line) and the residual distribution constructed from the experimental data (represented by circles).

Assumption 7) is that the outcome variable (wear) is independent. Each of the wear data in the sample must arise from a unique set of factor values. This is similar to the 1 to 1 correspondence criterion for an algebraic function, which stipulates that each value of x in the function domain corresponds to one (and only one) value of y. Although this may not be the case for the entire population of wear data (refer to section 5.2.4 for a discussion of the possibility of obtaining the same wear value for two different parameter sets, provided they correspond to the same value of the dimensionless number $\frac{\omega \ell}{\nu}$), it is true for the sample used to construct the regression model.

Assumption 8) requires that the relationship between the predictors and outcome variable be linear. In terms of logic, this may seem like a classic instance of "begging the question" (in which the proposition to be proved is assumed at the outset). The construction of a linear regression model, however, should be viewed as an exploratory test of the dependency between the predictors and the outcome variable rather than a formal proof of linearity. There are two ways to assess the validity of assumption 8). One is the R^2 metric: as a general rule of thumb, the data can be considered linear if R^2 is greater than 0.70. If R^2 is less than this value, the data are either uncorrelated or the relationship is of a higher order (in which case it is best fit by a quadratic, cubic, or higher degree polynomial function). The data is also linear if the mean values for each level of the predictor follow a straight line. This was verified as part of the SNR analysis performed in section 5.2.1 to rank the magnitude of factor effects and check for interactions. Since R^2 for our model is 0.81 and the factor means lie along a (sometimes roughly) straight line, we can conclude that the relationship between ω, ν, and ℓ is in fact linear. While validation of assumption 8) may seem unnecessary, it is important to

establish that the model we have chosen actually fits the data. A linear model that is constructed from a nonlinear data set obviously calls into question the legitimacy of any claims or conclusions made based upon it.

The nature of the relationship between each factor and the outcome (either direct or inverse), the regression model, and the dimensionless parameter derived from it form the foundation of the analyses which follow in chapters VI-VIII. While it is an often overlooked aspect of regression modeling, verification that the set of assumptions upon which the model is built are satisfied is critical to the results presented in subsequent chapters. Validation of the premises in 1) through 8) lends credibility to the conclusions drawn based on the model and expands its generalizability. However, even if all of these assumptions are verified, these criteria do not guarantee that the model will reflect trends in the general population as well as those in the sample. Cross-validation techniques and specialized statistical metrics (section 5.2.3) are needed to test the model's ability to predict the outcome variable for inputs which were not part of the sample used to construct it.

5.3 Conclusions and Additional Thoughts

5.3.1 Research Summary

To determine the dependence of wear on process variables, we used design of experiments in conjunction with regression modeling. A Taguchi design for three factors at three levels presented a directed experimental approach which minimized the number of experiments needed to ascertain the relationship between process variables in FSW of MMCs and tool wear. The symmetry of the data in orthogonal Taguchi designs makes them ideally suited for regression analysis: for the L_{27} array, the levels each factor can assume are restricted to three equidistant values which are equally represented in the design matrix. Both Taguchi design and regression modeling are tenable techniques for the study of wear phenomena, as evidenced by their prevalence in prior studies of wear (see chapter IV) . The relationship between the process variables (rotation rate, traverse rate, length of weld) and wear data in the Taguchi array is strongly linear. The regression model demonstrates excellent predictive capability when applied to data in the sample set. A cross-validation technique (in which the model is tested on several parameters separate from those used to construct it) was used to address concerns about scalability. While the model cross-validates well, there are still concerns about how the model will generalize to FSW of MMCs for manufacturing scenarios where the apparatus limits or possible range of process variable differ from those used in our study. The dimensionless number $\frac{\omega \ell}{v}$ is proposed as a scaling parameter to accommodate for variations in the range of values the factors can assume. The strong correlation of this number with wear simplifies the modeling of the relationship between wear and process parameters, as the

system response can now be estimated based on the value of this single, unitless number (rather than the weighted sum of the three variables which comprise it).

5.3.2 Comparison with results from machining

In chapter IV, the results of several studies of tool wear in machining MMCs were evaluated. The relationships between wear and process variables observed in cutting, milling, and drilling of MMCs are in many ways similar to the dependencies elucidated by our study of wear in FSW. In cutting of Al-MMCs, Davim and Baptista found that wear is directly proportion to cutting speed (analogous to rotation rate in FSW) and cutting time (a factor synonymous with length of cut), but inversely proportional to feed rate [64]. This result mirrors the relationships documented for wear on steel tools used in FSW of metal composites. In drilling of MMCs, wear is directly proportional to the rotation speed of the drill, the feed rate, and the number of holes drilled In both machining processes, the exact relationship between wear and feed rate is a source of some academic debate. There are researchers (references 75 and 76) who contest Davim and Baptista's claims that cutting tool wear varies inversely with feed rate and present experimental data to substantiate that the opposite case (in which wear of the cutting tool increases with increasing feed rate) is true. In chapter IV, we attempted to explain this discrepancy using the classic Taylor equation for tool life and the differences in the tool/workpiece materials used in the investigations cited. While we observed an inverse dependency between tool life and traverse rate in FSW of MMCs, the relationship between ν and wear was the weakest of the three factors included in the study. However,

the plots of Prado, et al. display similar trends, showing that for most parameters, wear decreases with increasing traverse rate [42-43]. The relationship between feed rate and wear is particularly important since it can provide insight into the wear mechanism. For instance, an inverse relationship between feed rate and wear indicates that the wear mechanism is driven by shearing, whereas a direct relationship between these variables suggests a drag phenomenon. The significance of the relationship between wear and ν is expanded upon in the next chapter, which develops a rotating plug model to explain wear in FSW of MMCs.

In machining, much concern is given to maintaining tool wear below some critical value: if the surface roughness of the finished part exceeds this value, then it no longer satisfies production criteria. Similarly, in FSW joints which must be qualified for usable hardware, wear must be constrained to levels which do not negatively impact the integrity of the finished joint (as discussed previously, the deterioration of the tool pin shape is associated with reduced vertical flow and subsequent void formation). Since the question of when wear begins to contribute to a reduction in joint strength is highly application specific, its discussion in this dissertation is relegated to the section on dimensionless analysis (which demonstrates how the dimensionless number can be used in conjunction with a classifier to determine which parameters will keep wear below some critical, experimentally determined value of volume loss). Additionally, there are many published parameterization studies on FSW of MMCs (summarized in chapter II) which determine an operating window (based on joint efficiency calculated from tensile tests) for a particular combination of tool geometry, tool material, workpiece material, joint configuration, and process variables. Unfortunately, these studies often neglect to

address the issue of wear and its potential impact on joint efficiencies. Bringing FSW of MMCs to a mature, technology ready level will require integration of these two pieces of research (tool wear vs. process parameters and process parameters vs. joint efficiency).

In machining of harder materials, operators can refer to a machinability map, a plot which estimates wear for common combinations of tool and workpiece materials in machining processes, to quickly assess the amount of wear which can be expected when running at a particular set of process parameters. The goal of the machinist is to find the parameters which both maximize tool life and maintain an acceptable surface texture for the finished part. This is an iterative process, calling upon the operator to perform the process first at parameters which correspond to the absolute minimum for wear and subsequently test the surface finish (either by visual examination or profilometry) for comparison with application-specific criteria. If the surface is too rough, the operator proceeds to choose parameters which correspond to a slightly greater level of wear, repeating this process until the surface texture requirement is satisfied. In instances where the surface texture criteria are never met, a change in the tool or workpiece material is recommended. We propose that a similar protocol be adopted for FSW of MMCs (the only difference being that for FSW, joint efficiency replaces surface roughness as the quantitative metric which determines acceptability). The dimensionless number and the three-dimensional plot in Figure 15 serve as a general guide to tool wear for FSW operators, providing rough estimates of expected volume loss for joining MMCs using conventional steel tools. In a production setting, initial friction stir welds of MMCs should be performed at the parameters which minimize wear (yet lie within apparatus limits). If tensile testing reveals the joint efficiency to be unsatisfactory, $\frac{\omega}{v}$ can be

gradually increased until an acceptable level of strength is achieved. It is anticipated that parameters in the low-wear regime will correspond to stronger welds, since tool wear in FSW has been linked to formation of wormholes (and the presence of void defects typically coincides with a significant reduction in ultimate tensile strength) [51]..

There are several machining studies profiled in chapter IV which consider the impact of material properties, rather than process variables, on tool wear. Obviously, wear is a dynamic process which does not occur in a vacuum. While we have chosen to isolate the effect of ω, ℓ, and ν for our initial investigation, there are a myriad of other factors which can potentially impact the wear process. Of these, the most frequently studied are the material properties of the tool and the workpiece. The amount of wear in machining MMCs can vary dramatically depending on the reinforcement material, the reinforcement inclusion percentage, the dimensions of the reinforcement material, the tooling material, and the heat treatment procedure for the tool. In the case of coated tools, wear is further influenced by the coating grain size and the degree of adhesion between the base tool material and the coating (which is determined by the compatibility of the substrate and coating material as well as the method for coating deposition). The influence of material properties on tool wear in FSW of MMCs is investigated in chapters VI and VII; the results from this research are intended to guide selection of materials which will inhibit or altogether eliminate the wear process in FSW of MMCs. The exclusion of the factors associated with material properties does not affect the validity of the regression model in equation 5-2. As discussed in the previous section, it is acceptable to omit factors (even if they are hypothesized to have a substantial effect on the system response) as long as the factors which are excluded are independent of those

which are included. The model presented in this chapter clearly satisfies this criterion: the choice of particle size, reinforcement material, percentage reinforcement, and tool material are independent of (i.e. do not restrict) the selection of process parameters (and vice versa). Material property considerations were initially neglected to 1) isolate the variation of process parameters and focus solely on determining their individual and collective impact on the amount of wear and 2) reduce the number of experiments. With regard to point (2), considering a small group of variables (≤ 3) presents a less experimentally-intensive alternative to attempting to account for all possible factors which may influence the outcome in a single study. Regardless of the DOE technique used, simultaneous variation of all of the eight factors hypothesized to influence wear in MMCs yields a very large array of experiments which would be difficult to execute under reasonable time and supply constraints. The precedent for separation of the effects of process variables and material properties on wear can also be found in the machining literature, where it is often the case that these two classes of factors are considered independently. This partitioning is necessary to isolate effects, thereby maximizing the efficiency of data acquisition.

Machining studies also discuss the potential existence of **latent variables** which may influence the wear process. The term latent refers to a variable which is not explicitly included in the study but may be influenced by the factors which are. For example, suppose we are trying to develop a model which will predict a brain surgeon's skill (percentage of surgeries performed successfully) based on quantifiable metrics. These metrics may include experience/knowledge (number of years in practice, difficulty of medical school and residency program) along with innate, measurable attributes which

gauge the ability of the surgeon to quickly recall information, assess situations, and make decisions (ex. concentration, timed response to stimuli). While all these factors probably have some influence on a surgeon's job performance, there are other factors which may play a role. Psychological factors such as happiness and quality of life may impact job performance directly or they may have some influence on the factors (such as concentration) that correlate with performance. These psychological factors illustrate a common characteristic of latent variables: they are not directly measured, but inferred from observed variables.

Time and temperature, the two most common latent variables discussed in machining literature, are also present in our study of wear in FSW. Although the duration of our process is not logged, the weld time (in minutes) can be calculated by dividing the length of weld (in inches) by the traverse rate (in inches per minute). The plot of wear versus weld time in Figure 70 suggests that wear is strongly correlated with time for the 1000 RPM and 1500 RPM data sets (R-squared values are 0.7158 and 0.7749, respectively), but the relationship is significantly weaker for the 2000 RPM cases ($R^2 = 0.43$). The trend in the 2000 RPM is difficult to characterize: on the whole, the wear increases with elapsed time, but there are a few points which deviate from this trend. It is possible that the increase in heat input associated with the 2000 RPM data inhibits the wear process, resulting in some instances where wear decreases as rotation speed or time is increased (a complete discussion of the effect of temperature on the width of the shear zone can be found in chapter VI). The link between wear and time does not provide any new information about the process, as the time dependency was implicitly included in the dimensional analysis through the $\frac{\ell}{v}$ term. Since $\frac{\ell}{v} = t$, the

dimensionless number can also be written as the product ωt. However, it is preferable to use the form $\frac{\omega \ell}{v}$ since the values of ω, ℓ, and v are explicitly chosen by the weld operator.

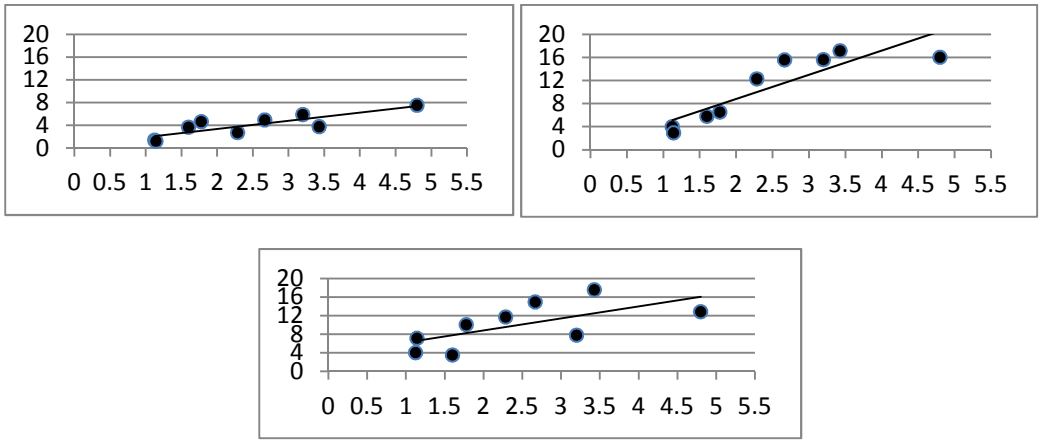


Figure 70 Plots of percent wear versus time tool is in contact with material for three rotation speeds (clockwise from the upper left): 1000 RPM, 1500 RPM, and 2000 RPM.

The calculation of temperature from the observed factors is a bit trickier. In machining (as in FSW), the temperatures experienced by the workpiece and the tooling are influenced by process parameters. It is generally true that welds with faster spindle speeds and slower traverse rates tend to be "hotter" than those with comparatively lower rotation speeds and faster traversal. There are numerical models (such as the equation developed by Schmidt, et al.) which attempt to estimate the average steady-state temperature for a friction stir weld based on tool dimensions and process parameters [100]. However, numerical temperature estimation is complicated by the fact that temperature varies across regions of the weld and the tool. Estimation of the temperature at any point in the three-dimensional flowfield is its own (incredibly complex) area of research, utilizing both CFD and finite element methods. While process parameters

unquestionably do exert some influence on temperature, there remains some question as to whether existing numerical models are able to accurately capture this relationship. It is generally accepted (based on the results of thermal imaging and data logged from thermocouples) that temperature in FSW is directly proportional to weld pitch (defined as the ratio of rotation speed to traverse rate). Thus Figure 70, which plots wear against $\frac{\omega}{v}$ for three weld lengths, is a way to broadly assess the effect of temperature on wear without making any direct temperature measurements. The heat input of the tool escalates as $\frac{\omega}{v}$ increases, creating a marked increase in steady-state temperatures. However, the plots in Figure 71 show that the correlation between wear and weld pitch is very weak, a likely indicator that wear is not strongly temperature-dependent. For the series of welds that are 8 inches in length, the correlation coefficient R^2 between weld pitch and wear is 0.3436. For longer welds, the coefficient decreases to 0.3128 (ℓ=16 inches) and again to 0.2981 for ℓ=24 inches. Temperature independence should not be a cause for alarm, since it is characteristic of the abrasive wear mechanism hypothesized to be responsible for wear in FSW of MMCs. However, the fact that there appears to be no discernable correlation between wear and temperature does not entirely rule out the possibility that the thermal environment exerts some (albeit small) influence on the wear process. In chapter VI, the idea that a significant amount of additional heat can widen the region of plasticized material which surrounds the probe, thereby inhibiting tool wear, is evaluated.

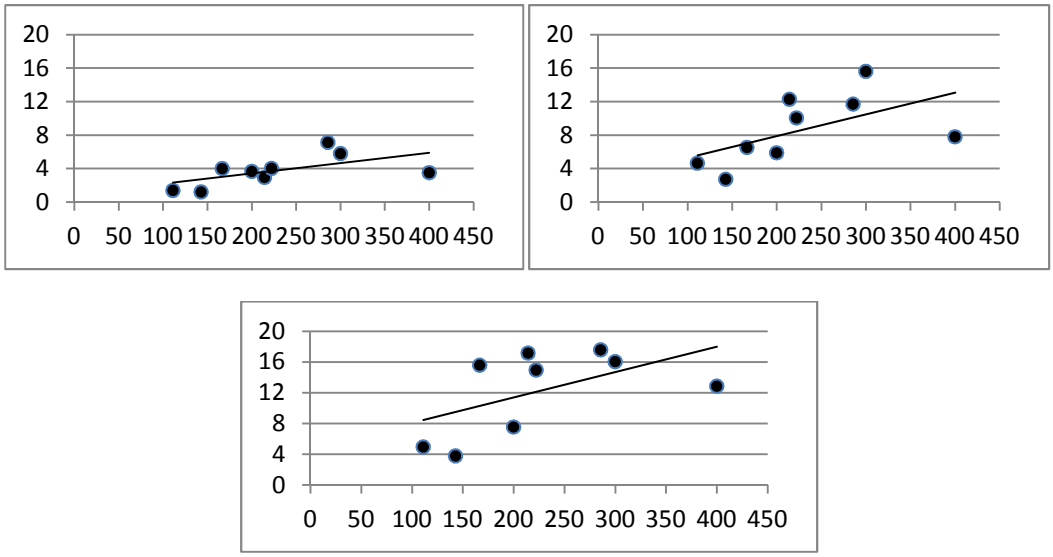


Figure 71 Plot of percent wear versus weld pitch (an indirect indicator for the amount of heat input to the weld by the tool) for weld lengths of (clockwise) 8 inches, 16 inches, and 24 inches.

Rather than plotting wear as a function of process parameters, a dimensionless number, time, or weld pitch, it may be useful to represent wear as a function of the number of rotations made by the tool during the course of a particular weld. The number of revolutions is synonymous with the dimensionless number $\frac{\omega \ell}{v}$ prior to its multiplication by the conversion factor $\frac{2\pi\ radians}{rotation}$. The plot of wear vs. number of rotations appears in Figure 72 (not surprisingly, wear generally increases with the number of revolutions). The slope of the line of best fit for this data (0.0022% per rotation) represents the average volume lost by the tool in a single revolution. For a ¼" cylindrical probe which extrudes 0.235" from the tool shoulder, this corresponds to a volume loss of 0.0115 cubic inches per revolution. This number illustrates the cyclical and propogatory nature of most wear processes: while a single wear event may correspond to infinitesimal material loss, these small losses can quickly add up to a substantial amount of degradation when the process

is permitted to proceed unchecked. Examining the data from this perspective provides a snapshot of the volume loss the tool experiences in an instant (the time it takes to complete a single rotation ranges from 0.06 seconds for the 1000 RPM cases to 0.03 seconds for 2000 RPM), an interpretation which can inform modeling of the wear process.

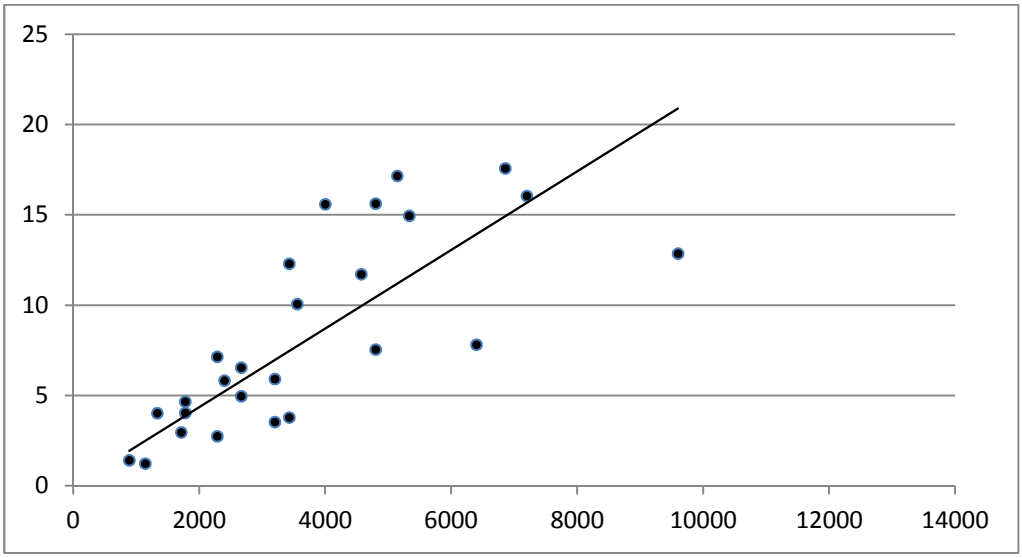


Figure 72 Plot of percent wear versus number of tool rotations

The literature concerning wear in machining MMCs serves as both a precedent and guide for our own analogous work on wear in FSW. We will continue to draw upon lessons learned from machining MMCs (from which process variables are the most influential, to strategies for combatting and sensing wear) as we proceed with our own investigations. Examining these studies (and associated methodologies) can provide valuable insight into wear in FSW, potentially reducing experimental time. We will return to machining literature in the chapters which follow, comparing the trends

documented for wear in FSW of MMCs with those previously observed in machining these materials where appropriate.

5.3.3 Final Thoughts/Moving Forward

The omission of parameterization (a topic which seems to dominate existing research on both machining and FSW) from this work should be addressed before moving forward. We have chosen to focus on characterizing the effect of various factors on tool life and modeling the physical phenomena which underlie the wear process. The parameterization studies which appear frequently in literature on machining are not incorporated in our analogous research into FSW of MMCs for two reasons:

1) There are many existing studies in the FSW literature which define the process envelope for MMC joints produced using FSW. The basis for these parameterization studies is a set of experiments with parameters which span the range of the apparatus limits. Tensile testing of the welded samples identifies those which satisfy design criteria – these parameters comprise the process envelope for one particular manufacturing scenario. This straightforward parameterization procedure can easily be implemented to optimize joint strength for any combination of tool geometry, workpiece material, and FSW equipment. The goal of our research is to contribute to a better overall understanding of the wear process in FSW of MMCs – we are concerned more with fundamentals, whereas parameterization research (for the most part) focuses on specifics. This is not to say that investigations of the latter type are not worthwhile (on the contrary, weld qualification is critical to implementation of FSW joints in mission-critical

structures), but the limited scope of parameterization is less aligned with our research goals.

2) There are significant scalability issues associated with parameterization. Even if we performed a parameterization study to determine which parameters result in joints with superior tensile strength for FSW in Al 359/SiC/20p, it is unlikely that our findings (which are specific to our welding apparatus, tool design, joint configuration, material dimensions, etc.) could be easily translated to another, even slightly different experimental setup. NASA Marshall Space Flight Center's manufacturing group, which uses FSW to join components for space shuttle and the agency's next generation launch vehicles, has frequently encountered issues with scalability when transitioning from the test bed to full-scale flight hardware. The process envelope for an application is defined using welds typically made using a smaller scale ISTIR Process Development System (PDS) welding robot. When these same parameters (which produce defect-free, strong joints on the PDS) are used to join the full-scale components (for example, a longitudinal weld along two barrel sections which form a fuel tank), the result is sometimes a defective joint. The unexpected emergence and growth of voids in the full-scale joint is probably attributable to differences in thermal environments between the PDS and the full-scale weld (either a lack of heat input or excessive heating can contribute to void formation). There has been some work devoted to development of a thermal scaling factor which can account for the changes in heat which accompany changes in the weld thermal environment [102]. However, until thermal scaling is better understood, there is little our research can offer in the way of parameterization which would assist in joining MMCs in large-scale structures,

Ultimately, the immediate relevance of parameterization to our work in tool wear is in question. While FSW tool wear almost certainly contributes to a degradation in joint strength, void formation is additionally influenced by process parameters, heat input, and tool geometry. Our research goal is to contribute to the (admittedly scarce) existing knowledge of the wear process in FSW of MMCs. This entails the development of a fundamental understanding of the wear process, an investigative undertaking which necessarily requires isolation of the wear phenomena. Parameterization, however, is only neglected temporarily – the reader will find that our results lend themselves to understanding process optimization in FSW of MMCs in unexpected ways. For instance, the research presented in this chapter pinpoint the parameters at which wear is minimized (the low-wear regime is bounded by comparatively smaller values of $\frac{\omega \ell}{v}$). Scenarios in which wear is mitigated (either by selecting process parameters which lie on the low end of the wear spectrum or using harder tool materials) should thus coincide with fewer weld defects. Less wear means that tool features (such as threads or flutes) which promote vertical stirring and guard against wormholes can be preserved. If voids persist at low-wear parameters, their formation can be attributed to factors other than erosion of features/an unfavorable change in tool shape.

There are many existing studies devoted to parameterization of FSW for alloys ranging from ordinary (Aluminum, steels) to exotic (MMCs). These studies sometimes neglect to posit explanations for the observed relationships between joint strength and process parameters (or account for the effect of wear in systems where it occurs). It is our philosophy that work on wear in MMCs should actually precede parameterization, since it is likely that the former has a profound effect on the latter. We have thus chosen to

shift the focus of our research from parameterization to the development of models which explain and predict wear in FSW of MMCs. The regression modeling and discriminant analyses presented in this chapter represent the beginning, rather than the end product, of this work. These analyses serve as a "jumping off" point: subsequent studies (chapters VI-VIII) elucidate the effects of material properties on wear, assess the feasibility of combatting wear using harder materials and/or coatings, and determine the effect of probe wear on process signals. A major objective of this research project is to equip end-users with enough knowledge to make informed decisions about parameters, tool materials, workpiece materials, etc. which will preserve tool features, promoting conditions which will increase the likelihood of creating a structurally sound weld. We seek to transition the study of tool wear in FSW of MMCs from an iterative, experimentally intensive process to a well- understood and predictable phenomena. A fundamental understanding of the wear process, including development of a physic-based model and theory to explain wear process dynamics, is essential to achieving these goals.

CHAPTER VI

THE ROTATING PLUG MODEL OF TOOL WEAR FOR FSW OF MMCS

6.1 Motivation: The Need for a Generalized Model of Wear in FSW of MMCs

As in previous chapters, our approach to the development of a generalized model is quite similar to the diagnosis and treatment of a patient's illness. Suppose a patient complains of back pain and presents with high blood pressure and elevated levels of creatinin in the bloodstream. In order to make a definitive diagnosis, the doctor must develop a theory which explains these symptoms. The triad of flank pain, elevated BP and creatinine buildup are classic hallmarks of a kidney problem. However, further tests are needed to pinpoint the disease mechanism since these symptoms are associated with a wide array of illnesses, ranging from the relatively innocuous (kidney stone) to life-threatening (kidney failure). The tests will either confirm the original diagnosis or, in the case that no abnormality is detected, point the way toward an alternative explanation/theory. The design of the test regimen is left to the discretion of the doctor/researcher, but tests should be chosen which are relevant to the proposed theory, provide maximum information, and minimize (to the extent that is possible without compromising an accurate and complete diagnosis) patient discomfort. For a non-specific kidney problem, a common diagnostic tool is a dynamic CT scan of the abdominal region. This test provides detailed images of the organs in the abdominal cavity and "traces" the flow of a radioactive dye (injected into the patient's bloodstream prior to the scan) through the kidneys. If cysts, scarring, dead tissue, or obstructions are

present, they will be clearly visible on the scan. The dynamic CT also measures the flow and filtration rates of blood as it circulates through the kidneys. The results of this test should thus pinpoint the root cause of the patient's symptoms and assist in the selection of a course of treatment which will maximize the likelihood of a successful outcome. While it is tempting to simply treat the symptoms (in this example, the doctor could administer medication to lower blood pressure or alleviate flank pain), this does little to counteract or halt the progress of the underlying illness.

The work detailed in the previous chapter successfully characterized the "symptoms" of wear in FSW of MMCs. We were able to determine that wear in FSW of MMCs is directly proportional to the process parameters rotation rate and length of weld, but inversely proportional to traverse rate. While wear appears to be axially symmetric, the amount of volume lost by the tool varies substantially with the location along the length of the probe. The greatest material loss is observed in the region at the probe tip, while little if any material is eroded in the immediate vicinity of the shoulder. The next step in our analytical schema is to develop a theory which will explain both the dependence of wear on process parameters and the variation of wear with distance from the shoulder. While we have been able to predict wear with some success based on process parameters, the regression model is highly specific and somewhat limited (the regression model tends to overpredict wear, a trend which manifests itself as an asymmetry in the residual distribution in Figure 69 -- the mean of the residuals is positive and nonzero). The dimensionless parameter and classification algorithm based on LDA demonstrates that the results are scalable, but offers little in the way of understanding the

physical phenomena which underlie the observed relationships, knowledge which is essential to mitigating (or even eliminating) wear in this process.

The defining problem of FSW in MMCs is the erosion of the tool probe during welding. The source of this volumetric deterioration is wear precipitated by contact between the tool and the much harder reinforcement material which gives the composite its enhanced properties. While we have presumed that the wear mechanism is abrasion due to the large discrepancy between the hardness values of the surfaces in contact, microscopic evaluation (an engineering analog to the CT scan in the illness example which began this chapter) is needed to confirm abrasive wear as the primary agent and rule out the influence of other wear mechanisms (such as adhesion, corrosion, and fatigue). Another characteristic (or "symptom") of the wear process in this application is a proportionality to $\frac{\omega l}{v}$. The peculiar inverse relationship which exists between wear and traverse rate complicates our diagnosis of the wear process. Intuitively, we might think that wear should increase with traverse rate, as a tool moving at faster speeds "plows into" the abrasive particles suspended in the metal matrix with greater force. Such a relationship would only be observed if wear were a drag phenomenon (in aerodynamics, drag varies with the square of the velocity). The fact that wear is less pronounced for faster traversal rates suggests that the wear process is instead governed by shearing of the material surrounding the tool. Since the wear mechanism is not consistent with classic abrasion, we must formulate a new model to capture the process dynamics unique to FSW of MMCs. This model forms the basis for a theory of wear which explains the variation of material loss with the ratio $\frac{\omega l}{v}$.

6.2 The Rotating Plug Model

The peculiarity of the inverse relationship between wear and traverse rate indicated by the numerical models in chapter V has been noted. While the direct variation of wear with weld pitch ($\frac{\omega}{v}$) suggests some degree of temperature dependence (since weld pitch is proportional to heat input), the influence of temperature was ultimately discounted due to the very weak correlation between wear and $\frac{\omega}{v}$ over all weld lengths. The inverse relationship between wear and v thus cannot be attributed to temperature effects alone. The ratio $\frac{\omega}{v}$, however, is not solely linked to temperature – the term appears in the formulation for the rotating plug model developed by Dr. Arthur Nunes of NASA MSFC to explain material flow in friction stir welding. It is worth investigating whether this model, which has been successfully applied to FSW in other contexts, can help to understand wear in MMCs [101].

Prior to the development of the rotating plug model, flow around the FSW pin tool in the x-y plane was visualized as a series of nested rings. The plan view of this model, sketched in Figure 73, resembles the ring structure of a tree trunk. The shear which results in plasticization of the workpiece is confined to this region: the outermost ring delineates the boundary between the plasticized material and the unaffected parent material. While the clearly defined demarcation between yielded and unyielded regions of the workpiece is an idealization, microstructural evaluations of friction stir welds have shown that the dimensions of the stir zone (the heavily deformed region of the weld) roughly coincide with the planar geometry of the pin tool. The rotational model captures steady-state conditions. Since the system is at equilibrium, the torque (τ_i) on the

outermost ring (which has radius r_i and moment arm l_i) should be equivalent to the torque (τ_1) on the inner ring (radius r_1 and moment arm l_1). The shear flow stress at these locations is designated by σ_i and σ_1, respectively. Equating τ_i with τ_1 and simplifying, we obtain the relationship $\sigma_1 {r_1}^2 l_1 = \sigma_i {r_i}^2 l_i$. Since $r_1 < r_i$ and $l_1 < l_i$, it follows that ${r_1}^2 l_1 < {r_i}^2 l_i$. Thus for the system to be in equilibrium, the flow stress at location 1 must be greater than the flow stress at location i. Since flow stress is a temperature dependent quantity (and σ is inversely proportional to T), this necessarily implies that temperature increases with radius r. The direct relationship between temperature and distance from the tool indicated by the "nested ring" model, however, is inconsistent with temperature data compiled from experiments and computational fluid dynamics (CFD) modeling. Temperature contours constructed based on these investigations reveal that temperature in the workpiece material dissipates as we move away from the tool, eventually reaching ambient temperature at some distance far removed from the region of plasticization. The discrepancies between accepted empirical data and the predictions of the "nested ring" model could be attributed to non-equilibrium flows. Nunes, however, argues that this explanation is not salient since there is no evidence of the instabilities or oscillations which characterize non-equilibrium conditions [101]. Experiments show that steady-state conditions prevail after the brief initial dwell period wherein the tool plunges into the workpiece: the temperature at a particular location $T(x, y, z)$ remains largely invariant during welding provided the process variables (tilt angle, rotation speed, traverse rate, plunge depth) are held constant. Nunes proposed the rotating plug formulation as an alternative to the "nested ring" model, as the

latter fails to capture the relationship between temperature, flow stress, and distance from the tool observed in experimental work.

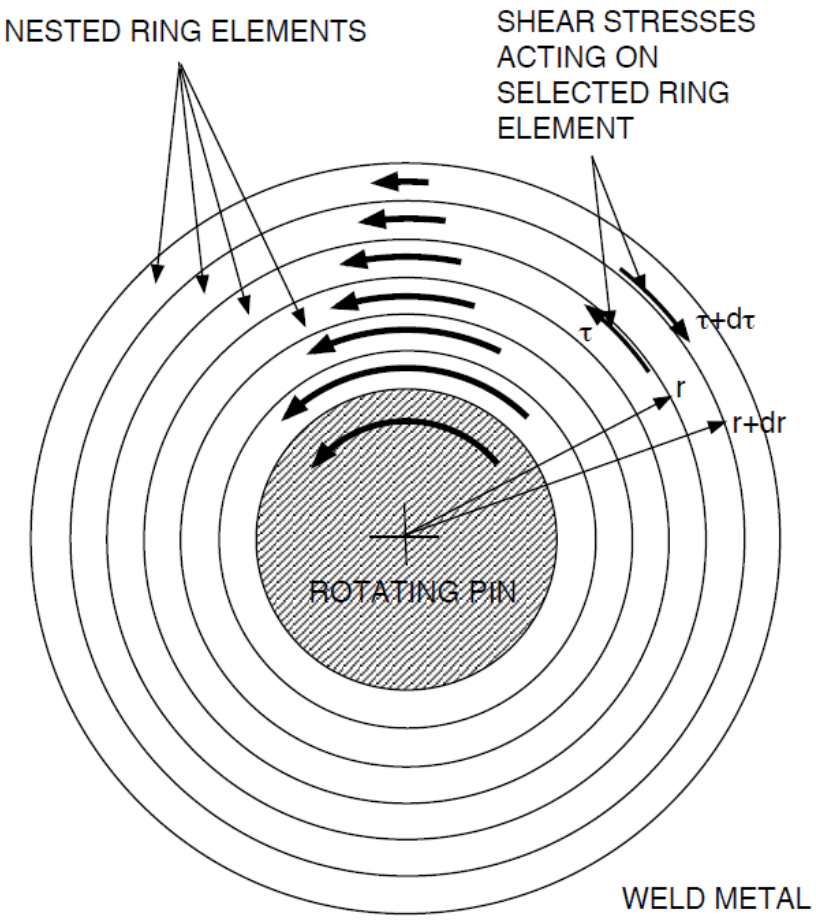


Figure 73 "Nested ring" model of equilibrium flow in friction stir welding. Plan view of workpiece in the x-y plane. The pin tool (represented by the solid circle) is surrounded by a series of nested rings of increasing radii [103].

In Nunes's model, the nested rings of plasticized material which surround the tool are replaced by a single plug of metal. This plug, which comprises the primary shear zone, rotates with the tool. The representation of material flow in FSW as a single slip

surface is substantiated by tracer experiments in which copper wire is embedded along the weld seam. Post-weld radiographs of the jointline appear to confirm the rotating plug hypothesis. The radiographs of Reynolds et al. (performed independently of Nunes's work) indicate that the tracer wire is swept around the tool and deposited on the retreating side. While Reynolds et al. assume that the FSW tool is responsible for this wiping motion, Nunes proposes that the tool never actually contacts the wire [104]. Rather, the wiping motion is initiated by the plug of plasticized metal which surrounds (and precedes) the tool as it advances through the workpiece. To illustrate this concept, Nunes draws an analogy between the rotating plug model and a merry-go-round. The tracer particle/wire steps onto the edge of the merry-go-round/plug and briefly rotates around, only to step off on the retreating side. Nunes argues that that the line of wire deposited behind the tool should parallel the line of approach [103]. The results of K. Colligan's shot-tracer experiments (in which spherical particle of steel shot, rather than wire, were embedded in a groove running the length of the weld seam), however, indicate that this is not quite the case. Some dispersion in the "exit path" of the tracer particles is evident in the characteristic plan-view radiograph in Figure 74 (adapted from Colligan) [105]. The scatter/lateral displacement of the tracer particles may be attributable to a secondary flow. Nunes postulates that the thickness of the rotating plug varies in the z-plane. For featured tools, it is probable that the tracer particles will be shifted up or down by the (secondary) flow of material in the vertical direction. Since the thickness of the plug varies with z-position, the particles "getting off" positions will differ slightly depending on their axial location. The difference in exit trajectories caused by vertical flow manifests itself as the particle dispersion in Figure 74.

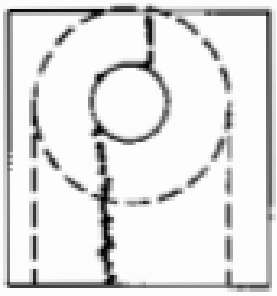


Figure 74 Plan view sketch of radiograph of friction stir weld containing steel shot tracer particles (adapted from Colligan et al.) [74]

From a conceptual standpoint, the rotating plug model seems to corroborate many of the observations documented in the literature on material flow in FSW. The model appears to explain the movement of particles around the tool during welding in a manner that is consistent with the results of tracer experiments. The anticipated variations in the thermal environment are agreement with thermocouple measurements. The flow stress σ is lowest in the rotating plug and increases with distance from this region. The major deficiency of the nested ring model lies in its inability to capture temperature dissipation as in the workpiece as we move away from the tool/heat source -- the rotating plug model is able to account for this temperature decay. The agreement of the general trends observed in the literature with the predictions of the rotating plug model should not be interpreted as a failsafe indicator of the model's "correctness." Rather, this concordance in trends is a sign that the rotating plug formulation is a more accurate reflection of FSW process dynamics than the nested ring model.

The rotating plug model is comprised of two flows: a primary flow, which consists of translation over a rotating rigid body, and a secondary "vortex" flow in the vertical direction that is facilitated by the probe features. These flows are sketched in

Figure 75 [102]. Since the tools considered in this dissertation are smooth cylinders, the secondary vortex flow can be neglected. The pin tool is represented by a cylinder– the tool rotates at speed ω and is simultaneously translated in the x-direction at velocity ν (motion denoted by the uniform straight arrows to the right of the pin). Uniform translational flow over a rotating disk conserves volume. At steady-state conditions, the trajectory of a particle of workpiece material that is captured by the rotating plug and "wiped" around the tool takes the shape of a circular arc (Figure 76). Nunes postulates that the vertical offset of this arc with respect to the tool is given by the ratio $\frac{\nu}{\omega}$ (offset is defined as the difference between the center of the arc and the center of the plug). As $\frac{\nu}{\omega}$ decreases (i.e. the traverse rate of the weld is slowed and the rotation speed is increased), the offset becomes smaller and the trajectory of the metal approaches the edge of the plug. Nunes terms the area bounded by the offset and the edge of the plug the *shear zone*; the shear zone distinguishes the weld region from the unaffected parent material.

Variables and regions pertinent to the rotating plug model are labeled on the plan view of an FSW weld in Figure 76. The tool in this weld was rotated clockwise at angular velocity ω. the advancing and retreating sides of the weld are defined relative to the direction of rotation. The labels in figure 4 correspond to clockwise rotation only (for counterclockwise motion, the labeling would be reversed). When the tool is extracted from the material, it leaves behind a "footprint" (otherwise known in FSW lexicon as a keyhole defect). In the x-y plane, this residual marking consists of an unbanded circle with an outer diameter equal to that of the tool shoulder. The tool pin, which plunges into the workpiece material, leaves behind a hole on exit. The shear zone is modeled as a thin boundary layer which is virtually coincident with the tool profile in the x-y plane.

The boundary of the shear surface is denoted in Figure 76 by a dotted line. δ, the width of the shear zone, is equal to the offset between the outer boundary of the shear surface and the rotating plug.

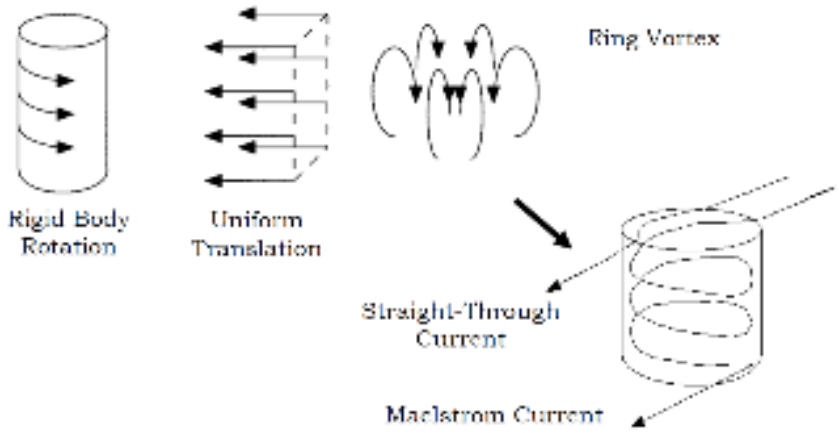


Figure 75 Sketch of primary and secondary flows in rotating plug model [1]

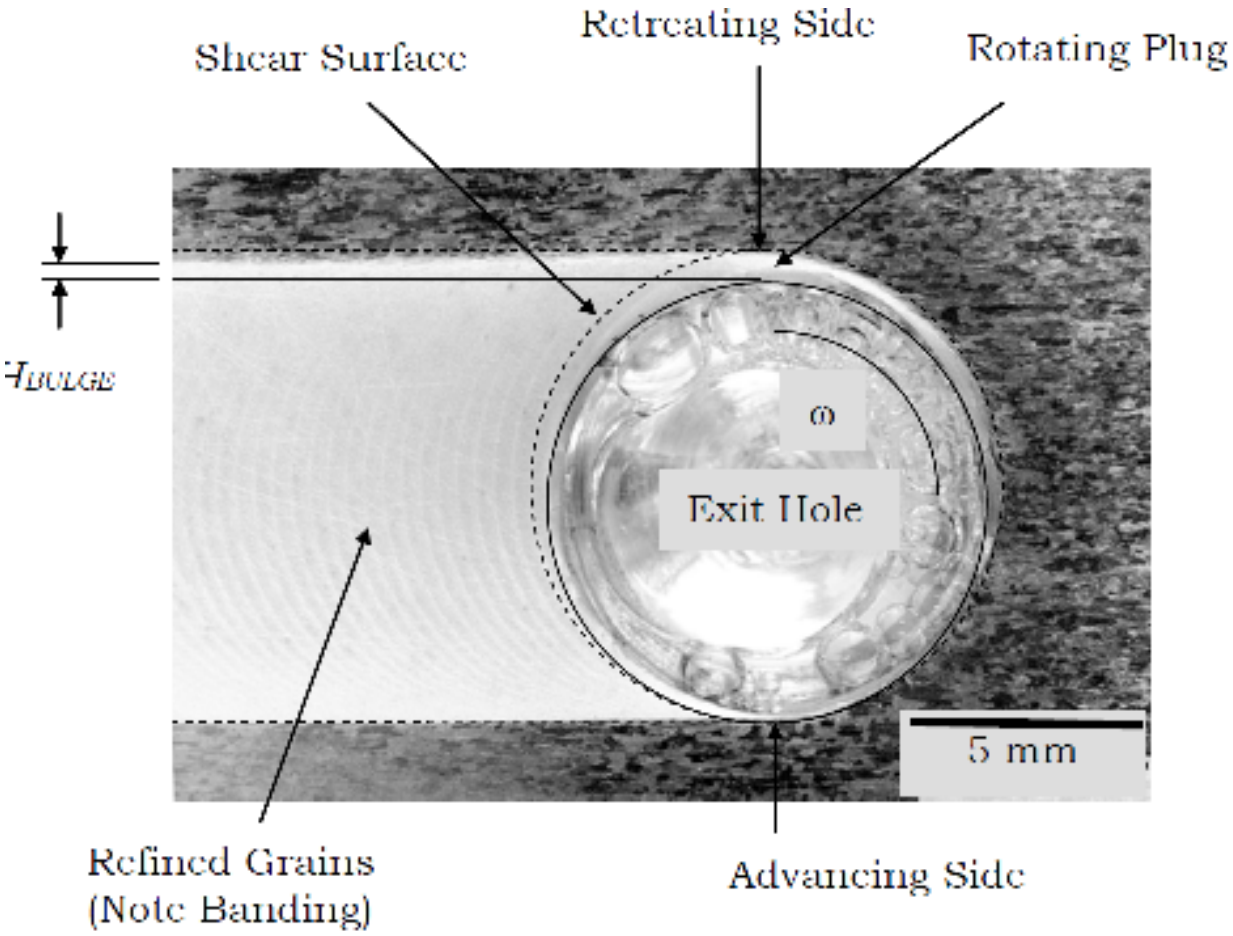


Figure 76 Plan view of FSW weld with regions of interest labeled [102]

Nunes details the development of governing equations for the rotating plug model in reference [101]. Of particular interest is an equation relating the thickness of the shear zone to process variables. Based on conservation of volume, the maximum thickness of the shear zone ($\delta_{\max}$) is given by $\frac{2\nu}{\omega}$. For process parameters which lie within the limits of the VUWAL apparatus (traverse speeds from 0.5 in/min to 14 in/min and rotation speeds ranging from 100 RPM to 2250 RPM) and a pin tool with a diameter of 0.25", $\delta_{\max}$ ranges from $2.789x10^{-6}$ in (for 100 RPM and 14 IPM) to $1.77x10^{-3}$ in (0.5 IPM and 2250 RPM). However, the parameters associated with these thicknesses are not

representative of FSW processing conditions – the former (low rotation speed and fast travel) would almost certainly result in probe fracture, while the latter (very slow traverse coupled with very fast rotation is a recipe for overheating). To get a better idea of the typical value for $\delta_{\max}$, we can calculate the average value of the function which defines it ($\frac{2v}{\omega}$) using equation 6-1. The average value of a general multivariate function $f(x,y)$ over the region with area A is given by:

$$f_{avg} = \frac{1}{A}\iint f(x,y)\ dA \qquad \text{(equation 6-1)}$$

In this instance, the area of interest is a rectangle bounded by [0.5, 14] on the ν axis and [100, 2250] on the ω axis. The characteristic value of the maximum thickness of the shear zone (sometimes referred to as the "nugget bulge") is a few orders of magnitude less than the diameter of the rotating plug. The shear zone is really little more than a thin boundary layer which surrounds the rotating plug -- the dimensions of the shear layer in Figure 76 have been exaggerated to make the region clearly identifiable.

The strain $\Delta\gamma$ is given by the ratio of the rotational velocity ($R\omega$) to the x-component of the velocity vector (ν $cos\theta$). This formulation (equation 6-2) is analogous to the representation of strain in mechanics as the stretch ratio (the ratio of the undeformed configuration to the deformed state). The strain rate ($\frac{\Delta\gamma}{\Delta t}$) is the derivative of strain with respect to time ($\Delta t = \frac{distance}{velocity} = \frac{\delta}{vcos\theta}$). The simplified form of this expression (equation 6-3b) reveals that, according to the rotating plug model, strain rate depends only on the radius of the plug R, the rotation rate ω, and the thickness of the shear zone δ (which in turn can be related to ω and ν). The shear strain rates predicted by the model

(for rotation speeds ranging from 500 RPM to 2000 RPM and traverse rates from 1 to 10 IPM) are on the order of 10^2 to 10^5. Strain rates of this magnitude are consistent with the dynamic recrystallization phenomena hypothesized to be responsible for the refined, equiaxed grain structures characteristic of FSW welds [106].

$$\Delta\gamma = \frac{R\omega}{\nu \cos\theta} \qquad \text{(equation 6-2)}$$

$$\dot{\gamma} = \frac{\Delta\gamma}{\Delta t} = \frac{\frac{R\omega}{\nu \cos\theta}}{\frac{\delta}{\nu cos\theta}} \qquad \text{(equation 6-3a)}$$

$$\dot{\gamma} = \frac{R\omega}{\delta} \qquad \text{(equation 6-3b)}$$

The rotation of the plug gives rise to a shear stress τ at the boundary. The total torque can be represented as the sum of the contributions from the probe bottom, probe sides, and tool shoulder (equation 6-4a). The force associated with each torque component is given by the product of the shear stress τ and the area over which the shear stress acts (the cross-sectional area of the probe $2\pi r^2$). The moment arm corresponding to the probe bottom and the tool shoulder is expressed as the differential dr. The moment associated with the probe bottom is calculated by summing the expression for torque ($2\pi r^2 \tau dr$) as the radius of the probe varies from $r = 0$ to $r = r$. The $\Omega_{shoulder}$ term uses the same expression for torque as $\Omega_{probe\ bottom}$, but the sum is taken over the limits $r = r$ to $r = R$ (where R is the radius of the tool shoulder and $R > r$). The contribution to torque from the tool sides is a single term ($2\pi r^2 \tau t$), where the moment arm *t* corresponds to the penetration depth of the probe (coincident with the thickness of the material for a full-penetration weld). The components of the total torque appear explicitly in equation 6-4b. Equation 6-4c is a simplified form of this expression .

$$\Omega_{total} = \Omega_{probe\ bottom} + \Omega_{probe\ sides} + \Omega_{shoulder} \quad \text{(equation 6-4a)}$$

$$\Omega_{total} = \int_0^r 2\pi r^2 \tau dr + 2\pi r^2 \tau t + \int_r^R 2\pi r^2 \tau dr \quad \text{(equation 6-4b)}$$

$$\Omega_{total} = \frac{2\pi R^3}{3}(1 + 3\frac{r^2 t}{R^3})\tau \quad \text{(equation 6-4c)}$$

The estimation of torque is complicated by the dependence of the shear stress term τ on temperature. According to [101], shear stress can be approximated using equation 6-5, where T_M denotes the melting temperature of the workpiece material and T is the weld temperature. Since FSW occurs at solid-state, it necessarily follows that $T_M > T$ and $T_M - T > 0$. The relationship between shear stress and the temperature difference is parabolic: τ and $(T_M - T)^2$ are related by the material constant β.

$$\tau \sim \beta (T_M - T)^2 \quad \text{(equation 6-5)}$$

The effect of process parameters on torque is captured by the τ term. Although ν (traverse rate) and ω (rotation speed) do not explicitly appear in any of the expressions for torque, these process parameters affect heat input/temperature, which in turn impacts the shear stress term τ. In the equations associated with the rotating plug model, the torque experienced by the tool during a weld is governed exclusively by the shear stress, the plunge depth, and the tool geometry (specifically the probe and shoulder diameter).

There is rough agreement between the torque predicted using the above equations and the torque measured using a Kistler Model 9124A rotating cutting force dynamometer [101]. The average discrepancy between the measured and computed torque for rotation speeds ranging from 350 RPM through 700 RPM is approximately 10 percent. The model tends to slightly overpredict the torque reading, an inaccuracy which

likely stems from the uncertainly inherent in estimating the steady-state weld temperature T in the temperature dependent flow stress term τ [101].

The nearly universally documented inverse relationship between torque and rotation speed in friction stir welding can be explained succinctly in terms of the rotating plug model. Since higher rotation speeds coincide with greater heat input, the temperature difference term $T_m - T$ in equation 6-5 is reduced. It follows that shear flow stress τ decreases with higher weld temperatures and, as a result, less power input is required from the tool to drive the material to plasticization. Thus the rotating plug model agrees with our intuitive understanding of FSW process dynamics as well as experimental observations. The relationships between tool geometry, plunge depth, and flow stress derived by Nunes can be exploited for the purposes of control. Longhurst et al. varied plunge depth (t in equation 6-4c) to maintain constant torque during welding [8]. The research suggests that torque control may be superior to control systems for FSW based on axial or lateral forces, as the torque signal is far more sensitive to variations in plunge depth than these measurements. According to the Nunes equations, torque should also exhibit a high sensitivity to changes in FSW tool dimensions (particularly the radial deterioration which accompanies tool wear in FSW of MMCs). Chapter IX explores whether changes in the torque signal can be used to gauge the amount of wear the tool has experienced in real-time. A robust method of wear estimation based on process signals would eliminate the need for off-line wear measurements in situations where the wear process has been extensively characterized.

The rotating plug model can also be used to estimate temperatures in the shear zone. The sole source of heat input to the weld (Q_{in}) is the mechanical power generated

by the rotation of the tool ($\omega\Omega_{total}$). In accordance with the first law of thermodynamics, the work input $\omega\Omega_{total}$ is precisely balanced by heat loss Q_{out} [101]. This thermal loss consists of four components:

a) Conduction within the workpiece
b) Conduction from the tool to the workpiece
c) Conduction from the workpiece to the backing anvil
d) Convective heat transfer from the tool to the surrounding air

(1)-(3) are approximated using various formulations of Fourier's law for steady state conductive heat transfer in one dimension ($Q = -kA\frac{dT}{dx}$). (1) corresponds to conduction through a cylindrical region with radius R, (2) is represented by conduction through a circular rod of radius R and height h, and (3) is modeled as heat flow through a sphere. The convective heat loss in (4) is expressed as the product of the workpiece's heat capacity (ρcVA) and the temperature difference ΔT. Substituting these expressions into the heat balance $Q_{in} = Q_{out}$ yields equation 6-6.

$$\omega\Omega_{total} = \frac{2\pi t k_w (T-T_0)}{ln\frac{R_0}{R}} + \pi R^2 k_t \frac{T-T_0}{1-h} + \frac{2\pi R k_a (T-T_0)}{1-\frac{R}{R_0}} + \rho CVA(T-T_o) \qquad \text{(equation 6-6)}$$

The ambient temperature is designated as T_0. R_o coincides with the radius (or distance h in the case of heat transfer through a circular rod) at which the material is in thermal equilibrium with its surroundings. The value of thermal conductivity k differs for each of the constituent terms. The subscript of the k term denotes whether the value of thermal conductivity corresponds to the workpiece (k_w), the tool (k_t), or the backing anvil (k_a). The heat capacity term ρCVA is associated with the workpiece material. By substituting

the expression for Ω_{total} (equation 6-4c) into the heat balance (equation 6-6), we obtain an expression for the temperature inside the shear zone [5]. This equation (6-7) estimates the peak temperature in the shear zone based on two process variables (rotation speed ω and plunge depth t), six material constants (k_a, k_t, k_w, ρ, C, and β), four geometric dimensions (R, R_0, R_a, and h), and the ambient temperature term T_0. The relation is transcendental and, as such, requires an iterative solution.

$$T = \sqrt{\frac{\frac{k_w t}{ln\frac{R_o}{R}} + \frac{R^2 k_t}{2h} + \frac{R k_a}{(1-\frac{R}{R_o})} + \frac{\rho C V A (T-T_o)}{\pi}}{\frac{R^3 \omega \beta}{3}(1+\frac{3r^2 t}{R^3})}} \qquad \text{(equation 6-7)}$$

In the Nunes formulation, the dependence of heat generation on tilt angle $\emptyset$ is captured in the multiplier *A* associated with the convection term ($\rho C V A \Delta T$), where *A* represents the cross-sectional area of the plug, a value given by $2rt + \pi R^2 \sin\emptyset$. The tile angle $\emptyset$ is selected by the operator – its value depends on the tool geometry, material thickness, and joint configuration. To ensure sufficient heating, the tilt angle $\emptyset$ and plunge depth *t* should be chosen such that at least 75% of the shoulder is in contact with the surface of the material. In general, the tilt angle in conventional friction stir welding lies within the range $-2 \leq \emptyset \leq 2$. More extreme tilt angles may be required for shoulderless tools (such as the conical geometry used by Lammlein et al. in [107]) or in cases of unusual geometries (spheres, cylinders, or curved plates). Tools with shoulder features (ex. scrolls) render the rake angle unnecessary. For this special class of tools, welds can be performed with the spindle head at 0 degrees (the *A* expression in that case reduces to *2rt*).

The expression for weld temperature in equation 6-7 represents a highly simplified model of heat transfer in FSW. The model deviates from reality in several important ways:

1) Representation of time and temperature-varying parameters as constants.

   In actual experiments, the total torque is a time-varying signal. However, since the torque response stagnates once steady-state conditions have been reached (oscillating slightly in time about some representative value Ω), it is reasonable to treat it as a constant for purposes of simplification. The same is true for the material constants (k in particular is expected to exhibit some degree of variation with temperature). For metals, thermal conductivity decreases with increasing T. As the material becomes hotter, the elevated velocity of the constituent particles contributes to enhanced lattice vibrations, a phenomenon that impedes the movement of the free electrons which conduct heat. This reduction in electron mobility reduces the efficiency with which energy/heat is transferred, thereby lowering the thermal conductivity. The temperatures encountered in FSW of Aluminum typically fall between 450 and 800 degrees K [108]. The accompanying shrinkage in k over this thermal range is small enough that the temperature dependence can be omitted from the analysis with little consequence [109].

2) Approximation of the thermal environment of FSW with standard heat transfer equations

In deriving the expression for heat output in FSW which comprises the right hand side of equation 6-6, Nunes used equations from well-understood heat transfer scenarios (for example, conduction from the tool to the workpiece was modeled as conduction through a rod). The difficulty of developing an equation for heat transfer in FSW stems from the complexity of the process. Even in the simplest formulation, FSW involves three distinct conduction modes, none of which precisely fit established formulations for one-dimensional, steady-state conduction. Rather than develop a host of new equations which exactly correspond to heat transfer in FSW (a task made immensely difficult by the multiple phenomena which simultaneously influence thermal transport), mathematical effort can be substantially reduced by selecting existing heat transfer equations which most closely resemble the scenarios under consideration. Heat flow in FSW is not as disparate from other, more familiar heat transfer situations as it might seem upon first examination. Returning to the transfer of heat from the tool to the workpiece via conduction, the tool geometry (or at least the portion which engages with the workpiece material) is sufficiently similar to a cylindrical rod that this aspect of heat transfer can be approximated as steady-state conduction through "a rod of [constant] area πr^2 to a region of ambient temperature T_0 attained at some distance h" [101]. There is some uncertainty regarding the distance from the heat source (R_o or h) at which a particular material in the system reaches ambient temperature. While Nunes does not detail the educated guesswork involved in these calculations, it is likely that the assumptions are based on thermocouple data or computer simulations published in

the literature on FSW. For example, it is probably safe to assume based on this data that in the case of radial conduction through a sphere or cylinder, a distance equal to 3 times the radius of the tool shoulder ($R_o = 3R$) is more than sufficient to dissipate the heat generated at the site of tool rotation.

The convection term in the heat input equation ($\rho CV\Delta T$) approximates heat loss to the surrounding air as the tool advances through the material, transitioning from a region of hot metal to cooler metal. This is an example of forced convection, as the circulation of heat from the material to the surrounding air does not entirely occur naturally but is instead induced by an external source (tool rotation). Although the argument could be made that convective heat loss in FSW belongs to a more complex set of problems (such as laminar or turbulent forced convection originating from a rotating cylinder), Nunes chooses to model convective phenomena in FSW using Newton's law of cooling. The validity of this powerful, albeit simple, equation is buoyed by its successful application to a wide-range of applications, from water boiling on a stove to wind power.

3) Omission of some heat transfer mechanisms.

-To further simply the heat loss equation, effects from the third mode of heat transfer (radiation) have been neglected. This is not to say that radiation does not occur (the material radiates heat at a rate equal to the Boltzman constant multiplied by the temperature difference to the fourth power), but its contribution

is anticipated to be much smaller than the conduction and convection terms. In Nunes's model, the contribution of plastic work (Q_p) to the heat ouput was also omitted [101]. Q_p represents the amount of heat generated in deforming the material; its contribution is anticipated to be much smaller than the sum of the heat losses from conduction and convection captured in the $Q_h$term. The exclusion of plastic work as a means of internal heat generation significantly simplifies temperature estimation based on equation 6-6.

-The FSW process is additionally assumed to take place at steady-state conditions. Thermocouple measurements documented in the literature indicate that, excepting an initial dwell period (typically 1 to 2 inches in length) in which the traverse speed is reduced by 10 to 50 percent of its steady-state value, the weld temperature remains virtually invariant with time [101]. This is not to say that a particular point in the temperature field does not undergo thermal cycling as the rotating tool approaches and retreats, or that temperature does not dissipate with distance from the tool. Rather, the steady-state assumption reflects the fact that the thermal profile *T(x,y,z,t)* is identical for all points in the field defined by (*x, y=c, z=c)*. Since the temperature profile is the same for every point in the x-plane, the process can be modeled as intransient with a constant, peak temperature of *T*. Setting the time-dependent term in the general heat diffusion equation, $\frac{\partial T}{\partial t}$, equal to zero gives rise to the expression in 6-6. While the steady-state assumption simplifies computation, it renders the governing heat transfer

equations (6-6 and 6-7) invalid for the dwell period since transient heating effects are present in this phase of the process.

Experimental verification of the temperature expression in 6-7 is documented by Nunes in reference [101]. The shear zone temperatures predicted based on the rotating plug model agree with temperature profile data derived from thermocouples embedded in the workpiece material during welding. The discrepancy between measured and predicted temperatures over rotation speeds ranging from 350 RPM to 700 RPM is consistently less than 10%. This strong correlation between measured and predicted temperatures suggest that the assumptions articulated in (1) through (3) are valid. The model's slight underestimation of shear zone thermal conditions may be attributable to the variation of the material constants with temperature, a dependence which was uniformly ignored in the interest of simplifying calculations. Despite these simplifications (steady state, temperature independent constants, etc.), the model appears to be a robust predictor of overall system behavior. In reference [101], Nunes goes on to derive expressions for translational, lateral, and axial forces experienced by the tool in the FSW process. The correlation between measured and computed values, however, is substantially weaker than that for torque and temperature data. While the computed values for F_x, F_y, and F_z are the same order of magnitude as the values recorded during welding, the observed trends (for instance, that lateral and translational forces increase with rotation speed) are not reflected in model predictions. Nunes postulates this discrepancy is due to interactions between the tool and the backing anvil which are not accounted for in the rotating plug model (for a full penetration weld, the bottom of the pin is in close proximity – within .01" – of the backing anvil) [101]. We have chosen to restrict our

evaluation of the rotating plug model to variables for which there is there is close agreement between model and experiment (temperature and torque).

Additionally, temperature and torque are dependent on tool geometry (R and r appear explicitly in the equations for both torque and temperature). This dependence suggests that erosion of the tool volume as a result of wear should translate into corresponding changes in torque and temperature. In this way, the rotating plug model may provide insight into how wear impacts the mechanics of the friction stir welding process. Decreases in probe radius r due to circumferential wear and/or a reduction in penetration depth t resulting from wear along the length of the probe should, according to the rotating plug model, produce a change in the thermal environment and (subsequently) a detectable deterioration in torque. The implications of these dependencies for on-line sensing of wear will be evaluated in chapter IX.

Before proceeding with a discussion on modifying the rotating plug model for MMCs (section 6.3), some clarification is necessary. Throughout the dissertation, terminology from fluid dynamics is occasionally used to describe the rotating plug model. For instance, the rotating plug was described as a boundary layer surrounding the tool profile in the x-y plane. We must emphasize, however, that the rotating plug model is *not* a fluids model. The "fluid" in this instance is not a fluid at all, but rather a metal in the plasticized state. Thus the concepts we typically associate with fluid flows (stagnation points, Bernoulli's equation, total pressure, etc.) are defined differently than in fluid dynamics (if they are applicable at all). If we borrow a term from fluid dynamics (such as boundary layer) to describe FSW, we do so only in the interest of convenience and to facilitate easy visualization of the process dynamics. In FSW, the metal in

question is *never* in a fluid state (this is the chief objection to the use of computational fluid dynamics in modelling FSW). The rotating plug model is particularly attractive to FSW researchers because its roots lie in plasticity theory. The surrounding metal in FSW is in a plasticized state – as such, it is our opinion that plasticity models will be more effective in capturing the deformation processes which underlie a solid-state welding process.

The rotating plug model is not beyond reproach: the model is a simplification which neglects some potentially significant interactions, a point evidenced by the disagreement between the measured and computed force values in reference . Despite its shortcomings, it serves as a much preferred alternative to CFD for modeling FSW of MMCs. While there are some papers on CFD modeling of FSW which exhibit good correlation with experimental data, work in this field is restricted to modeling of conventional Aluminum alloys, materials for which the temperature dependence of material properties is well-understood. Such detailed material property data does not yet exist for MMCs. Since the quality of a CFD analysis often hinges on the user's ability to accurately define the properties of the material being modeled, the development of a CFD model for MMCs would certainly be impeded by the lack of information readily available on the behavior of these materials at temperatures consistent with FSW. CFD is additionally complicated by the presence of reinforcement. In CFD software, the MMC would be modeled as a multi-phase material to accommodate this peculiarity – as we will see in the next section, a fluid flowing around solid, immobilized occlusions (which is what a CFD model of FSW of MMCs would look like) is not truly representative of the system. The rotating plug model is well-suited to FSW of MMCs precisely because its

predictions are largely independent of material properties (as indicated previously, temperature dependent material property data is sometimes difficult to come by for newer or more exotic materials). The merit to this unconventional approach is readily apparent: not only does it provide a broad, conceptual view of the process, but it also offers a path to analysis that is unencumbered by meshing, nodes, user defined functions for material properties, and the concerns about convergence which accompany CFD. The remainder of this chapter details the development of a rotating plug model for FSW of MMCs (section 6.3); the next chapter documents the steps in the experimental validation of this formulation.

6.3 Developing a Rotating Plug Model for Tool Wear in FSW of MMCs

6.3.1 The Basic Model

The previous section summarized Nunes's development of a rotating plug model for friction stir welding. Based on the strong correlation between predicted and observed temperature and torque data, the model seems to have captured the thermomechanical changes which take place in Aluminum alloys being joined by FSW. Nunes does not comment on the applicability of the model to materials other than the Aluminum alloy used in the verification experiments. However, because the welding process in the rotating plug formulation is modeled without a specific tool or workpiece material in mind, the model can be extended to FSW of MMCs with some modifications.

In the existing literature on FSW of MMCs evaluated in chapter II, little to no consideration was given to the fundamental physics which underlie the wear process. It is generally agreed upon that tool wear is negligible in Aluminum alloys which contain no reinforcement particles. The wear incurred in FSW of MMCs is assumed to be a result of the hard reinforcing particles which come in contact with the tool as it traverses through the workpiece. Initially, it seems difficult to reconcile the rotating plug model with this sort of abrasive wear process. In Nunes's model, the tool is surrounded by a plug of plasticized metal (shown in the x-y plane in Figure) which rotates with the tool. Nunes hypothesized that shearing occurs at the boundary of the plug and the parent material. This conjecture precludes wear, as the shearing which would normally result in removal of material cannot, within the constraints of the model, take place at the surface

of the tool. In light of this seeming incompatibility, why would we consider the rotating plug model as a candidate to represent the wear process in FSW of MMCs?

Research by Prater et al. established dependence of wear on the weld pitch ratio ($\frac{\omega}{v}$) [51-52,110]. It is this relationship which suggests the rotating plug model, since the width of the plug is partially a function of this ratio. For wear to occur within the context of the rotating plug model, however, there should be shearing at the particle/tool interface, a condition which is irreconcilable with the model assumption that restricts shearing to the plug boundary. Is there then a special case of the rotating plug model for FSW of MMCs, wherein the presence of the abrasive particle deforms the plug enough to permit the particle to indent the tool surface, plowing out a groove as the tool rotates? According to Nunes, it is unlikely that the velocity of the flow in the rotating plug is sufficient to facilitate removal of material in quantities that would explain the significant wear rates documented for FSW of MMCs. Nunes suggests that wear cannot be explained by flow inside the plug and should instead be considered in terms of rotational shear at the plug surface [111].

Figure 77 shows a cross-sectional view of the pin portion of the FSW tool in the x-y plane, rotating counterclockwise and flanked by a rotating plug of plasticized metal. As in Figure 4, the width of the rotating plug varies with angular position (this variation should be independent of workpiece material). The distinguishing feature of Figure 77 is the presence of the hard particle (represented by a circle), which is able to contact the tool at locations where the thickness of the rotating plug is less than the particle radius. In these regions, the particle is "captured" by the rotating plug and, as a consequence of its proximity to the tool surface and the disparity in hardness values between the abrasive

and the tool material, removes material from the tool surface. The bit of material removed by the action of the abrasive is referred to as a "wear particle." During the course of a weld, the generation of wear particles is necessarily equivalent to the volume lost by the tool.

*While the β, k, c, and ρ values in equation 6-5 through 6-7 are specific to the material properties of the workpiece and/or tool, these terms do not presuppose the choice of any specific metal.

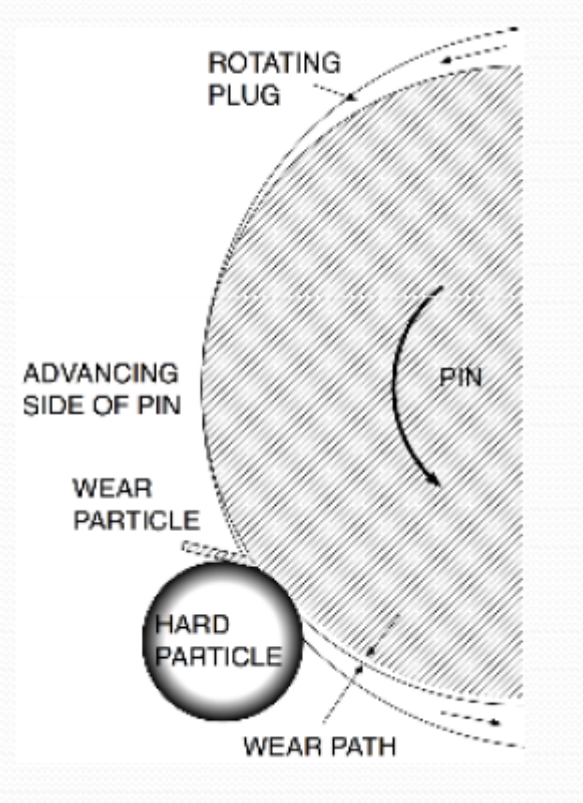


Figure 77 Schematic for rotating plug model of tool wear in friction stir welding of particle-reinforced metal matrix composites (MMCs) [111].

From this basic idea, the mechanics of the model can be mathematically formalized. Consider the spherical particle of diameter d impinging on the bottom of the

tool probe (z=0) with radius r in the x-y plane as shown in Figure 78. At this specific location in the z-plane, the width of the shear surface is 0 at $\theta = \frac{-\pi}{2}$. The thickness of the shear surface δ increases for angular positions in the range $\frac{-\pi}{2} < \theta < \frac{\pi}{2}$ (passing through $\theta = 0$) as the zone "picks up" plasticized material. δ reaches a maximum at $\theta = \frac{\pi}{2}$, then decreases over the interval $(\frac{\pi}{2}, \frac{-\pi}{2})$ as material is deposited on the retreating side. For clockwise tool rotation, the regions characterizing gain and loss of material are reversed: the shear surface accumulates material from $\theta = \frac{-\pi}{2}$ to $\theta = \frac{\pi}{2}$ (passing through $\theta = -\pi$) and deposits it from $\theta = \frac{\pi}{2}$ to $\theta = \frac{-\pi}{2}$ (rotating through $\theta = 0$). Although the width of the plug increases with vertical distance, the shear surface is symmetric about the y-axis at every location in the z-plane.

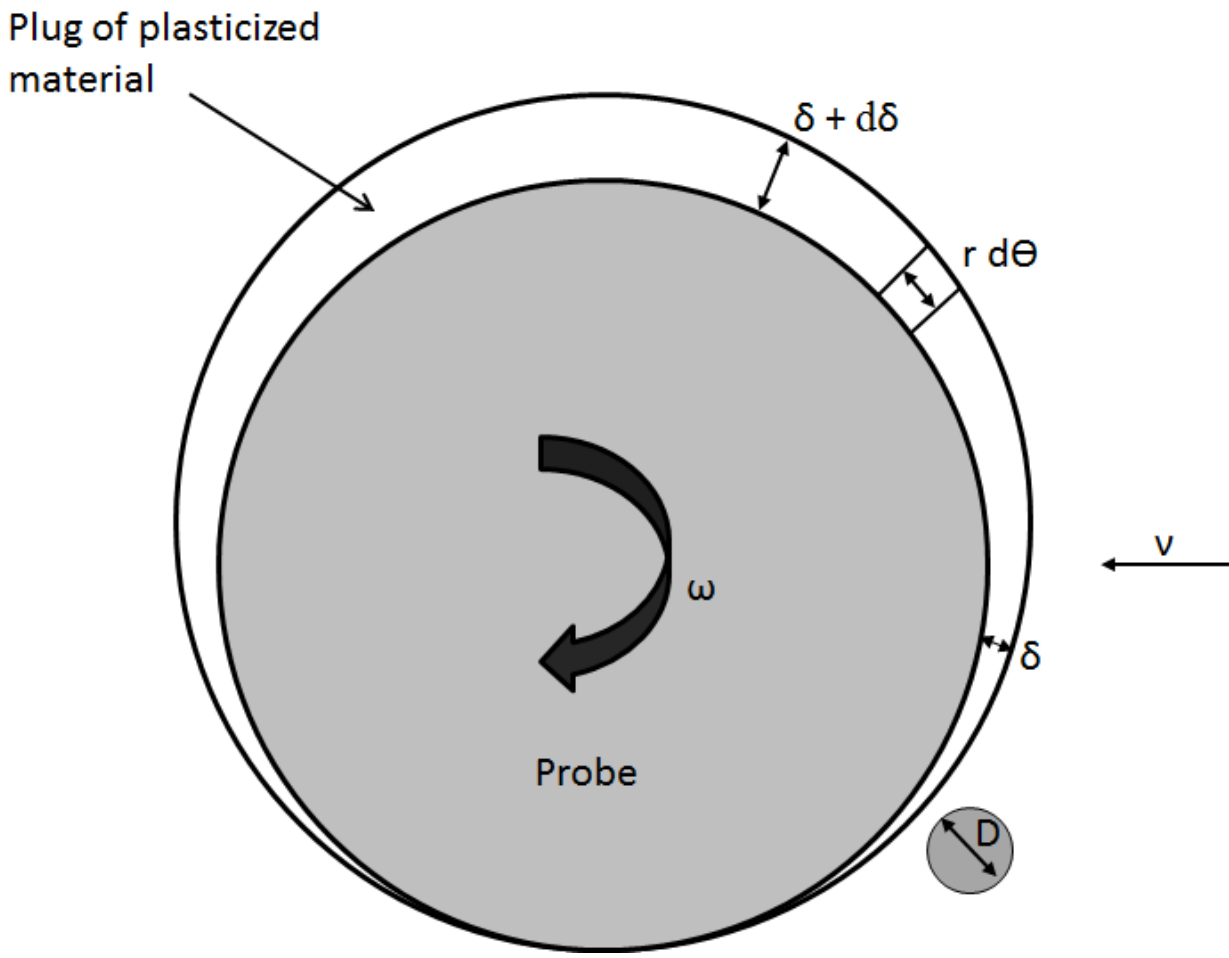


Figure 78 Rotating plug model of tool wear in FSW of MMCs in x-y plane at z=0.

The variation in the width of the shear zone with angular position can be obtained from a continuity equation. Suppose that the tool is rotating in a stationary position with spindle speed ω as the workpiece material advances at traverse speed v. The entry rate/inflow of a surface element at the bottom of the probe in this scenario (sketched in Figure 78) is given by $\cos\theta \; r \, d\theta$. Similarly, the outflow can be expressed as $-r\omega\delta +$

$r\omega(\delta + d\delta)$. Equating the inflow and outflow expressions (both have units of $\frac{volume}{time}$), we obtain:

$$\nu \cos\theta \; r \, d\theta = r\omega \, d\delta \qquad \text{(equation 6-8)}$$

$\nu \cos\theta$ represents the x-component of the velocity (ν_x); $d\delta$ is an infinitesimal change in thickness as we move counterclockwise toward the retreating side of the probe, where the additional material will be deposited. Separating variables yields the differential equation:

$$\frac{d\delta}{d\theta} = \frac{\nu}{\omega}\cos\theta \qquad \text{(equation 6-9)}$$

Integrating and applying the boundary condition $\delta(-\frac{\pi}{2}) = 0$ produces equation 6-10, which expresses the width of the shear surface δ as a function of rotation speed ω, traverse rate ν, and angular position θ.

$$\delta = \frac{\nu}{\omega}(1 + \sin\theta) \qquad \text{(equation 6-10)}$$

For a given angular position, the thickness of the rotating plug thus depends only on the process parameters ν and ω. The width of the shear surface is inversely proportional to the weld pitch (the ratio of rotation speed to traverse rate $\frac{\omega}{\nu}$). This dependency recalls the considerations which initially led us to consider the rotating plug model for FSW of MMCs. The observed relationship between tool wear and the process parameters ω and ν can be understood within the context of equation 3. Stated simply, parameters which correspond to higher weld speed (i.e. faster rotation speeds with comparatively slower rates of traverse) result in a thinner shear surface. Conversely, parameters with lower

weld pitches (slower rotation rates and higher traverse speeds) are associated with thicker shear surfaces. The width of the shear surface δ thus seems to have important implications for wear. For spherical abrasive particles with diameter D, abrasion should only occur for parameters and angular positions where the thickness of the shear surface δ (defined as the clearance between the shear surface and the pin boundary) is less than or equal to the particle radius ($\delta < \frac{D}{2}$). Under this conditions, the particle, which is still buried sufficiently in the surrounding non-rotating metal to be held fixed with respect to the rotating probe, is able to span the shear zone clearance δ and contact the probe surface. The particle is able to plough out material circumferentially along the probe until a condition is achieved where $\delta \geq \frac{D}{2}$. When the shear surface clearance equals or exceeds the radius of the spherical particle, the pull of the rotating plug on the particle is greater than its adhesive forces binding it to the surrounding metal matrix. At $\delta = \frac{D}{2}$, the particle is "captured" by the rotating plug. The particle moves at the same speed as the surrounding flow, $r\omega$, but does not contact or abrade the probe under these circumstances. The shear surface acts as a buffer when the plasticized metal in the shear region has sufficient relative velocity to dislodge the particle and sweep it around the tool, preventing contact between the particle and the comparatively softer tool material [111].

The above formulation of the rotating plug model of wear for FSW of MMCs is valid only at the bottom of the probe (z=0). As we advance vertically along the length of the probe, the shear surface becomes thicker. The axial growth of the shear surface is represented by the dz term in the volume balance of equation 6-11.

$$r\, d\theta\, dz\, v\, \cos\theta = r\omega\, dz\, d\delta \qquad \text{(equation 6-11)}$$

Integrating and applying the general boundary condition $\delta_0 = \delta(\theta = \frac{-\pi}{2}, z)$ yields equation 6-12, where δ_0 represents the minimum clearance between the shear region and the tool surface.

$$\delta = \delta_0 + \frac{v}{\omega}(1 + sin\theta) \qquad \text{(equation 6-12)}$$

At the bottom of the probe ($z = 0$), δ_0=0: in this special case, equation 6-12 reduces to $\delta = \frac{v}{\omega}(1 + sin\theta)$, the expression developed to describe the variation in the thickness of the shear zone with process parameters and angular position at the base of the tool. For $0 < z < H$ (where H is the height of the probe), the minimum clearance is positive, nonzero, and increasing with axial position z, reaching a maximum value at the probe/shoulder interface $z = H$. At this location, δ_0 is equal to the difference between the shoulder radius R and the probe radius r. The shear region at the "top" of the probe can thus be expressed as $\delta = (R - r) + \frac{v}{\omega}(1 + sin\theta)$.

The axial variation of the shear zone offers a potential explanation for the wear patterns observed in the experiments of chapter V, wherein wear is concentrated at the bottom of the probe and becomes less pronounced closer to the shoulder. In the context of the rotating plug model for wear in FSW of MMCs, abrasion will *always* take place at the bottom of the probe ($z = 0$) because there is a location ($\theta = \frac{-\pi}{2}$) where the boundary of the shear surface and the tool probe coincide ($\delta_0 = 0$). As the shear surface thickens with increasing distance from the probe bottom, it becomes more difficult for the reinforcing abrasive particles to contact the tool, resulting in less wear with increasing

axial distance z. The relatively large minimum clearance ($\delta_0 = R - r$) present at the location ($z = H$) where the probe meets the shoulder preclude wear from all but the largest particles. The rotating plug model dictates that volumetric wear should dissipate as axial distance from the shoulder decreases, a prediction which reflects the trends observed in previous experimental work.

According to the rotating plug model, wear can only occur at locations where the inequality $\delta_0 + \frac{v}{\omega}(1 + sin\theta) \leq \frac{D}{2}$ is satisfied. This expression is a generalized version of the $\frac{v}{\omega}(1 + sin\theta) \leq \frac{D}{2}$ condition developed to characterize abrasion at the bottom of the probe ($z = 0$). From this expression, the region over which wear occurs in the $z = 0$ plane is defined as $\sin\theta \leq \frac{\omega D}{2v}$ -1. With $\frac{-\pi}{2}$ (the angle where $\delta_0 = 0$) as the lowerbound, $\frac{-\pi}{2} < \theta < \sin^{-1}\frac{\omega D}{2v}$ gives the range of angular positions where particles can contact the probe. The cutting arc $\Delta\theta$ represents the difference between the maximum (θ_{max}) and minimum angle ($\frac{-\pi}{2}$) in this range. However, for all but the smallest particles and slowest rotation speeds, the length of the parallel grooves ploughed out by the abrasive particles should approximately coincide with the circumference of the tool. This confirms the results of the imaging study in chapter V, which indicated that wear is axially symmetric (i.e. deterioration in the probe cross-section is independent of the angular orientation of the pin with respect to the vertical). The fact that the measured material loss is virtually unchanged when the tool is rotated about the z-axis by some arbitrary angle θ suggests that while wear occurs over an arc $\Delta\theta$, the rotation of the tool effectively counteracts this effect, ensuring that wear is circumferentially equalized.

6.3.2 Size of Wear Particles

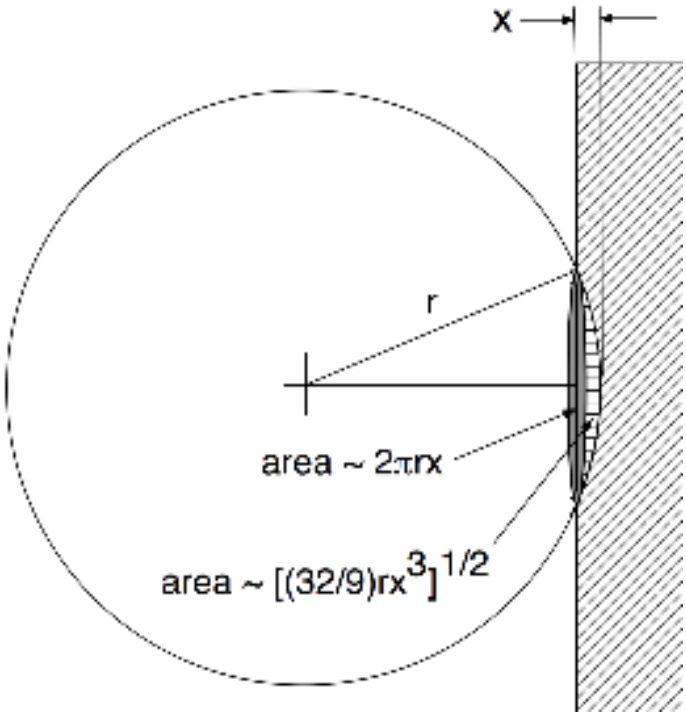


Figure 79 Hard particle of radius r impinges on tool surface, resulting in an indentation of depth x. Size of particle relative to tool is exaggerated for purpose of visualization [111].

With some algebraic manipulation, we can use the rotating plug model of wear to estimate the size of wear particles. If we assume that the total volume of the wear particles is precisely equivalent to the amount of material lost by the probe during the course of welding, the results of this analysis yields an approximation for the volumetric deterioration of the tool probe (expressed, as in chapter V, as a percentage of the probe's initial volume). The geometry which serves as the basis for this formulation is drawn in Figure 7. The abrasive particle impinges on the surface of the cylindrical probe

(represented as a flat surface) and indents the probe by some depth x. The particle radius r forms the hypotenuse of a right triangle with base $r - x$ and height y. The base, height, and hypotenuse are related by the Pythagorean Theorem (equation 6-13). Since y is also the radius of the contact region, the area of particle contact can be expressed as πy^2 (equation 6-14). While the depth x of the indentation ultimately depends on a number of factors (including the material properties of the tool and the particle), it can be reasonably assumed that x in any instance is much less than the radius of the particle. Taking the discrepancy in magnitudes between x and r into account, the expression for the contact area reduces to $2\pi rx$.

$$r^2 = (r - x)^2 + y^2 \qquad \text{(equation 6-13)}$$

$$\pi y^2 = \pi(\sqrt{2rx - x^2})^2 = 2\pi rx(1 - \tfrac{x}{2r}) \qquad \text{(equation 6-14)}$$

From the contact area, the area swept out by the particle (a rectangle with length $2y$ and width dx) is calculated from the integral in equation 6-15. This integral can be computed from a table of integrals and further approximated using a Taylor series expansion. The swept area A_s over the interval 0 to x is thus estimated as $\sqrt{\frac{32}{9}rx^3}$ (equation 6-16), the simplified output of the integration.

$$\int_0^x 2y\, dx = \int_0^x 2\sqrt{2rx - x^2}\, dx \qquad \text{(equation 6-15)}$$

$$A_s = \sqrt{\frac{32}{9}rx^3} \qquad \text{(equation 6-16)}$$

To calculate the size of the wear particle, the swept area is multiplied by the distance over which the particle contacts the tool ($r\omega\Delta t$), where Δt denotes the time the particle is

engaged with the tool surface. The expression for the volume of the wear particle V_W appears in equation 6-17.

$$V_W = r\omega \Delta t A_s \qquad \text{(equation 6-17)}$$

In Nunes's original rotating plug model for FSW, the workpiece material moves in the x-direction with velocity $v_x = v\cos\theta$. By virtue of this motion, the particles which are embedded in the Aluminum matrix of an MMC should also be translated at velocity v $\cos\theta$. Movement of the particle continues in this manner until one of two conditions is reached: 1) the particle is swept away in the surrounding flow (this occurs where $> \frac{D}{2}$) or 2) the particle impinges on the tool surface ($\delta \leq \frac{D}{2}$). These conditions can also be interpreted in terms of the flow stresses in shear of the pin (τ_P) and the weld matrix (τ_M). According to plasticity theory, a normal stress equivalent to $2\tau_P$ to shear the pin along a plane at 45 degrees with respect to the swept area A_s. For the metal matrix, the tensile flow stress (σ), is equal to twice the shear flow stress τ_M. Rabinowicz estimates that the stress necessary for indentation is three times the tensile flow stress (or $6\tau_M$) [49]. In turn, the normal stress required for material to move past the pin is twice the indentation stress (or $12\tau_M$). Flow stress can be converted to force by multiplying by the swept area $\int 2y\,dx$. For scenario (1), wherein the particle is swept away in the surrounding flow, to occur, the forces outside the plug boundary/metal matrix must either exceed or balance the forces at the plug boundary. The expression in equation 6-18 represents the condition which must be present for the particle to be swept away without impinging on the tool surface (in which case the force balance is zero).

$$2\tau_p \int_0^x 2ydx + 12\tau_M \int_x^\delta 2ydx = 12\tau_M \int_\delta^r 2ydx + 12\tau_M \int_0^r 2ydx \qquad \text{(equation 6-18)}$$

Integrating and simplifying, we obtain:

$$\left(\frac{x}{r}-1\right)\sqrt{2\frac{x}{r}-\left(\frac{x}{r}\right)^2}+\sin^{-1}\left(\frac{x}{r}-1\right)=-\frac{\pi}{2}-\frac{12\tau_M}{\tau_p-6\tau_p}\left[\left(\frac{\delta}{r}-1\right)\sqrt{2\frac{\delta}{r}-\left(\frac{\delta}{r}\right)^2}\text{ -}\sin^{-1}\left(\frac{\delta}{r}-1\right)\right] \quad \text{(equation 6-19)}$$

We previously established that since $x \ll r$, the ratio $\frac{x}{r}$ is approximately 0. It is also true that the flow stress necessary to shear the pin (τ_p) is much greater than the flow stress required to shear the metal matrix (τ_M). If this were *not* the case, wear would be observed in FSW of aluminum alloys with steel tools (since the flow stress needed to shear the material would exceed that which results in shearing of the probe). Applying these conditions ($\frac{x}{r} \ll 1$ and $\tau_p \gg \tau_M$) to and solving for $\frac{x}{r}$ results in:

$$\frac{x}{r}=72\left(\frac{\tau_M}{\tau_p}\right)^2\left[\left(\frac{\delta}{r}-1\right)\sqrt{2\frac{\delta}{r}-\left(\frac{\delta}{r}\right)^2}\text{ -}\sin^{-1}\left(\frac{\delta}{r}-1\right)\right] \qquad \text{(equation 6-20)}$$

If the clearance between the boundary of the tool and the rotating plug δ is much smaller than the particle radius r, equation 6-20 reduces to:

$$\frac{x}{r}=178\left(\frac{\tau_M}{\tau_p}\right)^2 \qquad \text{(equation 6-21)}$$

Equation 6-21 is only valid when $\frac{\delta}{r}$ is approximately 0. According to Nunes, this condition is satisfied when $\frac{\tau_M}{\tau_p} \leq 0.024$ [111]. For flow stress ratios greater than 0.024, the clearance between the rotating plug and the probe δ is large enough relative to the particle radius to render the $\frac{\delta}{r}=0$ approximation invalid.

Another special case of equation 6-20 occurs when the clearance δ is on the same order as the particle radius r ($\frac{\delta}{r} \approx 1$). In this scenario, the ratio of indentation distance to

particle radius ($\frac{x}{r}$) is approximately zero. Under this condition, the particle is unable to indent the tool, as it is carried away from the tool by the rotating plug. This result has important implications for parameter selection. The rotating plug model of wear for FSW of MMCs suggests that one avenue to reduce wear is to choose parameters capable of producing plug clearances δ which exceed the radius of the particle, an idea previously articulated in the discussion of conditions for abrasion in section 6.3.1.

Now that we have defined a relationship between indentation depth x, particle radius r, plug clearance δ, and shear flow stresses τ_M and τ_p, we can develop an expression for the volume of tool material removed by an abrasive particle in a time increment, otherwise known as the wear rate $\frac{dW}{dt}$. Since $\frac{dW}{dt}$ can be directly compared with experimental data, the output of this expression will provide some idea as to how well the model predicts system behavior.

The time it takes for a particle to indent the tool surface to some depth x is simply the indentation depth divided by the velocity at which the particle is moving ($v\cos\theta$). Written as a differential, the indentation time dt is $\frac{dx}{v\cos\theta}$. The volume removed by the particle is the product of A_s (the swept area) and $r\omega\Delta t$. Equation 6-22 expresses the differential volume dv removed by a particle of radius r in time dt (this is an extension of equation 10, with the evaluation of the integral $\int_0^x 2ydx$ substituted for A_s).

$$dv = \{[(\frac{x}{r}-1)\sqrt{\frac{2x}{r}-(\frac{x}{r})^2}+\sin^{-1}(\frac{\delta}{r}-1)]+\frac{\pi}{2}\}r^2R\omega\frac{dx}{v\cos\theta} \qquad \text{(equation 6-22)}$$

There is, however, a constraint on the time component. The time it takes the particle to indent the tool surface is less than the time it takes the tool to complete half a revolution (180 degrees or π). The time available for material removal in a given rotation is $\frac{\pi}{\omega}$. Substituting for dx and applying this condition yields the inequality:

$$\int_0^x \frac{dx}{v\cos\theta} \approx 72(\frac{\tau_M}{\tau_p})^2[(\frac{\delta}{r}-1)\sqrt{2\frac{\delta}{r}-(\frac{\delta}{r})^2} - \sin^{-1}(\frac{\delta}{r}-1)]^2 < \frac{\pi}{\omega} \quad \text{(equation 6-23)}$$

Rearranging and solving for $\cos\theta$ produces equation 6-24.

$$\cos\theta > \frac{72}{\pi}(\frac{r\omega}{v})\,(\frac{\tau_M}{\tau_p})^2[(\frac{\delta}{r}-1)\sqrt{2\frac{\delta}{r}-(\frac{\delta}{r})^2} - \sin^{-1}(\frac{\delta}{r}-1)]^2 \quad \text{(equation 6-24)}$$

Consider the case where $\frac{\delta}{r} \approx 0$. The expression for $\cos\theta$ can then be simplified to

$$\cos\theta > 18\pi(\frac{r\omega}{v})\,(\frac{\tau_M}{\tau_p})^2 \quad \text{(equation 6-25)}$$

Suppose that the particle has a radius of 6 μm = 0.00024" (this is the particle size corresponding to the F500 abrasive inclusions in the composites manufactured by mc21 and used in all experiments in this work). The flow stress ratio is estimated to be 0.02. At very slow rotation speeds, there is less opportunity for the particle to score the tool (owing to the thicker plug produced at these parameters). However, as we concluded in section 6.3.1, scoring seems to be equalized by the rotation of the tool. In experiments, wear was observed to occur all around the circumference of the tool – material removal is not limited to a small segment of the probe perimeter (as a strict interpretation of this analysis would indicate).

The simpler equation derived in section 6.3.1 ($\sin\theta \leq \omega(\frac{r-\delta_0}{v})$ -1) similarly defines an angular range over which the particle can come in contact with the tool surface. At the bottom of the probe ($z = 0$), δ_0=0 and the region over which scoring can occur for the parameters 1000 RPM and 3 IPM is +/- 60.16 degrees. For the 200 RPM/3 IPM case, this range decreases to +/- 26 degrees. While these values are larger than those predicted by equation 18, they are still inconsistent with experimental results. From this point forward, we will assume that rotation speeds are sufficient for particles to abrade the probe over its entire circumference.

When $\delta \ll r$, the ratio of the indentation depth to the particle radius is given by equation 14. If we take the average distance the particle indents the tool to be approximately half the x value, then the expression for swept area becomes:

$$A_{s,avg} = \sqrt{\frac{32}{9} r {x_{avg}}^3} = \sqrt{\frac{32}{9} r [89 \left(\frac{\tau_M}{\tau_p}\right)^2 r]^3} \qquad \text{(equation 6-26)}$$

If the wear region coincides with the circumference of the tool ($2\pi R$), then the volume of an average wear particle size is:

$$(2\pi R) A_{s,avg} = 9900 r^2 R \left(\frac{\tau_M}{\tau_p}\right)^3 \qquad \text{(equation 6-27)}$$

The wear rate $\frac{dW}{dt}$ is calculated by multiplying the average size of a wear particle by the frequency with which wear particles are produced. The latter quantity depends on a number of factors: the workpiece material volume fraction n (calculated based on the diameter of the particle and the percentage reinforcement), the traverse rate v (which

governs the speed at which abrasive particles approach the rotating plug), the radius of the tool probe, and the variation of the minimum plug clearance δ_0 with z. At z=0 (the bottom of the probe), $\delta_0 = 0$ (wear particles will always be produced at this location) δ_0 monotonically increases along the length of the tool, eventually reaching a point where the minimum clearance δ_0 is equal to the radius of the particle. Above this location, wear particles do not form since the abrasive is unable to contact the tool surface. The volumetric wear rate is thus

$$\frac{dW}{dt} \approx 9900r^2R\left(\frac{\tau_M}{\tau_p}\right)^3\{n\nu R\int_{\delta_{0(z)=0}}^{\delta_{0(z)=r}}(1-\frac{[r-\delta_0(z)]\omega}{\nu}+1)dz\} \quad \text{(equation 6-28)}$$

The integral which accounts for the decrease in the frequency of wear particle production with vertical position z can be approximated by a right triangle with height Δz and base $R(1-\frac{r\omega}{\nu}+1)$. At Δz the minimum distance between the shear surface and the probe is r. The scored area depicted in Figure 80 is calculated using equation 6-29. Substituting this expression for the integral component of $\frac{dW}{dt}$ gives rise to equation 6-30.

$$Scored\ area = \frac{1}{2}R(2-\frac{r\omega}{\nu})\Delta z \qquad \text{(equation 6-29)}$$

$$\frac{dW}{dt} \sim 4950nr^2R^2\left(\frac{\tau_M}{\tau_p}\right)^3\Delta z(2-\frac{r\omega}{\nu})\nu \quad \text{(equation 6-30)}$$

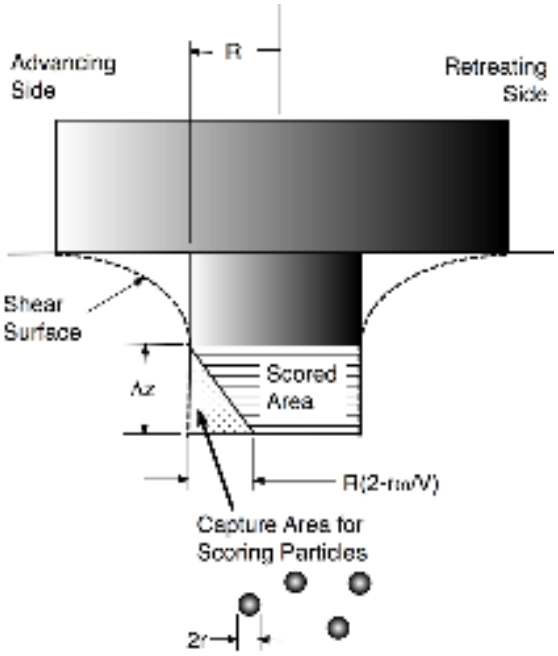


Figure 80 Illustration of shear region and scored area for friction stir welding tool probe [111].

The wear rate calculated based on equation 6-30 can be directly compared with the amount of wear observed in experiments. The composites used in the experimental study in chapter V had 17.5% Silicon Carbide particulate reinforcement. Assuming a median particle diameter of 0.00025" (the dimension provided by the material manufacturer), the particle density for Al 6061/Sic/17.5 by volume is $2.58\ x\ 10^6/in^3$. The shear flow stress ratio $\frac{\tau_M}{\tau_p}$ for a steel tool used to join an Aluminum alloy is approximately 0.025 (τ for Aluminum is 2.5 ksi, while τ for steel is 100 ksi). The probes used in the chapter V wear study had a Trivex geometry – to simplify the wear rate calculation, we will consider these tools to be right circular cylinders with a diameter equal to the 0.25" (pin radius R=0.125"). The height of the scored area (Δz) is assumed to be approximately equal to half the length of the probe (0.125"). If the tool traverses at 5 IPM and rotates at 1000

RPM (6,280 radians/minute), then the rate of wear (based on equation 23) is $2.1\ x\ 10^{-4}$ $\frac{in^3}{min}$. The volumetric wear rate of the probe can be expressed in terms of a percent by dividing $\frac{dW}{dt}$ by the volume of the probe which can be scored by abrasive particles ($\pi r^2 h = \pi(0.125)^2(0.125)$=$0.0061 in^3$). For the 1000 RPM, 5 IPM parameter set, $\frac{dW}{dt}$ becomes $\frac{3.41\%}{min}$. To calculate the percent change in volume per unit distance, multiply this value by the reciprocal of the traverse rate ($\frac{1}{v}$). The result is the percent change in volume per inch of weld ($\frac{0.68\%}{in}$ for 1000 RPM and 5 IPM). Table 28 lists the wear rates predicted by the model for the parameters considered in the chapter V study. Table 29 compares the amount of wear predicted for each case with some values observed in experiment.

Table 28 Volumetric wear rates predicted by the rotating plug model of wear.

Rotation speed (RPM)	Traverse rate (IPM)	$\frac{dW}{dx}$ (%) model
1000	5	0.68
1000	7	0.72
1000	9	0.74
1500	5	0.62
1500	7	0.67
1500	9	0.70
2000	5	0.55
2000	7	0.63
2000	9	0.67

Table 29 Comparison of wear predicted by the model with some experimental values.

Rotation (RPM)	Traverse (IPM)	Distance welded (inches)	Wear (experimental)	Wear (predicted)
1000	5	8	3.69	5.47
1000	5	16	5.91	10.94
1000	5	24	7.55	16.41
1000	7	8	1.24	5.77
1000	7	16	2.75	11.53
1000	7	24	3.79	17.30
2000	5	8	3.54	4.44
2000	5	16	7.81	8.87
2000	5	24	12.86	13.31

Although the wear rates predicted by the model are on the same order of magnitude as experimental values, the predicted wear can differ by as much as three times the observed value. This discrepancy in scale can be corrected by incorporating a scalar multiplier (equal to the ratio of the experimental to predicted wear) in equation 6-30. However, the factor by which the predicted and observed values differ is nonconstant. Using an average scaling factor can correct the predicted values so that they lie closer to the observed wear rates, but the correction is non-uniform since the errors (*predicted value minus actual value*) are not equal for all parameters.

Some of the disagreement between the model and experimental values may be attributable to differences in tool geometry – the probe in the model is a right cylinder, while the tool used to generate the experimental data in Table 29 had a Trivex geometry. The volume of the portion of the tool which can be scored is greater for the cylinder considered in the model – per equation 6-30, this larger volume should translate into a higher wear rate for the cylindrical tool. One implication of this result is that threaded

tools should exhibit a lower wear rate than unthreaded tools, as the latter have slightly more volume. Since the Trivex has a smaller volume which can be abraded, the wear rates associated with the Trivex should be less than those predicted for the smooth cylinder. While this is true, the discrepancy between experimental and predicted values exceeds the variation which can be attributed to differences in tool geometry.

Another red flag is that the numerical model does not reflect the trends indicated by both the statistical model in chapter V and the conceptual model in section 6.3.1, which claim that wear rate increases with rotation speed and decreases with traverse rate. Rather, the numerical model predicts that wear rate will rise with traverse speed and fall with rotation rate, trends which are opposed to those reported in the data from both chapter V and Prado et al. [42-43]. The antithetical trends in the numerical model are much more alarming than order of magnitude discrepancies. While errors of the latter variety can be accounted for with a properly-placed constant, disagreement in the former points to either: 1) errors in experimentation or 2) an incorrect theoretical basis and/or assumptions made in model development (neither of which are easily remedied).

The approach chosen to develop the theoretical wear rate seems sound: the indentation depth x to which the tool can indent the particle is calculated, extrapolated to a swept area, and this area is revolved 360 degrees to yield an expression for the volume carved out by each particle per unit length of translation (dV/dx). This quantity is multiplied by a frequency factor (related to the percentage reinforcement and the particle size) intended to reflect the number of abrasions which can occur over a unit distance. In the interest of simplification, a number of factors which could potentially impact the wear rate have not been considered. For instance, the frequency factor is not a constant, but varies axially

(the shear surface surrounding the pin grows in thickness as we move closer to the shoulder). In equation 6-30, the location Δz at which the minimum clearance between the shear surface and the probe $\delta_0$exceeds the radius of the particle is estimated as half the probe length. While the variation of δ_0 with z remains to be characterized, it follows that the precise axial location at which abrasive action ceases should depend on process parameters: parameter sets associated with thicker shear surfaces should produce smaller value of Δz. This is corroborated by experimental data – for welds with lower rotation speeds, wear is confined to the bottom of the probe, manifesting itself as little more than a slight rounding of the initially sharp probe edges. For welds with higher rotation speeds and/or slower traversal rates, the shear surface is thinner and the extent of wear is greater – radial deterioration is observed at locations z that are some distance from the probe bottom. Observationally, the amount of abrasion dissipates as we move along the probe's vertical axis toward the shoulder. The rate of this dissipation (dW/dz) seems to hinge on the process parameters (which in turn determine the thickness of the shear surface and the degree to which it can function as a buffer against potentially abrasive particles). The trends in wear patterns described here are evident in the photos of the tool probes at various stages of wear which appear in Appendix A. The theoretical model for $\frac{dW}{dx}$ could be improved upon by defining a $\delta(z)$ function to capture axial wear dependence. Our formulation assumed that the wear rate remains constant in the z direction until $\Delta z = \frac{H}{2}$ (where h is the height of the cylindrical pin), at which point $\frac{dW}{dx}$ becomes 0. Since this is not quite the case, this assumption may inhibit the model's predictive capability. If the axial dependence of the plug thickness is taken into account, wear rates should generally be less than those reported in "model" column of Table 2.

6.3.2.1 Compensation for evolution of tool shape with wear

The model does not take into account changes in the volume of the probe which accompany progressive wear. As the distance welded increases, the probe's dimensions are radially and axially reduced as a consequence of abrasion. Prater et al. mechanically gauged these dimensions (the diameter measured at the bottom of the pin and the pin length) and plotted them as a function of weld distance [51]. For all parameters considered in the study, circumferential wear was much greater than axial wear. The tool associated with the case that exhibited the highest wear rate (1350 RPM/4IPM) experienced a 16 percent reduction in radius over 36 inches of weldment, but only a 3.5 percent reduction in pin length over this same distance. For parameters with less wear, the radial material loss, while still comparatively greater, is closer to the length reduction. The ratio of diameter to length reduction for the 1000 RPM/10 IPM case over 36 inches of weldment is 1 (a marked decrease from the 4.6 ratio of diameter to length reported for the 1350 RPM/4 IPM experiments). Based on these results, the tool radius R and the probe length h (Δz is a fraction of this length) are not constants, but functions of weld distance x. The $R(x)$ and $h(x)$ deterioration functions are specific to the process parameters. Changes in the tool dimensions as a function of distance welded can be defined for a specific welding condition from experimental results. Decline should occur more rapidly for process parameters with thinner shear regions (e.g. high rotation rates, slow rates of travel). The data from Prater et al. indicates that for high weld pitches, $R(x) \gg h(x)$. Thus in instances where radial deterioration is much more pronounced than wear which occurs along the probe's vertical axis, $h(x) = h$.

The change in diameter also declines with axial position z (recall that abrasive action is inhibited by the thickening of the shear surface as we move closer to the shoulder). In the truest sense, the diameter of the probe at any time during an MMC weld is a function of both distance welded x and axial position z. The impact of progressive changes in the volume of the probe on the amount of wear can be accounted for by: 1) experimentally determining the deterioration functions $R(x,z)$ and $h(x)$ and 2) substituting these functions for $R$and Δz in equation 23. This method is computationally intensive, requiring multiple experiments and regression analysis to generate deterioration functions which relate probe radius and length to distance welded and axial position. Precise determination of the $R(x,z)$ function requires a method (such as shadowgraph or circumferential profilometry mappings) that enable precise delineation of the vertical axis. However, it may be possible to improve the model's predictive capability without going to the trouble of developing functions which characterize the evolution of tool shape with wear. In reference [51], Prater et al. used calipers to periodically gauge dimensional changes in the tool with wear. This work reports that for friction stir welding of an Aluminum MMC with 17.5 percent SiC reinforcement, the length of the probe is reduced by anywhere from 0 percent to 3.5 percent (a rate of 0.1% per linear inch of weld), depending on the process parameters and distance welded (this data is plotted in Figure 81). The circumferential loss shown in Figure 82 (also taken from Prater et al.) is much more dramatic: the cumulative deterioration in probe diameter is estimated at 16% (a rate of 0.44% per linear inch welded) for the 1350 RPM/4 IPM parameter.

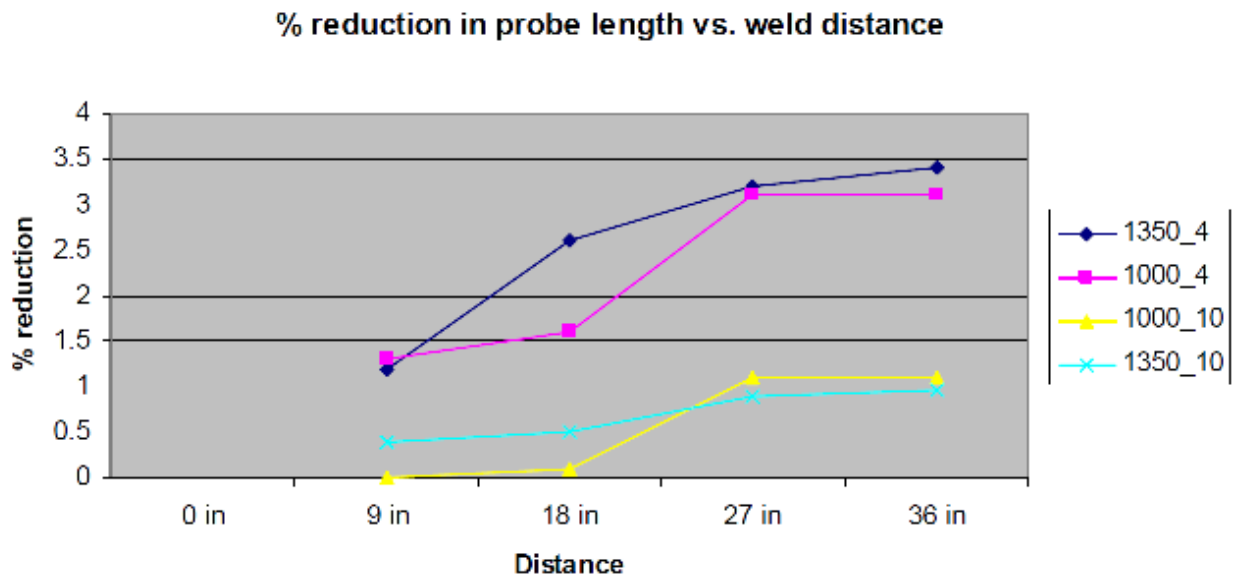


Figure 81 Percent reduction in probe length for friction stir welding of Aluminum MMC with 17.5 percent SiC reinforcement [51].

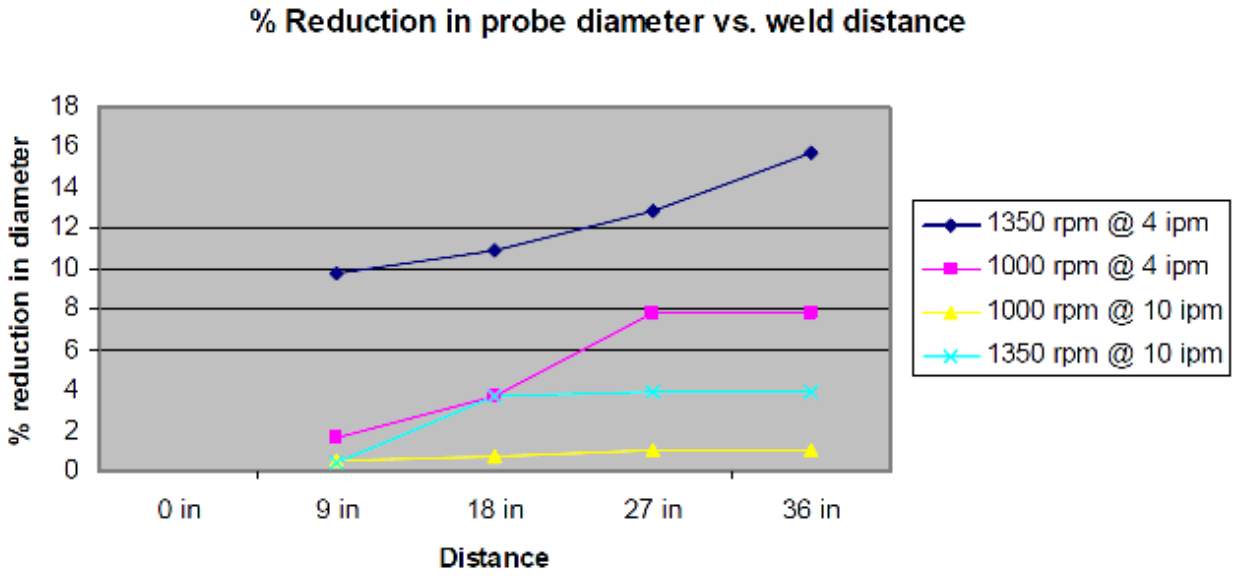


Figure 82 Percent reduction in probe diameter for friction stir welding of Aluminum MMC with 17.5 percent SiC reinforcement [51]

It seems that the data from these plots, when used in conjunction with equation 23, could account for the effect of progressive changes in tool shape on wear. While it is certainly true that the amount of material lost by the tool radially and axially is dependent on process parameters, we can use the average value of the data from Prater et al. to roughly estimate a factor which, when incorporated into the right hand side of equation 6-30 as a multiplier, yields values of wear that are closer to those observed experimentally. For shorter welds ($\leq$ 18 inches) with low weld pitches ($\frac{\omega}{v} \sim 100$), the tool experiences an average length reduction of 0.25% and a 1.75% reduction in diameter. For longer welds (>18 inches) with the same weld pitch, the length is reduced by 1% while the diameter undergoes a 2.5% reduction. The average reductions in length and diameter for shorter welds with high weld pitches are 1.6% and 6.75%, respectively. When the distance welded increases beyond 18 inches, the probe dimensions of length and diameter are reduced by an average of 3.25% and 11.25%. What effect do these dimensional reductions have on the $\frac{dW}{dx}$ (and W) values tabulated in Table 28? Rather than calculate $\frac{dW}{dx}$ for the reduced dimensions associated with each of the four classes in the Prater et al. study (high weld pitch/short weld distance, low weld pitch/short weld distance, high weld pitch/long weld distance, low weld pitch/long weld distance), we instead use the maximum and minimum percent reduction for each dimension to define a range of values which the length and radius may assume during the course of welding MMCs. The lower and upper bound for the reduced length are taken as 0.25% and 3.25% (values which, for a probe that is 0.25" in length, represent losses of 0.000625" and 0.008125", respectively). Reduction in diameter is more substantial, with a lower limit of 0.004" (1.75%) and an upper limit of 0.028" (11.25%). With this compensation for wear, the

probe length ranges from 0.242" to 0.249" and the diameter is estimated to be between 0.222" and 0.246." Table 2 compares the $\frac{dW}{dt}$ (in $\frac{in^3}{min}$) calculated directly from the rotating plug model with the range of values expected for $\frac{dW}{dt}$ when progressive changes in the tool shape with wear are accounted for.

As expected, the adjusted values for $\frac{dW}{dt}$ are smaller than those predicted by the original model. As the volume of the probe decreases, $\frac{dW}{dt}$ (the volume of material removed per unit time) is also reduced. The adjusted values for transient volumetric wear (column 4 in Table 3, intended to estimate the effect of changes in tool dimensions on the wear rate), represent a reduction of 11% (for the smallest change in dimensions) to 72% (for the maximum dimensional deterioration) from the original (unadjusted) values output by the model. The prediction that the wear rate will decrease as the probe is eroded by abrasion is consistent with the asymptotic wear behavior observed in each of the studies by Prater et al. and Prado et al. [42-43,51-52]. For every parameter in these studies, the rate at which material removal occurs decreases with distance welded. For longer weld distances (>2 feet), the reduction is so substantial that it led Prado et al. to suggest a novel means to reduce wear in FSW of MMCs: use a probe whose geometry coincides with the "self-optimized shape" (the shape taken on by the tool when the wear rate has plateaued). While this geometry does contribute to reduced wear, the self-optimized shape is associated with weld defects. The tendency of the self-optimized tool to produce voids was documented by Prater et al. in reference [51].

Table 30 Adjusted wear rates (expressed as $\frac{in^3}{min}$) which take into account dimensional reduction as a result of wear

Rotation speed (RPM)	Traverse rate (IPM)	$\frac{dW}{dt}$ model (unadjusted)	$\frac{dW}{dt}$ model (with dimensional compensation for wear)
1000	5	$2.1x10^{-4}$	$(5.96x10^{-5},1.87x10^{-4})$
1000	7	$3.1x10^{-4}$	$(8.79x10^{-5},2.76x10^{-4})$
1000	9	$4.1x10^{-4}$	$(1.16x10^{-4},3.65x10^{-4})$
1500	5	$1.9x10^{-4}$	$(5.39x10^{-5},1.69x10^{-4})$
1500	7	$2.9x10^{-4}$	$(8.23x10^{-5},2.58x10^{-4})$
1500	9	$3.9x10^{-4}$	$(1.11x10^{-4},3.47x10^{-4})$
2000	5	$1.7x10^{-4}$	$(4.83x10^{-5},1.52x10^{-4})$
2000	7	$2.7x10^{-4}$	$(7.66x10^{-5},2.4x10^{-4})$
2000	9	$3.7x10^{-4}$	$(1.05x10^{-4},3.29x10^{-4})$

The expression $\frac{dW}{dx}$ (where dW is expressed as a percent of the probe volume) is obtained from $\frac{dW}{dt}$ by dividing $\frac{dW}{dt}$ by the product of the probe volume ($\pi R^2 \Delta z$) and the traverse rate v. If we use the original (unworn) dimensions of the tool probe to calculate this volume*, the wear rates predicted by the model decrease to the range of values tabulated in Table 31. The amount of wear extrapolated from these rates aligns more closely with the measured values (Table 5). That the wear rate changes with tool dimensions is not a surprising result, but it is an important one since it confirms the asymptotic behavior of wear observed in previous studies. The linear model constructed in chapter V is lacking in this capacity.

*Using the reduced dimensions to calculate the volume in this context has no effect on the wear rate, $\frac{dW}{dx}$, since the volume expression $\pi R^2 \Delta z$ also appears in the expression for $\frac{dW}{dt}$ ($\frac{dW}{dx} = \frac{dW}{dt} \div \pi R^2 \Delta z$ v). The original dimensions (rather than those adjusted to compensate for wear) are more appropriate for this situation, as the volume in the denominator of the $\frac{dW}{dx}$ equation represents the volume of the probe that is available for scoring.

Overall, the slope of the $W(\nu, \omega)$ curve for a particular value of ℓ is constant over the domain considered in the previous chapter ($1000 \leq \omega \leq 2000, 0 \leq \ell \leq 24, 5 \leq \nu \leq 9$). Cross-validation demonstrated that a linear model, while not perfect, yields reasonably accurate predictions of tool wear based on process parameters. However, if we break down the wear plots constructed from actual data (not a model) into segments, it becomes clear that the wear rate $\frac{dW}{dx}$ and the changes in dimensions $\frac{dR}{dx}$ and $\frac{dl}{dx}$ are greater for segments whose lengths span distances that fall farther along the x-axis. This behavior is explained by equation 6-30: as the dimensions of the probe are reduced as a result of wear, the wear rate decreases because $\frac{dW}{dt}$ is proportional to the probe volume. Since wear rates are initially high and then decrease, the FSW probe in an MMC weld will retain its original dimensions only briefly. Thus the range of values for material removal which take into account dimensional changes are likely to be a better representation of the "true" wear rates than the unadjusted rates output by the original model. This is demonstrated in Table 5, which compares the amount of volumetric wear observed in experiment with the range of wear values predicted by the rotating plug model (with and without dimensional compensation for wear). Each of the data points falls well within or relatively close to the limits of wear output by the rotating plug model when the model is adjusted to account for maximal and minimal dimensional changes.

On the whole, the wear model in its original form tends to overpredict the amount of volume loss, so its predictive capability may be slightly improved by using the midpoint of the range obtained from the model with dimensional compensation (the final column of Table 32) as the wear estimator.

Table 31 Adjusted wear rates (expressed as $\frac{\%volume\ loss}{in}$) which take into account dimensional reduction as a result of wear

Rotation speed (RPM)	Traverse rate (IPM)	$\frac{dW}{dx}$ model (unadjusted)	$\frac{dW}{dx}$ model (with dimensional compensation for wear)
1000	5	0.68	(0.19, 0.61)
1000	7	0.72	(0.20, 0.64)
1000	9	0.74	(0.21, 0.66)
1500	5	0.62	(0.18, 0.55)
1500	7	0.67	(0.19, 0.60)
1500	9	0.70	(0.20, 0.63)
2000	5	0.55	(0.16, 0.49)
2000	7	0.63	(0.18, 0.56)
2000	9	0.67	(0.19, 0.60)

Table 32 Comparison of amount of wear predicted by rotating plug model (with and without dimensional compensation) with experimental values

Rotation (RPM)	Traverse (IPM)	Distance welded (inches)	% Wear (experimental)	% Wear model (unadjusted)	%Wear model (with dimensional compensation)
1000	5	8	3.69	5.47	(1.52,4.88)
1000	5	16	5.91	10.94	(3.04,9.76)
1000	5	24	7.55	16.41	(4.56,14.64)
1000	7	8	1.24	5.77	(1.80,5.12)
1000	7	16	2.75	11.53	(3.60,10.24)
1000	7	24	3.79	17.30	(5.40,15.36)
2000	5	8	3.54	4.44	(1.28,3.92)
2000	5	16	7.81	8.87	(2.56,7.84)
2000	5	24	12.86	13.31	(3.84,11.76)

Though the model is lacking in some respects, one predicted outcome that is consistent with experiment is that parameters with approximately equal weld pitches (the ratio of rotation speed to traverse rate) yield similar values of wear (the effect of weld pitch on wear was previously explored in chapter V). The model developed here has the

potential to predict wear more precisely if we can develop a method to accurately account for changes in shape with time (for instance, development of the aforementioned deterioration functions which express the probe radius R as a function of time and axial position) Such functions would be highly specific to the experimental conditions (such as tool geometry and dimensions, material properties of the tool and workpiece, process parameters, tilt angle, joint configuration, etc.) and, as such, would require extensive data collection. The functions $R(z,t)$ and $h(t)$ can be substituted into equation 23 for the R and Δz terms, resulting in an expression which captures the transient reductions in probe dimensions and its effect on the wear rate. As with the linear discriminant analysis in chapter V, our goal in developing this model is not to characterize wear for every possible scenario involving FSW of MMCs, but rather to provide a general framework upon which a predictive physics-based model for a specific case can be constructed. The work in this chapter demonstrates that this general formulation can be applied to our own data with some degree of success (and thus could potentially be extrapolated to predict wear in other, slightly disparate, settings). The discussions which follow examine the addition of two other factors which may improve upon the existing model: the consideration of temperature effects (6.3.2.2) and the cutting arc (6.3.2.3).

6.3.2.2 Temperature Effects

In the force balance, the shear flow stress terms for the matrix (τ_M) and the tool (τ_P) were assumed to be constant and equal to the values associated with Aluminum and steel, respectively. For steel, $\tau_p \approx 100\ ksi$; for aluminum alloys, $\tau_M \approx 2.5\ ksi$. The

ratio of these terms ($\frac{\tau_M}{\tau_P}$ in equation 23) is assumed to remain constant at 0.025. While the shear stress for the workpiece is assumed to be the same as the Aluminum matrix which comprises 80% of the material, a better approximation could be obtained from the law of mixtures or existing property data on Aluminum MMCs [112] As expected, the inclusion of Silicon Carbide particles improves the temperature resistance of the Aluminum alloy (as with other material properties such as density, the improvement is approximately proportional to the percentage reinforcement). The magnitude of the increase, however, even for MMCs with high reinforcements, is small enough that Al-MMCs can still be considered low flow stress materials (even though they have higher flow stress inclusions). While the slightly increased value of shear stress associated with the MMC (as compared to the unreinforced Aluminum alloy) increases the wear rate values output by the model, it actually reduces the model's predictive capability, as the predicted wear rates calculated using the stress value for the composite are much larger than those observed in experiment. Thus our initial choice to model the workpiece material as a homogeneous Aluminum alloy is a sound assumption and yields estimates of wear which closely mirror empirical data (at least with respect to order of magnitude). The idea that the composite could (in some instances) be accurately represented by its matrix properties is not altogether surprising, as the metal alloy is virtually always the dominant constituent in a reinforced composite (for example, the proportion of Aluminum to Silicon Carbide in an MMC with 20% reinforcement is 4 to 1).

The estimation of flow stress is complicated by the fact that is: a) highly dependent on temperature and b) exhibits a nonlinear response. For a typical Aluminum alloy, a 100 degree increase in temperature corresponds to a 50% reduction in flow stress

[113]. The relationship between flow stress and temperature for commonly used materials (such as steel) are well characterized and can be found in any metals handbook. Data on flow stress in specialty materials such as MMCs, however, is scarce. A research article by Kalaichevi et al. uses neural networks to predict the flow stress for a 6061 Aluminum alloy with 15% Silicon Carbide reinforcement (a material which is very close to the Al 6061/SiC/17.5p material used in the wear study of chapter V) [114]. Kalaichevi et al. claim that the traditional Zener-Holloman formulation, which estimates flow stress based on strain rate, the activation energy (the minimum energy input required for deformation), and material properties (ex. the stress exponent *n*), proves an insufficient estimator for flow stress in metals because it fails to account for the nonlinear behavior of the materials at high temperatures. Discrepancies between experimental flow stress values and those calculated from a regression of the Zener-Holloman equation as great as 30% are common. Kalaichevi's neural network approach, proposed as an alternative to regression modeling methods which assume a linear relationship between flow stress and temperature, is a highly accurate predictor of flow stress for Aluminum alloys undergoing extrusion. Since Friction Stir Welding is also a hot deformation process and the material used to test the Kalaichevi model is virtually the same as that in our experimental work, his reported values for shear stress at temperature may prove useful for our calculations. Similar data and predictions can be extracted from sources such as reference 17 which characterize the flow stress behavior of Metal Matrix Composites. If we know the temperature of the workpiece and tool materials, we can use the existing experimental data/literature on this subject (whether in the form of a table or a neural network prediction) to estimate the flow stress.

Although the FSW process is steady-state (i.e. the temperature fields for the tool and workpiece stabilize after the initial dwell period), the variation in weld temperature between parameter sets is substantial enough to have a significant impact on flow stresses (and thus influence the wear rates predicted using equation 6-30). To account for the effect of temperature on wear, we need a way to estimate the weld temperature at steady-state. Once this temperature is known, the flow stresses for the metal matrix composite and the steel tool which correspond to this temperature can be substituted into equation 6-30. Essentially, extracting a wear rate from the rotating plug model for FSW of MMCs with thermal considerations consists of 4 steps:

1) estimate weld temperature for process parameters and tool geometry
2) determine flow stresses which correspond to these temperatures
3) calculate flow stress ratio
4) calculate wear rate from equation 23

6.3.2.2.1 Estimation of shear stress ratio at temperature using the Schmidt model and steady state conduction

There is an existing formulation which can be used to estimate steady-state weld temperatures for the tool and workpiece [100]. Schmidt's analytical model of heat generation in friction stir welding, a landmark and often-cited paper in the field of FSW, provides a concise and simple method to calculate the amount of heat generated at the tool/material interface. Schmidt defines three state variables which describe the three possible contact conditions in FSW. $\delta = 1$ is the full sticking condition, where the velocity of the matrix material precisely coincides with that of the tool's traversal velocity. In partial sticking/sliding ($0 < \delta < 1$), the velocity of the tool exceeds that of

the matrix. For the sliding condition ($\delta = 0$), the velocity of the tool is much greater than the matrix, effectively rendering the surrounding material stationary [100]. In the Nunes model, the shearing does not occur at the contact interface, but instead takes place in a layer that is separate from (but close to) the contact interface known as the shear zone [101]. This supposition of the Nunes model corresponds to the sticking condition ($\delta = 0$). A detailed explanation of the derivation behind the Schmidt model for heat generation can be found in reference [100]. Essentially, Schmidt proposes that heating in FSW originates from three sources: the probe shoulder (Q_1), the probe sides (Q_2), and the probe tip (Q_3). The sum of these contributions yields an expression for the total heat Q. For a tool with a flat shoulder, the amount of total heat generated at the tool/material interface is given by:

$$Q_{total} = \frac{2}{3}\pi\tau_{contact}\omega({R_{shoulder}}^3 + 3{R_{probe}}^2 H_{probe}) \qquad \text{(equation 6-31)}$$

For a probe with a shoulder diameter of 0.75" (0.009525 m), a probe diameter of 0.25" (0.003175 m), and a probe length (height) of 0.25" (0.00635 m), Q_{total} reduces to

$Q = 2.198\ x\ 10^{-6}\ m^3(\tau_{contact}\omega)$. Unfortunately, the shear stress term for the workpiece material, one of two values we are trying to estimate, appears in the equation for heat generation. However, this term can be replaced by an empirically derived equation which expresses shear strength for Aluminum 6061 (the matrix alloy) as a function of temperature ($\tau \approx \frac{\sigma}{\sqrt{3}}$). The data which serves as the basis for this equation is extracted from the plot in reference [115]. Shear stress as function of temperature for Aluminum 6061 is given by the equation τ =0.75T^3-374T^2-5.8$x10^5T$+4$x10^8$, where T is in degrees Kelvin and τ has units of Pascals ($\frac{N}{m^2}$). The R value associated with the

polynomial regression for this data is 0.975. Initially, the degradation in yield strength with temperature is dramatic, but grows less pronounced as the material heats up, eventually reaching 0 Pa at the material's melting point (855 K). The relationship between shear stress and temperature is approximately linear in the high temperature range (500 K – 866 K). Since peak FSW temperatures generally fall in this range (data from thermocouples reported in the literature place weld temperatures at 70 to 90 percent of the melting temperature), the τ term can be linearized as $\tau_M = -0.0157T + 13.4$ (T is in Kelvins and τ is in ksi).

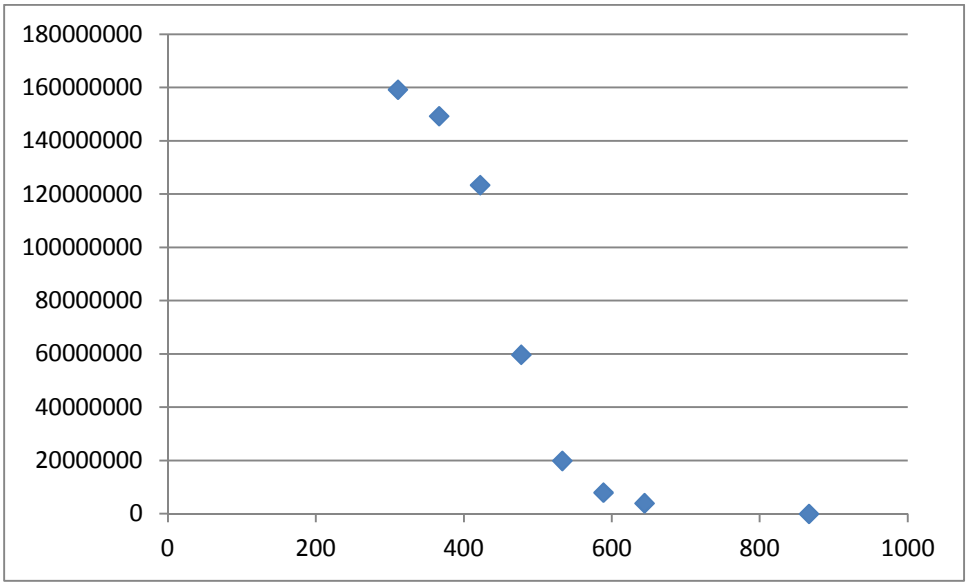


Figure 83 Plot of shear stress versus temperature for Aluminum 6061 (adapted from reference 115)

The variable of interest in our application of the Schmidt formulation is the temperature term, since it can be used to estimate the shear stress values needed for the wear equation in (6-30). Replacing the left hand side of equation (6-31) with a known

heat flux formulated based on Fourier's law of steady state conduction results in an equation which can be solved for T for any rotation speed ω. The tool/material interface can be modeled as a disk with diameter D (0.75"/0.01905 m) adjacent to a semi-infinite medium (the Aluminum workpiece) with temperature T_2 and thermal conductivity k (Figure 84). Conduction is assumed to be steady-state and two-dimensional (heat transfer occurs in the x-z plane). The system could also be modeled as a vertical cylinder in a semi-infinite medium, but application of the associated equation is restricted to cases where length of the cylinder is at least an order of magnitude larger than its diameter (a condition which is not satisfied for FSW). Another alternative is to represent the system as a circular cylinder with diameter D and height L centered in a solid with width W – for this case, w must be greater than D and L must be greater than W(the former criterion is satisfied, but the latter is typically unmet for FSW). While the disk on a semi-infinite medium is far from a perfect representation of heat transfer in FSW, it can be applied without restrictions. Additionally, the temperature at of tool shoulder (T_1) will be reasonably close to the temperature at the tool/material interface which appears in the right hand side of equation (6-31). The temperature, T_2, of the workpiece material at some distance from the tool should roughly coincide with the temperature of the surroundings (the ambient temperature for FSW is taken as 293 K – room temperature). It is reasonable to assume that the workpiece material retains significantly less of the generated heat than the tool. Anyone who has ever watched or performed a friction stir weld quickly realizes that the tool in FSW functions as a heat sink. Immediately after a weld, the tool is too hot to touch, yet the workpiece material remains comparatively cool. (For very high rotation speeds, a secondary cooling mechanism, such a fan or water

immersion, may be necessary to reduce the time between welds and protect the electronics in the dynamometer from overheating). That material outside the vicinity of the tool remains unaffected by heating is corroborated by the results of thermocouple and tensile test data as well as microscopy (it is evident from transverse FSWed cross sections that the material outside the weld zone is virtually unaltered by the process and shows no signs of undergoing a heating/cooling cycle).

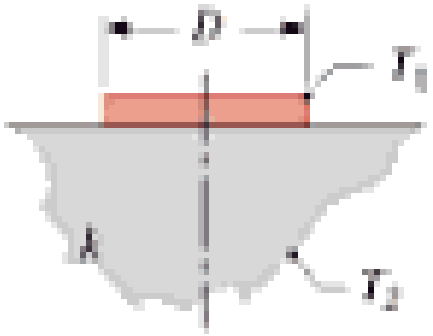


Figure 84 Heat transfer schematic for two dimensional, steady-state conduction of disk at temperature $T_1$to a semi-infinite medium with thermal conductivity k and temperature T_2 [116].

In its most general form, Fourier's law for steady-state, two-dimensional conduction is expressed as $Q = Sk(T_1 - T_2)$. S is a shape factor specific to system geometry. For the case in Figure 9 (a disk on a semi-infinite medium), the shape factor is twice the diameter of the shoulder ($2D = 1.5'' = 0.0381$ m). The thermal conductivity k of any material varies with temperature. However, unlike the highly variable thermal conductivity associated with solids such as steel, the k value for Aluminum alloys remains relatively constant over the 200 K-600 K range. The conductivity of the

Aluminum alloy can thus be represented as a constant ($k \approx 200 \frac{W-m}{K}$). The temperature of the surrounding workpiece material (T_2) is estimated as the ambient temperature (293 K). Equating the two-dimensional steady state conduction equation with the Schmidt formulation (equation 6-31), the resulting equation can be solved for weld temperature T at any rotation rate (ω should be expressed in Hertz to ensure dimensional consistency). The predicted weld temperatures associated with rotation rates ranging from 250 RPM to 1000 RPM are tabulated in Table 33. The shear stress of the matrix material at temperature is calculated from the regression equation. Since tensile test data for O1 tool steels is rare, the shear stress of the tool must be estimated using a linear approximation. According to the material property data for O1 steel in reference [117], the yield strength for the O1 steel hardened to 50 HRC at room temperature is 196 *ksi* (which, by the $\tau \approx \frac{\sigma}{\sqrt{3}}$ approximation, corresponds to a shear stress of 113 *ksi*). Assuming that the yield strength of the material is 0 at its melting point (1810 K) and decays linearly, shear stress (in *ksi)* as a function of temperature (in degrees Kelvin) for the tool pin can be estimated using equation (6-32).

$$\tau = -0.07T + 135.5 \qquad \text{(equation 6-32)}$$

Table 33 Rotation speed, predicted weld temperature, and corresponding shear stress values for workpiece and tool.

Rotation speed (RPM)	Predicted Weld Temperature (T)	Shear stress of matrix τ_M (ksi)	Shear stress of tool τ_p (ksi)	Shear stress ratio
250	416 K	14.8	104.3	0.14
500	465 K	10.4	100.7	0.10
750	492 K	8.0	98.7	0.081
1000	509 K	6.5	97.4	0.067
1250	521 K	5.5	96.5	0.057
1500	529 K	4.8	95.9	0.050
1750	536 K	4.2	95.4	0.044
2000	542 K	3.7	94.9	0.039

The temperature estimates and corresponding stress values calculated based on the Schmidt formulation in conjunction with the classical model of 2-D heating for a cylinder on a flat surface are weak at best. The temperature estimates reported for the range of rotation speeds in Table 85 severely underestimate the actual working temperatures encountered in friction stir welding. For conventional FSW, temperatures generally lie between 75 to 90 percent of the melting point of the workpiece (for an Al 6061 alloy, this would place steady-state temperatures in the range of 650 to 780 K, depending on process parameters). Our model actually cannot predict temperatures in this region, since the regression equation used to estimate shear stress in Al 6061 is negative for this range. Thus the initial estimates for shear stress values used in section 6.3.2 are probably more representative of the materials' behavior at temperature than the information extracted from the preceding thermal model. This is because the temperatures in Table 6 are smaller (and the shear stress ratios are larger) than those we

would anticipate for FSW based on the empirical thermal data published in the literature. Because the shear stress ratios are higher, the thermal model leads to significant overprediction of wear. For instance, the shear stress ratio for 2000 RPM in Table 6 (0.04) suggests that the tool probe will wear at a rate of 5% per inch (a 40% volume loss over the course of an 8 inch weld), a value which is not realized in an experimental setting. The largest value observed in the wear experiments of chapter V is approximately 20% over a 24 inch weldment.

The Schmidt model is sound and has been used as the basis of numerous methods in the literature to estimate weld temperature. Schmidt's formula (equation 6-31) is used frequently as a constituent equation for CFD thermal models, which rely upon it to calculate the power input of the tool to the material [107]. The flaw in our temperature estimates then, must lie in our approach to estimating heat transfer using Fourier's law and equating it to the more rigorous estimation technique developed by Schmidt. The 2-D steady-state conduction case was chosen for its computational simplicity (and because it can be applied without restrictions), but the geometry for this formulation is not truly representative of heating in FSW. Although the FSW tool shoulder is cylindrical and rests on the surface of the material as depicted in Figure 84, the tool is not a stationary heat source, but an active source of heat generation as it rotates and simultaneously traverses through the workpiece. The 2-D steady-state conduction model fails to account for the impact of traversal and rotation on heating as well as the contributions from the tool pin.*

*It should be noted that the Schmidt model also fails to consider the effect of traverse rate on heat generation. Although spindle speed is the dominant factor in determining heat input, the temperature at the tool/material interface should be inversely proportional to traverse rate. For the same rotation speed, a tool with a comparatively faster traverse rate should generate less heat than a tool moving forward with a slower velocity. The Schmidt model does not account for heating effects originating from process variables other than spindle rotation and tool geometry.

In summary, the Schmidt analysis and the steady state 2-D conduction model for a cylinder on a flat plate are different models – while each has its own set of strengths and weaknesses, neither is a catch-all representation of heating phenomena in FSW. While equating the Q values obtained from each model is convenient since it yields a closed-form solution for temperature, the assumption of equivalency misrepresents the process and as such, leads to a severe underestimation of weld temperature for FSW of Al 6061 (we would expect temperatures in FSW of an Al-MMC to be higher since MMCs have a comparatively greater thermal resistance).

6.2.2.2 Estimation of shear stress ratio at temperature from Raghulapadu et al. thermal model

A better estimate of temperature (and the shear stress ratio) can be obtained from Raghulapadu et al.'s model of heat transfer in friction stir welding, a formulation which is perhaps more appropriate for our application since it was specifically developed based on the rotating plug model [118]. Raghulapadu et al. obtain an approximate analytic estimate of maximum weld temperature from the equation for heat conduction in three dimensions (with radial, transverse, and axial heating components). Like the Schmidt model, the Raghulupada formulation is a "stick" model which assumes that the portion of the workpiece material closest to the tool surface adheres to it. Overall, it seems that stick models are superior (at least in terms of predictive capability) to slip/sliding models. Details of the derivation for the Raghulapadu model can be found in reference [118].

This temperature estimation technique relies on a number of dimensionless quantities (Peclet number P, Biot number B, and mechanical dissipation ε) to account for the various thermal effects at play in FSW. These variables are summarized in Table 34. The Peclet number P captures the effect of the traverse rate on heating: it is the ratio of the product of the traverse speed and the pin radius to the thermal diffusivity (α) of the workpiece (for Al 6061, $\alpha \approx 6.2 \frac{in^2}{min}$). There are two Biot numbers of interest, one for the surface of the workpiece (B_H) and another for the bottom (B_O). h_H and h_0 represent the heat transfer coefficients for the top and bottom of the workpiece, respectively. The Biot numbers account for convection of heat from the both sides of the workpiece to the surrounding air – as such, h_H should be greater than h_0 since the region associated with the former term is closer to the heat source. The mechanical dissipation term ε includes the thermal conductivity of the workpiece k, the melting temperature of the material T_M, the pin radius, the angular velocity of the tool, and the shear stress of the workpiece at room temperature $\tau_{0,0}$. The expression for dissipation also includes a "correction factor" f which depends on the geometry of the tool. f accounts for discrepancies between the surface area of the tool and the rotating plug (as illustrated in Figure 78, the cross-sections of the tool and the plug in the x-y plane are not coincident). The correction factor is calculated from the equation (6-33). f is highly specific to the welding conditions – equation (6-33) estimates f for a full-penetration weld (wherein the penetration depth of the tool is equal to the thickness of the workpiece).

$$f = 1 + \frac{\left(\frac{R}{H}\right)\left(\frac{R_S}{R}\right)^3 - 1}{3} \quad \text{(equation 6-33)}$$

The approximate closed-form solution for the differential equation governing heat diffusion in FSW (as derived by Raghlapadu et al.) reduces to a Bessel function. The small-argument form of this Bessel function, combined with the expression for axisymmetric steady state heat conduction in a cylindrical shell with inner radius *R* and outer radius $\frac{2}{P}$, yields equation (6-34). This expression can be used to estimate the temperature near the pin. The Raghlapadu et al. thermal model was used to predict the temperatures for the parameters in the chapter V wear study. The intermediate values used for Raghlapadu et al. temperature estimation technique, along with the final predicted temperature value, appear in Table 35. The correction factor *f* for the configuration is 5.33.

$$T = T_\infty + \frac{(T_M - T_\infty)\ln\left(\frac{2}{P}\right)}{\varepsilon + \ln\left(\frac{2}{P}\right)} \qquad \text{(equation 6-34)}$$

Table 34 Dimensionless numbers in Raghulapadu et al. thermal model

Peclet number	P	$\frac{VR}{\alpha}$
Biot number (workpiece surface)	B_H	$\frac{h_H H}{k}$
Biot number (bottom of workpiece)	B_O	$\frac{h_o H}{k}$
Mechanical dissipation	ε	$\frac{kT_M}{f\tau_{0,0}\omega R^2}$

Table 35 Temperature estimates obtained from Raghlapadu et al. thermal model

Traverse rate (in/min)	Rotation speed (RPM)	P	ε	T (K)	τ_M(ksi)	τ_P(ksi)	$\frac{\tau_M}{\tau_P}$
5	1000	0.110	2.74	591	4.12	91.18	0.045
5	1500	0.110	1.83	647	3.24	86.96	0.037
5	2000	0.110	1.37	684	2.66	84.17	0.032
7	1000	0.154	2.74	574	4.40	92.49	0.048
7	1500	0.154	1.83	630	3.50	88.22	0.040
7	2000	0.154	1.37	669	2.90	85.34	0.034
9	1000	0.198	2.74	559	4.62	93.58	0.049
9	1500	0.198	1.83	616	3.73	89.29	0.042
9	2000	0.198	1.37	655	3.11	86.35	0.036

The three rightmost columns of Table 34 list the corresponding shear stress values (τ_M and τ_P) which correspond to the temperatures calculated using the Raghlapadu model. The $\tau_M$values in the look-up table are quelled from the linearization of the shear stress for the matrix ($\tau_M = -0.0157T + 13.4$) – this expression was chosen over the regression equation because the temperatures predicted by the Raghlapadu model lie close to the range where the shear stress values predicted by the polynomial function are negative. The shear stress of the pin (made of O1 steel) is also calculated from a straight-line idealization: $\tau_P = -0.075T + 135.5$.

Compared with the Schmidt formulation, the temperature predictions output by the Raghlapadu et al. model are preferable since they correspond to values of shear stress which produce wear values on the same order of magnitude as those observed experimentally. The difference in the predicted and experimental wear values, however, is actually much greater when thermal effects are taken into account. The overprediction is a consequence of our assumption that the material behavior of an Al-MMC will closely

resemble that of its unreinforced counterpart. In reality, the presence of the SiC reinforcement renders the MMC significantly harder than its base alloy. It is thus expected that weld temperatures associated with an MMC will be hotter than those observed in FSW of conventional aluminum alloys. This is certainly the case for FSW of steels, where temperatures are great enough to induce melt wear of the pin tool. Weld temperatures for the MMC are expected to be higher than those predicted in Table 6 (from the Schmidt et al. model) or Table 8 (from the Raghlapadu et al. model). Elevated temperatures (in the 700-800 K range) correlate with smaller shear stress ratios which correspond to wear values more closely aligned with those recorded in experiments. These values of shear stress are more closely aligned with our initial estimate of τ for the tool at temperature (2.5 ksi).

It is perhaps ironic that our initial estimate of the shear stress ratio appears to be more representative of the behavior of the system at temperature than the thermal models we have taken great pains to derive here. This suggests that temperature effects are not as important as initially theorized: their inclusion adds a level of complication which ultimately does not enhance the predictive capability of the model. A legitimate criticism of the model is that it places too much emphasis on thermal effects. Because the shear stress ratio term in equation 23 is cubed (and the shear stress values which comprise it are highly sensitive to temperature), any fluctuation in temperature will translate to a substantial change in the predicted wear rate. A cursory evaluation of the impact of temperature on wear in chapter V revealed that the correlation between the two variables appeared to be weak (temperature, a function of rotation speed and traverse rate, was discarded as a potential latent variable by virtue of this analysis). This behavior is

consistent with abrasive wear, a mechanism which is largely insensitive to temperature. The formulation in equation 6-30 is more indicative of melt wear, where a tool becomes more susceptible to wear as a consequence of material softening. This is not to say that some melt wear does not occur in friction stir welding (the yield strength of the tool will decrease with tool temperature, enabling abrasive particles to more easily indent its surface), but it is in all likelihood not the primary wear mechanism. The next section adjusts the model based on the interaction between the rotating plug and wear particles. The inclusion of process dynamics (specifically, the degree to which the plug acts as a "buffer zone" in which the particle can be swept up prior to indenting the tool surface) is an attempt to make the model more reflective of the system behavior recorded in multiple references.

6.3.2.3 Cutting Arc Compensation

Our original model is satisfactory in that it a) produces estimates that are on the same order of magnitude as experimental data and b) explains the asymptotic wear behavior observed in the experiments of Prater et al. and Prado et al [42-43,51-52. The major flaw in the model is that it fails to reproduce (and explain) the trends documented in the aforementioned references. For instance, the experiments of Prater et.al. and Prado et al. indicate that tool wear in FSW of MMCs decreases with traverse rate and increases with rotation speed, trends which are opposed to those predicted by equation (6-30). Sections 6.3.2.1 and 6.3.2.2 investigated whether the failure of the model to predict trends was a consequence of factors which were disregarded in its initial derivation. The

potential effect of in-process changes in the shape of the tool on wear rate was assessed in section 6.3.2.1. Though the compensation for changes in length and diameter of the tool as a result of wear was able to explain the asymptotic behavior observed in the wear process, the predicted variation of the wear rate with process parameters was unaffected. This was also the case for the thermal considerations added to the model in section 6.3.2.2. The discrepancy between predicted and experimental trends are exacerbated, rather than aligned, by the inclusion of temperature effects.*

The major problem with the model lies in its inability to mirror experimental trends, a deficiency that is likely attributable to its basis in indentation theory. As mentioned in the section on temperature, any indentation model of wear will be strongly temperature dependent since it features a yield strength term (in our model, the yield strength term is converted to shear stress using the approximation $\tau \approx \frac{\sigma_y}{\sqrt{e}}$). Indentation is very sensitive to temperature: as the tool material is heated, it becomes softer as a consequence of the reduction in yield strength, making it easier for particles to indent the tool (as well as increasing the indentation depth). In this scenario, melt wear (wear facilitated by softening of the material at elevated temperatures) may be the primary mechanism. Wear in FSW of MMCs, however, is different from the conventional indentation model. Although a portion of the wear incurred in this process can indelibly be attributed to tool softening/melt wear (just as the tool experiences some small, possibly negligible, amount of adhesive wear whenever it is in contact with the workpiece material), the thermal component of wear is outweighed by abrasion.

*As discussed previously, the model appears to be much more sensitive to temperature than the actual system.

The trends in the observational data point to abrasion (i.e. the marked decrease in wear rate with distance and the dependence of wear on the time term $\frac{\omega}{v}$), and the microscope images presented in chapter VII (which clearly show parallel grooves on the surface of worn tool specimens) confirm its role as the dominant mechanism.

Abrasive wear in FSW of MMCs is complicated by the presence of the rotating plug, the plasticized layer of material which surrounds the tool during welding. As discussed previously, in order to indent the tool, a particle must be of sufficient size (D) to span the width of this region (δ). Hence abrasion is only possible when $\delta \leq \frac{D}{2}$. This unique aspect of the process dynamics (which explains the nonintuitive variation of wear with the process parameters ω and ν) is not accounted for in the indentation model. The model stands to be improved by the addition of a variable which accounts for the degree to which the plug can act as a "buffer zone", preventing particles from contacting the tool when the $\delta > \frac{D}{2}$ condition is satisfied. We propose either 1) the inclusion of a corrective factor, known as the cutting arc compensation factor f_c, to equation 6-30 or 2) a new formulation to predict wear rates based on the width of the cutting arc. The f_c metric would be analogous to the forcing factor/feedback mechanisms used in many CFD and analytic models. Forcing factors augment the predicative capability of a model by ensuring that it reflects physical observations. In this case, f_c would ensure that kinematic effects (the width of the rotating plug) supersede thermal effects. This is not reflected in equation (23), where the highly-temperature dependent shear stress terms have the greatest influence on the outcome.

An estimate of the volumetric wear rate of the tool for a particular parameter set can also be calculated based on the span of the cutting arc. The cutting arc is defined as the region where abrasion can occur (it is bounded by the angles θ_1 and θ_2 and exists only where $\delta \leq \frac{D}{2}$). The width of the cutting arc, like the rotating plug, varies with axial position along the tool length. The arc will be widest at z=0 (the bottom of the tool pin), and deteriorates with increasing z until the plug becomes wide enough to preclude abrasion at all locations around the tool circumference. We approximate that this condition is satisfied at $z = \frac{H}{2}$ (halfway along the tool length). The cutting arcs ($\theta_1 \leq \theta \leq \theta_2$) for several combinations of rotation and traverse rates are tabulated in Table 36. The cutting arcs calculated in this table correspond to the axial location $z = 0$ (where δ_o, the minimum clearance between the rotating plug and the tool, is 0 at θ=-$\frac{\pi}{2}$).

Table 36 Parameter set and corresponding cutting arc ΔC

Rotation rate	Traverse speed	θ_1 (degrees)	θ_2 (degrees)	ΔC (degrees)
1000	5	257.2	282.8	25.7
1000	7	259.2	280.8	21.7
1000	9	260.4	279.6	19.1
1500	5	254.3	285.7	31.5
1500	7	256.7	283.3	26.6
1500	9	258.3	281.7	23.4
2000	5	251.8	288.2	36.4
2000	7	254.6	285.4	30.7
2000	9	256.5	283.5	27.1

It is evident from Table 35 that the cutting arc is proportional to $\frac{\nu}{\omega}$. When the traverse rate increases *or* the rotation speed decreases, the angular range for which $\delta \leq \frac{D}{2}$ decreases (and the region for which abrasion is possible shrinks). The opposite occurs when the traverse rate is slowed *or* the rotation rate becomes more rapid – for these cases, the region susceptible to abrasion becomes larger as the angular range for which $\delta \leq \frac{D}{2}$ grows. The trends predicted by the cutting arc calculations are consistent with those observed experimentally and reflect the mechanistic effect of the rotating plug on wear. The question now becomes how best to incorporate the cutting arc calculation into our original wear model as a correction factor.

The cutting arc can be expressed as a proportion of the entire tool circumference. The width of the cutting arc, however, is dependent on axial position z. ΔC is at a maximum when $z = 0$ and shrinks to 0 at approximately $z = \frac{H}{2}$. The ΔC values in Table 35 are only valid at the bottom of the tool probe ($z = 0$) and thus cannot be applied directly to equation 23 since it expresses the wear rate as a percentage of the probe volume. The decay of the cutting arc can be approximated as the linear function in equation (6-36), and the total cutting arc (in degrees) ΔC_V over the tool volume can be calculated from the integral in equation (6-35). ΔC_V is divided by 180 to determine the proportion of the tool's surface area which can potentially contact an abrasive particle in a single revolution. The surface area of the tool ($2\pi rh$) multiplied by this proportion yields the maximum area that can be scored by the abrasive. If the average depth to which a particle indents the tool surface is taken as half its diameter, then the maximum volume removed by particles per revolution of the tool can be expressed as πRHD. This

volume divided by the original probe volume ($\pi R^2 H$) to obtain the percent volume loss associated with a revolution, a value which is in turn converted to percent volume loss per inch. This series of calculations is summarized in Table 37. $\frac{dV\%}{dx}$, the wear rate of the tool expressed as a percent of its initial volume per unit length, appears in the final column.

$$\frac{\Delta C}{360} = \frac{\theta_2 - \theta_1}{360} \qquad \text{(equation 6-35)}$$

$$\Delta C(z) = \frac{-2z\Delta C_{max}}{H} + \Delta C_{max} \qquad \text{(equation 6-36)}$$

$$\Delta C_V = \int_0^{\frac{H}{2}} z\left(\frac{-2z\Delta C_{max}}{H} + \Delta C_{max}\right)dz \qquad \text{(equation 6-37)}$$

Table 37 Estimating $\frac{dV\%}{dx}$ based on the cutting arc ΔC

Rotation rate	Traverse speed	ΔC	% volume scored in single revolution*	$\frac{\%dV}{dx}$
1000	5	25.7	0.0017	0.36
1000	7	21.7	0.0015	0.22
1000	9	19.1	0.0013	0.15
1500	5	31.5	0.0021	0.66
1500	7	26.6	0.0018	0.40
1500	9	23.4	0.0016	0.27
2000	5	36.4	0.0025	1.01
2000	7	30.7	0.0021	0.61
2000	9	27.1	0.0018	0.41

*assuming that indentation depth is equal to the particle radius

The calculations based on the cutting arc more closely resemble the empirical data than the rates calculated using the original indentation model. Because thermal effects are not emphasized, the predictions of the cutting arc approximation reflect the trends documented in references 42-43 and 51-52. Table 38 compares experimental data with predicted wear rates for both the Nunes indentation model and the cutting arc approximation. While the difference between observed wear and that obtained from Nunes' indentation technique may be less than the discrepancy between observed and predicted wear calculated from the cutting arc method for some cases, the trends predicted by the Nunes' model are opposed to those documented in the aforementioned references. Based on this assessment, the cutting arc method is preferred because it is better able to capture the process dynamics. We thus recommend that the cutting arc approximation be used to predict wear rates in FSW of MMCs. Two assumptions which belie the indentation model are called into question:

1) There are no circumferential restrictions on where a particle can score the tool *or* that such restrictions only apply for very slow
2) The wear rate is strongly dependent on shear stress (and thus closely correlated with temperature)

There is, however, one important aspect that is shared by both models, as the dependence of wear on particle size is an important component of both formations. An experimental investigation of the relationship between wear and particle size is the subject of the next chapter. The results of these experiments provide additional key insights into the modeling of wear phenomena in FSW.

Table 38 Comparison of best estimates obtained from the Nunes model and the cutting arc approximation with observed wear

Rotation (RPM)	Traverse (IPM)	Distance welded (inches)	Wear (experimental)	Wear (Nunes model)	Wear (cutting arc approximation)
1000	5	8	3.69	5.47	2.85
1000	5	16	5.91	10.94	5.71
1000	5	24	7.55	16.41	8.56
1000	7	8	1.24	5.77	1.72
1000	7	16	2.75	11.53	3.44
1000	7	24	3.79	17.30	5.16

CHAPTER VII

AN EXPERIMENTAL EXPLORATION OF THE ROTATING PLUG MODEL FOR TOOL WEAR IN FRICTION STIR WELDING OF MMCS

7.1 The Rotating Plug Model for Wear

Chapter VI focused on the development of a predictive process model for tool wear in friction stir welding (FSW) of Metal Matrix Composites (MMCs). This model ultimately seeks to predict the amount of volumetric wear the probe experiences during the course of a weld, given the process parameters (rotation rate ω and traverse speed ν), the initial volume of the tool probe, and the length of the joint. The formulation is based on Nunes's rotating plug model for FSW of Aluminum alloys – as indicated in the preceding chapter, this model (which works well in other contexts) can be extended to FSW of metal composites with some modifications. The magnitude of the wear rates output by the rotating plug model match those recorded in experiments. The estimates obtained using the approximation in section 6. 5 are also on the same order of magnitude as empirical values, but they also accurately reflect experimental trends (as wear is predicted to increase with rotation rate and linear distance welded but decrease with traversal speed). Both the original rotating plug model of wear and the ancillary cutting arc approximation offer tremendous insight into process dynamics, potentially explaining the physical phenomena which underlie the strong relationship between wear and weld pitch documented in chapter V.

The wear rate equation derived based on the rotating plug model (equation 7-1) expresses the amount of wear per unit time as a function of volume fraction n, particle radius r, process parameters (ω and ν), tool geometry (shoulder radius R and pin length H), and the flow stress of the workpiece and tool (τ_M and τ_p, respectively). $\%dW/dx$ (the percent volume loss per unit length) is obtained by dividing $\frac{dW}{dt}$ by the product of the traverse rate ν with the original volume tool volume ($\pi R^2 H$).

$$\frac{dW}{dt} \sim 4950 n r^2 R^2 \left(\frac{\tau_M}{\tau_p}\right)^3 \Delta z (2 - \frac{r\omega}{\nu})\nu \qquad \text{(equation 7-1)}$$

When this model failed to predict the trends observed in experiment, an alternative, but closely related, formulation based on the cutting arc was introduced (section 6.5). The inclusion of the cutting arc accounts for the ability of the rotating plug to act as a buffer zone – when the distance between the tool surface and the plug exceeds the radius of the abrasive, the particle is is swept up in the surrounding flow of plasticized material and unable to impinge on the tool surface. Abrasion is thus prevented at all locations where $\delta > \frac{D}{2}$. The cutting arc approximation is also preferred because it de-emphasizes the influence of temperature on wear rate, a modification that, while it does not affect the magnitude of the estimate, is able to accurately reflect the empirical relationship between wear and process parameters. The τ terms in equation 7-1 are highly temperature dependent. Although there is certainly a thermal component to wear, experimental results suggest that kinematic effects (i.e. the growth of the rotating plug) dominate. As such, the width of the clearance between the rotating plug and the tool surface (δ) is hypothesized to be the determining factor as to when (or how much) abrasion can occur during the joining process. The model of the system suggests that two

major factors contribute to tool wear: 1) the width of the rotating plug (determined primarily by parameter selection) and 2) particle size. Since (1) was evaluated in the initial wear study in chapter V, we turn our attention in this chapter to evaluating the effect of particle diameter on wear. The results of this analysis will either confirm or dispute the predictions of the rotating plug model of wear.

Some of the key trends forecasted by the model are summarized below (a more complete assessment of the outcomes predicted by the model and the underlying mathematics appear in chapter VI):

1) It is possible to reduce, but not eliminate, contact between particles and the tool surface by controlling the width of the rotating plug δ (δ is the clearance between the boundary of the rotating plug and the perimeter of the tool). This is accomplished through parameter selection, since $\delta = \frac{\nu}{\omega}(1+\sin\theta)$. A wider clearance decreases the size of the region over which scoring can take place, effectively acting as a buffer zone between the tool and the harder particles embedded in the workpiece material. For conditions which correspond to a thicker plug, particles are captured by the plug of plasticized material and transported around the tool without pressing into it. Thus less wear should be observed for parameters which correspond to comparatively larger values of δ. This predicted behavior is consistent with the results of the wear experiments summarized in chapter V, where higher rates of wear were observed for

parameters with larger ratios of $\frac{v}{\omega}$. Parameters associated with greater wear correspond to thinner shear zones.

2) For each parameter set, there exists a critical abrasive particle diameter, D_{cr}, equal to twice the width of the rotating plug ($D_{cr} = \frac{2v}{\omega}(1+sin\theta)$. D_{cr} represents the minimum particle size which is able to score the tool (by spanning the width δ) for a particular value of θ. Based on the variation of δ with angular position, it is apparent that D_{cr} should increase with θ: as the clearance between plug and tool grows wider, abrasive action is impeded unless the particle is large enough to span δ and press into the tool surface. The relationship between wear and particle size is the subject of this chapter.

3) For a specific particle diameter, the region over which scoring can occur is defined by the range of θ values for which the inequality $sin\theta \leq \frac{\omega D}{2v} - 1$ is satisfied. This expression only applies to interactions in the x-y plane at $z = 0$, the axial location coincident with the bottom of the tool probe. The angular range over which abrasion can occur (θ_1, θ_2) is smaller for parameters associated with larger δ values. The relationship between process parameters and the cutting arc was previously explored in chapter VI, where it was incorporated into the rotating plug model to align wear estimates with values and trends

observed in experiments. The expression in equation 7-2 can be used to convert the span of the cutting arc ($\theta_2 - \theta_1$) to an arc length s. In theory, the value of s should correspond to the length of a single abrasive groove at $z = 0$. Grooves should become smaller (both in arc length and depth) as we move upward along the vertical axis of the tool probe. The widening of the plug and corresponding shrinkage of the cutting arc with increasing z combine to reduce the likelihood of abrasion. Reduced contact between the tool and abrasive at these locations should translate into less pronounced (or even a complete absence of) grooving in the upper region of the probe surface, near the tool shoulder (where δ, even at its minimum thickness, should be large enough to preclude interaction between the tool and the much harder inclusions).

$$s = 2\pi R \frac{\theta_2 - \theta_1}{360} \qquad \text{(equation 7-2)}$$

In summary, the rotating plug model of wear predicts that the amount of abrasion is governed by the clearance between the plug boundary and the probe, a quantity defined as δ. While careful selection of process parameters and abrasive particle diameter may prove an effective strategy to limit wear, the model suggests that it is not possible to eliminate wear completely because of the manner in which the plug develops. As set forth in the Nunes model, the distance between the plug and the tool is 0 at $\theta = \frac{-\pi}{2}$ and grows slowly in the region surrounding this location. Thus a particle of any diameter

would be able to score the tool in the vicinity of $\theta = \frac{-\pi}{2}$, regardless of the process parameters selected by the weld operator. An experimental test of the predictions of the model pertaining to particle size are evaluated in sections 7.3 and 7.4. The next section focuses on confirming an underlying, key assumption of the rotating plug model of wear: that wear in FSW of MMCs is primarily a result of abrasion.

7.2 Identifying the Wear Mechanism

In his text Wear of Materials, Rabinowicz writes that wear generally falls into one of four major categories: surface fatigue, adhesion, abrasion, and corrosion [49]. These categories are not mutually exclusive: a system in which wear is observed may exhibit characteristics associated with any or all of the major mechanisms. Though multiple wear mechanisms which act on a system simultaneously is a fairly common occurrence, a dominant mechanism (one that is responsible for the majority of the observed wear) usually emerges during analysis. An addendum to Rabinowicz's classification scheme further classifies wear in any category according to severity. Unfortunatley, the boundaries which distinguish light, moderate, and severe wear are not well-defined (or are defined relative to wear observations specific to a particular system).

The process of identifying the primary wear mechanism in a system is relatively straightforward, despite the potential overlap between wear regimes. Each of the wear processes identified by Rabinowicz (surface fatigue, adhesion, abrasion, and corrosion) has unique characteristics. Hence all that is required to pinpoint a mechanism as the primary source of wear is to determine which of these characteristics the system behavior

is most closely aligned with. Three such characteristics (as identified by Rabinowicz) are the size of the wear fragments, the microscopic appearance of the worn surface, and the rate of wear. Occasionally these characteristics, such as fragment size, are defined relative to other mechanisms. The fragments produced by abrasive and surface fatigue wear are identified as "large" – however, these wear particles can only be deemed "large" when compared with fragments produced by corrosion and adhesion. Asymptotic wear means that the total volume of material loss converges to finite limit as the sliding distance/time of contact increases. This steep dropoff in the wear rate is present only in abrasive systems and corrosion (the latter only when film formation does not occur).

There are few published papers on tool wear in Friction Stir Welding of metal matrix composites. Researchers who address this subject in their work, if they discuss wear mechanisms at all, assume that abrasion is the primary mode of wear. This is a reasonable assumption – after all, abrasive particles are embedded in the workpiece material. The presence of the abrasive is all that distinguishes MMCs from conventional Aluminum alloy: since tool wear does not occur in FSW of the latter material but is a major problem in the former, it makes sense to attribute wear to the ceramic inclusions present in MMCs. However, a tribological evaluation of the worn surface is necessary to confirm abrasion (and rule out other possible wear modes) as the primary cause of wear.

As part of the experimental work documented in this chapter, numerous worn tool specimens were examined using a Scanning Electron Microscope (SEM). The SEM image of the tool surface prior to welding is shown in Figure 85. This image was taken at 100X magnification at an excitation voltage of 5 kV. The surface features visible on the tool probe are residual tool markings from the turning process used to fabricate the tool.

These markings are distinct from those which appear in the post-weld image of the probe in Figure 86. The surface features in this image are grooves (presumably) created when SiC particles contact the probe surface and plough out material as the tool rotates; these circumferential, parallel grooves are the microscopic signature of abrasion. Specifically, the underlying wear mechanism observed for this process is three-body abrasion, as the abrasive grain is sandwiched between two sliding surfaces (the tool and the Aluminum matrix of the tool).

Other aspects of the system that are consistent with abrasive wear include:

1) Rapid wear rate. Abrasion is consistent with the rapid wear rates documented in chapter V. For parameters associated with high wear, the tool probe can lose as much as 1 percent of its volume per unit length of weld. The wear particles/fragments produced must be relatively large since material lost by the probe is visible to the naked eye after only a brief weld period.

2) Asymptotic wear rate. Asymptotic behavior of wear is not readily evident in the wear plots of chapter V, as the distances welded are not long enough to permit a clear emergence of an asymptotic trend. However, the asymptotic behavior becomes apparent when we examine the intermediate wear rates associated with a particular parameter: the wear over the course of the weld distance is not constant, but decreases as length is increased. Evaluating the data from chapter V, we see that in general the wear rate is greatest over the 0 to 8 inch distance , while the smallest change in wear is usually observed over the range 16 to 24 inches (the longest weld length considered in the study).

The extrapolation of these results suggests that the wear rate would eventually decay to zero at some weld distance. Asymptotic wear in FSW of MMCs has been corroborated by previous researchers. Prado et al. found that tool shape stabilizes for weld distances on the order of meters. (a shape they term "self-optimized" since no subsequent changes in volume are observed even as distance is increased).

3) Insensitivity to temperature. Temperature independence (or at least the comparatively weak correlation between wear and temperature) suggests that the wear mechanism is abrasive. Although there is undoubtedly some component of the wear observed in FSW of MMCs which can be attributed to melt wear (wherein the yield strength of the tool is reduced at elevated temperatures, enabling abrasive particles to indent the surface to a greater depth), the analyses and discussions in chapter V and VI indicate that kinematic effects (rather than the thermal environment) have the greatest influence on the amount of wear a tool experiences.

The potential presence of other mechanisms can be ruled out based on microscopy. The rough and grooved tool surface seen in Figure 86 is not consistent with the smooth and rounded surfaces produced by corrosive action. Some rounding is visible at the tip of the probe, but this change in tool shape is almost certainly a result of material removal by abrasive particles which contact the tool at this location rather than corrosion. There is additionally no evidence of wear originating from surface fatigue, as this variety

of wear is usually commensurate with failure (i.e. breakage of the tool during a machining process). Even in its most severe state of wear (the "self-optimized" shape identified by Prado et al.), the FSW tool can still be used to join material. Though the reduction in vertical material flow which accompanies the self-optimized shape may contribute to wormhole formation, the tool in this state is not characterized as a failed component since it can still produce welds. Surface fatigue wear would manifest itself as a fissure in the tool surface which ultimately leads to shearing of the tool probe along the fracture line. There is no evidence of fatigue wear for either the steel tools used in the chapter V wear study or the tracer experiments detailed in this chapter.

Unlike surface fatigue, abrasion, or corrosion, adhesive wear is present in any system where materials are in direct contact. While adhesive wear is unquestionably present in this system, the direction of material transfer is from the Aluminum matrix of the workpiece to the tool surface, and, as such, is not considered a contributing factor to the wear of the tool. The conglomerations on the tool surface in Figure 86 may be miniscule clumps of Aluminum which persistently adhere to the tool surface, even after etching. The wear rate reported for FSW tools used to join MMCs far exceeds that associated with adhesive wear. Taken together, these observations suggest that adhesive wear is not a contributing mechanism to tool wear in this system.

The results and arguments presented in this section pinpoint abrasion as the primary cause of tool wear in FSW of MMCs, confirming the abrasive wear assumption made by previous researchers. Existing mathematical models of abrasive wear processes are usually based on sliding wear tests -- where a ball is repeatedly slid along a surface in a motion similar to polishing. While sliding wear models prove insufficient for wear

prediction in the context of FSW, the variables which appear in them (particle size, sliding speed, sliding distance, specimen volume, etc.) are nearly identical to those considered in the rotating plug model for wear developed in chapter VI. The next sections test whether a central hypothesis of abrasive wear models, that volume lost by the tool is proportional to the size of the abrasive particles, holds true for FSW of MMCs.

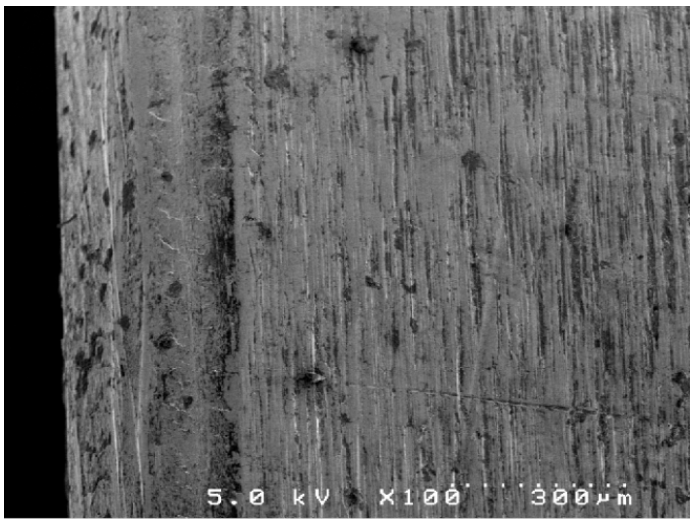


Figure 85 SEM image of tip of tool probe prior to welding.

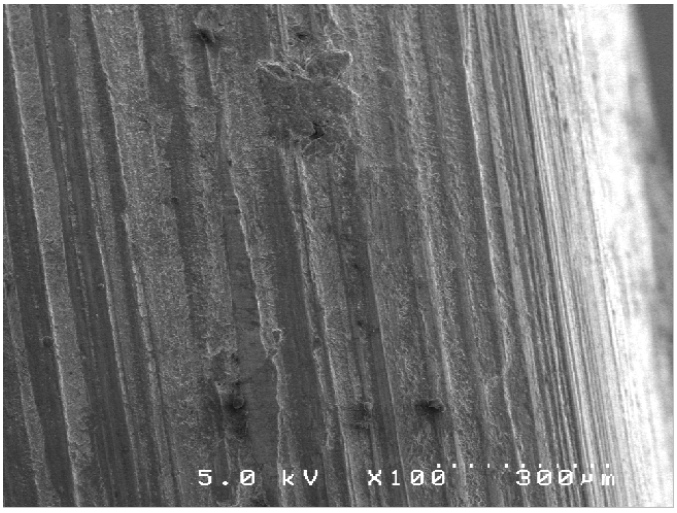


Figure 86 SEM image of tip of tool probe after use in joining Aluminum alloy reinforced with Silicon Carbide particles

7.3 Tracer Experiments: Testing the relationship between Wear and Particle Size

The parallel, regularly spaced groove apparent on the magnified tool probe surface (Figure 86) are a hallmark of abrasive wear. The wear model formulated in chapter VI predicts that wear rate is proportional to the particle radius, a relationship which is also characteristic of abrasion. In the context of this model, particle size is hypothesized to be a critical parameter for determining under what conditions (and with what frequency) wear can occur. The process parameters (tool rotation rate ω and traverse speed v) govern the growth of the rotating plug of plasticized material that surrounds the tool. δ, the clearance between the tool surface and the boundary of the rotating plug , and the size of the particle are hypothesized to be the key factors in determining the amount of material which can be removed from the tool during the course of a single rotation. Since a particle is only able to span the plug and impinge on the tool surface when the radius of the particle exceeds δ, it seems that wear could be reduced (but not eliminated altogether since the clearance distance $\delta \approx 0$ at several locations near the probe tip) by selecting smaller particles for inclusion in the MMC. According to the model, smaller diameter particles should result in comparatively less wear since there are fewer opportunities for them to contact the tool surface and plough out material. This section details the experimental procedure designed to test the effect of particle size on wear in FSW of MMCs. The results of these experiments will either provide additional evidence for the rotating plug model or point the way toward sounder modeling concepts.

Varying the size of particles in an MMC is problematic. In our previous experiments, MMCs were purchased off the shelf from a commercial manufacture (mc21, inc). The purchased MMCs were surplus from large-volume orders, manufactured specifically to meet customer requirements (for instance, an Aluminum 6061 matrix reinforced with Silicon Carbide particles with a diameter of 0.0005" at a volume fraction of 20 percent). Ideally, composites could be obtained (or custom made by the manufacturer) which have the same Aluminum matrix and percentage reinforcement but different particle sizes. The material supply requirements for a particle size study, for example, might consist of three classes of Al 6061/SiC/20p MMCs with particle diameters of 0.05" (large), 0.005" (medium), and 0.0005" (small). It is difficult to acquire off-the-shelf surplus composites which differ only in reinforcement particle size (since said composites must have the same matrix material, reinforcement material, reinforcement type, percentage reinforcement, and fabrication protocol). If cost were not a factor, the obvious solution would be to contract with a commercial MMC manufacture to fabricate a series of composites which meet the particle-size criteria specified above. Regrettably, such a venture would far exceed the scope of the budget for this project.

In light of budget constraints, we propose a methodology used by Colligan et al. as a more cost-effective alternative to custom manufacturing of specific materials. In his work on modeling of FSW, Colligan et al. was able to characterize material flow in FSW of Aluminum alloy using steel-shot tracer particles [105]. For these experiments, spherical steel shot (diameter 0.38 mm) was implanted in a groove cut along one face of the weld seam. Colligan performed a friction stir weld along this seam, rapidly extracting the tool in the midst of the shot after some distance welded (a procedure commonly

referred to as an "emergency stop"). A subsequent plan view radiograph of the plate revealed the flow streamline of material into, around, and out of the tool. By varying the height of the groove on the weld seam and the lateral displacement of the seam with respect to the tool, the streamline from any desired point of origin on the weld cross section could be traced. Colligan's tracer technique is illustrated in Figures 87 and 88. The top half of Figure 87 shows the groove in the portion of the joint where the steel shot is embedded; the lower portion is an exploded view of the complete workpiece (with groove along one face) in a butt weld configuration. Figure 88, also from Colligan et al., is adapted from a plan view radiograph of the workpiece material taken post-weld. The dots represent tracer particles displaced by the tool's motion.

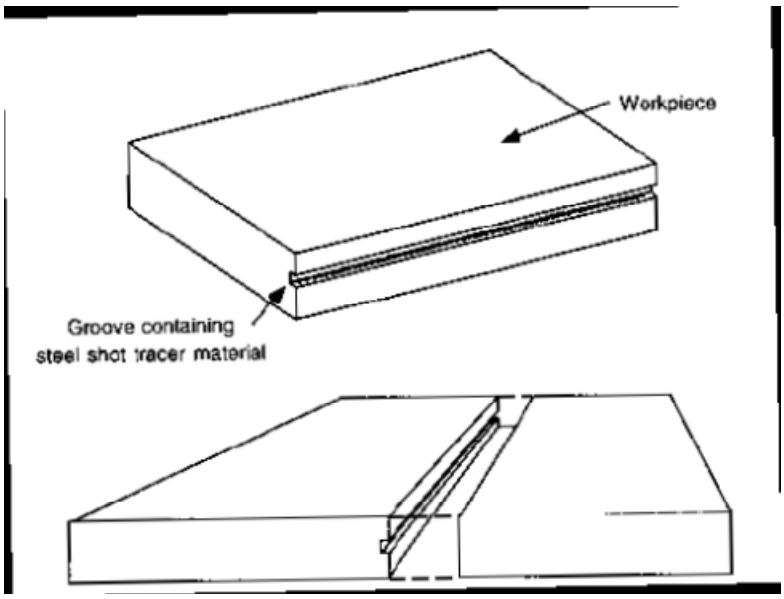


Figure 87 Tracer technique devised by Colligan et al. Steel shot is embedded in a groove which runs along one side of the weld seam.

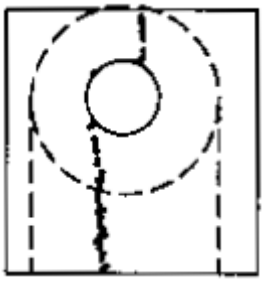


Figure 88 Plan view sketch of radiograph of tracer sample adapted from Colligan et al.

This tracer technique can be easily adapted to study the effect of particle size on tool wear in Friction Stir Welding of Metal Matrix Composites. The degree to which abrasive particle diameter influences material removal in this process can be tested by implanting the reinforcing particles in a square groove machined along the seam of an Aluminum alloy and measuring the corresponding rate of wear. Silicon Carbide particles were selected as the reinforcement material to facilitate easy comparison of results from this study with wear values obtained from earlier experiments involving SiC reinforced MMCs. Abrasive particles are purchased in the form of a powder and their size is expressed as a grade on the Federation of European Producers of Abrasive (FEPA) scale. Table 39 lists the most common (commercially available) FEPA grades and the particle diameters (in micrometers and inches) corresponding to each class. One anamoly of the FEPA designation is that the particle diameters cited are not average measurements. Rather, the specified diameter for each grade designates that all but 10 percent of the particles in the lot have diameters which are less than this value. For example, in an F12 Silicon Carbide powder, only 1 in 10 particles will have a diameter larger than 1400 μm (0.055”). The scale is also unusual in that larger grade numbers correspond to smaller particle sizes. The particles typically used in commercial grade MMCs are in the F100-

F500 range [119]. Thus even if the results of the tracer experiments demonstrate a strong, direct link between particle size and wear, it is unlikely that a significant reduction in wear could be obtained by further decreasing the size of the particulate since the size used in MMCs already lies at the lower end of the spectrum.

Table 39 FEPA Grade and Particle Size [120]

FEPA Grade	Particle Diameter (μm)	Particle Diameter (inches)
F12	1400	0.055
F14	1180	0.046
F16	1000	0.039
F20	850	0.033
F24	600	0.024
F30	500	0.020
F36	425	0.017
F40	355	0.014
F46	300	0.012
F54	250	0.010
F60	212	0.008
F70	180	0.007
F80	150	0.006
F90	125	0.005
F100	106	0.004
F500	12	0.0005

The particles chosen for this study were FEPA grades F14, F60, and F150. These selections are representative of the range of particle sizes available and each increase in grade corresponds to an order of magnitude change in diameter. For example, 90% of the particles in an F14 lot have diameters below 0.055", while 90% of particles in F60 (the next highest grade considered) have diameters which do not exceed 0.008." 1 lb. each of F14, F60, and F150 grade Silicon Carbide particles were acquired from UK Abrasives. The Aluminum 6061 (in which the particles are to be embedded via a groove) measured 1

½" in width, 16" in length, and 0.25" in thickness. A groove with dimensions 16"x0.10" x0.10" was machined into the seam surface of each Aluminum piece. This groove was positioned 0.01" inches from the bottom of the workpiece in order to ensure that contact between the particles and the tool occurs in the region where wear is expected to be the greatest (the distance between the shear surface and the tool is at its smallest for the bottom third of the tool). Since the concentration of particles with respect to the total volume of material is very small, it makes sense to position the groove where it will maximize the likelihood that particles will contact the tool.

The Al 6061 sample is positioned so that the groove is facing upward; adhesive* is injected into the groove using a needle. Immediately following the application of the adhesive, the abrasive particles are spread along the surface of the seam.** The width and depth of the groove is large enough to accommodate the largest particles considered in the study (F14 grade SiC, which has an expected particle diameter of 0.055"). Particles which are dispersed along the seam fill the groove and are held in place via the adhesive. After twenty minutes of curing time, the Aluminum alloy is turned on its side – particles inside the groove remain glued in place, while those outside the groove (which have not come in contact with any adhesive) fall away.

*The adhesive used in this work was wood craft glue. There is little danger of flammability since the flash point of this substance is higher than temperatures encountered in welding and the glue is present in a relatively small amount. The volume of injected adhesive was approximately 25 cubic centimeters over the length of the sample.

**Adhesive is not injected along the first 2" of the 16" sample. During welding, this region of the jointline is comprised primarily of the dwell period. The absence of particulate in this location is intended to ensure that any wear does not occur until the weld process has attained steady state. As discussed in chapter V, it may be the case that wear is accelerated in the dwell region since the traverse speed is significantly slower than the tool velocity during the weld.

It is sometimes helpful (particularly for smaller particles) to inject some additional adhesive after the particles have been placed in the groove – this helps the particles to adhere to one another in addition to the sides of the slot. Using the adhesive as an inter-particle binding agent helps ensure that particles will remain in place during fixturing and clamping of the sample prior to welding.

The experiments outlined in this section consider the effect of three particle sizes (F14, F60, and F150) on the amount of wear. It is also important to evaluate the influence of parameter selection within each particle size category. The ability of a particle of a specific size to score the tool is closely related to parameter selection. Within the context of the rotating plug model, the clearance between the tool and the rotating plug boundary at steady-state (the distance δ which governs whether a particle is of sufficient size to abrade the tool) depends upon the process parameters ω and ν selected by the operator. The plan of experiments designed for this study thus considers three factors: rotation speed ω, traverse speed ν, and particle size (specified in terms of the expected particle diameter D). Summarized in Table 40, the experimental matrix may initially appear limited in scope. Parameters were chosen to provide variation in the weld pitch (the ratio of traverse rate to rotation speed), produce welds with acceptable appearance, and isolate effects (permitting a test of the independent effect of traverse speed or rotation speed on wear for each particle size). All of the experiments were repeated as a test of consistency and samples from each experimental test case were subject to thorough and extensive analysis.

Table 40 Test cases for tracer experiments

Particle Size (FEPA grade)	Rotation speed	Traverse speed
F14	1000	5
F14	1500	5
F14	1500	9
F14	2000	5
F14	2000	9
F60	1000	5
F60	1500	5
F60	1500	9
F60	2000	5
F150	1000	5
F150	1500	5
F150	1500	9
F150	2000	5

The plan of experiments in Table 40 consists of 9 test cases (since each test is repeated, this translates into 18 total experiments). The tool design chosen for this study was a featureless cylindrical probe 0.25" in diameter and 0.235" inches in height. This length ensures that when the probe is engaged in the material (at a plunge depth of 0.005" and a tilt angle of 1 degree), the vertical position of the probe tip approximately coincides with the bottom wall of the slot containing abrasive particles. The tool shoulder measures 0.75" in diameter.

Rather than fabricate 18 separate tools, an economical and efficient alternative is to modify the standard tool design to permit easy exchange of the probe. The only difference between this design (shown in Figure 90) and the conventional FSW tool is the presence of a hole (with a diameter equal to that of the probe) bored along the centerline of the tool. A tool probe in the form of a cylindrical rod can be inserted through this bore-hole and adjusted until the portion of the rod which extrudes from the shoulder

matches the desired probe length (0.235” for these experiments). Once the length is set, the probe is secured in place by means of two set screws. The set screw on the side of the tool shank interfaces with a flat machined on the surface of the probe. A second set screw, inserted from the back of the tool that is opposite the probe tip, prevents vertical movement of the probe during welding. The threaded holes in Figure 90 (one located on the side of the tool and the other at the bottom) serve as the points of insertion for the set screws.

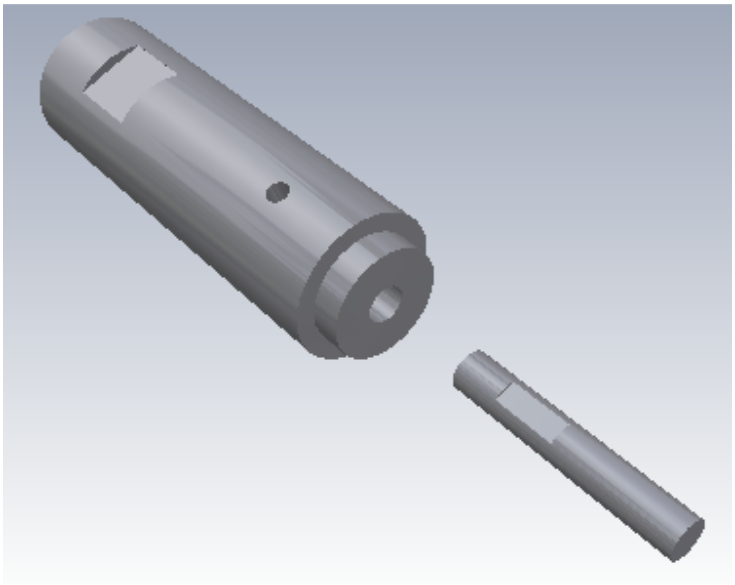


Figure 90 Friction Stir Welding tool design for tracer experiments

The interchangeability of the probes possible with this tool configuration is also beneficial from an analysis standpoint. The two-part tool design opens the door for additional methods of gauging wear that were previously precluded by the size and/or mass of the monolithic tool A high resolution mass balance can detect even the most

subtle changes in the weight of the probe inserts due to wear. The mass of the tool inserts (15-50 grams, depending on the probe material) falls well within the range of these devices, which can resolve differences in mass on the order of micrograms. The reduced size of the probe insert also facilitates microscopic analysis of its surface. Unlike the monolithic tool, the geometry of the insert is compatible with the specimen chamber for the SEM housed in the Vanderbilt Institute of Nanoscience and Engineering (VINSE) laboratory. Images of the probe surface topology acquired using this instrument offer insights which could not be readily obtained using other methods. For instance, the microscopic examination of the surface in section 7.2 conclusively confirmed abrasion as the primary mechanism of wear in FSW of MMCs.

The procedure for each tracer experiment is as follows:

1) The cylinder which functions as the tool probe was inserted through the bore hole until the portion extruding from the tool shoulder was 0.235" (as previously indicated, this length was chosen to ensure that the bottom of the probe coincides with the lower boundary of the slot containing abrasive when the tool is engaged in the material). Once the desired length is obtained, the probe's position is fixed by tightening the set screws.
2) The weld is performed at the parameters specific to the experimental case (Table 40). The tool enters the material from the side. The first inch of weld is set aside for the dwell period, wherein the tool rotates at the spindle speed indicated in the plan of experiments, but traverses at a slower rate (1 inch/minute) to generate sufficient heat for plasticization. After dwelling for one minute, the linear velocity of the tool is increased to the traverse speed

associated with welding. The first two inches of the weld do not contain abrasive – this ensures that the tool does not encounter abrasive particles (and that wear does not occur) until the system has reached steady-state. The tool is extracted from the material one inch from the end of the jointline (the total portion of the weld containing abrasive is 14 inches in length). tool is extracted rather than permitted to "run out" (traverse past the joint line) because the keyhole defect left by extraction may prove useful in analysis. Colligan was able to discern the path of embedded material around the tool from plan view x-rays of the keyhole (his emergency stop procedure.

3) The cylindrical insert is removed from the tool using a drill press. The tool and the insert are then etched in a solution of water and Sodium Hydroxide (NaOH) to remove excess Aluminum which adheres to the surfaces of the probe and shoulder during welding. The wear of the insert is gauged based on differences in its pre and post-weld weight as well as microscopic changes in surface topology.

While our experimental procedure is very similar to that put forth by Colligan, the primary objective of our experiments (to quantify the effect of particle size on wear) necessitated some departures from Colligan's technique. Colligan initiated the emergency stop procedure (in which the tool is extracted from the material *in media res*) comparatively early in the weld process. The stop procedure for our experiments, described in (2), is not initiated until the weld is nearly complete. The longer weld distance is intended to maximize the amount of wear which occurs during the weld –

because the particles are small and concentrated in only one portion of the material, a longer weld distance is warranted to ensure that the tool undergoes a detectable level of volume loss. In the Colligan study, the position of the groove containing steel shot was varied to examine material flow at different locations with respect to the tool. Groove placement in our study is the same for every experiment and was chosen so that any potential contact between the tool and the abrasive will take place in the region where δ (the clearance between the plug boundary and the probe surface) is expected to be thinnest. Confining abrasive to this region also enables us to conclusively determine whether fretting (wear induced by vertical oscillatory movement of the particles with respect to the tool) contributes to material removal. Relocating the groove containing abrasive closer to the surface of the material has interesting implications for the wear model (which predicts that wear will not occur in this scenario since the thickness of δ in the vicinity of the tool shoulder is large enough to prevent particles from contacting the tool). The effect of groove position on wear may be examined at a later stage.

The results of these experiments are discussed in sections 7.4 (which quantifies the effect of abrasive particle size on the volume lost by the tool probe) and 7.5 (which examines the degree to which particle size impacts surface texture). The findings of these analyses are particularly important with regard to the rotating plug model of wear, which indicates that the size of the particle strongly influences the amount of material lost by the tool. Data which supports this hypothesis serves as further evidence that the rotating plug formulation is an accurate representation of the fundamental process dynamics which underlie wear in this system.

7.4 Effect of Particle Size on Volume Loss

Prior to each experiment in Table 40, the weight of the associated probe insert is measured using a microgram mass balance (part of the suite of instruments in the Nanocrystal Fabrication Laboratory at the Vanderbilt Institute for Nanoscale Science and Engineering). This instrument has a display resolution of 0.1 milligrams with +/- 0.1 milligram repeatability. A sliding panel is used to isolate the specimen and minimize the influence of air buoyancy/flow of air under the weighing tray. The probe insert is weighed again post-weld after undergoing an etching process to remove adherent Aluminum. The pre and post-weld weights of the probe inserts are compared for each experiment as a way to measure wear. The changes in weight are very small relative to the original weight of the tool insert.

The reported density of O1 steel (the material for both the tool and the insert) is 127 g/in^3. The density specific to our specimens can be calculated from the product of the original mass and volume of the insert. The original volume of the tool insert is approximately (this value represents the volume of a right circular cylinder with a radius of 0.25" and a height of 3" minus the volume of the x x rectangular flat which, in combination with a set screw, holds the insert in place during welding). The density of the tool material is calculated to be -- any discrepancy between the reported and calculated value of density can probably be attributed to the heat treatment inserts undergo prior to use in welding.

This value of density can be used to scale the mass difference so that it represents the mass lost *only* by the portion of the cylindrical insert which comprises the probe. The probe (a right circular cylinder with a diameter of 0.25" and a length of 0.235") initially

has a volume of 0.00367 in^3. The theoretical mass of the probe (independent of the insert) is estimated as 0.47 grams (the density multiplied by the initial probe volume). It is assumed that any change in the weight of the insert is a result of interaction between the probe and abrasive particles – any wear which occurs is thus limited to the region of the cylindrical insert that is engaged in the material during welding. While it is possible that some additional material loss occurs when the insert is separated from the tool using a drill press, the contribution of this extraction to wear is assumed to be negligible. The post-weld mass of the probe is the difference between the initial probe mass (the theoretical value calculated based on the density and volume of the probe) and the change in mass measured using the mass balance. Since all of the inserts have slightly different initial weights, it is more useful to express the mass lost by the probe as a percentage of its original mass. The percent reduction in probe mass is calculated from equation 7-1. m_i denotes the initial mass of the probe and Δm is the change in mass of the probe insert.

$$\%\, reduction\ in\ probe\ mass = \frac{m_i - \Delta m}{m_i} \text{x} 100\% \quad \text{(equation 7-1)}$$

Information relating to changes in the volume and mass for the insert and the portion of the insert which constitutes the tool probe is summarized in Table 41.

Table 41 Data for particle-tracer experiments

Particle Size (FEPA)	Rotation speed (RPM)	Traverse speed (IPM)	Pre-weld mass of insert (g)	Post-weld mass of insert (g)	Delta m (g)	% reduction in mass of probe
F14	1000	5	17.78001	17.6530	0.12701	8.60%
F14	1500	5	17.77493	17.46795	0.30698	20.78%
F14	1500	9	17.75787	17.55120	0.20667	14.00%
F14	2000	9	17.745315	17.453475	0.29375	19.92%
F14	2000	5	17.93785	17.7717	0.16615	11.14%
F60	1000	5	17.799955	17.7335	0.066455	4.50%
F60	1500	5	17.78722	17.624525	0.162895	11.02%
F60	1500	9	17.79782	17.74915	0.04867	3.29%
F60	2000	5	17.9446	17.9078	0.0368	2.47%
F150	1000	5	17.8607	17.8133	0.0474	3.19%
F150	1500	5	17.80278	17.7265	0.07628	5.16%
F150	1500	9	17.78488	17.773	0.01188	0.80%
F150	2000	5	17.944	17.9149	0.0291	1.95%

The plot in Figures 91 shows the relationship between particle size and the volume lost by the probe during welding for each parameter set. For a given parameter, wear generally increases with particle size, a result that is consistent with the plug model – since larger diameter particles are more likely to be able to span the clearance between the rotating plug and the tool, there is greater opportunity for them to remove material from the tool surface during a rotation. It follows that larger-size inclusions will produce comparatively greater wear rates. Figure 91 shows that for each parameter set, the probe wear associated with the largest particle size (FEPA grade 14 with a particle diameter of 0.046") can be as much as three times greater than the wear observed for the smallest particle size (FEPA grade F150 with particle diameter 0.0025").

For a constant particle size, the model predicts that wear will generally increase with weld pitch (the ratio of rotation rate to traverse speed), a relationship initially derived

based on the data from the wear study in chapter V. According to the rotating plug model for wear, as the distance between the boundary of the rotating plug and the tool surface decreases as the weld pitch increases—the thinning of this zone enables the particles to contact the probe surface at more locations, contributing to an increase in wear. This phenomenon is observed to some degree in Figure 92, which plots wear for each particle size as a function of weld pitch As the weld pitch is increased from 167 (the 1500 RPM, 9 IPM parameter) to 300 (the 1500 RPM, 5 IPM parameter) for each particle size, the wear of the probe generally becomes more pronounced.

The highest weld pitch (400 for the 2000 RPM, 5 IPM parameter) represents a noticeable deviation from this trend, as the wear values for w/v = 400 are far below those associated with the next highest weld pitch (300 for the 1500 RPM/5 IPM parameter). This observation, which is consistent across all particle sizes, is opposed to the behavior predicted by the rotating plug model of wear, a discrepancy which may be attributable to thermal effects. As the heat input from the tool to the material increases, the corresponding elevation in the temperature of the workpiece contributes to the growth of the rotating plug [101]. Since the plug can act as a buffer (preventing smaller particles which are unable to span the distance between the plug and the tool from contacting the probe), thermally-induced expansion of the plug can impede the wear mechanism. Thermal effects seemed to have little impact on the wear rates reported for FSW of MMCs in chapter V, where weld pitch was determined to be directly proportional to wear. The relationship between wear and process parameters in FSW of MMCs was believed to be largely governed by kinematic effects – the plug became thinner as weld pitch increased, permitting increased removal of material by the abrasive. The key

difference between the tracer experiments and the previous wear studies is the workpiece material. The particle tracer experiments use Al 6061, a material which is less temperature resistant than its MMC counterpart. It is possible that the development of the rotating plug in Al 6061 is accelerated by heat input to a degree greater than what is possible in the thermally rigid composite, resulting in reduced wear rates for high temperature parameters. If this were indeed the case, we would expect material lost by the tool in the particle tracer experiments to be mitigated by parameters with fast rotation rates and comparatively slow traverse rates. The effect of the thermal environment on wear (and the degree to which it can exacerbate or impede wear in materials of varying thermal resistance) is a topic which merits further study.

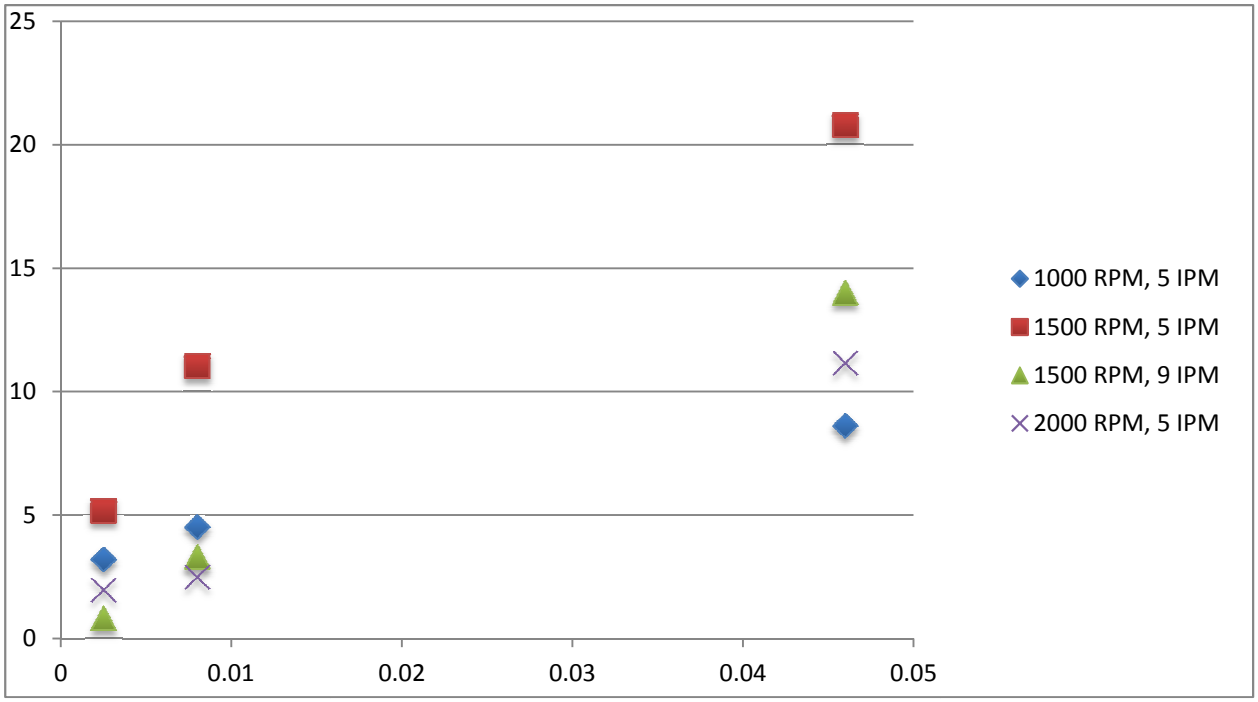


Figure 91 Plot of percent volume loss (of probe) versus particle size. Each data series represents a separate set of parameters.

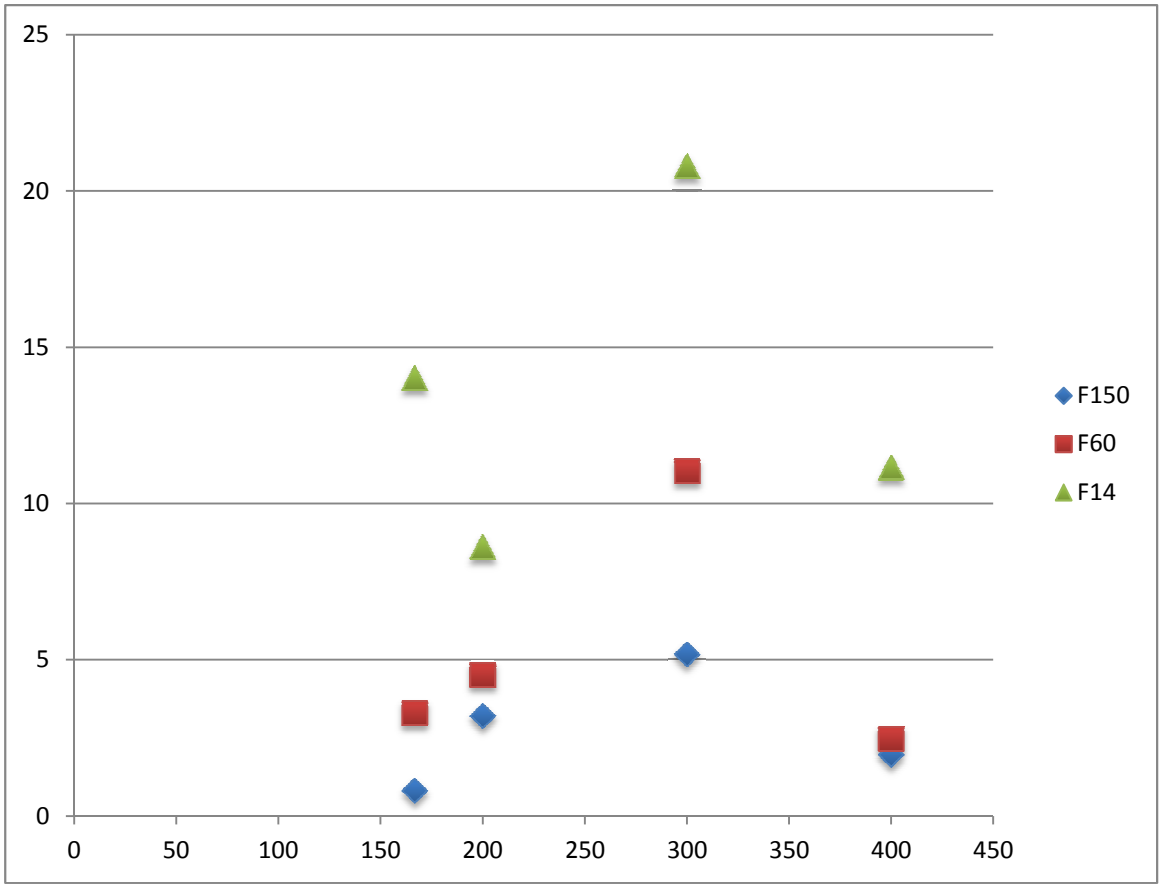


Figure 92 Plot of percent volume loss (of probe) versus weld pitch for F14 and F60. Each data series represents a specific particle size.

The expression for the time rate of material removal derived from the rotating plug model indicates that wear is directly proportional to the square of the particle diameter. To test this hypothesis, linear, quadratic, and logarithmic regressions were performed on the data sets associated with each parameter. These regression equations express the percentage of the original volume lost by the probe as a function of particle size. The results of the regression analyses are summarized in Table 42. An initial concern was that the range of particle sizes considered in our study is sufficiently small to make the data appear linear when this may not be the case -- a mathematical concept

known as local linearity. As the domain of a polynomial shrinks, the function will begin to resemble a straight line. In this way, the macroscopic behavior of the function is obscured when it is examined over a narrow domain. The nonlinearity of the relationship between wear and particle diameter, however, is apparent even on this small scale. Nonlinear effects can be surmised from a cursory analysis of the data in Figure 7, where an 18-fold increase in particle size produces anywhere from a 3 to 17 fold increase in wear, depending on the weld pitch. The regression analysis bears this out, as both the quadratic and logarithmic functions derived for a given parameter have higher R^2 values than a linear regression performed on the same data.

Table 42 Summary of regression analyses for data from tracer experiments. The regression equations relate volume lost by the tool during welding to the diameter of the SiC inclusions.

Parameter	Quadratic regression	R^2 value for quadratic regression	Logarithmic regression	R^2 value for logarithmic regression
1000 RPM, 5 IPM	$-11600D^2+700D$	0.849	$1.9(\ln D)+14$	0.999
1500 RPM, 5 IPM	$-25800D^2+1640D$	0.986	$5.4(\ln D)+37$	0.999
1500 RPM,9 IPM	$-2530D^2+420D$	0.999	$4.7(\ln D)+28$	0.948
2000 RPM, 5 IPM	$-3000D^2+380D$	0.978	$3.3(\ln D)+20$	0.878

For each parameter set in Table 42, there are two candidate functions which express wear as a function of particle diameter D. Because the R^2 values for each regression equation are so close in magnitude, we revert to our fundamental

comprehension of abrasive wear processes as a way to determine which function (quadratic or logarithmic) best captures the effect of particle size on material loss. In every case, the quadratic function is a concave down parabola with a y-intercept of 0. The y-intercept condition is imposed to signify that wear can only occur when particles are present ($D > 0$). The quadratic equations in the second column of Table 42 predict that wear will increase with particle size until it reaches a maximum (for the parameters in this table, D_{max} ranges from 0.06" to 0.12", particle diameters which roughly correspond to a FEPA grade of F12). For particle diameters which exceed D_{max}, wear decreases with increasing particle size . The predicted behavior of the wear curve beyond the extremum is problematic, as it means that while volume loss increases from the smallest particle size up to F12 (a prediction which is consistent with experimental observations), it will decrease for particles with diameters larger than this size. The y-axis symmetry of the parabolic function suggests that wear for the largest particles will be equal to wear produced by the very smallest particles, a result which is not compatible with the historical understanding of abrasion.

In light of these considerations, it is more likely that the function relating wear and particle size is logarithmic. A logarithmic $W(D)$ function means that wear is always increasing with particle size, but the magnitude of this increase is less dramatic for larger diameter particles (the limiting effect of particle size on wear predicted by the logarithmic function is analogous to the asymptotic relationship between wear and sliding distance documented by Rabinowicz). If the overall trend is indeed logarithmic, it is not surprising that a quadratic function is also a good fit for the particle size/wear data. The region of interest is small enough that the logarithmic function may appear locally

quadratic (just as a polynomial function can be locally linear when the domain is sufficiently narrow). Close to the maximum of the curve, the equation for a parabola with downward concavity has virtually the same values as a logarithmic function (or vice versa). A conclusive determination as to whether the fit is logarithmic or quadratic (it could be either, although the trends predicted by a logarithmic equation are more intuitive) requires additional data points.* The central finding of this research is that smaller diameter particles generally translate into less volume loss, a result that corroborates the relationship between inclusion size and wear initially indicated by the plug model.

As the results summarized above demonstrate, the tracer methodology adapted from Colligan affords key insight into the process dynamics of wear in FSW of MMCs. The method, however, has some limitations which merit further discussion. Chief among these is that the particulate inclusions in the tracer samples are confined to a small portion of the workpiece volume (rather than homogeneously dispersed throughout the material, as they would be in an MMC). The high concentration of particles at the jointline and near the bottom of the workpiece ensures that wear occurs only at the probe tip.

*The limited range of particle sizes considered in this study means that there are only three data points for each series (which correspond to grit sizes of F150, F60, and F14). While it is unlikely that MMCs would contain reinforcements with particle diameters smaller than F500 or larger then F12, the inclusion of intermediate data points can potentially clarify the relationship between particle size and wear. Assuming the same quality of data, an n-point regression will offer a more accurate representation of trends in the data than the 3-point regression presented here.

The wear patterns produced by positioning the particles in this manner are not very different from those observed in a "true" MMC – this is because the particles are situated in the region where the clearance between the rotating plug and the tool surface is smallest (permitting them to impinge on the tool surface and remove material). Even if particles were present in other regions of the tracer sample, they would be too distant to contribute to wear of the tool (or prevented from doing so by the widening of the plug in the axial direction). While the arrangement of particles in this manner appears to neither impede nor accelerate the wear process, it does decrease the percentage reinforcement for the sample as compared to an MMC. The percentage reinforcement by volume for each tracer sample, %, is independent of particle size. Any variation in particle diameter changes the number density, n (the number of particle per unit volume), a term which appears explicitly in the rate equation for wear derived from the plug model. Number density is inversely proportional to particle diameter and, since number density depends on the particle size, the terms cannot be changed independently of one another. The reduction in n which accompanies an increase in particle size is balanced (or more accurately, overridden) by the change in the radius of the particle, whose contribution to wear manifests itself in the r^2 term of equation 6-22. While we might initially expect wear to decrease with particle size simply because there are fewer particles available to abrade the tool, the prior analysis (combined with the modeling efforts detailed in chapter VI), indicate that just the opposite is true.*

*The number density term does not appear in the cutting arc method (section 6.3), where wear is modeled solely as a function of particle size and process parameters. As a result, the direct correlation between wear and particle size is more easily discerned in this formulation.

The tracer technique used by Colligan and adapted here to study wear in FSW has important implications for machining as well; this method presents a novel, economic, and efficient way to study the effect of particle size on wear in any machining process. The advantage of the technique is twofold: 1) it is much cheaper than purchasing/manufacturing multiple lots of composites which differ only by particle size and 2) it isolates the effect of particle size (and thus naturally lends itself to testing wear models which include reinforcement size as a variable). With regard to (2), the tracer technique is superior because it does not impose limitations on the size of the particles. For the machining results summarized in section 4.2.2.2, the scope of the studies was limited to the particle sizes included in the off-the-shelf composites purchased from the manufacturer. With the Colligan method, the experimenter is free to select particle sizes from the entire spectrum of grain sizes used in manufacturing applications.

What we have learned about the effect of particle size on wear in FSW of MMCs (and the competing effects of particle size and number density on wear) is relevant (and hopefully also transferrable) to machining this class of materials. The results of our analysis are closely aligned with published research on wear and particle size in machining MMCs, an indication that the FSW process is not as far-removed from traditional machining processes (cutting, drilling, milling, etc.) as it may initially seem. The Kannan study (which found varying the diameter of the reinforcement by as little as 5 micrometers produced a change in tool flank wear on the order of 0.1 mm) , along with the results of Sahin/Sur and Kishawy et al., demonstrates the variation of wear with particle size in machining is direct and detectable [63-64, 82].*

*Refer to section 4.2.2 2 and 4.2.2.3 for a more detailed summary of the collective research on wear in machining of MMCs.

Tools in contact with particle diameters corresponding to the extrema of grain sizes used in manufacturing applications exhibit large variations not only in the amount of wear but also the pattern of abrasive tracks left behind at the particle/tool interface, a result which is also borne out by our research on wear in FSW of MMCs.

7.5 Profilometry

The weighing method used to estimate wear in section 7.4 uses the mass lost by the tool to gauge the effect of particle size and process parameters on wear processes. The observed direct relationship between particle size and volume loss is consistent with the rotating plug model for wear developed in chapter VI.

In the context of this model, wear is possible only when an abrasive particle embedded in the Aluminum matrix is able to contact the tool surface (this occurs at locations where δ, the distance between the tool surface and the plug boundary, is less than the particle radius). When conditions develop which enable the particle and the surface to come into contact, the abrasive grain ploughs out material as the tool rotates past it, leaving behind a groove which extends around the probe. These circumferential grooves produced by abrasive action are evident on microscope images of the probe surface (Figure 86). The total volume of these grooves is approximately equivalent to the volume lost by the tool during all of the particle/tool interactions which take place during the joining process. Because the wear mechanism takes the form of abrasion, the surface topology of the tool sample after welding will be markedly different from its pre-weld texture. Based on the images collected in section 7.2, this appears to be the case. For

instance, the tool probe in Figure 2, characterized by parallel trenches cut into the surface during abrasion, clearly has a rougher surface texture than its as-machined, pre-weld counterpart shown in Figure 1. In theory, the degree to which the pre and post-weld surface textures differ should be indicative of the amount of wear which has occurred. The size of the reinforcement particles in the MMC should thus have a substantial impact on surface texture, as larger particles are able to impinge on the tool surface with greater frequency (and to a greater depth) than smaller particles. While all specimens used to join an MMC will experience some abrasion (and thus have rougher surfaces than they did prior to use), the subtle differences in texture between specimens may hold further information about particle size and its impact on the wear mechanism.

Surface texture is another metric (in addition to the change in weight of the probe and the volumetric deterioration calculated based on the change in the probe's cross-sectional area) which can be used to gauge wear. Data relating particle size to surface texture is obtained from mechanical gauging of the surfaces of the probe specimens from the tracer experiments in section 7.3. Prior to gauging of the specimen surfaces using contact profilometry, images of some of the inserts were collected using a Hitachi S4200 Scanning Electron Microscope (SEM) housed within the VINSE (Vanderbilt Institute for Nanoscience and Engineering) Electron Optics Laboratory. The purpose of this imaging study was to provide a qualitative assessment of the relationship between the depth of the abrasive grooves and particle size before proceeding with a quantitative analysis. The Hitachi S4200 is a high resolution cold field emission electron source with backscatter electron imaging and digital image acquisition. Each probe insert (a cylindrical rod measuring approximately 3" long and 0.25" in diameter) was mounted in a PELCO

Straight Bulk Specimen holder and held in place by 4 set screws. Fixturing of the sample for insertion into the SEM is shown in Figure 93. An M4 screw at the bottom of the holder affixes to the specimen sled, which in turn attaches to the insertion rod for positioning the probe/holder assembly in the vacuum chamber for imaging. Images of each sample were collected at a working distance of 15 mm, an accelerating voltage of 5kV, and magnifications ranging from 50X to 200X.

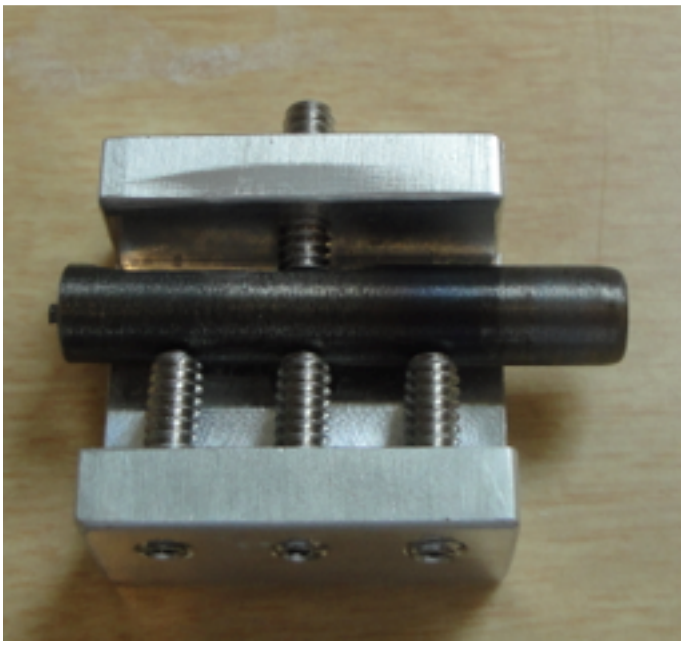


Figure 93 Probe insert mounted in specimen holder for imaging

For the same particle size, the appearance of the specimen surface is quite similar regardless of the process parameters. Figure 94 compares the post-weld probe from the F14/1000 RPM/5 IPM experimental case with the probe from the F14/1500 RPM/5 IPM and F14/1500 RPM/9 IPM experiments. The images show that there is some similitude

between the specimens' surface appearances, even though the rotation rates and traverse speeds associated with each sample are different. In contrast, changes in the diameter of the inclusion produce (what appear to be) substantial variations in surface texture. Figure 95 compares images of tool specimens at the 1000 RPM/5 IPM parameter set for the three different particle sizes. For the F14 image, the grooves are comparatively wider and the tip of the probe (initially a right circular cylinder) has been rounded by abrasion. For the smallest particle size, the grooves are more closely spaced and occur with greater frequency, but appear to be shallower than those on the surface of samples used to join material with the larger (F14 or F60) grade inclusions. The images in Figure 94 and 95 are a good representation of the visual trends in surface texture observed in the particle-tracer experiments: surface appearance is highly sensitive to particle size but less affected by changes in process parameters.

All the samples exhibit grooving, but it is worth noting that grooves are only present in the region near the bottom of the probe. The left image of Figure 96 shows the appearance of the probe tip in the region coincident with the slot containing abrasive; the right image is taken farther "up" the tool, near the probe/shoulder interface (approximately 0.20" from the tip), where no abrasive is present. It is clear from these images that abrasive grooving is the defining feature of surface texture at the probe tip, but becomes less pronounced (eventually disappearing altogether) as we move toward the shoulder. While this result is to be expected since the abrasive particles were confined to the bottom of the weld seam, it is significant because it indicates that abrasive action in the tracer experiments was limited to vertical positions on the tool which coincide with particle placement. This localization of wear is evidence that fretting, in which vertical

movement of the particles is induced when a particle is grasped in the shear zone and undergoes oscillation, does not occur. If fretting were present, evidence of abrasion well above or below the position of the particles should be observed. The rotating plug model of wear does not take into account the possible oscillatory behavior of reinforcement particles; thus it is good news for our model if fretting can be eliminated as a possible contributor to wear.

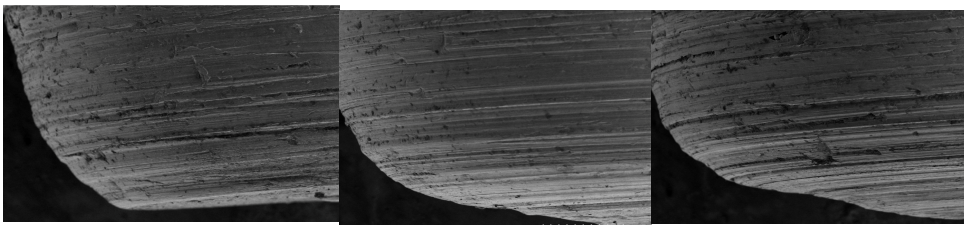


Figure 94 Post-weld images of probes used to join Aluminum containing F14 SiC inclusions at the following process parameters (from left to right) 1000 RPM/ 5IPM, 1500 RPM/ 5 IPM , and 1500 RPM/ 9 IPM.

Figure 95 Post-weld images of probes used to join Aluminum containing (from left to right) F14, F60, and F150 SiC inclusions at 1500 RPM and 5 IPM.

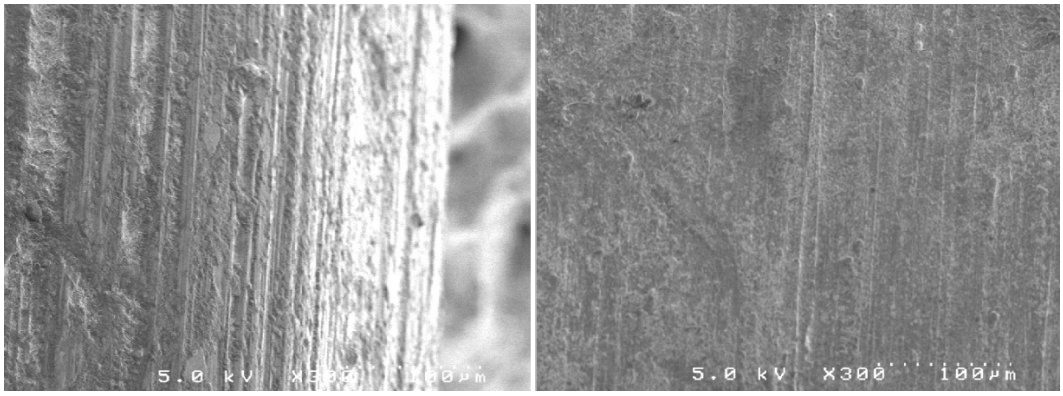


Figure 96 A side-by-side comparison of SEM images taken at the probe tip (left) and approximately 0.20" from the tip, near the probe/shoulder interface (right).

Although examination of the tool surface under a high magnification microscope can aid in identification of the wear mechanism and provide a visual estimate of the degree of material loss, we rely on profilometry techniques for quantitative data regarding the surface topography. Profilometry is the most common method used to characterize the topography of a surface. While there are many techniques which can be used to measure variations in surface height, the methods are generally classified into two broad categories (each of which can be further categorized as two-dimensional or three-dimensional):

1) Contact methods, in which a stylus touches and traverses over the part surface
2) Non-contact methods, in which a light sensor detects and measures changes in the height of the surface

Selection of a method hinges on the part geometry and the resolution needed. The highest profile resolutions are associated with non-contact methods, as the resolution for contact profilometry is limited by the size of the stylus (which can be large relative to

variations in surface height). The cylindrical specimens in this study are well-suited for stylus profilometry: the surface area to be analyzed is small (a few mm in length) and the surface features are large enough to be visible to the naked eye (rendering the higher resolution of optical methods unnecessary). The procedure for acquiring data in this case is also relatively straightforward, requiring only a straight-line trace or scan – simpler operations are better suited to contact methods.

Profilometry data was collected using a Taylor-Hobson form Talysurf PGI 1230, a stylus tracing instrument that is part of the Metrology Laboratory at NASA Marshall Spaceflight Center. This apparatus uses a conisphere diamond-tipped stylus with a 2 micrometer radius and can detect changes in surface height as small as 0.8 nanometers [121]. The photograph in Figure 97 shows the instrument in use, as the stylus traces along a part and the computer software package (Taylor Hobson Ultra) records height values (the software was set to record 1750 measurements per mm). The stylus essentially functions as a moving height gauge, recording subtle variations in the height of the part surface as it traverses across it. The part being examined in Figure 13 is a flat surface; obtaining surface texture data for a cylindrical part proves to be slightly more complex. To take the measurements reported in this study, the part (a right circular cylinder with height 3" and diameter 0.25") was fixtured in a 3 jaw chuck controlled by a rotary motor. To perform a measurement, the stylus (which has a 12.5 mm vertical range) is lowered until it contacts the probe surface; initial contact is established slightly beyond the base of the flat nearest the probe tip. Since the part is round, the location of the cylinder crest must be identified prior to the trace operation. An iterative technique, which consists of translating the stylus in the positive and negative y-direction over

successively smaller ranges, is used to determine the location of the local maximum that coincides with the cylinder crest. After cresting is complete, the position of the stylus in the yaw direction is fixed. To make a trace, the stylus translates along the specimen surface, recording data from a distance of 0.27" above the location of the probe/shoulder interface (just past the edge of the flat) to slightly past the probe tip (a total trace length of 0.50" or 12.7 mm). The reason for not beginning the trace precisely at the probe/shoulder interface is two-fold: 1) the trace along the portion of the insert which was not engaged in the material during welding is intended to serve as a baseline/point of reference for comparison with the data collected from the probe surface and 2) there is a short run-up period of approximately 1 mm (the exact distance required for the run-up depends on the stylus) associated with each trace -- thus starting the measurement at a point before the feature of interest ensures that the stylus has completed the run-up and is collecting viable data well before it reaches the targeted surface).

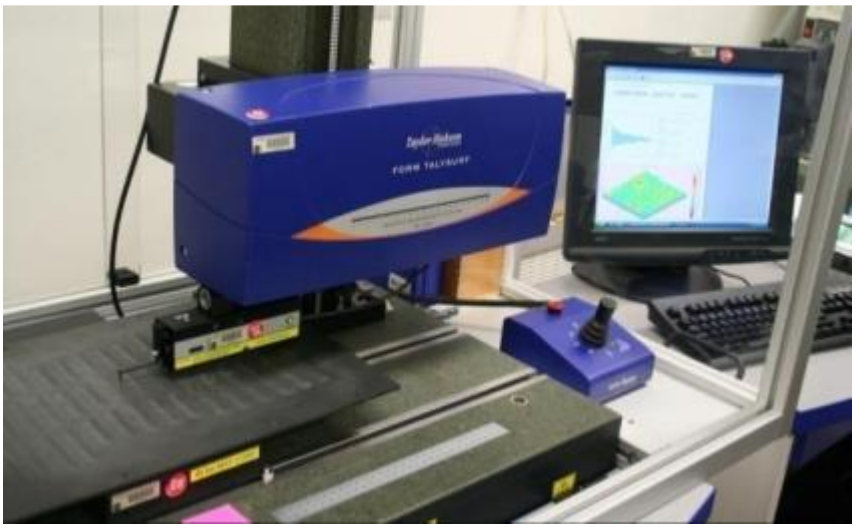


Figure 97 Talysurf PGI 1230

Once a trace is complete, the data can be plotted and analyzed within the Ultra software or exported to Microsoft Excel. The Excel file is simply a list of points (x,z), where x denotes the linear position of the stylus and z is the height of the surface at a particular location. Figure 98 shows an example of a plot of height versus distance for a straight-line trace along the crest of the cylinder of an insert. The arrow designates the location of the probe/shoulder interface. In some instances (such as Figure 99), the probe/shoulder interface can be clearly identified because it coincides with a ridge created when the insert was extracted from the FSW tool holder post-weld. This artifact corresponds to a spike in surface height.

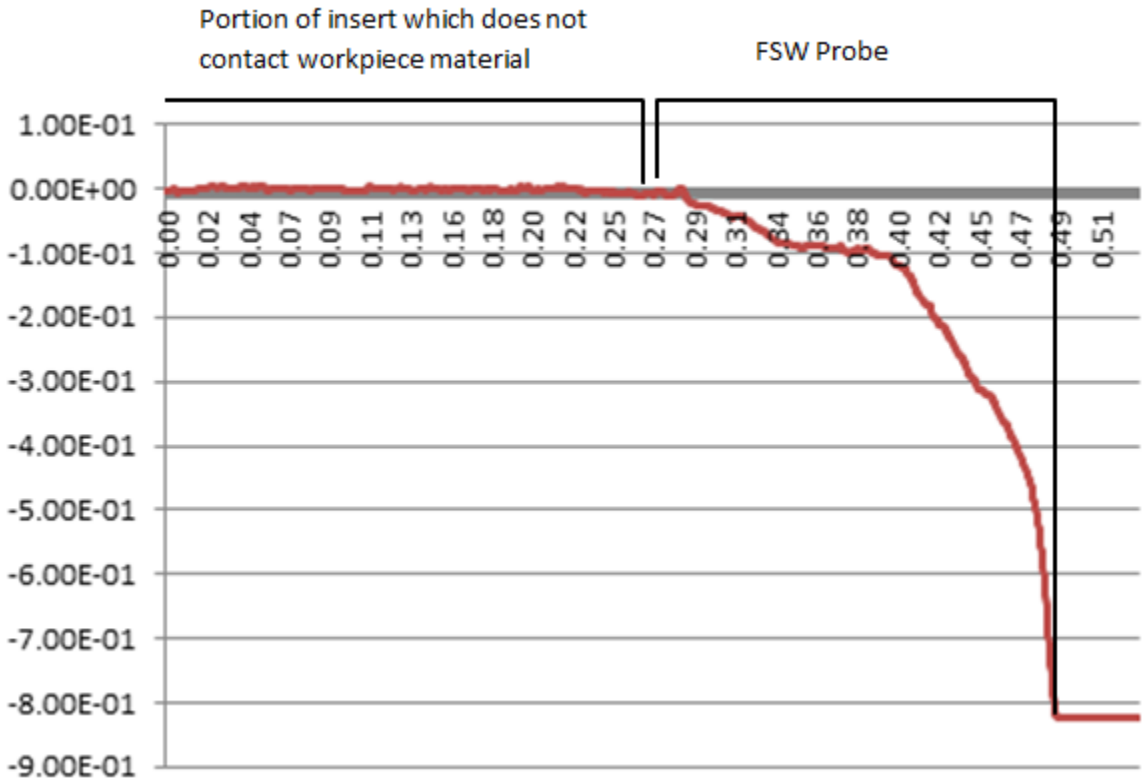


Figure 98 Plot of straight-line trace along the crest of a probe insert used to join Al 6061 containing F60 SiC at 1500 RPM and 5 IPM. The x axis represents the linear position

of the stylus (in inches); the y axis is the corresponding height (in inches) at each location.

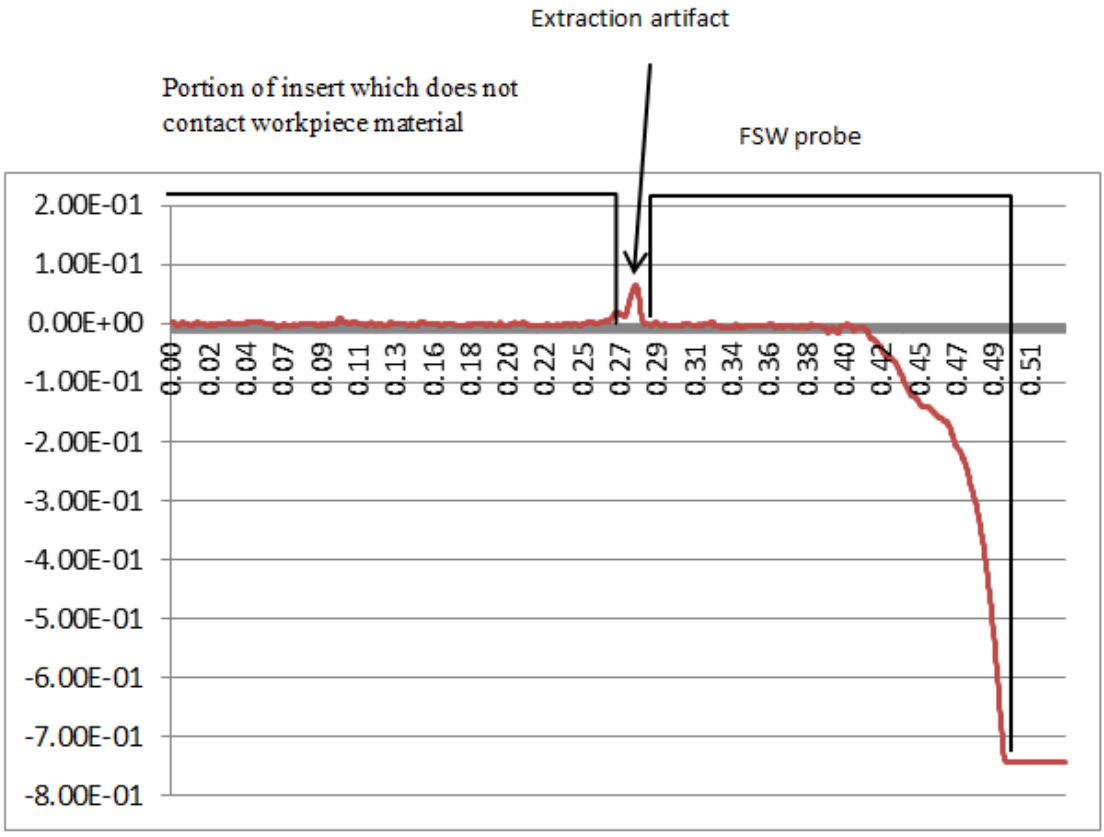


Figure 99 Plot of straight-lint trace along the crest of a probe insert used to join Al 6061 containing F60 SiC at 1000 RPM and 5 IPM. Distance (in inches) is plotted on the x axis; surface height (also in inches) is plotted on the y-axis. The rise in the surface profile corresponding to the ridge created on extraction is indicated.

After each longitudinal straight-line trace, the tool is rotated 90 degrees (clockwise) and another measurement is taken. The orientation of the specimen at each angular position can be defined relative to the rectangular flat machined on the insert surface. 0 degrees describes the part orientation when the flat is facing upward; at 90 degrees, the flat faces the negative y-direction (toward the wall in the setup shown in

Figure 97); the flat points down at 180 degrees; at 270, the flat faces outward; the 360 configuration returns the part to its original position. A comparison of these four traces for each specimen confirms the symmetry of the wear process in FSW of MMCs. Figure 100 plots the 0, 90, 180, and 270 degree traces for an insert on the same axes -- the traces are virtually coincident. The stylus-stop height is slightly different for each orientation, but the overall shape (i.e. the slight curvature of the probe resulting from wear) varies little with insert orientation.

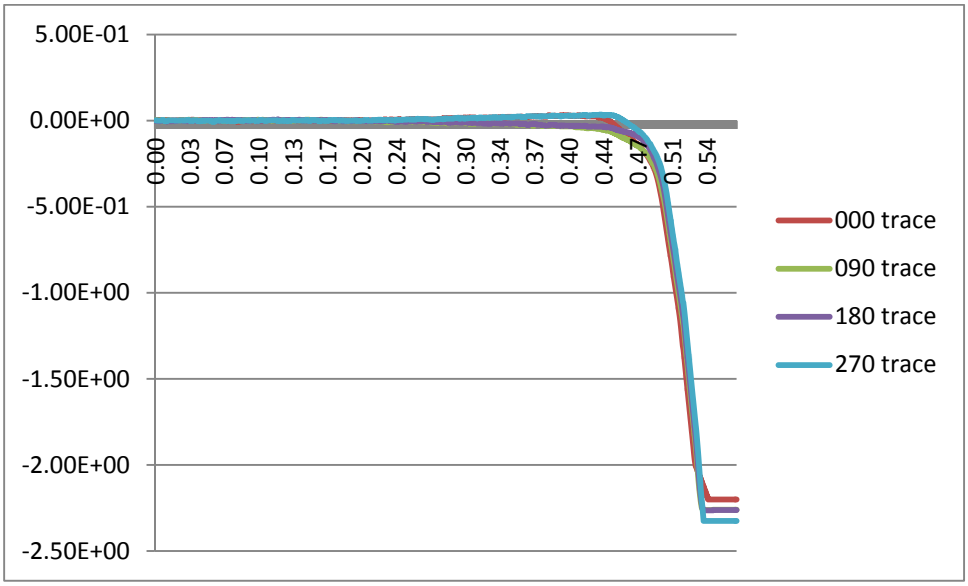


Figure 100 Comparison of longitudinal traces at the 0, 90, 180, and 270 degree orientation for an insert used to join a workpiece reinforced with F150 SiC particles at parameters 1000 RPM and 5 IPM. Surface height is on the y-axis and linear distance is plotted on the x-axis.

The impact of wear on tool shape is clearly evident from the profile plots, but there appears to be little meaningful variation in the surface height. Subtle variations in the height are concealed by the scale of the plots. Because the range over which the data

is plotted is large compared to the range at which changes in surface texture become visible, the straight-line trace along the insert appears flat up to the point where the probe tip has been rounded by wear processes. Figure 101 illustrates that the macroscopic scale on which the trace is initially viewed belies the underlying surface structure. The top image (101a) shows the portion of the trace for 1500 RPM/5 IPM/F60 SiC corresponding to the region of the insert that does not engage with the workpiece material (spanning the x values from 0" to 0.27"). The height scale for the plot in 101a ranges from -0.7" to 0.2" -- the surface height appears to be largely invariant over this range. Contrast this with the bottom image (101b), which displays the same trace over a much narrower height range: -0.015" to 0.010." The surface texture is easily discerned on this smaller scale, as ridges, pits, and oscillations in the surface profile emerge. The scale reduction makes it apparent that even in a region which is unaffected by wear, the surface of the insert is not flat. It is anticipated that changes in surface texture will be even more pronounced over the region of the insert which comprises the tool probe. Characterizing the impact of particle size on probe surface texture is the subject of the next section.

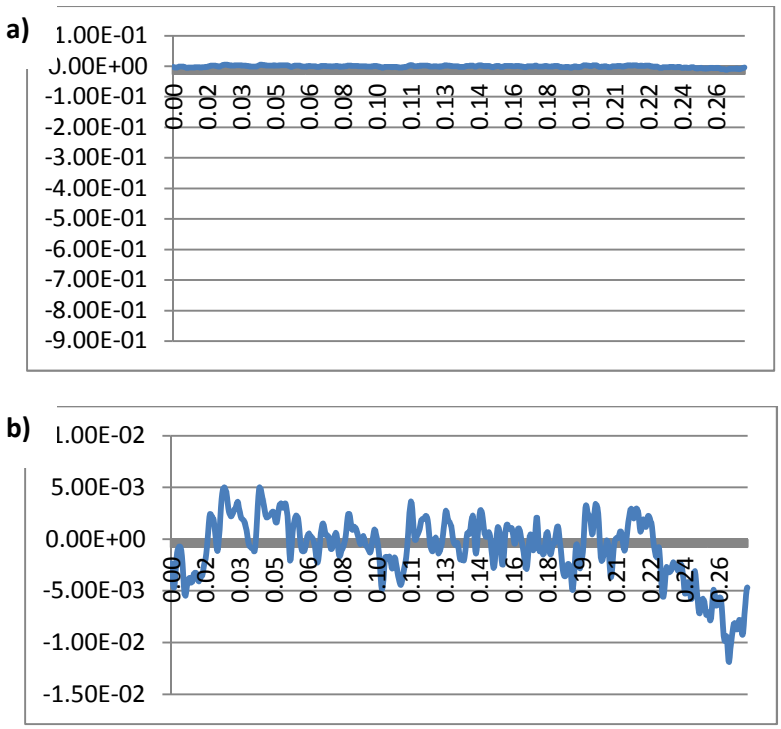


Figure 101 a) single profile trace of unused insert surface with vertical axis range scaled from -0.7" to 0.2" b) the same trace with vertical axis scale changed to -0.015" to 0.01"

7.6 Effect of Particle Size on Surface Texture

The x,z data collected from longitudinal, straight-line traces of the inserts used in the tracer experiments hold important information about surface texture and how it is impacted by changes in particle size and/or process parameters. Characterizing the topography of the insert can provide additional insight into the mechanics of the wear process in FSW of MMCs. For abrasion, the characteristic depth of the grooves and the frequency with which they occur on the probe surface roughly coincides with the amount of material removed (we thus expect grooving to be more pronounced for surfaces which experience more wear). The features of the tool surface apparent in visual inspection of

the tool using microscopy (i.e. circumferential grooving) can be quantified using several key surface texture parameters. The results of this analysis will confirm and/or shed additional light on three key trends observed in the imaging study:

1) For the same parameter, grooves appear to be deeper for larger particle sizes
2) Surface texture appears to be more sensitive to changes in particle size than process parameters.
3) Grooving does not extend all the way up the probe surface, but is concentrated in the region coincident with the groove containing abrasive.

The surface texture parameters of interest include R_a, R_q, R_t, R_p, and R_z. The physical meanings of these parameters (and how each is calculated) are summarized below [122]:

- R_a, the average roughness of a surface, is defined by the integral $\frac{1}{L}\int_0^L abs(z)\ dx$. z represents the height of each point in the data set and dx is the scale increment/horizontal distance between two successive data points. L is the total length of the profile. R_a can be interpreted as the area under the surface profile curve per unit length. R_a is thus expected to be larger for rougher surfaces (where the deviation of the profile from the centerline is more pronounced).

- R_q is the root-mean square value of the average roughness. Calculated from the integral $\frac{1}{L}\int_0^L z^2 dx$ (or using the approximation $1.25R_a$ when the surface heights fit a random Gaussian distribution*), R_q represents the deviation of the height values about the mean height (R_a). While the parameters R_q and R_a are closely related, the former is considered the more meaningful since it conveys the overall closeness of the data to a central point (the mean).
- R_t is the peak to valley roughness. Mathematically, it is the vertical distance from the highest peak in the data set to the lowest valley.
- R_p is the distance from the mean height of the surface to the highest peak.
- R_z is the distance between the average height of the five highest peaks and the average height of the five lowest peaks in the data set.

The Ultra software associated with the Taylor-Hobson profilometer can calculate these metrics (as well as many others) automatically. In the Ultra software, the data to be used in the analysis is selected using a visual method, wherein a window is expanded or contracted until it includes the entire region of interest (and excludes all other points). The alternative representation of the trace as a list of points (x,z) enables more precise selection of data for the subsequent surface texture calculations; for this reason, we have chosen to use Excel to compute the desired metrics.

*Gaussian distribution means that 68% of the heights lie within one standard deviation of the mean surface height, 95% of the data fall within two standard deviations, and 99% are bounded by an upper and lower limit of three standard deviations.

There are three regions of interest for each specimen surface: 1) the portion of the insert which does not contact the workpiece material (roughly spanning from the edge of the flat to the probe/shoulder interface), 2) the portion of the insert (the probe) engaged in the workpiece during welding, and 3) the region of the probe which roughly coincides with the placement of the abrasive in the workpiece sample. In terms of the distance along the x-axis of the profile trace plots, region (1) spans x=0" to approximately x=0.27" (the exact delineation between the probe and the unused portion of the insert, usually detectable from a spike in the surface height, may vary from specimen to specimen); region (2) ranges from x=0.27" to x=0.505" (in instances where the length of the specimen has been reduced by significant wear, the location of the probe tip may be less than 0.505"); region (3) corresponds to x-values between x=0.395 and x=0.505."

The initial surface texture of the probe is characterized by the profile data collected in region (1). The impact of wear on surface texture can be determined by comparing the metrics $R_a, R_q, R_t,$ and R_z for this region with the value of the surface texture parameters at the probe tip (region 3, which coincides with the placement of the abrasive). The ratio of the value of a particular surface texture metric in region 3 ($R_{i,3}$) to the corresponding value in region 1 ($R_{i,1}$) is denoted as R'_i. The index i can be any of the five letters (a, q, p, t, or z) associated with the roughness parameters described previously. A R'_a value of 45, for instance, means that the surface roughness at the tip of the specimen post-weld is 45 times rougher than the initial (pre-weld) surface. In general, the surface texture parameters in the region of the abrasive are much greater than those in the portion of the insert (region 1) which does not contact the material. The degree to which the surface texture is affected depends on particle size. Figure 102

compares the average ratio R'_a for three classes of particle size: F14, F60, and F150. It is clear that the disparity in the pre and post-weld surface roughness R_a is extremely sensitive to the diameter of the particulate reinforcement. For the smallest particles (F150), the average ratio R'_a across all four sets of process parameters (2000 RPM/5IPM, 1500 RPM/9 IPM, 1500 RPM/5 IPM, and 2000 RPM/5 IPM) is 47. Isolated from the rest of the data, this value may seem large, but it is approximately 30% smaller than R'_a for the F60 class ($R'_a \approx 75$) and represents a 75% reduction from the ratio associated with the largest particle size (for F14, $R'_a \approx 175$). Figure 103 and 104 plot the effect of particle size on the change in the rms surface roughness R_q and the peak to valley roughness R_t, respectively. The relationship between particle diameter and the distance between the average profile height and highest peak R_p (as well as R_z) is shown in Figure 105. For all surface roughness parameters, the magnitude of the disparity between $R_{i,1}$ and $R_{i,3}$ decreases with particle size. This deterioration is most dramatic for R_q. On average, the rms surface roughness for the tool in the region of the F14 abrasive is 30,000 times greater than it was initially; for F150, R'_q reduces to 1600 (a 95% decrease from the F14 class).

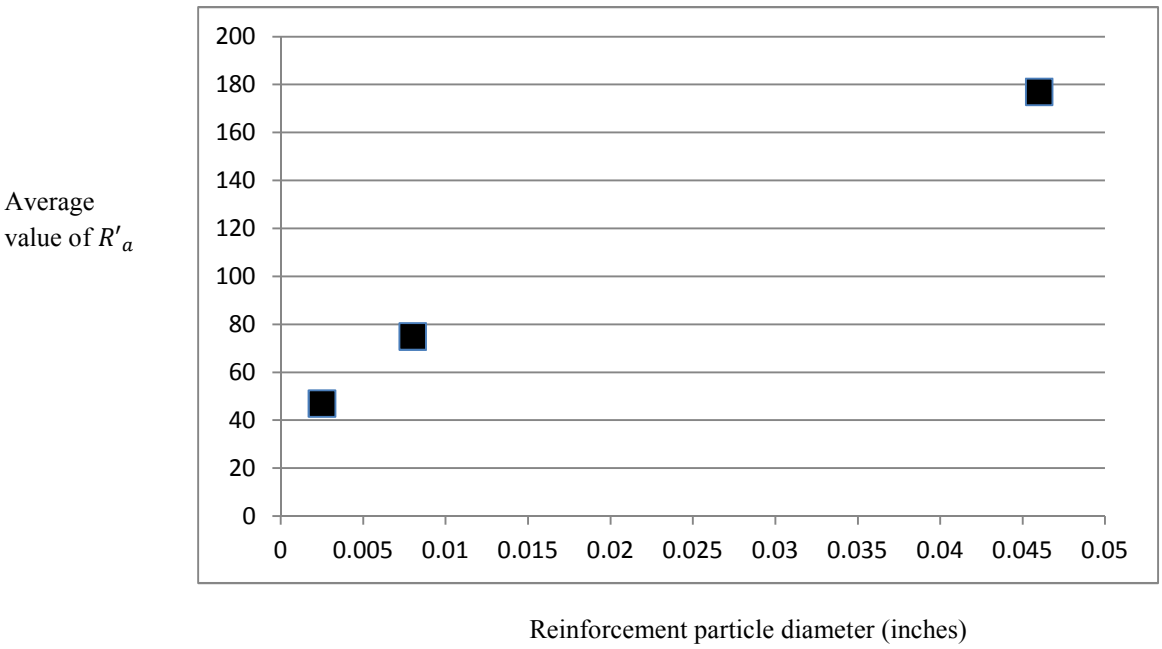


Figure 102 Relationship between particle size and average R'_a*

*Each point on the plot represents the average of the ratio of the surface roughness of the tip of the probe after welding to the initial surface roughness of the specimen for four parameter sets.

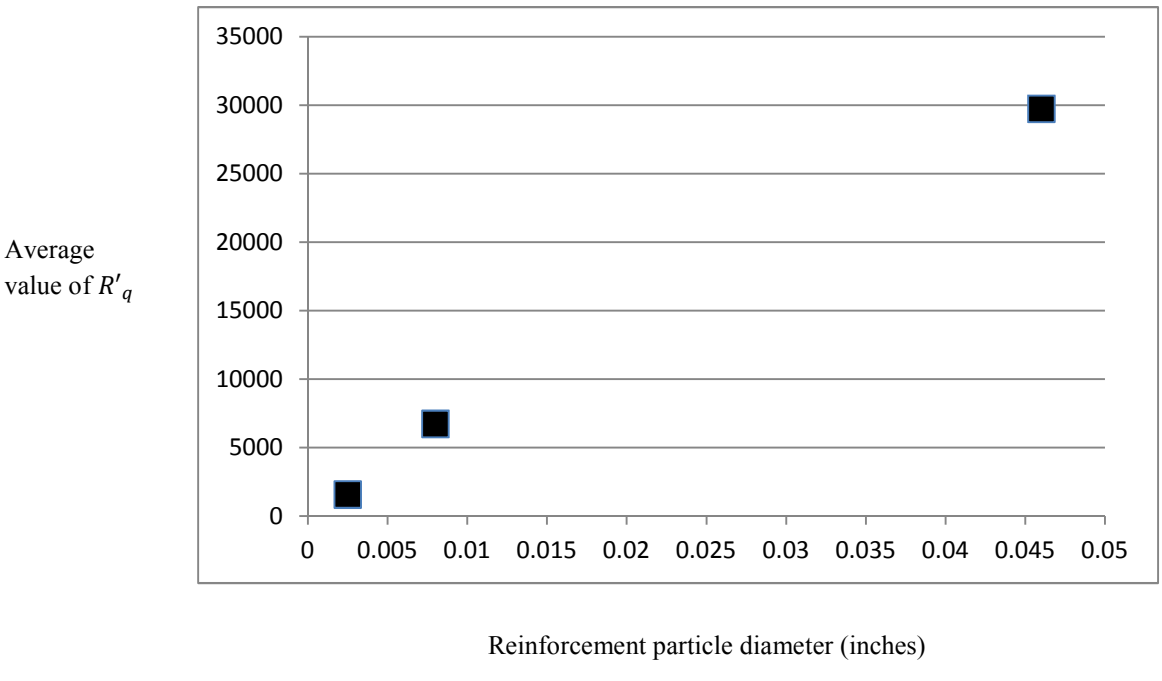


Figure 103 Plot of average R'_q versus particle size

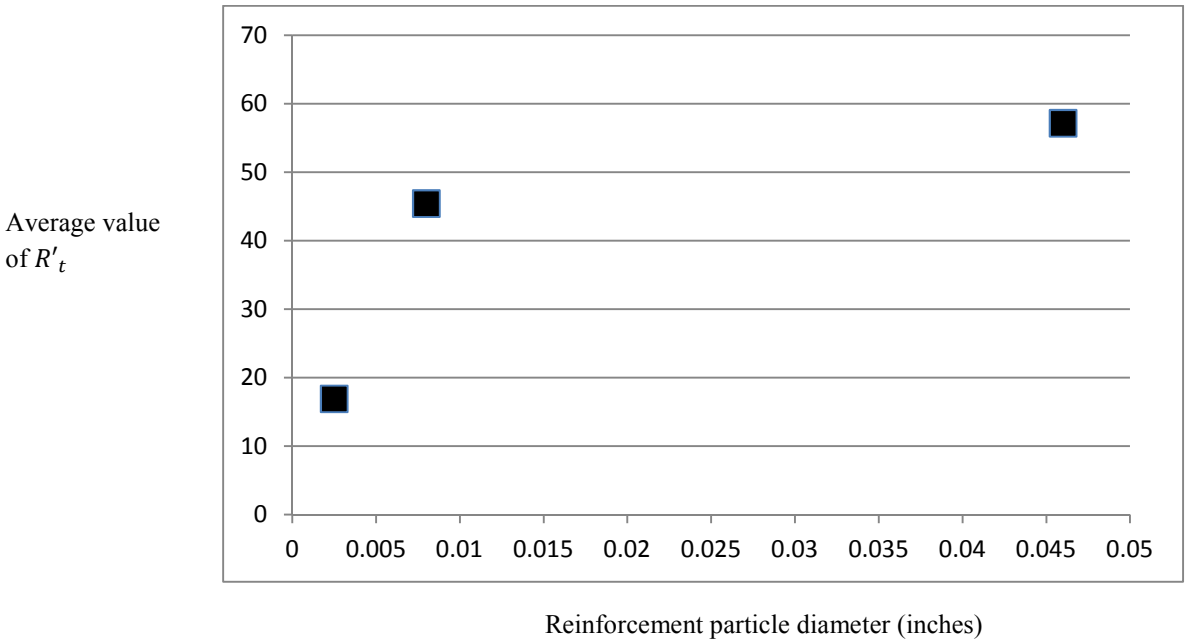


Figure 104 Plot of average R'_t versus particle size

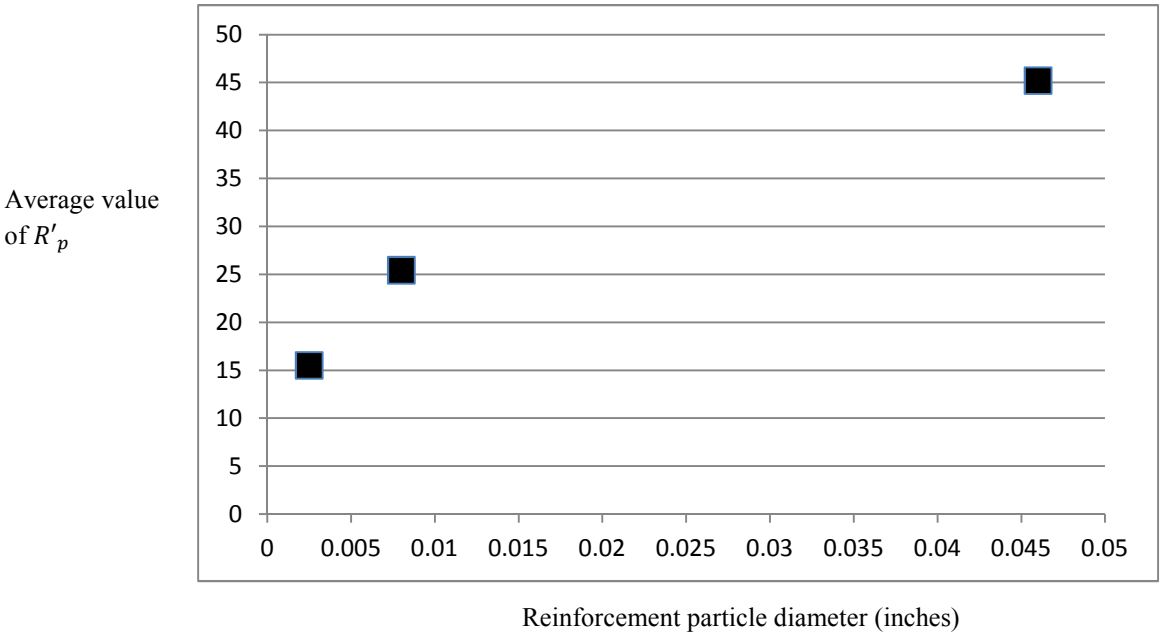


Figure 105 Plot of average R'_p versus particle size

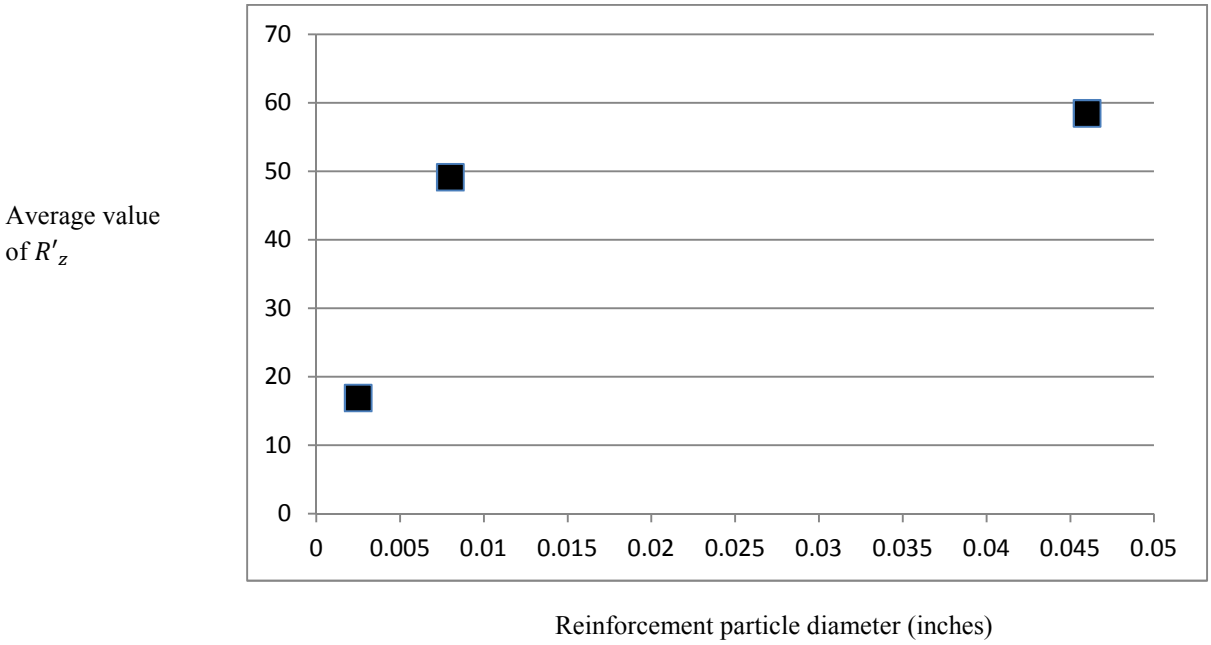


Figure 106 Plot of average R'_z versus particle size

The effect of process parameters on surface texture is less obvious. The surface profile data can be used to test a key observation of the imaging study: that, for a given particle size, different parameters produce specimens with roughly the same topographical characteristics. Consider the plot of surface profiles for specimens used to join Aluminum containing F60 size particles for four process parameters in Figure 107: the profiles are very similar in shape and magnitude, despite substantial changes in weld pitch. Contrast this with Figure 108, which plots the surface profiles associated with the 1000 RPM/5 IPM parameter for three different particle sizes. It is clear that abrasive action is dramatically enhanced by the use of the largest particle size, but is comparatively less affected by changes in process parameters within a given particle class. These observations are consistent with the predictions of the wear model developed in chapter VI – the squared particle size term indicates that wear should be extremely sensitive to the size of the inclusions in the MMC, a prediction which is borne out by the results of this study.

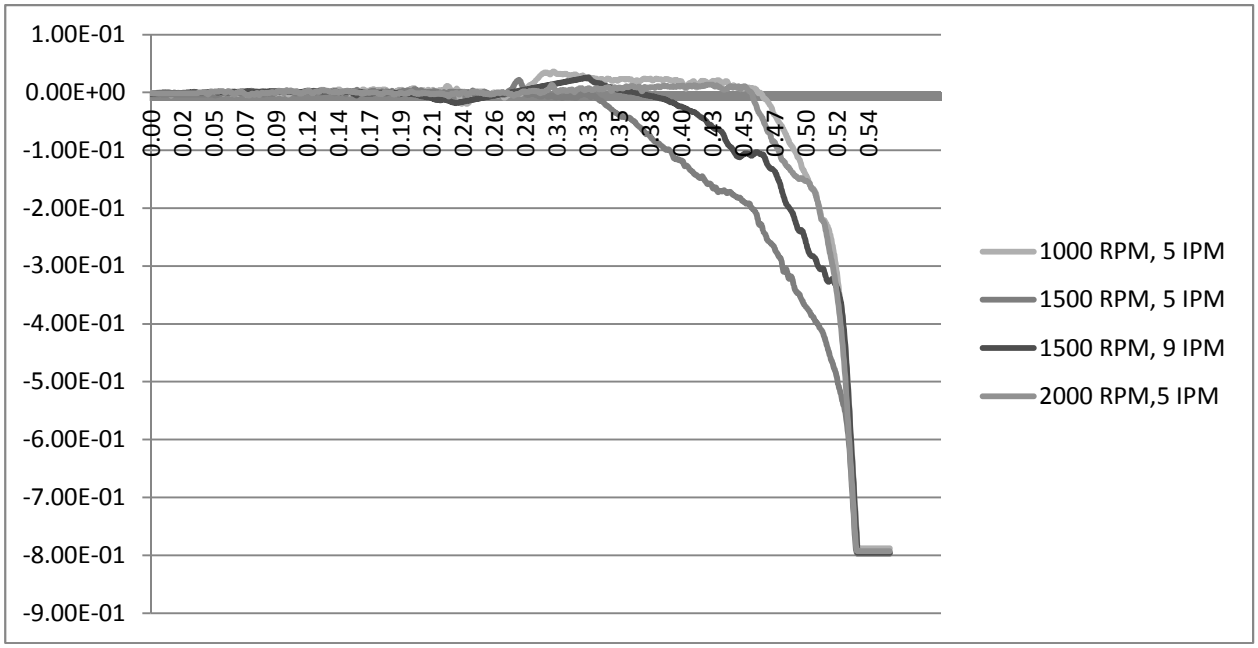


Figure 107 Surface profiles for inserts used to join Aluminum containing F60 SiC particles at 1000 RPM/5 IPM, 1500 RPM/5 IPM, 1500 RPM/9 IPM, and 2000 RPM/5 IPM

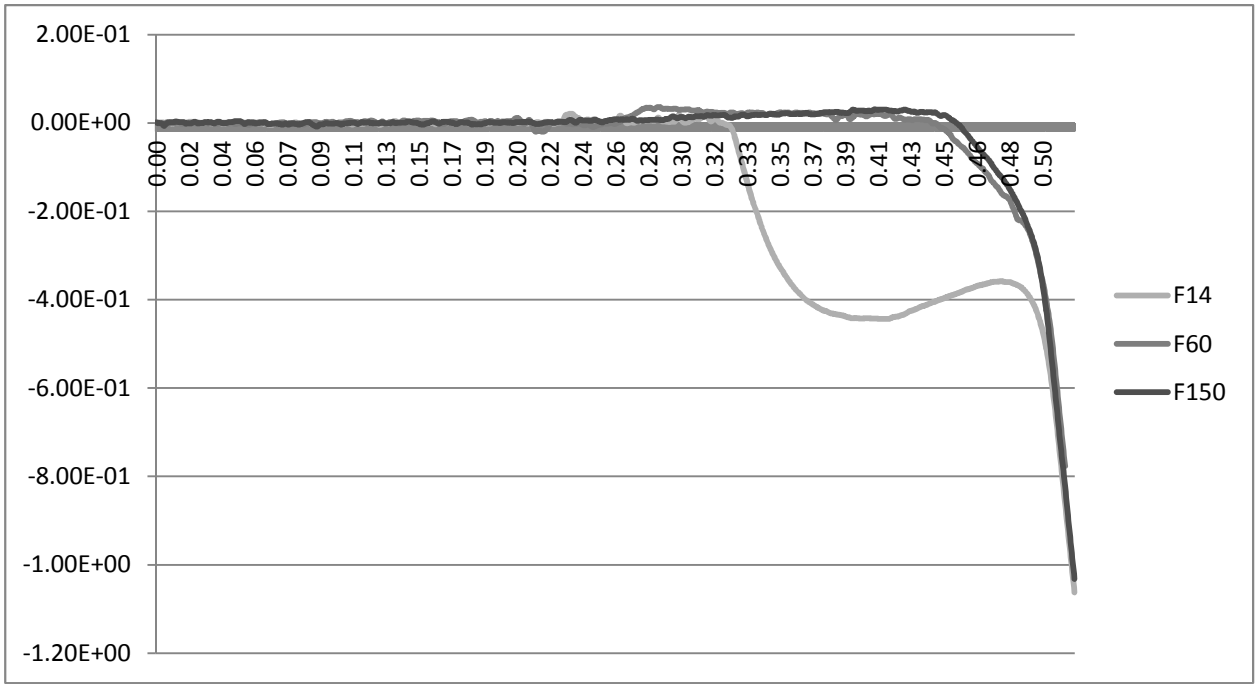


Figure 108 Surface profiles for specimens used to join Aluminum containing F14, F60, and F150 grade SiC particles at 1000 RPM/5 IPM

A note about filtering:

The (x,z) data files output by the Talysurf software can contain as many as 30,000 data points for a trace along the length of the insert. The data density can be reduced through a simple filtering operation known as a moving average. In this technique, the original data is divided into subsets and each subset is replaced by a single value equal to the average of all the numbers in the subset. For instance, if the data set consists of 21,000 points, it can be divided into 2100 subsets with 10 data points each. The data set after filtering is 2100 points (a ten-fold data reduction), where each point represents the average of 10 other points. The x-value corresponding to each average is taken to be the location of the first point in the subset. The filter makes the data set more manageable

but does not significantly impact the shape of the data or the surface texture parameters calculated from it. For example, the average roughness R_a over the virgin region of the tool for the 1000 RPM/5 IPM/ F14 data set before filtering is 1.52E-3. After filtering, the average roughness is 1.49E-3 (which translates to a 2% difference between the surface texture parameter R_a for the pre and post filtered data sets).

7.7 Conclusions

The objective of the work detailed in this chapter is to quantify the effect of particle size on tool wear in FSW of MMCs. The phenomenological model of wear developed in chapter VI for FSW of MMCs predicts that wear can only occur when the reinforcing particle is of sufficient size to span the width of the rotating plug in the x-y plane and impinge on the tool surface, carving out material as the tool rotates past it. The experiments undertaken in this chapter were intended to test the sensitivity of wear to particle size (thereby, depending on the results, either offering support for the plug model or directing us to sounder modeling concepts which more accurately reflect process dynamics). The outcome of the tracer experiments confirm two key predictions of the rotating plug model for wear in FSW of MMCs as described in the previous chapter:

1) The primary wear mode is abrasion. Circumferential, parallel grooves (the hallmark of abrasive action) are clearly visible on microscopic examination of the tool surface post-weld. This grooving is concentrated at the probe tip, where the model predicts that the clearance between the tool surface and the rotating plug is at a minimum, facilitating increased contact between the tool and the much harder reinforcement in this region.

2) Wear is sensitive to changes in particle size. The fundamental wear equation derived from the rotating plug model predicts that wear varies with both process parameters and the diameter of the reinforcement particles (the effect of process parameters was previously investigated in chapter V). The impact of particle size on wear was ascertained using two techniques: weighing and profilometry. To measure the amount of material loss, the pre and post-weld weights of the insert associated with a particular experiment were compared -- any change in mass was assumed to be a consequence of wear processes acting on the portion of the insert engaged in the workpiece during joining. Regression analyses of the data indicate that the relationship between wear and particle size is logarithmic.

The effect of wear on surface texture was evaluated using contact profilometry. Abrasive action at the probe tip creates grooving -- this alteration of the surface by wear manifests itself as an increase in surface roughness. Perhaps not surprisingly, larger particles create deeper grooves (the depth of the grooves in a particular region of the tool is characterized by the peak to valley roughness). The ratio of the post-weld surface texture to the pre-weld characterization of the surface (a metric denoted by R_i') was found to be directly proportional to particle size. The selection of smaller particles for inclusion in MMCs is advantageous because it minimizes the disturbance in surface texture by limiting the depth to which particles can penetrate the tool surface during joining. Specimens which experience a large amount of wear have rougher surfaces. That larger particles produce comparatively greater amounts of wear may seem like an obvious (or even trivial) result, given the disparity in hardness between the tool and the reinforcement and the initial surface morphologies of the materials in contact (the as-

machined tool surface is smooth and rounded, while the particles have sharp edges and asperities). However, the strong relationship between wear, surface texture, and particle size serves to bolster the predictions based on the rotating plug model of wear, suggesting that wear in this process is governed by the width of the plug (and as such, is a shear, rather than drag, phenomena). Abrasive wear seems to occur in the manner we imagined it in chapter VI using Nunes's rotating model for material flow as a foundation.

While there are a few published studies on the variation of tool wear in FSW of MMCs with process parameters, the effect of particle size on wear for this application has never been investigated. Because the particle size is usually selected by the manufacturer without input from the customer, its potential influence on wear is often overlooked. The relationship between particle size and wear, however, can provide important insight into the behavior of the wear mechanism (as we have demonstrated here through our own study of particle size effects). The significance of these results and the experimental techniques used to obtain them extends beyond just FSW of MMCs. The insertion of particles in grooves along the jointline represents a novel and economical method to test the impact of the material properties of an MMC on wear.

While we have relied on profilometry and weighing to analyze specimens from these experiments, the tracer technique opens up possibilities for additional analyses which can provide further insight into material flow. In Colligan's work with steel shot tracers, the workpiece was x-rayed post-weld -- the radiographs clearly indicate the path of the steel shot particles around the tool (Figure 109a). A similar radiographic analysis was attempted for our samples (Figure 109b), but the SiC particles and the surrounding material (Al 6061) are too close in density (3.1 $\frac{g}{cm^3}$ and 2.7 $\frac{g}{cm^3}$, respectively) to produce

sufficient contrast. A transverse X-ray of the weld cross-section also fails to produce the level of clarity documented by Colligan. The size of the density difference can be improved slightly by choosing an abrasive with a larger density (such as Aluminum Oxide, which has a density of 3.69 $\frac{g}{cm^3}$), but the disparity in density is still small enough to render radiographic imaging techniques impractical.

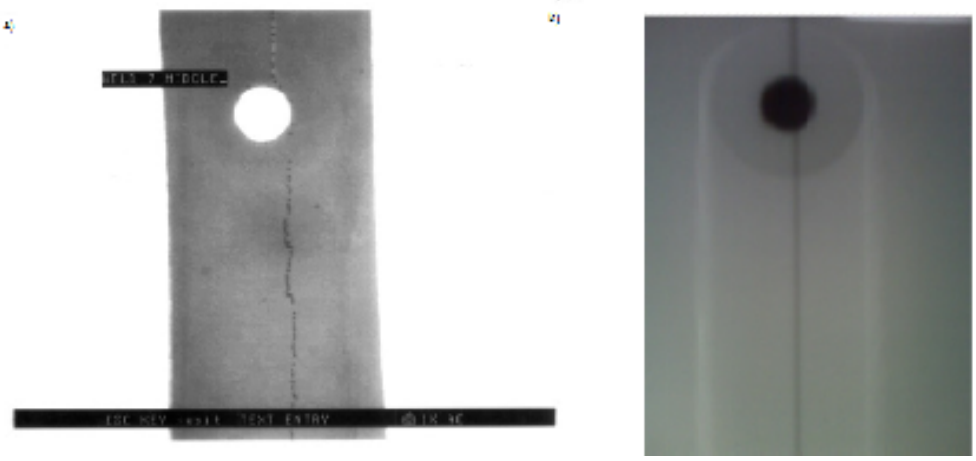


Figure 109 Comparison of plan-view radiographs of welded Al 6061 samples with steel shot tracer (left) and SiC tracer particles (right).

Radiographic techniques are a powerful tool for visualization, but they can only be applied when there is a substantial density difference between the tracer particles and the base material. All ceramics typically used as reinforcements in metal composites have density values comparable to Aluminum. A method which is better suited to the proposed analysis is energy dispersive X-ray spectroscopy (EDX), which generates an elemental breakdown of the chemical composition of a specimen. EDX represents one possible pathway for further study in this area.

CHAPTER VIII

USING HARDER MATERIALS TO COMBAT WEAR

8.1 Introduction

The research documented in the three preceding chapters demonstrates that wear in FSW of MMCs is abrasive. The wear process is somewhat unusual in that it is a shear phenomenon – the clearance between the plug of plasticized material surrounding the tool during welding and the tool surface limits the amount of wear that can occur. The plug functions as a buffer zone, as wear occurs only at locations where the particle is of sufficient size for its radius to span the plug width δ. The growth of the plug is governed by the equation $\delta = \frac{\nu}{\omega}(1 + sin\theta)$. Wear can be mitigated through parameter selection – parameters which create wider plugs (low rotation rates and high traverse rates) are associated with less wear. However, wear cannot be altogether eliminated because of how the plug develops. According to the plug model, there will always be locations on the tool surface (regardless of the rotation speed ω or traverse rate ν) where the clearance between the plug boundary and the tool surface (δ) is zero. This boundary condition means that wear will always occur whenever there are abrasive particles/fibers present in the workpiece material. The previous chapter showed that selection of smaller-size particles reduces wear for two reasons: 1) smaller particles are generally less likely of spanning the width of the clearance δ and b) when small inclusions do contact the surface, they penetrate it to a comparatively shallower depth than larger particles, removing less material. However, even the smallest particles will be able to impinge on

the tool at the locations where $\delta \approx 0$ (strictly speaking, this region consists of the bottom portion of the tool probe in the vicinity of $\theta \approx \frac{-\pi}{2}$, although rotation speeds are high enough in our experiments to circumferential equalize the action of the abrasive particles). Thus careful selection of process parameters and particle size is an effective strategy to mitigate, but not altogether eliminate, wear in FSW of MMCs.

While process parameters in FSW are controlled by the operator, the size of the particulate inclusions in an MMC is usually the domain of the manufacturer. Particle size is selected to facilitate homogenous mixing of the matrix and reinforcement while minimizing curing time and maximizing the improvement in material properties which accompany their inclusion. In general, manufacturers tend to choose particle sizes from the higher end of the FEPA spectrum (the FEPA grade of the particles included in the composites provided by mc21, inc. are F500, which corresponds to a particle diameter of roughly 0.000025"), as these are more compatible with typical MMC manufacturing processes (for example, stir casting or powder metallurgy) than larger particles. Although the results of the preceding chapter demonstrated the influence of particle size on the amount of wear, this variable generally falls outside the control of the end-user unless he or she has the ability to manufacture composites to desired specifications. Small lot production of MMCs solely to produce materials with a specific particle size (unless said size proves critical to material performance) is generally not cost-effective. Additionally, since most manufacturers already use very small particles, decreasing the size further is unlikely to have any substantive effect on tool wear and could, depending on the manufacturing process, contribute to a reduction in the enhanced material properties which make MMCs structurally attractive. The only certain means of

preventing tool wear in an abrasive scenario (short of eliminating the particles altogether) is to employ tool materials which are harder than the reinforcement. The study detailed in this chapter assesses the effectiveness of a variety of materials to combat wear in FSW of MMCs containing both 20% and 30% reinforcement.

8.2 Selection of Candidate Materials

Rabinowicz's principle of abrasive wear for materials in contact states that if material A is harder than material B, material A will be scratch material B but will not be scratched in return [49]. This suggests that the remedy for abrasive wear is simply to select a material which is harder than the material responsible for the wear. Figure 1 compares the hardness of the SiC inclusions with several potential tool materials. There is a large discrepancy in hardness between the conventional tool materials used for FSW (a variety of steels with hardness values in the shaded region) and the ceramics (Al_2O_3 and SiC) used as reinforcement in MMCs. Figure 110 introduces several candidate tool materials which may prove effective in mitigating (or preventing altogether, in the case of materials which are harder than the reinforcement) wear in FSW of MMCs. In our previous research, tools were fabricated from O1 steel, which has a hardness value (post-heat treatment) of approximately 60 HRC. O1 steel is a common FSW tool material and exhibits no signs of wear when used to join conventional (unreinforced) Aluminum alloys. In FSW of MMCs, O1 steel has a high susceptibility to wear, a consequence of the substantial difference (approximately 30%) between the hardness of O1 steel (~60 HRC) and the hardness of the SiC reinforcement (~85 HRC). The surface morphologies

of the materials in contact (the tool surface is smooth and rounded, while the particles are angular and have a high level of asperity) make it easy for the reinforcements to scratch the tool surface when the tool material is much softer than the impinging particle.

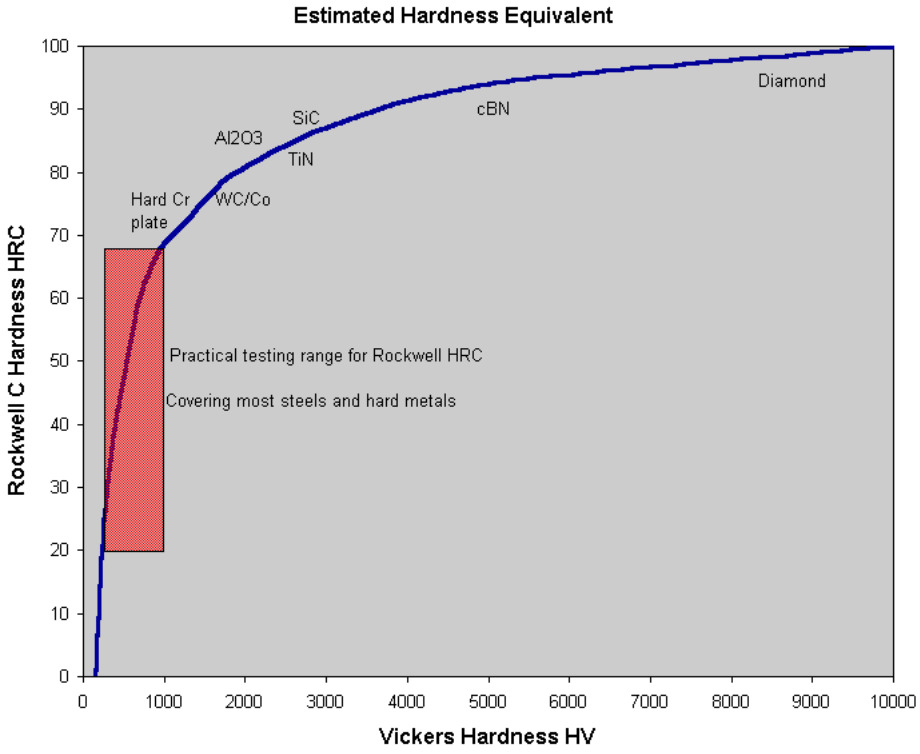


Figure 110 Comparison of hardness values (plotted on both the Rockwell C and Vickers hardness scale) for candidate tool materials [].Figure 24 Chart used to convert between Vickers and Rockwell C hardness scales. The plot can be used to compare the hardness values for tool materials typically used in FSW (denoted by the shaded square corresponding to steels and metals) with the hardness of the reinforcement phase in MMCs (usually either Al_2O_3 or SiC). It is apparent that CBN and diamond, which have hardness values greater than the reinforcement particles, should demonstrate improved wear resistance in FSW of metal composites [123].

Tungsten Carbide (WC) is a potential alternative to the steel/hard metal tools traditionally used as tool materials in FSW. WC is relatively cheap and commonly used to fabricate tools in machining applications where wear is a concern. From the perspective of wear in MMCs, the WC's primary advantage is that its hardness value is more closely aligned with that of the ceramic reinforcement. While the selection of WC as a tool material will not preclude wear (since the reinforcing material is still harder than the tool itself), the parity in hardness should make it more difficult for contacting particles to remove material from the tool surface. Unfortunately, WC in its pure form is very brittle and prone to fracture under the loads encountered in machining (and presumably, FSW). WC can be alloyed with Cobalt to improve ductility and reduce the likelihood of tool failure. Since there is some evidence from machining that the grain size of the Cobalt binder phase influences wear resistance, WC tools with both micrograin and submicrograin Cobalt binders should be considered as candidate tool materials for FSW of MMCs [85]. While it is to be expected that a machine tool made of Tungsten Carbide (WC) used to machine a metal reinforced with Silicon Carbie would exhibit some wear (since WC is only 75 percent as hard as SiC at room temperature), it shows significantly better wear resistance steel tools which are an orde. The relatively low expense of WC along with its improved wear resistance and ductility have made the use of WC drill bits and cutting tools for machining abrasive materials common practice.

Although WC/Co tools have a longer life than the metals typically used for machine tooling, they still exhibit substantial wear in machining of metal composites reinforced with SiC or B_4C [85,79]. For this reason, most researchers recommend that even harder tool materials are essential for this application. The obvious candidate for

wear-resistant tooling is diamond, the hardest known material. While diamond is much more expensive than WC, it represents the best means of eliminating tool wear in machining of ceramics (or materials, such as MMCs, with a ceramic constituent). As an added benefit, harder tools have proved effective in lowering forces during machining; in cutting, reduced forces associated with diamond tools have the potential to improve surface finish as well as enable better overall control of the machining process. The scarcity and expense of diamond in its natural state has led to the development of a variety of processes to grow diamond in a laboratory setting. Synthetic (industrial) diamond has a nearly identical crystalline structure to naturally occurring diamond and can be produced *en masse*. Unfortunately, synthetic diamond (like its counterpart in nature) is still very expensive, owing to the complex machinery and processes required to create continuous structures of this material. Synthetic diamond blanks are very difficult to machine and require sophisticated machining techniques (such as electron discharge machining or laser cutting) to produce net shape parts, further compounding the cost of the product to the end-user.

An economic alternative to monolithic diamond tools is the use of diamond coatings, in which grains of diamond are deposited on a substrate. The diamond technology research group at Vanderbilt University uses Chemical Vapor Deposition (CVD) to synthesize diamond particles from methane (CH_4) and hydrogen gas (H_2). In this process, CH_4 and H_2 are dispersed into a high-temperature microwave chamber. A series of activation reactions takes place within the chamber: molecules of H_2 are split into two Hydrogen atoms, which then react with methane to produce methylium (CH_3) and hydrogen gas. After some time, this flow of reactions results in diamond grains,

which then diffuse and deposit on the surface of the substrate, a process illustrated in Figure 111. According to Deryagin and Fedosayev, a diamond nucleus is formed when the surface concentration of carbon on the substrate reaches some critical value (Figure 112). The mechanism of diamond film formation is a field of research unto itself [124]. The reactions which govern the formation, deposition, and growth of diamond grains in CVD are still being explored and there is no single accepted model for how these processes occur. The manner in which the film forms can vary depending on the chamber conditions, the substrate material, and the geometry of the part to be coated.

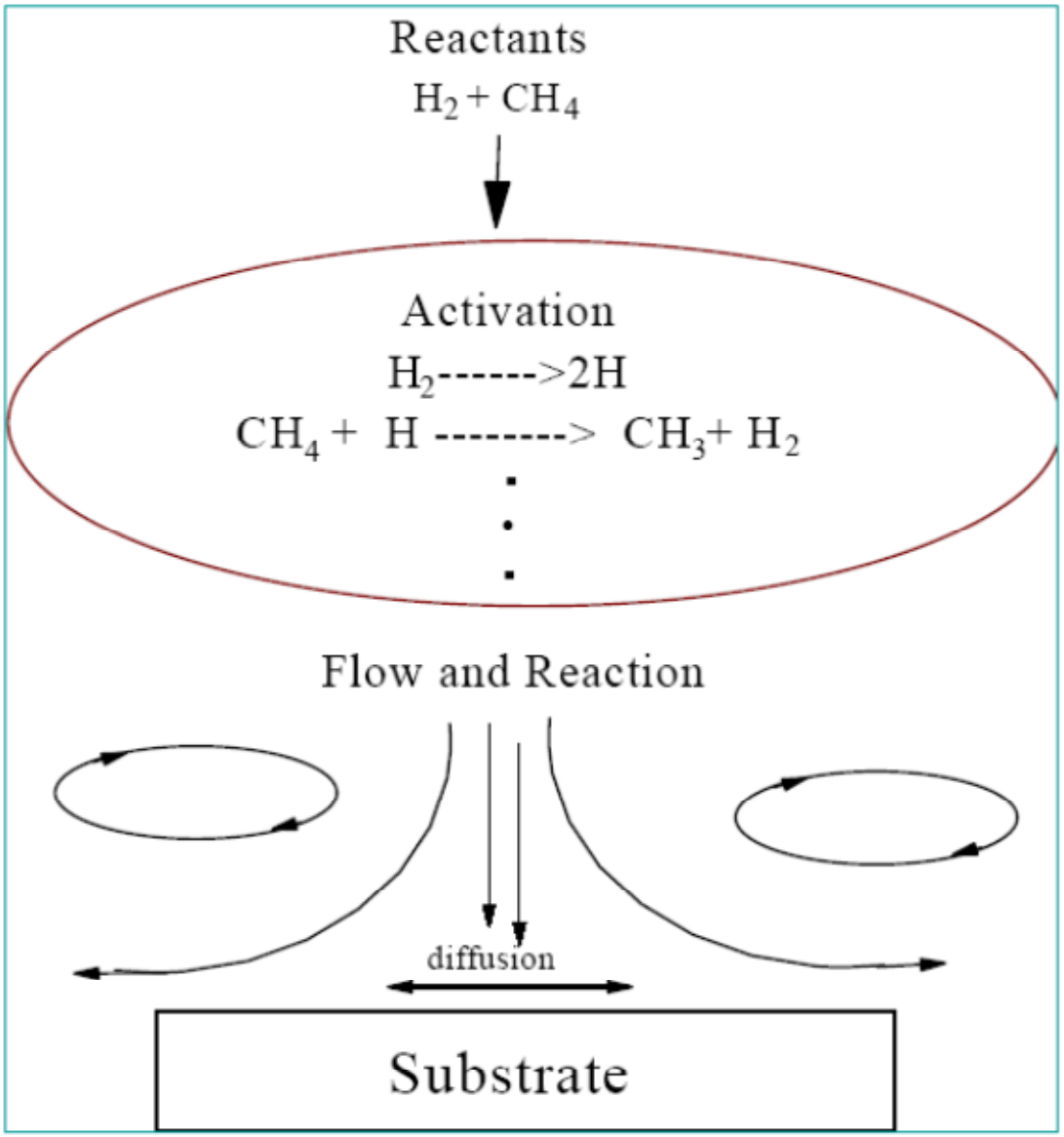


Figure 111 Schematic of CVD deposition process [125]

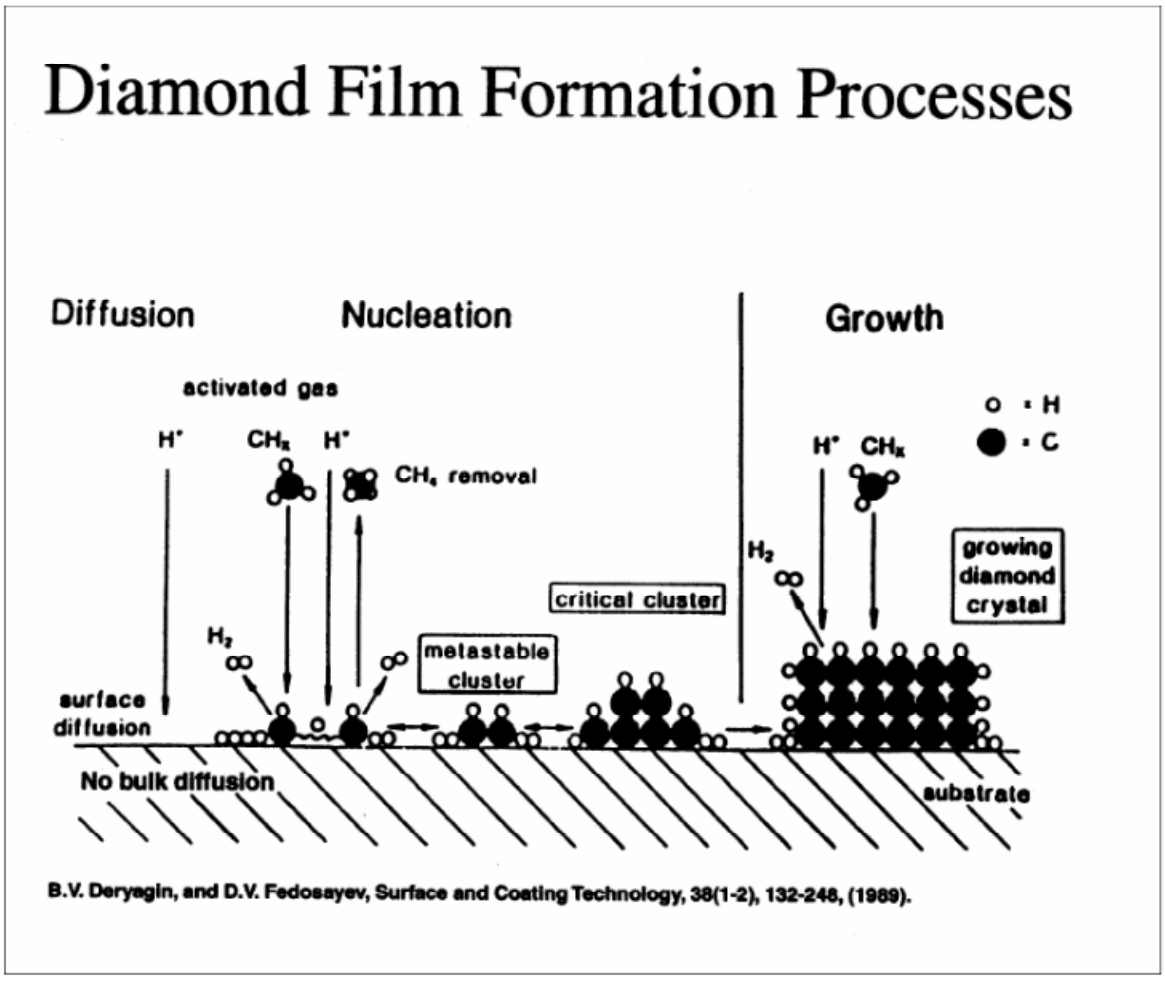


Figure 112 Deryagin and Fedosayev model of diamond film formation [124]

The application of diamond technologies to FSW is fraught with challenges, in large part due to the challenges encountered in substrate selection. During CVD, the sample is exposed to microwave radiation of 500 W and experiences temperatures ranging from 800 to 1200 degrees Celsius, conditions which can at least soften most hard metals conventionally used in FSW tooling. A larger concern, however, is that carbon-

based substrates such as steel are fundamentally incompatible with a deposition process such as that shown in Figures 111 and 112 -- the efficacy of these processes hinges on the adherence of carbon to the substrate surface in amounts that are sufficient to initiate diamond nucleation and growth. Since steel tools cannot be coated using CVD without modifying the substrate surface, we shift our consideration to materials which a) can withstand the conditions of the coating chamber and b) have an affinity for carbon.

In electronics, Molybdenum wafers are commonly coated with diamond using deposition techniques. Our first attempts at applying diamond technology to FSW of MMCs used Molybdenum as a tool material. Since the FSW tool designs compatible with the VUWAL apparatus are too large to fit in the CVDD chamber in their monolithic form, a two-part tool design was adopted. The pin and shoulder of the tool were machined from Molybdenum by Midwest Tungsten. A cylinder which extrudes from the back of this piece can be press-fit into a tool shank (made from h13 steel) to form a complete FSW tool. A diamond coating was deposited on the probe and shoulder of the Molybdenum piece via CVD. This coating, shown in the SEM images of Figure 113, is only a few micrometers thick and covers the surfaces of FSW tool which come in contact with the workpiece material during welding.

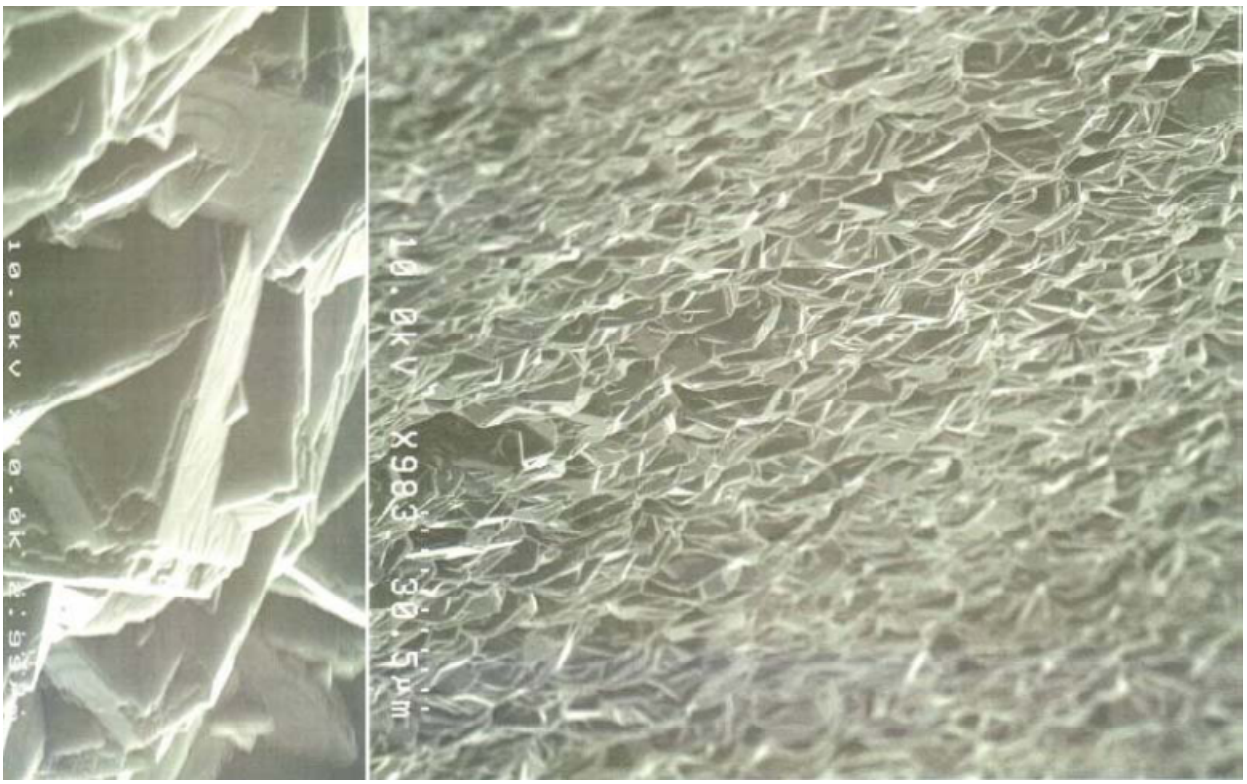


Figure 113 SEM images of diamond coating on surface of Molybdenum tool

While Molybdenum is a highly desirable substrate for diamond coatings, its brittleness renders it prone to fracture. Molybdenum's susceptibility to fracture makes it difficult to produce welds using the coated tool. Failure commonly occurs in the early stages of the weld (usually just after the dwell period, when steady-state traverse speed is reached): the probe shears off cleanly along the plane where the probe extrudes from the shoulder surface. This shearing occurs even sooner in uncoated Molybdenum, suggesting that the coating process actually improves the ductility of the specimen. During coating, the Molybdenum undergoes a temperature cycling analogous to heat treatment, a process which appears to delay the onset of fracture for coated tools. One strategy to combat tool fracture in FSW is to extend the time the tool spends in the dwell phase, the initial portion of the joint where the tool enters the material and traverses at a reduced speed – 0.5

in/min to 1 in/min -- in order to generate sufficient heat to plasticize the workpiece. Adopting this practice can delay (or for the proper parameters, altogether prevent) tool fracture. Another option is to mill a slot slightly larger than the probe diameter at the start of the joint line. This will ensure that the shoulder fully contacts the material before the probe, generating enough heat to plasticize the localized material and reduce the stress on the probe as it creeps into the joint. On the rare occasion when the Molybdenum diamond-coated tools did not fracture, they were successful in significantly reducing process forces, an effect also noted in machining of MMCs using diamond or diamond-coated tools [97]. Ultimately however, the brittleness of the substrate, its propensity for fracture, and the expense of replacement outweighed any benefits in tool performance.

Silicon Carbide (SiC) is another common material for CVD substrates. The probe/shoulder piece of an FSW tool was fabricated from SiC and press fit into a steel tool holder. The brittleness of SiC, however, hastens tool failure. In FSW using SiC tools, the probe shatters upon entering the material, behavior which renders SiC tools highly unsuitable for use in FSW.

Though Molybdenum and Silicon Carbide remain the best materials from a coating perspective (i.e. these materials can withstand the heat of the coating chamber and diamond adheres to them readily), their brittleness makes them inappropriate for FSW applications. Steel was initially discarded as a substrate candidate because of its carbon content, but it is possible to coat steel with diamond if an intermediate layer is deposited first. As a test of this concept for FSW, the probe and shoulder surfaces of an O1 steel FSW tool was coated with Chromium Nitride (CrN). A layer of diamond was then deposited on top of the intermediate CrN layer via the CVD method described

previously. Initially steel tools coated with CrN and diamond exhibits superior wear resistance in FSW of MMCs reinforced with 20% SiC. After 16 inches of weld at 1500 RPM and 5 IPM, the tool probe has lost only 3 percent of its volume, a marked improvement over the 15 percent wear rate documented for a steel tool at the same parameters in chapter V. Unfortunately, the coating is likely to delaminate (separate from the substrate) for welds beyond this length. Once the coating is lost, the wear performance of the tool returns to that associated with the base material (O1 steel). Tools with multiple coatings are particularly susceptible to delamination because the bond between the coatings (i.e. CrN and diamond) is significantly weaker than the bond between the substrate and the first coating (O1 steel and CrN). The occurrence of delamination is apparent when coated tools are examined post-process: for coatings which have delaminated, the tool exhibits substantial wear characteristic of steel tools at the same parameter and SEM images of the tool reveal that the coating has been completely eroded. Delamination also manifests itself in the force data recorded during the weld. As noted in the literature on machining, the use of harder tool materials coincides with reduced process forces. Table 43 compares the average, steady-state axial force values for three segments of a weld of a 20% SiC Al-MMC at 1500 RPM and 5 IPM. For the 0-8 inch and 8-16 inch portions of the weld, the mean steady-state axial force Fz is around 2500 N. In the 16-24 inch region of the weld, the average Fz is 5500 N, a 200% increase over the preceding values. This dramatic increase in axial force indicates that the coating has delaminated sometime during this portion of the process. The relationship between force and the state of the coating is potentially useful from a sensing standpoint, since a sudden spike in forces is consistent with loss of the coating,

Steel tools with multiple coatings are successful in two respects a) the tool does not fracture and b) the tool exhibits very little wear until the coating delaminates.

Table 43 Comparison of steady-state axial force values for three segments of a 20% Al-MMC weld

Length of joint	Force
0-8 in.	2624.435 N
8-16 in.	2365.172 N
16-24 in.	5512.174 N

The exploration of diamond coatings for use in FSW of MMCs has been an intensely iterative process. After several unsuccessful attempts to select a substrate that was ductile enough to avoid fracture yet also facilitate strong bonding with diamond grains grown on its surface, we opted to seek out commercial alternatives. CVD Diamond, Inc. provided probe inserts (like those described in chapter VII) of WC/Co coated with diamond. Details of the coating process are considered proprietary, and as such, we can offer little information about how the coating was obtained. Although the commercial tools were still somewhat prone to fracture (particularly in MMCs with 30% or greater reinforcement), they exhibited the best performance of all the diamond-coated tools considered in our investigation.

8.3 Plan of Experiments

The amount of wear a particular tool will experience during an MMC weld is hypothesized to vary inversely with the hardness ratio H, a dimensionless metric defined as the hardness of the tool (Ht) material to that of the reinforcement (Hr). When H is less than 1, the hardness of the reinforcement exceeds the hardness of the tool (Hr > Ht). For these cases, an increase in the hardness ratio (accomplished by decreasing the hardness of the reinforcement or increasing the hardness of the tool) should correspond to a proportional decrease in the amount of wear the tool experiences. Tool wear cannot occur when the hardness ratio is greater than 1, as the hardness of the tool is greater than that of the reinforcing material (Ht > Hr). Research by Weiner on tool wear emphasizes the importance of the material properties of the tool relative to those of the reinforcement, declaring it to be the determining factor in the abrasive wear process [83]. While the amount of wear can be dramatically affected by changing one or both of these materials, it is generally easier to limit wear by modifying the tool, as the ceramics most commonly used as reinforcements in MMCs all possess similar hardness values.

This study seeks to quantify the variation of tool wear in FSW of MMCs with the hardness of the tool material (and the hardness ratio H). The materials considered are O1 steel, micrograin WC, sub-micrograin WC, and WC coated with diamond, selected based on the exploratory work detailed in section 8.2. As in chapter VII, a modified tool design in which probe inserts can be exchanged was used for these experiments. The tool holder is made of steel –cylindrical rods of each material can be inserted through a bore-hole along the tool's axial centerline and adjusted until the portion of the rod which extrudes from the cylinder coincides with the desired probe length. Probe length is fixed using a

set screw located on the shank of the holder which comes to rest on a flat machined in the insert. A second set screw positioned on the end of the tool shank opposite the probe prevents vertical movement of the insert during welding. Most studies of wear in FSW of MMCs (including our own in Chapter V) conclude that wear of the shoulder is nonexistent or negligible. In the rare instance where shoulder wear is observed, the holder/insert configuration offers a distinct advantage over monolithic FSW tools. In the two-part design, wear of the shoulder does not have to result in an increase in probe length; rather, probe length can be adjusted between experiments to compensate for any shoulder wear which occurs.

The inserts made of O1 steel and WC coated with diamond are cylindrical rod stock measuring 0.25" in diameter and 2.75" in height. The WC inserts, however, are ¼" diameter modified ball end mills: the rounded end (opposite the end used for milling) functions as the probe. Both micrograin and submicrograin end mills were used in the experiments, as there are some studies of wear in machining MMCs which suggest that finer WC grains inhibit wear resistance. Aluminum MMCs in the form of 0.20" thick flat plates were purchased from mc21, inc. Two varieties of MMC were used: Al 359 with 20% SiC reinforcement (by volume) and Al 359 containing 30% SiC reinforcement. The higher reinforcement percentage was included to test the performance of the tools under accelerated wear conditions. Parameter selection, particularly in the case of the 30% reinforcement, proves challenging. While steel tools exhibit rapid wear, they have an advantage over harder tool materials in that they afford a comparatively large operating window. The decreased ductility of harder tools (such as WC or Molybdenum) significantly narrows the range of rotation rates and traverse speeds at which welds can

be performed [97]. The temperature resistance of the MMC workpiece, combined with the brittleness of the harder tools, means that welds must be performed at speeds which will generate enough heat to a) plasticize the material and b) reduce the likelihood of tool failure (while the WC used in these experiments is alloyed with Cobalt to improve its ductility, fracture remains a concern). The high thermal conductivity of WC and diamond relative to the Aluminum matrix further complicates parameter selection. The large discrepancy in the thermal conductivity of the tool and the workpiece means that little of the heat generated by the tool is transferred to the material. The frictionless factor of Debroy et al., f, which characterizes how heat is transferred between the tool and the workpiece during welding, indicates that a diamond-coated tool retains 63 percent of the heat generated [97]. Because such a large proportion of heat goes into the tool, rotation speeds should be increased to maximize heat transfer efficiency and facilitate plasticization. Based on this analysis, the best parameter for these experiments is a rapid rotation rate (>500 RPM) coupled with a traverse speed slow enough to prevent tool failure. Intially, a rotation rate of 1000 RPM was chosen and the traverse speed was set to 5 inches/min. Unfortunately, the WC tool fractured during the weld at these parameters. The traverse speed was gradually reduced (by 0.5 in/min increments) to 3 inches/min, at which point a weld could be completed using the WC tool without incident.

The experiments in this study consist of a series of 14" long butt joints of Al 359 containing either 20% or 30% SiC particulate reinforcement. The length of the probe for each experiment is set at 0.185." Welds are performed at the 1000 RPM/3 IPM parameter. The first 1" of the jointline is set aside for the dwell period (in this region the

tool rotates at 1000 RPM, but traverses at 1 IPM for 1 minute, then increases to the steady-state traverse speed of 3 IPM). The probe length for each experiment is set at 0.185" and the tool is tilted 1 degree with respect to the workpiece. The plan of experiments is summarized in Table 44. There are eight probes (two of each material). Each probe is used to make three welds of either Al 359/SiC/20p or Al 359/SiC/30p. After each weld, the insert can be removed material loss measured to the nearest microgram using the weighing technique described in the previous chapter. Because Aluminum accumulates on the probe surface during welding, inserts must be etched prior to analysis. The insert is immersed in a solution of NaOH and water and allowed to soak until all the Aluminum has eroded from the surface. The measured change in weight can be converted to percent volume loss by exploiting the relationship between mass, density, and the volume of the tool probe (a conversion calculation detailed in chapter VII). Probes are also imaged post-weld using the optics bench constructed for the study in chapter V.

Table 44 Plan of experiments for hardness study

Probe material	Percentage Reinforcement
O1 Steel	20
WC/Co (micrograin)	20
WC/Co (submicrograin)	20
Diamond coated WC	20
O1 Steel	30
WC/Co (micrograin)	30
WC/Co (submicrograin)	30

8.4 Results

The results of the experiments in the hardness study are represented graphically in Figure 5. It is clear from the plot that harder materials experience less wear in both the 20% and 30% reinforcement classes (a trend previously explained in terms of Rabinowicz's "hardness ratio"). As expected, the highest wear values are associated with the O1 steel tools. The wear resistance of WC/Co micrograin and WC/Co submicrograin in FSW of MMCs is vastly superior to that of steel (for the 30% case, replacing the steel probe with WC/Co micrograin reduces the volume loss by a factor of 9!). Not surprisingly, the tool material which exhibits the best performance in terms of wear is WC coated with diamond. Unfortunately, both the diamond coated tool and the WC/Co tools are highly susceptible to fracture during welding. In the 30% class, the diamond tool failed during the second weld – the wear value displayed in Figure 114 represents the observed wear after 14" of weld. The diamond tool used for Al 359/SiC/20p was able to complete two welds (for a distance welded of 28") without fracture. The third weld in this series was forgone to preserve the tool for examination. There were two incidents in which the WC/Co submicrograin insert fractured, but these welds were able to completed using replacement end mills. Because only two diamond coated tools were provided by CVD Diamond, the 30% weld series could not be repeated. The performance of diamond in Figure 114 is slightly inflated since the distance welded with the diamond tools is shorter than that for the WC/Co and steel tools. If we extrapolate the wear for diamond based on the existing data, the diamond coated tool would experience 0.25% volume loss in the 20% reinforcement and 1.53% volume loss in the 30% reinforcement

(values which are both substantially less than the wear documented for other materials in the study).

There are some reports in the literature on machining that wear of WC/Co can be affected by the grain structure. Based on the data in Figure 5114, there does seem to be a slight benefit associated with the use of WC/Co micrograin in FSW of metal matrix composites. The WC/Co submicrograin insert experiences more wear than the micrograin equivalent in both the 20 and 30% reinforcement classes. As in machining, this curiosity can be explained in terms of grain size: finer, submicrograins are more easily stripped from the tool by reinforcing particles than their coarser (micrograin) counterparts. Removal of the WC grains makes the tool more susceptible to abrasion. If wear occurs in FSW of a reinforced composite using a WC tool, it is in part because the abrasive particles are able to remove the cobalt binder phase through adhesion, thereby liberating the WC particles. By this theory, a finer WC grain structure actually accelerates wear. This prediction seems to be bourne out by FSW experiments comparing the wear performance of WC/Co with micrograin and submicrograin structures. The data for the WC/Co tools is summarized in Table 45.

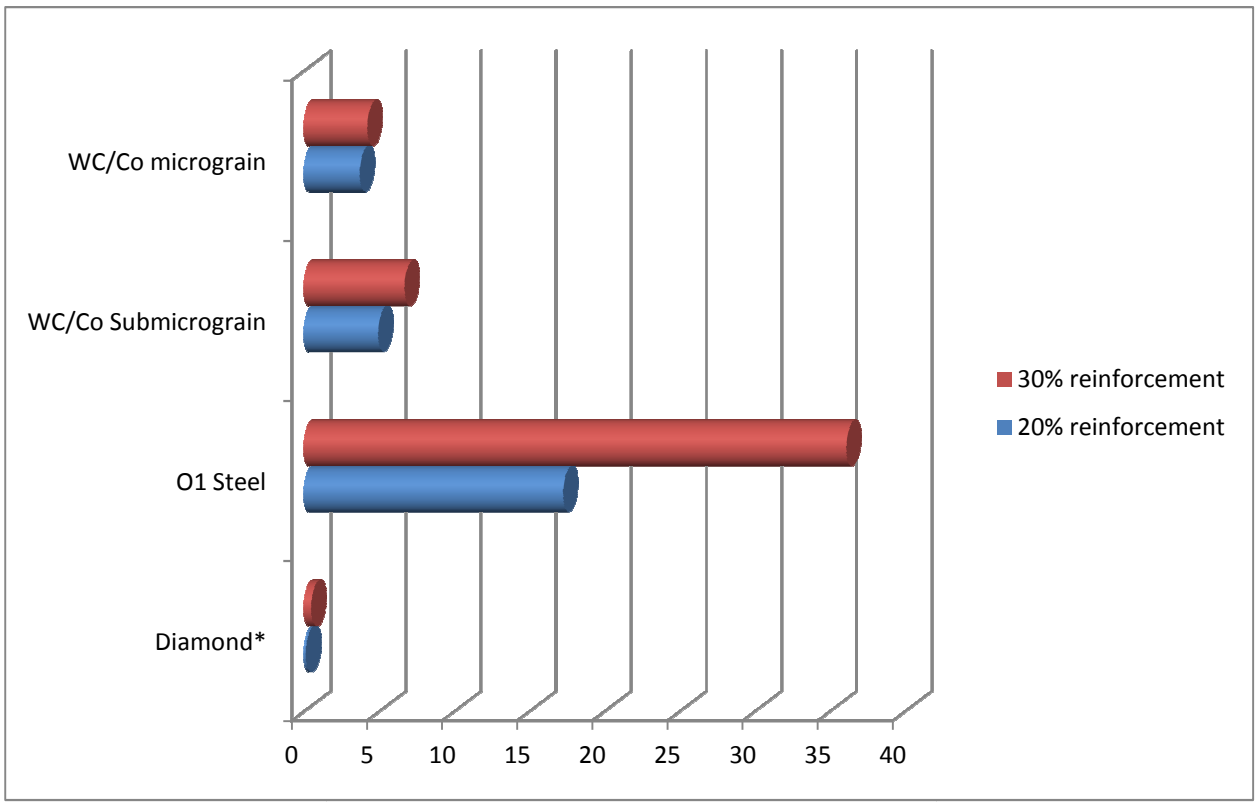


Figure 114 Plot of % wear versus tool material. Each cylinder represents the % wear recorded for the corresponding tool material after three 14" long welds (at 1000 RPM/3 IPM) in an Al 359 MMC with either 20% or 30% SiC reinforcement. *For the 30% reinforcement, the diamond tool fractured during the second 14" pass. The reported wear is the value measured at the completion of the first weld sequence. In the 20% reinforcement, the tool completed two 14" passes without incident.

Table 45 Comparison of volume loss for WC/Co micrograin and submicrograin tools used in FSW of Al 359/SiC with 20% and 30% reinforcement

	20 percent reinforcement	**30 percent reinforcement**
Submicrograin	4.94%	6.76%
Micrograin	3.72%	4.23%

The degree to which an increase in the reinforcement percentage impacts the amount of wear depends on the tool material. For O1 steel, increasing the amount of reinforcement 10% (by volume) produces a 110% increase in the amount of wear over a 42" weld distance. In the case of WC/Co tools, the wear values for the 20% and 30% reinforced composites are very close in magnitude. In WC/Co submicrograin, the 10% increase in reinforcement elevates probe wear from 4.94% to 6.76% (a 37% increase). For WC/Co micrograin, the increase in wear with additional reinforcement is smaller (only 13.7%). The impact of percentage reinforcement on wear is a subject which merits further investigation. The important finding of this study is that the relationship is not 1 to 1 relationship (i.e. it is incorrect to assume that an x% increase in reinforcement necessarily produces an equivalent increase in wear). To illustrate this point, close-up images of probe profiles prior to welding and after completing the series of Al MMC welds with either 20% or 30% reinforcement are compared in Figures 115 through 118 for the steel, WC/Co micrograin, WC/Co submicrograin, and diamond coated inserts, respectively. Wear for WC/Co and the diamond coated specimens is very subtle. Wear (and the increase in wear with percentage reinforcement) is dramatic for the O1 steel inserts.

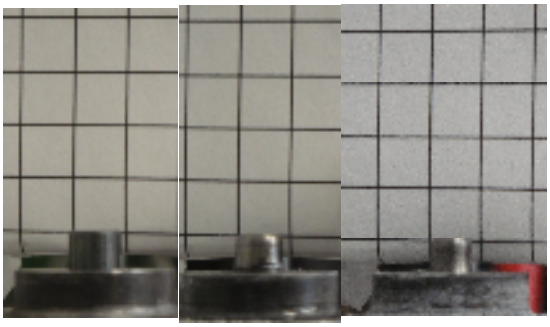


Figure 115 Tool profile for O1 steel (left to right): pre-weld, post-weld of 20% SiC reinforced composite, and post-weld of 30% SiC reinforced composite.

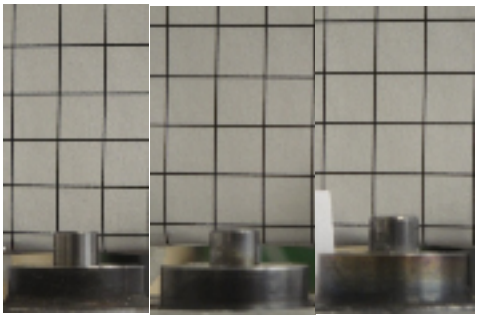


Figure 116 Tool profile for WC/Co micrograin: pre-weld, post-weld of 20% SiC reinforced composite, and post-weld of 30% SiC reinforced composite.

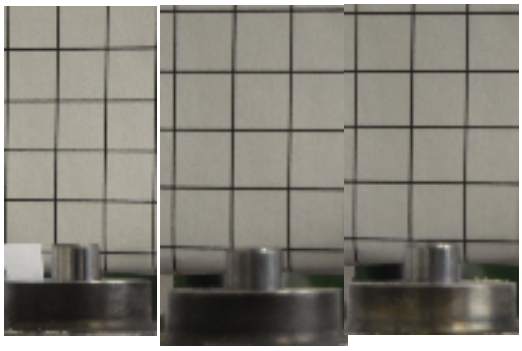


Figure 117 Tool profile for WC/Co submicrograin: pre-weld, post-weld of 20% SiC reinforced composite, and post-weld of 30% SiC reinforced composite.

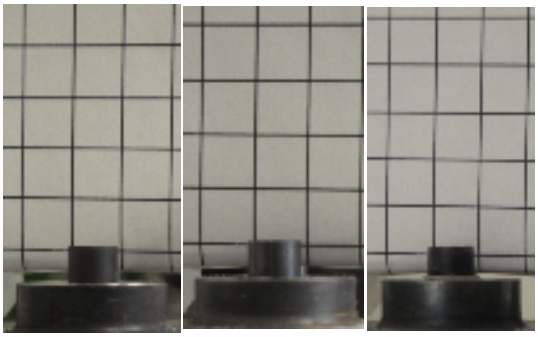


Figure 118 Tool profile for WC with diamond coating: pre-weld, post-weld of 20% SiC reinforced composite, and post-weld of 30% SiC reinforced composite.

8.4 Conclusions

Our intuitive understanding of wear processes involving abrasive particles is that to combat wear, operators must select tool materials with hardness values which approach or exceed those of the abrasive reinforcement. The closer the ratio of the hardness of the abrasive reinforcement to the hardness of the tool lies to 1, the less wear is observed. To combat or altogether eliminate wear, operators should select tool materials with hardness values (such as WC/Co or diamond) which approach or exceed those of the abrasive reinforcement. The concern with such materials is three-fold: 1) they are brittle and may fracture under typical FSW forces, 2) coatings (especially double or triple coated substrates) are susceptible to delamination, and 3) the discrepancy in the conductivities of the tool and workpiece material (when the tool is not steel) can contribute to the formation of surface defects. With regard to (3), this effect is evident in the welds of WC/Co in Al 359/SiC/20p, where a "trench" forms on the weld surface (probably) as a consequence of the mismatch in thermal conductivities. This can be

compensated for by increasing (or decreasing, when appropriate) the heat input to the weld.

This research demonstrates that the wear resistance of the tool in FSW of MMCs can be improved by selecting harder tool materials. The reduction in wear effected by materials such as WC/Co and diamond-coated tools has the potential to expand the usability (and weldability) of MMCs. However, there may be cases where cost is prohibitive or the weld properties attainable with a more exotic tool material are unacceptable. In these instances, the operator will either need to regularly take the tool off-line and measure the amount of wear or have some means of gauging wear in-process. The latter idea is the subject of the next chapter.

CHAPTER IX

SENSING WEAR IN-PROCESS

9.1 Introduction

Chapters V-VIII characterized the influence of several variables on the amount of wear the tool experiences in Friction Stir Welding of Metal Matrix Composites: rotation rate, traverse speed, length of weld, the size of the reinforcement particles, the percentage in which the reinforcement is present, and the hardness of the tool material relative to the reinforcement particles. Wear can be mitigated through explicit control of these variables. For instance, Reducing the rotation rate and increasing the traverse speed can reduce wear, but ultimately the steady, progressive action of the abrasive particles over time/distance will degrade the probe to a degree that necessitates replacement. Another option is to reduce the size of the reinforcement, but this proves difficult since a) particle grades are generally selected by the manufacturer rather than the customer and b) the particle sizes found in off-the-shelf MMCs are already among the smallest available, so any benefits gained from further size reduction are likely to be marginal. The investigation in the previous chapter established the selection of harder tool materials as the best means to combat wear. However, the materials with the highest wear resistance (Molybdenum, Silicon Carbide, Tungsten Carbide) are also among the most brittle, making them highly susceptible to fracture under typical FSW process forces. The process parameters necessary to prevent fracture (high rotation rates coupled with very slow traverse speeds) create what can be an unfavorable thermal environment, as heat

generation beyond what is necessary to plasticize the workpiece has a tendency to produce joints with voids. As the hardest known material, diamond represents the best candidate to eliminate tool wear. Unfortunately, diamond in bulk form is prohibitively expensive and very difficult to machine, requiring sophisticated techniques such as laser cutting or electron discharge machining. The use of diamond coatings is also fraught with challenges: selection of a substrate is difficult, since the materials which diamond adheres readily to (ex. Mo or SiC) are very brittle and prone to failure. Growing diamond on an intermediate coating of a steel substrate is temporarily effective, but the tool's wear resistance reverts to a level consistent with the base material once the coating has delaminated.

The results of the previous chapters indicate that there is no easy solution to the wear problem in FSW of MMCs. Returning to the medical analogy first posited in chapter III, wear for this process is very similar to the progression of a chronic disease. Consider chronic kidney disease (CKD) as an example. In CKD, kidney failure is an inevitability. Though its onset can be delayed with disease management (diet, exercise, medication, interventionary treatments such as dialysis), eventually the patient will die from organ failure unless the kidney is replaced. A transplant becomes necessary when kidney function has deteriorated to a critical level. The presence of waste products in the blood is an indication of the health of the organ. When these waste products (which would ordinarily be filtered out of the bloodstream by the kidneys) attain critical levels, the need for a transplant is imminent.

CKD was chosen as an example because it closely parallels the progression of wear in FSW of high-performance materials. Like kidney failure in CKD, tool wear in

FSW of MMCs is an eventuality. We can slow the progression of volume loss to some degree by selecting process parameters in the low-wear regime or using harder tool materials, but the best strategy is to monitor the amount of wear and replace the tool when it exceeds some critical level. While this "critical level" is not defined since (unlike machining) there are no existing standards for tool wear in FSW, it is regarded for the purposes of this work as the point beyond which degradation in tool shape negatively impacts the structure of the finished joint. The level of wear that coincides with the formation of voids (a consequence of reduced vertical flow associated with the erosion of tool features) is expected to vary widely with FSW applications.

Currently, wear of the tool in FSW is evaluated off-line using weighing or optical techniques. This chapter investigates the feasibility of using process signals (force or torque) as an indicator of the amount of wear the tool has experienced. In-process sensing of wear presents an alternative to offline evaluation of the tool, a technique which is more disruptive to the welding process than interchanging tools only when wear has grown severe enough to necessitate replacement. Our attempt to determine what relationship, if any, exists between process signals and wear builds on the previous work of two researchers: Brian Gibson, who simulated control of wear in FSW of MMCs based on the torque signal, and Russ Longhurst, who explored torque control as a means to improve weld quality [126].

9.2 Exploring the Relationship between Torque and Wear Using the Rotating Plug Model

An expression for the steady-state torque experienced by the tool during welding can be derived based on Nunes rotating plug model (summarized in chapter VI) [101]. Nunes represents the total torque as the sum of the contributions from the pin bottom, pin sides and shoulder (equation 9-1). The torque on the pin bottom is calculated as $\int_0^r 2\pi r^2\,\sigma\,dr$. Force in this equation is represented as the stress σ multiplied by the probe area $2\pi r^2$ ($\sigma = \frac{F}{A}$). The distance at which the torque acts (the moment arm) is dr. $T_{pin\ bottom}$ is thus the sum of the contributions $2\pi r^2 \sigma dr$ as the probe radius *r* varies from $r = 0$ to $r = r$. Similarly, the torque which acts on the shoulder ($T_{shoulder}$) can be written as $\int_r^R 2\pi r^2\,\sigma\,dr$, where R is the radius of the shoulder ($R > r$). The torque for the pin sides ($T_{pin\ sides}$) is $2\pi r^2 t\sigma$, where *t* is the length of the probe, Substituting these components of torque into T_{total}, evaluating the integrals, and simplifying yields equation 9-2.

$$T_{total} = T_{pin\ bottom} + T_{pin\ sides} + T_{shoulder} \qquad \text{(equation 9-1)}$$

$$T_{total} = \frac{2\pi R^3}{3}\left(1 + 3\frac{r^2 t}{R^3}\right)\sigma \qquad \text{(equation 9-2)}$$

As in chapter VI, the shear flow stress in the torque expression is approximated from equation 9-3. β is a material constant, T_m is the melting temperature of the workpiece,

and T is the weld temperature at the shear surface. According to Nunes, the temperature at the shear zone (the thin region just beyond the rotating plug where deformation takes place during FSW) can be calculated from equation 9-4.

$$\sigma = \beta(T_m - T)^2 \qquad \text{(equation 9-3)}$$

$$T = 530°C - \sqrt{\frac{9765(T-30°C)}{\omega}} \qquad \text{(equation 9-4)}$$

The calculation of T_{total} can be broken down into three steps:

1) Solve equation 9-4 for T
2) Substitute T into equation 9-3 to calculate the associated flow stress σ
3) Substitute σ into equation 9-2 to obtain T_{total}

The expression for torque derived from the rotating plug model (equation 9-2) thus predicts that the tool geometry has a significant impact on torque.. According to this formulation, the t radial deterioration of the probe during joining of MMCs by FSW should correspond to a decrease in the magnitude of the torque experienced by the tool. This result raises the possibility as to whether *in situ* estimates of wear can be made based only on the torque signal. Unfortunately, the torque signal is also sensitive to flow stress, temperature, and plunge depth as well as geometry. These additional (and sometimes coupled) dependencies make it difficult to isolate changes in the torque signal which can be attributed solely to the radial loss of tool material. The experimental work detailed in section 9.3 seeks to: 1) collect and analyze process signals from welds in which wear

occurs and 2) offer an assessment as to whether changes in these signals detected in-process can be used to estimate the amount of wear.

9.3 Torque and Wear: Experimental Results

To investigate the relationship between wear and torque, a series of longer welds (with weld distances ranging from 3 feet to 9 feet) were performed. The material selected for the panel welds was an Aluminum 359 alloy reinforced with 20 percent Silicon Carbide particles. The tool design was modeled after Longhurst's prototype FSW tool, shown in Figure 119. The relationship between plunge depth and torque for this tool is linear, a result of its spherical, convex shoulder. The dimensions of the tool are as follows: the spherical radius is 4", the probe is 0.25" in diameter and 0.200" in length, the shoulder has a diameter of 1", and the height of the shoulder is 0.05." The scrolls on the shoulder surface offer the added benefit of containing weld flash, thereby improving the surface appearance of the joint. There is also evidence that this tool design decreases the sensitivity of the torque signal to the thermal environment. Longhurst found that increasing the plunge depth by 0.001 inches produces a 5 to 6 times greater change in the torque signal than a 1 degree Fahrenheit rise in temperature. Minimizing the impact of temperature on torque is particularly useful for our investigation, which seeks to isolate the effect of tool wear on the weld torque.

Figure 119 Spherical shoulder tool used for panel welds. Left image is a closeup of the spherical shoulder (with scrolls). Right image shows the tool in profile [127].

The optimal parameters for this tool in Aluminum 6061 were determined to be 1400 RPM and 9 IPM. While there is no guarantee that these parameters will produce quality joints in Al 359/SiC/20p, they were chosen to serve as a starting point for this investigation based on their success in unreinforced Aluminum. The weld lengths proposed for the wear and torque experiments (3 feet, 6 feet, and 9 feet) exceed the dimensions of the VUWAL test bed. The largest size panel the apparatus can accommodate measure 36" in length, 10" in width, and 0.25" in thickness. To complete these welds, a routine was developed wherein the tool makes 6 passes of equal length. At the completion of each pass, the tool is translated laterally 1 inch before making another pass in the opposite direction. The total length of the 6 passes in this "zigzag" pattern is equivalent to the weld distance called for in the experiment. For example, a 3 ft long weld consists of six 6" passes; likewise, six 12" passes produce 6 feet of weld; and the longest weld, at 9 ft, can be divided into six 18" passes.

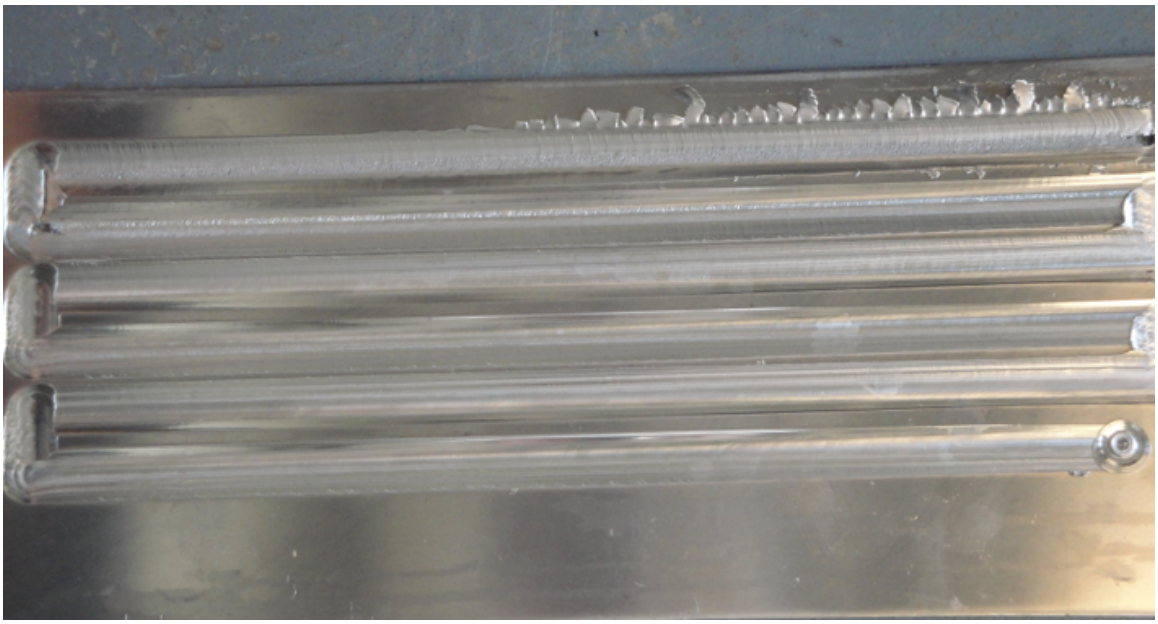


Figure 120 Weld of Al 359/SiC/20p with spherical tool. Weld is 9 feet in length and divided into six 18" linear passes. A 1" lateral movement is used to reposition the tool between passes.

For these experiments, weld torque was measured using the spindle motor current. The current signal is recorded during welding by a clamp-on current transducer (at a frequency of 50 Hz). While the precise relationship between torque and current for the VUWAL test bed is yet to been determined, research by Longhurst suggests that the current (in amperes) is approximately 60 percent of the torque value (in N-m). The plots of current versus time for the 3 ft, 6 ft, and 9 ft welds are shown in Figures 121, 122, and 123, respectively. The values of current recorded during the passes are easily distinguished from the current signal during the 1" lateral repositioning segments. The latter movements coincide with the lower plateaus in the data, while the higher (and

wider) “steps” correspond to each of the six passes whose lengths sum to the specified weld distance. The regions of the weld are labeled for the plot in Figure 3.

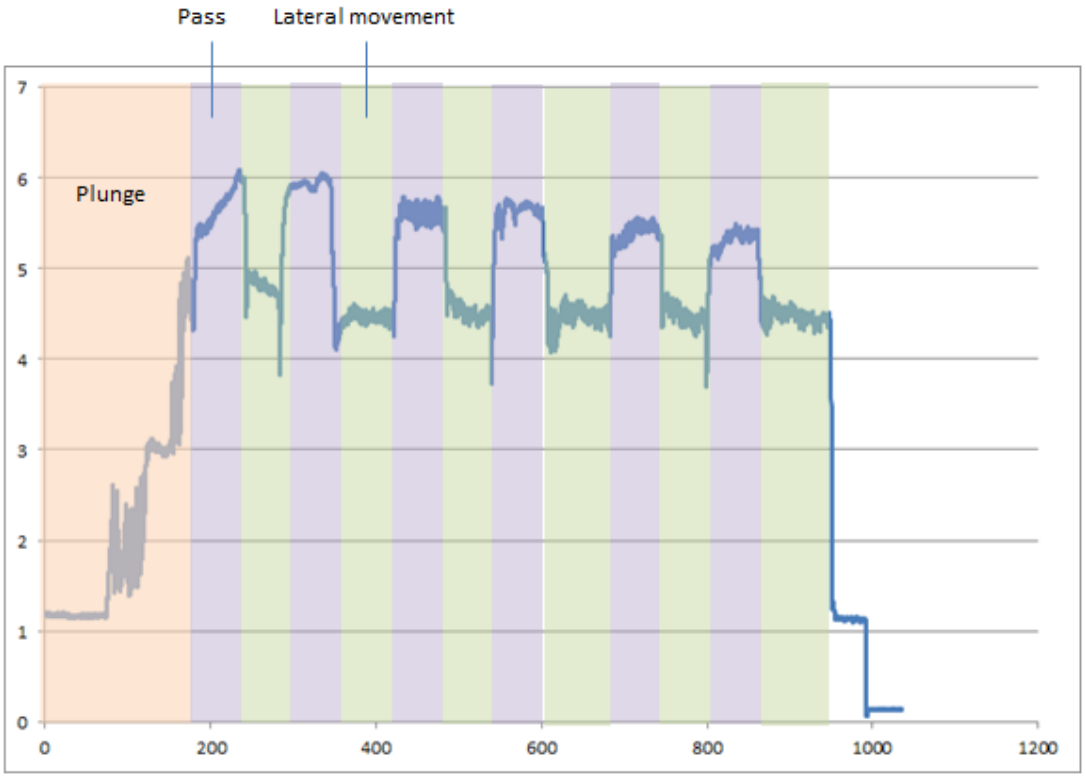


Figure 121 Plot of current versus time (in seconds) for 3 ft weld of Al 359/SiC/20p. Red shaded are corresponds to the plunge, purple shading indicates the 6” passes, and the areas in green represent the lateral movements for repositioning the tool between passes.

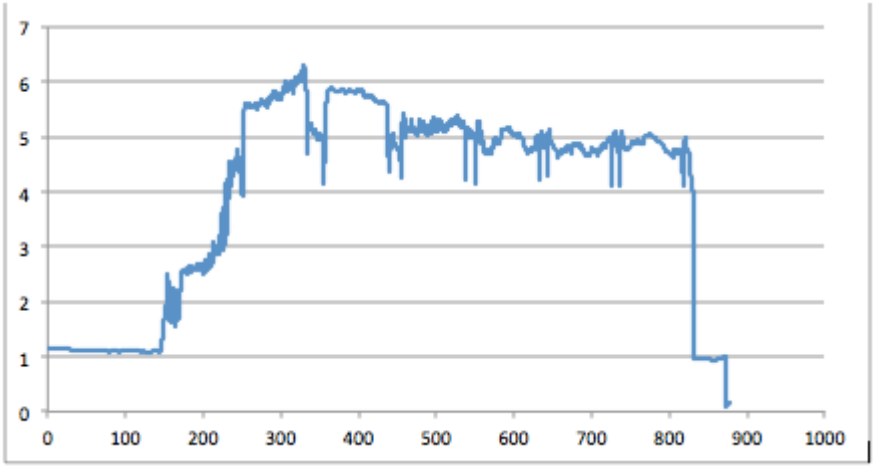


Figure 122 Plot of current versus time for 6 ft weld of Al 359/SiC/20p at 1400 RPM and 9 IPM.

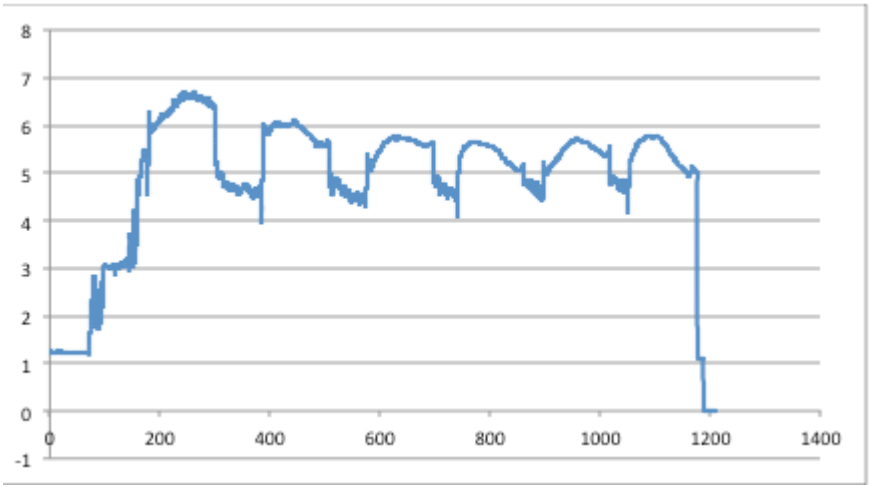


Figure 123 Plot of current versus time for 9 ft weld of Al 359/SiC/20p at 1400 RPM and 9 IPM.

Based on these plots, it appears that the current signal (and thus the torque) in FSW of MMCs experiences a significant decay over time. Since the only variable (aside from the thermal environment, whose contribution to the variation in torque has been minimized through tool design) is the volume of the tool, this deterioration is probably attributable to the erosion of tool material by abrasive action during welding. Figure 124

compares the shape of the three tools used to produce bead on plate welds on the MMC panels. The leftmost image is a spherical tool prior to use. The three images which follow correspond to (from left to right): 1400 RPM/9 IPM/3 ft, 1400 RPM/9 IPM/6 ft, and 1400 RPM/9 IPM/9 ft. The dramatic changes in tool shape evident in these images can be quantified using either weighing or optical techniques. The difference in the pre and post-weld weights of the tool is assumed to be a result of wear and wear is further assumed to only act on the probe. Because the tools in this study are monolithic (the tools in chapter VII had probe inserts which could be analyzed separate from the larger specimen), the difference in weight (even in cases of extreme wear) is very small. For instance, the tool for the 1400 RPM/9 IPM/9 ft panel weld loses only a minute amount of mass (~1 grams) relative to its original weight (~ 200 g). Additionally, the threaded geometry of the probe makes estimation of its initial volume (a value which is needed to convert the mass lost by the tool during welding to a percent volumetric decrease) difficult. Based on these considerations, the optical technique first introduced in chapter V was more appropriate for gauging wear in this application. Close-up images of the tool probe were taken in an optics bench and the cross-sectional areas of the probe pre and post-weld were compared in the imaging software Photoshop. Table 46 tabulates the estimated percent volume loss for each tool based on this method.

It is clear from the images in Figure 124 and the data in Table 46 that wear increases with distance welded. The evolution of tool shape for these welds is distinct from the wear progressions documented in the parameterization, particle size, or hardness studies. where the lower region of the probe was slightly rounded through contact with abrasive particles. For the 6 ft and 9 ft lengths, the region of the probe farthest from the

shoulder begins to resemble a needle – the diameter of the probe at this location is approximately 1/5 the probe diameter at the base of the shoulder. Wear for the panel welds is much more dramatic than the wear observed in previous studies, a result which can be explained by two factors: 1) the weld lengths are substantially (as much as six times) longer than the distances welded in previous investigations and 2) the tools have threads (wear is more pronounced for featured tools). The rate of wear (expressed as the percent volume loss per unit length in the rightmost column of Table 46) decreases with weld distance. As noted earlier, such asymptotic behavior is characteristic of abrasive wear mechanisms; in FSW of MMCs, the decay in wear rate with distance manifests itself in the emergence of the "self-optimized" probe shape discussed in chapter II.

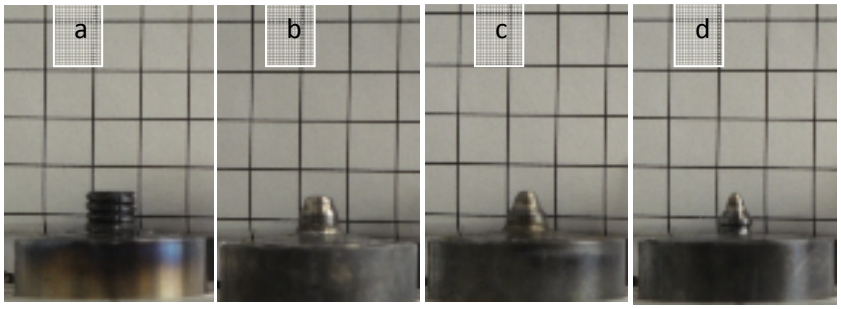


Figure 124 Comparison of tool shape for spherical tools used in panel welds. From left to right: a) virgin tool profile, b) tool after welding 3 ft of Al 359/SiC/20p, c) a tool used to weld 6 ft, and d) tool after 9 ft of weldment.

Table 46 Estimated percent volume loss of probe in panel welds of Al 359/SiC/20p

Weld length	Percent volume lost by probe	Rate of volume loss (% per inch)
3 ft	33.3	0.925
6 ft	39.8	0.552
9 ft	55.6	0.515

The original purpose of these experiments was to cast light on a possible link between changes in torque and wear. In FSW of MMCs, the torque value decreases with time (a trend predicted by the rotating plug model, since T_{total} is directly related to the radius *r*, which decreases with time/weld distance, by equation). But is the magnitude of this decrease over the length of the weld proportional to the amount of wear which has taken place? If we make the convenient (but not altogether valid) assumption that wear is linear, the amount of wear which occurs during each pass can be obtained by dividing the values in Table 46 by the number of passes. For each weld, the volume lost by the tool

can then be plotted against the average current value for each pass. Graphs of current versus distance welded for the 3 ft, 6 ft, and 9 ft welds are shown in Figure 125, 126, and 127, respectively.

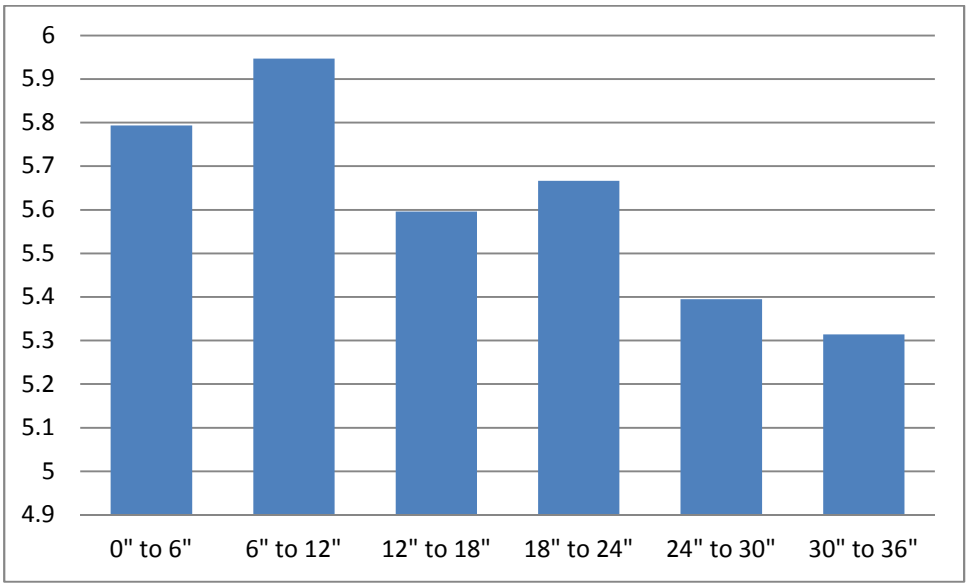


Figure 125 Plot of average current versus distance welded for 3 ft long weld of Al 359/SiC/20p at 1400 RPM/9 IPM.

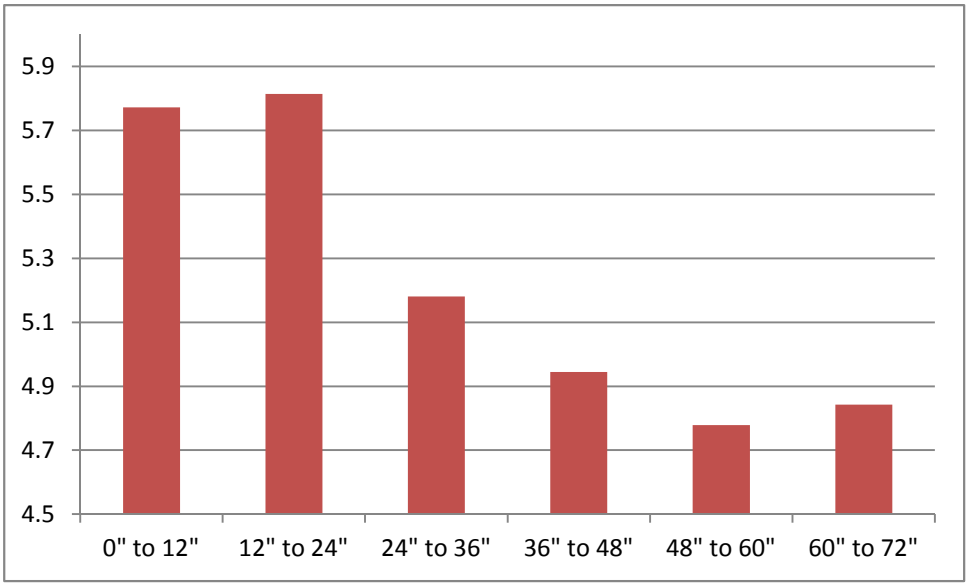


Figure 126 Plot of current versus distance welded for 6 ft long weld of Al 359/SiC/20p at 1400 RPM/9 IPM.

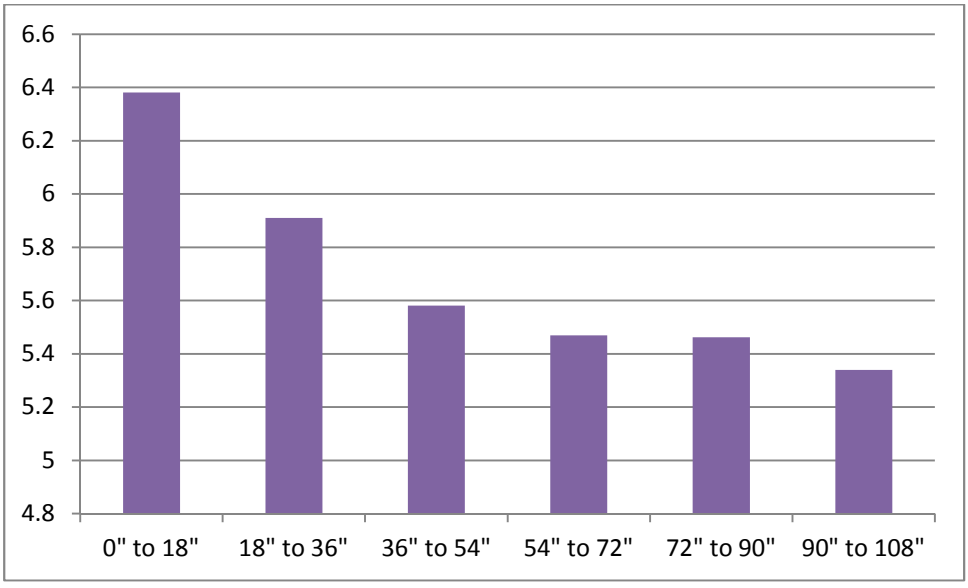


Figure 127 Plot of current versus distance welded for 9 ft long weld of Al 359/SiC/20p at 1400 RPM/9 IPM.

Based on the above plots, there appears to be an inverse relationship between distance welded and current/torque. Since wear is proportional to distance welded, this means that the amount of wear in an MMC weld can be linked to changes in current/torque. This is consistent with the predictions of the rotating plug model: when abrasive particles contact the tool in an MMC weld, the dimensions of the probe (radius and length) are decreased. By equation 9-3, this dimensional deterioration should produce a corresponding decrease in the torque signal. The experimentally observed relationship between torque and wear opens the door to the development of in-process sensing. For an MMC weld, the decay in the torque signal can be correlated to the amount of volume lost by the probe. A "decay function" can be constructed based on empirical data gathered from experiments similar to those detailed in this section. The "decay function" for the 9 ft long weld of Al 359/SiC/20p (shown in Figure 128) is $C(W) = -0.6\ln(W) + 7.6$, where C is the current in amperes and W is percent wear. If the goal of the operator is to keep the wear volume loss below 20% (suppose for the sake of argument that 20% is the level beyond which the reduced vertical flow contributes to defect formation), then this equation indicates that the tool should be replaced when the current value has decayed to 5.8 amperes (a value that is 10% less than the steady-state value for current recorded during the first pass). It is anticipated that decay functions will be highly specific to weld conditions, as the torque signal (and the degree to which it decays as a result of wear) is sensitive to process parameters, tool geometry, workpiece material, and joint configuration, among other factors. The decay function for any FSW process can be determined experimentally using the technique presented here.

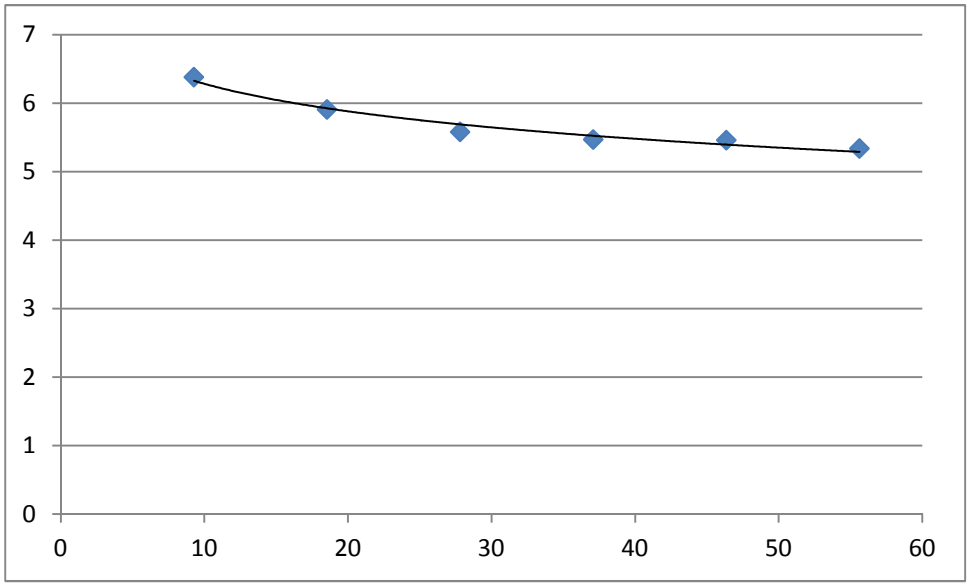


Figure 128 "Decay function"(current versus percent wear) for 9 ft weld of Al 359/SiC/20p at 1400 RPM and 9 IPM. $C(W) = -0.6\ln(W) + 7.6$

9.4 Moving Forward

The observed relationship between current/torque and wear mirrors that predicted by the rotating plug model. The link between wear and torque opens the door to in-process sensing of wear using process signals. *In situ* wear estimation is valuable because it can provide information about tool condition without off-line measurement of tool shape, enabling the operator to initiate replacement of the tool before wear progresses to a level that will affect the quality of the joint.

A potential issue is that when the deterioration in the process signal indicates that the tool has experienced enough wear to warrant replacement, extraction of the tool from the sample introduces a keyhole. This defect is unavoidable since the tool must be

extracted from the material before it can be exchanged for a new part. If an in-process wear sensor were to be implemented in a production setting for FSW, we propose two solutions to this problem: either weld over the keyhole (i.e. after replacement the tool is reengaged in the material but begins its traversal slightly before the keyhole, welding over the point of retraction) or eliminate the keyhole defect altogether by using a retractable pin tool (RPT) or a conical tool [128-129].

The decay of the torque signal as the tool loses volume presents a potential avenue for control of the wear process. Gibson et al. simulated adaptive torque control for the purpose of wear estimation in FSW of MMCs [130]. The proposed controller estimates the probe radius based on the torque signal and adjusts the controlling variable (plunge depth) to maintain a constant, desired value of torque. This controller, shown in the block diagram in Figure , would make it possible for the end-user to know the amount of wear that has taken place at any point during the weld.

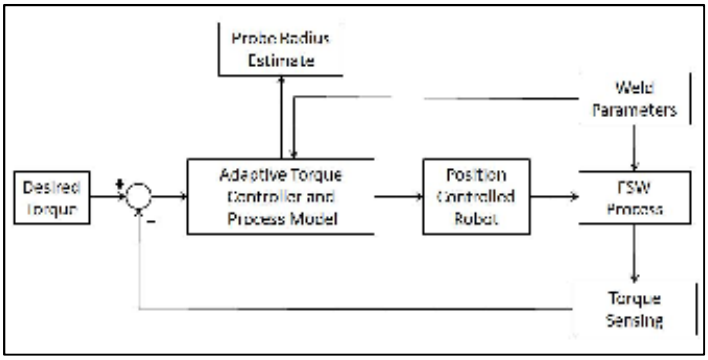


Figure 129 Block diagram for controlling wear using the torque signal [130]

As a preliminary test of torque control in applications where the tool shape is changing due to wear, the torque controller developed by Longhurst et al. for panel welds of Aluminum 6061 was incorporated into a 9 foot long weld of Al 359/SiC/20p. This test weld was performed at the process parameters 1400 RPM and 9 IPM using the same spherical, convex tool described in section 9.3. The controller maintains the value of torque recorded at the completion of the first inch of weldment. As in Gibson's controller design, this constant value of torque is maintained by varying the plunge depth. Longhurst et al. has demonstrated the sensitivity of torque to plunge depth – a major advantage of the convex, spherical tool (in addition to flash containment and enhanced stirring) is the linear relationship between torque and the vertical position of the tool in the material [126]. As the tool wears, the torque value decreases. The controller makes small adjustments in the plunge depth to maintain a desired value. In cases of extreme wear, changes in the torque signal may be substantial enough that the controller initiates a downward motion of the tool which exceeds the limits of the test bed. This is a major concern for the implementation of torque control in systems where wear is present. As a cautionary measure for our experiment, limits were placed on the distance the tool can be moved vertically. Ultimately, wear will effect a decrease in the torque signal and the controller will plunge the tool deeper and deeper in order to maintain desired torque. An 0.003" limit on tool movement in the downward direction ensures that the tool will not contact the backing anvil under any circumstances (the material is 0.25" thick, the probe is 0.200" in length).

While Longhurst's controller proves reasonably effective in ensuring that the surface quality of the entire joint is comparable to the surface over the first inch of weld,

the addition of control has little to no influence on wear. Figure 130 compares the tool used in the control weld (c) with the appearance of the probe prior to welding (a) and after the experiment (1400 RPM/9 IPM/9 ft) in section 9.3. The amount of wear for (c) -- calculated using the optical technique discussed previously – is estimated as 49.8%, a value which is only slightly less than that recorded for the weld without control.

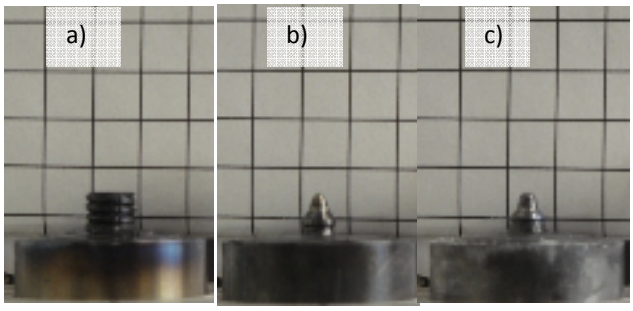


Figure 130 Comparison of spherical, convex tool profile: a) pre-weld, b) after a 9 ft long weld of Al 359/SiC/20p at 1400 RPM/9 IPM, c) after 9 ft long weld of Al 359/SiC/20p at 1400 RPM/9 IPM with torque control.

The controller developed by Longhurst et al. for torque control in FSW was not expected to be completely successful in this application, as the program was tuned specifically for use of the spherical, convex tool in panel welds of Al 6061. It was never anticipated that this controller would be used for MMCs. If torque control were to be either a) used as a means of wear estimation (as proposed in the work by Gibson et al.) or b) implemented in systems where wear occurs, then a new controller designed specifically for this application should be developed. While the block diagram for torque control in FSW of MMCs would be similar to that in Figure 130, it is unlikely that

Longhurst's exact controller could be used without substantial modification, as a controller tuned for a particular set of process parameters/tool geometry/workpiece material is unlikely to maintain full functionality when one or more of these variables is changed. The design and implementation of an adaptive controller for torque control in FSW of MMCs is the natural extension of the initial research into the relationship between torque and wear presented in this chapter.

CHAPTER X

CONCLUSIONS AND FUTURE WORK

Space is weightlifting. For the past half-century, ballistic launch methods have been used to transport cargo from the earth's surface to orbit. Weight was always foremost in Werner von Braun's philosophy of spacecraft design: initially skeptical that lunar missions could ever be accomplished using a one-shot approach (in which crew and cargo would be launched together), Von Braun's proposed launching vehicles carrying crew and cargo separately (an idea which later resurfaced as part of NASA's now-defunct Constellation program) [131]. The lift capacities of vehicles based on current propulsion technologies (multi-stage rockets which rely on conventional fuels and solid-rocket motors for propulsion) vary widely. The Atlas V can transport 64,000 pounds (approximately the weight of a large satellite) into low-earth orbit (LEO), while the Delta IV (a rocket with a similar design) has a significantly smaller lift capacity of 49,000 pounds. For a man-rated space transport system, the space shuttle has a comparatively low carrying capacity (53,000 pounds). The Saturn V was rated for 262,000 pounds of cargo, a figure surpassed only by the proposed Ares V (which, if built, would have had a lift capacity of 410,000 pounds) [132]. These cargo weights comprise only a small portion (less than 5 percent) of the vehicle's weight at launch; the vast majority of a rocket's launch weight is derived from structural components and fuel. The low ratio of cargo to total vehicle weight indicates the inefficiency of ballistic launch systems. One

way to improve this efficiency is by decreasing the vehicle's dry weight (the weight prior to the addition of fuel). The result is an increase in cargo capacity that is equivalent to the reduction in structural weight. Concerns about weight are also rooted in economics, as the cost of a launch is in direct proportion to the weight of the vehicle.

The best way to achieve weight reduction is through the selection of lighter materials. The aerospace industry primarily relies on Aluminum alloys, materials which are lightweight but strong. The development of advanced materials, in particular composites, has opened the door to even lighter structures which satisfy (and in many cases exceed) the mechanical criteria established for flight-rated hardware. In 2001, NASA changed the material used for the space shuttle's external tank from Al 2219 to an Aluminum-Lithium composite (Al-Li 2195) developed by Lockheed Martin, a substitution which reduced the total weight of the external tank by 7,500 pounds [11]. This reduction enabled the space shuttle to transport the heavier components of the International Space Station slated for the transport system's final flights (and gave NASA the option to consolidate multiple components into a single flight, a significant cost savings over the alternative of multiple launches). Unfortunately, the application of traditional fusion welding techniques to join the composite material resulted in mechanically deficient joints. Friction Stir Welding (FSW), a solid-state joining technique pioneered by The Welding Institue (TWI) of Great Britain, was shown to yield defect-free Al-Li 2195 joints with superior mechanical properties. In light of these results, NASA converted much of its manufacturing to utilize the FSW process. In 2006, Al-Li 2195 was selected as the material for the Ares I upper stage and the Orion Crew Exploration Vehicle, both

elements of the Constellation program. All welds on these structures were to be made using FSW.

The selection of lightweight, strong materials continues to be a guiding consideration in design of aerospace vehicles. Since the American space program will be reliant on ballistic launch architectures for the forseeable future, the aerospace industry will continue its push for lighter materials, particularly composites, which may require advanced material processing and welding techniques. While FSW is considered a mature process for many Aluminum alloys (including the 2000, 6000, and 7000 series), there is considerable interest in expanding the process to other materials, such as steels, Magnesium alloys, and even Titanium. Another class of materials which are especially amenable to FSW are metal matrix composite (MMCs). MMCs are dual phase materials which consist of a ceramic reinforcement embedded in a metal alloy (the matrix). They are classified according to the type of reinforcement (reinforcements are typically ceramics such as Silicon Carbide or Aluminum Oxide, but may be in the form of either particulates or fibers), the amount of the reinforcement material that is present (expressed as a percentage of the material's total weight or volume), and the metal alloy which comprises the matrix. The advantages of metal composites lie in their very high strength to weight ratio (which may be more than twice that of conventional Aluminum alloys), temperature resistance, wear resistance, and fatigue life (all properties which make them ideal for use in aerospace structures). Fusion welding of these materials produces joints characterized by porosity and cracking [32]. Additionally, the mechanical properties of the joint are negatively impacted by the presence of Al_4C_3, an undesirable precipitate formed by the reaction of molten Aluminum with the reinforcement [32]. Since FSW

occurs below the melting point of the matrix alloy, the deleterious theta phase is absent in welds produced using this process. The major barrier to FSW of MMCs is rapid and severe wear of the tool, a consequence of the contact between the tool (typically fabricated from a steel alloy) and the comparatively harder reinforcement particles. Progressive wear of the tool removes features which facilitate material stirring, an effect which increases the likelihood of void development.

Though the study of wear is often relegated to the specialized field of material science known as tribology, wear affects virtually every mechanical system in which materials are in contact. In the vernacular sense, the word "worn" describes something that is out of fashion, old-looking, or dilapidated [5]. The engineering definition of wear is somewhat different. According to Rabinowicz, wear is "the removal of material from solid surfaces as a result of mechanical action" [49]. While the term "wear" has a negative connotation in both engineering and ordinary usage, there are rare instances in which it can be beneficial. Polishing and sharpening are examples of wear processes that result in an improved surface texture. In FSW, wear is a significant problem for tools used to join high-strength and/or high-temperature alloys (such as steel, Titanium, or reinforced composites). The wear processes in these scenarios do not serve to enhance the weld, but rather contribute to defect formation in the joint.

An example of a wear problem in aerospace vehicles which (had it gone undiagnosed and unaccounted for) could have resulted in catastrophe is found in the ball bearings used in the turbopumps of the space shuttle main engine (SSME). Each orbiter in the space shuttle fleet is equipped with twelve turbopumps which deliver the liquid oxygen and liquid hydrogen propellants housed in the external tank (ET) to the shuttle's three main

engines. The ball bearings used in the turbopumps are sandwiched between the rotor and the impeller and routinely operate at rotation speeds exceeding 30,000 RPM [133]. The location of the bearings and the extreme operating conditions of the turbopump make these components susceptible to fatigue (attributed to the cyclical loadings the bearings undergo during normal operation) and adhesive wear (precipitated by intimate and prolonged contact between the bearing and adjacent components). Bearing wear can result in the formation of subsurface cracks, which may precede the catastrophic failure of the component. Not surprisingly, bearings fabricated from stainless steel (hardness RC 58) required frequent replacement [133]. Based on the results of a comprehensive tribological evaluation conducted in 2001, engineers at NASA MSFC (Marshall Spaceflight Center) recommended Silicon Nitride as a candidate material for future bearings, a change which significantly increased the life of the component and alleviated concerns about crack formation and propagation during engine operation [133].

The case study of wear in the SSME turbopump bearings illustrates the importance of understanding wear phenomena in mission-critical structures. FSW of MMCs is distinct in that the wear mechanism acts on the tool used to machine the component rather than the component itself. This does not mean that the component is unaffected by wear: as previously noted, erosion of tool features is a contributing factor to void formation. Since large defects typically coincide with a deterioration in mechanical properties, it is important to preserve the tool shape which promotes material stirring and diminishes the probability of a void defect developing. When a defect becomes larger as wear progresses, the potential for mechanical failure of the weld may also increase. Preventing unacceptable defects which stem from wear is critical if MMCs are to be used as

structural materials for aerospace applications, as a failure in any of the welds used to assemble flight hardware could be disastrous. To prevent weld-related structural failures, NASA invests significant time and research into post-process inspection and non-destructive testing techniques (such as ultrasonics and X-rays) to qualify welds of flight hardware. Unfortunately, post-process evaluation of MMC joints produced using FSW is an insufficient means of qualification. Traditional qualification is based on the results of parameterization studies, sets of experiments designed to identify weld parameters which produce joints that meet the criteria for acceptable weld properties established by the researcher. While various parameterization studies for the FSW of MMCs have demonstrated that an operating window of parameters which correspond to defect-free welds can be established for virtually any tool/workpiece combination, these investigations fail to account for defects which may arise as a consequence of the wear mechanism(s) which affect the system. Even if parameters used to produce the MMC joint lie within the operating window, there is always a possibility that a defect may develop as the tool loses volume and the workpiece experiences an accompanying reduction in the flow of plasticized material. For this reason, successful integration of MMCs into aerospace structures requires control of the wear process during joining. Mitigation of wear can be accomplished through control of the process parameters (rotation speed, traverse rate, length of weld, tilt angle) and material properties (percentage reinforcement, type of reinforcement, particle size, and tool material). In instances where wear is inevitable (this may be the case for longer welds or scenarios in which coatings fail to guard against wear), in-process monitoring is needed. As we have emphasized, in-process wear detection is often synonymous with in-process fault

detection for FSW of MMCs. Techniques used to sense wear in this setting may be applied to in –process quality control for FSW of conventional alloys

The ultimate goal of this research is to extend the usability of metal composites to aerospace structures, an outcome which can only be achieved by solving the problem of wear encountered in joining MMCs to themselves or other materials. MMCs possess a wide range of properties which make them well-suited for aerospace applications. Their material properties (strength, fatigue life, wear resistance) represent a significant improvement over both Magnesium and Aluminum alloys [11]. Additionally, MMCs are customizable materials: the constituency of an MMC (the matrix alloy, the reinforcement material, and the percent in which this reinforcement is present) can be manipulated to obtain the desired value of a mechanical property. While the weight of an unreinforced Aluminum alloy and an MMC are virtually the same, the greater strength associated with the MMC reduces the volume of material required for a specific design. The properties of MMCs reflect their status as prime materials for use in aerospace structures, but weldability (not cost) is the primary barrier to their inclusion in aerospace structures. The work presented in this dissertation seeks a fundamental understanding of this wear process, knowledge which is paramount to controlling and/or eliminating wear in FSW of MMCs (and thus expanding the applications in which these materials can be used).

Despite its impact on joining metal composites and other high-strength alloys, little is known about tool wear in FSW. The studies presented in this dissertation serve as a starting point for understanding this neglected aspect of friction stir welding research. The most important aspects of this work are: 1) identification of the wear mechanism, 2) development of empirically based models which predict the amount of tool wear based

on process parameters and material properties, 3) advancement of tool materials and coatings which prevent and/or mitigate wear, and 4) an evaluation the feasibility of sensing techniques which can provide information regarding the condition of the tool in-process. The aim of this inquiry into wear is to develop a fundamental understanding of wear phenomena in FSW and utilize this knowledge to control wear processes, thereby extending the use of these materials to aerospace industry applications where they would be of benefit.

10.1 Significance of the Study

This dissertation presents several separate but related studies on the wear process in FSW of MMCs: modeling the variation of wear with process variables, assessing the impact of material properties of both the tool and workpiece material on wear, and developing an in-process sensor for wear using force, torque, and acoustic feedback signals. Together these original studies represent a significant contribution to the literature on materials joining, composite materials, and tribology. Since FSW occurs below the melting point of the workpiece material, the reaction between molten Aluminum and SiC particles which leads to the formation of the undersirable theta phase in fusion welding of these materials is unable to proceed in FSW. Rather, the primary complication encountered when joining MMCs by FSW is severe tool wear. While abrasive wear processes have been extensively studied for machining of MMCs, there are scarcely any investigations on wear phenomena in FSW of these materials. Reducing (or ideally, altogether eliminating) the consumable nature of tools in FSW of MMCs has the

potential to increase the usability of these materials in applications where their material properties would be of immense benefit. The high strength to weight ratio, temperature resistance, and fatigue life of MMCs make them ideal for aerospace structures where these properties are regarded as mission-critical and often the determining factors (rather than cost) which govern material selection. The major barrier to the integration of metal composites into aerospace structures is the successful joining of MMCs to themselves or other materials (and subsequent qualification of these joints). The contribution of this work lies chiefly in its potential to extend both the applicability of MMCs as well as the classes of materials which can be successfully welded using FSW. In order to do this, we must first forge a fundamental understanding of the wear process in FSW of MMCs. This is challenging because FSW is distinct from previously studied processes in that it involves both structural deformation of the tool and plasticization (but not melting) of the surrounding material. The objectives of this dissertation were to:

1) investigate the variation of wear with process variables
2) determine the dependence of wear on material properties of both the tool and the workpiece
3) construct phenomenological model(s) of the wear process
4) evaluate the feasibility of using process signals to detect changes in tool shape caused by wear

Our work on this problem assumes that the process being studied is largely path independent. In keeping with this idea, most of the material presented in this dissertation is intended as proof of concept. For instance, the action statements indicated in

objectives (1) and (2) do not call for construction of a series of machinability maps which predict the variation of wear for every forseeable combination of tool/workpiece/geometry, but rather seek a "big-picture" understanding of these dependencies and a determination as to whether wear can be predicted and controlled through their selection. We assume that the empirical models constructed through the work in (1) and (2) are representative of trends across a wide range of scenarios (statistical metrics are used to test the validity of this assumption). If a particular manufacturing case requires a more specific model, the path independent nature of the process means that this specific model can be derived using the same methodology used to generate the empirical models presented in the dissertation.

Path independence is a common theme in the literature on parameterization of Friction Stir Welding, having been extensively studied by D. Buford of the National Institute of Aviation Research. Buford suggests that the body of literature on FSW of butt joints is sufficient evidence for path independence since "many aluminum alloys have been joined with a wide variety of weld tool designs. Therefore, an unspecified number of tool designs are expected to make equally sound joints with independently developed process windows" [134]. A similar argument can be made for wear in machining of MMCs: As documented in the literature review on tool wear in machining, there are many studies which use disparate tool designs and workpiece materials, yet observe the same trends in the system response (wear) as the process variables are changed. Applying path independence to our own study significantly narrows its scope, allowing us to focus solely on elucidating the relationships between wear and process variables/material properties/feedback signals which can best inform modeling and

sensing. Inherent in this work is the idea that while we have only considered a few of the hundreds of possible workpiece material/tool material/tool geometry/joint configuration permutations, the trends observed for these cases can be easily translated to other nondescript (but ultimately similar) manufacturing scenarios. Concerns about scalability (i.e. whether these relationships are preserved over ranges which differ from those considered in the initial studies) are addressed through the dimensionless analysis incorporated into objective 3.

The contribution of this work extends beyond just FSW of MMCs. Though many of the techniques used to diagnose and study the wear process in objectives 1-3 are established methods, the particle-tracer experiments represent a unique means of controlling for abrasive particle size (and more economical than the alternative of purchasing custom composites). In a departure from many classical theories of abrasive wear, the wear process in this application is studied and modeled as a shear phenomenon. This approach to physical modeling of systems impacted by wear has the potential to explain wear phenomena in other processes where the behavior of the wear mechanism is not accurately captured by existing models. While the classification techniques applied to the signals that will be collected in objective (4) are not new, this is the first time they have been exploited as a means to sense wear in FSW.

From an industry perspective, the results of this work have practical significance as well. Knowledge of how wear varies with material characteristics and process variables permits operators and material manufacturers to make informed decisions about parameters, tooling, inclusion sizes, and reinforcement percentages which can minimize wear while still producing acceptable joints which satisfy the material properties required

by the application. FSW of MMCs is another area of manufacturing where diamond coatings have the potential to improve performance and prolong tool life (though it remains to be seen whether these coatings are cost effective, since most common FSW tool materials are relatively inexpensive). The primary benefit of coated tools lies in preserving the quality of the weld by guarding against wear. As the probe deteriorates both radially and longitudinally, the likelihood of joint defects increases as a consequence of the accompanying reduction in material stirring. Erosion of material at the tool tip reduces the probe length -- unless this loss of material is compensated for by plunging deeper into the workpiece (which would require both sensing and adaptive control), at a least a portion of the resulting weld will be partial penetration (while partial penetration welds are desirable in some instances, they leave behind a defect known as a "joint line remnant" which may coincide with decreased strength). A control measure which increases the plunge depth to counteract the reduction in probe length runs the risk of excessive shoulder penetration, which produces excessive flash at the weld surface and exacerbates wear in the vicinity of the tool shoulder (a region typically unaffected by abrasion). An in-process wear sensor is a better option than adaptive control for two reasons: 1) it avoids the potential joint issues introduced when the plunge depth is increased beyond its original and 2) it offsets the high costs associated with diamond/wear-resistant tool materials. In cases where wear cannot be eliminated or reduced to an acceptable level by other means, an *in situ* sensor is needed to provide feedback about the rate of deterioration of the probe. The value reported by such a sensor can be used to determine when volumetric loss is large enough to necessitate replacement of the tool. An in-process sensor represents a means of evaluating tool shape that less

disruptive to the weld process than offline examination and gauging of the tool's volume change.

The studies presented in this dissertation are potentially valuable to the fields of industrial manufacturing, tribology, modeling, composites, and welding. This work contributes to a more complete understanding of wear processes in both manufacturing and materials joining.

10.2 Future Work

10.2.1 Wear in FSW of Metal Matrix Composites

Throughout the dissertation, the changes in tool shape that accompany wear have been documented. Imaging techniques were used to monitor (and quantify) shape changes in chapters V, VIII, and IX. In chapter VII, contact profilometry was used to characterize the effect of particle size on surface texture. The evolution of the tool shape with wear, however, has yet to be fully characterized. The Taylor-Hobson Talyrond 295 RSU, which performs three-dimensional mappings of cylindrical surfaces, would be the ideal instrument for such an analysis. In cylinder mapping, the tool is mounted on a stage which can rotate 360 degrees; the part (tool specimen) rotates with respect to a stationary stylus that traces along the tool circumferentially at evenly spaced intervals. If axial traces are desired, the stylus can translate vertically along the tool surface, recording a vertical trace for each n degree increment. The associated software constructs a 3-dimensional plot of these data in the r- θ- z plane (where r is the radius of the specimen, θ represents angular position, and z is the height of the specimen). Cylinder mapping is the

ideal tool for visualizing the evolution of tool shape with wear. Unlike a shadowgraph or longitudinal profilometry traces of the tool surface, this technique provides quantitative measurements of the roundness of the specimen as well as how the circumference varies with height. The data from cylinder mappings of worn tools can be to construct r(z) functions which characterize the variation of the probe radius with vertical distance z from the shoulder for various wear regimes. Shape functions are needed to evaluate whether the self-optimized shape (the tool shape which may coincide with the cessation of wear) is a consistently observable phenomenon or an experimental anomaly. Cylinder maps can be used to generalize the approximate shape of the probe for low, medium, and high levels of material loss. These analyses also inform modeling of the wear process, since the width of the shear zone is predicted to vary in the x-z plane (along the length of the probe) as well as the x-y plane (the latter variation was verified indirectly using the particle-tracer experiments). The change in the width of the shear zone along the z-axis could explain why (in most cases) the region of high wear is confined to the tip of the probe, yet little to no wear is observed in the region nearest the shoulder. In this way, shape functions obtained from cylinder mappings would enhance the phenomenological model for wear in FSW developed in chapter VI.

Chapter IX assessed the feasibility of exploiting changes in process signals to monitor wear in-process. When wear cannot be eliminated (or maintained below some critical level), a wear sensor could provide information about the amount of wear experienced by the tool. An in-process wear sensor presents an alternative to offline evaluation of the tool, a technique which is more disruptive to the welding process than interchanging tools only when wear has grown severe enough to necessitate replacement.

Dr. Nunes's rotating plug model of wear indicates that torque values are proportional to the radius of the tool probe; since most of the volumetric deterioration due to wear occurs radially rather than longitudinally, the torque signal can potentially be exploited for sensing purposes [130]. Further work is needed to 1) correlate the deterioration in the torque signal with the wear for various combinations of tool geometry/workpiece material/process parameters and 2) design and implement a sensor for wear estimation. There may be some additional value in incorporating acoustic signals into work on wear sensing. Process signals collected from successive welds (forces, torque, and, when an accelerometer is present, the acoustic energy *E* and the normalized acoustic power *P*) can be assembled into a field of row vectors with dimensionality $1\ x\ n$. Vectors should originate from experiments performed at the same parameters and using the same tool to ensure that the only factor which distinguishes the vectors collected at the onset of the experiment from those collected later is the linear distance the tool has traversed (and consequently, the amount of wear experienced by the probe). The vector field forms the basis for a classification algorithm which uses process signals to discriminate between two tool states: 1) "worn", which denotes the condition of the tool when wear is severe enough to significantly impact weld quality (for most cases, a tool is considered "worn" once it has experienced a volume loss in excess of 20 percent) and 2) "unworn" , applied to tools with material loss below this cutoff value. The designation of a tool as "worn" or "unworn" in this instance is somewhat misleading. For example, "unworn" does not necessarily indicate that no volume has been lost due to wear; since "unworn" includes all values of wear which lie below a specified value, tools in this category may have experienced some wear, but not enough to be classified as "worn." While this binary

classification scheme lacks the nuance of a continuum approach, it is necessary for discriminant analysis. As an alternative to grouping the wear data into these two broad categories (some may argue this scheme misrepresents the true degree of tool wear which has occurred), the "unworn" and "worn" schema can instead be interpreted in terms of the action statements "The tool does not require replacement" and "The tool needs to be replaced", respectively.

The objective of linear discriminant analysis (LDA) is to construct a function which maximally discriminates between two classes of data (the mathematics which serve as the basis for this technique will be discussed in the chapter on sensing). The result of LDA is a function which identifies the group a particular set of data belongs to based on a feature vector of the form $x = [x_1\, x_2 \ldots \ldots x_n]$. For the proposed analysis, the feature vector would consist of the process signals F_x, F_y, F_z, M_z, E, and P (or some combination thereof). The value that is obtained when this vector is input into the discriminant function $L(x)$ is used to classify it. The classification criteria based on LDA is essentially a piecewise function. If $L(x)$ evaluated for the vector $x = [x_1 = c_1, x_2 = c_2, \ldots \ldots x_n = c_n]$ yields a value that is greater than zero, then the specimen associated with x should be classified in group I; otherwise, it belongs to group II. Mathematically:

$$L(x) = \begin{cases} > 0, & x \in \text{group I} \\ < 0, & x \in \text{group II} \end{cases}$$

The group designations are arbitrary – in this study, group I corresponds to the "worn" condition, while group II represents the class of "unworn" specimens. The LDA function is similar to a regression equation in that it consists of a combination of weights that map

the input data onto a subspace which describes some characteristic of the system. LDA assumes that the independent variables are continuous, normally distributed, and that the classes are linearly separable (in instances where the classes of data do not meet this criterion, a discriminant function may be derived using an ad-hoc solution such as the kernel trick to force convergence). The sensor developed based on this algorithm uses process signals and LDA to make a real-time inference about the amount of material lost by the tool. The accuracy with which the discriminant function classifies the data can be tested by substituting the data points used in the LDA (referred to as the "test set") into $L(x)$ and comparing the classification indicated by the discriminant to the known condition of the tool. The number of misclassifications in the test set is an indication of the predictive accuracy of the discriminant function. A high degree of misclassification in the test set used to train the classifier suggests that LDA is not a reliable method of classification. For systems in which LDA fails to classify the data with an acceptable degree of accuracy, more complex strategies for in-process monitoring and control, such as bin sorting or genetic algorithms, can be implemented.

LDA analysis is more difficult to execute in transient systems, where the value of the feature vector changes in time. The problems inherent in integrating a classifier into the architecture of a dynamic system are not unique to LDA-based sensing. With any classification technique, there will be a delay between the time the "event" (for example, component failure) occurs, the time it takes for this event to produce a detectable change in the process signal, and the time at which the end-user receives an indication from the sensor (based on an analysis of process signals) that this event has taken place. Time lag can be minimized by decreasing the relay time between signal collection and signal

analysis. Increasing the sampling rate decreases the time it takes to accumulate the number of data points needed for analysis (for instance, a feature vector which consists of the average of 500 force and torque readings) another factor which contributes to lag time is the speed with which the computer can manipulate the data into a form compatible with LDA (the raw process signals originating from the sensing elements may require filtering and/or undergo Fast Fourier transforms to extract information which reflects the condition of the component). While lag time can never be completely eradicated (since the classifier needs to collect some minimum amount of data in order to make its determination), efforts should be made to reduce sensor lag, thereby making the classification as close to real-time as possible. The ultimate goal of LDA (and similar classification techniques) is to assess the state of the system in-process and provide information to the end-user about the condition of the component with a high degree of reliability.

10.2.2 FSW of Thermoplastic Matrix Composites

The work in this dissertation could be extended to friction stir welding of other varieties of composites. Another potential advanced material in which FSW would be of benefit is thermoplastic matrix composites (TPMCs), materials used in missile structures as a weight-saving alternative to thermoset polymer matrix composites (which cure irreversibly). TPMCs have carbon fiber inclusions and are currently joined to one another using fusion bonding, ultrasonics, induction welding, or resistance welding. The foremost concern with any of these techniques is that the joining process will disrupt the

orientation of the fibers, which are carefully positioned in the material fabrication stage to bear the anticipated maximum load the member may experience. If the fiber orientation is disturbed in joining, the load along the joint will be carried by the matrix, which could (depending on magnitude of the discrepancy between the load and the strength of the matrix material) result in failure of the component at the joint line. FSW has the potential to join thermoplastics with minimal disturbance of the fibers, resulting in joints with mechanical properties that are comparable to the (unwelded) base material. The shear strength of the lap joint can be evaluated using tensile and fatigue tests in concordance with ASTM standards; the post-weld orientation of the fibers in the cross-section of the welded specimen (and the degree to which this orientation differs from that in the parent material) can be examined using scanning electron microscopy and/or energy dispersive x-ray spectroscopy (which constructs an elemental mapping of the specimen surface). The mechanical data collected through these analyses could be compared against TPMC joints produced using fusion bonding techniques. For TPMCs, FSW may be able to: 1) solve problems associated with fusion bonding of TPMCs (namely deterioration in mechanical properties at the jointline attributed to reorientation or degradation of fibers during joining process) and 2) reduce weight of structure (FSW does not require additional material such as filler or adhesive to join components).

Currently, thermoplastic matrix composites are welded using some variation of fusion bonding, in which the components are pressed together and heated to a temperature greater than the glass transition temperature of the TPMC. The material interface dissolves under the applied heat and cures as it cools, creating a bond between

two initially separate components. Ultrasonic, induction, and resistance welding (which have each been considered as potential joining processes for TPMCs) operate on the same principle: energy (in the form of heat or high frequency vibration) directed into the material induces local melting at the points of contact and subsequent mechanical interlocking at the interface [135]. The problems encountered in joining thermoplastic matrix composites are well-documented. Welds of TPMCs produced using these techniques are often characterized by void growth (a consequence of incomplete bonding at the interface that can be exacerbated by rapid thermal cycling of the substrate), softening of the matrix material (a result of heating the substrate past the melting point to create a bond), and/or reorientation of the reinforcing carbon fibers to an extent which degrades the mechanical properties of the joint [135]. Additionally, the quality of the bond is highly dependent on the surface texture of the interface (necessitating careful preparation of the workpiece surface prior to welding to minimize roughness) [135,136]. Because plastics are nonconductive, it is necessary to place a metal interlayer or metal inclusions (subceptors) between the material interface to absorb and transmit the heat energy in induction welding. These ferromagnetic meshes, analogous to filler material in arc welding, add weight to the structure. While the additional material may seem insignificant, mass in aerospace and defense applications is a critical design parameter, as mass determines both fuel requirements and cargo capacity (and is thus closely tied to cost). Inclusion of metal meshes to facilitate heat transfer can also weaken the bond between components. Most of the problems observed in TPMC joints stem from the large amount of heat that is input to the workpiece material. Since FSW occurs below the melting point (and the glass transition temperature for polymers), the process should

minimize if not altogether eliminate these effects [136, 137]. Because FSW does not require filler material or intermediary conductive layers, the process of joining components does not add any weight to the structure.

While there is historically little to no research on friction stir welding of thermoplastics or thermoplastic matrix composites, related investigations into FSW of metal matrix composites can lend insight into the potential advantages of using FSW to join materials which contain reinforcement. The precedent for the proposed work on TPMCs is a study by Storjohann et al. which compared fusion and friction stir welds of aluminum composites reinforced with either Silicon Carbide and Aluminum Oxide particulates and fibers [32]. In the composites with fibrous reinforcement, Storjohann et al. found that when the fibers were short, discontinuous, and aligned unidirectionally along the x-direction of the workpiece (parallel to the direction of welding), FSW of the composite resulted in only slight reorientation of the fibers along the jointline. The reorientation is most pronounced in the dynamically recrystallized region of the joint (sometimes referred to as the weld nugget), where some of the fibers have been rotated 90 degrees. Storjohann et al. did not perform mechanical tests to determine whether this reorientation was significant enough to negatively impact the material properties of the joint. A GE patent for friction stir welding of fiber reinforced composites suggests that FSW can be used to join TPMCs with minimal disturbance of the orientation [138]. Figure 131, adapted from this patent, compares a plan view of the post-weld fiber distribution for a thermally bonded and friction stir welded TPMC, respectively. In the fusion bonded thermoplastic, the flow of material results in the movement of fibers away from the jointline, leaving the weaker matrix as the load-bearing element. In the friction

stir welded thermoplastic, the fiber distribution in the joint resembles that observed in the unwelded base material – there is no weaker, fiber-depleted region which could contribute to structural failure. While the patent suggests that FSW is best-suited for use with materials where the fibers are initially discontinuous and randomly distributed, work on FSW of fiber-reinforced composites suggests that it would also be appropriate for materials with discontinuous, unidirectional fiber reinforcements. The research proposed here would represent the first application of FSW to thermoplastics matrix composites with fiber reinforcement.

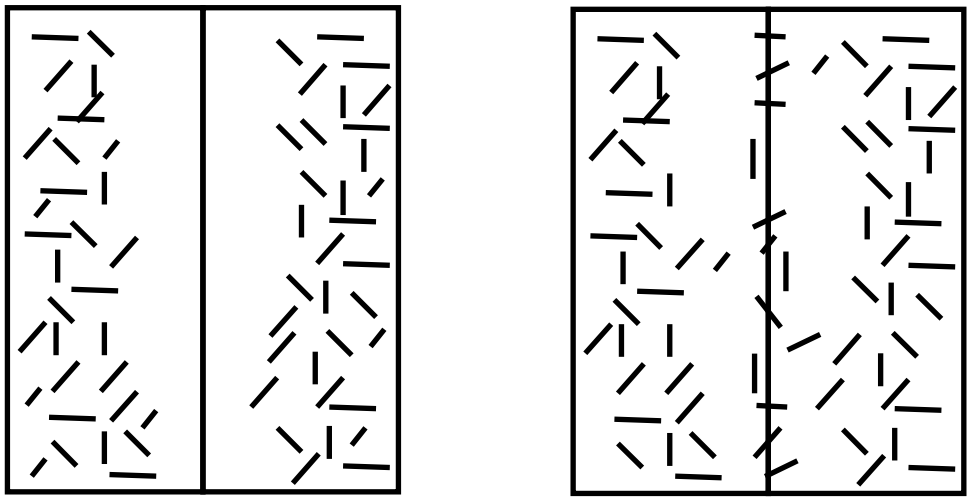


Figure 131 Comparison of post-weld fiber distribution (plan view) for thermally bonded (left) and friction stir welded (right) TPMC joints [138]. Fibers are short, discontinuous, and initially randomly oriented.

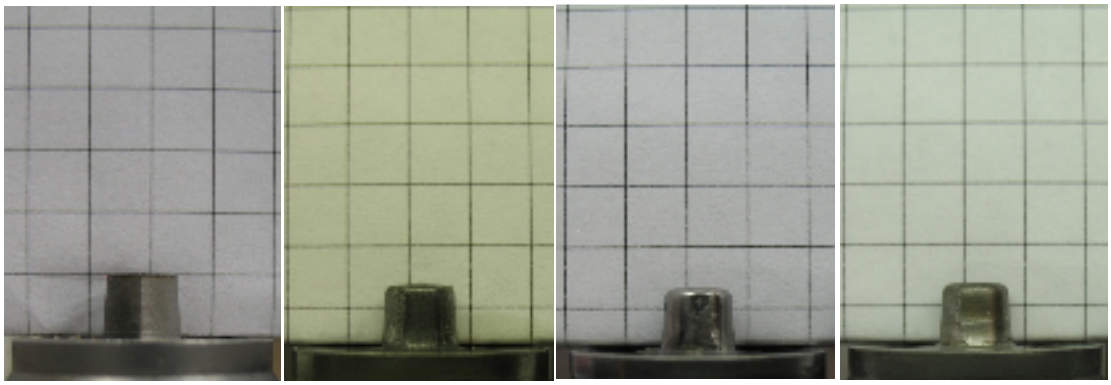
10.2 Final Thought

The limits placed on component size by manufacturing techniques ensure that engineers must inevitably confront the task of joining components in order to assemble a larger structure. Welds are typically the weakest part of a structure. With FSW, the drop-off in strength and structural integrity that typically occurs at the joint interface is substantially reduced (assuming the process has been optimized to produce a joint which is free of defects). The most significant aspect of this work lies in its potential to develop a reliable joining method for metal matrix composites by reducing tool wear. Mitigation techniques for wear (controlling process parameters, using harder materials, using changes in process signals to monitor tool condition) are based on the fundamental understanding of the wear process developed through the research in this dissertation. Knowledge of the process dynamics of FSW of MMCs enhances the applicability of these materials to lightweight structures (such as aerospace components) where they would be of maximum benefit.

APPENDIX A

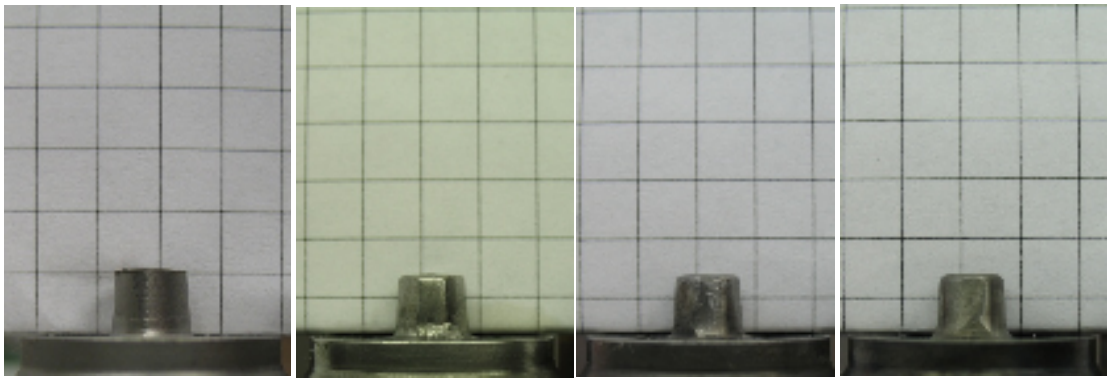
IMAGES OF WORN FIRCTION STIR WELDING TOOLS

1000 RPM, 5 IPM



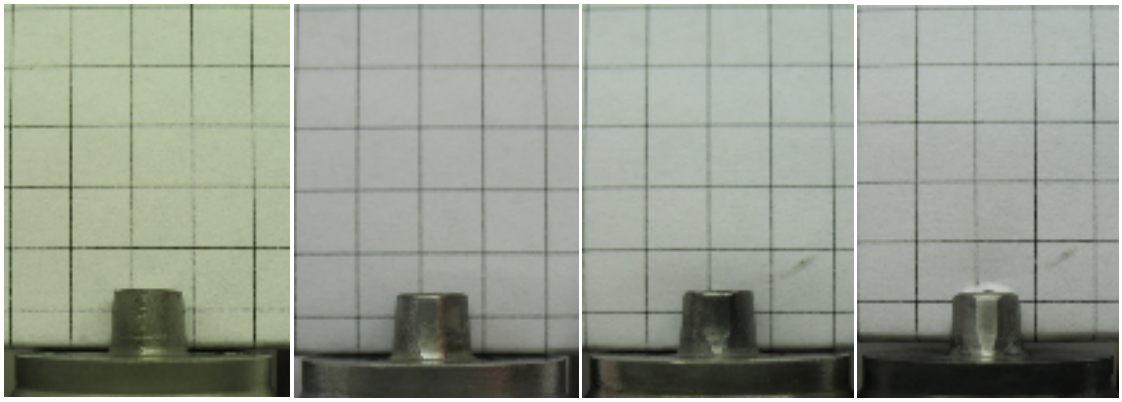
Pre-weld Post-weld 1 Post-weld 2 Post-weld 3

1000 RPM, 7 IPM



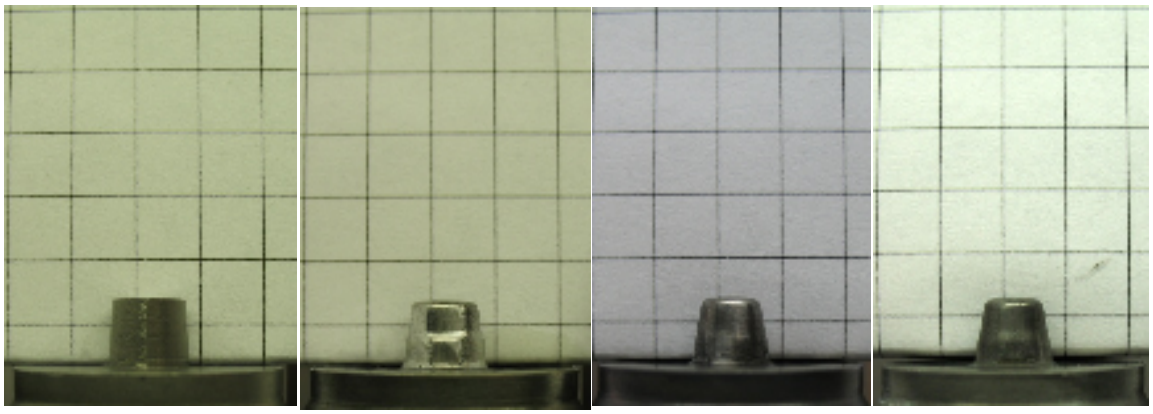
Pre-weld Post-weld 1 Post-weld 2 Post-weld 3

1000 RPM, 9 IPM



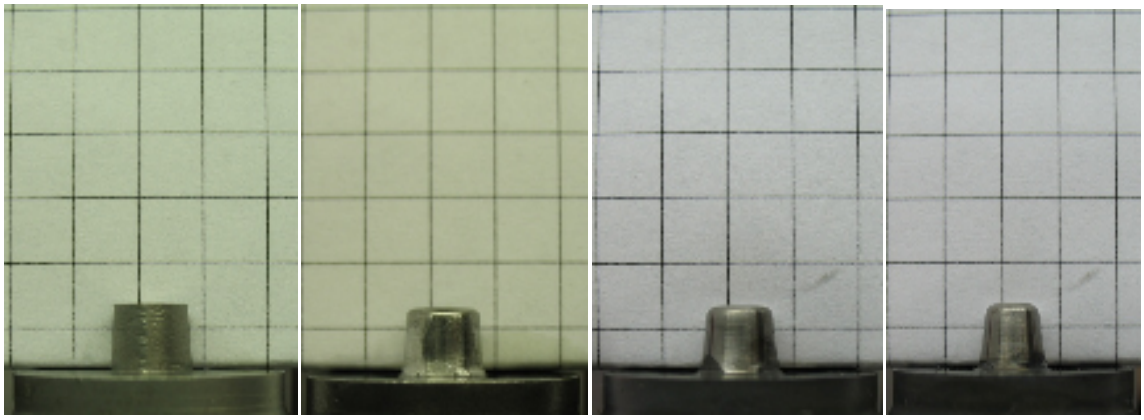
Pre-weld Post-weld 1 Post-weld 2 Post-weld 3

1500 RPM, 5 IPM



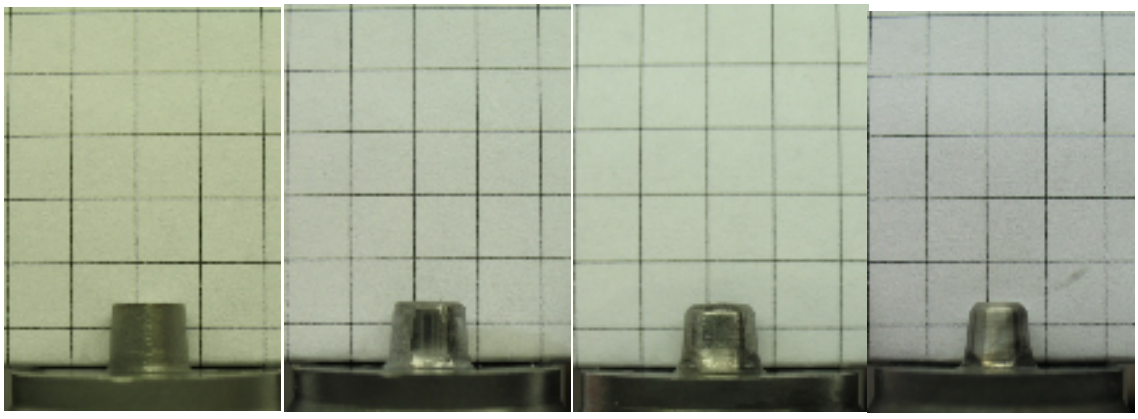
Pre-weld Post-weld 1 Post-weld 2 Post-weld 3

1500 RPM, 7 IPM



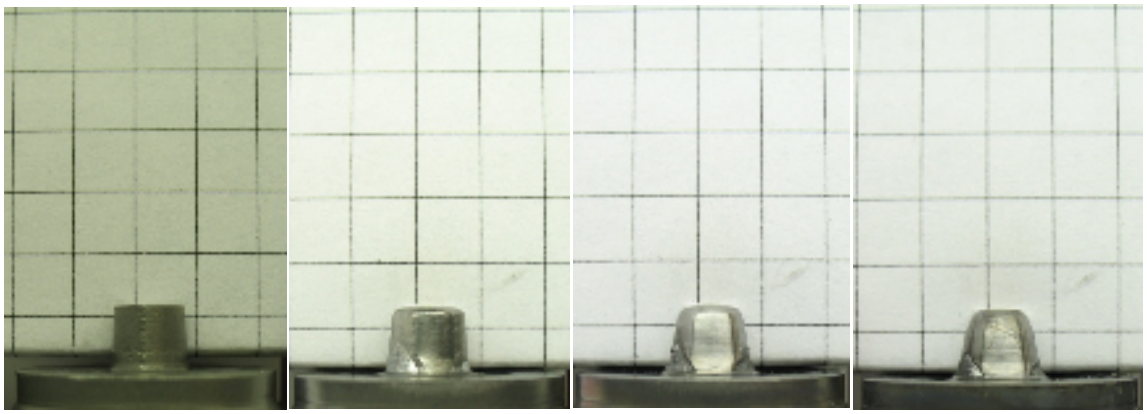
Pre-weld Post-weld 1 Post-weld 2 Post-weld 3

1500 RPM, 9 IPM



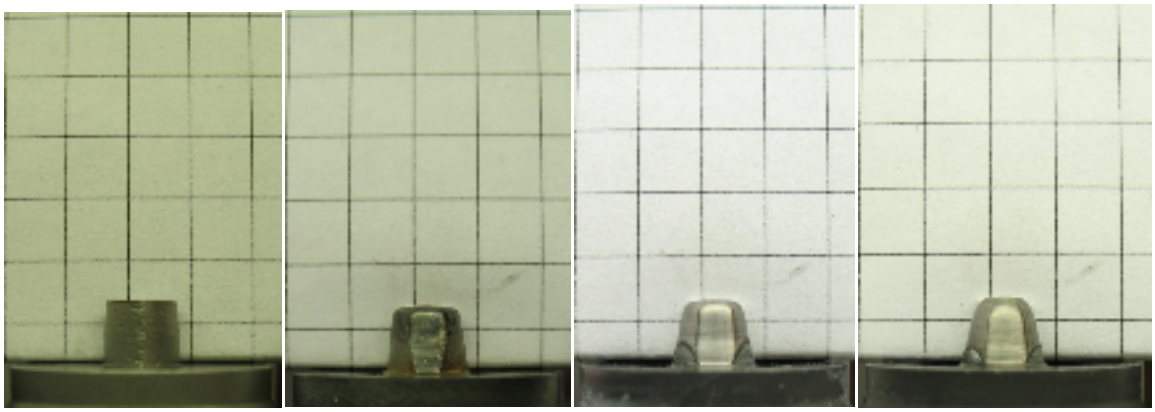
Pre-weld Post-weld 1 Post-weld 2 Post-weld 3

2000 RPM, 5 IPM



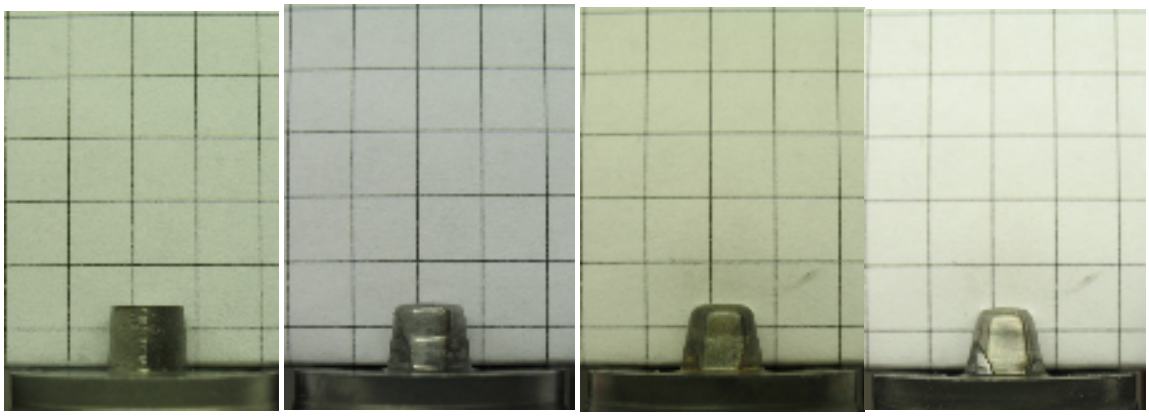
Pre-weld Post-weld 1 Post-weld 2 Post-weld 3

2000 RPM, 7 IPM



Pre-weld Post-weld 1 Post-weld 2 Post-weld 3

2000 RPM, 9 IPM



Pre-weld Post-weld 1 Post-weld 2 Post-weld 3

WORKS CITED

1. Thomas, W.M., et al. "Friction Welding." US Patent 5460317. 24 Oct. 1995.

2. Mishra, R.S. and Z.Y. Ma. "Friction Stir Welding and Processing." Materials Science and Engineering: Research Reports 50 (2005): 1-78.

3. Threadgill, P.L. "Terminology in Friction Stir Welding." Science and Technology of Welding & Joining 12 (2007): 357-360.

4. Shercliff, H.R. and P.A. Colegrove. "Development of Trivex friction stir welding tool part I: two-dimensional flow modelling and experimental validation." Science and Technology of Welding and Joining 9 (2004): 345-351.

5. Mishra, R.S. and M.W. Mahoney, eds. Friction Stir Welding and Processing. Materials Park, OH: American Society of Materials International, 2007.

6. MatWeb Material Property Data. 2009. MatWeb. Oct. 2009 <http://www.matweb.com>.

7. Threadgill, P.L., A.J. Leonard, H.R. Shercliff and P.J. Withers. "Friction stir welding of aluminum alloys." International Materials Reviews 54 (2009): 49-93.

8. Ding, R.J. National Aeronautics and Space Administration. Evaluation of Forces on the Welding Probe of the Retractable Pin-Tool (RPT). Washington: GPO, 1999.

9. Trapp, T.J., et al. "Method of Friction Stir Welding and Retractable Shoulderless Variable Penetration Friction Stir Welding Tool for Same." US Patent 0086775 A1. 27 April 2006.

10. Fleming, P.A. "Monitoring and control in friction stir welding." PhD dissertation, Vanderbilt University, 2009.

11. "Friction Stir Welding." Space Shuttle Technology Summary. 2001. NASA. Oct. 2009 <http://www.nasa.gov/centers/marshall/pdf/104835main_friction.pdf>.

12. Ray, Jay. "Delta 4 fleet goes from 'Medium' to 'Heavy.'" Spaceflight Now. 12 Nov. 2002. Oct. 2009 < http://spaceflightnow.com/delta/d296/030304delta4.html>.

13. Heston, Tim. "Rocket Science, Entrepreneur Style: SpaceX Takes a Fresh Approach to Rocket Fabrication." The Fabricator. 11 Nov. 2008. Oct. 2009 <http://www.thefabricator.com/FabStories/FabStories_Article.cfm?ID=2045 >.

14. . Christner, Coury, Higgins. Proc. 4th International Symposium on Friction Stir Welding. Park City, UT, USA. May 2003, TWI.

15. Delany, Fred, Stephen Kallee, and Mike Russell. "Friction Stir Welding of Aluminum Ships." The Welding Institute (TWI). June 2007. TWI. Oct. 2009 <http://www.twi.co.uk/content/spswkjune07.html>.

16. Mortimer, J. "Jaguar 'Roadmap' rethinks self-piercing technology." Industrial Robot: An International Journal 32 (2005): 209-213.

17. Uzun, Huseyin. "Friction Stir Welding of SiC particulate reinforced AA 2124 aluminum alloy matrix composite." Materials and Design 28 (2007): 1440-1446.

18. Leonard, Lockyer. Proceedings of the Second Symposium on Friction Stir Welding. Gothenburg, Sweden: 27-29 June 2000.

19. Leonard, Lockyer Proceedings of the Fourth Symposium on Friction Stir Welding. Park City, Utah: 14-16 May 2003.

20. Thomas, W.M., D.G. Staines, I.M. Norris and R. De Frias. 'Friction Stir Welding Tools and Developments." Welding in the World 47 (2003): 10-17.

21. Kim, Y.G., H. Fujii, T. Tsumura, T. Komazaki and K. Nakata. "Three defect types in friction stir welding of aluminum." Materials Science and Engineering A 415 (2005): 250-254.

22. Khaled, T. Federal Aviation Administration. An Outsider Looks at Friction Stir Welding (Report Number ANM-23N-05-06). Washington: GPO, 2005.

23. Magnusson, L. and L. Kallman. Proceedings of the Second Symposium on Friction Stir Welding. Gothenburg, Sweden: 27-29 June 2000.

24. Dickerson, T.L. and J. Przydatek. "Fatigue of friction stir welds in aluminium alloys that contain root flaws." International Journal of Fatigue 25 (2003): 1399-1409.

25. Krishnan, K.N. "The effect of post weld heat treatment on the properties of 6061 friction stir welded joints." Journal of Materials Science 37 (2002): 473-480.

26. Stellwag, W.L. and T.J. Lienert. EWI. Friction Stir Welding of Aluminum Metal Matrix Composites Progress Report. Columbus: 2001.

27. Matthews, F.L., G.A. Davies, D. Hitchings, and C. Soutis. Finite Element Modelling of Composite Materials and Structures. Cambridge: Woodhead Publishing Limited, 2003.

28. Sahin, M. "Joining of stainless-steel and aluminium materials by friction welding." The International Journal of Advanced Manufacturing Technology 41 (2009): 487-497.

29. Kunze, J.M. and C.C. Bamptom. "Challenges to developing and producing MMCs for space applications." Journal of the Minerals, Metals and Materials Society 53 (2001): 22-25.

30. Gugger, M.D. "Machining of DRA-MMCs: yesterday, today, and tomorrow." Journal of Advanced Materials 33 (2001): 7-12.

31. DWA Aluminum Composites. 2009. DWA-DRA. Oct. 2009 <http://www.dwa-dra.com>.

32. Storjohann, D., O.M. Barabash, S.S. Babu and S.A. David, et al. "Fusion and Friction Stir Welding of Aluminum Metal Matrix Composites." Metallurgical and Materials Transactions: A: Physical Metallurgy and Materials Science 36A (2005): 3237-3247.

33. Lee, J.A., R.W. Carter and J. Ding. National Aeronautics and Space Administration. Friction Stir Welding of Aluminum Metal Matrix Composites (MSFC Center Director's Discretionary Fund Final Report, Project No. 98-09). Washington: GPO, 1998.

34. Feng, A.H., B.I. Xiao and Z.Y. Ma. "Effect of microstructural evolution on mechanical properties of friction stir welded AA2009/SiCp composite." Composites Science and Technology 68 (2008): 2141-2148.

35. Ceschini. L., L. Boromei, G. Minak, A. Morri and F. Tarterini. "Effect of friction stir welding on microstructure, tensile, and fatigue properties of the AA7005/10 vol. %Al2O3p composite." Composites Science and Technology 67 (2007): 605-615.

36. Ceschini, L., I. Boromei, G. Minak, A. Morri and F. Tarterini. "Microstructure, tensile and fatigue properties of AA6061/20 vol. % Al_2O_{3p} frictions stir welded joints." Composites Part A: Applied Science and Manufacturing 38 (2007): 1200-1210.

37. Lee, Won-Bae, et al. "Microstructures and wear property of friction stir welded AZ91 Mg/SiC particle reinforced composite." Composites Science and Technology 66 (2006): 1513-1520.

38. Humphreys, F.J. and M. Hatherly. Recrystallisation and Related Annealing Phenomena. Oxford: Elsevier Limited, 2004.

39. Marzoli, L.M., A.V. Strombeck, J.F. Dos Santos, C. Gambaro and L.M. Volpone. "Friction stir welding of an AA6061/Al_2O_{3p} reinforced alloy." Composites Science and Technology 66 (2006): 363-371.

40. Shindo, D.J., A.R. Rivera and L.E. Murr. "Shape optimization for tool wear in the friction stir welding of cast Al 359-20% SiC reinforcement." Journal of Materials Science 37 (2002): 4999-5005.

41. Cavaliere, P., E. Cerri, L. Marzoli and J. Dos Santos. “Friction stir welding of ceramic particle reinforced Aluminium based metal matrix composites.” Applied Composite Materials 11 (2004): 247-258.

42. Prado, R.A., L.E. Murr, K.F. Soto and J.C. McClure. “Self-optimization in tool wear for friction-stir welding of Al 6061+20% Al_2O_3 MMC.” Materials Science and Engineering 349 (2003): 156-165.

43. Prado, R.A., L.E. Murr, D.J. Shindo and K.F. Soto. “Tool wear in the friction-stir welding of aluminum alloy 6061+Al_2O_3: a preliminary study.” Scripta materialia 45 (2001): 75-80.

44. Fernandez, G.J. and L.E. Murr. “Characterization of tool wear and weld optimization in the friction-stir welding of cast aluminum 359+20% SiC metal matrix composite.” Materials Characterization 52 (2004): 65-75.

45. Liu, F.J., J.C. Feng, H. Fujii and K. Nogi. “Wear characteristics of a WC-Co tool in friction stir welding of AC4A+30% vol SiCp composite.” International Journal of Machine Tools and Manufacture 45 (2005): 1635-1639.

46. Crawford, R., G.E. Cook, A.M. Strauss, D.A. Hartman, and M.A. Stremler. “Experimental defect analysis and force prediction simulation of high weld pitch friction stir welding.” Science and Technology of Welding and Joining 11 (2006): 657-665.

47. Diwan, Ravinder. National Aeronautics and Space Administration. Investigation of Friction Stir Welding of Al Metal Matrix Composite Materials. Huntsville, AL: 2002.

48. Wert, J.A. "Microstructures of friction stir weld joints between an aluminium-base metal matrix composite and a monolithic aluminium alloy." Scripta Materialia 49 (2003): 607-612.

49. Rabinowicz, Ernest. Friction and Wear of Materials. New York: John Wiley & Sons, 1965.

50. Rodwell, Andy. "Engine bearing failure: detection improved with new technology." Flying Safely August 2002. May 2010 <http://findarticles.com/p/articles/mi_m0IBT/is_8_58/ai_90304762/>

51. T. Prater, G.E. Cook, A.M. Strauss, J. Davidson, M. Howell. Parameterization of Friction Stir Welding of Al 6061/SiC/17.5p for Various Tool Materials. 8th International Conference on Trends in Welding Research , Pine Mountain, GA, 2008.

52. T.J. Prater, A.M. Strauss, G.E. Cook, C. Machemehl, P. Sutton, and C.D. Cox. **"**Statistical modeling and Prediction of Wear in Friction Stir Welding of a Metal Matrix Composite (Al 350/SiC/20p)**."** Journal of Manufacturing and Technology Research 2 (2010): 1-13.

53. Davim, J.P. "Design of optimization of cutting parameters for turning metal matrix composites based on the orthogonal arrays." Journal of Materials Processing Technology 132 (2002): 340-344.

54. Davim, J.P. "Study of drilling metal matrix composites based on the Taguchi techniques." Journal of Materials Processing Technology 132 (2003): 250-254.

55. Archard, J.F. "Contact and Rubbing of Flat Surfaces." Journal of Applied Physics 24 (1953): 981-988.

56. Kruschov, M.M. and M.A. Babichev. "An Investigation of the Wear of Metals and Alloys on an Abrasive Surface." Friction and Wear in Machinery 11 (1956): 1-12.

57. Mulhearn, T.O. and L.E. Samuels. "The abrasion of metals: A model of the process." Wear 5 (1962): 478-498.

58. Aleinikov, F.K. "The Influence of Abrasive Powder Microhardness on the Values of the Coefficients of Volume Removal." Soviet Physics – Technical Physics 2 (1957): 505-511.

59. Adachi, K. and I.M. Hutchings. "Wear mode mapping for the micro-scale abrasion test." Wear 255 (2003): 23-29.

60. Cavanaugh, Mary. "3d surface profilometry of AA7075-T651 exposed to 0.1M NaCl for 1 week." Online image. Fontana Corrosion Center. May 2010. <http://www.matsceng.ohio-state.edu/~frankel/FCC/labs.htm >.

61. Abowd, Richard. Probing the Causes of Piston Ring Wear by the Radiotracer Technique. New York: Society of Automative Engineers, 1959.

62. Sahin, Y. and G. Sur. "The effect of Al_2O_3, TiN, and Ti (C,N) based CVD coatings on tool wear in machining metal matrix composites." Surface and Coatings Technology 179 (2004): 349-355.

63. Kishawy, H.A., S. Kannan, and M. Balazinski. "Analytical Modeling of Tool Wear Progression During Turning Particulate Reinforced Metal Matrix Composites." CIRP Annals – Manufacturing Technology 54 (2005): 55-58.

64. Davim, J. Paolo and A. Monteiro Baptista. "Relationship between cutting force and PCD cutting tool wear in machining silicon carbide reinforced aluminum." Journal of Materials Processing Technology 103 (2000): 417-423.

65. Quigley, O., J. Monaghan, and P. O'Reilly. "Factors affecting the machinability of an Al/SiC metal-matrix composite." Journal of Materials Processing Technology 43 (1994): 21-36.

66. Marinov, V. "Experimental study on the abrasive wear in metal cutting." Wear 197 (1996): 242-247.

67. El-Wardany, T.I., E. Mohammed, and M.A. Elbestaw. "Cutting temperature of ceramic tools in high speed machining of difficult-to-cut materials." International Journal of Machine Tools and Manufacture 36 (1996): 611-634.

68. Barker, Thomas B. Quality by Experimental Design. New York: Marcel Dekker Inc., 1994.

69. Berger, Paul D. and Robert E. Maurer. Experimental Design with Applications in Management, Engineering, and the Sciences. Belmont, CA: Duxbury Press, 2002.

70. Field, Andy. Discovering Statistics Using SPSS. New York: Sage Publications, 2005.

71. Wackerly, Dennis, William Mendenhall, and Richard L. Scheaffer. Mathematical Statistics with Applications. Belmont, CA: Duxbury Press, 2008.

72. Groover, Mikell P. Fundamentals of Modern Manufacturing: Materials, Processes, and Systems. New York: John Wiley and Sons, 2010.

73. Pramanik, A., L.C. Zhang, and J.A. Arsecularatne. "Prediction of cutting forces in machining of metal matrix composites." International Journal of Machine Tools and Manufacture 46 (2006): 1795-1803.

74. Kilickap, Erol and Ali Inan. "A study on machinability of Al Si7Mg2/SiCp metal matrix composite." International Journal of Machining and Machinability of Materials 1 (2006): 463-475.

75. Ozben, Tamer, Erol Kilickap, and Orhan Caku. "Investigation of mechanical and machinability properties of SiC particle reinforced Al-MMC." Journal of Materials Processing Technology 198 (2008): 220-225.

76. Davim, J.P. "Diamond tool performance in machining metal-matrix composites." Journal of Materials Processing Technology 128 (2002): 100-105.

77. Coelho, R.T., S. Yamada, D.K. Aspinwalt and M.L.H. Wise. "The application of polycrystalline diamond (PCD) tool materials when drilling and reaming aluminium based alloys including MMC." International Journal of Machine Tools and Manufacture 35 (1995): 761-774.

78. Roy, G.G., R. Nandan and T. DebRoy. "Dimensionless correlation to estimate peak temperature during friction stir welding." Science and Technology of Welding & Joining 11 (2006): 606-608.

79. Li, X.P. and K.H.W. Seah. "Tool wear acceleration in relation to workpiece reinforcement percentage in cutting of metal matrix composites." Wear 247 (2001): 161-171.

80. Faulring, Gloria M. and S. Ramalingain. "Oxide inclusions and tool wear in machining." Metallurgical and Materials Transactions A 10 (1979): 1781-1788.

81. Hrushtchove, M.M. and M.A. Babitchev. Abrasive Wear. Moscow: 1970.

82. Kannan, S., H.A. Kishawy and M. Balazinski. "Flank Wear Progression During Machining Metal Matrix Composites." Journal of Manufacturing Science and Engineering 128 (2006): 787-791.

83. Weinert, K. and W. Konig. "A Consideration of Tool Wear Mechanism when Machining Metal Matrix Composites (MMC)." CIRP Annals – Manufacturing Technology 42 (1993): 95-98.

84. Ding, X., W.Y.H. Liew and X.D. Liu. "Evaluation of machining performance of MMC with PCBN and PCD tools."Wear 259 (2005): 1225-1234.

85. Lee, S.-Tong, Zhangda Lin, and Xin Jiang. "CVD Diamond Films: nucleation and growth." Materials Science and Engineering 25 (1999): 123-154.

86. Xing, Ai, Jianfeng Li and Chuanzhen Huang. "Recent Devlopments in Tool Materials for High Speed Machining." Materials Science Forum 471 (2004): 438-442.

87. Murakawa, M., S. Takeuchi, H. Miyazawa and Y. Hirose. "Chemical vapour deposition of a diamond coating onto a tungsten carbide tool using ethanol." Surface and Coatings Technology 36 (1988): 303-310.

88. Shibuki, K., K. Sasaki, M. Yagi, T. Suzuki and Y. Ikuhara. "Diamond coating on WC-Co and WC for cutting tools." Surface and Coatings Technology 68-69 (1994): 369-373.

89. Andrews, C.J.E., H.Y. Feng and W.M. Lau. "Machining of an aluminium/SiC composite using diamond inserts." Journal of Materials Processing Technology 102 (2000): 25-29.

90. Oles, E.J., A. Inspektor and C.E. Bauer. "The new diamond-coated carbide cutting tools." Diamond and Related Materials 5 (1996): 617-624.

91. Davim, J.P. "Turning particulate metal matrix composites: experimental study of the evolution of the cutting forces, tool wear and workpiece surface roughenss with the cutting time." Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 215 (2001): 371-376.

92. Jacoby, Henry. *House* and Philosophy: Everybody Lies. Hoboken, NJ: John Wiley & Sons, 2008.

93. Herling, Darrell R., Glenn J. Grant and Warren Hunt, Jr. "Low-cost Aluminum Metal Matrix Composites." *Advanced Materials and Processes* July 2001: 37-40.

94. mc21. mc21, Inc. July 2011 < http://www.mc21inc.com/capabilities.html>.

95. iStir Process Development System. 2004. MTS Systems Corporation. July 2011 <http://www.mts.com/ucm/groups/public/documents/library/dev_002515.pdf>

96. Rosio Friction Stir Welding Robot. ESAB. July 2011

< http://products.esab.com/Templates/T041.asp?id=177852>.

97. Prater, Tracie. An Investigation Into the Friction Stir Welding of Al 6061 and Al 6061/SiC/17.5p Using Diamond Coatings. M.S. thesis, Vanderbilt University, 2008.

98. Crawford, Reginald. A Mechanistic Study of the Friction Stir Welding Process. PhD dissertation, Vanderbilt University, 2007.

99. Ipsen, D.C. Units, Dimensions, and Dimensionless Numbers. New York: McGraw Hill, 1960.

100. Schmidt, H., J Hattel and J Wert. "An analytical model for the heat generation in friction stir welding." Modelling and Simulation in Materials Science and Engineering 12 (2004): 143-157.

101. Nunes, A.C., E.I. Bernstein and J.C. McClure. "A rotating plug model for friction stir welding." 81st American Welding Society Convention, Chicago, IL, 2000.

102. Querin, Joseph. Deconvoluting the link between weld tool geometry and process parameters. PhD dissertation, Mississippi State University, 2010.

103. Nunes, Arthur. Personal communication. Summer 2010.

104 Reynolds, A.P., Seidel, T.U., Simonsen, M. 1999. Visualization of Material Flow in an Autogenous Friction Stir Weld. 1st International Symposium on Friction Stir Welding, Rockweel Science Center, Thousand Oaks, CA. June 14-16.

105. Colligan, K. "Material flow behavior during friction stir welding of aluminum." Supplement to The Welding Journal (1999): 229-237.

106. McQueen, H.J., and J.J. Jonas. "Recovery & Recrystallisation During High Temperature Deformation." Plastic Deformation of Materials (1975): 393-493.

107. Lammlein, David. Friction Stir Welding of spheres, cylinders, and T-joints: design, experiment, modeling, and analysis. PhD dissertation, Vanderbilt University, 2011.

108. Hamilton, C., S. Dymek and A. Sommers. "Characteristic Temperature Curves for Aluminum Alloys during Friction Stir Welding." Supplement to The Welding Journal (2010): 189-194.

109. Aluminum. Efunda. July 2011

<http://www.efunda.com/materials/elements/TC_Table.cfm?Element_ID=al>.

110. Prater, T., C. Cox, B. Gibson, A. Strauss, G. Cook. "Dimensional analysis and a potential classification algorithm for tool wear in friction stir welding of metal matrix composites." Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science (accepted, publication pending).

111 Nunes, Arthur. "A Theory of Wear in Friction Stir Welding of Metal Matrix Composites." Unpublished manuscript, 2010.

112 Kumar, Vereesh, C.S.P Rao, N. Selvaraj, M.S. Bhagyashekar. "Studies on Al 6061-SiC and Al 7075 Al_2O_3 Metal Matrix Composites." Journal of Minerals and Materials Characterization & Engineering 9 (2010): 43-55.

113. Bauser, M., G. Sauer and Klaus Siegert. Extrusion. Dusseldorf, Germany: ASM International, 2006.

114. Kalaichelvi, V., D. Sivakumar, R. Karthikeyan, K. Palanikumar. "Prediction of the flow stress of 6061 Al-15% SiC-MMC composites using adaptive network based fuzzy interference system." Materials & Design 30 (2009): 1362-1370.

115. Yield Strength. Burns Stainless, LLC. January 2012 <http://www.burnsstainless.com/yieldstrength.aspx>.

116. Incropera, Frank. Fundamentals of Heat and Mass Transfer. Hoboken, New Jersey: John Wiley & Sons, 2006.

117. Tool Steel Facts. Bohler-Uddeholm. June 2010 <http://www.bucorp.com/files/aisi_o1.pdf>.

118. Raghulapadu, J.K., J. Peddieson, G.R. Buchanan, A.C. Nunes. "A Rotating Plug Model of Friction Stir Welding Heat Transfer." Heat Transfer Engineering 29 (2008): 321-334.

119. Schuster, David. Personal communication. May 2009.

120. FEPA Powder Size Specifications. Reade. May 2009 < http://www.reade.com/resources/particle-measurement>.

121. Moore, Lewis. Personal communication. Summer 2011.

122. Ludema, Kenneth C. Friction, Wear, Lubrication: A Textbook in Tribology. New York: CRC Press, 1996.

123. "Estimated Hardness Equivalents between Vickers and Rockwell C." Gordon England Surface Engineering Forum. 2008. Gordon England Company. Oct. 2009 <http://www.nasa.gov/centers/marshall/pdf/104835main_friction.pdf>.

124. Deryagin, B.V. and D.V. Fedosayev. "The Growth of diamond and graphite from the gas phase." Surface and Coatings Technology 38 (1989): 131-248.

125. Davidson, Jimmy. Personal communication. 2008.

126. Longhurst, R.A., A. Strauss, G. Cook, P. Fleming. "Torque control of friction stir welding for manufacturing and automation." The International Journal of Advanced Manufacturing Technology 51 (2010): 905-913.

127. Longhurst, Russ. Unpublished results. March 2012.

128. Ding, J., et al. "System for controlling the stirring pin of a friction stir welding apparatus." US Patent 6497355. 24 Dec. 2002.

129. Trapp, et al**. "**Method of friction stir welding and multi-section faced shoulderless retractable variable penetration friction stir welding tool for same." US Patent 7416102. 26 Aug. 2008.