SENSOR-BASED MONITORING AND INSPECTION OF SURFACE MORPHOLOGY IN ULTRAPRECISION MANUFACTURING PROCESSES

By

PRAHALADA KRISHNA RAO

Bachelor of Engineering in Production Engineering
Victoria Jubilee Technical Institute (VJTI),
Bombay University
Mumbai, Maharashtra, India
2003

Master of Science in Industrial Engineering and Management
Oklahoma State University
Stillwater, Oklahoma, United States
2006

Submitted to the Faculty of the
Graduate College of
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2013

SENSOR-BASED MONITORING AND INSPECTION OF SURFACE MORPHOLOGY IN ULTRAPRECISION MANUFACTURING PROCESSES

Dissertation Approved:

Dr. Satish T.S. Bukkapatnam
Dissertation Adviser

Dr. Zhenyu (James) Kong
Dissertation Co-adviser

Dr. Kenneth E. Case

Dr. William J. Kolarik

Dr. Don A. Lucca
External member

Acknowledgements

I thank my adviser Dr. Satish Bukkapatnam for his kind guidance, patience, and help over the years encompassing my MS and Ph.D. study. I have known him for close to a decade and am indebted to him for providing me the opportunity to accomplish my education. It is due to my adviser that I have obtained the highest human pleasure of learning. At the risk of sounding platitudinous, without him I would have achieved nothing of intellectual value in my life.

I thank Dr. Kenneth E. Case for being the ever-present mentor, right from the day I attended my first lecture at OSU long years ago. Dr. Case was the first to suggest that I should pursue a doctoral degree, and he has stood by me despite my numerous failings. I would have never known of my abilities if he had not mined them out of me. Dr. Case has taken the time at every important juncture in my professional and personal life to ensure that I not only do *things right*, but also that I do the *right things*. Suffice to say, he is much more than a teacher, lot more than a mentor, and a good deal more than a friend to me.

I express my gratitude to Dr. Zhenyu Kong who has been a close mentor and adviser during my Ph.D. research. I have learned a lot by observing his amenable, helpful, and approachable demeanor. I am lucky to have such gentlemen as mentors and guides.

My thanks are due to Dr. William Kolarik for his patience and understanding approach to mentoring. His encouragement has certainly helped in my profession-

al development. I am grateful to Dr. Don Lucca for his inputs, help, and for agreeing to be the outside member on the dissertation committee.

I would like to thank Mr. John Gage (Lab Manager, DML) for his guidance, time, and help with fabrication of the experimental setups used in this research, and most importantly, for his friendship, prayers, and concern for me.

I also wish to thank my colleagues M/s O. Beyca, A. Fields, S. Byalal, A. Ohri, Z. Wang and C. Cheng. My wife, Shilpa has been solely responsible for keeping me sane and motivated when things were not going smoothly. There are numerous friends and colleagues who have helped me out, in ways both small and big, my sincere thanks to you, you are certainly not forgotten.

The preponderate measure of my gratitude is credited to Dr. Ranga Komanduri (1942-2011). My ears will forever yearn for his approval. Those who know me have often said that a part of me has forever changed since his passing. I will not dispute that observation, and I will mourn his loss every moment of my life.

Every aspect of this research stems from Dr. Komanduri's vision. Whether the outcomes satisfy his standards or not, is something for which I have neither the wherewithal nor the audacity to speculate even in the vaguest. I only know that he is with me always. I have no words to describe my affection for Dr. K, except with this pithy epitaph.

Mere platitudes you never gave,
empty words you never uttered,
shy of a score I collected;
I wish only one more.

This dissertation is dedicated to the memory of Dr. R. Komanduri.

This research is funded by the following grants from the

National Science Foundation (NSF).

CMMI 0700680, CMMI 0729552, CMMI 0927557, CMMI 1000978

Acknowledgements reflect the views of the author and are not endorsed by committee members or Oklahoma State University.

Name: PRAHALADA KRISHNA RAO

Date of Degree: JULY, 2013

Title of Study: SENSOR-BASED MONITORING AND INSPECTION OF SURFACE MORPHOLOGY IN ULTRAPRECISION MANUFACTURING PROCESSES

Major Field: INDUSTRIAL ENGINEERING AND MANAGEMENT

Abstract: This research proposes approaches for monitoring and inspection of surface morphology with respect to two *ultraprecision*/*nanomanufacturing* processes, namely, ultraprecision machining (UPM) and chemical mechanical planarization (CMP). The methods illustrated in this dissertation are motivated from the compelling need for *in situ* process monitoring in nanomanufacturing and invoke concepts from diverse scientific backgrounds, such as artificial neural networks, Bayesian learning, and algebraic graph theory. From an engineering perspective, this work has the following contributions:

1. A combined neural network and Bayesian learning approach for early detection of UPM process anomalies by integrating data from multiple heterogeneous *in situ* sensors (force, vibration, and acoustic emission) is developed. The approach captures process drifts in UPM of aluminum 6061 discs within 15 milliseconds of their inception and is therefore valuable for minimizing yield losses.
2. CMP process dynamics are mathematically represented using a deterministic multi-scale hierarchical nonlinear differential equation model. This process-machine interaction (PMI) model is evocative of the various physio-mechanical aspects in CMP and closely emulates experimentally acquired vibration signal patterns, including complex nonlinear dynamics manifest in the process. By combining the PMI model predictions with features gathered from wirelessly acquired CMP vibration signal patterns, CMP process anomalies, such as pad wear, and drifts in polishing were identified in their nascent stage with high fidelity ($R^2 \sim 75\%$).
3. An algebraic graph theoretic approach for quantifying nano-surface morphology from optical micrograph images is developed. The approach enables a parsimonious representation of the topological relationships between heterogeneous nano-surface features, which are enshrined in graph theoretic entities, namely, the similarity, degree, and Laplacian matrices. Topological invariant measures (e.g., Fiedler number, Kirchoff index) extracted from these matrices are shown to be sensitive to evolving nano-surface morphology. For instance, we observed that prominent nanoscale morphological changes on CMP processed Cu wafers, although discernible visually, could not be tractably quantified using statistical metrology parameters, such as arithmetic average roughness (Sa), root mean square roughness (Sq), etc. In contrast, CMP induced nanoscale surface variations were captured on invoking graph theoretic topological invariants. Consequently, the graph theoretic approach can enable timely, non-contact, and *in situ* metrology of semiconductor wafers by obviating the need for reticent profile mapping techniques (e.g., AFM, SEM, etc.), and thereby prevent the propagation of yield losses over long production runs.

Executive Summary

This research proposes approaches for monitoring and inspection of surface morphology in two *ultraprecision/nanomanufacturing* processes, namely, ultraprecision machining (UPM) and chemical mechanical planarization (CMP) (see Graphic I). The methods illustrated in this dissertation are motivated from the compelling need for *in situ* process monitoring in nanomanufacturing, and invoke concepts from diverse scientific backgrounds, such as sensor-based process modeling, digital signal processing, artificial neural networks, Bayesian learning, and algebraic graph theory. The underlying principles are nonetheless envisioned to be extensible to other nanomanufacturing processes, and by induction, conventional manufacturing processes (particularly those involving material removal) not addressed in this research. The objectives, socioeconomic impact, and synopsis of this research are presented in Chapter 1.

A combined neural network and Bayesian learning approach for early detection of UPM process anomalies by integrating data from multiple heterogeneous *in situ* sensors (force, vibration, and acoustic emission) is illustrated in Chapter 2 (summarized in the northwest quadrant of Graphic I). This approach captures process drifts in UPM of aluminum 6061 discs (Sa ~ 20 nm) within 15 milliseconds of their inception and is therefore valuable for minimizing yield losses in UPM processes.

Chapter 3 details *in situ* sensor-based monitoring approaches for CMP. We have designed a wireless sensor network consisting of miniature MEMS vibration and sound sensors that enable close-proximity data acquisition in a CMP process. Extensive tests are conducted (documented in Appendix II) in order to isolate sensor data features that can capture CMP process states from those sensor signal components which are merely the effect of extraneous noise. As a consequence, we identify certain signal features which are valuable for process monitoring in CMP.

Next, the CMP process dynamics are mathematically represented using a deterministic multi-scale hierarchical differential equation model (summarized in the northeast quadrant of Graphic I). This mathematical model (documented in Appendix III) is evocative of the various physio-mechanical aspects in CMP (termed process-machine interaction (PMI) model) and closely emulates experimentally acquired vibration signal patterns, including complex nonlinear dynamics manifest in the process. By combining the PMI model predictions with features from the wirelessly acquired experimental signals, CMP process anomalies, such as pad wear, and drifts in polishing load, can be predicted at an early stage with high fidelity ($R^2 \sim 75\%$).

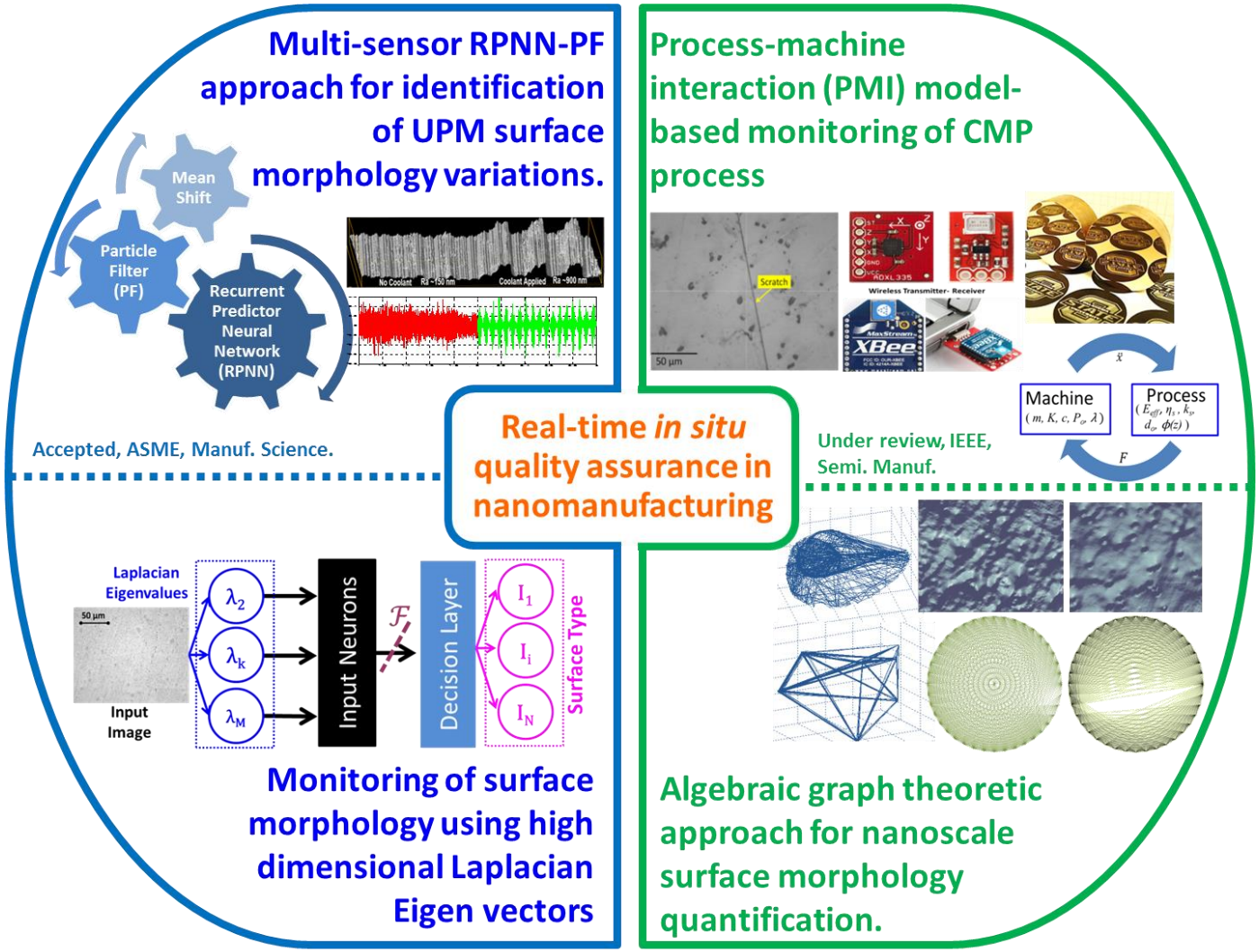
In Chapter 4, we present an algebraic graph theoretic approach (depicted in the southeast quadrant of Graphic I) for quantifying nano-surface morphology from optical micrograph images. This approach overcomes some of the lacunae associated with conventional statistics-based surface metrology. The graph theo-

retic approach is not only capable of capturing nanoscale topographical aspects (e.g., scratches, pits, and ridges) but also obviates reticent surface profile scanning methods, such as laser interferometry AFM, SEM, etc.

The algebraic graph theoretic approach is essentially a parsimonious representation of the underlying topological relationship between heterogeneous surface features. These relationships are enshrined in graph theoretic entities, namely, the similarity, degree, and Laplacian matrices. The properties of these graph matrices (e.g., eigenvalues and eigenvectors) are found to be sensitive to evolving nano-surface morphology.

This is pertinent because it was observed that nanoscale morphological changes (for instance, on CMP processed Cu wafers) were not tractably quantified using statistical features, such as arithmetic average roughness (Ra, Sa), root mean square roughness (Rq, Sq), and other such distribution parameters (skewness, kurtosis). For instance, we show that by invoking the graph theoretic invariant Fiedler number (viz. the smallest non-zero eigenvalue of the graph Laplacian matrix), the morphology of CMP processed Cu surfaces can be assessed from optical micrographs. This can enable timely, non-contact, and *in situ* metrology of semiconductor wafers.

Finally, future research plans (summarized in the south-west quadrant of Graphic I) and conclusions are discussed in Chapter 5 and Chapter 6, respectively.



Graphic I: An Overview of the dissertation.

North-West quadrant: The RPNN-PF methodology for capturing incipient process anomalies in UPM (Chapter 2).
North-East quadrant: PMI model approach integrating wireless sensor data and multi-scale CMP phenomena for detection of process induced wafer defects (Chapter 3).
South-East quadrant: An algebraic graph theoretic approach for capturing surface morphology evolutions in nanomanufactured ultraprecision components (Chapter 4).
South-West quadrant: Future research directions for algebraic graph theoretic quantification of surface morphology (Chapter 5).

Table of Contents

List of Figures

List of Tables

1 Introduction

This dissertation describes approaches for monitoring and inspection of surface morphology in ultraprecision/nanomanufacturing processes by incorporating *in situ* quality assurance capabilities in two such vital processes, namely, ultraprecision machining (UPM) and chemical mechanical planarization (CMP) [1-4]. UPM and CMP are often used in the manufacture of components with nanometer (10^{-9} m) level dimensional specifications (1 nm to 100 nm) [2, 4]. Since surface quality is a critical determinant of functional performance for nanomanufactured components, defect-free surface quality with characteristics in the nanometer range, e.g., arithmetic mean roughness (Ra) < 10 nm (generally referred to as mirror or specular finished surface – contingent on the processed material), is an often mandated requirement for such components [5-9].

Table 1-1 juxtaposes UPM and CMP alongside conventional manufacturing techniques in terms of the achievable surface finish (Ra) [1, 3, 10]. From Table 1-1, it is evident that the (typical) surface quality (with reference to surface roughness, Ra) that can be realized with UPM and CMP is almost three orders of magnitude smaller (smaller is better) to comparable conventional manufacturing processes (e.g., turning and lapping, respectively). In a similar vein, Table 1-2 shows representative manufacturing tolerances and functionally critical facets for various engineering components, with tolerance < 100 nm being the domain of nanomanufacturing processes, such as UPM and CMP.

Table 1-1: Typical surface finish ranges for various manufacturing processes (after DeGarmo, *et al.* [10]).
Representative surface finish values for UPM [1] and CMP [3] processes have been appended. Darker shades indicate typical ranges; the lighter shades represent less typical ranges.

AVERAGE SURFACE FINISH (Ra) (MICROMETERS, µm) →	50	25	12.5	6.3	3.2	1.6	0.8	0.4	0.2	0.1	0.050	0.025	0.012	0.006	0.003	0.001
CLASSIFICATION →	**CONVENTIONAL**						**PRECISION**				**ULTRAPRECISION**					
MACHINING (METAL CUTTING)																
SAWING																
PLANING, SHAPING																
DRILLING																
MILLING																
BORING, TURNING																
BROACHING																
REAMING																
ULTRAPRECISION MACHINING																
ABRASIVE MACHINING																
GRINDING																
BARREL FINISHING																
HONING																
ELECTRO-POLISHING																
ELECTROLYTIC GRINDING																
POLISHING																
LAPPING																
SUPER-FINISHING																
CHEMICAL MECHANICAL PLANARIZATION																
MOLDING																
SAND CASTING																
PERMANENT MOLD CASTING																
INVESTMENT CASTING																
DIE CASTING																
METAL FORMING																
HOT ROLLING																
FORGING																
EXTRUSION																
COLD ROLLING, DRAWING																
ROLLER BURNISHING																
NON-TRADITIONAL MACHINING																
FLAME CUTTING																
CHEMICAL MILLING																
ELECTRON BEAM CUTTING																
LASER CUTTING																
ELECTRO-DISCHARGE MACHINING																

Table 1-2: Part tolerance and emphasis with respect to different applications (after Taniguchi [4]).

PART TOLERANCE ↓		MECHANICAL PARTS	ELECTRONIC PARTS	OPTICAL PARTS
EMPHASIS →		SHAPE (3D) PRECISION	PATTERN (2D) ACCURACY	SURFACE FINISH (1D) FINENESS
CONVENTIONAL PROCESSING	200 µm	Conventional machine and homeware parts	General purpose electronic parts (switch, motor, connector)	Camera and Telescope bodies
	50 µm	General-purpose mechanical parts (gear, thread), typewriter parts, engine parts	Package (electronic parts), micrometer, transistor, diode, magnetic head	Camera shutter, lens holder (camera, microscope)
	5 µm	Mechanical watch parts, accurate gears, threads, machine tool bearings, ball screw, rotary compressor parts, shaver blade	Electric relay, resistor, condenser, disc memory, silicon wafer, TV color mask, video head cylinder	Lens, prism, optical fiber and connector
PRECISION PROCESSING	0.5µm (500 nm)	Ball and roller bearings, precision drawn wire, flapper servo valve, gyro-bearing, air bearing, precision die, roll thread die, ink jet nozzle	Magnetic head (VCR), magnetic scale, CCD, quartz oscillator, magnetic bubble memory, IC, magnetron, thin-film pressure transducer, thermal printer head	Precision lens and prism, optical scale, IC-exposure mask (photo, X-ray), laser polygon mirror, elastic deflection mirror
ULTRAPRECISION PROCESSING	0.05 µm (50 nm)	Block gage, diamond indenter, high precision X-Y table, high precision stamper and die, microtome cutter (diamond)	IC memory, electronic video disc, LSI IC chip, micro-vacuum tube TFT-LCD	Optical flat, precision Fresnel lens, optical diffraction grating, optical video disc (CD)
	0.005 µm (5 nm)	Ultraprecision parts (plane, ball, roller, thread)	VLSI IC chip, supper lattice (synthesis) thin film	Ultraprecision diffraction grating

1.1 Motivation and Research Objectives

UPM (often also referred to as diamond turning (DT)) is typically a single-point turning process, which uses a single crystal diamond (SCD) tool for crafting precision components, such as precision molds, miniature lenses, aspheric mirrors, micro-channels, etc., which are functionally critical to the electro-optical, aerospace, biomedical, and defense industries [1-4, 11, 12]. Characteristic tolerances (waviness) for UPM finished parabolic mirrors used in radiometry are of the order of ± 6 nm, with surface finish (Ra) in the range of 2 – 25 nm [13, 14]. We report near-specular[1] surface finish (arithmetic mean areal surface roughness, (Sa) 15–20 nm) obtained on aluminum 6061 discs using our experimental UPM setup (see the left side of Graphic II, which summarizes these outcomes).

The UPM process is observed to be exceedingly sensitive to minute instabilities from extraneous sources, such as temperature fluctuations, vibrations, material inconsistencies, etc. [15-17]. It is noted that an infinitesimal change or drift in process conditions, for instance, due to heat radiated from the human body (of the operator), vibration from nearby machinery, perturbation from opening and closing doors, material crystal defects, etc., that might be inconsequential in conventional machining, manifest almost immediately on the surface of the work-material in UPM [1, 18].

[1]The surface produced has close to mirror finish quality, but is not defect-free due to certain sub-optimal conditions stemming from microstructural defects in the material. The Al 6061 sample used in our UPM study has hard material inclusions (aluminum silicide) due to the presence of silicon as an alloying element. This significantly impedes the homogeneity of the microstructure, and consequently the uniformity of cutting forces and conditions (depth of cut).

If such process drifts/anomalies are not detected in a timely manner, they can render waste several hours of labor, expensive tools, and substrates. **Therefore, detection of incipient anomalies is important for quality assurance in a UPM process.** Consider for example, an extract (produced verbatim below) from a recent (March 2013) job posting by II-VI Inc., of Saxonburg, PA, which describes the responsibilities of a process engineering manager for its ultraprecision machining operations (referred as diamond turning (DT) by the company).

> *"**Errors in DT can have significant impact on profits of Diamond Turning Business Unit**. Attaining and maintaining satisfactory yields, on time delivery throughput and product quality is critical to II-VI profitability. Operational errors can jeopardize II-VI reputation in the market and can result in lost growth opportunities."*

From the above excerpt, it is evident that control of key process output variables (dimensional accuracy and surface integrity of the component) in UPM is a pertinent engineering problem consequential to profitability.

CMP is a free abrasive process similar to lapping, which is widely used in the semiconductor industry for finishing dielectric and metal interconnect layers patterned on semiconductor wafers [19]. The dimensional consistency of these interconnect layers is crucial to the functional performance of semiconductor devices and, therefore, tightly controlled during manufacture [19, 20]. Since a typical modern semiconductor integrated circuit (IC) has 7 to 10 interconnect layers, with each layer requiring a CMP step, CMP induced variations in layer thickness can severely impede device functionality [20, 21].

Pertinently, with semiconductor devices with feature sizes less than 22 nm soon to be commercially introduced, nanoscale wafer surface morphology defects (pits, scratches, corrosion) that occur during a CMP step can have deleterious effects on microprocessor performance [20]. **Consequently, monitoring of CMP process is vital for ensuring functional integrity of semiconductor devices**. In this research, copper wafers with near-specular[2] characteristics and surface finish (Sa) in the range of 5 – 8 nm were obtained using our experimental CMP setup (depicted in the right half of Graphic II).

CMP process induced wafer defects have significant financial implications. For instance, revenue losses from damage to a single 300 mm (12 inch) semiconductor wafer due to CMP defects are appraised at over $100,000 [22]. **Industry practitioners have noted that wafer defects resulting from poor CMP process control are among the top five reasons inhibiting semiconductor yield** [23]**.**

For quality assurance purposes, semiconductor manufacturers spend close to $9 billion annually on wafer metrology related operations [24]. The current emphasis has been towards development of responsive wafer inspection methods, since *ex situ* offline metrology can lower production yield by as much as 35% [25]. For example, results from off-line wafer inspection may often take more time than a production shift to be reported. Such delays inhibit detection and

[2]This is because the copper wafers used in our study have tellurium (Te) added as an alloying element. Consequently, the particular copper alloy used this study (C14500 series) has a higher machinability rating (~ 80%) compared to pure copper (~ 20%). A higher machinability rating was necessary for ease of shaping (sawing and CNC machining) of wafers commensurate with the dimensional constraints of the apparatus, and in order to mount sensors on the wafer backside. During CMP these Te inclusions tend to be removed as flakes, which impedes the surface finish that can be achieved (in comparison to pure copper).

subsequent prevention of process related defects [26]. **Thus, poor control and reticent inspection in nanomanufacturing processes, such as CMP, can result in significant revenue losses.**

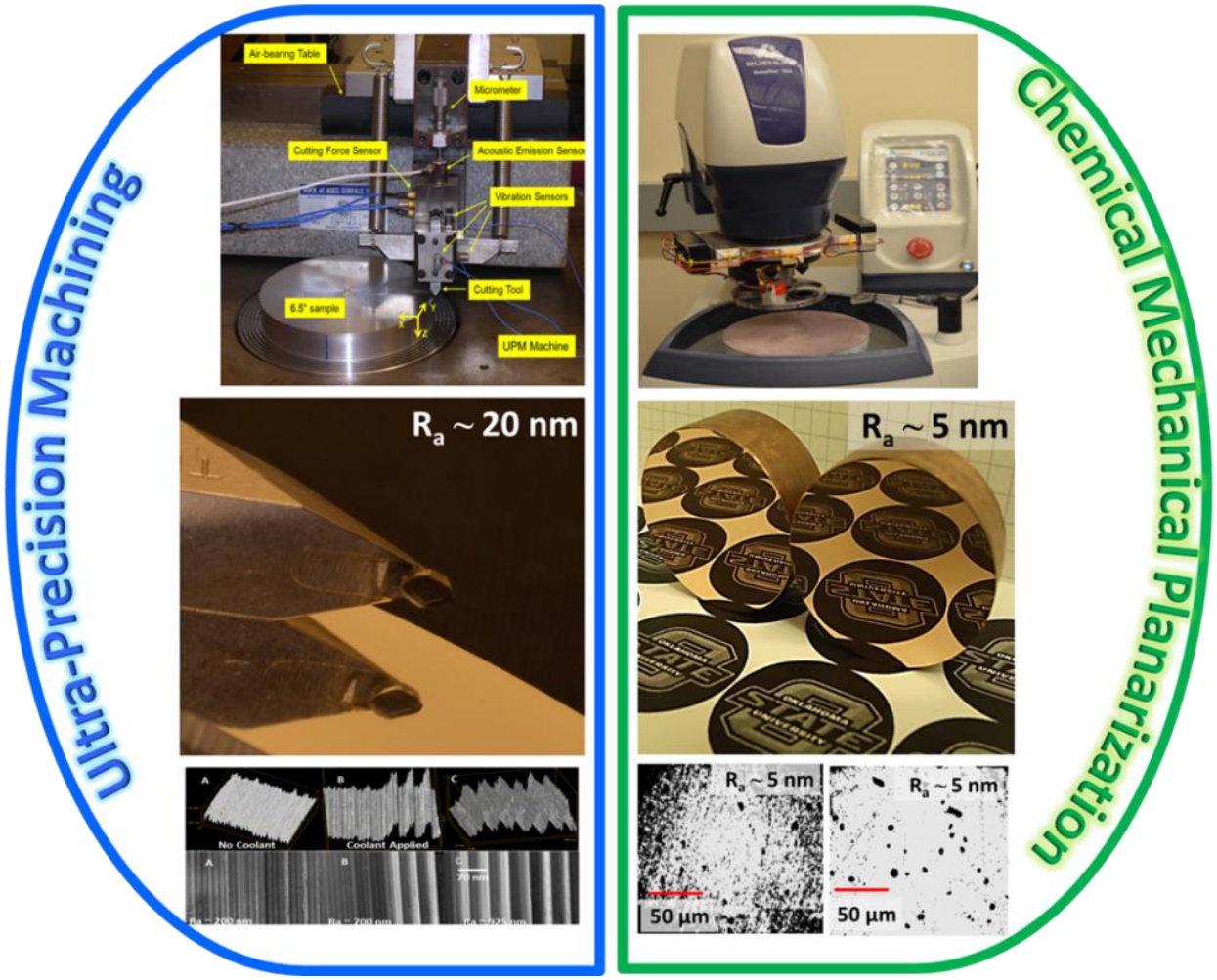
As noted by Bennet [24], one of the evolving trends, particularly in the semiconductor industry, is the ***increasingly greater emphasis on faster characterization of manufacturing processes, assisted by increased modeling and simulation for nanotechnologies to offset the greater increase in costs associated with measurements*** **(**quoted verbatim from [24]). Given these concerns, traditional post process statistics-based monitoring and inspection techniques are of limited utility, because, due to their inherent reticence they may fail to detect some of the subtle process drifts manifest in nanomanufacturing operations [27, 28].

Therefore, there is a compelling need to suggest approaches that can capture evolving (incipient) process anomalies at an early stage in nanomanufacturing applications so that opportune corrective actions can be taken and yield losses can be minimized. With these concerns in perspective, we distill the objectives of this research as follows:

Research Objectives

1. To forward approaches for monitoring and identification of incipient anomalies in nanomanufacturing process, specifically UPM and CMP, based on *in situ* sensing techniques.
2. To suggest methods for surface quality assessment of nanomanufactured ultraprecision components that can be incorporated *in line* and are capable of capturing subtle nanoscale surface morphology variations.

Approaches for real-time *in situ* quality assurance in ultra-precision machining (UPM) and chemical mechanical planarization (CMP).



Graphic II: Summary of research concerning ultraprecision machining (UPM) and chemical mechanical planarization (CMP) processes.

Left: The experimental UPM setup instrumented with multiple miniature sensors, and the typical specular finish obtained on Al 6061 workpieces using this setup. The bottom portion shows profiles of abnormal surfaces due to UPM process drifts.
Right: A CMP machine instrumented with multiple sensors, and the mirror-like surface obtained as a result of CMP on copper discs. The bottom portion shows the micrograph of the surface before (as lapped, left) and after CMP (right); however note that the improvement in surface quality due to CMP that is readily evident to the naked eye is not captured using statistical measurements, such as Ra (which is accessed at 5 nm).

1.2 Challenges and Proposed Approaches

Meeting these research objectives requires surmounting many challenges, some of which were enumerated by experts from academia, industry, and national labs at a recently organized (Nov. 2009) National Science Foundation (NSF) workshop on nanotechnology. These expert findings, which appear in a topical article authored by Bukkapatnam, *et al*. [27], can serve as a roadmap for future nanotechnology related research.

These expert recommendations have inspired the approaches developed in this research to address some of the challenges associated with quality assurance in nanomanufacturing. Some of the challenges and expert recommendations in perspective of this research are reproduced (verbatim) herewith. They are also summarized in Graphic III (adapted from a CIRP review article by DeChiffre, *et al* [8]).

Challenges in Nanomanufacturing

(as listed in [27])

1. *In situ* sensing

Nanoscale processes and systems pose many challenges for sensing:

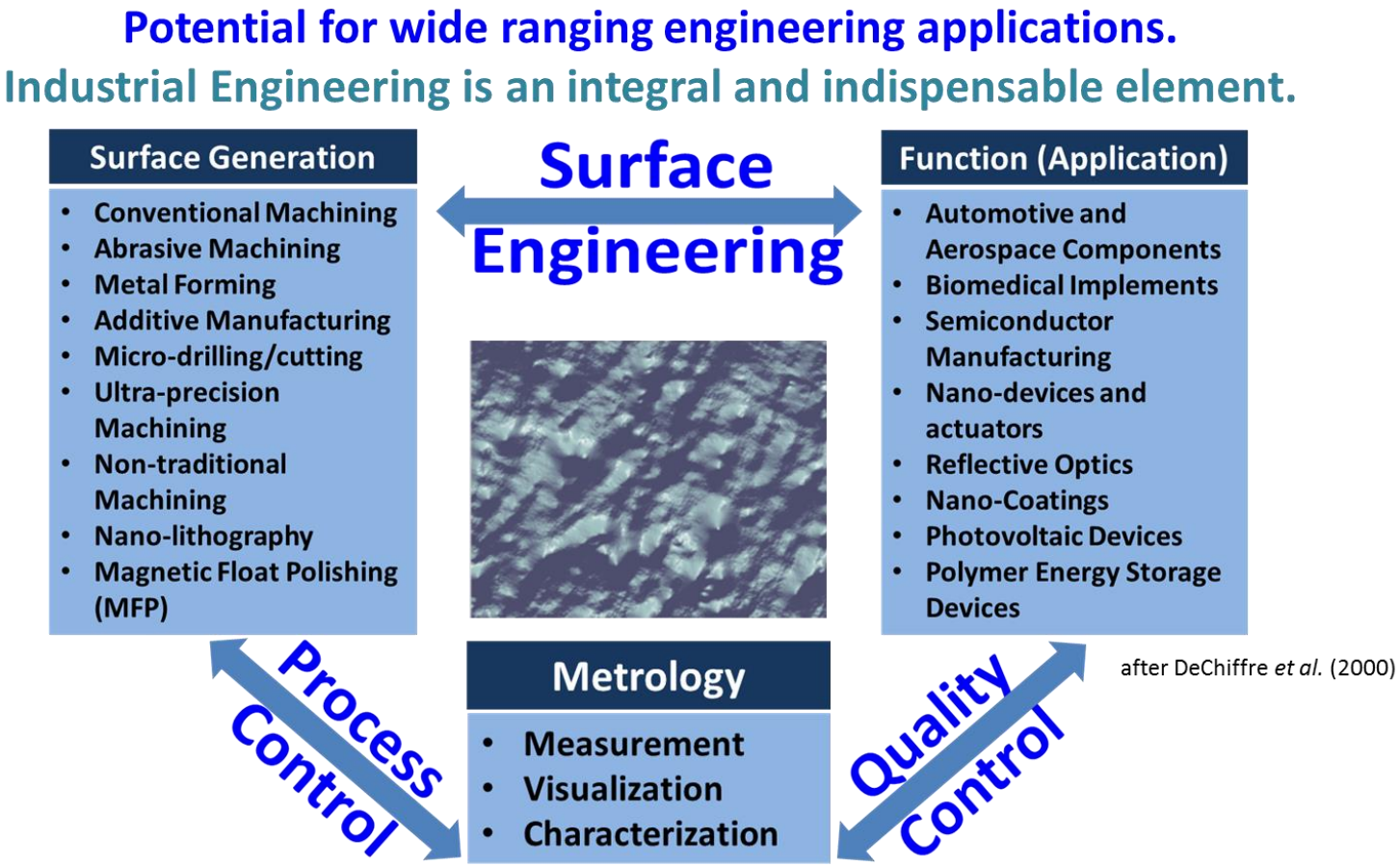
- accessibility to signal source is not easy;
- in situ sensing is almost impossible;
- signals are short, evanescent, and weak;
- quantization of signals makes transduction difficult; and
- signal-to-noise ratio is low.

2. Complex process physics

Mechanisms that cut across multiple scales make observation and characterization of nano-materials and nano-processes difficult.

3. Product quality assessment

Tools currently used in nanotechnology research labs offer atomic-level resolution for characterizing nanoscale surfaces but are barely adequate to meet the requirements of high-volume nanomanufacturing. For example, an Atomic Force Microscope (AFM) gives nearly atomic-level surface resolution, albeit at a very slow rate; it would be impossible to use AFM to characterize surfaces in commercial-scale high-rate operations.



Graphic III: Issues and challenges in manufacturing, particularly in the emerging nanomanufacturing realm involving surface engineering, process control, and quality control aspects which present opportunities for industrial engineers to make relevant contributions [27]. Graphic adapted from DeChiffre, *et al.* [8].

Recommendations for Nanomanufacturing Research
(as listed in [27])

1. A systematic set of methodologies of quality engineering should be developed to provide the following:
   - guidelines for the design and analysis of experiments to optimize nano-process settings;
   - **on-line monitoring and diagnosis techniques** to reduce nanoprocess variation and downtime during production; and
   - strategies for continuous improvement for high yield and quality.
2. **Physics-based statistical models** considering nanoscale hierarchical physics and nonlinearities should be developed. These realistic models will enable monitoring, diagnostics, prognostics, and reliability analysis.
3. **Instrumentation and analytical tools for comprehensive characterization of nano-materials for on-line process control** should be developed and characteristics required for specific applications defined.

1.2.1 Scope of the Research

Summary of the research satisfying objective 1

The first objective viz., to **develop approaches for monitoring and identification of incipient anomalies in nanomanufacturing** is accomplished by integrating *in situ* heterogeneous miniature sensors, such as force, vibration, and acoustic emission (AE) in close proximity to the (UPM and CMP) process. However, as noted in the challenges listed in the foregoing, due to the low signal-to-noise ratio (S/N) of acquired sensor signals and the sensitivity of the process to extraneous noise, the signal patterns depict complex nonstationary and nonlinear behavior. Consequently, the sensor signal patterns acquired from UPM and CMP

processes may not be amenable to application of conventional process monitoring approaches [27].

To overcome these impediments, we use concepts from various domains, such as chaos theory, neural networks, and Bayesian inference, in order to develop approaches for analyzing the acquired sensor signal patterns, and consequently detect anomalous process variations. For example, an integrated Recurrent Predictor Neural Network, Bayesian Particle Filter (RPNN-PF) prediction algorithm is used to detect incipient defects in UPM processing of aluminum discs, by invoking data from heterogeneous wired piezoelectric sensors.

Using this approach, process drifts leading to deterioration in surface quality of the machined sample were detected within 15 milliseconds (ms) of their inception. In comparison, conventional signal processing methods were largely reticent (> 30 ms) and prone to Type 1 (false alarm) and Type 2 (failing to detect) errors. Using this approach, we have addressed expert recommendations exhorting the **development of on-line monitoring and diagnosis techniques to reduce nano-process variation and downtime during production** [27].

For CMP process monitoring, we have designed a close-proximity wireless miniature (3 mm × 3 mm) MEMS multi-sensor array consisting of tri-axis vibration and sound sensors. Commercially available piezoelectric sensors, such as those used for the previous UPM study, are ill suited for CMP monitoring. This is because the presence of multiple rotating machine elements does not allow for easy integration of wired piezoelectric sensors. Although wireless sensors enable

close-proximity monitoring of the CMP process, their utility is impeded, as noted by Bukkapatnam, *et al*, due to low signal-to-noise ratio, propensity to quantization errors, low signal sampling rates, etc. [27].

Therefore, based on expert recommendations suggesting that **physics-based statistical models considering nanoscale hierarchical physics and nonlinearities should be developed** [27], the sensor signal patterns are integrated with a deterministic nonlinear differential equation process-machine interaction (PMI) model (documented in Appendix III) evocative of the various multi-scale phenomena that culminate in the complex process dynamics. By using the sensor signals in conjunction with a physical model, the effect of extraneous noise and inconsequential artifacts are negated.

Pertinently, the PMI model allows evolving process anomalies to be anticipated *a priori*. For example, after extensive verification of the PMI model solutions with experimentally acquired sensor data, accurate (R^2 > 75%) statistical models capable of timely identification of process drifts, such as polishing pad wear are suggested.

Summary of the research satisfying objective 2

In order to satisfy the second objective calling for development of **methods for surface quality assessment of nanomanufactured ultraprecision components,** we propose an algebraic graph theoretic approach for quantification of

surface morphology variations in nanomanufactured components. This approach represents the nano-surface morphology as a network graph.

As a result, the underlying topological relationships of the surface are preserved and can be subsequently quantified using graph theoretic invariants. This is advantageous because traditional surface metrology is restricted to statistical parameters, which do not elucidate the nano-surface topography [29]. Researchers have also reported that statistical quantification of nanomanufactured ultraprecision surfaces are often incongruous [30].

Additionally, since this graph theoretic approach primarily uses optical micrographs for characterization of the surface morphology, it overcomes several lacunae associated profile mapping techniques, such as laser interferometry, atomic force microscopy (AFM), scanning electron microscopy (SEM), etc., that can be time consuming, destructive and restricted to assessment of small areas. In contrast, contemporary optical inspection systems can scan (*ex situ*, off-line) a 300 mm semiconductor wafer in less than 1 minute [26].

Given these considerations, the graph theoretic approach is amenable to in line integration and can therefore be valuable as an **analytical tool for comprehensive characterization of nano-materials**, as suggested by experts [27]. We now motivate the possible socioeconomic benefits of this research.

1.3 Socioeconomic Impact of the Research

Why cannot we write the entire 24 volumes of the Encyclopedia Britannica on the head of a pin?

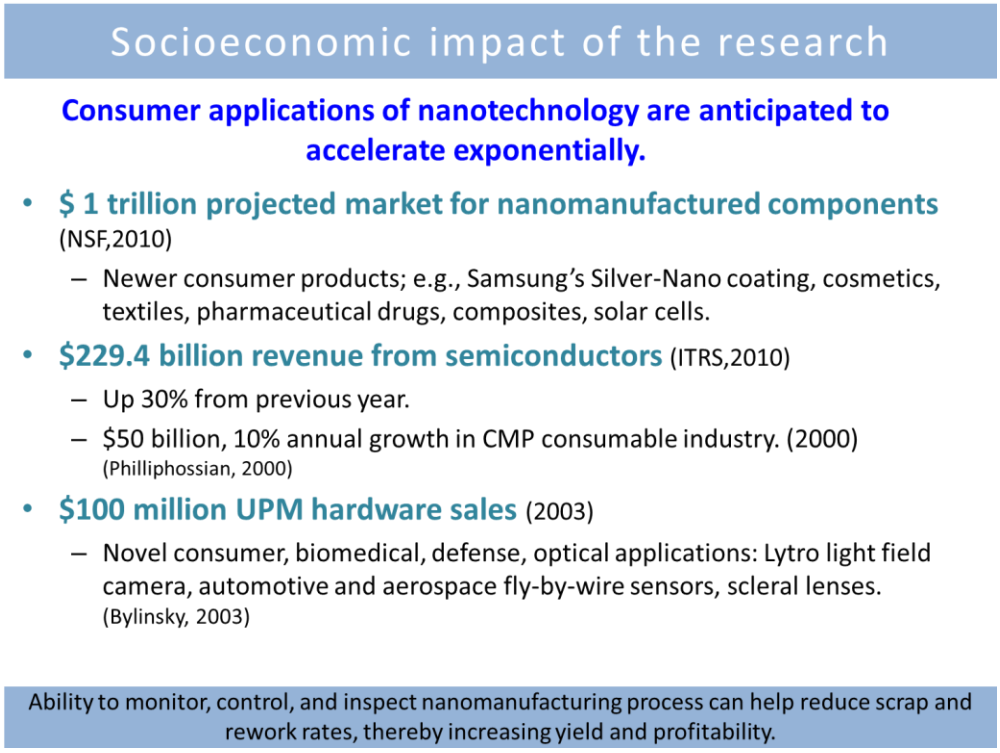
− Richard Feynman, *"There is plenty of room at the bottom"* [31].

If the automobile industry advanced as rapidly as the semiconductor industry, a Rolls Royce would now get half a million miles per gallon, and it would be cheaper to throw it away than to park it.

− attributed to Gordon Moore

Although *nanomanufactured* components, such as computer chips, hard drive components, precision optics, etc., influence our daily life in the same profound way as an automobile, most consumers barely think about how these articles are made. Indeed, nanotechnology enabled materials are increasingly being incorporated into mundane consumer products, such as cosmetics, textiles, anti-bacterial coatings in refrigerators and air purifiers [32, 33] (e.g., Samsung's Silver-Nano® coating), etc. Pertinently, the ability to successfully accomplish nanomanufacturing related objectives is vital to the national interest and impacts several consequential areas, from semiconductor devices in the electronics industry, biomedical implants in the healthcare industry, optical components for space and defense applications, to efficient photovoltaic cells for harnessing solar energy [34]. Some of these potential impact areas are listed in Table 1-3.

From a pecuniary perspective, NSF projects a worldwide market of $1 trillion in the next 10 years for nano-enabled products in the manufacturing, electronics, healthcare, and transportation sectors of the economy [35] (see also Graphic IV).



Graphic IV: Socioeconomic impact of the research in context of the emerging nanotechnology paradigm.

At this juncture, we introduce some terms related to nanotechnology and nanomanufacturing for the benefit of the general audience. The US government National Nanotechnology Initiative (NNI) strategic plan [36] quotes the following with respect to nanotechnology and nanomanufacturing:

Nanotechnology is the understanding and control of matter at dimensions between approximately 1 and 100 nanometers*, where unique phenomena enable novel applications. Encompassing* nanoscale *science, engineering, and technology, nanotechnology involves imaging, measuring, modeling, and manipulating matter at this length scale.*

A nanometer is one-billionth of a meter. A sheet of paper is about 100,000 nanometers thick; a single gold atom is about a third of a nanometer in diameter. Dimensions between approximately 1 and 100 nanometers are known as the nanoscale. Unusual physical, chemical, and biological properties can emerge in materials at the nanoscale. *These properties may differ in important ways from the properties of bulk materials and single atoms or molecules.*
United States National Nanotechnology Initiative Strategic Plan (pg.3). [36]

The same document states the following goal identified by the US government as being vital to nanomanufacturing related research:

> ***R&D aimed at enabling scaled-up, reliable, and cost-effective manufacturing of nanoscale materials, structures, devices, and systems. Includes R&D and integration of ultra-miniaturized top-down processes and increasingly complex bottom-up or self-assembly processes.***
> *United States National Nanotechnology Initiative, Strategic Plan (pg.5).* [36]
>
> ***Provide, facilitate the sharing of, and sustain the physical R&D infrastructure for nanoscale fabrication, synthesis, characterization, modeling, design, computation,*** *and hands-on training for use by industry, academia, nonprofit organizations, and state and federal agencies…*
> *United States National Nanotechnology Initiative Strategic Plan (Goal 3, pg.29).* [36]

Though *nanotechnology* is a relatively new (1983) term, the concept was suggested in 1960 by Richard Feynman in a seminal article – *There is plenty of room at the bottom* [31]. In this article, Feynman challenges, *'Why cannot we write the entire 24 volumes of the Encyclopedia Britannica on the head of a pin?'* Feynman not only envisions atomistic scale manufacturing (printing, manipulation, machining, bionics), but also suggests how these could be achieved, and the constraints we might face (e.g., he devotes a section on lubrication at the atomic scale).

Table 1-3: Potential impact of nanotechnology (after Corbett, *et al.* [34]).

Technology	Present Impact	Potential Impact
Dispersion and Coatings	• Thermal barriers • Optical (visible and UV) barriers • Imaging enhancement • Ink-jet materials • Coated abrasive slurries • Information-recording layers	• Targeted drug delivery/gene therapy • Multifunctional nano-coatings
High Surface Area Materials	• Molecular sieves • Drug delivery • Tailored catalysts • Absorption/desorption materials	• Molecule-specific sensors • Large hydrocarbon or bacterial filters • Energy storage • Grätzel-type solar cells
Consolidated Materials	• Low-loss soft magnetic materials • High hardness, tough WC/Co • cutting tools • Nano-composite cements	• Superplastic forming of ceramics • Ultrahigh-strength, tough structural materials • Magnetic refrigerants • Nano-filled polymer composites • Ductile cements
Nano-devices	• GMR read heads	• Terabit memory and micro-processing • Single molecule DNA sizing and sequencing • Biomedical sensors • Low noise, low threshold lasers • Nanotubes for high brightness displays
Additional Biological Aspects	• Bio-catalysis	• Bioelectronics • Bio-inspired prostheses • Single-molecule sensitive biosensors • Designer molecules

1.3.1 The need for nanoscale precision

Research in the last century has enabled the continual tightening in manufacturing accuracy specifications. For example, in the early 1980s an exponential trend in machining accuracy was observed by Taniguchi [2]. This observation, depicted in the so-called Taniguchi curves (Figure 1-1), suggested that machining accuracies in the nanometer range (10^{-9} m) would be achievable before the onset of the new millennium. Professor Taniguchi consequently coined the manipulation of material in the nanometric range as *nanotechnology* [34].

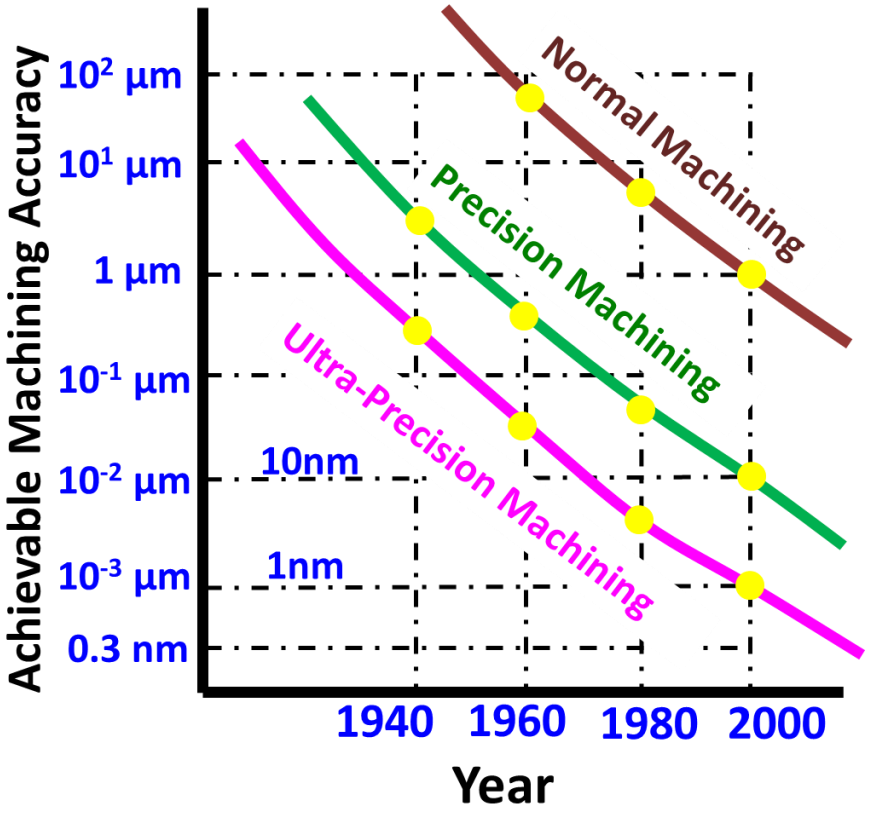


Figure 1-1: Taniguchi curve [2] (1983) predicting the exponential improvement in machining accuracy.

McKeown [12] motivates the need for precision manufacturing from economic and scientific perspectives. McKeown cites six reasons as prime movers in the continual tightening of manufacturing accuracy specifications:

i. Promote ease of assembly.

ii. Reduce scrap, rework, and inspection.

iii. Promote interchangeability.

iv. Improve wear life and fatigue life.

v. Achieve miniaturization.

vi. Advance science and technology.

As an illustrative example, it was observed that significant improvement in performance of aerospace components, such as jet engines and turbine gearboxes could be achieved by incorporating greater levels of precision during their manufacture [12]. For example, as shown in Figure 1-2, on tightening the surface roughness (Ra) specification of turbine blades from 500 nm to 300 nm, the compressor efficiency of later Rolls-Royce Trent series engines improved by almost 2%, which is comparable to the improvement achieved between design generations (e.g., Rolls-Royce RB series vs. Trent series) [12]. Similarly (see Figure 1-3), on reducing the individual tooth error in an aerospace transmission from 3 - 4 μm to below 1 μm, increase in torque efficiency from 50% to close to 90% was achieved [12].

Another example of such an exponential trend (Figure 1-4) in manufacturing specifications, is the doubling of number of transistors on integrated circuits approximately every two years (Moore's Law) [37].

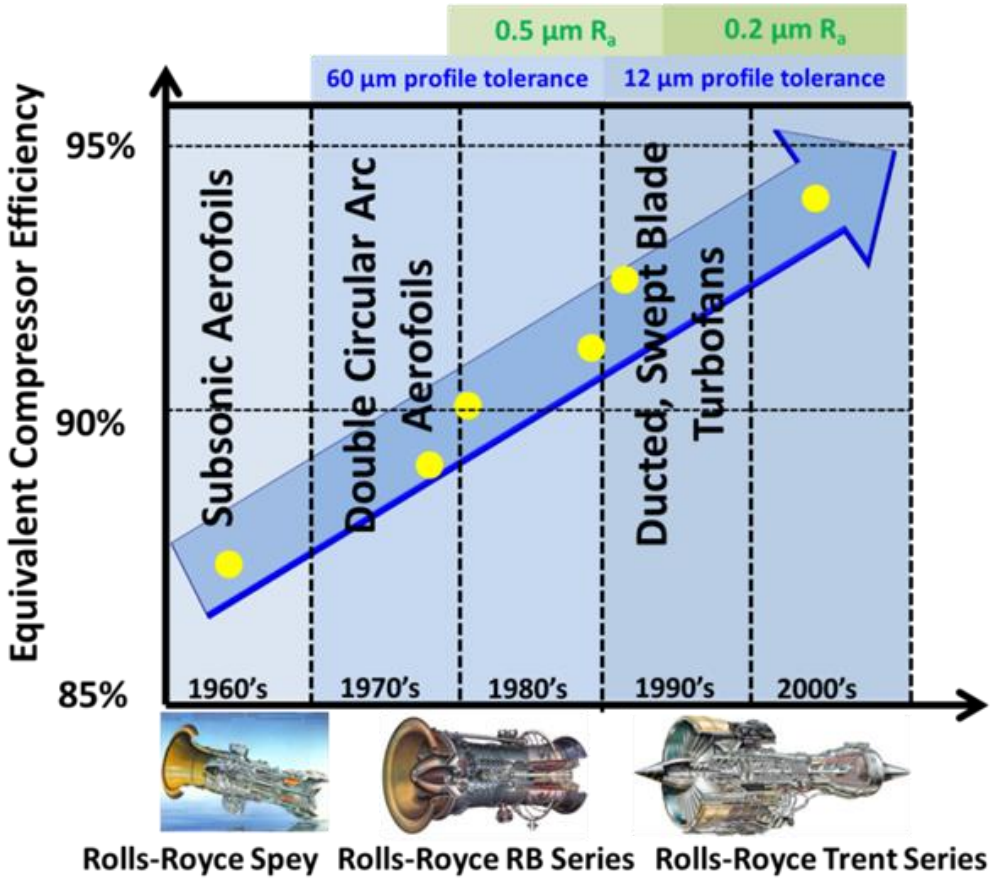


Figure 1-2: Improvement in compressor efficiency over last five decades.
The compressor efficiency of jet engines has continuously improved over the last five decades due to advancement in design and technology. However, a significant improvement in compressor efficiency can also be achieved by tightly controlling the surface quality and profile tolerances of jet engine blades (Rolls-Royce corporation, graphic adapted from [12]).

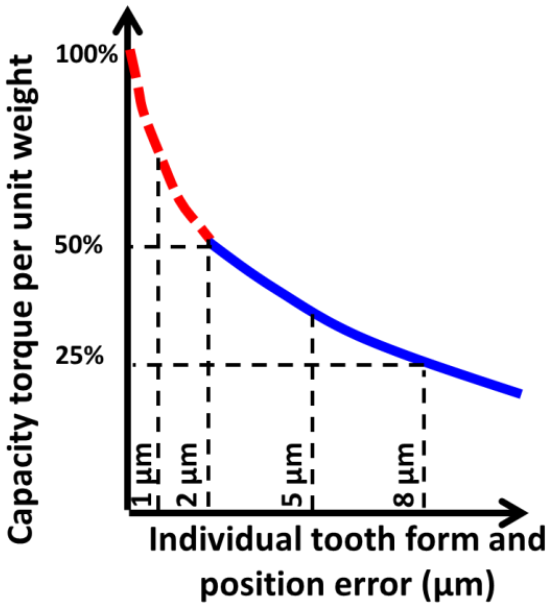


Figure 1-3: Improvement in torque capacity due to increased precision in manufacturing.
The torque capacity of an aerospace gearbox can be improved (almost exponentially) by refining the precision of the gear tooth profile and minimizing errors in loci without extensive redesign. (Rolls-Royce corporation, graphic adapted from [12]).

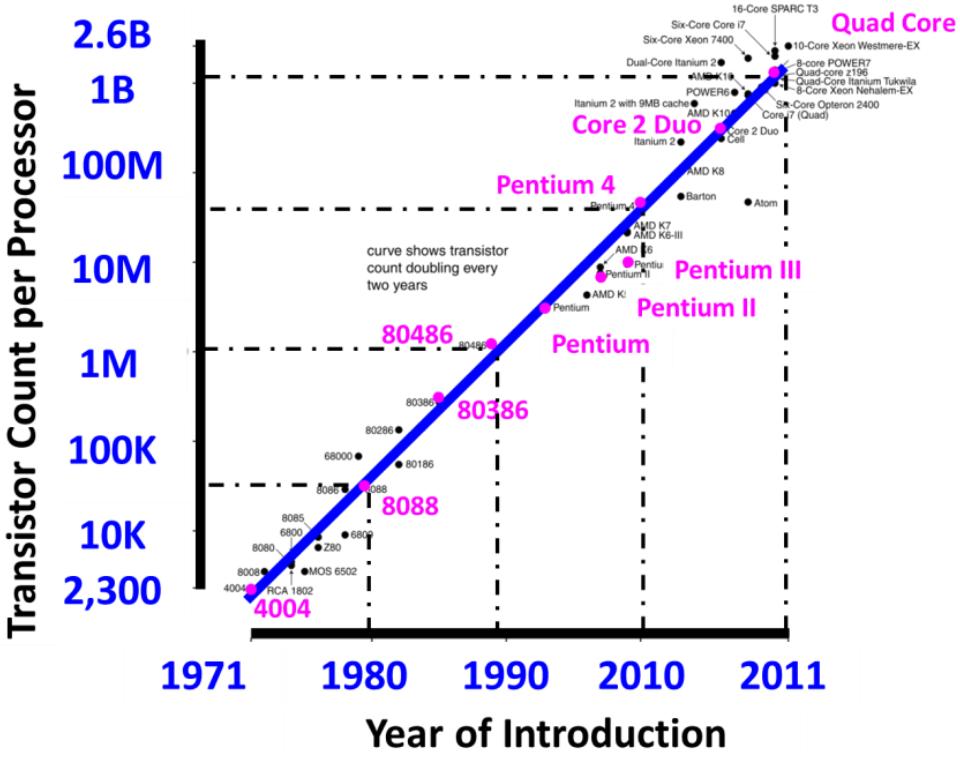


Figure 1-4: An illustration of Moore's law.
The doubling of semiconductor transistor counts on an integrated chip approximately every two years [38].

1.3.2 The need for novel process monitoring approaches in nanomanufacturing

Despite the nanomanufacturing processes studied in this research being distinctive from conventional processes, such as turning or milling, which one encounters on a daily basis, the issues of the broadest consequence are almost identical. Notwithstanding the sophistication involved in nanomanufacturing, the classic industrial engineering concern, *get it done faster, cheaper, and better*[3] is not outmoded [39]. It is indeed even more significant in the case of nanomanufacturing, because as noted previously, the economic stakes are many orders of magnitude greater. Additionally, our physical understanding of some of these processes is not yet complete [27].

Our dilemma is further compounded when we find that owing to the complex physical phenomena manifest at the nanoscale [27, 35], few techniques from the generic IE toolkit, such as statistical process control, time series analysis, linear programming, etc., are at best useful for a peripheral understanding of the problems encountered [27, 28]. Therefore, in the case of nanomanufacturing, the industrial engineer has to devise novel ways of getting it *done faster, cheaper, and better.*

The central theme of this dissertation stems from the quality assurance perspective in nanomanufacturing processes.

[3]F. W. Taylor, "On the art of cutting metals," *Transactions of the ASME, Annual Meeting,* 1907.

A brief introduction to UPM and CMP processes

Ultraprecision machining (UPM) and chemical-mechanical planarization (CMP) are employed for achieving surface finish in the nanometric range (< 100 nm) **by subtractive means** on a variety of materials encompassing the domain of metals, non-metals, semiconductors, and composites. However, any similarity between the twain ends with their classification as subtractive processes. A delineation of the method of material removal is perhaps most conducive to contrast UPM and CMP.

The volume of material removed is significantly smaller (few tenths of a milligram per minute) in the case of CMP compared to UPM (tens of grams per minute). In general, UPM is analogous to turning operations conducted on a specialized lathe using natural diamond cutting tools; while CMP is similar to loose abrasive material removal processes such as lapping.

In UPM, material is removed (typically) as a single point cutting operation. A single crystal diamond (SCD) tool is used to remove material in the form of fine chips at depths in the 2 – 50 micrometer (μm) range on an extremely precise lathe having the following salient characteristics [1, 17, 18, 40]:

- **Enhanced vibration and noise isolation**: Stiff granite machine structure, pneumatically elevated machine supports, and acoustic/vibration isolators are used to eliminate errors due to extraneous disturbances. In some cases, the machine might be (quite literally) buried in a ‘pit’ instead of being elevated on the floor for vibration and acoustic isolation [13, 14, 40].

- **Minimization of errors due to inertia of machine elements**: The machine elements are designed in a manner such that their inertial moment is balanced or compensated. Precision straightedges are mounted to measure the deflection of machine slides, so that corrective action can be taken [13, 41].
- **Minimization of errors due to backlash and friction**: Air-bearing spindles mechanically coupled (via precision flexible couplings) to brush-less DC motors, and precision air (cushion) slides are used to eliminate backlash and friction errors. Conventional ball screws and hydrodynamic journal bearings are avoided [4].
- **Integrated precision metrology**: Multiple laser interferometers, encoders, and process condition monitoring sensors (e.g., temperature) are incorporated to ensure precision positioning and movement (of the tool and workpiece).
- **Precision work-holding arrangements**: For error free location and clamping of the workpiece, vacuum or magnetic chucks are used. These prevent the possibility of imbalance, warping, and misalignment of the workpeice often observed with mechanical fastening.
- **Precision cutting tools**: Ground and lapped single crystal diamond (SCD) tools with stringently controlled signature (i.e., geometry) and surface quality are used.
- **Minimization of thermal errors**: Machine elements (tool slides, tool holders, fasteners, etc.) are constructed from low thermal expansion materials,

such as Invar; coolant showers and temperature-controlled enclosures are used to minimize thermal deflection of the machine.

In the United States, Lawrence Livermore National Laboratory (LLNL) has been the pioneering institution in the design and development of UPM-capable machines [18]. Ultraprecision machines for turning 4 inch diameter spherical geometry lenses and optics, to 84 inch diameter non-cylindrical mirrors used for X-ray radiometry applications have been designed at LLNL [13].

For example, a UPM machine designed by LLNL is the vertical Large Optics Diamond Turning Machine (LODTM) [14, 40], which is purported to be one of the most accurate UPM machine available (2001) (see Figure 1-5). This machine is capable of turning 64-inch diameter workpieces for optical applications to a surface finish of 4.2 nm (Rq) and 28 nm form (geometric) accuracy. This machine utilizes multiple interferometers (7 in all) for tool movement with respect to the machine structure.

In addition, capacitive gauges are used to precisely locate the workpiece in reference to the machine coordinates. Aerostatic bearings are employed for the slideways (14 in all), and water-cooled pressurized oil radial and thrust bearings stabilize the rotational motion of the spindle. Connectors are made from materials with low coefficient of thermal expansion (e.g., Invar) to ensure thermal stability. Suppression of environmental noise is achieved using pneumatic isolators, independently mounted air compressors, and air conditioning equipment. The machine

operates in an acoustically isolated area where the air temperature is kept within 0.001° F.

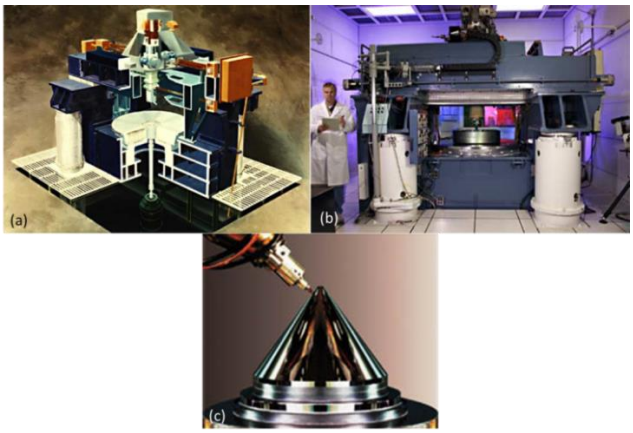


Figure 1-5 64 inch Large Optics Diamond Turning Machine (LODTM) at LLNL [14, 40]. (a) Schematic of the machine (b) Photograph of machine with workpiece loaded (c) Aspherical mirror for astronomy applications machined on LODTM.

In CMP, the physio-chemical action of nano-particulate abrasives (< 100 nanometer (nm) grain size), such as alumina (Al_2O_3) and silica (SiO_2), suspended in colloidal form in a chemically active solvent (typically, potassium hydroxide (KOH) or sodium hydroxide (NaOH)), are used to remove material as grains (too small to be visible to the naked eye) with a polishing-like mechanical action [19, 21].

Modern semiconductor integrated circuits (IC) are built layer by layer, with trenches and via(s) (connections) etched into each preceding layer using ultraviolet photolithography processes [20, 21]. Dielectric or metal is subsequently deposited on the patterned layers, usually by vapor diffusion methods. After deposition, it is necessary to polish each layer to specification using CMP. Control and monitoring of CMP is crucial, because if planarization is not accomplished to precision, then the dimensional thickness of subsequent layers will fall out of specifi-

cation, and consequently affect the functional integrity of the device [21]. For example, if the preceding layers are crooked (called *bird beak* structure in semiconductor manufacturing parlance [19]) the ultraviolet light used in photolithography patterning of following layers will be exceedingly out of focus. The bird beak structure is but one of several anomalies that can be induced during CMP [20].

Modern current integrated circuits have 7-10 such interconnect layers [19, 20], with each layer requiring a CMP step. Considering that the worldwide revenue from sale of semiconductors for financial year 2010 was recorded at $299.4 billion, up ~ 30% from the previous year [42], it is readily apparent that the economic implications from minimizing CMP related defects and anomalies are profound.

In contrast, UPM is a relatively niche market, with (hardware) sales in 2003 recorded at ~ $100 million, and is in the midst of a transition from the realm of national laboratories to large-scale commercial deployment [43]. Ultraprecision diamond turning machines, which were until recently designed and housed in national labs, and primarily meant for manufacturing components for so called *big science* projects, such as X-ray telescope mirrors, are now being scaled down and sold to small-medium private businesses [13, 14, 40].

Scaled-down UPM machine tools in the hands of private entrepreneurs are now being applied in previously unimagined consumer driven areas [43-46]. For example, a Boston area ophthalmologist uses ultraprecision diamond turning for

making scleral eye lenses customized to fit the ocular characteristics of the patient, and crafted with miniature grooves which can hold lubricating fluid [43]. In addition, these lenses are designed with complex spline geometries in order to dilate and contract with the eye. The complex geometry and micro-scale grooves enable the lens to be placed touching the cornea.

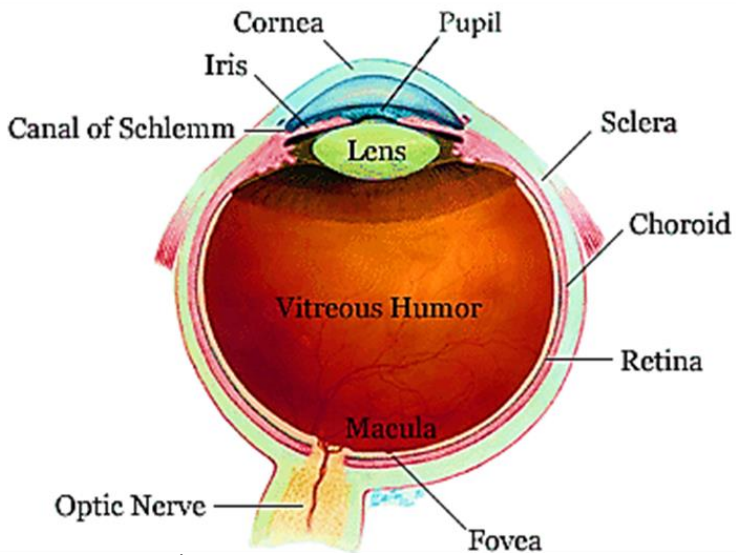
Pertinently, without the lubrication provided by the fluid held in the UPM-cut grooves, the lens sticks to the cornea. Hitherto the spline geometry and grooves had to be painstakingly cut by hand, scleral lenses were therefore prohibitively expensive. The specialized UPM process adapted for this application is credited with restoring eyesight to the functionally blind (see Graphic V). One also notes, the potential health hazard due to a miniscule defect caused by poor process control during the manufacture of such lenses.

Thus, given the wide applicability of UPM and CMP, **expeditious process prognosis can help alleviate inherent low yield/high rejection rate problems.** Significant socioeconomic incentives are consequently envisioned from preventing anomalies in their incipient stages in CMP and UPM.

Societal impact of UPM

Potential to impact life, not just lifestyle.

Ability to craft complex nano-scale features and geometries integral to bio-medical devices.

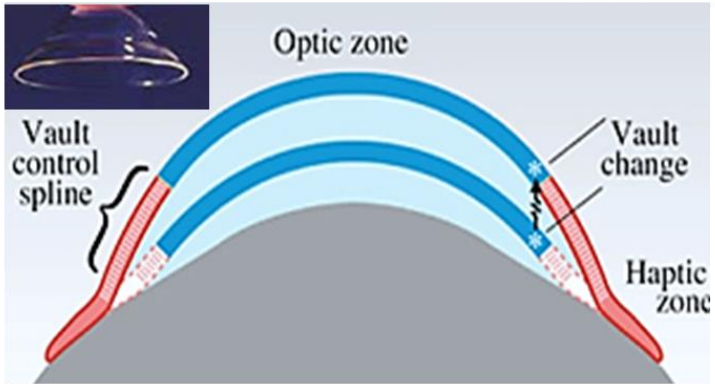


Courtesy, National Eye Institute

Closing In On Perfection Ultraprecision machine tools are putting manufacturers within nanometers of absolute accuracy.

By Gene Bylinsky
June 23, 2003

FORTUNE



P.Rosenthal, *Evolution of an Ocular Surface Prosthesis,* Contact Lens Spectrum Magazine. 12/01/2009

Nano size grooves cut into surface of scleral lenses using UPM are able to hold tears. Avoids the lens from sticking to the eye and injuring it.

Graphic V: The societal impact of UPM. Ultraprecision lathes are used for shaping scleral lenses bearing intricate splines and grooves. These lenses can potentially impart vision to blind people with damaged corneas

1.4 A Synopsis of the research

In freshman engineering shop class, we were perhaps asked to turn a steel bar to shape (as done here at Oklahoma State) on a lathe. To check whether the part was 'in spec'; we regularly stopped the lathe and measured the relevant dimensions using a vernier caliper. If the part was 'off spec', we tried to rework it. If unfortunately the errors were unfixable, we started anew on a fresh steel bar. In the worst case, our total loss amounted to a few dollars.

Unlike conventional manufacturing, where we might have the wherewithal to rework or scrap a part often entailing moderate losses, we rarely have such liberty when dealing with nanomanufactured ultraprecision components. For example, an undetected non-conforming telescope mirror (due to surface and geometry defects) might cost many millions to replace [47]. The thrust, therefore, is to **minimize the *ex situ* monitoring of the process and expeditiously detect the onset of process faults in order to avoid expensive failures**.

One of the most viable means to achieve this goal is to integrate *in situ* sensors in the process and examine the emerging signals. Nevertheless, *which type sensors are suitable for our application?*

The above question can be answered based on experience of previous researchers [48-52]. Particularly, the works of Dornfeld, *et al.* [48-50] indicate that accelerometers and Acoustic Emission (AE) sensors are best suited for applications where subtle changes, such as surface finish and sub-surface damage, which occur in the nanometer range, have to be monitored (see Figure 1-6). Therefore, in

this work, signals acquired from accelerometers (vibration sensors) are used widely.

The next step is to correlate the characteristics of the output signals with process anomalies. By doing so, we can begin to consider corrective actions as the process shows signs of malfunction or *out of control condition*.

However, these aims hinge on the central assumption that the *acquired sensor signals are evocative of the process* and not largely the result of accompanying noise from factors extraneous to the process dynamics, such as kinematic errors from machine elements and environmental conditions. Once we are assured of a *high signal to noise ratio* from the sensor data, we might even think ahead and consider *predicting when the process is about to malfunction*.

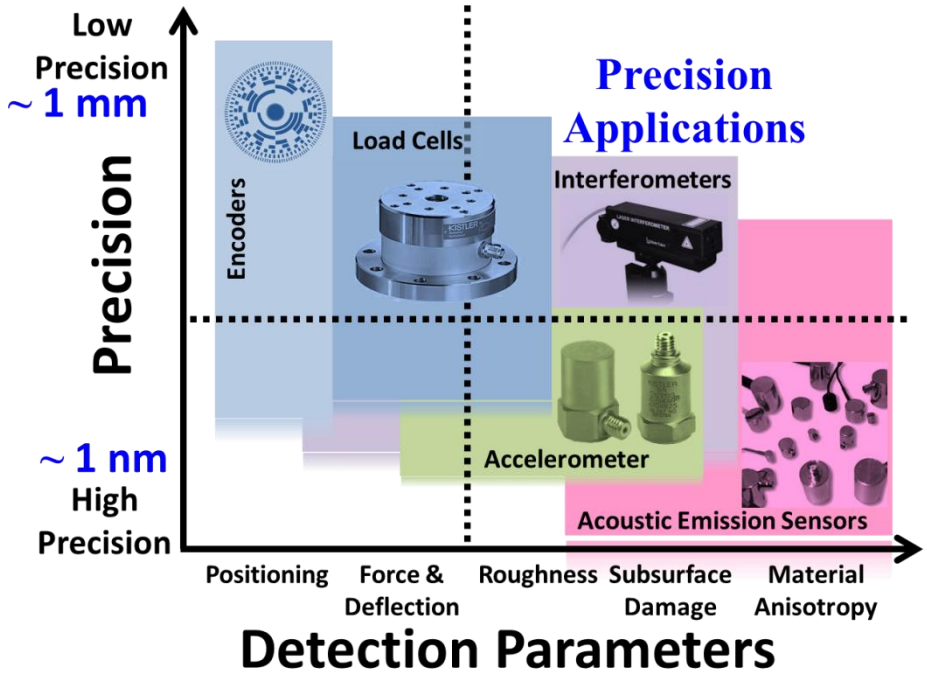


Figure 1-6: Sensor types and concomitant detection regimes (after Dornfeld, *et. al*.[50, 53])

In summary, to reliably monitor the process *in situ* using sensors, one has to ensure the following conditions are met:

1. Acquired sensor signals should have sufficiently high signal to noise ratio.
2. Sensor data should depict consistent and replicable trends.
3. Algorithms for signal monitoring should correlate actual physics of the process with relevant signal characteristics.

Optimistically, when we have achieved these prerequisites, the components that are produced would be largely conforming to specification. Nevertheless, *how could we be sure that the component is still 'in spec'?*

Perhaps, we will proceed to inspect the part that was produced. Analogous to when we used a vernier caliper in freshman engineering, we might examine the surface of the nanomanufactured component, albeit using sophisticated methods, such as laser interferometry, stylus-based (contact) profiling, optical microscopy, atomic force microscopy, etc.

What happens if we find that some of the quantifiers we have inherited from traditional manufacturing practices are not suitable for quality assessment of nanomanufactured ultraprecision components? Therefore, it is **essential to devise a reliable quantification approach that can capture aspects of the surface morphology critical to the functional performance of the component.**

Unfortunately, given the complexity and uniqueness of the processes studied, a universal strategy cannot be applied. For example, in UPM the material removal action is akin to single point plastic deformation (assuming material removal in

UPM to be a similar to conventional metal cutting constitutes a grave error, nonetheless this analogy is (tenuously) defensible for broad didactic purposes), while in CMP the material removal action is mainly due to abrasive wear.

It is rather obvious that apart from the process mechanics, the physical setup and constraints in terms of the machine tools, process conditions, and environmental noise for these processes also differ. These considerations influence the nature and type of sensors that can be used for monitoring. For example, the CMP process uses chemically active slurries, which can easily damage expensive wired piezoelectric sensors. Hence, in CMP it is imperative to design a sensor network that is wireless, non-contact, and uses easily replaceable inexpensive sensors.

One of the essential premises of this dissertation is that nanomanufacturing processes, such as UPM and CMP, can be monitored with high fidelity using sensor data.

We now provide a brief overview of the dissertation encompassing Chapter 2 through Chapter 4 that essentially enshrines the main contributions of this research.

1.4.1 Real-time Identification of Incipient Surface Morphology Variations in Ultraprecision Machining (UPM) Processes

In chapter 2, we address expert recommendations concerning on-line monitoring techniques for real-time diagnosis of nanomanufacturing process variations [27]. We have devised an approach that invokes neural network and Bayesian learning techniques for *in situ* detection of UPM process variations, by integrating data from heterogeneous sensors, such as force, vibration, and acoustic emission (AE).

Challenges in nanomanufacturing research [27]

Nanoscale processes and systems pose many challenges for sensing:

- Accessibility to signal source is not easy;
- *In situ* sensing is almost impossible;
- Signals are short, evanescent, and weak;
- Quantization of signals makes transduction difficult; and
- Signal-to-noise ratio is low.

Expert recommendations for nanomanufacturing research [27]

A **systematic set of methodologies of quality engineering should be developed** to provide the following:

- Guidelines for the design and analysis of experiments to optimize nanoprocess settings;
- **On-line monitoring and diagnosis techniques to reduce nanoprocess variation and downtime during production**; and
- Strategies for continuous improvement for high yield and quality.

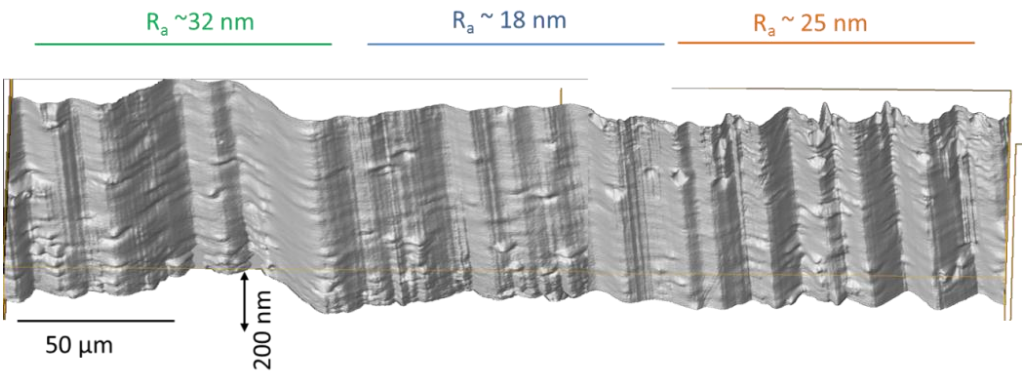


Figure 1-7: Surface roughness profile of a section of UPM processed Al 6061 disc (obtained using a laser interferometer).

Consider Figure 1-7, which shows a 300 µm (1 µm = 10^{-6} m) long contiguous surface profile of a UPM-processed aluminum 6061 (Al 6061) disc. It is evident that the surface roughness (Ra) changed significantly in a short span amounting to a few milliseconds (ms) of processing time, in this case approximately 30 ms.

Unlike, in conventional manufacturing such evanescent *out of control* process states have a substantial impact on the quality of parts produced in UPM. In this context, this particular Al disc could be a telescope mirror measuring several inches (if not feet) in diameter, with allowable deviation in surface roughness specifications amounting to Ra < 10 nm [13, 41].

Under such circumstances, we might have to rework, or worse, scrap the part, leading to waste of several man-hours of specialized labor, costly tools (diamond tools), machine downtime, setup, materials, etc.,– in all a very expensive preposition.

Clearly, such abrupt changes in surface morphology are indicative of process anomalies, which can have deleterious consequences and lead to considerable capital losses if not detected and stopped at an early stage. It would be valuable, therefore, if we could find approaches to avoid such eventualities by *predicting the onset of process anomalies so that timely corrective action can be taken.*

Goal 1
(Satisfying Objective 1)

To suggest a method capable of real-time in situ detection of incipient anomalies in UPM process**.**

However, *in situ* prediction of developing process anomalies from sensor data is not easily accomplished, mainly because:

i. Traditional statistical signal processing techniques are reticent in detecting short time drifts (sampling window is small).

ii. Typical control charting approaches (e.g., X-bar chart) are particularly stymied if the underlying distribution of the signal is non-Gaussian, time-dependent, nonlinear, and nonstationary (i.e., IID assumptions are violated).

iii. Conventional signal processing techniques, such as Fourier transform spectra are easily contaminated by noise, and are therefore ill suited for analyzing complex signals with broadband characteristics, such as those observed in UPM.

iv. Data from more than one type of sensor may be needed to predict the onset of UPM process anomalies. However, incorporating information

from multiple heterogeneous sensors is not tractable using traditional methods.

Challenges

1. Traditional statistics-based quality control and spectral analysis methods cannot capture evanescent, nonlinear, and nonstationary signal patterns in a timely manner, and therefore do not lend toward incipient real-time detection of ultraprecision machining anomalies.
2. Information from multiple heterogeneous sensors is not easily integrated using traditional statistical methods.

Research approach and outcomes

In Chapter 2, we integrated heterogeneous sensors, such as piezoelectric vibration, force, and acoustic emission (AE) *in situ* into an ultraprecision diamond turning machine (DTM), in an attempt to detect the onset of anomalies (e.g., Figure 1-7). The machine itself is similar to a lathe, albeit several orders of magnitude more precise and mechanically rigid to minimize vibrations.

Al 6061 discs are finished to a surface roughness (Sa) in the 15 – 25 nm range using a natural diamond single point cutting tool (single crystal diamond (SCD)) on this machine. The physical setup of the DTM allowed for acquisition of high fidelity wired piezoelectric vibration, force and AE sensor signals, with sampling rates > 10 kHz.

In order to surmount the foregoing enumerated challenges, we developed a method (Figure 1-8), which uses a combination of a neural network called recur-

rent predictor neural network (RPNN) and Bayesian particle filtering (PF) techniques capable of *online prediction of UPM process anomalies using inputs from multiple heterogeneous sensors*, and hence prevents production of defective parts.

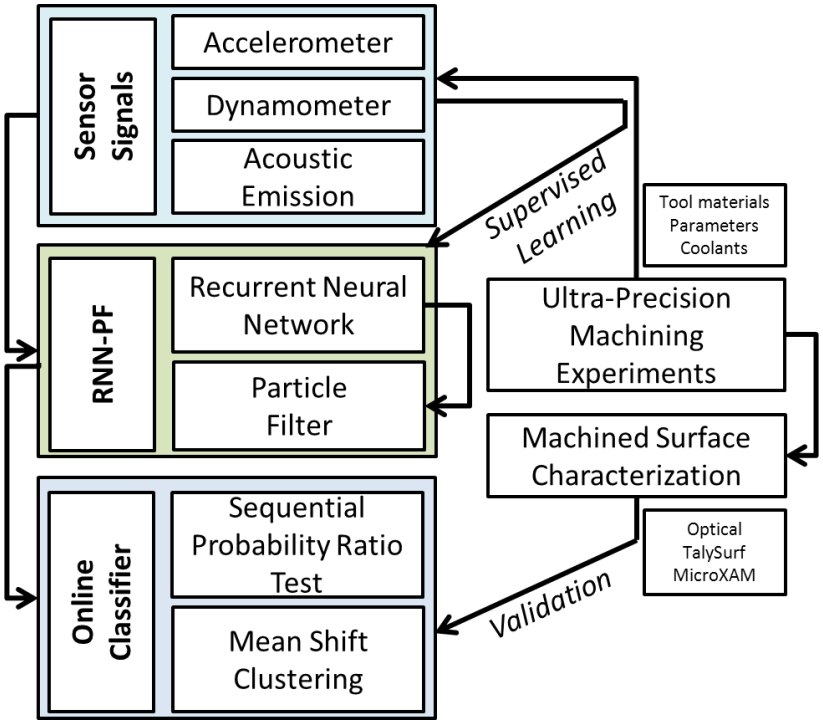


Figure 1-8: Summary of the approach used for *in situ* detection of UPM process anomalies.

We (artificially) introduced defects into the process in a controlled manner and studied their influence on the process (chip type) and component surface quality, while simultaneously acquiring heterogeneous sensor data.

The combined RPNN-PF prediction approach imparts processing advantages and allows responsive prediction of incipient process anomalies. We show that the **RPNN-PF approach significantly outperforms traditional statistical signal processing methods**. Typically, the delay in capturing process anomalies was ~ 15 ms using this RPNN-PF approach, compared to over 30 ms using conventional

statistical techniques. In this study, we assumed extraneous noise effects were negligible; this is reasonable given:

i. The close proximity and high sampling rate of the sensors used.

ii. The precise nature of the machine (air bearings, air slides, and electrically isolated brushless DC motors are built into the setup).

iii. The care taken in maintaining a controlled environment during processing – experiments are conducted when other machinery (lathes, milling machines) in vicinity of the setup are not operational (the machine rests on a vibration-isolated three metric ton granite base).

iv. The consistency observed in the acquired signal patterns and surface morphologies on replication of experimental conditions.

Hence, in this particular study little attempt was made to delineate or model the process dynamics. The specific tasks executed to accomplish the research outcome from this chapter are listed herewith.

Specific Tasks

1. Ensure surface finish in the nanometric range (Sa < 25 nm) with near-specular characteristics for UPM processed Al 6061 discs.
2. Instrument UPM setup with vibration, force, and AE sensors in close proximity to the tool-chip interface.
3. Develop a combined recurrent predictor neural network – Bayesian particle filtering prediction (RPNN-PF) technique that incorporates data from heterogeneous sensors, such as force, vibration, and AE sensors for real-time identification of incipient anomalies.
4. Test the approach using synthetic signals, and subsequently (statistically) characterize the sensitivity of the approach to different types of signal variations.
5. Design UPM experiments to introduce controlled surface defects, and subsequently study their effects on sensor signals.
6. Apply the RPNN-PF approach for early detection of UPM process anomalies.

1.4.2 Process-Machine Interaction (PMI) Model-based Monitoring of the Chemical Mechanical Planarization (CMP) Process using Wireless Vibration Sensors

In Chapter 3, we invoke a deterministic process-machine interaction (PMI) model for identifying evolving process anomalies (the PMI model formulation is documented in Appendix III). The PMI model combines multi-scale phenomena active in the CMP process in the form of a two-degree of freedom nonlinear differential equation.

We invoke and verify the PMI model using experimentally acquired MEMS wireless vibration signals. By using this physics-based model for explaining CMP process dynamics, we identify evolving CMP process anomalies and therefore avoid low yield situations. Thus, in this chapter we have addressed two expert recommendations as stated below, ultimately culminating in hierarchical physics-based models for explaining dynamics of nanomanufacturing processes [27].

Challenges in nanomanufacturing research [27]

Nanoscale processes and systems pose many challenges for sensing:

- Accessibility to signal source is not easy;
- *In situ* sensing is almost impossible;
- Signals are short, evanescent, and weak;
- Quantization of signals makes transduction difficult; and
- Signal-to-noise ratio is low.

Mechanisms that cut across multiple scales make observation and characterization of nano-materials and nano-processes difficult.

Expert recommendations for nanomanufacturing research [27]

A **systematic set of methodologies of quality engineering should be developed** to provide the following:

- **guidelines for the design and analysis of experiments to optimize nanoprocess settings;**
- **on-line monitoring and diagnosis techniques to reduce nanoprocess variation and downtime during production**; and
- strategies for continuous improvement for high yield and quality.

Physics-based statistical models considering nanoscale and hierarchical physics, and nonlinearities should be developed. **These realistic models will enable monitoring, diagnostics, prognostics, and reliability analysis.**

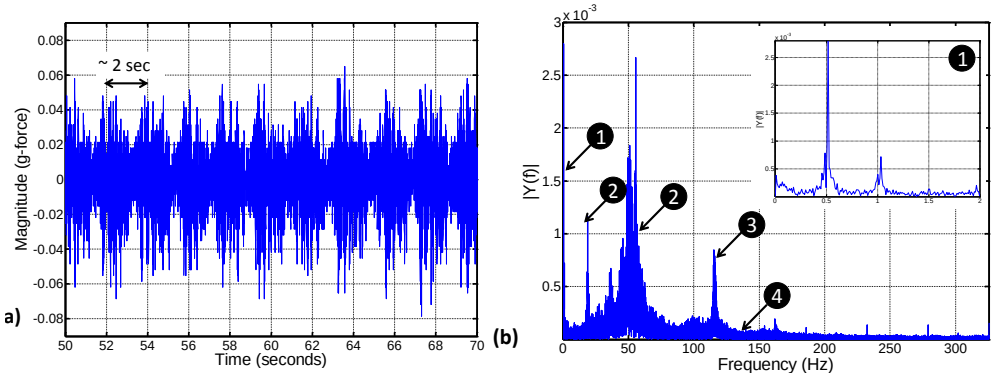


Figure 1-9: Typical vibration sensor data in the tangential (V_X) direction obtained during CMP.

Conditions are as follows: 30 RPM spindle speed, 150 RPM platen speed, and 2 lb. down force.
(a): Time series plot showing ~ 20 sec of the total (180 sec) data, period between prominent low frequency components ~ 2 seconds.
(b): FFT frequency spectrum of the data, showing presence of four main components. Low frequency component (0.5 Hz – 1 Hz) seen in (a) is marked by 1. A zoomed-in view of component 1 is shown in the inset.

Consider Figure 1-9, which shows the MEMS wireless vibration signal patterns (sampling rate ~ 700 Hz) obtained during CMP. The time series in Figure 1-9(a) shows prominent low frequency periodic behavior interlaced with high frequency components. These frequency components are evident in the fast Fourier transform (FFT) of the signal shown in Figure 1-9(b), some of which are broadband in nature (e.g., component marked 2 in Figure 1-9(b)), while others occur at relatively higher frequency regions (component 3).

Sensor-based process monitoring with signals depicting such complex nonlinear and nonstationary facets can be challenging, because:

i. As with the UPM case, traditional statistical process monitoring techniques are of limited utility.

ii. We are not certain what aspects (components 1 through 4 in Figure 1-9(b)) of the sensor signal patterns are evocative of the process. That is, **we cannot tell for sure which frequency components of the signal are due to noise, and which components are representative of the process and therefore useful for tracking process variations.**

For example, a curious observer might ask:

- *Why does the signal in Figure 1-9(a) show prominent low frequency beat-like patterns?*
- *What causes these low frequency patterns?*
- *Are they relevant to the process? How do you know?*[4]

Goal 2
(Satisfying Objective 1)

Identify and statistically quantify those aspects of MEMS wireless sensor signal components that are relevant to the CMP process dynamics, and therefore useful from a real-time monitoring perspective.

In other words, we need to address the following questions:

i. How do we ensure that the acquired vibration sensor signal has some useful information pertaining to the process?

ii. Which particular aspect (or frequency component) of the signal is sensitive to changes in process inputs, and is therefore relevant for process monitoring?

[4] Readers acquainted with this research might recollect that these questions were first put by (late) Dr. R. Komanduri. This footnote is meant to serve as a reminder, especially to this student, of his profound impact (even in absentia) on every aspect of this research.

iii. What process conditions correspond to the different aspects of the signal, i.e., if the polishing load is changed, which (one or more) frequency component(s) change(s)?

iv. Are the signal features sensitive to changes in process conditions, and are these features *statistically significant and replicable*?

The challenges that arise in answering these questions are listed herewith, and the empirical tests conducted for delineating relevant signal characteristics are documented in Appendix II.

Challenges

1. Wireless vibration signals obtained during CMP process have low signal to noise ratio and depict complex nonlinear, nonstationary, and broadband characteristics. Under such circumstances, conventional statistical process monitoring techniques are of limited value.
2. Given the complex signals emerging from the process, it is not known what aspects (features or frequencies) of the signal are relevant to the process.

Consider Figure 1-10(a) which shows a deliberately glazed CMP polishing pad, i.e., the surface of the polishing pad was caked with slurry debris, used for processing of Cu wafers. Polishing with such a degraded pad causes damage to the wafer surface, as evidenced by the nanoscale scratch shown in Figure 1-10(b). When we compare the frequency spectra of MEMS vibration signals acquired while polishing under normal pad conditions (Figure 1-11(a)) to the glazed pad condition (Figure 1-11(b)), we notice an increase of ~ 35% to 40% in the magni-

tude in the 115-120 Hz region (shaded region). *How can we explain this behavior in terms of the process dynamics?*

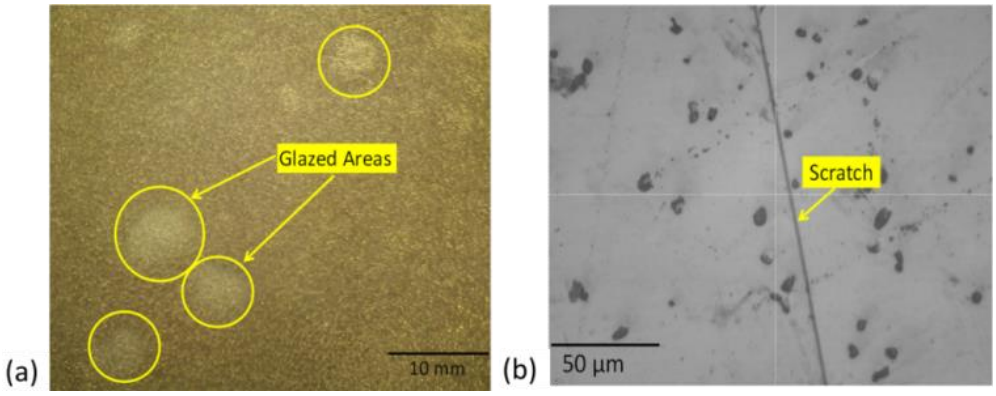


Figure 1-10: (a) Glazed pad after 12 minutes of CMP, and (b) Scratches on wafer observed at the end of 12 minutes of CMP.

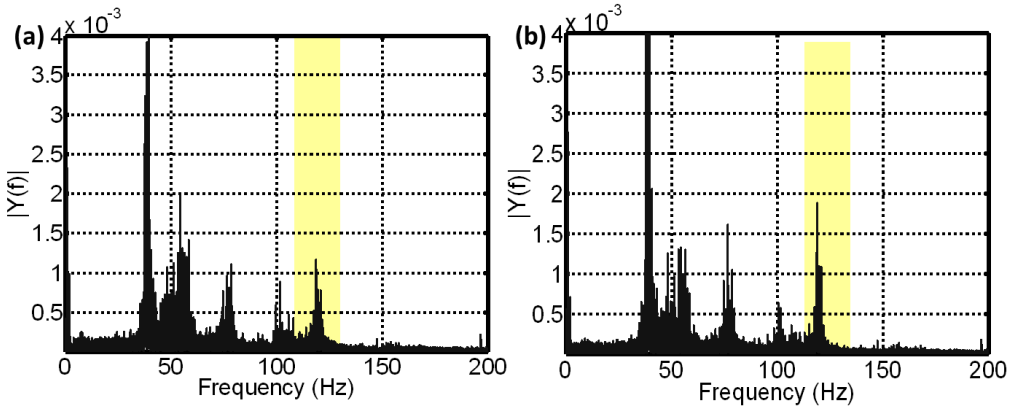


Figure 1-11: FFT of experimental vibration data obtained for (a) new pad vs. (b) 12 minutes used pad (glazed).

Oftentimes, there is no dearth of high fidelity time-series sensor data occupying several gigabytes of memory. Given the large volume of information, the approach typically used is to analyze the data applying sophisticated signal processing tools in order to correlate sensor signal patterns with process conditions.

Although, correlations between the sensor data features (i.e., statistical and spectral quantifiers) and process dynamics (see Appendix II) are valuable for

understanding the source of sensor signal components, we are nonetheless essentially *listening to one side of the story*. As *correlation does not imply causation*, it would therefore be valuable to gain an insight as to *why* the sensor data depicts certain types of behavior.

This implies we must develop the capability to identify the specific physical reasons (connected with the process dynamics) sensor patterns. For example, this could mean delineating the process dynamics that led to higher energy content in a certain bandwidth of the vibration signals because of controlled variations introduced in the process (e.g., Figure 1-11).

As we will explain shortly, the ability to provide the physical reason for particular signal patterns can enable *a priori* identification of process anomalies from sensor signals. The goal therefore is stated as follows.

Goal 3
(Satisfying Objective 1)

Explain based on process dynamics and physical phenomena, the physical reason for certain vibration sensor patterns observed in CMP and thereby facilitate real-time process prognosis.

In Chapter 3, we invoke a physical model-based justification for behaviors observed in MEMS vibration signal features identified using empirical tests (Appendix II). This closes the loop from both the process dynamics and signal behavior perspectives.

However, the challenge in attempting to capture sensor signal patterns from the process dynamics stems from various *multi-scale* phenomena active in the

process (as alluded in [27]). Apart from the material removal regime at the nanoscale, there is a significant influence of bulk-level dynamics (due to pad elastic behavior) and machine-level kinematics on the sensor signal patterns. The vibration patterns emerging from each of these different multi-scale aspects may be coupled (i.e., influence each other) leading to complex aggregated vibration signal patterns. Therefore, it is important to integrate the contribution of each multi-scale aspect in order to obtain a *close corroboration with observed sensor data*.

Challenges

Interaction between multi-scale phenomena in CMP manifest in complex vibration signal patterns. Integration of such multi-scale process aspects is necessary in order to identify the physical causes influencing vibration sensor signal patterns.

Research approach and outcomes

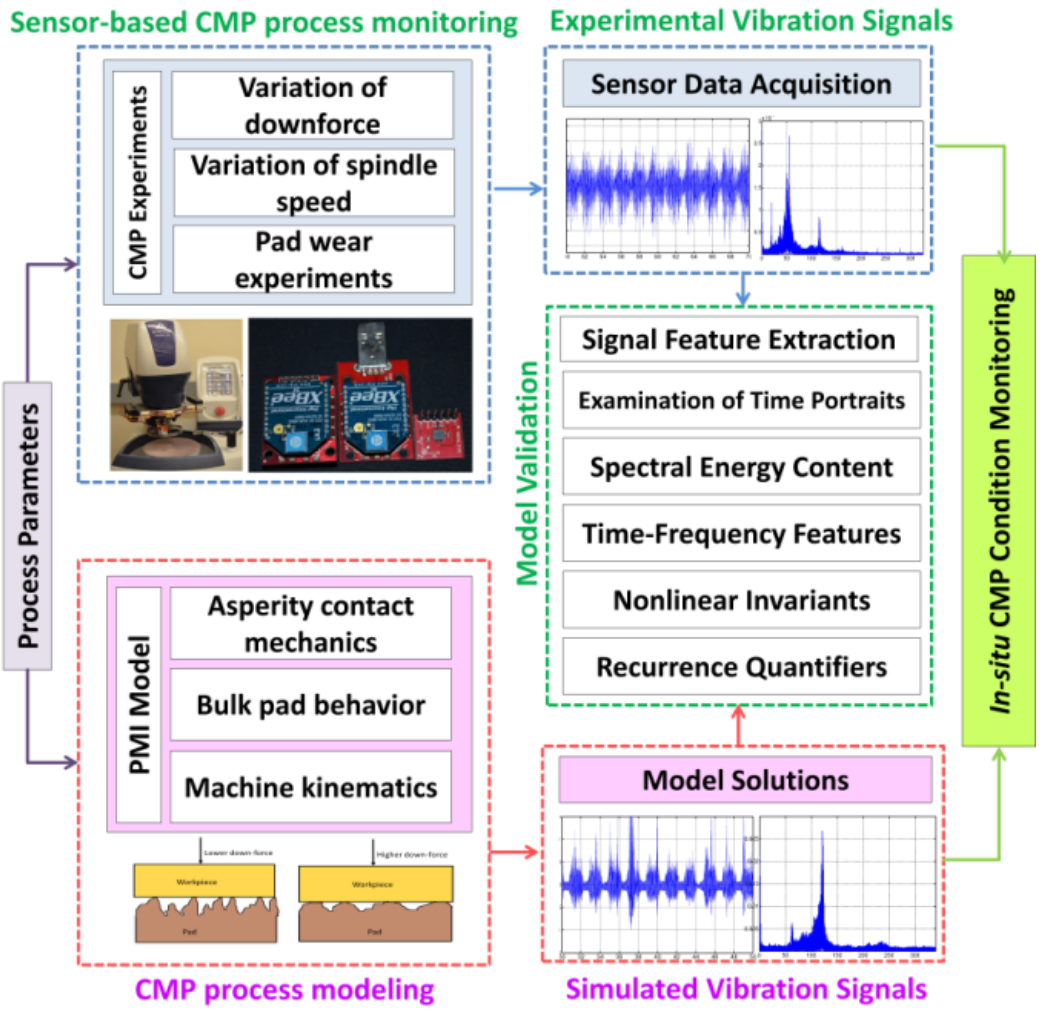


Figure 1-12: Schematic of the approach used for *in situ* condition monitoring in CMP.

Figure 1-12 shows the overall approach used in Chapter 3, where unlike the UPM case, the physical setup of the CMP machine does not allow for easy integration of wired piezoelectric sensors due to the presence of rotating components. Instead, we use wireless MEMS sensors for monitoring in CMP. Although, we have not detailed the sensor integration process, we note that off-the-shelf solu-

tions cannot be readily modified to suit the application. Therefore, we first designed a data acquisition system for CMP process monitoring, using IEEE 802.15.4 protocol radios for wireless MEMS vibration sensor signal transmission, and reception.

Nevertheless, as noted in the challenges, wireless sensing comes with the attendant problem of low sampling rate and therefore *lower signal to noise ratio* compared to piezoelectric sensors. Hence, in contrast to the UPM case, where we could proceed to analyze the signals directly once they are acquired, here we have to *judiciously ensure that signal features relevant to the process dynamics was captured*. Nine experimental tests are detailed in Appendix II. These tests led to identification of signal components that are relevant for process monitoring in CMP. We also statistically quantified the sensitivity of these signal features with respect to changing process conditions. Based on these studies we inferred (referring to Figure 1-9(b)):

1. Component 1 results from eccentricity errors in the polishing head (spindle).
2. Component 2 is most likely a consequence of sensor characteristics, electromagnetic interference from machine elements, and structural vibration.
3. Component 3 is responds to changing downforce (polishing load) conditions, and process state, such as pad wear, and is therefore useful for process monitoring applications.

4. Component 4 is largely due to environmental noise.

The *deterministic* process-machine interaction (PMI) model invoked for explaining the physical reasons for the patterns summarized above incorporates multi-scale aspects of the process. The PMI model represents the effects at the machine, bulk, and nanoscale levels on the overall CMP vibration patterns in the form of a 2 degree of freedom nonlinear differential equation. The detailed formulation of the PMI model is documented in Appendix III.

Using the PMI model and vibration sensor signals in conjunction, as shown in Figure 1-12, facilitates an explanation of the physical reason for observed signal patterns. This can be valuable under sparse data conditions, because, the PMI model can be used to predict anomalous process states, and thereby reduce wafer defects. For instance, we can simulate *offline* using the PMI model the vibration patterns that emerge when the pad is close to being glazed (see Figure 1-11). From the simulated vibration patterns, we can then anticipate *a priori* the features and characteristics to monitor in sensor signals and therefore predict the onset of process anomalies.

In other words, using the PMI model, we know *beforehand* those features in the actual signal that should be tracked closely to predict anomalous behavior. For instance, we used the PMI model generated signals to predict process anomalies due to pad wear with an accuracy of $R^2 \sim 75\%$. Specific tasks toward fulfilling the outcomes of this chapter are listed herewith.

Specific Tasks

1. Ensure surface finish in the nanometric range (Sa < 5 nm) with near-specular characteristics for CMP processed Cu wafers consistent with the state-of-art.
2. Instrument CMP polishing experimental setup with *in situ* MEMS wireless vibration sensors in close proximity of the wafer and thus acquire data remotely.
3. Isolate signal components evocative of process dynamics as opposed to extraneous noise by conducting a battery of empirical tests, such as machine modal analysis, sensor response characterization studies, structural noise isolation assessments, etc.
4. Conduct designed experiments to correlate relevant signal components with process conditions. For example, correlate magnitude change in the frequency spectra of CMP wireless vibration signals that respond to variations in polishing load by conducting designed experiments.
5. Invoke a multi-scale process-machine interaction (PMI) model evocative of multi-scale process mechanics in CMP, in order to explain the behavior of vibration signal patterns.
6. Corroborate PMI solutions with experimentally gathered data. For example, show that the model simulated vibration signal patterns and actual vibration patterns depict identical time-frequency behavior.
7. Identify features gathered from simulated vibration patterns and use them in conjunction with experimentally acquired signals for predicting the onset of process anomalies.

1.4.3 A Graph Theoretic Approach for Quantification of Ultraprecision Surface Morphology

In Chapter 4, we present an algebraic graph theoretic approach for characterization of CMP wafers processed to near-specular finish. This was motivated from the observed incongruity of statistical parameters in quantifying such ultraprecision surfaces. The approach is based on treating various heterogeneous features (e.g., pits, ridges, scratches) as a stationary mixture of random fields that allows for compact graph theoretic representation of the surface morphology as measured from optical micrograph images.

The graph theoretic approach, by primarily employing optical microscopy, eschews relatively reticent profile mapping techniques, such as laser interferometry and atomic force microscopy. Thus, we have addressed the expert concerns regarding characterization of nano-materials in this chapter.

Challenges in Nanomanufacturing Research [27]

Tools currently used in nanotechnology research labs offer atomic-level resolution for characterizing nanoscale surfaces but are barely adequate to meet the requirements of high-volume nanomanufacturing. For example, an Atomic Force Microscope (AFM) gives nearly atomic-level surface resolution, albeit at a very slow rate. It would be impossible to use AFM to characterize surfaces in commercial-scale high-rate operations.

Expert Recommendations for Nanomanufacturing research [27]

Instrumentation and **analytical tools for comprehensive characterization of nano-materials for on-line process control should be developed** and characteristics required for specific applications defined.

Motivation and goal

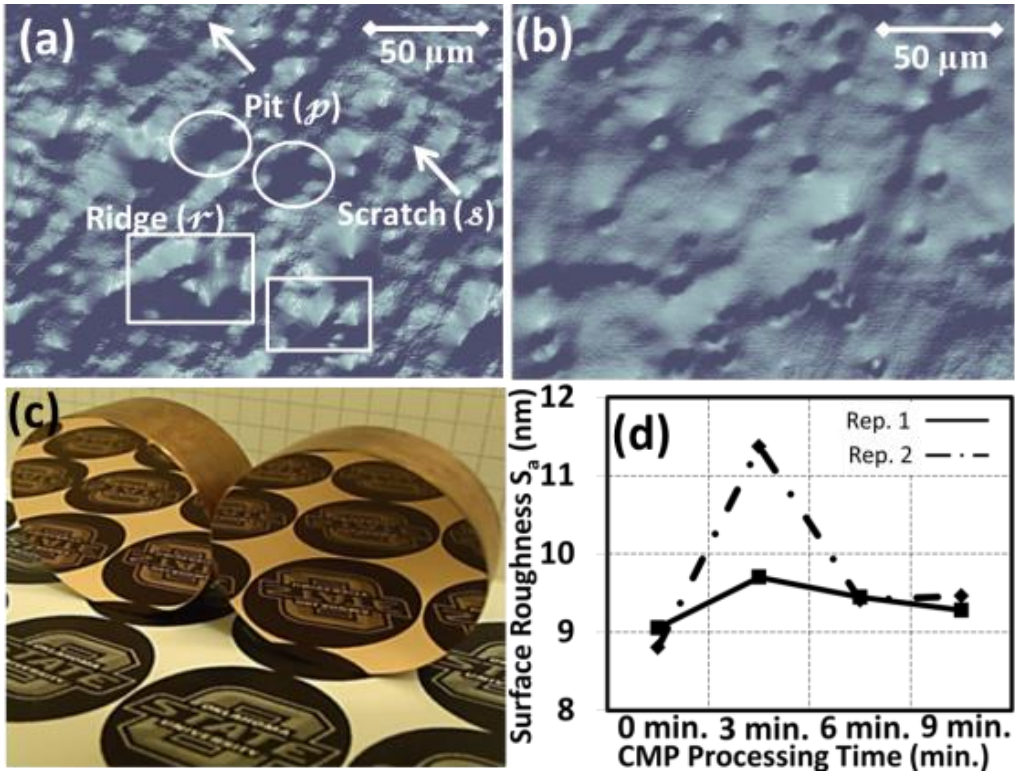


Figure 1-13: The surface profile of a copper wafer obtained using a laser interferometer.

(a): The surface profile prior to CMP operations shows heterogeneous features, such as scratches (*s*), ridges (*r*), pits (*p*), etc.
(b): The surface profile after CMP shows significant improvement in surface morphology within 3 min of CMP operations.
(c) Specular finish obtained on CMP processed blanket copper wafers.
(d): The mean areal surface roughness (Sa) obtained over 3 min intervals of CMP for two replications fails to reflect these improvements in surface morphology.

Evident in Figure 1-13(a) are heterogeneous surface features, such as ridges, scratches, and pits occurring over different length scales (multi-scale behavior) present on a copper (Cu) wafer surface. Such surface morphology is considered

deleterious to the performance of a semiconductor device and is therefore improved using the CMP process. However, we found that although the surface morphology improved within 3 min of CMP (Figure 1-13(b)) and specular finish ultimately achieved (Figure 1-13(c)), traditional statistical quantifiers, such as average roughness (Ra, Sa), root mean square roughness (Rq, Sq), etc., did not track these morphological variations (Figure 1-13(d)).

Surface morphology is a critical determinant of functional performance in ultraprecision components, such as semiconductor ICs, optical mirrors, MEMS devices, etc. [5, 6, 8, 9, 54]. However, ultraprecision surfaces possess multi-scaled heterogeneous features (such as those seen in Figure 1-13(a)), which make quantification of the surface morphology a compelling challenge. This is mainly because statistical parameters that are traditionally used for surface quantification can only capture a few of these morphological features. Therefore, in order to characterize the surface morphology, several parameters may be needed. This has led to a situation where several dozens of parameters have been standardized. Experts in the area call this tendency to continually define newer statistical parameters as the *parameter rash* [41].

Additionally, statistical parameters tell us very little about the relationship connecting various morphological aspects of the surface. To use a geodesic analogy, though we can describe the average height, and depth of summits and valleys for an area, we know nothing about how these features (valley and ridges) are connected. Although, fractal and wavelet analysis techniques have been proposed

for quantification of multi-scale morphological aspects, these methods require sifting through different length scales and are therefore computationally demanding [6, 55-57].

Furthermore, ultraprecision surface characterization approaches, such as atomic force microscopy (AFM), scanning electron microscopy (SEM), laser interferometry, stylus-based profilometry, etc., can be time consuming, restricted to assessment of small areas, and sometimes destructive [56]. Hence, there is a compelling need for more efficient means for surface characterization. For example, in the semiconductor industry offline characterization can take several hours [26]. This consequently stymies opportunities to correct process anomalies at an early stage. Also, offline characterization using test wafers can lead to significant yield losses estimated by some researchers to be close to 35% [25].

Consequently, an approach that can parsimoniously capture surface morphology variations using efficient means, such as optical micrographs can be valuable for *in situ* quality assessment of ultraprecision components. The goal and challenges with respect to quantification of nano-surface morphologies are stated below.

Goal 4
(Satisfying Objective 2)

Develop an approach that can quantify ultraprecision surface morphologies in a rapid and parsimonious manner.

Challenges

1. Ultraprecision surfaces depict heterogeneous multi-scale aspects, which are not captured using traditional statistical quantifiers.
2. Traditional statistical parameters are not evocative of the underlying morphological relationships.
3. Characterization approaches, such as AFM, SEM, and laser interferometry are reticent, restricted to small areas, destructive, and expensive.

Research approach and outcomes

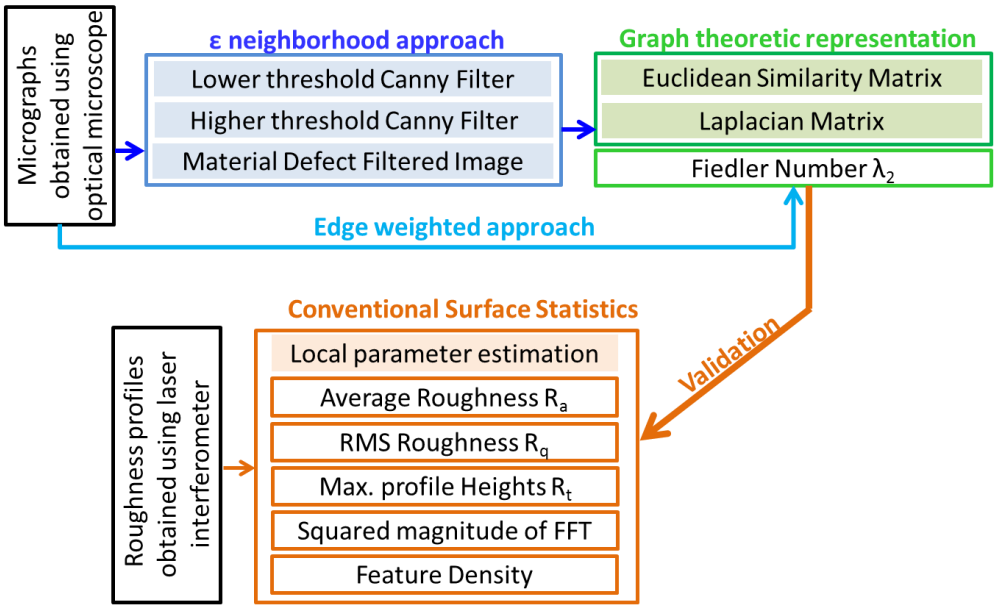


Figure 1-14: Schematic of graph theoretic approach used for quantification of ultraprecision surface morphology.

In order to surmount these challenges, we proceeded to represent the surface morphology as a network graph. This graph theoretic approach for surface morphology quantification implicitly represents ultraprecision surfaces as a convolution of random fields associated with heterogeneous multi-scaled features present on the surface. As evident from the overall schematic of the approach in Figure

1-14, two symbiotic methods for graph theoretic representation of ultraprecision surface morphology are developed. Consequently, graph theoretic topological invariants, such as Fiedler number of the network graph, are estimated to capture the evolving ultraprecision surface morphology (for example, the morphology illustrated in Figure 1-13).

We note that by using optical micrographs for quantification of the surface morphology, the graph theoretic approach precluded inefficient characterization methods such as laser interferometry. We verify that the Fielder number is able to capture the evolving morphology of an ultraprecision surface with high fidelity ($\rho > 80\%$). The specific tasks to be accomplished in Chapter 4 are listed below.

Specific Tasks

1. Represent ultraprecision surface morphology using a graph theoretic approach based on optical micrograph images.
2. Illustrate that the graph theoretic topological invariant, Fiedler number, is capable of capturing heterogeneous multi-scale surface morphology features, which are not adequately tracked using statistical quantifiers.
3. Verify the Fiedler number trends for CMP surfaces with scale-limited local surface roughness statistics.

A note for the reader

Given the diversity of processes and approaches studied, a literature review chapter is eschewed in this dissertation. Instead, we take a contextual approach by providing a literature review in each chapter. Likewise, since the approaches used are contingent on the process, these are detailed at appropriate junctures. In principle, each chapter is meant to be a standalone description. Indeed, the majority of the chapters stem from peer reviewed journal papers either; in press (Chapter 2)[58], editorial acceptance (Chapter 3) [59], or manuscript preparation stages (Chapter 4).

2 Real-time Identification of Incipient Surface Morphology Variations in Ultraprecision Machining (UPM) Processes

"Errors in DT can have significant impact on profits of Diamond Turning Business Unit. Attaining and maintaining satisfactory yields, on time delivery throughput and product quality is critical to II-VI profitability. Operational errors can jeopardize II-VI reputation in the market and can result in lost growth opportunities."

– An extract from a job posting for a Diamond Turning Process Engineering Manager position by II-VI Inc., of Saxonburg, PA.

Real-time monitoring and control of surface morphology variations in their incipient stages is vital for assuring nanometric range finish in the ultraprecision machining (UPM, often also termed as diamond turning (DT)) process. A real-time monitoring approach, based on predicting and updating the process states from sensor signals (using advanced neural networks and Bayesian analysis) is reported for detecting the incipient surface variations in UPM. An ultraprecision diamond turning machine is instrumented with three miniature accelerometers, a 3-axis piezoelectric dynamometer, and an acoustic emission (AE) sensor for process monitoring.

Goal

To suggest a method capable of real-time in situ detection of incipient anomalies in UPM process**.**

The diamond turning machine tool is used for face turning aluminum 6061 discs to a surface finish (Ra) in the range of 15-25 nm. While the sensor signals (especially the vibration signal in the feed direction) are sensitive to surface varia-

tions, the extraneous noise from the environment, machine elements, and sensing system prevents direct use of raw signal patterns for early detection of surface variations. In addition, the nonlinear and time-varying nature of the process dynamics does not lend conventional statistical process monitoring techniques suitable for characterizing UPM-machined surfaces. Consequently, instead of just monitoring the raw sensor signal patterns, the nonlinear process dynamics wherefrom the signal evolves are more effectively captured using a Recurrent Predictor Neural Network (RPNN).

The parameters of the RPNN (weights and biases) serve as the surrogates of the process states, which are updated in real–time, based on measured sensor signals using a Bayesian Particle Filter (PF) technique. We show that the PF-updated RPNN can effectively capture the complex signal evolution patterns. We use a mean-shift statistic, estimated from the PF-estimated surrogate states, to detect surface variation-induced changes in the process dynamics. Experimental investigations show that variations in surface characteristics can be detected within 15 ms of their inception using the present approach, as opposed to 30 ms or higher with the conventional statistical change detection methods tested.

The following challenges and expert recommendations for nanomanufacturing research as listed by Bukkapatnam, *et al.* [27] provides the vista for this work[5].

[5] P. Rao, S. Bukkapatnam, O. Beyca, Z. J. Kong, and R. Komanduri, "Real-time identification of incipient surface morphology variations in ultraprecision machining process," *Transactions of the ASME, Journal of Manufacturing Science and Engineering*, (In-Press) 2013.

Challenges in nanomanufacturing research [27]

Nanoscale processes and systems pose many challenges for sensing:

- Accessibility to signal source is not easy;
- *In situ* sensing is almost impossible;
- Signals are short, evanescent, and weak;
- Quantization of signals makes transduction difficult; and
- Signal-to-noise ratio is low.

Expert recommendations for nanomanufacturing research [27]

A **systematic set of methodologies of quality engineering should be developed** to provide the following:

- guidelines for the design and analysis of experiments to optimize nanoprocess settings;
- **on-line monitoring and diagnosis techniques to reduce nanoprocess variation and downtime during production**; and
- strategies for continuous improvement for high yield and quality.

2.1 Introduction

Aluminum and copper disks are widely used in the manufacture of mirror finish surfaces (arithmetic average surface roughness (Ra) in the nanometric range) for diverse engineering applications, including hard drives and memory discs in the computer industry, precision aluminum mirrors in lasers, rotating mirrors in copy machines, as well as optical elements critical to the aerospace and defense industries [1, 2].

Such optical surfaces are often realized using ultraprecision machining (UPM) process, where depth of cut is in 2-50 μm range and a single crystal dia-

mond tool is employed [1]. An UPM-machined surface typically does not need an additional finishing process, such as lapping or polishing [1, 60]. For most industrial UPM processes, assurance of consistent surface finish is crucial. **Adverse variations of surface characteristics can lead to heavy scrap rates and rework. Therefore, detection of surface variations in their incipient stages would be crucial for effective control, so that rework can be significantly reduced.**

Surface finish in UPM depends on the machine stiffness and precision [1], chip generation process [61], and other factors, such as machine tool vibration [62], and thermal instabilities [63] that affect dynamic stability. Unlike in conventional machining, the uncut chip thickness in UPM is comparable to tool edge radius, and of the same order of magnitude as the grain size of the workpiece material (see Graphic VI). These contrasts with conventional machining give rise to distinctly different chip-formation process, magnitude of cutting and thrust forces and their ratio, and surface generation mechanisms [1, 64]. These mechanisms can affect the surface properties [65] and functional behavior of a machined component [1]. These factors also make the prediction and control of surface quality and its variations an enormous challenge.

For instance, at the uncut chip thickness typical to UPM (viz. tens of nanometers), complex physical phenomena which are distinctive from conventional machining are evident. Below a critical uncut chip thickness, termed as minimum chip thickness [61, 66], material removal in the conventional sense (shearing of

material in form of chips) is not observed [64, 67]. Instead, material is dislocated as the tool *ploughs* through the surface [64, 67]. Consequently, the magnitude of forces and specific energy observed empirically in ultraprecision machining are not in accordance with conventional machining models [67].

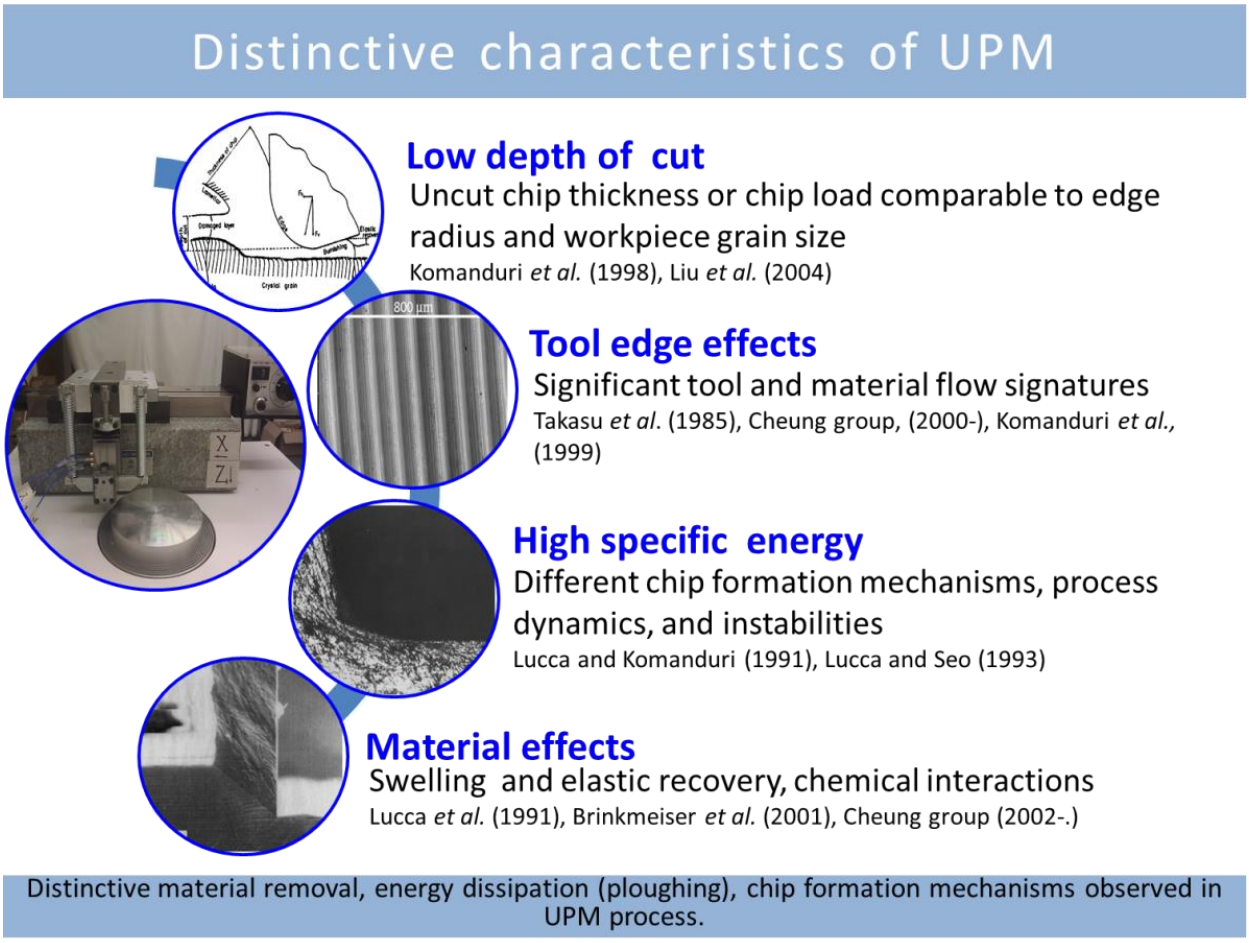
In this context, experts have noted that the tool edge radius is a critical factor governing the minimum chip thickness (and thereby the surface quality, material flow characteristics, and the nature (magnitude and direction) of resultant forces in UPM process) [61, 64, 66, 68]. Typically, the ratio of minimum chip thickness to tool edge radius is observed to be approximately 0.20 – 0.30 [69]. A (nanometric) change in the geometry of the tool edge, e.g., due to wear, which might be inconsequential in conventional machining, may completely alter the cutting dynamics in ultraprecision machining [68].

However, despite the significant effect of tool edge radius on UPM process mechanics, commercial diamond tool suppliers often restrict the explicit specification of the tool signature to the nose radius, clearance and rake angles. This is because the measurement of the edge radius is a challenging preposition. Researchers have therefore developed novel approaches using precision instruments, such as AFM and SEM for measurement of SCD tool edge radius.

For example, Asai, *et al.* [70] used a SEM setup integrated with two electron detectors for accurately estimating the tool edge geometry. Lucca, *et al.* [68] use an AFM-based setup (compensating for the geometry of the AFM probe) for measuring the tool edge radius. Evidently, the measurement of tool edge geome-

try is not a tractable/trivial exercise, and is considered to be outside the scope of this research.

For this work, we wish to clarify at the outset that we have not attempted to characterize the tool edge geometry and its effects on UPM mechanics. More importantly, we implicitly assume that the tool geometry, particularly the edge radius, remains unchanged during our experiments. This is a tenable assumption because the experiments documented in our study are used for cutting few feet (if not inches) of material under conservative processing/machining conditions.



Graphic VI: Distinctive characteristics of the UPM process.

Some research efforts have focused on the mechanisms of material removal in UPM as well as the use of sensors for monitoring and control of the process [50]. For example, Cheung, *et al*. [17, 71] developed a surface topography simulation model considering the effects of tool geometry, cutting parameters, and the vibration between the cutting tool and the workpiece in UPM operations. Abouelatta, *et al*. [72] modeled the effects of process parameters and vibration in both radial and feed direction of machined surface characteristics. Beggan, *et al*. [73] used acoustic emission (AE) sensor to predict surface roughness.

Hayashi, *et al.* [74] used a miniature tool temperature sensor composed of platinum electrodes coupled with an integrated PID type feedback controller that minimizes excessive temperature variations around the tool tip by adjusting the spindle speed. The approach was demonstrated in UPM of aluminum alloy discs as follows: for a fixed feed rate of 15 μm/rev, depth of cut 5 μm, and maximum allowable spindle speed of 5000 RPM, an arbitrary rise of 7.1 °C was selected as the permissible level of temperature change around the tool tip from steady state. When a temperature rise greater than 7.1 °C was detected, the spindle speed was gradually reduced to ~ 2000 RPM via a feedback controller. However, the authors caution that parameters of the feedback controller require extensive tuning in order to be effective for early detection.

Similarly, Yoshioka, *et al.* [75] attempted to minimize the thermal deformations in the aerostatic bearing spindle of an UPM machine by monitoring the temperature of air at the spindle exit, and adaptively adjusting the temperature of

the intake air. Shinno, *et al.* [76] reported the use of a strain gage sensor in a Wheatstone bridge formation for the detection of anomalies, such as chatter and winding of chips around tool shank in machining. Although, these prior works report the use of sensors to detect well-developed anomalies, approaches for early stage detection of nascent process drifts, critical for effective process control in UPM, have not been reported to date.

The real-time monitoring approach presented in this research is also sensor-based, and it is demonstrated on an ultraprecision diamond turning machine instrumented with three miniature accelerometers, a 3-axis piezoelectric dynamometer, and an acoustic emission (AE) sensor for process monitoring. The machine tool is used for face turning aluminum (Al) 6061 discs to a surface finish (Sa) in the range of 15-25 nm.

For example, Figure 2-1 shows a sample surface of an Al 6061 disk UPM-machined to an average Sa of 20 nm (see also, Figure 2-3)[6]. As a ~ 300 μm long contiguous surface profile obtained on the same sample in Figure 2-2 indicates, changes in surface characteristics can take place rather abruptly, especially near the edges – the Sa value changes from ~ 18 nm to 30 nm in a span of ~ 100 μm.

Prediction of the onset of such surface defects in their incipient stages can be valuable for assuring product quality and minimizing subsequent rework. The sensors can be used to discern the complex processes variations that cause and/or

[6] Representative 3D profiles and statistical metrology parameters for this UPM-processed surface are shown in Figure 2-3 and Table 2-1, respectively

result from such sudden surface finish variations, which are often difficult to predict using the current physical models [77].

Much of these complexities and sharp variations in the surface morphology can be attributed to nonlinear dynamics [78] and the inherent nonstationarity (i.e., time-varying nature) of the process.

Challenges

1. Traditional statistics-based quality control and spectral analysis methods cannot capture evanescent, nonlinear, and nonstationary signal patterns in a timely manner, and therefore do not lend toward incipient real-time detection of ultraprecision machining anomalies.
2. Information from multiple heterogeneous sensors is not easily integrated using traditional statistical methods.

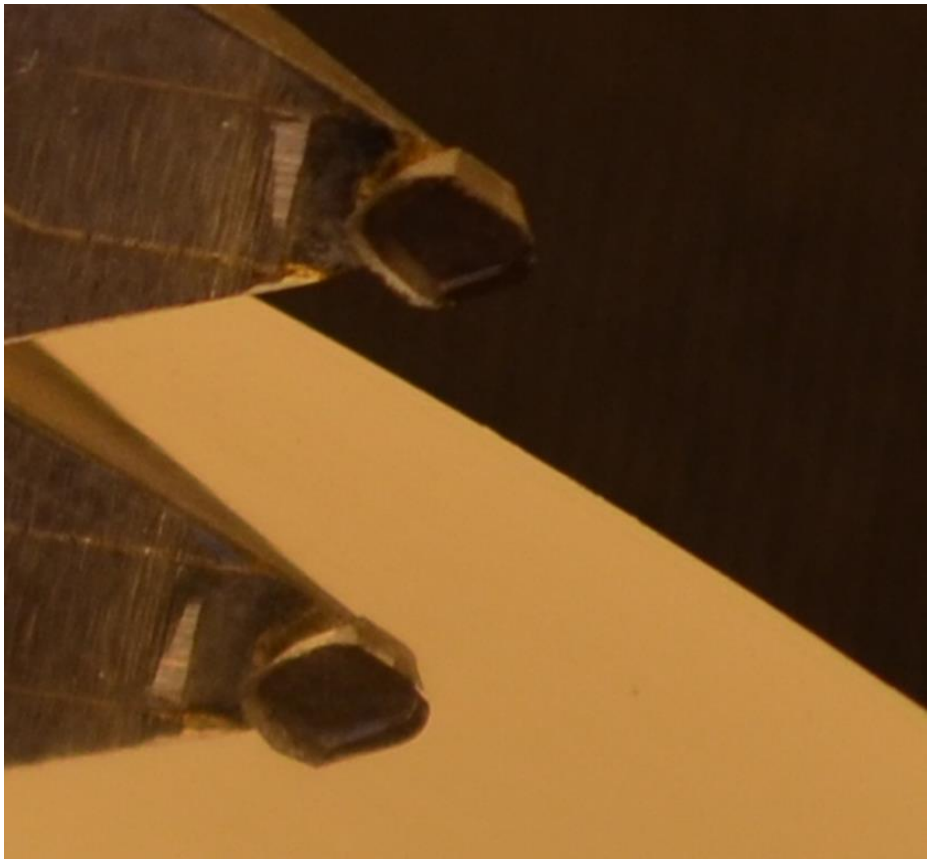


Figure 2-1: SCD tool used in UPM experiments and its image on a typical surface (Sa ~ 20nm).

Surface was obtained on an Al 6061 sample workpiece machined with the UPM setup. The sample was machined with the SCD tool under the following conditions: spindle speed 2000 RPM, feed rate 15 mm/min, and depth of cut 4 µm.

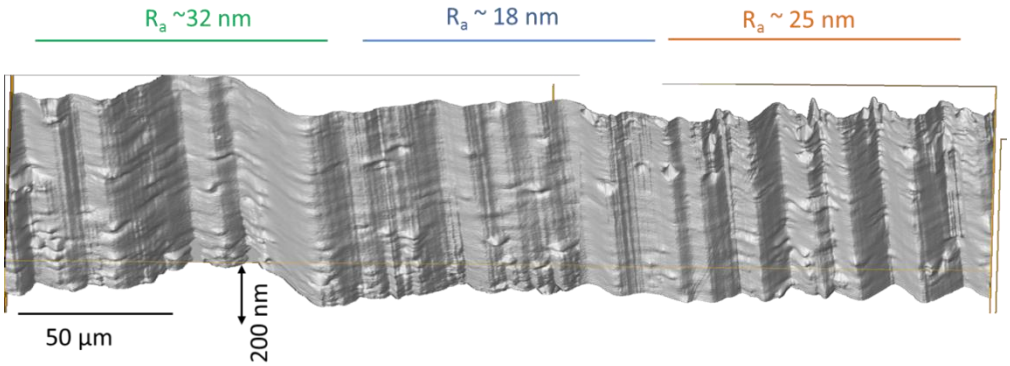


Figure 2-2: MicroXAM 3D profile of the workpiece shown in Figure 2-1.

Measurements were taken near the outer edge showing abrupt changes in the surface finish.

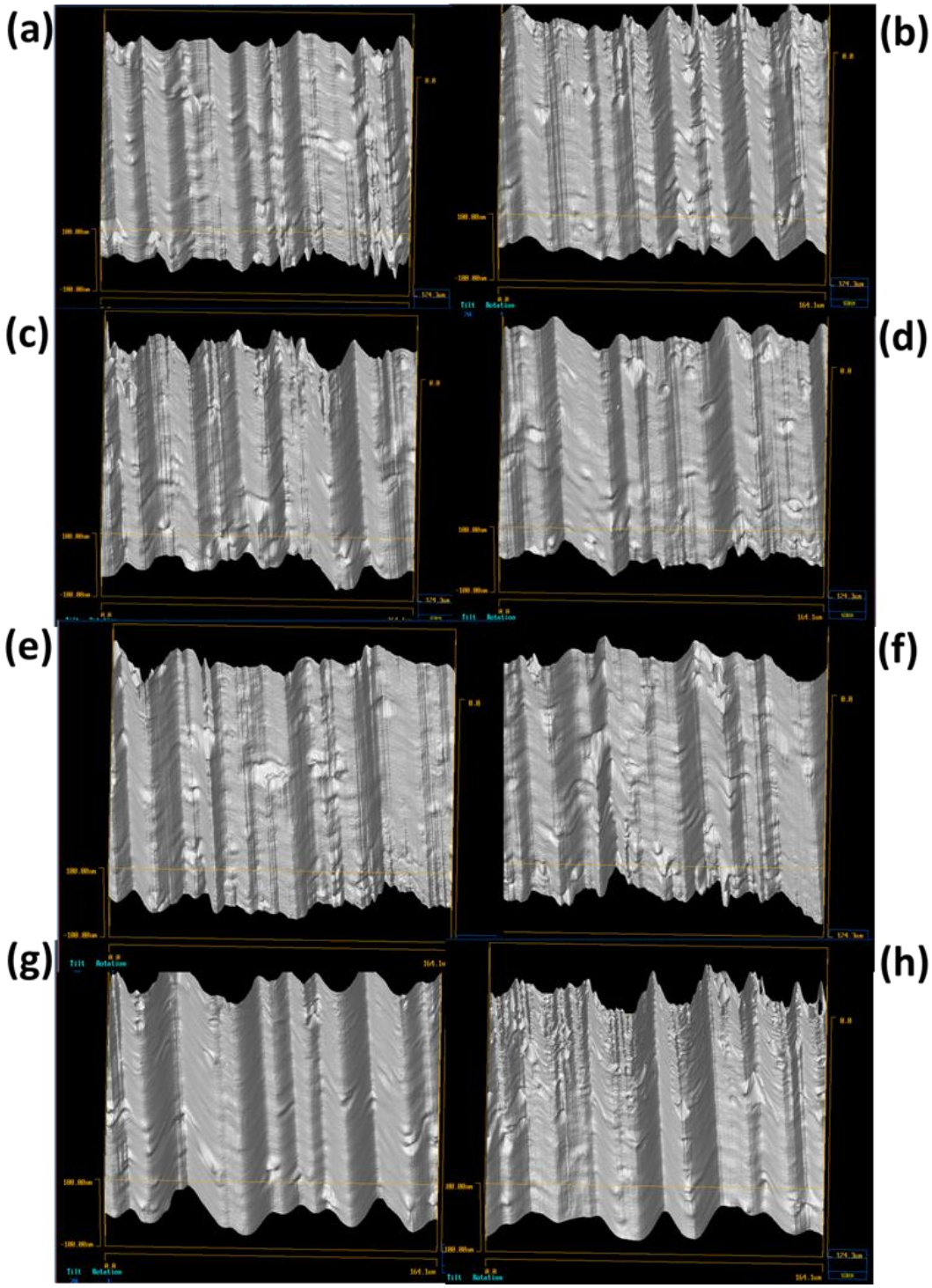


Figure 2-3: Representative 3D profiles obtained using the MicroXAM laser interference microscope for the surface shown in Figure 2-1.
The profile sections are 164 nm × 124 nm × 100 nm. The feed direction is along the long axis. A green light source (λ = 547.8 nm) is used; magnification is set at 50X (see also Appendix I).

Table 2-1: Some areal statistical characteristics corresponding to the surface profiles shown in Figure 2-3 [6, 56].

(all units are in nanometers, Ssk and Sku are dimensionless parameters.

PROFILE	ARITHMETIC AVERAGE SURFACE ROUGHNESS	ROOT-MEAN SQUARE SURFACE ROUGHNESS	MAXIMUM PROFILE PEAK HEIGHT	MAXIMUM PROFILE VALLEY DEPTH	PROFILE SKEWNESS	PROFILE KURTOSIS
	Sa	**Sq**	**Sp**	**Sv**	**Ssk**	**Sku**
(a)	12.9	16.1	52.1	59.4	0.18	-0.41
(b)	16.1	19.9	59.0	49.1	0.23	-0.66
(c)	23.4	28.8	68.9	86.6	-0.38	-0.36
(d)	18.7	22.8	63.7	61.2	0.21	-0.33
(e)	17.2	21.9	74.0	68.4	0.36	-0.31
(f)	22.9	30.4	94.3	85.4	-0.35	0.36
(g)	26.2	33.0	110.1	74.9	0.44	-0.15
(h)	27.9	34.4	93.6	200.1	-0.10	-0.19
MEAN	**20.6**	**25.9**	**76.9**	**85.6**	**0.07**	**-0.25**
STD. DEV	**5.2**	**6.6**	**20.2**	**48.0**	**0.31**	**0.29**
RANGE	**15**	**18.3**	**58**	**140.7**	**NA**	**NA**

In the present approach, the complex evolution patterns discernible from multiple sensors are compactly captured by a dynamic version of a Neural Network (NN) called a Recurrent Predictor Neural Network (RPNN) [79, 80]. The parameters of the RPNN (i.e., weights and biases) are used as the surrogate states of the process underlying the measured sensor signals. With every new signal measurement, the surrogate states (i.e., weights and biases) are updated using a Bayesian prediction technique known as Particle Filter (PF) method [81]. The use of RPNN

weights was found to accentuate surface variation-induced changes in signal characteristics.

The PF is able to capture the effects of complex nonlinear and non-stationary UPM process dynamics on the signal variations, and hence the weight patterns. The combined RPNN-PF is applied to predict surface variations in their early stages. The experimental investigations indicate that the present multi-sensor-based approach can effectively predict surface variation on the machined work-piece about 15 ms compared to 30 ms or higher with conventional statistical change monitoring techniques.

The remainder of this chapter is organized as follows: a brief review of the literature is presented in Sec. 2.2; the overall research approach in Sec. 2.3; the implementation details and results in Sec. 2.4 – 2.6.

2.2 Review of the Literature

Considerable research has been conducted in the past two decades to investigate the effects of various process parameters on forces and surface characteristics in UPM. For example, Ikawa, *et al.* [1] have provided a historic overview of UPM. They also delineated the critical issues encountered during processing, including the process stability under different machine settings and tool geometries. Dornfeld, *et al.* [60] reviewed recent developments in micromachining process physics, contemporary modeling efforts, metrology, and machine tool

errors. In this section, we review some of the relevant experimental and theoretical works focusing on modeling and prediction of surface characteristics in UPM.

Liu, *et al.* [64] noted that the tool might encounter different material phases, such as ferrite, pearlite in micromachining of ferrous materials that can significantly alter the cutting process mechanics and obtained surface finish. They also noted that for depth of cuts typical to UPM, the tool may not continuously encounter material, thus giving rise to surface inconsistency. Ikawa, *et al.* [61] showed that an uncut chip on the order of 1 nm can be obtained using a specially prepared diamond tool. The minimum obtainable uncut chip thickness is strongly influenced by the cutting tool geometry, and the machine tool system.

Moriwaki, *et al.* [82] conducted UPM tests on copper using a diamond tool and found that as the nominal depth of cut decreases from 1 μm to 2.5 nm, the specific cutting energy increases sharply from 2 GPa to 150 GPa, suggesting the limited utility of conventional orthogonal machining models to quantify the UPM machined surfaces. Lucca, *et al.* [67] suggested that the combined effects of sliding at the tool-workpiece interface and plowing (due to the effective rake angle being negative) may play a significant role in explaining the large rise in specific cutting energy observed by Moriwaki, *et al.* [82].

Takasu, *et al*. [62] proposed a physical model to predict the surface roughness of workpieces obtained from UPM that considers interactions between vibration characteristics (amplitude and phase) at the tool-workpiece interface, geometry of the tool cutting edge (nose radius), and machining parameters (feed and

depth of cut). They delineated conditions where tool interference is said to occur, i.e., when the cross section of the cut made by the cutting edge of the tool at the current pass overlaps with the preceding pass. Tool interference is reported to have a deleterious effect on the surface finish. Cheung, *et al.* [17, 71] considered tool interference as part of a computational model to predict the topography of surfaces from UPM processes.

Lee, *et al*. [77] used the Taylor's theory of plasticity (see ref. [83]) to predict the variations in the cutting force, and their influence on surface characteristics. They conducted turning experiments with a single crystal diamond tool (774 µm nose radius), on both copper single crystal and alloy workpieces at depth of cut in the range of 5 to 50 µm, 3000 RPM speed, and 20 mm/min feed rate. The cutting and thrust forces were measured during machining using a piezoelectric dynamometer (9252A from Kistler). They observed that compared to conventional machining, the crystallographic orientation of the workpiece material greatly affects the cutting forces, chip generation process, and surface finish at low depths of cut typical to UPM.

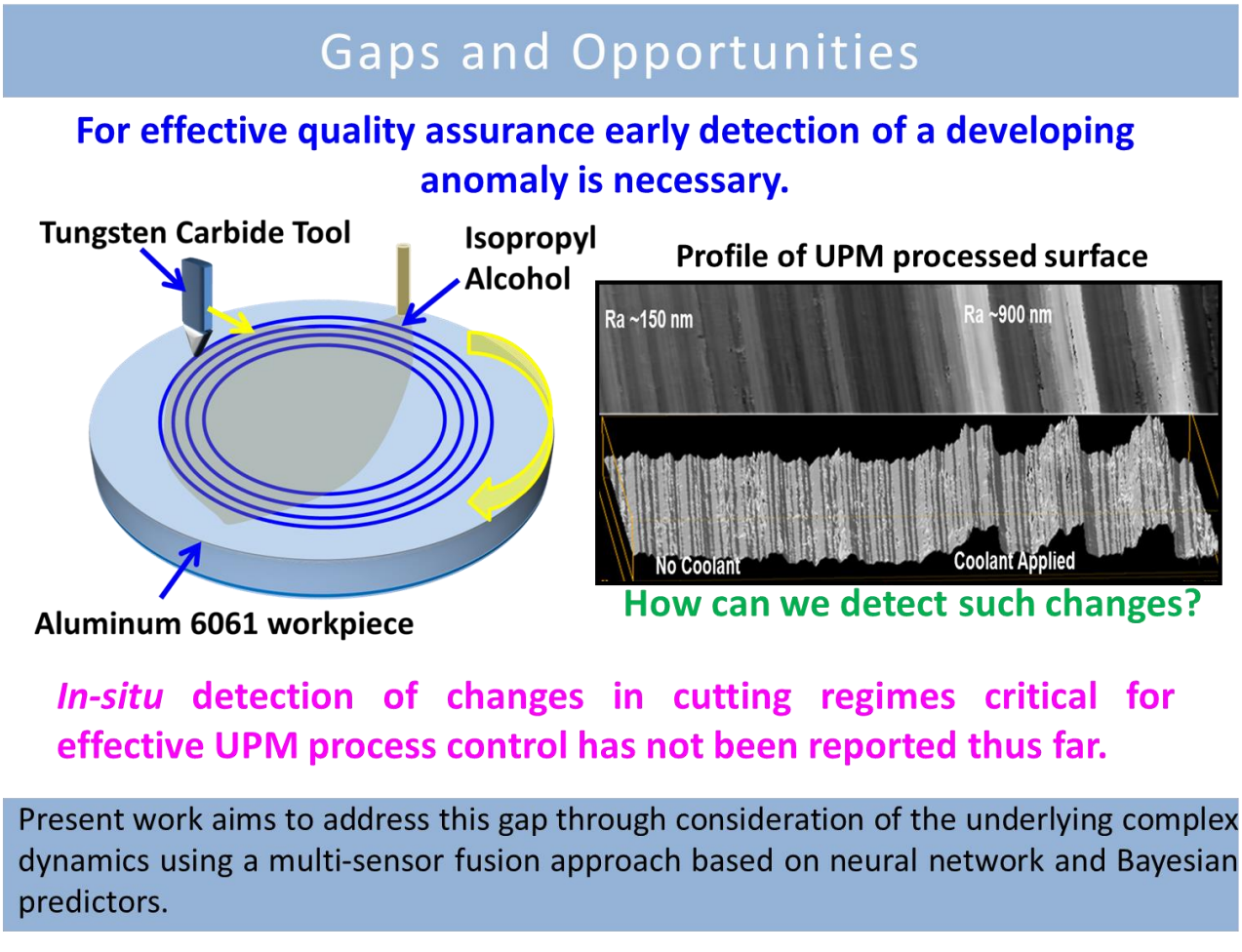
Kong, *et al.* [84] presented a phenomenological model to explain the characteristics of surfaces obtained from UPM of ductile materials, such as electroless nickel-phosphorus and aluminum based on elastic recovery and swelling effects of the material. They noted that when the depth of cut is small (~ 2 µm) and tool clearance angle is shallow (~ 7°), the elastic recovery, and subsequent material swelling has a significant impact on the surface characteristics.

Lee, *et al.*[53] used AE signals to investigate material anisotropy ahead of the tool (single crystal diamond) in UPM of single crystal as well of polycrystalline copper. The presence of material irregularities is determined as a significant cause of deterioration in the machined surface.

Cheung, *et al.'s* [85] framework for virtual machining and inspection of optical surfaces from UPM process consisted of a simulation module, which predicts the surface roughness based on estimated errors of tool location, wear, and spindle run out. Chan, *et al.* [17, 86] presented a finite element method (FEM) model considering process parameters, tool signature, and workpiece material aspects (swelling and elastic recovery) to predict the surface topography of UPM-machined Al-SiC metal-matrix composites.

Also, Hocheng, *et al.* [87] described a method to predict Ra in phosphor-bronze used for generating lens molds. They determined the surface profile using a stylus based instrument (TalySurf®), to investigate the frequency spectrum of the surface profile. Their analysis showed three major groupings for surface profile components on UPM-machined samples; (i) low frequency vibrations from the machine tool, (ii) from the tool geometry, and (iii) noise effects from the measuring instruments. They conducted a statistical analysis of variance (ANOVA) relating the speed, feed, depth of cut, and tool nose radius to the spectral intensity of the machined surface features. Their model can be used in a regression setting for the prediction of surface roughness.

These methods are able to detect prominent anomalies. However, early *in situ* stage detection, highly desirable for effective process control in UPM has not been reported thus far. The present work aims to address this gap (summarized also in Graphic VII) through consideration of the underlying complex dynamics using an NN and Bayesian update method.



Graphic VII: Gaps and opportunities in UPM research.

2.3 Research Approach

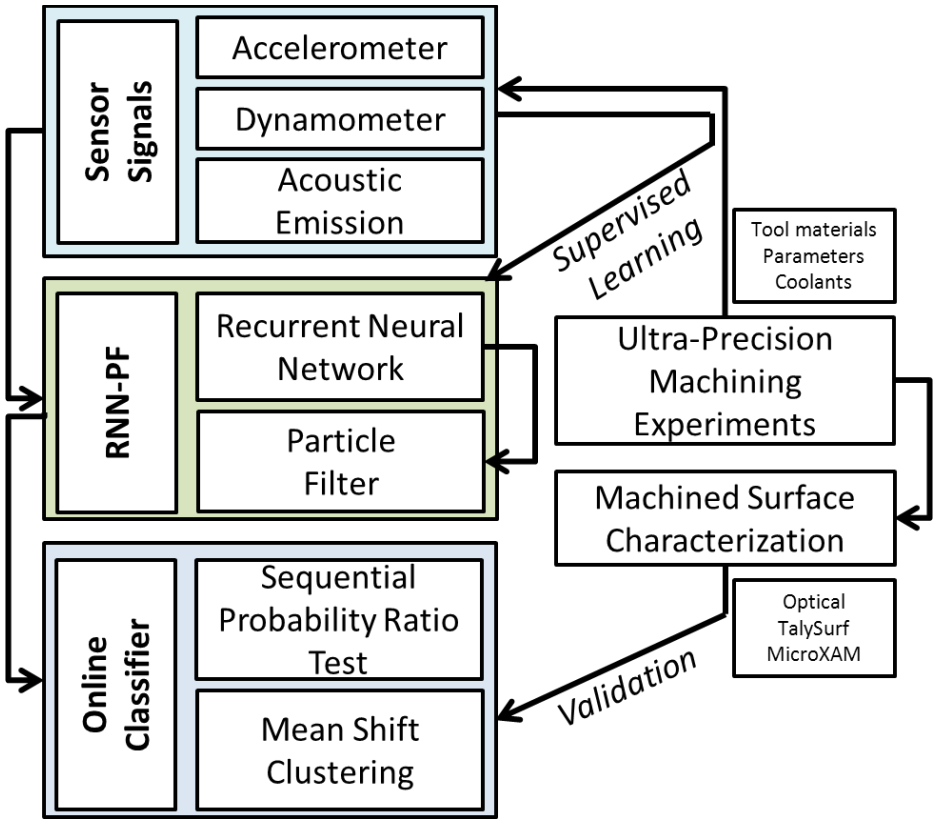


Figure 2-4: Summary of the proposed approach.

As summarized in Figure 2-4, we used information from multiple sensor signals for early stage detection of surface variations. The four main aspects of our approach are as follows:

2.3.1 Sensor Signal Acquisition

We integrated an UPM setup with multiple sensors, including, force, vibration, and AE as detailed in Sec. 2.4 (see also Table 2-2). This experimental UPM setup (Figure 2-5) is capable of face turning flat disks to nanometric finish in the range of 5 to 100 nm. The machine is equipped with an aerostatic spindle bearing

(model Block-Head® 4R) and air-slide tool carriage manufactured by Professional Instruments Inc. The setup rests on a 2 metric ton granite base for additional stability.

2.3.2 Machined Surface Characterization

The finished surfaces are first characterized using an optical microscope. Subsequently, quantitative characteristics of the surface waveforms are extracted using a high-resolution optical interference microscope (MicroXAM®, see Appendix I for a brief explanation of the operating principle of the instrument), and a stylus-based profilometer (TalySurf®). The analytical methods described in the following (RPNN-PF) are developed and validated for relating the signal features with certain incipient changes in the surface characteristics.

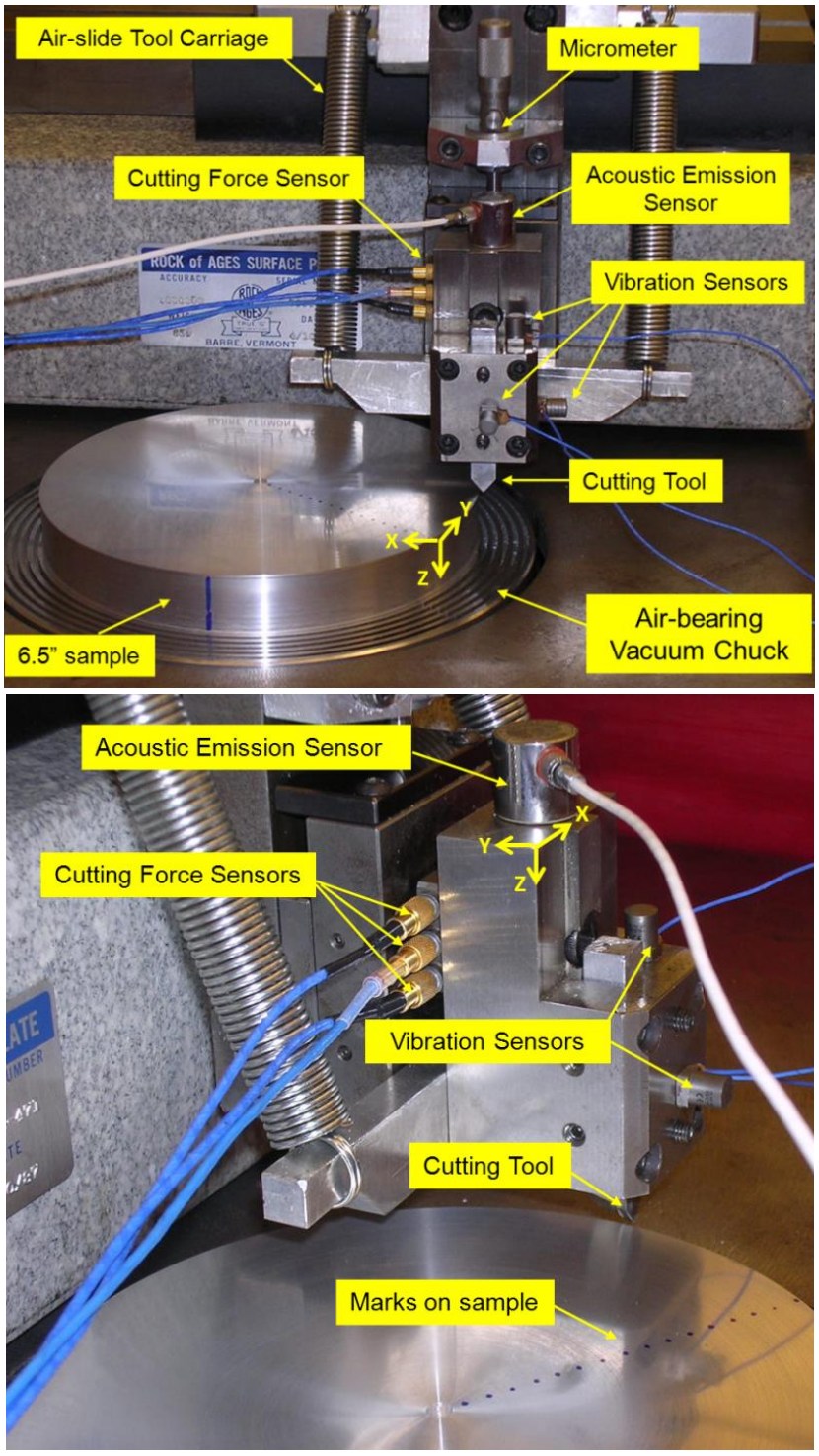


Figure 2-5: Two views of the Ultraprecision Machining (UPM) experimental setup. Setup shows the force (3 axis), vibration (3 directions) and AE sensors.

2.3.3 RPNN-PF

The measured multi-channel vibration and force sensor signals, although sensitive to variations in surface characteristics, contain significant amount of noise. Under such conditions, the signals can at best be used to only detect well-developed surface variations as opposed to predict or detect incipient surface variations. For incipient detection, it is important to separate useful information from the signals, as well as, impart some predictive capability, so that an appropriate mitigating control action can be executed.

Since the signals are nonlinear [88] and time varying [77], the use of traditional prediction approaches, including classical time series [89] and Kalman Filtering [81] methods, are unlikely to be adequate. Furthermore, since the raw signals contain significant extraneous information, it is desirable to transform these signals so that the relevant components of the signals are accentuated. We therefore, take a new approach that uses a Recurrent Predictor Neural Network (RPNN) [79, 80] to capture the nonlinear dynamic evolutions of signal patterns (see Figure 2-6). In other words, the parameters (i.e., weights and biases) of the RPNN serve as the surrogate states of the process dynamics.

While the RPNN can capture effectively the signal evolutions over a short term, its long-term predictability is limited by the non-stationary nature of the process (the process conditions are constantly changing), and sustained retraining of RPNN involves significant computational overhead.

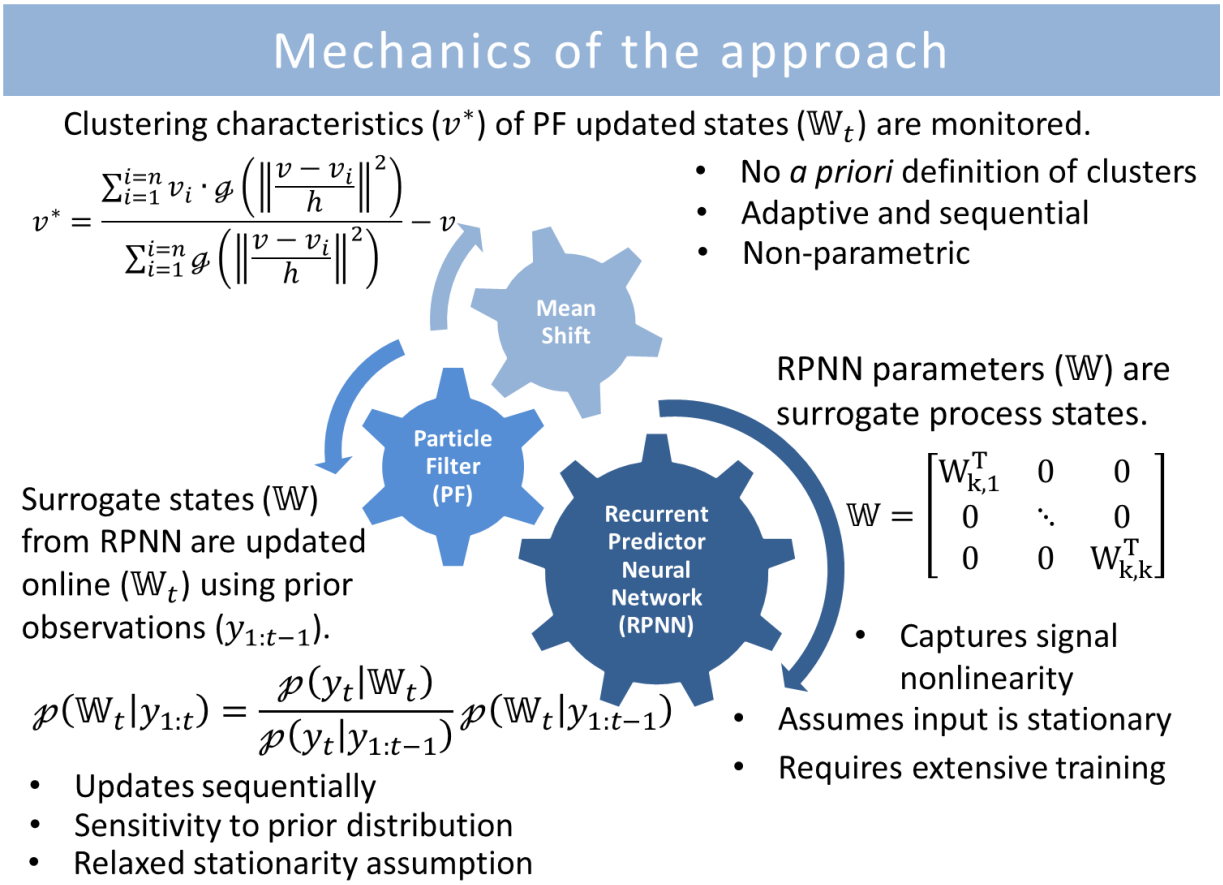
Therefore, an on-line method to adjust the weights and biases of the RPNN based on real-time sensor measurements is necessary to sustain prediction accuracies. In this context, the weights and biases can serve as the surrogates of the UPM process states that can be adjusted on-line using sensor measurements. Thus, the use of RPNN ensures that the predictions made are based on inherent variations in the process dynamics underlying the measured signals as opposed to mere statistical (and possibly noise-induced) variations in signal patterns. Sequential Bayesian techniques, including Particle Filters (PFs) are being increasingly sought for on-line adjustments in such complex prediction applications [81].

The effectiveness of these methods depends on the prior distributions of the variables being estimated. The distribution of the signal as a function of the surrogate state (RPNN parameters) can be used as the prior for a PF. Therefore, as shown in Figure 2-4, PF updates the weights and biases of a trained RPNN (i.e., the surrogate states), in real-time based on the signals measured on-line.

2.3.4 On-line Classifier

Oftentimes, surface-induced variations in the signals cause significant changes in the distribution of updated RPNN parameters, which have been used as the surrogates of the process states in this work. This change in the RPNN parameter distribution can be detected in their early stages using an on-line mean shift method. Compared to traditional change detection methods, this method does not assume stationarity of states, which makes it more suitable for the present applica-

tion. The relationship between RPNN, PF, and mean-shift stages is schematically depicted in Graphic VIII.



Graphic VIII: RPNN-PF approach for prediction of surface morphology variations in UPM.

2.4 Experimental Procedure

The present study involved face turning of Φ 16.25 cm × 3.75 cm cylindrical aluminum alloy (Al 6061) disk-shaped workpieces on the UPM setup (Figure 2-5). The following three cutting tools were used: natural single crystal diamond (SCD) tool with 750 µm nose radius, polycrystalline diamond (PCD) cutting tool with 400 µm nose radius, and brazed tungsten carbide chevron type cutting tool

with 396 μm nose radius. All the above tools have 0° rake angle and ~10° cylindrical clearance angle.

As shown in Figure 2-5, three vibration sensors (model 8728A500 from Kistler) were mounted along three orthogonal directions on the tool holder near the PCD cutting insert to measure vibration signals, and a 3-axis piezoelectric dynamometer (Kistler 9251A) was mounted on the underside of tool holder to measure force signals. Additionally, an AE sensor from Physical Acoustics (model R80) was mounted on top of the tool holder.

For experiments with the PCD tool, 36 different process parameters altogether were obtained from varying the spindle speed (500, 1000, 2000 RPM), feed rate (1.5, 3, 6 mm/min), and depth of cut (5, 10, 20, 25 μm), so that we can quantify the effects of factors and interactions among the process parameters and signal features.

With the carbide tool, experiments were performed at 2000 RPM spindle speed, 8 μm depth of cut, and six different feed rates (6, 15, 30, 60, 150, 300 mm/min). We have also investigated the effects of coolant (KoolMist®) on the surface and the corresponding signal characteristics. Sampling rates of 10 kHz were chosen for vibration and force signals and 1 MHz for the acoustic emission (AE) signals (see Table 2-2).

To provide consistent initial surface characteristics, all workpieces were initially finish-turned on a conventional lathe prior to the commencement of an UPM

test. They were then machined on the UPM setup at 1500 RPM, 60 mm/min feed rate (40 μm/rev feed), and 12 μm depth of cut in two successive steps.

First, we used a carbide tool to achieve surface flatness in 300 nm range. Next, we used PCD tool to obtain overall surface finish Sa ~ 100 nm. Thereafter, the SCD tool was mounted for finish UPM experiments. The surface shown in Figure 2-1 (Sa ~ 22 nm) was obtained with SCD tool in the UPM process conducted under spindle speed of 2000 RPM, feed rate 15 mm/min and depth of cut 4 μm. The sample had the following surface characteristics (12 point measurement): Sa ~ 20 nm (5 nm standard deviation) and Sq ~ 27 nm (6 nm standard deviation).

The results are comparable to those reported in the prior literature for Al 6061. For example, Dahlgren and Gerchman [16] have reported Ra of 7 – 10 nm in UPM of aluminum alloy series 201, 6061, 771 and 713 by tightly controlling the chemical integrity of the raw material during alloying and melt casting stages. Other researchers [84, 86, 90-92] have reported Ra in 12 – 25 nm range with Al 6061 workpieces.

The experiments exemplify some of the challenges with the UPM process for Al 6061, which are mainly due to presence of hard inclusions (from alloying elements such as Mg and Si),and trace iron impurities [15, 62, 90, 93]. These have also been noted to limit the life of SCD tools in UPM process.

Much of the present experimental studies on surface variation sensing are focused on machining with carbide and PCD tools. The SCD tool was used sparingly in this study, since much of the experiments were conducted under suboptimal

process conditions, such as sudden introduction of coolants, so that surface anomalies can be introduced and studied (see Sec. 2.6 for details). These conditions as well as the reasons noted in the foregoing significantly impede the life of SCD tools in UPM of Al 6061.

At every process parameter setting, the UPM facing operation was performed on a separate workpiece with the tool moving from the periphery towards center. It may be noted that the cutting speed continuously decreases during the process with the maximum speed near the periphery and almost zero speed near the center.

Therefore, the workpiece surface is divided into 16 concentric strips (henceforth referred to as zones) each measuring ~ 0.5 cm in width. Each zone is marked by almost constant process settings. Signals collected at each zone are synchronized in time with the corresponding machining location on the workpiece surface that forms a source of signal generation. On completion of machining, the workpiece surface is cleaned with isopropyl alcohol.

Subsequently, the surface characteristics, including the complete profile, surface roughness (Sa) of the workpiece were measured at three random locations on each of the 16 zones using multiple instruments, namely, optical laser interference microscope (MicroXAM® from KLA-Tencor), optical microscope, and a stylus-based profilometer (TalySurf®) manufactured by Rank-Taylor-Hobson. Among these, we mostly report the profiles measured from MicroXAM®. The surface

roughness values were determined by the Talysurf® and corroborated with MicroXAM® readings.

Table 2-2: Details of the sensing system mounted on the UPM setup

Signal Type	Sensor Orientation	Symbol	Sampling Rate	Sensor Type
Vibration	Along the feed direction	V_X	10 KHz	Kistler 8728A500
	Across the feed in the XY plane	V_Y		
	Vertical	V_Z		
Force	Along the feed direction	F_X	10 KHz	Kistler 9251A
	Across the feed in the XY plane	F_Y		
	Vertical	F_Z		
AE	Vertical	AE	1 MHz	Physical Acoustics R80

2.4.1 Recurrent Predictor Neural Network to Capture Signal Evolution in UPM process

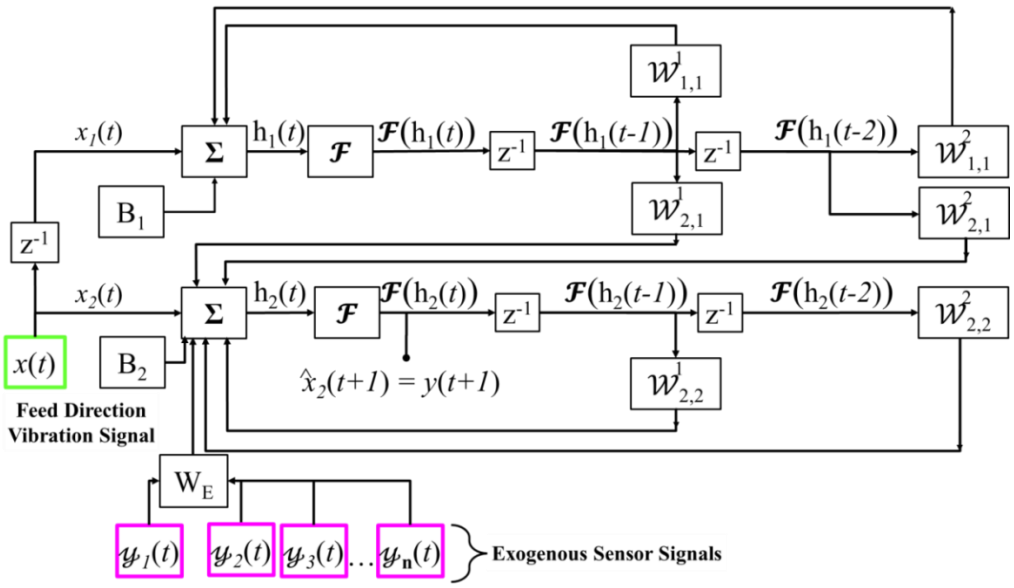


Figure 2-6: The architecture of an RPNN with input node I = 2, hidden node P = 2, and n exogenous inputs.

The complex evolutional patterns of the sensor signals from UPM process cannot be adequately captured using conventional prediction approaches. Recurrent Predictor Neural Networks (RPNN) [79, 80] offer a means to compactly capture the complex evolution of process underlying the measured vibration signals.

Unlike conventional NNs (also referred to as feed-forward NNs), RPNN captures the temporal relationship between successive realizations of the output variable (e.g., vibration sensor signals in our case). In other words, in conventional NN, the output *y* is a function of certain input variable *x,* and does not depend on the previous realizations of *y.* In contrast, an RPNN captures the signal evolution as a nonlinear difference equation.

Formally, this difference equation uses the **I** previous realizations of *y* (called nodes) or inputs, and **P** delays (hidden nodes) of neural network activations with each of the **I** nodes, i.e., for an RPNN with **I** nodes and **P** activation delays (also see Figure 2-6). We represent the signal in terms of difference equations involving the RPNN weights and biases as follows[7]:

$$\begin{gathered} x_{\mathbf{I}}(t) = y(t) \\ x_{\mathbf{I}-1}(t) = \mathbf{z}^{-1}x(t) = x(t-1) \\ x_{\mathbf{I}-2}(t) = \mathbf{z}^{-2}x(t) = x(t-2) \\ \vdots \\ x_{\mathbf{I}-\mathrm{k}}(t) = \mathbf{z}^{-\mathrm{k}}x(t) = x(t-k) \\ \vdots \\ x_1(t) = x_{\mathbf{I}-(\mathbf{I}-1)}(t) = \mathbf{z}^{-(\mathbf{I}-1)}x(t) = x(t-\mathbf{I}+1) \end{gathered} \tag{2-1}$$

[7] We also include **n** *exogenous* inputs ($\mathcal{y}$) independent of the primary input *x*. For example, if *x* is chosen to represent the feed direction vibration sensor (V_X) data and its time lags, then $\mathcal{y}$ could represent a combination of data from other sensors such as vertical direction vibration (V_Y), feed direction force (F_X), AE, etc.

The output at node *k* is given by,

$$\hat{x}_k(t+1) = \mathcal{F}\left(h_k(t)\right) \tag{2-2}$$

Since the one-step prediction is obtained at node **I** (the last node),

$$\hat{y}(t+1) = \hat{x}_I(t+1) = \mathcal{F}\left(h_I(t)\right) \tag{2-3}$$

*w*here $h_k(t)$, *k=1*... **I** is written as,

$$\begin{aligned} h_k(t) = x_k(t) + B_k \\ + \mathrm{tr}\left(\begin{bmatrix} W_{k,1}^T & \cdots 0 \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots 0 \cdots & W_{k,k}^T \end{bmatrix} \begin{bmatrix} H_1(t) & \cdots 0 \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots 0 \cdots & H_k(t) \end{bmatrix}\right) \\ + W_E^T \times \mathcal{y}(t) \end{aligned} \tag{2-4}$$

Each element $W_{i,k}$ in Eqn. (2-4) has the following form,

$$W_{i,k} \stackrel{\text{def}}{=} \begin{bmatrix} \mathcal{W}_{i,k}^1 & \mathcal{W}_{i,k}^2 & \cdots & \mathcal{W}_{i,k}^P \end{bmatrix}^T \tag{2-5}$$

Each $\mathcal{W}$ is one of the **P** x **I** (**I** +1)/2 weights in the RPNN. For example, $\mathcal{W}_{2,1}^3$ is a weight connecting from the third delay in the first node to the second node. Similarly, each element $H_k(t)$ in Eqn. (2-4) has the following form,

$$H_k(t) \stackrel{\text{def}}{=} \begin{bmatrix} \mathcal{F}\left(h_k(t-1)\right) & \mathcal{F}\left(h_k(t-2)\right) & \cdots & \mathcal{F}\left(h_k(t-P)\right) \end{bmatrix}^T \tag{2-6}$$

We define the exogenous components in Eqn. (2-4) as having the following form,

$$\begin{aligned} W_E &\stackrel{\text{def}}{=} \begin{bmatrix} W_E^1 & W_E^2 & \cdots & W_E^n \end{bmatrix}^T \\ \text{and, } \mathcal{y}(t) &\stackrel{\text{def}}{=} \begin{bmatrix} \mathcal{y}_1(t) & \mathcal{y}_2(t) & \cdots & \mathcal{y}_n(t) \end{bmatrix}^T \end{aligned} \tag{2-7}$$

Where elements in W_E are weights connecting to the exogenous inputs $\mathcal{y}(t)$, there being **n** such weights. For example, W_E^1 connects to $\mathcal{y}_1(t)$, W_E^2 to $\mathcal{y}_2(t)$, etc. Thus, the total number of weights ($\mathcal{W}$ and W_E) in the RPNN is given by **P** x **I** (**I** +1)/2 + **n**, and the number of biases (denoted as B) is **I**. The RPNN represents the process dynamics as a 4-dimensional nonlinear difference equation (i.e., **I** = 4). We use **P** = 4 for the signals obtained from the process. Such an RPNN has

been applied for the prediction of complex phenomena in diverse scientific and engineering domains [79, 80].

The key difference between conventional multilayer NNs and RPNN is that unlike conventional NNs where the training set elements are presented independently of each other without regard for their sequence or order, the training set elements in RPNN are treated as time-shifted versions of each other.

In addition, as shown in Figure 2-6, the layers of the networks are uniquely interconnected to each other. Such rich interconnectivity between layers requires a different approach towards calculation of network derivatives for minimizing the error between the RPNN outputs and the corresponding target values. Therefore, in contrast to the Back Propagation (BP) method [94] common to conventional NN, the RPNN uses the Back Propagation Through Time algorithm (BPTT) [95].

2.4.2 Online Particle Filter (PF) Update of the Process State

The evolution of surrogate process state $\mathbf{w}_k$, as captured from the p autoregressive terms of every RPNN parameter (weights and biases)[8] at time k is estimated using online vibration sensor data $\boldsymbol{y_k}$. The underlying relationship between the dynamic state $\mathbf{w}_k$ and measurement $\boldsymbol{y_k}$ can be represented in the state-observer form,

$$\mathbf{w}_k = f_k(\mathbf{w}_{k-1}, \omega_{k-1}) \tag{2-8}$$

[8] Here, for each RPNN parameter w, we created a separate state vector consisting of the current and its order-p autoregressive terms, i.e., for each RPNN parameter, $\mathbf{w}_k = [w_k\ w_{k-1} \ldots\ w_{k-p+1}]^T$, where the superscript $[\ ]^T$ denotes the transpose operator.

$$\boldsymbol{y}_k = \mathrm{h}_k(\mathbf{w}_{k-1}, \mathbf{v}_k)$$

where f_k is the state transition function, ω_{k-1} is white process noise, h_k is the observation function, and $\mathbf{v}_k$ is the white observation noise. The Bayesian approach is used to dynamically estimate and predict the state evolution, from a time-series of observations (vibration sensor signals in the feed direction), $y_1, y_2 \cdots y_k$, in terms of the conditional probability density functions (pdfs) $p(\mathbf{w}_k|y_{1:k})$, and $p(\mathbf{w}_{k+1}|y_{1:k})$. Based on the Bayesian theory, these conditional pdfs can be determined as,

$$\begin{aligned} p(\mathbf{w}_k|y_{1:k}) &= \frac{p(y_k|\mathbf{w}_k)}{p(y_k|y_{1:k-1})} p(\mathbf{w}_k|y_{1:k-1}) \\ p(\mathbf{w}_{k+1}|y_{1:k}) &= \int p(\mathbf{w}_{k+1}|\mathbf{w}_k)\, p(\mathbf{w}_k|y_{1:k}) \mathrm{d}\mathbf{w}_k \end{aligned} \tag{2-9}$$

This is a recursive process of estimation and prediction evolves over time with every new signal observation [81]. However, due to the nonlinearity of UPM process dynamics Eqn. (2-9) cannot be solved analytically. This is because the nonlinear state transition function f_k will cause the posterior distribution of state $\mathbf{w}_k$ to be non-Gaussian.

To surmount this challenge, a sequential Bayesian analysis based on a numerical Monte Carlo method (also called, particle filter) is used in this study to recursively solve Eqn. (2-9) [81]. Recently, the authors have used this method to compute the evolution of material removal rates in Chemical Mechanical Planarization (CMP) process [96].

The basic concept behind particle filter method is that any probability density function (pdf) can be approximated by a set of random samples $w(i)$ *i=1.. N,*

called *particles* that are drawn from another (analytically more tractable than $p(\cdot)$ given in Eqn. (2-9)) distribution $q(\cdot)$ (also called the importance pdf) as,

$$p(w) \approx \sum_{i=1}^{N} \xi(i)\delta\left(w - w(i)\right)$$
$$\omega(i) \propto \frac{p(w_i)}{q(w_i)}\xi(i) \propto \frac{p(w_i)}{q(w_i)} \tag{2-10}$$

where δ is the Dirac delta function and $\xi(i)$ is related to the ratio between the posterior pdf $p(w_{\mathrm{i}})$ and the selected importance pdf $q(w_{\mathrm{i}})$. For scalar signal observations y

$$\xi_k \propto \frac{p(y_k|w_k) \cdot p\left(w_k\middle|w_{k-p:k-1}\right) \cdot p(w_{0:k-1}|y_{0:k-1})}{q\left(w_k\middle|w_{k-p:k-1}\right) \cdot q(w_{0:k-1}|y_{0:k-1})} = \xi_{k-1}\frac{p(y_k|w_k) \cdot p\left(w_k\middle|w_{k-p:k-1}\right)}{q\left(w_k\middle|w_{k-p:k-1}\right)} \tag{2-11}$$

At every k we choose the importance pdf $\left(w_k\middle|w_{k-p:k-1}\right) = p\left(w_k\middle|w_{k-p:k-1}\right)$, referred to as the prior for time step k. Thus, Eqn. (2-11) becomes,

$$\xi_k \propto \xi_{k-1} \cdot p(y_k|w_k) \tag{2-12}$$

where $p(y_k|w_k)$ is the likelihood of measurement y_k, which can be estimated by the measurement noise $\mathbf{v}_k$ in Eqn. (2-8).

In summary, using Eqn. (2-8), the process state can be predicted one step ahead; and then using Eqn. (2-10) – (2-12) the weights of the particles of the process state can be updated by using the measurement data.

By using this sequential Monte Carlo sampling method, Eqn. (2-9) can be sequentially and recursively solved to estimate and predict the state vector. Also, akin to our recent application of PF (see Ref. [96]), we capture the nonlinear state transition using the state space representation given in Eqn. (2-1) – (2-7), and the

parameters were estimated by the same joint Bayesian estimation scheme used in this recent investigation.

The PF prediction accuracies critically depend on the initial prior $p(y_0|w_0)$. We had used the distribution of the parameters and the outputs of the trained RPNN to determine the initial prior distribution. As will be evident in the following sections, the use of this prior, as well as the appropriate structure to capture the dynamics helps avoid instability issues and convergence to local minima in PF.

2.4.3 On-line Change Detection

A non-parametric clustering technique, called, the mean-shift algorithm [97] is used to detect surface-variation induced changes in the dynamics of surrogate process state (i.e., the PF-updated RPNN parameters). Due to the non-parametric nature of this method, the shape as well as the number of clusters need not be known *a priori*.

Also, the method can relax some of the stationarity and Gaussian assumptions that underpin many of the statistical change detection techniques. For the presented application, we have noted that the surface variations cause changes in the second moment of the weights of the output node of the RPNN.

Therefore, we had considered a time-series v_i, formed by the 1000 samples (~100 ms) moving variance computed from the output node weight $\mathcal{W}_{4,4}^1$ (see Sec. 2.6.1, similarly, $\mathcal{W}_{2,2}^1$ for the numerical example in Sec. 2.5, and multi-sensor monitoring case in Sec. 2.6.2) to determine a possible surface induced shifts in the

weight pattern. At every time step *k*, we determined the centroid v_k^* of the time-series v_i as,

$$v_k^* = \arg\min_{v} m_h(v) = \frac{\sum_{i=1}^{n} v_i g\left(\left\|\frac{v - v_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{v - v_i}{h}\right\|^2\right)} - v \tag{2-13}$$

where $g(\cdot)$ is a kernel function that weights distances of different samples from their centroid v, and h is the parameter used to scale the distribution of the samples. Every v_k^* is compared with its previous value v_{k-1}^*, and if the difference is larger than a threshold a change is detected at time step *k*. Further intuition behind the overall approach is evident from the following numerical illustration.

2.5 Numerical Example

For illustrating the methodology, we have used a synthetic signal composed of sinusoidal waves of multiple frequencies (see Figure 2-7(a)). The first half (2,000 time steps) of the signal was composed of four frequencies (Figure 2-7(b)) and the next half had five frequencies (Figure 2-7(c)), but the amplitudes were adjusted to make the signal energies the same for both signal halves.

As one can gather from the time portrait (Figure 2-7(a)), the two signal halves have the same first and the second moment, akin to subtle signal variations resulting from incipient surface variations during UPM.

In such cases, the process drift, or the change point (here the data point #2,000) is not detectable with the use of conventional statistical change detection techniques. We trained an RPNN using the first 100 points of the signal as de-

scribed in Sec. 2.3.3. We used 2 nodes with 2 delay terms (**I** = **P** = 2) for the RPNN. Evident from Figure 2-7(d) is that the trained RPNN can accurately predict the evolution of the first half of this synthetic signal.

Next, we used a PF to update the RPNN parameters (weights and biases) as presented in Sec. 2.4.2. As shown in Figure 2-7(e), the PF predicts signal evolutions satisfactorily for both signal halves ($R^2 > 0.95$)[9]. The PF updated values of the RPNN output node weight was used for on-line change detection. Here, we compared the performance of mean-shift statistic with that of a conventional statistical monitoring technique based on Sequential Probability Ratio Test (SPRT).

An SPRT statistic can be used to detect changes in the first and the second moment of the data. Figure 2-7(f) summarizes the results of SPRT with different data sets. The values of the SPRT statistic that lay above the threshold (horizontal dashed line) have a 99% chance that a change or a drift has occurred in the computed statistic (i.e., Type I error or false positive rate, $\alpha < 1\%$). It may be noted that this study evaluates how well the changes in the statistics are indicative of the actual change in the signals (which occurs at the data point #2,000). First, we applied SPRT on the raw data sets (dashed line).

As shown in the figure, the SPRT statistic falsely crosses the threshold much earlier (near data point #1,000) than the real change point. But it fails to detect the change (the statistic values remain below the threshold) even after 1,000 data

[9] R^2 statistic ($0 \leq R^2 \leq 1$) quantifies the fraction of the variation in the signals that the model captures. The higher (closer to 1) the R^2 statistic values the more accurate is the model.

points following the actual change. Such a mode of failure of the method is referred to as a Type II error (failing to detect).

In contrast, the SPRT statistic values of the RPNN weight cross the threshold at 2,266th data point, i.e., the SPRT of the weights is able to detect change in the signal characteristic 266 data points after the actual change that occurs at the data point #2,000. This result indicates that the use of RPNN parameters as surrogates of the process state can be a more effective means to detect process variations compared to tracking statistical variations in the raw signals.

Although the RPNN was able to detect the change for the presented case, latency of 266 data points following the actual inception of the change may not be early enough for effective mitigating control in precision manufacturing applications. Comparatively, we applied the mean-shift algorithm described in the previous section on the raw data sets and RPNN weights.

Using this method the change was detected at the 2,285th and 2,010th data point (see Figure 2-7(g)), for the raw data set and RPNN weights respectively. The latter is within 10 data points following the actual change inception. This outcome indicates that compared to the use of SPRT statistic, the use of mean-shift with RPNN weights can lead to earlier detection of process-induced changes in the signal patterns.

Next, we studied a case where we had the same signal characteristic (i.e., the same 4 frequencies) in as the previous case for the entire length of the time-series.

Interestingly, the SPRT statistic of the raw signal wrongly detected changes (false positive or Type I error) at two instances as indicated in Figure 2-7 (h). The SPRT of the raw weights also detects a nonexistent change in the signal characteristic. This false positive rate needs to be lowered for effective change monitoring. Pertinently, the mean-shift algorithm does not yield any false alarms.

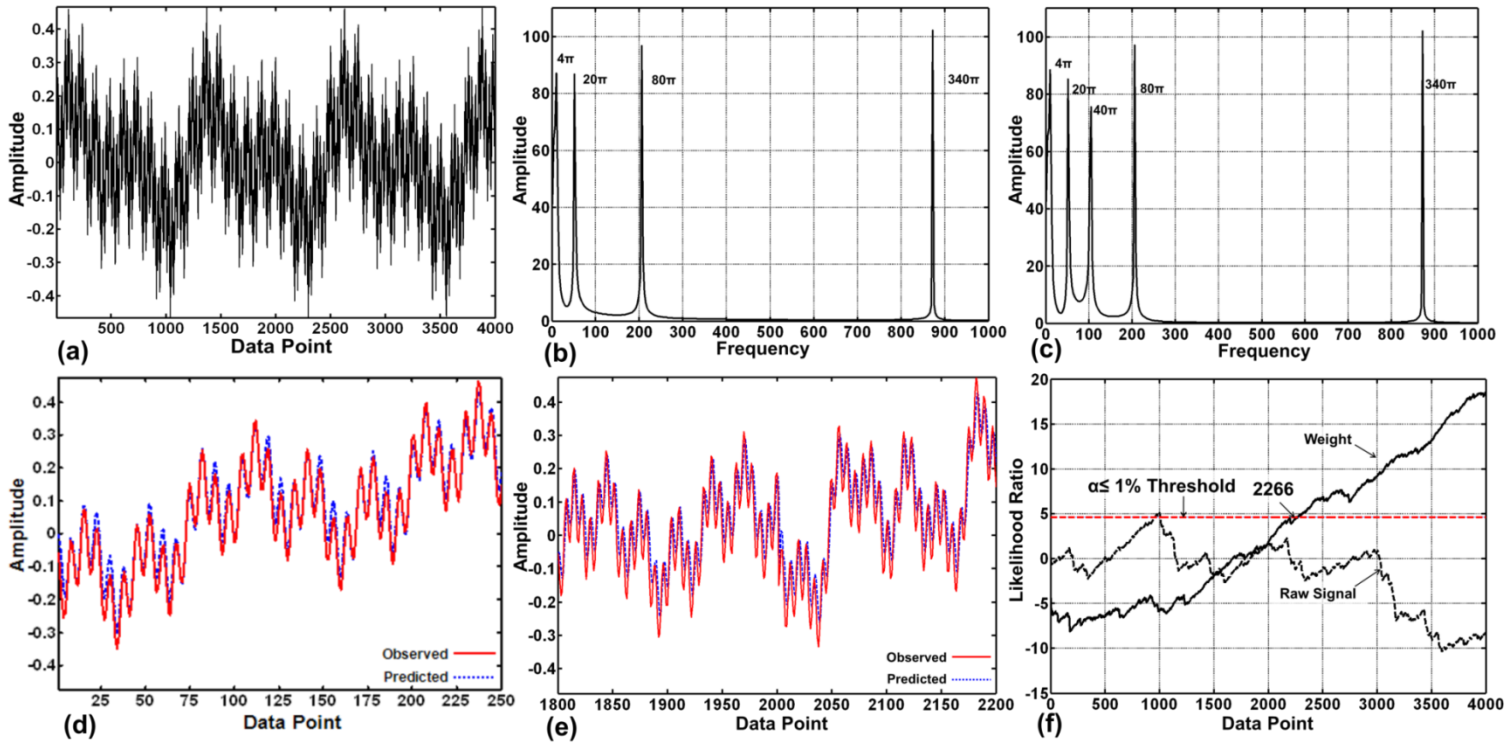
The values of the mean-shift statistic were below the threshold for all the time steps when applied to the raw data set, as well as the RPNN weights (see Figure 2-7(i)). Thus, the numerical study indicates:

i. RPNN and PF are effective in capturing the evolution of the process underlying the signals,

ii. the use of RPNN parameters are adequate surrogates to track the process dynamics instead of using the statistical signal patterns alone, and,

iii. mean-shift algorithm applied to the moving variance of the PF-updates of RPNN weight is effective in early detection of subtle changes in signal characteristics (no apparent mean or variance shift), and minimizing the false positive rates compared to the use of conventional statistical change detection techniques, such as SPRT.

The following section describes the application of the present approach for monitoring of precision turned surfaces.


Amplitude
Data Point
(a)
Frequency
(b)
4π
20π
80π
340π
(c)
40π
Observed
Predicted
(d)
(e)
Likelihood Ratio
Weight
α≤ 1% Threshold
2266
Raw Signal
(f)

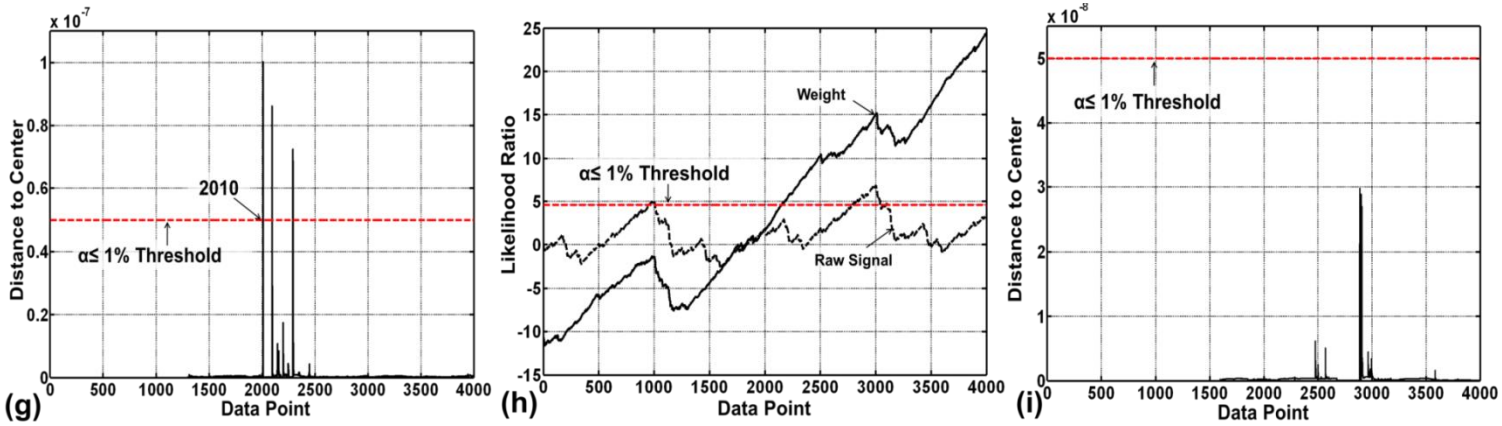


Figure 2-7: Summary of results from the numerical example.

(a): Concatenated (with change case) sinusoidal simulated time series with the first 2,000 points consisting of 4 periods, and last 2,000 points with 5 periods.
(b) & (c): Frequency portraits (~2 kHz sampling rate) of the two sinusoidal time series shown in (a) with 4 and 5 periods respectively.
(d): Performance of trained RPNN for the with change case.
(e): Performance of PF-updated RPNN for the with change case.
(f): SPRT applied to the with change case simulated time series showing both Type I and Type II error and SPRT applied to the RPNN weights.
(g): Behavior of the mean-shift statistic when applied to RPNN weight time series with change.
(h) SPRT applied to the RPNN weights and raw signal consisting of 4 periods alone (no-change case), both instances depict Type I error.
(i): Behavior of the mean-shift statistic when applied to RPNN updated weights for the no-change case.

2.6 Change Detection in Ultraprecision Machining Process using the RPNN-PF approach

2.6.1 Change Detection Using One Sensor Signal (vibration signal along the feed direction)

As noted in the foregoing, early detection of surface variation can be vital for minimizing scrap rate and rework in an industrial UPM process. Figure 2-8 summarizes the overall result from the application of RPNN-PF for detecting changes in the surface characteristics during UPM using vibration signals in the feed direction. Figure 2-8(a) shows a surface zone after UPM, and Figure 2-8(b) shows the feed direction vibration signal obtained during the machining of that zone. It can be seen that the signals bear information sensitive to machined surface characteristics.

For example, Figure 2-9 shows a portion zoomed on the surface obtained during the ultraprecision facing operations with a carbide tool (~0.4 mm (396 μm) nose radius), at 2000 RPM, 60 mm/min feed rate (30 μm/rev feed), and 12 μm depth of cut. As stated in Sec. 2.4 the carbide and PCD (instead of SCD) tools was used as they present less risk for damage due to sudden introduction of the coolant in these experiments.

During the process, the surface speed varied from 1021 m/min at the periphery (φ 16.25 cm) to near zero at the center. Two notable process characteristics reflected in the surface are (i) the feed marks occurring at 30 μm intervals, and (ii) the tool radius chord length of ~200 μm. The figures also show the feed vibration

signal obtained concurrently during the process. It can be noted that the signal peaks appear to be aligned with the valleys of the surface, and the signal appears to be modulated at a wavelength of 200 μm. This examination of the signals and the surface indicates that the measured signals are sensitive to variation in the material removed under different surface characteristics. Similarly, Wang, *et al.* [92] have noted a correlation between workpiece surface characteristics and force sensor signals in UPM of Al 6061.

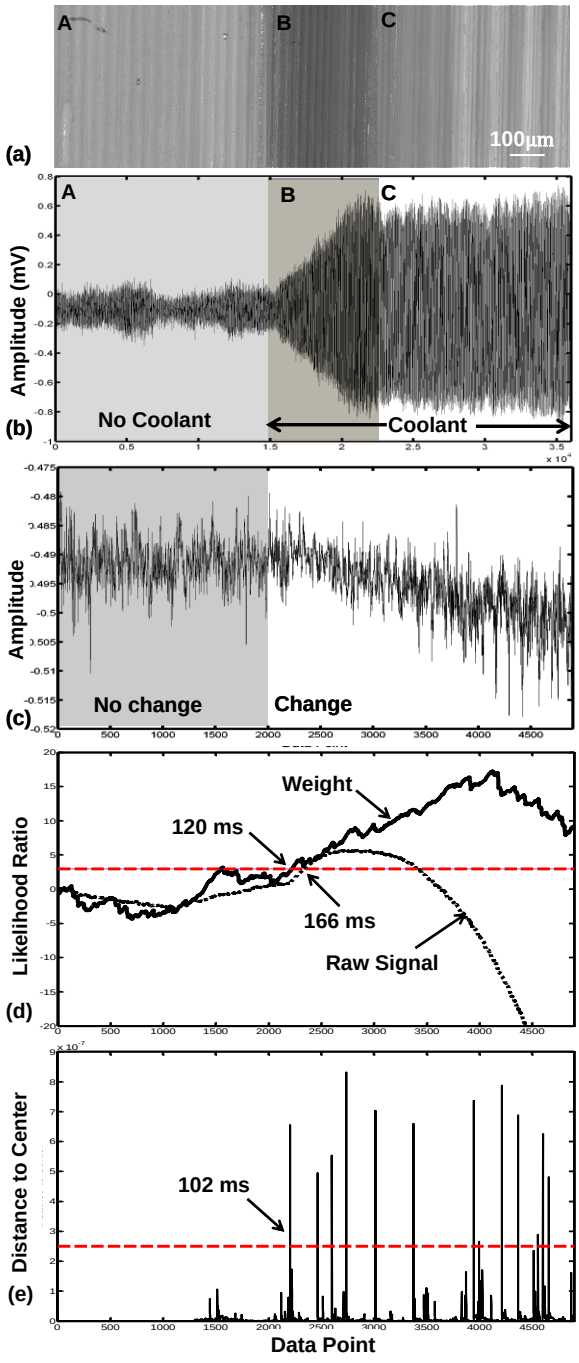


Figure 2-8 : Overview of the RPNN-PF method applied to precision face-turning (facing) process.

(a): Surface obtained during machining of aluminum before application of coolant (portion A) and after application of coolant (portions B and C). (b): Feed direction vibration signal (V_x) observed isochronously during machining of portions A, B and C. (c): Behavior of one particular PF-updated RPNN weight ($\widetilde{w}^1_{4,4}$) before and after application of coolant. (d) SPRT applied to the raw times series shown in (b) and the RPNN weight shown in (c). (e): Mean-shift clustering applied to the RPNN weight.

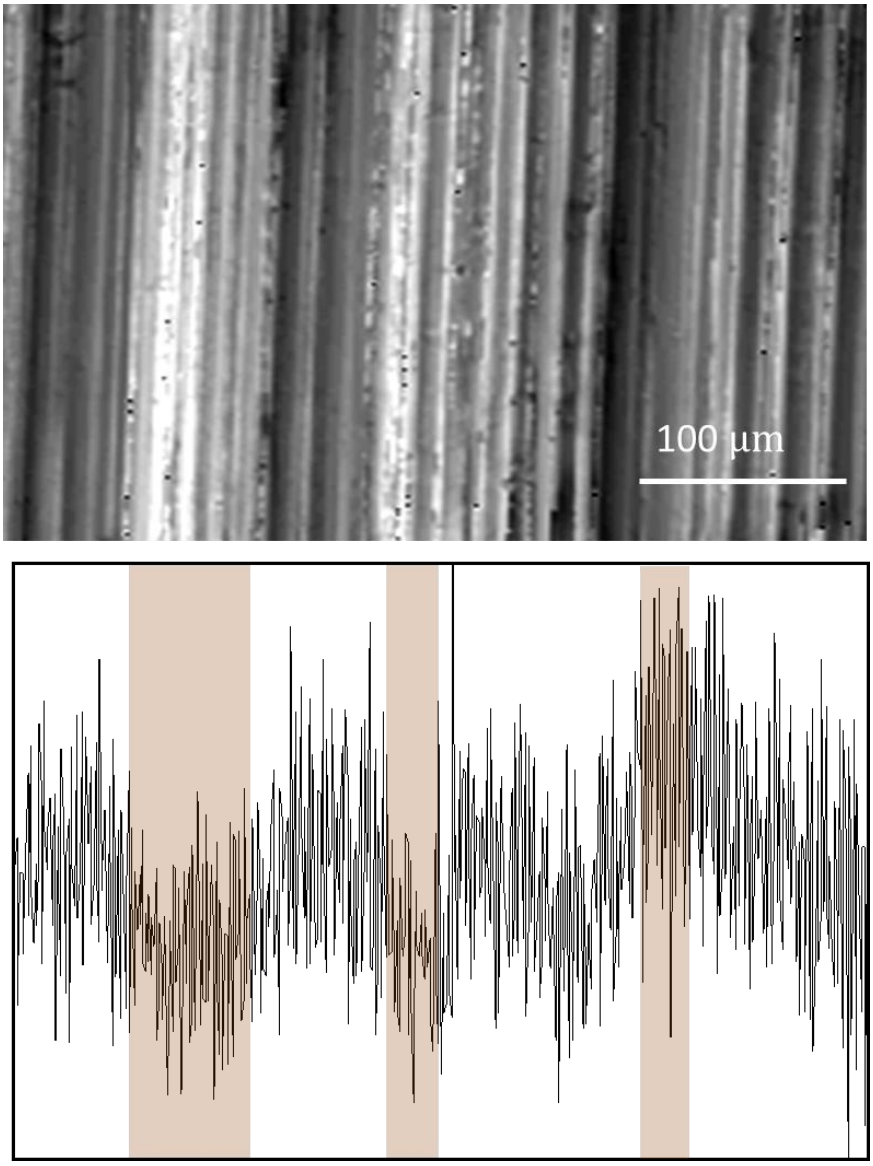


Figure 2-9 : Surface and concurrent signal obtained during facing test

Test condition are: 2000 RPM, 60 mm/min feed rate and 12 µm depth of cut. Signal down sampled from 10 kHz to 100 Hz.

In Figure 2-8(a), the initial portion (labeled A, lies predominantly in the 6th zone as measured from the periphery) was realized under dry cutting conditions with spindle speed of 2000 RPM, 6 mm/min feed rate (3 μm/rev feed), and 8 μm depth of cut. The resulting surface roughness was Ra ~ 200 nm.[10] When the end of portion A was reached (start of 7th zone), a coolant (Kool Mist-78®) was introduced and continuously applied for the remainder of the operation.

We note that the application of the coolant in this study was not aimed at optimizing the surface characteristics, but to introduce abrupt changes in the surface characteristics. Portions labeled B and C were obtained with the use of this coolant. Soon after the coolant was applied, the surface characteristic changed to reveal certain dominant periods (with wavelengths of about 40 μm) as shown in Figure 2-8(a) for portion B, and the chip morphology changed from discontinuous to continuous type.

Continued application of the coolant in mist form for further 6 sec leads to a wavy surface with a wavelength of about 100 μm, as shown in Figure 2-8(a), portion C. Evidently, the surface characteristics deteriorate significantly from A to C, to Ra > 300 nm. This is likely a consequence of chemical [98, 99] and thermo-mechanical interactions [100] that take place in the cutting zones in the presence of a coolant.

[10] The reason for obtaining this surface roughness value is because we used a tungsten carbide tool instead of a single crystal diamond tool. It is well known that single crystal diamond tool would give a surface finish in the nanometric range. The main objective of this study was to investigate a multiple sensor fusion approach (with RPNN and PF) for early detection of changes in surface characteristics.

Wang, *et al.* [92] have suggested that the effect of fluid damping due to introduction of coolant may cause changes in surface characteristics. Such oscillations in cutting forces and temperature may lead to the deterioration in surface characteristics as seen in Figure 2-8 (a) portion C. Figure 2-8(b), shows that the vibration peak to peak amplitudes progressively increase with the application of the coolant during machining over portion B, reaching a steady state when portion C is reached.

While there is a noticeable change in the signal characteristics with surface variations, the use of time domain vibration signals may not lead to early stage surface detection; see Figure 2-8(d). In UPM, earlier detection of defects even by a few milliseconds can significantly avert excessive rework. Towards this end, we use the PF updates of the trained RPNN[11] (Figure 2-8(c)) as surrogates for the raw signal. The RPNN chosen consisted of four nodes (**I**) and four delays (**P**), implying a total of 40 (= **P** x **I**(**I**+1)/2) weights ($\mathcal{W}$) and 4 bias (B) parameters. This peculiar choice is made for the purpose of symmetry.

Pertinently, the embedding dimension and auto-mutual information of the signal [101] were observed to provide a value close to the selected **I** and **P**, respectively. It should be noted that the raw signal is averaged using a rectangular moving window of length 5, and down sampled to $1/5^{th}$, to suppress the high frequency contents that mostly emerge from extraneous sources. The length of the rectangular window is selected based on spectral frequency estimates. The largest

[11] Since we use only one sensor signal, the term $W_E^T \times \mathcal{Y}(t)$ will become zero in Eqn. (2-4).

window length, which does not excessively attenuate the dominant signal frequency characteristics, is deemed as an appropriate window length. Such down sampling did not have a deleterious effect on prediction delay, and was in turn, observed to aid in faster RPNN convergence. A representative of this down sampled signal is shown in Figure 2-8(b).

For this representative case, a total of 180,000 ($\boldsymbol{v}_{\mathbf{x}}$) data points are gathered at 10 kHz sampling rate from the accelerometer mounted in the feed direction (V_x) during the machining operation conducted at 2000 RPM, 6 mm/min feed rate and 8 μm depth of cut on the aluminum workpiece described in preceding sections. The data set $\boldsymbol{v}_{\mathbf{x}}$ temporally corresponds to 18 sec epoch of operation, or equivalently, 600 workpiece revolutions, or 1800 μm on the workpiece surface. A $1/5^{th}$ down sampling of this data set yields 36,000 data points long data set $\tilde{\boldsymbol{v}}_{\mathbf{x}}$ shown in Figure 2-8(b). This data is in synchronization with the workpiece surface segment in Figure 2-8(a) as follows:

- First 7.25 sec or corresponding to ~ 725 μm of surface machined is captured in data points 1 through 72,500 $\in \boldsymbol{v}_{\mathbf{x}}$, or data points 1 through 14,500 $\in \tilde{\boldsymbol{v}}_{\mathbf{x}}$ in portion A of Figure 2-8(a).
- Next 4 sec (~ 400 μm machining ending at 1125 μm total length), is synchronous with data points 72,500 through 112,500 $\in \boldsymbol{v}_{\mathbf{x}}$, or data points 14,500 through 22,500 $\in \tilde{\boldsymbol{v}}_{\mathbf{x}}$, represented by portion B
- Last 6.75 sec (~ 675 μm machining, ending with 1800 μm total length) is synchronous with data points 112,500 through 180,000 $\in \boldsymbol{v}_{\mathbf{x}}$, or data points 22,500 through 36,000 $\in \tilde{\boldsymbol{v}}_{\mathbf{x}}$, as represented by portion C

From this down sampled dataset, 6,000 points, representing 3 sec of machining, are chosen for our analysis, these are taken in the vicinity of the detected change isochronously with portion B, i.e., when the coolant was first introduced. With respect to the machined surface, these points are equivalent to 300 μm in length.

Out of the 6,000 points (data point 12,700 – 18,700 in Figure 2-8(b)) in the down sampled signals, the first 3,300 points are the down sampled vibration measurements taken before the change (no coolant) state while the remaining 2,700 points were taken after the change (with coolant). The first 3,300 points are further partitioned into two parts, the initial 1,100 points are used for training the RPNN while the remaining 2,200 points are concatenated with the 2,700 points taken with coolant on, and used for testing and PF updates. The 4,900 points shown in the abscissa of Figure 2-8(c) - (e) correspond to the time steps at which these 4,900 vibration measurements were taken**.**

The RPNN is initialized with Gaussian distributed random weights of the order of 10^{-1}. The learning rate is adjusted heuristically, as suggested by Hagan, *et al.* [94]. We set a primary stopping criteria of $R^2 = 0.85$, and secondary criteria threshold of 2500 iterations for the RPNN training process. If the primary stopping criteria is not met at the end of the iteration threshold, the training process is re-started and proceeded till convergence, or until the set number of iterations are reached, or the learning rate reaches 0.001% of the initial value.

The training process is continued until four converged results are obtained. The parameters obtained over these four RPNN training loops are arithmetically averaged, this averaged RPNN parameters are taken as the surrogates for the process state. The RPNN parameters are updated by the PF for optimal one step ahead prediction (see Figure 2-8(c)) as each data point from the 4900 points long data set is presented in sequence.

These updated parameters set ($\widetilde{\mathcal{W}}$) are now considered surrogates of the process state during those time intervals. It may be noted that $\widetilde{\mathcal{W}}$ is dynamic, and identical in length (4900 data points) to that of the signal segment used in the PF update.

Figure 2-8(d) shows results from the Sequential Probability Ratio Test (SPRT) applied to both the raw signal $\boldsymbol{v_3}$ and the one particular PF updated weight ($\widetilde{\mathcal{W}}_{4,4}^{1}$) shown in Figure 2-8(c). SPRT detects the change in the signal type within 120 ms when applied to $\widetilde{\mathcal{W}}_{4,4}^{1}$. In contrast, the shift in the raw signal alone takes over 166 ms to identify.

Figure 2-8(e) depicts the detection performance of the mean-shift statistic based change detection applied to variation of the shown weight (same as Figure 2-8(c)). The change detection delay due to mean-shift clustering statistic reduces to 102 ms. This reduction in the latency of change detection can be attributed to the less restrictive assumptions that underpin the mean-shift statistic compared to the SPRT statistic.

Also, we have tested the present approach to cases where the surface characteristics were fairly consistent, and no significant surface changes were noticeable (see Figure 2-10 for a summary of a representative result). In this case, the mean-shift statistic values were well below the threshold for the entire time span over which the surface segment was precision finished.

A similar pattern was noted for SPRT statistic of raw signal as well as PF-updated RPNN weight values, reinforcing that this statistic is fairly insensitive to small signal variations. However, as noted from the previous example, lack of sensitivity with the conventional methods can lower the accuracy and earliness of detection of surface variations, and the present approach involving the use of PF-updated RPNN weights along with the mean-shift statistic offers a robust means for the detection of incipient surface variations in the process.

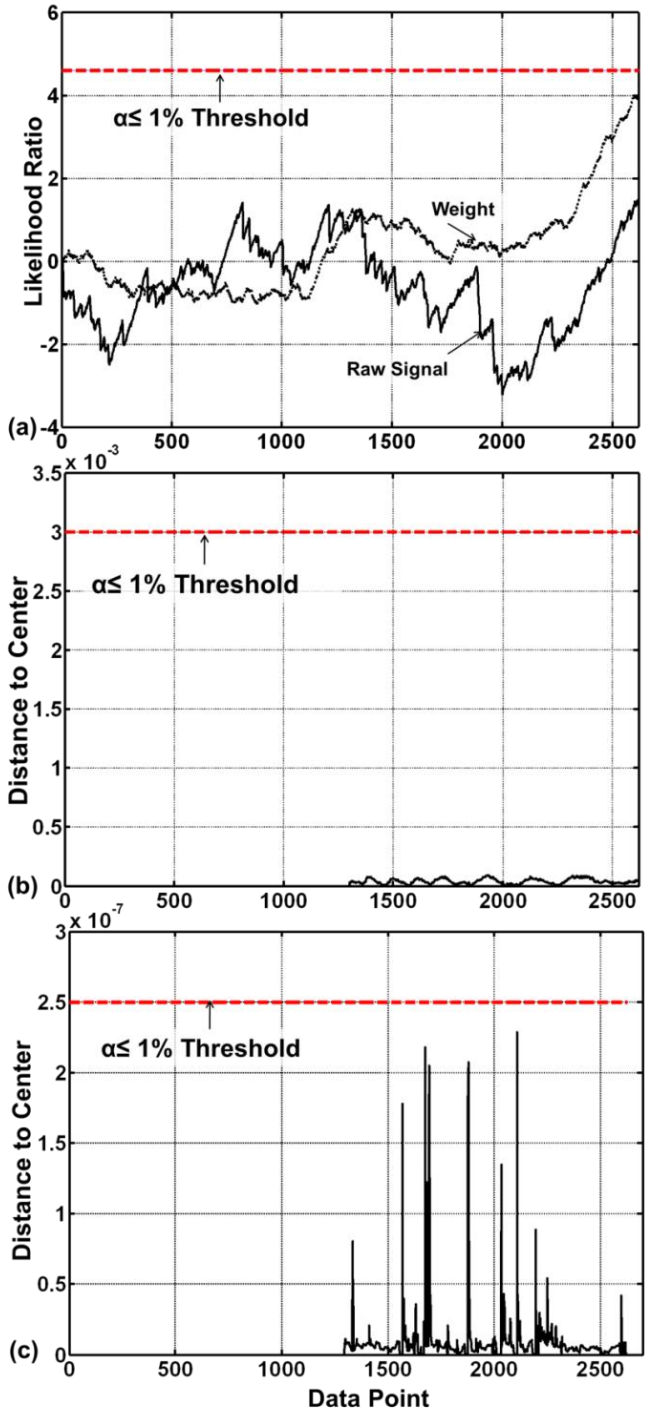


Figure 2-10 : Summary of the application of the present approach to signals from the feed direction vibration sensor for the portions where no change is apparent.

(a) variation of the SPRT for weights and raw signal statistics. (b): Behavior of the mean-shift statistic as applied to the raw signal. (c): Behavior of the mean-shift statistic as applied to RPNN weights.

2.6.2 Fusion of Information from Multiple Sensors

Next, we continued with the same procedure using two (vibration and force in the feed direction), as well as all sensors simultaneously. The signals from sensors other than the feed direction vibration sensor were taken as exogenous inputs as explained in Sec. 2.3.3. These exogenous inputs influence the $W_E^T \times \mathcal{y}(t)$ term in Eqn. (2-4) which was cancelled out for the case presented in Sec. 2.6.1. It is surmised that UPM process signals contain significant amount of redundant information. Increasing the number of sensor signals, albeit, increasing the redundancy, would contribute to increase the robustness of change detection.

We conducted four replications of the facing tests using a PCD cutting tool under cutting conditions identical to Sec. 2.6.1, but with isopropyl alcohol as coolant. Surface finish (Sa) in the range of 80 to 100 nm was obtained during these tests.

The signals gathered from these tests were used to train an RPNN. The feed direction vibration sensor signal was used as the primary signal, with other sensor signals in the exogenous role. This is because the feed vibration signals, as stated in Sec. 2.6.1, was most sensitive to process variations that alter the surface profiles.

Thereafter, we had used the trained RPNN weights to initiate a PF to track and detect possible process variations as captured by the dynamics of these weights. Both SPRT and mean-shift statistics were used for change detection.

As indicated in Figure 2-11, fusion of signals from two sensors, namely feed vibration (as the main variable for RPNN), and cutting force in the feed direction[12] (as an exogenous input) with mean shift statistics, reduces the time-delay in detecting a change in UPM process dynamics by a half (~ 15 ms) of that with one sensor signal (~ 27 ms).

In addition, the variation (range) of the detection times is reduced from about 20 ms to ~ 5 ms with multiple sensors, leading to a more consistent performance. We also observed that using more than two sensor signals did not significantly improve the mean detection time. In general, the more number of sensors the more accurate and robust would be the predictions. Also, the saturation of prediction accuracies with two sensors may be peculiar to the surface characteristic changes studied in the present investigation. One may need a different combination and number of sensors to facilitate early detection of other process anomalies.

Using exogenous sensor inputs allowed for a more tractable RPNN structure. An RPNN consisting of only two nodes (**I** = 2), and two delays (**P** = 2), i.e., with six weights and two biases (and additional weights for exogenous inputs depending on the number of sensor signals used) was seen to suffice (compared to 40 weights and 4 biases), thus significantly reduces the computation load.

Thus, using more than one sensor signal for change point detection allows for the following noteworthy outcomes in comparison to the no exogenous input case:

[12] We have tested several combinations of primary and exogenous inputs to arrive at this conclusion.

i. reduction in mean change detection delay from ~ 27 ms to about 15 ms,

ii. significant improvement in reliability of detection with roughly 5 ms range of uncertainty (95% confidence interval) compared to 20 ms with one sensor signal, and

iii. reduction in the computational load due to the use of 6 weights and 2 biases, instead of 40 weights and 4 biases.

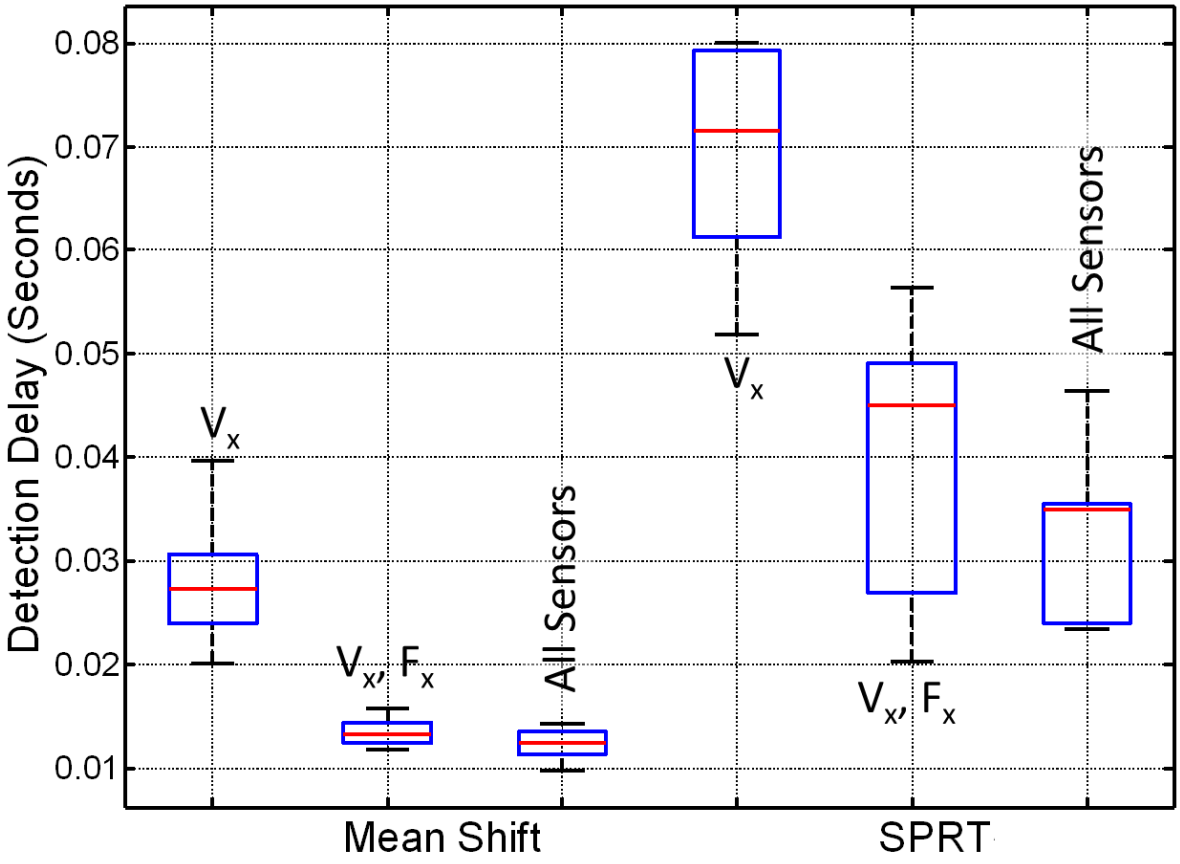


Figure 2-11: Effect of multiple sensor fusion on change detection delay.

comparing performance of the mean-shift and SPRT applied to RPNN weights with (i) feed direction vibration sensor (V_X) alone, (ii) with feed direction vibration and force sensor (V_X, F_X), and (iii) all sensors. The saturation of prediction accuracies with two sensors may be peculiar to the current situation where the change in surface morphology is prominent.

2.7 Summary

We have reported a new approach, namely, Recurrent Predictor Neural Network – Particle Filter (RPNN-PF) for real time monitoring of incipient surface variations in the UPM process. This approach can effectively capture the nonlinear and non-stationary dynamics of the process, instead of just tracking the statistical patterns of the signals. Consequently, it was able to detect surface variations within 15 ms of its inception. Specific conclusions are summarized in the following:

1. The experimental UPM machine tool was integrated with multiple sensors, namely, a three component accelerometer to measure tool vibration along these orthogonal directions, a three component force piezoelectric dynamometer, and an AE sensor. Among these, feed vibration sensor was found to be most sensitive to the variations in the surface characteristics. Surface finish on Al 6061 discs varies over Ra ~ 15-25 nm range with single crystal diamond (SCD) tools. Experiments also suggest that surface morphology in UPM can undergo sudden and almost abrupt variations. The information discerned from multiple sensors was used for predicting these variations in their incipient stages.
2. The complex evolution of signal patterns, emerging from the underlying nonlinear process dynamics was effectively captured using an RPNN. The RPNN was trained for one-step ahead prediction of the feed vibration signal. After training, the RPNN was thus able to capture the nonlin-

ear process dynamics, and the RPNN outputs matched well with the vibration signal evolutions (R^2 ~ 85%). Therefore, the parameters of the RPNN, including the weights and biases, were treated as surrogates of the process states, and their time-bound variation was tracked to monitor the process state changes.

3. The RPNN parameters were continually updated with every new vibration signal measurement using the PF. The temporal evolution of the PF-updated RPNN parameters, i.e., the surrogate process states, was found to reflect the inherent complexity in the raw signal. This provides a means for tracking surface vibration-induced changes in the process dynamics as opposed to mere statistical signal patterns.
4. Since the UPM process dynamics exhibits an inherent time varying [64, 77] and nonlinear [88] (non-Gaussian) nature, a nonparametric mean-shift statistic of moving variance of the output node weight values was applied to detect process drifts. This mean-shift statistic based clustering method is shown to correctly identify surface-induced changes in the process dynamics. It was also found to be robust to false positives (false alarm or Type I error), and provides for early stage detection in comparison to conventional sequential statistic based methods, such as SPRT. On the other hand, conventional and mean-shift statistics applied upon the raw signal itself were found to yield high false positive (failing to detect or Type II error), and low detection rates.

5. Another noteworthy aspect of the presented method, evident in the numerical case studies is that the present approach was able to detect subtle harmonic distortions taking place without significant changes in the mean or the variance. Such changes were detected within 2 dominant cycles. In comparison, the experimental studies have shown that change detection based on SPRT statistic on the extracted features from RPNN-PF can sustain both Type I (false alarm) and Type II (failure to detect) errors.
6. The use of multiple sensor signals can improve both the timeliness and robustness of change detection compared to the use of one sensor signal alone. Furthermore, multi-sensor fusion allowed for reduction in the computation load involved in training the RPNN.
7. By incorporating the force sensor signals in the feed direction, along with the feed direction vibration signals, changes in process dynamics was detected, on an average, within 15 ms of its inception, with an uncertainty of ± 2.5 ms (95% confidence interval). Peculiar to the defects considered in this study, the inclusion of information from more than two sensors does not seem to further enhance the timeliness or robustness of identification. It may, however, provide other defect information not possible with a single sensor. These results indicate that the sensor signal information may be coupled, and much of the information pertaining to process dynamics responsible for the surface variations considered in this study may be captured in two signals. However, from a practical stand-

point, fusion of information all sensors enhances the redundancy and therefore robustness of a monitoring approach for early detection of changes in the process.

Thus, the present approach can lead to a reliable monitoring system for detecting incipient surface characteristic variations in the UPM process. Future investigations will attempt to combine information from other signals, as well as use longer-range predictions to improve the reliability as well as timeliness of incipient surface variation detection.

3 Process-Machine Interaction (PMI) Model-based Monitoring of the Chemical Mechanical Planarization (CMP) Process using Wireless Vibration Sensors

We use a deterministic process-machine interaction (PMI) model that can associate different complex time-frequency patterns, including nonlinear dynamic behaviors that manifest in vibration signals measured during a chemical mechanical planarization (CMP) process for polishing blanket copper wafer surfaces to near-optical finish (Sa ~ 5 nm) to specific process mechanisms. The model captures the effects of the nonuniform structural properties of the polishing pad, pad asperities, and machine kinematics on CMP dynamics using a deterministic two degree of freedom nonlinear differential equation.

Goals

- Identify and statistically quantify those aspects of MEMS wireless sensor signal components that are relevant to the CMP process dynamics, and therefore useful from a real-time monitoring perspective.
- Explain based on process dynamics and physical phenomena, the physical reason for certain vibration sensor patterns observed in CMP and thereby facilitate real-time process prognosis.

The model was validated using a Buehler (Automet 250) bench top CMP machine instrumented with a wireless (XBee IEEE 802.15.4 RF module) multi-sensor unit that includes a MEMS 3-axis accelerometer (Analog Devices ADXL 335). Extensive experiments suggest that the deterministic PMI model can capture

such significant signal patterns as aperiodicity, broadband frequency spectra, and other prominent manifestations of process nonlinearity.

Remarkably, using the deterministic PMI model we were able to explain not only the physical sources of various time-frequency patterns observed in the measured vibration signals, but also their variations with process conditions. The features extracted from experimental vibration data, such as power spectral density over the 115 − 120 Hz band, and nonlinear recurrence measures were statistically significant estimators ($R^2 \sim 75\%$) of process parameter settings. The model together with sparse experimental data was able to estimate process drifts resulting from pad wear with high fidelity ($R^2 \sim 85\%$). The signal features identified using the PMI model can lead to effective real-time *in situ* monitoring of wear and anomalies in CMP process.

The following challenges and recommendations noted by Bukkapatnam, *et al.* [27] provides the contextual motivation for this work[13].

[13] P. Rao, M. B. Bhushan, S. Bukkapatnam, Z. Kong, S. Byalal, O. Beyca, A. Fields, and R. Komanduri, "Process-machine interaction (PMI) modeling and monitoring of chemical mechanical planarization (CMP) process using wireless vibration sensors," *IEEE Transactions in Semiconductor Manufacturing*, (Accepted, 1st Review**)** 2013.

Challenges in nanomanufacturing research [27]

Nanoscale processes and systems pose many challenges for sensing:

- Accessibility to signal source is not easy;
- *In situ* sensing is almost impossible;
- Signals are short, evanescent, and weak;
- Quantization of signals makes transduction difficult; and
- Signal-to-noise ratio is low.

Mechanisms that cut across multiple scales make observation and characterization of nanomaterials and nanoprocesses difficult.

Expert recommendations for nanomanufacturing research [27]

Physics-based statistical models considering nanoscale and hierarchical physics, and nonlinearities should be developed. These realistic models will enable monitoring, diagnostics, prognostics, and reliability analysis.

3.1 Introduction

"Data contain both signal and noise. To be able to extract information, one must separate the signal from the noise within the data."

– Walter Shewhart

CMP is a vital back-end-of-line (BEOL) process in semiconductor manufacturing for obtaining both local and global planarity on a variety of materials [19, 21]. CMP is often the last step before device testing and packaging stages [21]. As a consequence, wafer anomalies resulting from CMP operations will lead to high losses [22].

Advent of copper (Cu) semiconductor interconnects as a viable alternative to tungsten (W) and aluminum (Al) poses additional challenges to yield [102]. This

is because tantalum (Ta) and tantalum-nitride (TaN) barrier layers designed to prevent diffusion of Cu into the neighboring silicon dioxide (SiO_2) - low k dielectric have low selectivity with respect to Cu, and Cu interconnects are easily damaged during CMP due to their relative softness compared to W and Al [103]. Stringent control of operating conditions is therefore considered essential for defect-free realization in CMP process (see also Graphic IX) [19].

Industry predominantly uses off-line statistical process control (SPC) methods by surface characterization of test wafers for process quality assurance in CMP [104]. However, the use of test wafers for process monitoring can lead to ~ 35% reduction in throughput, and cause as much as 100% increase in cost of ownership [25]. These traditional SPC methods may fail to detect some of the subtle process drifts inherent to CMP [22, 105-107]. Consequently, real-time *in situ* sensor-based approaches have been pursued for CMP monitoring [22, 25, 108-111].

Contemporary CMP monitoring approaches primarily use piezoelectric vibration and force [112-124], acoustic emission (AE) [53, 125-129], laser [130, 131], electro-chemical [132, 133], and thermo-optical [134-137] sensing elements. Apart from cost, these sensing systems require careful attention to calibration and location. *In situ* optical sensing systems for CMP endpoint detection typically need specially designed polishing pads with optical filter windows [108, 131, 135].

Also, the added bulk and high power consumption of these sensing systems limits their applicability to non-intrusive close proximity monitoring. Hence, the majority of these approaches are mostly limited for detection of CMP endpoint as opposed to tracking anomalous process variations [109]. Wireless MEMS sensors have been used towards alleviating this issue [117, 118, 137]. These sensors also facilitate close proximity monitoring of the process.

However, the sensor signals acquired from the CMP process exhibit certain complex time-frequency patterns [116]. For instance, Figure 3-1(a) shows a time portrait of a representative signal acquired from a wireless MEMS vibration sensor mounted on our CMP apparatus (see Sec. 3.4). The signal exhibits a beat-like pattern with prominent periodic low frequency component, superimposed with aperiodic high frequency components.

Such inherently complex signals can manifest broadband frequency spectra as seen in Figure 3-1(b). Traditional methods use features, such as band limited energies, statistical moments, etc., that quantify the statistical patterns in a signal and not the inherent process dynamics in CMP.

Challenges

1. Wireless vibration signals obtained during CMP process have low signal to noise ratio and depict complex nonlinear, nonstationary, and broadband characteristics. Under such circumstances conventional statistical process monitoring techniques are of limited value.
2. Given the complex signals emerging from the process, it is not known what aspects (features or frequencies) of the signal are relevant to the process.

A physical model capable of elucidating the multi-faceted aspects of CMP process dynamics can be used to define features that can track the process, as opposed to the mere signal variations. Existing physical models, however are largely focused on the wafer-pad asperity level mechanics, and explain the material removal regimes dominant at such a scale, such as hydrodynamic, mixed, and direct contact modes [138-143].

These models overlook the interaction among mechanics active at different scales, such as bulk pad structure, and machine kinematics. Consequently, the model solutions cannot be associated with complex patterns in the measured vibration signals, and therefore are not suited for process monitoring.

We invoke a deterministic process-machine interaction (PMI) model which incorporates the effects of the nonuniform structural properties of the polishing pad and machine kinematics on CMP dynamics at the wafer-pad interface using a deterministic two degree of freedom nonlinear differential equation. The PMI model is used to explain the physical sources of various time-frequency patterns observed in the measured vibration signals, as well as, their variations with CMP process conditions. The signal features identified based on the PMI model can track inherent variations in CMP process as opposed to just statistical signal patterns, and thus would lead to effective real-time *in situ* monitoring of drifts (wear) and anomalies in CMP process.

The remainder of this chapter is organized as follows: a review of the relevant literature is provided in Sec. 3.2, an overview of the research approach is

presented in Sec. 3.3, experimental validation of the PMI model in Sec. 3.4, and the application of the PMI model for condition monitoring in CMP is discussed in Sec. 3.5.

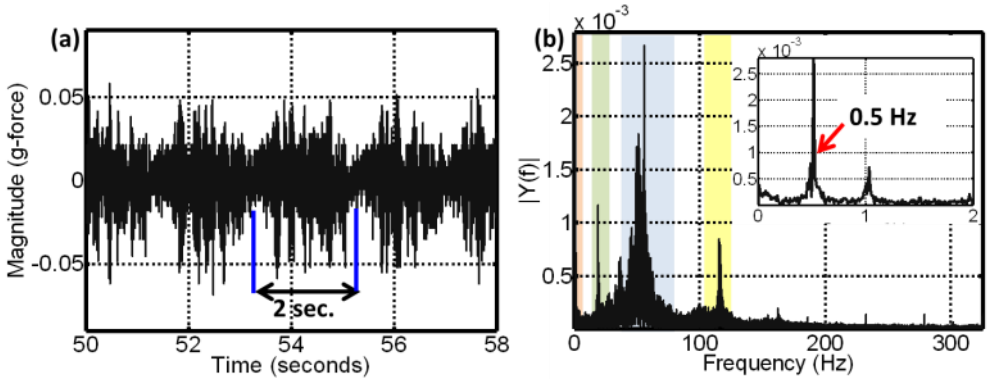
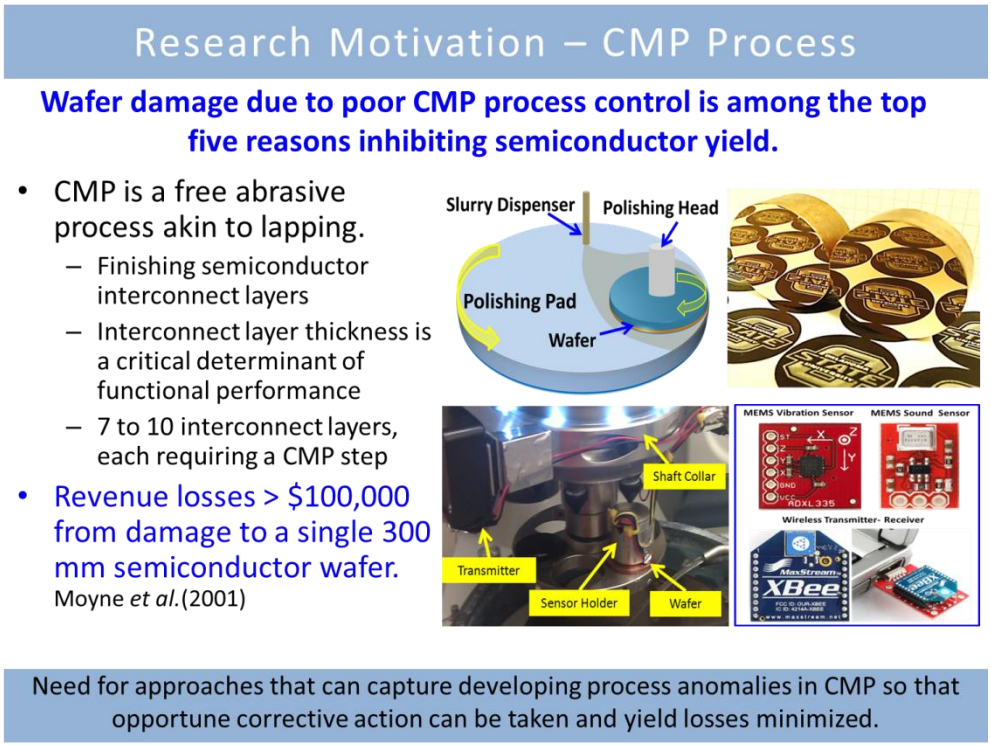


Figure 3-1: Typical vibration sensor data in the tangential (X) direction obtained during CMP.

Polishing of copper surfaces under the following conditions: 30 RPM spindle speed, 150 RPM platen speed, and 2 lb. (8.9 N) down force. (a): Time series plot showing 8 sec of the total (180 sec) data, time period between prominent beat-like low frequency components is ~ 2 sec (b): FFT frequency spectrum of the data, showing presence of 4 main frequency bands (see Sec. 3.4). Low frequency (2 sec) region seen in (a) is shown in inset.



Graphic IX: Challenges in CMP process

3.2 Review of the Relevant Literature

Vibrations from the CMP process result from a dynamic interplay among mechanisms taking place at three different scales, namely wafer-pad asperity [138, 139, 144], bulk pad structure [145, 146], and machine kinematics. Much of the literature has focused on deriving analytical models of the CMP process at pad asperity and bulk scales [138-141, 143-149].

For example, Luo, *et al.* [138] and Wang, *et al.* [139] proposed models incorporating pad asperity effects for predicting material removal rate (MRR) in CMP. They showed the influence of pad topography and slurry particle size distribution on the contact load at the wafer-pad interface, and as a consequence the MRR. Significant work has been done to model the effects of slurry agglomeration [147], chemical action [150], pad porosity [151], and hydrodynamic effects of the slurry [142, 152] on forces and MRR in CMP.

At the bulk pad level, Bastawros, *et al.* [145] invoked the effect due to elastic pad bending to delineate the different contact modes at the wafer-pad interface. Fu and Chandra [146] considered the effect of viscoelastic pad behavior to model within-wafer-nonuniformity. Bajaj, *et al.* [153] correlated material properties of the polishing pad, such as shear modulus, asperity density, etc. with MRR. They explained the deterioration in MRR over time because of pad degradation at the asperity and bulk levels.

Borucki [149] invoked pad degradation effects, such as thinning, and asperity wear to model MRR. While CMP pads are essentially multi-material structures

[20], the heterogeneous characteristics have not been considered in earlier models. Lu, *et al.* [154] considered heterogeneity of pad structure and showed via experiments that multi-material composite stacked pads can depict complex dynamics. Stavreva, *et al.* [155] conducted polishing experiments using both stacked and un-stacked pads, their results indicated that composite stacked pads offered significant improvement in MRR and uniformity characteristics compared to un-stacked pads.

Integration of models capturing pad-asperity effects with those addressing pad structural non-uniformity may be necessary to delineate the physical sources of the spatio-temporal patterns in vibration signals from CMP and thereby facilitate process monitoring.

From the process monitoring viewpoint, prior sensor-based works for CMP process monitoring have used vibration [112-115, 119, 121], thermal [134, 136, 156], friction (including AE) [122, 123, 125, 129], and fluid pressure [142, 152] measurements. A common theme in these efforts has been to relate statistical features from the sensor signals, such as vibration amplitude, signal RMS, power spectral density (PSD), etc., with CMP process conditions – typically endpoint and MRR.

For example, Carter, *et al.* [113] used a piezoelectric displacement sensor (Micro-epsilon S601-0,5) mounted on the wafer carrier shaft to implicitly measure the drag (friction) force at the wafer-pad interface, together with an infrared (IR) pyrometer (Mikron Infrared MI-N500) to monitor oxide CMP process. They

report close to 85% increase in PSD of displacement signals over a broadband 50 − 90 Hz frequency region as downforce is increased from 14 kPa (2 psi) to 42 kPa (6 psi), however, the trend is not linear. They also noted continuous bands of active frequencies, and significant interaction among process parameters, which pose challenges for CMP process monitoring.

Hetherington, *et al.* [114] mounted piezoelectric accelerometers (Endevco 7259A-500) on a CMP polisher (IPEC 472) spindle head. Using a 2^{3-1} fractional factorial design of experiment (DoE) to vary wafer carrier (spindle) speed, platen speed, and downforce for polishing plasma enhanced chemical vapor deposition tetraethyl orthosilicate glass (PE-TOS) coated SiO_2 wafers, they noted that vibration signal PSD in the 800 − 900 Hz region increases from -50 dB to -30 dB as downforce is varied from 2 psi (14 kPa) to 10 psi (70 kPa), decreases from -30 dB at 20 sec of polish to -60 dB as polishing proceeds beyond 185 sec, and remains unaffected by changes in platen or wafer speed.

Jeong, *et al.'s* [112] multi-sensor monitoring system for CMP uses a tethered piezoelectric force sensor attached to the spindle head, a Hall Effect sensor to monitor the current drawn by the spindle servo motor (as a consequence of polishing load), and an AE sensor attached to the backside of the wafer. These sensors essentially record variations in friction force at the pad-wafer interface due to material layer transitions during polishing. They detected a transition point from copper to tantalum (barrier) layer during CMP, at which point the force signal energy over the 80 – 100 Hz range increased 8 fold, and AE RMS signals showed

a 3 fold increase at the material transition point. The Hall Effect sensor signal was found to be contaminated with extraneous sources, mainly motor impedance, and was therefore reticent in detecting these transitions.

Tang, *et al.* [125] mounted AE sensors on two different setups for polishing of spin-on glass, and PE-TOS wafer films. The AE RMS data showed large spikes whenever they induced defects on the wafer by using contaminated slurries (e.g., with 1 μm size diamond particles) and polishing pad having particulate residue. Ganesan, *et al*. [126] used wavelet coefficients of AE RMS signals to track the process drifts, such as delamination defects, within a statistical process control (SPC) framework for CMP of patterned Cu wafers.

Park, *et al.* [128] mounted tethered and wireless AE sensors on a CMP apparatus. They conducted experiments in order to compare AE signal characteristics between Cu and oxide CMP process. Due to relative softness of Cu, scratches were observed on Cu deposited wafers when polishing conditions were maintained identical to those for oxide polishing. Consequently, the intensity in the 210 – 250 kHz frequency range was approximately 3 times greater for Cu-CMP.

Chi, *et al.* [124] tracked the residual errors from the Kalman filter predictions of piezoelectric friction sensor signals to determine the polishing endpoint based on a predetermined statistical threshold. Allen, *et al.* [157] mounted probes on the polishing platen that concurrently induce and sense eddy currents in the rotating wafer. The polishing endpoint is estimated by correlating eddy current intensity with wafer thickness. Meloni's [133] endpoint detection method is based on

estimating the concentration of reaction species, such as hydroxyl ions, which are typical byproducts of the chemical action in Cu-CMP, by relating the concentration of relevant reaction species with wafer thickness.

Kojima, *et al.* [121] used band limited frequencies from piezoelectric vibration sensors signals to detect endpoint in Cu-CMP. They correlated the intensity of vibration signals in the 1.5 – 4 kHz region with different phases of Cu layer thickness. Yamada, *et al.* [120] used an *in situ* dual axes strain gage probe in contact with the polishing pad to detect polishing endpoint in Cu-CMP. Their system is similar to a pin-on-disk tribometer and is designed to detect transitions in pad coefficient of friction. For example, the coefficient of friction decreased gradually with Cu removal (due to deposition of polishing byproducts on the pad) and reached a minimum value (approximately 30%) near the endpoint.

A majority of the research in CMP monitoring relies on statistical, as opposed to physically motivated features for correlating sensor data with process conditions and outcomes. A PMI model that can relate the complex signal patterns with process mechanisms can facilitate extraction of signal features that are sensitive to variations in CMP process, and not just to signal patterns.

3.3 Overview of the Research Approach

Our approach to invoke and experimentally validate a PMI model of CMP is summarized in Figure 3-2. The PMI model combines the mechanics at the pad asperity level [149, 158] with the effects of the bulk pad material [153] and machine kinematics in the form of a deterministic two degree of freedom nonlinear differential equation model. The model solutions are corroborated with data acquired from experiments. The detailed formulation of the PMI model is documented in Appendix III.

One of the main objectives of this chapter is to verify the model-derived vibration $\ddot{x}$ with experimentally acquired wireless MEMS accelerometer data patterns.

The experiments were conducted on our Buehler CMP machine instrumented with wireless MEMS accelerometers. Process parameters, such as downforce (2 lb. (8.9 N) – 8 lb. (35.6 N)), spindle speed (30 RPM – 60 RPM), and pad condition were varied in these experiments. At each of the experimental conditions the PMI model parameterized with variables identical to the experimental conditions was validated using conventional time series techniques and frequency domain analysis [101], as well as, nonlinear invariants, such as recurrence quantification measures [159].

Next, the sensitivity of these PMI model-directed features for *in situ* monitoring of variations in process parameters (e.g. downforce, speed) and process conditions (here, pad wear) was assessed.

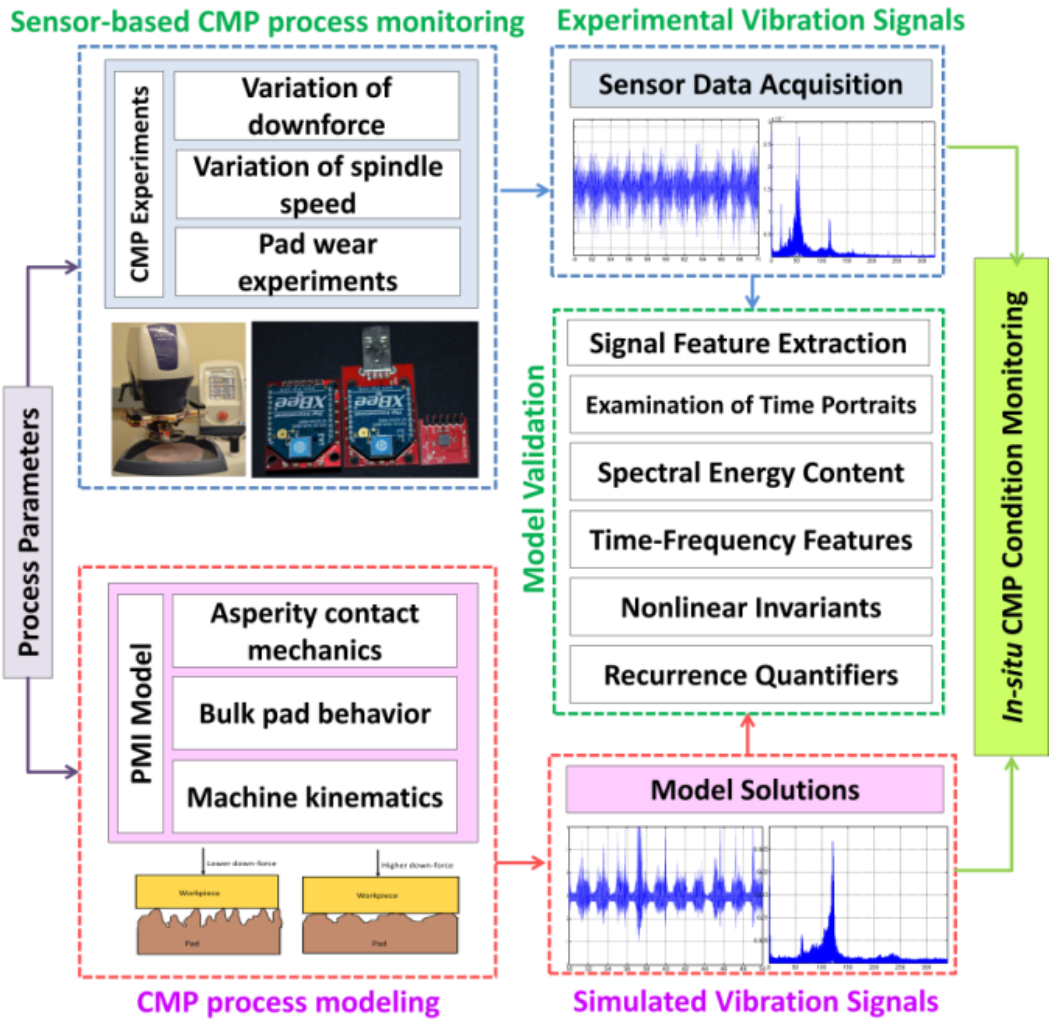


Figure 3-2: Schematic of the proposed approach for *in situ* condition monitoring in CMP.

3.4 Experimental Validation of the PMI Model

A Buehler (model Automet 250) metallographic polishing apparatus is instrumented with MEMS vibration (model ADXL 335 tri-axis accelerometer) and sound sensors (model ADMP 401 microphone) from Analog Devices [160]. Signals gathered from the accelerometer were employed for model validation. The accelerometer is capable of measuring vibration between ± 3g, and has a maximum sampling rate of 1600 Hz for each axis. The signals are sampled at ~ 685 Hz using a XBee (IEEE 802.15.4 Protocol RF module) unit with an onboard analog to digital converter and transmitted wirelessly to a desktop computer having a coupled XBee receiver unit (see Ref. [160]). We mainly use the signals gathered from the tangential (V_X) direction (X-Y plane with respect to the rotating spindle) vibration sensor. The entire wireless sensing platform rotates with a wafer carrier, and therefore travels through the error in the spindle head (Figure 3-3 (a) and (b)).

Cylindrical copper (free-machining C14500 series[14]) discs (wafers) of diameter 40.6250 mm ± 0.1 mm (1.625 in.), and thickness ~ 12.5 mm ± 2 mm (0.5 in.) are polished on this apparatus [62]. Scratch free, near-optical finish with Ra ~ 5 nm is reported in this work (Figure 3-3). Experiments were conducted in accordance by varying downforce between 2 lb. (8.9 N) and 8 lb. (35. 6 N), with spindle speed in the range of 30 RPM – 60 RPM, and platen speed fixed at 150 RPM. We used a KOH-based alkaline colloidal silica slurry supplied by Eminess Technolo-

[14] This particular copper series is 99.5% pure with tellurium (Te) as an alloying element. Te improves the machinability rating of copper but limits the surface finish that can be achieved.

gies (Ultra-Sol S17, particle size 70 nm, 10 pH) with slurry flow rate maintained constant at 20 ml/minute Typical CMP operations were conducted for duration of 9 minutes in three stages each lasting 3 minutes.

Since the initial surface roughness of the blanket copper wafers can significantly affect the acquired CMP vibration signal patterns, the wafers used for CMP experimental studies were first lapped to a surface finish in the range of Ra ~ 12.5 nm ± 2.5 nm. By tightly controlling the initial wafer surface roughness we ensure that the vibration signal patterns are consistent across experimental replications. The wafer surface is therefore considerably smoother than the pad surface and thereby satisfies one of the key assumptions of the GW Hertzian contact formulation [158] (Eqs. (15) through (17) in Appendix III).

Consequently, the vibrations resulting from the wafer surface morphology can be considered to be of negligible concern and dominated by vibrations from the pad asperities and bulk-pad structure

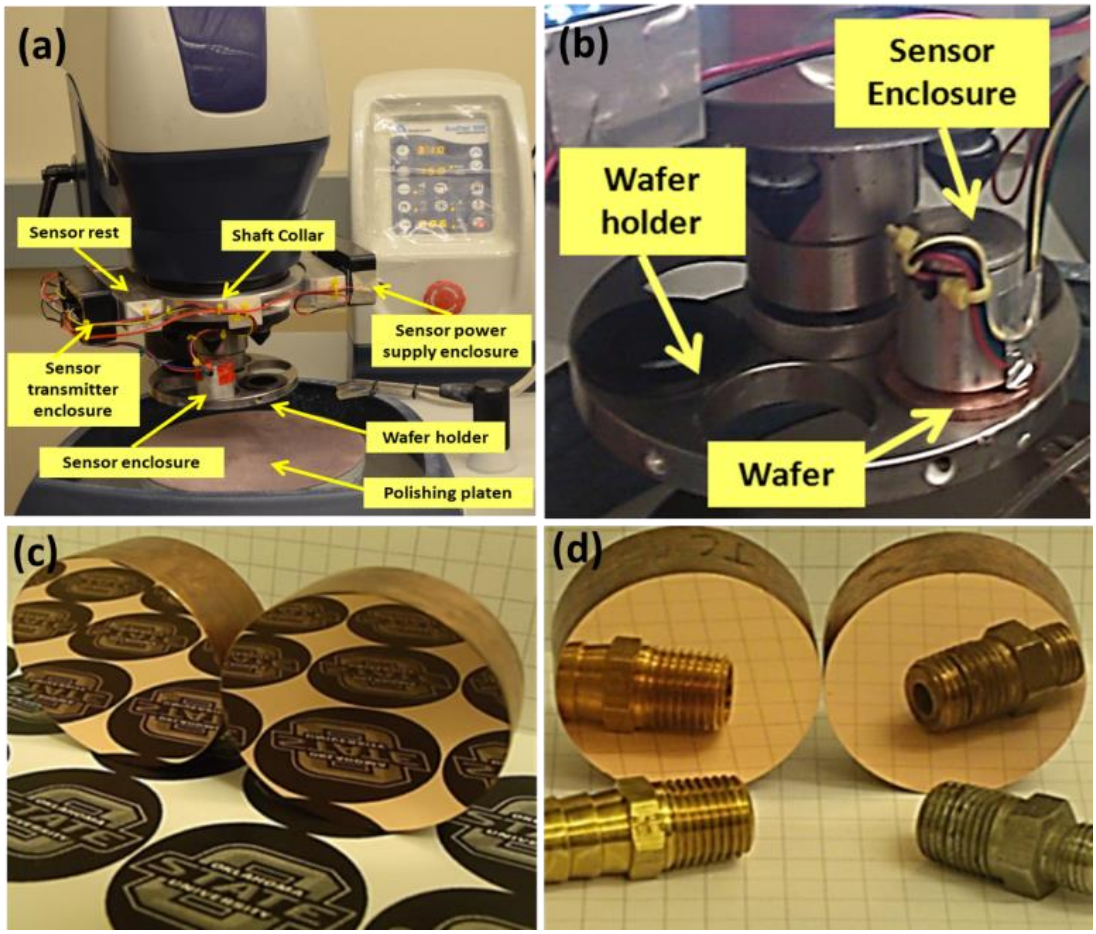


Figure 3-3: Blanket copper wafers after 12 minutes of CMP with 70 nm colloidal silica slurry.

(a) Buehler polishing apparatus instrumented with wireless vibration sensors. (b): Close-up view of the wireless vibration sensing setup. (c) and (d): blanket copper wafers after 12 min. of CMP with 70 nm colloidal silica slurry.

3.4.1 Examination of Time Portraits

We compared the time portraits of vibration signals obtained from experimental tests and the PMI model. Figure 3-4(a) shows a 6 sec long vibration signal, gathered at ~ 685 Hz sampling rate from CMP tests conducted at 2 lb. (8.9 N) downforce, 150 RPM platen speed, and 30 RPM (0.5 Hz) spindle (head) speed.

We note a characteristic low frequency pattern recurring every 2 seconds. This period corresponds to the spindle head speed. This low frequency pattern is

replicated in the simulated vibration signal shown in Figure 3-4(b). From these time portraits we also note that the high amplitude segments last for ~ 1.25 seconds.

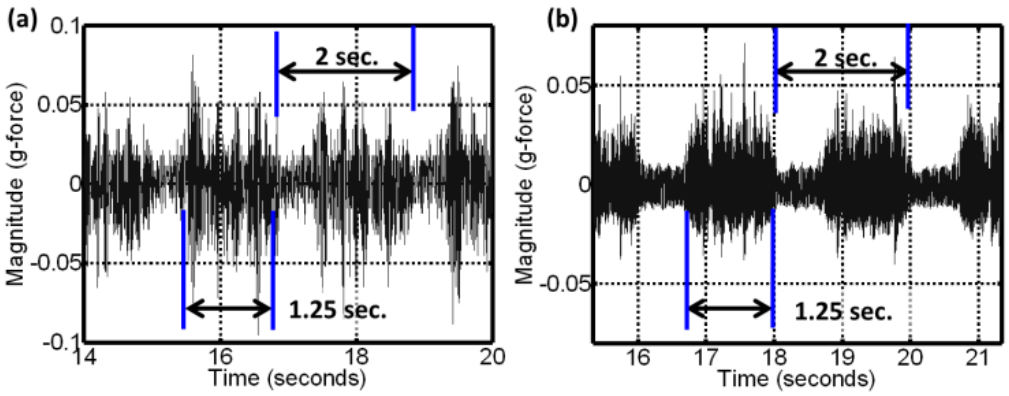


Figure 3-4: Representative time portraits of: (a) experimental and, (b) simulated vibration signals for CMP process.

(a) 6 sec long experimental vibration signal for CMP tests conducted at 2 lb. (8.9 N) downforce and 30 RPM (0.5 Hz) spindle speed, showing characteristic low frequency pattern occurring at 2 sec intervals (0.5 Hz), and high amplitude portion of the signal lasting for ~ 1.25 seconds.
(b) The corresponding simulated vibration time series, showing presence of low frequency pattern occurring at 2 sec intervals, and high amplitude portion lasting for ~ 1.25 seconds.

3.4.2 Frequency Domain Analysis of Vibration Signals

Evident in Figure 3-1(b) are four major frequency bands. The signal content in the 0.5 Hz – 1 Hz range is likely a result of spindle shaft eccentricity. The two broadband frequency regions centered around 25 Hz and 50 Hz are found to be a conjoined effect from sensor ambient characteristics, electromagnetic interference from machine elements, and extraneous vibration from the machine structure. The fourth, observed in the vicinity of 120 Hz was found to be sensitive to applied downforce and pad wear, and can therefore be useful for process monitoring [160].

Next, the effect of varying downforce on the frequency characteristics of the vibration signals was studied. Representative frequency portraits of experimental and simulated vibration signals are shown in Figure 3-5. Here the CMP tests were conducted under two different downforce conditions; 2 lb. (8.9 N, low downforce) and 8 lb. (35.6 N, high downforce). The spindle (head) and platen speed were maintained at 60 RPM and 150 RPM, respectively. We observe a 50% increase in the energy in 115 Hz − 120 Hz region when downforce is increased from 2 lb. (8.9 N) to 8 lb. (35.6 N). These tests were replicated 9 times for each downforce setting, and a statistically significant (p-val. < 0.01) increase in spectral energy content of (sum of squares of FFT magnitudes) 115 Hz – 120 Hz region with increasing downforce was observed.

Figure 3-5(b) and (d) show corresponding FFTs for simulation conditions imitating low (2 lb., 8.9 N), and high downforce (8 lb., 35.6 N) at 60 RPM spindle speed conditions. The model vibration data is in close agreement with the experimental data for the 115 Hz − 120 Hz frequency region. The magnitude of this region increases by roughly 40% when down force is increased from 2 lb. (8.9 N) to 8 lb. (35.6 N). The increase in magnitude of the 115 Hz − 120 Hz region is explained based on the PMI equations, which are documented in Appendix III. From Eqn. (14) it is evident that an increase in downforce P_0, leads to an increase in magnitude of $F_N(\nu)$. Consequently, F_X in Eqn. (8) also increases.

The observed increase in spectral energy, i.e., sum of squares of FFT magnitudes, is statistically corroborated using ANOVA (Table 3-1). We used a non-

overlapping moving window of length 3 sec to gather spectral energy values from simulated and experimental signals. ~ 30 sec of data each representing low (2 lb. (8.9 N)) and high (8 lb. (35.6 N)) downforce conditions were analyzed (since there are 3 replications, we have a total of 30 measurements for each condition).

As shown in Table 3-1, there is a significant difference in spectral energy contained in the 115 Hz − 120 Hz frequency band of both experimental as well as model-derived vibrations signals at the two downforce levels analyzed. The ANOVA regression R^2 (adj.) values are close to 75% for experimental, and 90% for the simulated cases.

Table 3-1: ANOVA results comparing spectral energy contained in 115 Hz − 120 Hz region at high and low down force conditions

Quantifier	**Simulated signal**	**Experimental signal**
Spectral energy of 115 Hz − 120 Hz region for low down-force (2 lb. (8.9 N)) vs. high downforce conditions (8 lb. (35.6 N))	Difference in mean: 20.77% p-val.: < 0.001	Difference in mean: 54.57% p – val.: < 0.001
Regression R^2	92.22% 91.73% (adj.)	78.24% 76.43% (adj.)

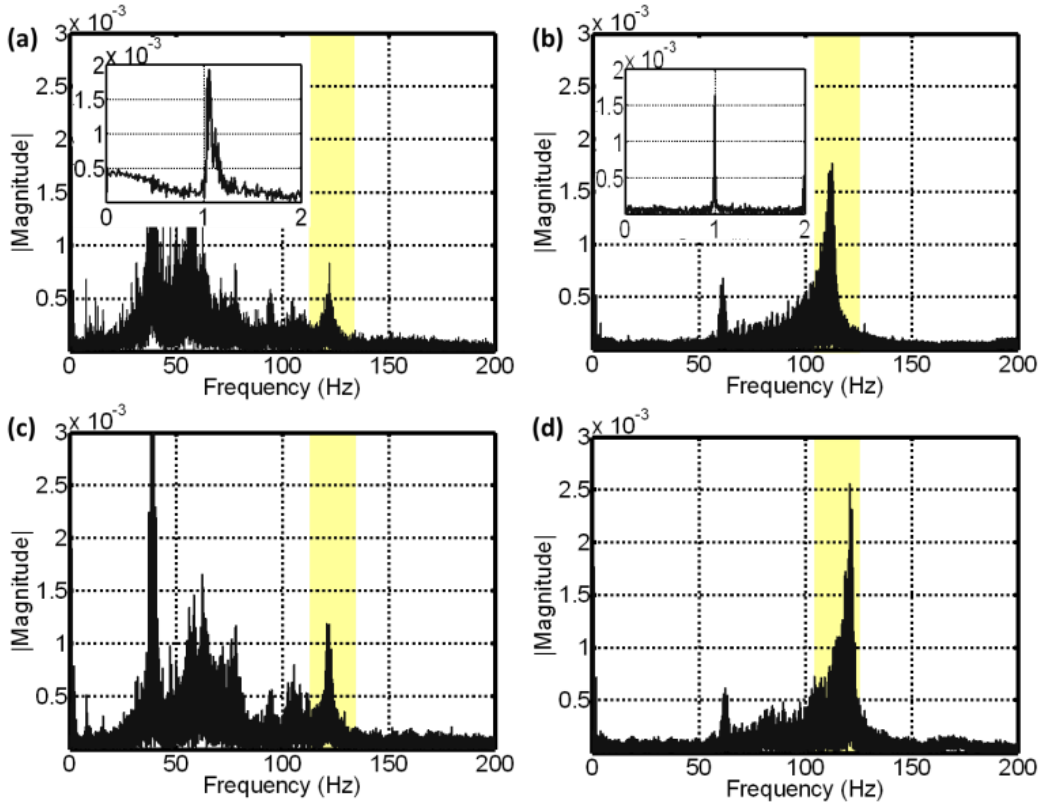


Figure 3-5: Representative frequency domain fast Fourier transform (FFT) portraits.

(a and c) experimental, and (b and d) simulated vibration signals for CMP process obtained under different downforce conditions (spindle speed identical, at 60 RPM, i.e., 1 Hz).

(a) FFT portrait of a ~ 180 sec long experimental vibration signal (685 Hz sampling rate) for CMP tests conducted at 2 lb. (8.9 N) downforce and 60 RPM (1 Hz) spindle speed, showing the different characteristic regions. Inset shows the 0 - 2 Hz region zoomed in, with peaks corresponding to spindle speed (1 Hz).
(b) The corresponding simulated vibration time series, showing presence of low frequency peaks (inset) as in (a), and 115 Hz − 120 Hz region prominently replicated.
(c) FFT portrait of a ~ 180 sec long experimental vibration signal for CMP tests conducted at 8 lb. (35.6 N) downforce and 60 RPM (1 Hz) spindle speed. 115 Hz − 120 Hz region shows an increase of ~ 50% compared to (a).
(d) Simulated vibration signal corresponding to (c). 115 Hz − 120 Hz region shows an increase of ~ 40% compared to (b).

3.4.3 Time-frequency Analysis

Figure 3-6 compares the time-frequency *spectrogram* portraits of the experimental (Figure 3-6(a)) and simulated (Figure 3-6(b)) vibration signals. These plots were obtained by taking a short-time Fourier transform (STFT) of the time

portrait with an overlapping moving window of length 0.5 sec (345 data points), with an overlap of 0.0125 sec (8 data points). The time-evolution is shown along the abscissa, while the frequency in Hz is plotted along the ordinate axis. The highest magnitude portions of the STFT are colored red, and the lowest magnitude portions take a dark blue hue. Thus, the spectrogram plot allows for visualization of the signal in both time and frequency domains.

For spectrogram plot of a vibration signal obtained from CMP tests at 2 lb. (8.9 N) downforce, 150 RPM platen speed, and 30 RPM spindle speed (Figure 3-6(a)), two distinct regions can be discerned − (i) region marked in black corresponds to the high amplitude portions of the time series, and (ii) region marked in green, corresponds to the low amplitude portions.

We notice that the high amplitude portions (dark red hue) appear at every 2 sec intervals corresponding to the spindle speed and last for ~ 1.25 sec (Figure 3-4(a)). One also notes the prominent presence of significant signal components in 25, 50 Hz and 120 Hz bands, along with another component in 240 Hz region. The spectrogram for the equivalent simulated vibration signal is shown in Figure 3-6(b). Comparison between the experimental and simulated vibration signals attests to the ability of the model in capturing the CMP process dynamics, both in the time and frequency domains.

We compare the length of the low amplitude portions of the time series for different downforce conditions – 2 lb. (8.9 N) and 8 lb. (35.6 N). Visual examination of the 120 Hz region of the spectrogram plots indicated that the length of the

low amplitude region is shorter for the high downforce case. The interval lengths were determined based on a statistical clustering methods to segment signals based on dynamic behavior [161]. Here, we used a recurrent Dirichlet classifier [162] to estimate the duration of the low amplitude portions.

Time duration information corresponding to the low amplitude portions is extracted for low (2 lb. (8.9 N)) and high downforce (8 lb. (35.6 N)) conditions (thirty measurements for each condition) from simulated as well as experimental data. The ANOVA results are summarized in Table 3-2. We note a statistically significant difference in the length of the low amplitude portion between low and high downforce conditions, with mean difference of 10.78% for the model vs. 16.31% for experimental data.

Table 3-2: ANOVA results comparing low amplitude portion length at high and low down force conditions

Quantifier	**Simulated Signal**	**Experimental signal**
Length of low amplitude segment for low (2 lb. (8.9 N)) and high downforce (8 lb. (35.6 N)) conditions	Difference in mean:10.78% p-val.: 0.003	Difference in mean: 16.31% p- val.: < 0.001
Regression R^2	15.26% 13.69% (adj.)	56.10% 55.35% (adj.)

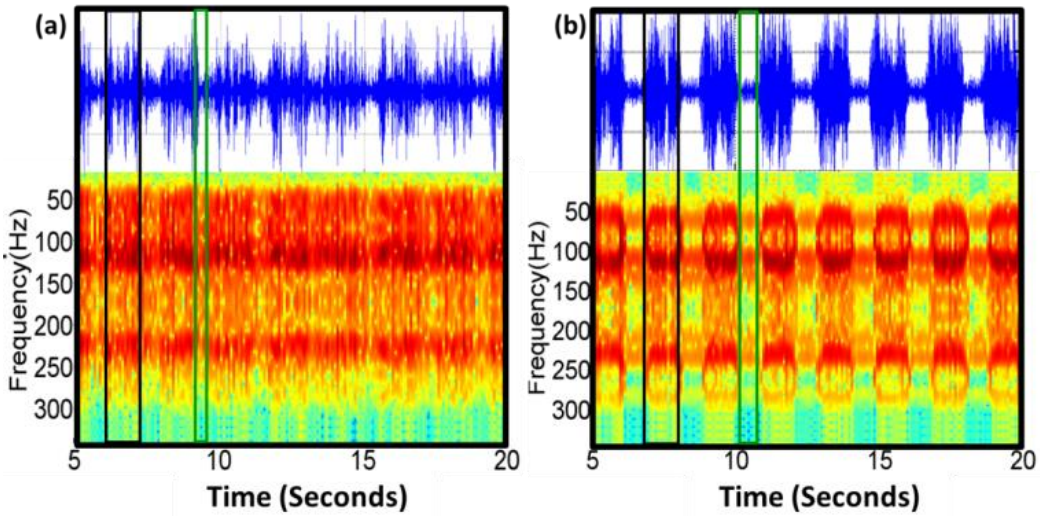


Figure 3-6: Representative time-Frequency domain analysis – spectrogram portraits.

(a) experimental, and (b) simulated vibration signals for CMP process obtained under identical conditions (2 lb. (8.9 N) downforce, spindle speed 30 RPM, i.e., 0.5 Hz).

(a) Spectrogram of a ~ 30 sec long experimental vibration signal (685 Hz sampling rate) for CMP showing two distinct regions: (i) high energy portions corresponding to high amplitude sections in the overlaid time series (marked in black), and (ii) low energy portions corresponding to the low amplitude sections of the time series (marked in green).
(b) The corresponding simulated vibration time series, showing similar characteristics.

3.4.4 Comparison of Nonlinear Dynamic Quantifiers

Next, we compared certain topological properties, such as recurrence [159] and space-time separation characteristics [101] of the state-space of the dynamics derived from experimental data vs. the PMI model. Application of a battery of tests suggested that the model correctly captures the dimensionality (m = 4) of the CMP process state space [88, 163]. The acquired vibration time series were embedded in a four dimensional state-space using a delay reconstruction procedure [101, 164].

The Euclidean distance of each data point in the reconstructed state space with every other data point is evaluated (i.e., pairwise measurements) and color mapped. The regions of higher magnitude are mapped with light (yellow-red) hue,

while points with relatively lesser pairwise Euclidean distance (i.e., neighbors in the state space) are marked with dark (black-blue) hue. Although the recurrence plot in Figure 3-7(a) for the experimental data is contaminated by extraneous noise, the overall characteristics (called *texture* in recurrence parlance) appears to be similar to the recurrence plot from simulated data (Figure 3-7(b)). The high energy regions, marked by yellow hues in the plots, seem to agree in temporal spacing, they occur at roughly 1 sec intervals, corresponding to the set spindle speed (60 RPM). Also, for both cases the high energy region persists for close to 0.60 seconds.

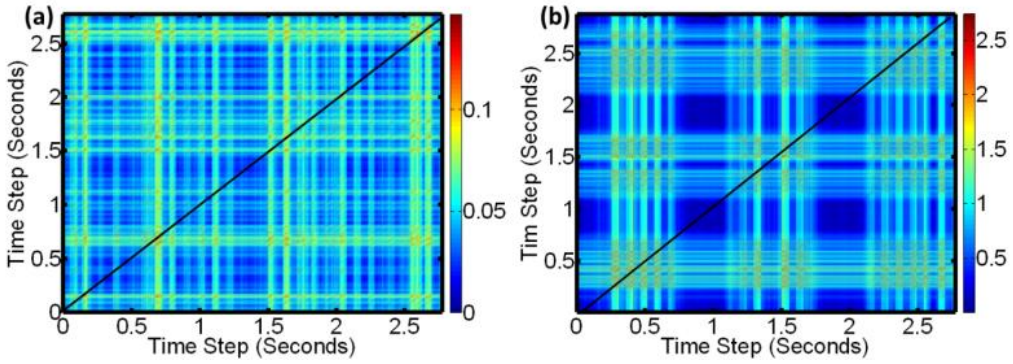


Figure 3-7: Representative un-thresholded recurrence plot.

(a) experimental and (b) simulated vibration signals with CMP data under identical conditions (down-force 2 lb, spindle speed 60 RPM) for ~ 3 sec long time series acquired at 685 Hz sampling rate. The yellow to red hues are considered as regions of high magnitude in time domain (i.e., high Euclidean distance in state-space).

(a) High magnitude patterns are ~ 1 sec apart (~ 685 data points) which corresponds to spindle speed (1 Hz), and lasts for ~ 0.6 sec Signal is relatively contaminated with noise, mostly emerging from 25, 50 Hz frequency regions.
(b) High magnitude patterns are also ~ 1 sec apart corresponding to spindle speed, and lasts for ~ 0.6 sec Simulated signal is relatively less contaminated with noise.

A space-time separation plot (Figure 3-8) depicts the distribution of recurrence behavior of the attractor topography in a manner analogous to the recurrence plot [101]. The plot captures for every point in the state space (denoted by

its time index), the neighborhood size (ε) around the point at which, a specified density of measure (roughly the specified fraction of points in the state space) is reached. In other words, we evaluate the size ε of the Euclidean ball, required to accommodate a fixed percentage of points in the state-space.

For example, consider a representative space-time separation plot in Figure 3-8(a), obtained from a CMP experiment conducted at low downforce (2 lb. (8.9 N)) and 60 RPM spindle speed. There are eight prominent lines in the plot; each of the eight lines uses a specific fraction between 0.60 and 0.95 with a step increment of 0.05. The bottom most line (cyan), has a fraction specified at 0.60, while the second (yellow) has 0.65, and so on.

In our case, comparison of the space-time separation plots for experimental (Figure 3-8(a)), and simulated (Figure 3-8(b)) vibration patterns, shows a close corroboration at various specified fractions. For example, we notice the recurrent periodic behavior for the 0.95 fraction (top most line, cyan) where the distance between successive periods is ~ 4 time steps in state-space. Similar corroborating patterns can be discerned for fraction size down to 0.75 (5^{th} line from top, black line).

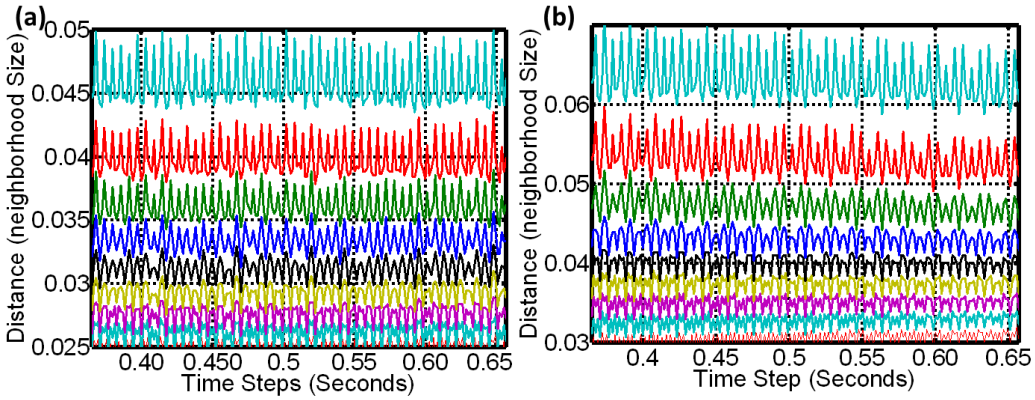


Figure 3-8: Representative space time separation plots.

(a) experimental, and (b) simulated vibration signals with CMP data under identical conditions (downforce 2 lb, spindle speed 60 RPM) for ~ 10 sec long time series acquired at 685 Hz sampling rate.

Next, we quantitatively compared the characteristics of the experimental and simulated vibration signal dynamics using features obtained from the recurrence plot [159]. Consider the *un-thresholded* recurrence plot shown in Figure 3-7, the Euclidian distance measured from these plots is now converted into a matrix of ones and zeros applying a Heaviside step function, i.e., if the Euclidean measurement is greater than a set fraction it is assigned a value 1, and 0 otherwise.

The appropriate threshold (neighborhood size) is selected based on the space-time separation plot (Figure 3-8). We selected a fraction of 0.75 (fifth line from top, black line), and note the corresponding average neighborhood size (ε, along the ordinate axis) for both cases. For the experimental case the neighborhood size, $\varepsilon = 0.03$, and for simulation $\varepsilon = 0.04$ are selected. These neighborhood sizes are used as the thresholding for the Heaviside step function.

A number of recurrence quantifiers discussed in Table 3-3 were extracted from the thresholded recurrence plot as suggested by Marwan, *et al.* [159]. First,

~ 7.25 sec long (5000 data points) non-overlapping sliding window is applied to the vibration signals obtained from both experiment and simulations. Recurrence measures were then extracted for each window. We apply this moving window 5 times.

The results are tabulated in Table 3-3. We observe from Table 3-3 that for 9 of the total 14 recurrence measures, the error between the experimental and simulated case is < 30%. This is noteworthy considering the noise content in the experimental data, and that the model is purely deterministic.

Table 3-3: Recurrence quantifiers for experimental and simulated signal for low downforce (2 lb. (8.9 N)), 60 RPM spindle speed, and 150 RPM platen speed CMP condition. There are total of five measurements (windows) for each recurrence measure, each totaling 5000 data points (~ 7 sec).

Recurrence Measure	From experiment		From simulation	
	Mean	Std. dev	Mean	Std. dev
Recurrence Rate	0.40	0.026	0 .30	0.039
Determinism	0.62	0.009	0.65	0.030
Determinism/Return Rate	1.54	0.099	2.17	0.197
Lamilarity	0.81	0.007	0.79	0.023
Lamilarity/Determinism	1.30	0.009	1.21	0.020
Maximal Diagonal Line Length	562	76.895	632	18.575
Mean Diagonal Line Length	23	0.870	55	9.518
Entropy of Diagonal Lengths	3.67	0.048	3.55	0.048
Divergence	0.001	0.0002	0.001	< 0.0001
Max Vertical Line Length	417	12.774	595	24.212
Trapping Time	6.43	0.272	13.02	1.573
Entropy of Vertical Line Lengths	1.76	0.041	1.70	0.062
Recurrence Times First Type	2.43	0.153	3.19	0.452
Recurrence Times Second Type	7.83	0.473	11.80	0.618

3.5 Condition Monitoring of CMP using PMI Model

The developed PMI model was shown to capture some of the salient dynamics of the CMP process. We now illustrate the application of the PMI model for condition monitoring applications in CMP. Apart from using spectral features, we used various recurrence measures gathered from the signal, as discussed in Sec. 3.4 (Table 3-3) to monitor the following three types of variations in CMP process:

(i) downforce is varied from 2 lb. (8.9 N) to 8 lb. (35.6 N),

(ii) polishing pad gradually deteriorates, and glazed portions become apparent, and

(iii) both downforce and pad condition vary.

3.5.1 Effect of Varying Downforce

We used experimental and simulated vibration data for two downforce conditions, 2 lb. (8.9 N, low downforce), and 8 lb. (35.6 N, high downforce). For each downforce condition 9 data sets, each representing 3 sec of vibration signals (30 sec total) are considered. From these data sets,16 different quantifiers including 14 recurrence measures, spectral energy content in 115Hz – 120 Hz range, and duration of low amplitude portions (see Sec. 3.4) were used as candidate features.

First, a best subset regression analysis was conducted to determine the subset of these quantifiers, which can differentiate between varying downforce conditions. Then, conventional linear regression models are constructed with the selected candidate features (identified from the best subset regression step) regressed on

downforce (dependent variable). We can thereby identify the common parsimonious feature set capable of explaining the process variation (due to varying downforce) for experimental and simulated signals.

The regression results are briefly summarized in Table 3-4. Compared to cases where only the signal spectral and time features were used (Table 3-1 and Table 3-2 of Sec. 3.4), a significant improvement in predictability on using recurrence-based features is observed. Furthermore, the simulated signals seem to capture over 95% (R^2) of the process variation, and can provide a means to anticipate process anomalies.

3.5.2 Effect of Pad Wear

The effect of pad wear has been experimentally studied by Bajaj, *et al.* [153]. They observed the evolution of morphology of polyurethane polishing pads to explain the decay in material removal rate with polishing over time. They observed blockage of pad pores in worn pads, which subsequently hinders slurry flow to wafer-pad interface. Byrne, *et al.* [165] observed that the polishing pad undergoes thinning overtime, with a worn pad being ~ 10% thinner than a fresh pad. They also note that while worn IC 1000 pads are more compliant than a fresh pad with the elastic modulus decreasing by ~ 13%, the pad hardness (shore D) for worn pad is nearly 3% higher.

Lu, *et al.* [166] observed significant changes in surface roughness, pore geometry and spectral properties of the polishing pad due to wear. In comparison to a fresh pad, the average surface roughness (Ra) of worn pads decreased to ~ 6.5

μm from ~ 8.5 μm, the pore geometry became more elongated along the direction of rotation with use, and worn pads showed almost two fold increase in infrared absorbance magnitude.

Models proposed by Wang, *et al.* [139, 140] and Borucki [149] incorporated the gradual degradation observed in the pad asperity distribution (mean and standard deviation of asperity heights decrease) to explain the decay in removal rate with time. The degradation of the pad asperity distribution affects the wafer-pad separation distance ($d_0(v)$ in Eqn. (18) of Appendix III). For worn pads $d_0(v)$ reduces in comparison to fresh pads. Borucki [149] estimated a decrease of close to 2 μm and 8 μm in d_0 after 8 minutes and 45 minutes of polish, respectively.

Table 3-4: Regression analysis for detection of variation in downforce in CMP

Factor	Statistical significance (p-val.)	
	Simulation	Experimental
Spectral energy of 115 Hz − 120 Hz region	0.00	0.0350
Entropy of vertical line lengths	0.002	0.282
Trapping time	0.012	0.0380
Segment length	0.634	0.0220
Regression R^2	97.5%, 96.8% (adj.)	91.4% 88.7% (adj.)

Lapped copper wafers (Ra ~ 10 nm − 15 nm) were polished on the CMP set-up in 3 minutes intervals with silica slurry. The polishing conditions were as follows: platen speed 250 RPM, head speed 60 RPM, and downforce 4 lb. (17.83 N) The platen speed is deliberately increased to this high value in order to accelerate pad wear. Such high platen speed is not advisable for long runs, since we observed a significant vibration of the machine rests and workbench. These extra-

neous vibrations manifested in a dominant peak at around 75 Hz in the frequency spectrum (Figure 3-10).

After 3 minutes of CMP the average wafer Ra improved to nearly 7 nm. Subsequently, pad wear was accelerated by soaking the pad in slurry for 45 minutes, followed by drying in air. By soaking and subsequent drying of the pad, the silica particles in the slurry tend to crystallize in the gaps between the asperities (pad glazing) [153].

This constrains the flow of slurry at the wafer pad interface, and deprives the wafer of adequate slurry. Secondly, by employing high relative velocity the hardened asperities (due to pad glazing) are easily sheared off (pad wear). As a consequence, we observed sheared pad material residue in the slurry reservoir.

The CMP process is carried out in 4 stages of 3 minutes each with the pad soaked and dried in the interim. After the end of 12 minutes of CMP, significant glazing of the polishing pad is observed (Figure 3-9(a)). In the same interval, prominent scratches were seen on the wafer (Figure 3-9(b)), and Ra increased to approx. 22 nm.

The FFT of the tangential direction (V_X) vibration sensor obtained after 3 minutes, and at the end of 12 minutes (when glazing of pad is observed) are compared in Figure 3-10(a) and (b) respectively. The magnitude in the 115 Hz − 120 Hz region increases by nearly 30 – 40% at the end of 12 minutes (Figure 3-10(b)) of CMP, indicating the effect of pad deterioration on vibration data.

To emulate the effect of pad wear in the PMI model, the static separation distance ($d_0(v)$ in Eqn.(18), Appendix III) was reduced by 10% from ~ 36 μm for the unused pad to ~ 33 μm for a moderately worn pad case [149, 165]. The results from the simulation are shown in Figure 3-11. We note an increase of 20 – 30% in magnitude of FFT for the worn pad case (Figure 3-11(b)) compared to fresh pad (Figure 3-11(a)).

Next, we proceeded to construct a linear regression model for data sets representing experimental, as well as, simulated vibration patterns. The experimental vibration patterns obtained from the first 3 minutes of CMP represent the fresh pad case, whereas the signals obtained between 9 to 12 minutes was chosen to represent the worn pad case. For each of the two representative data sets, we extracted 9 non-overlapping segments each measuring 20 seconds. As in the previous case, we computed 16 different quantifiers for each of these segments. Similar steps were taken with simulated vibration patterns from the PMI model.

We then combined the simulated and experimental data sets in order to compensate for the sparseness of available experimental data (see for example, Ref. [167]). Thus, we have 36 different data points representing fresh and worn pad cases (18 data points each).

We used integer indicator variables (1 and 2 for fresh pad vs. worn pad case, respectively) to differentiate between pad conditions. Similar to the previous case, a best subset regression analysis was first conducted to determine the subset of the candidate features which can differentiate between varying pad conditions. Then,

conventional linear regression models are constructed with these selected candidate features regressed on pad condition. The regression results are shown in Table 3-5.

Close to 75% of the variation in experimental vibration signals as a result of pad wear can be captured using only 6 features (30 degrees of freedom for residual error). The PMI model can thus help identify subtle drifts in process conditions from pad wear and can also compensate for sparse experimental data. This can enable early detection of incipient process anomalies with a parsimonious data set so that timely corrective action can be applied to prevent yield losses.

Table 3-5: Regression analysis for detection of pad wear in CMP

Factor	**Statistical significance (p-val.)**
Spectral energy of 115 Hz − 120 Hz region	0.688
Determinism	0.00
Lamilarity	0.00
Maximum diagonal line length	0.00
Maximum vertical line length	0.00
Recurrence times first type	0.00
Regression R^2	77.1% 72.3% (adj.)

3.5.3 Effect of Changing Downforce and Pad Condition

Anomaly causing conditions in CMP are oftentimes confounded, i.e., pad wear can occur at both high and low downforce conditions. In order to ascertain whether the PMI model can differentiate between such confounded states, we consider the four cases resulting from the combination of (i) low (2lb., 8.9 N) and high down force (8 lb., 35.6 N), and (ii) fresh pad vs. worn pad.

The four different process conditions were coded with integer indicator variables. For example, the ordinal 1 representing low down force – no pad wear, 2 high downforce – no pad wear, and so on. As before, we extracted 8 non-overlapping segments from each condition, giving a total of 32 (= 8 × 4) data points. We then computed the candidate features for each segment. Thereafter, the simulated and experimental data sets were combined (64 total data sets) and the candidate features regressed on integer coded process states.

Five features, as listed in Table 3-6, are found to be significant, with R^2 in the vicinity of 88%. We also note that the set of statistically significant features identified from regression analysis for each of the three cases illustrated in this section are largely unique (Table 3-4 − Table 3-6). This uniqueness of feature sets sensitive to different types of anomalies mitigates the possibility of confounding. For example, while spectral energy in 120 Hz range undergoes similar variation with increase in downforce as well as pad wear, the recurrence feature *lamilarity* is found to be significant for the pad wear case (Sec. 3.5.2), but not a relevant predictor of downforce (Sec. 3.5.1). In other words, based on the behavior of different features sets, a robust inference can be made regarding the type of anomaly.

Furthermore, by using the vibration patterns from the PMI model different types of anomalous conditions can be simulated offline. Consequently, the statistical feature set indicating the onset of such anomalies can be anticipated *a priori* from the simulated data. Therefore, instead of merely tracking a fixed set of statistical features, some of which may not necessarily be sensitive to process varia-

tions; the PMI model can be used as a means to select the most cogent features based on understanding the process dynamics.

Table 3-6: Regression analysis combining both simulated and experimental data

Factor	Statistical significance (p-val.)
Spectral energy of 115 Hz − 120 Hz region	0.019
Recurrence rate	0.000
Mean diagonal line length	0.004
Divergence	0.000
Trapping time	0.004
Regression R^2	88.8% 87.8% (adj.)

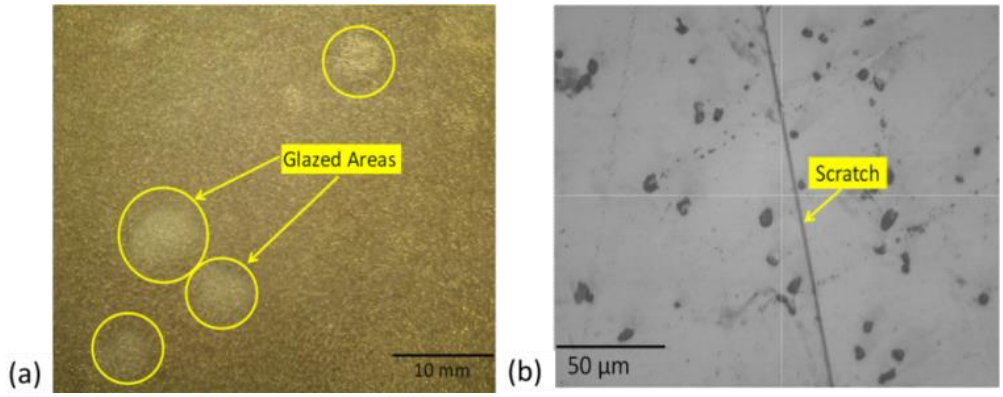


Figure 3-9: (a) Glazed pad after 12 minutes of CMP, and (b) Scratches on wafer observed at the end of 12 minutes of CMP.

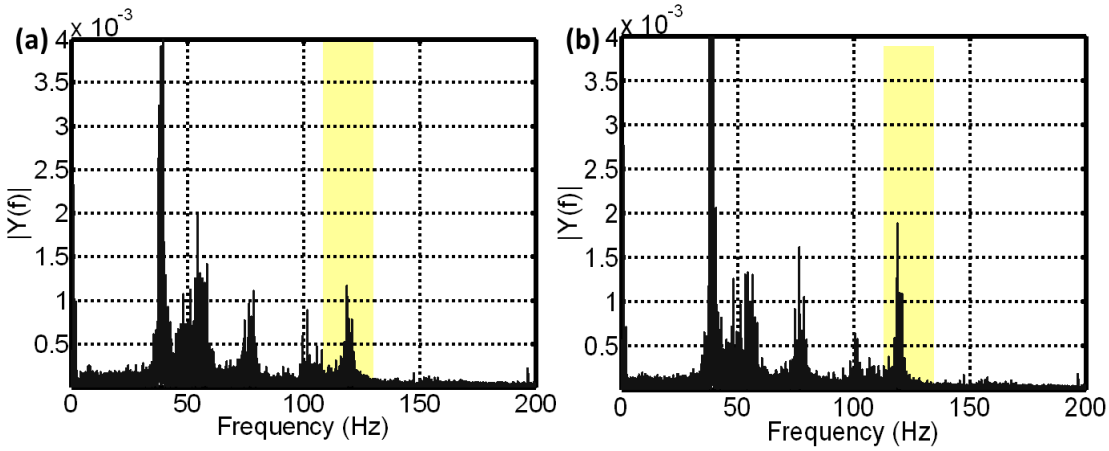


Figure 3-10: FFT of experimental vibration data obtained for (a) new pad vs. (b) 12 minutes used pad (glazed).

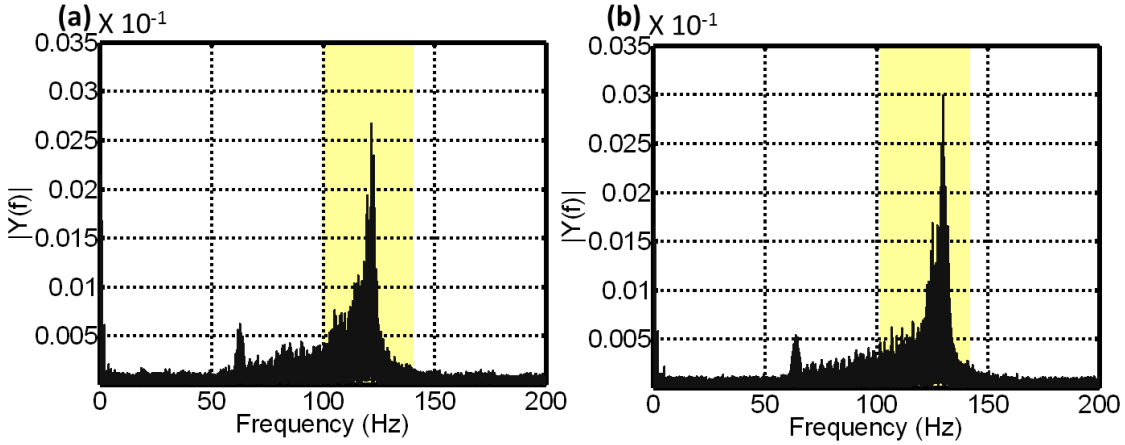


Figure 3-11: FFT of simulated vibration data obtained for: (a) new pad vs. (b) used pad.

3.6 Summary

We have forwarded a deterministic multi-scale process-machine interaction (PMI) model of CMP in the form of a two degree of freedom differential equation, which in combination with experimentally acquired vibration signals can help identify onset of process anomalies. Specific contributions of this work are as follows:

1. A deterministic two degree of freedom nonlinear differential equation process-machine interaction (PMI) model for CMP was formulated encompassing the responses at three different levels, namely: (i) pad-asperity, (ii) bulk pad structure, and (iii) machine kinematics levels. The model was validated on a CMP apparatus instrumented with multi-channel wireless vibration sensors. Despite being deterministic, the PMI model simulated vibration patterns closely emulated (R^2 ~ 90% for some cases) the signals obtained from CMP tests. Remarkably, apart from capturing the spectral aspects of the measured

vibration signals, the PMI model solutions was also able to replicate complex time-frequency and nonlinear topographical aspects of experimentally acquired vibration signals.

2. The PMI model solutions were used off-line to simulate different types of process drifts. The resulting simulated signal patterns were analyzed *apriori* for identifying the appropriate statistical feature set responsive to process variations. The features so extracted are closely related to physical changes in the process as opposed to mere signal statistics. Consequently, signal features identified based on PMI simulated signal patterns were observed to capture the process variation and resulting anomalies with R^2 in the range of 80 − 90%.
3. The PMI model solutions, due to their close dynamic similarity with measured vibration signals, can augment the sparse experimental data typical to CMP. For example, in case of process defects such as pad wear, availability of prior observations can be rare or at best evanescent. Under such conditions, the features extracted from the PMI simulated signals can be used as surrogates to experimental data.

It may also be noted that for reasons of tractability, the CMP experiments in this work were conducted on blanket copper (dia. 40.625 mm, thickness 12.5 mm) wafers, whereas the wafers used in industry are both significantly larger and thinner (dia. 300 mm, thickness < 1 mm), and composed of multi-material phases.

The effect of these different wafer dimensions on the vibration signal patterns may be non-negligible.

For example, we have observed that the raw magnitude of vibration signals in the 115 − 120 Hz region is close to 60-70% lower when 25 mm (1 in.) dia. copper wafers were used for preliminary tests. With larger diameter wafers used in industry, we contend that the vibration signal patterns will be dependent on the radial location of the sensors. In order to isolate and thereby model the effect of sensor location, several sensors can be mounted at different radial locations on the wafer carrier in close proximity to the substrate. We are currently investigating Bayesian-based analytical approaches for data fusion from multiple sensors.

Also, since industry uses a multi-step CMP approach for copper interconnects, it is reasonable to anticipate a change in the vibration signal patterns as the process evolves from blanket copper removal phase to copper clearing and barrier removal stages. It is therefore expected that additional experimental efforts may be necessary towards extending the concepts presented in this work to an industrial semiconductor production scenario.

These practical challenges notwithstanding, we contend that the overall approach of modeling the various multi-scale PMI phenomena in CMP and subsequently integrating the model with observed signal patterns, as presented in this work, can be valuable from a quality assurance perspective.

4 A Graph Theoretic Approach for Quantification of Ultraprecision Surface Morphology

We present an algebraic graph theoretic approach for quantification of nanoscale ultraprecision surface morphology. Two complementary methods to realize graph theoretic representation and subsequent quantification of nanoscale surface morphology variations are reported.

Investigations with CMP processed copper wafers suggest that the graph-based topological invariant Fiedler number (λ_2) was able to quantify and track variations in surface morphology more effectively compared to other quantifiers reported in literature. Using this approach, heterogeneous, multi-scaled aspects of nano-surface morphology can be captured from optical micrographs as opposed to reticent profile mapping techniques, and can therefore be valuable for *in situ* real-time assessment of surface quality.

Goal

Develop an approach that can quantify ultraprecision surface morphologies in a rapid and parsimonious manner.

4.1 Motivation

"But there is no easy way of predicting how summits join together to form ridges, or how valleys link up. In geographical terms, it is as if we knew the average area of a valley, but had no way of finding the easiest pass to the next valley, or the most direct route through a range of hills by means of interconnecting valleys."

– T.R. Thomas, "Trends in surface roughness," *International Journal of Machine Tools and Manufacture*, (**38)**5–6, pp. 405-411, 1998.

Modern semiconductor microelectronics consist of nanoscale dielectric/metal interconnect layers [20, 21]. The dimensional and topographical integrity of these interconnect layers is a critical determinant of device performance, and is therefore tightly controlled during fabrication [20]. In semiconductor manufacturing, layer topography specifications are met using multi-step chemical mechanical planarization (CMP) process [19]. Unimpeded CMP induced deviations in interconnect layer topography dimensions can compromise device functionality and cause high yield loss [19-21]. Hence, monitoring and characterization of CMP processed surfaces is vital for quality assurance of semiconductor devices [19, 20, 168]. In this context, effective wafer metrology can help prevent propagation of catastrophic defects over long production runs, and thereby minimize yield loss.

The semiconductor industry has prioritized development of responsive surface characterization approaches, and spends an estimated $9 billion per annum on metrology related operations [24]. Several lacunae have been identified with current characterization approaches, such as atomic force microscopy (AFM), X-ray fluorescence, and mass spectrometry methods. These approaches are noted to be reticent, expensive, and destructive [26]. Some of the surface metrology related

challenges and recommendations motivated by Bukkapatnam, *et al.* are used as a segue for this chapter [27].

Challenges in Nanomanufacturing ([27])

Tools currently used in nanotechnology research labs offer atomic-level resolution for characterizing nanoscale surfaces but are barely adequate to meet the requirements of high-volume nanomanufacturing. For example, an Atomic Force Microscope (AFM) gives nearly atomic-level surface resolution, albeit at a very slow rate; it would be impossible to use AFM to characterize surfaces in commercial-scale high-rate operations.

Recommendations for Nanomanufacturing Research ([27])

Instrumentation and analytical tools for comprehensive characterization of nanomaterials for on-line process control should be developed and characteristics required for specific applications defined.

Experts have noted that statistical quantifiers, such as arithmetic mean roughness (Ra, Sa), root mean square roughness (Rq, Sq), etc., used for quantification of conventional surface morphologies are not amenable at the nanoscale [6, 8, 9]. For example, consider Figure 4-1, which shows morphological evolutions of a copper (Cu) wafer surface over CMP processing intervals. The surface morphology changes are captured using a profile scanning laser interferometer (Figure 4-1(a) and (b)) and optical microscope (Figure 4-1(c) and (d)).

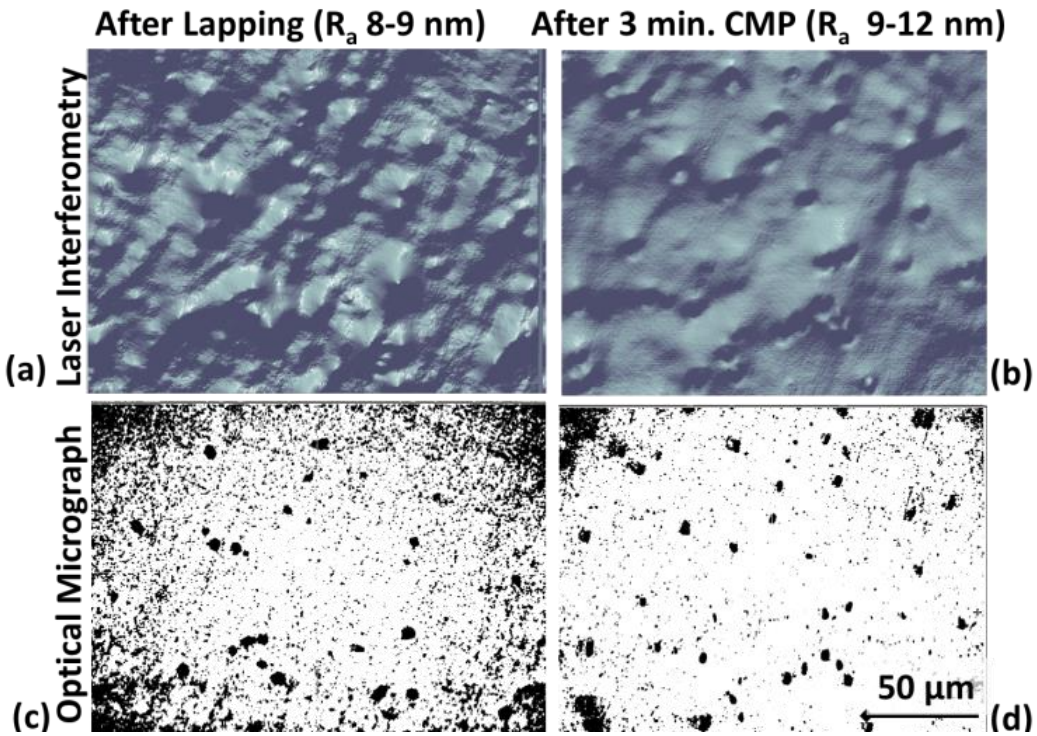
The surface quality of the Cu wafer visibly improved within a short CMP processing interval of 3 minutes, and the ridged texture of the lapped surface (Figure 4-1(a) and (c)) was polished to a distinctly smoother topography with near-specular surface finish with CMP (Figure 4-1(b) and (d)). However, the measured Sa values (Figure 4-1(e), and similar surface statistics such as Sq) failed to capture these visibly prominent changes. This intractability in quantification of surface morphology variations in Cu-CMP using conventional statistical parameters (such as Sa and Sq) has motivated us to pursue an alternative approach (to surface characterization).

Our approach aims to overcome these lacunae by invoking graph theoretic quantification of surface morphology using optical micrographs. This can be advantageous from an application standpoint, because, although optical microscopy is insensitive in detecting surface chemistry and sub-surface related defects (e.g., surface corrosion, sub-surface damage, and chemical etching), it is significantly faster compared to profile mapping (such as AFM, and laser interferometry) techniques [33, 169]. In industrial settings, semiconductor wafers (> 200 mm. dia.) are typically scanned (in-line) in less than 30 seconds using optical techniques. Additionally, optical inspection approaches are easier to integrate in-line, non-contact, and non-destructive [26, 170].

It is interesting to note that Thomas [29] recognized the need for a *connectivity* based surface characterization approach, since it can be a better exposition of the overall topography in terms of the relationship among peaks and valleys on a

surface. Our approach aims to provide such a *connectivity*-based perspective for quantification of nanoscale surface morphology using graph theoretic methods.

The rest of this chapter is structured as follows: the challenges for surface metrology, particularly at the nanoscale, are briefly reviewed in Sec. 4.2; the research approach summarized in Sec. 4.3; mathematics of graph theoretic representation for surface morphology explained in Sec. 4.4; followed by case studies demonstrating the potential of graph theoretic invariants for quantifying surface morphology variations in Sec. 4.5; and further verifications with near-optical CMP finished copper wafer surface measurements are provided in Sec. 4.6.



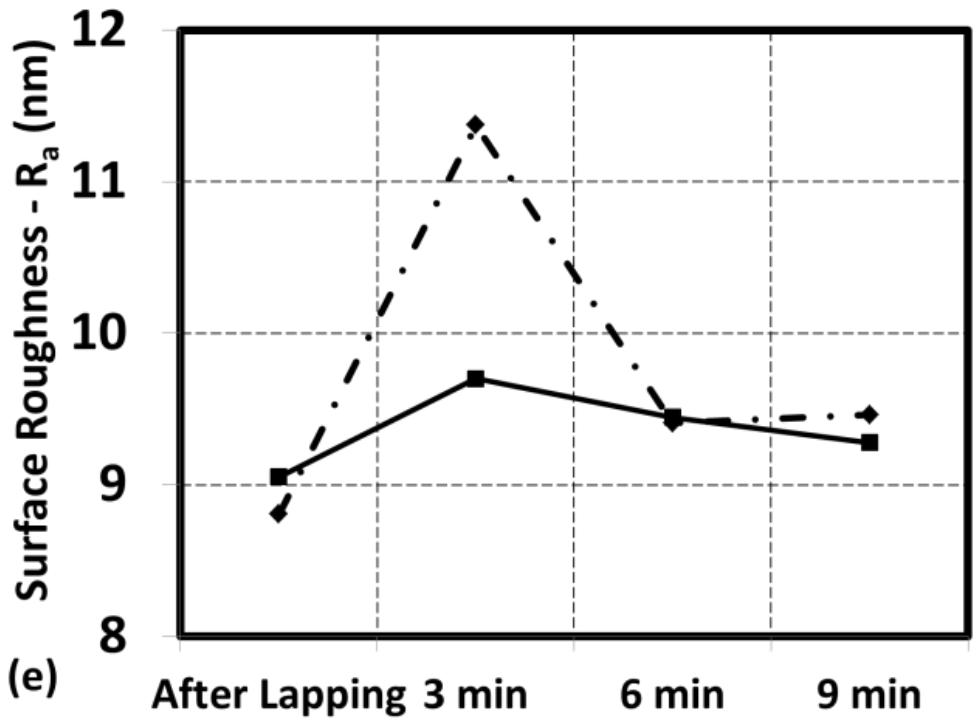


Figure 4-1: Surface morphology evolution of CMP processed Cu wafer.

(a): The surface profile obtained using a laser interferometer of a lapped wafer surface prior to CMP. (b): The surface profile obtained after 3 minutes of CMP. (c) – (d): The optical micrographs corresponding to surface profiles in (a) and (b) respectively. (e): The mean surface roughness (Ra) obtained across 3 intervals of CMP for two replications.

4.2 Challenges for Surface Metrology at the Nanoscale

Surface metrology of nanomanufactured semiconductor microelectronic components is challenging due to several factors, namely:

- **Reticence of measurement,** which constrains in-line use, for example, Jiang [6] notes that AFM-based topography mapping of a 2 mm diameter shell can take more than 2 hours.
- **Lack of quantification approaches for complex surfaces**, such as re-entrant surfaces, freeform splines, nanoscale ridged surfaces, etc., which relegates assessment of such morphologies to qualitative criteria [6, 9, 26, 56].

- Effect of **instrument parameters**, such as spatial resolution of the stylus (contact profilometers and AFM), scan length, sampling interval, digitizing filters, errors from transduction, drift, hysteresis, etc. [6, 9, 30]. For example, Poon and Bhushan [30] have recorded a difference of as much as 40% in surface measurement statistics (e.g., Ra) due to instrument factors.
- **Scale dependency,** which restricts measurements to small local areas**, and inability to characterize global morphology** [6], such as surface texture, with popular statistical quantifiers used for assessment of surface quality, e.g., arithmetic average (Ra, Sa) roughness, maximum peak-valley height (R_t, S_z), and root mean square (Rq,Sq) roughness.
- The last four decades have seen **profusion in parameters available (*parameter rash*)** for surface characterization [8, 54]. Several quantifiers have been standardized in an attempt to incorporate aspects of functional behavior of engineering components; for example, bearing ratio parameters for assessing seal integrity [8]. An unintended consequence of this profusion in parameters is the lack of coherence in surface metrology practice across industry [29, 171].

- Scale independent techniques such as fractal dimensions have been investigated in order to overcome some of these shortcomings [172-174]. Surface characterization using fractal dimensions involves mapping a statistic of the surface at various sampling lengths [173]. The gradient of the measured statistic over the sampling length (in logarithmic scale) is the evaluated fractal dimen-

sion. **A unifying statistical quantifier for measurement of fractal dimensions has not been suggested thus far**.

Moreover, **fractal dimensions can be ambiguous**, the difference between rough and polished surfaces is less than a few tenths [173, 175]. Whitehouse [176] notes that for parts manufactured using abrasive processes (such as grinding and polishing), **the fractal dimension could be an artifact** resulting from the stochastic nature of material removal (in abrasive finishing) [3].

Recently, multi-resolution wavelet-based approaches for characterizing surfaces at different length scales have been proposed [8, 55]. However, these newer methods require profile-based mapping (which is an off-line process), and extensive post-measurement analysis (which can be computationally demanding). Graphic X summarizes these challenges.

Challenges

1. Ultraprecision surfaces depict heterogeneous multi-scale aspects, which are not captured using traditional statistical quantifiers.
2. Traditional statistical parameters are not evocative of the underlying morphological relationships.
3. Characterization approaches, such as AFM, SEM, and laser interferometry are reticent, restricted to small areas, destructive, and expensive.

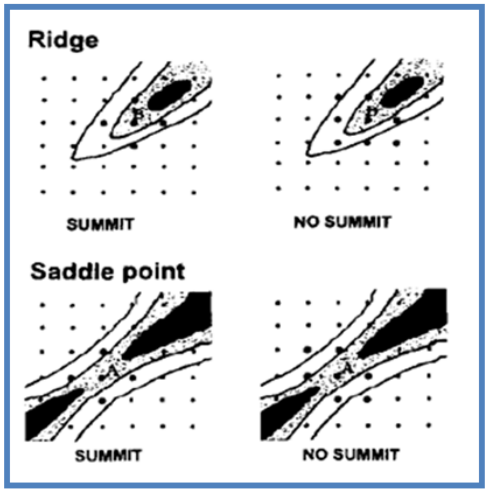
Statistical parameters are insufficient descriptors of surface morphology.

Statistical parameters occlude the underlying morphological relationships

- Cannot distinguish between patterned and randomly occurring morphological features
- Requires estimation of surface statistics over several length scales (Fractal mapping and wavelet decomposition)
- Need many quantifiers (parameter rash)

Characterization approaches are not well suited for *in situ* real time inspection

- Slow (AFM of 2 mm ball > 2.5 hr.)
- Only qualitative assessment of small local areas
- Destructive



*..But there is no easy way of predicting how summits join together to form ridges, or how valleys link up. In geographical terms, it is **as if we knew the average area of a valley, but had no way of finding the easiest pass to the next valley**, or the most direct route through a range of hills by means of interconnecting valleys* - T. R. Thomas, "Trends in Surface Roughness," *International Journal of Machine Tools and Manufacture,* **(38)**5–6, pp. 405-411, 1998.

Graphic X: Challenges in surface metrology with currently available approaches.

4.3 Summary of the Research Approach

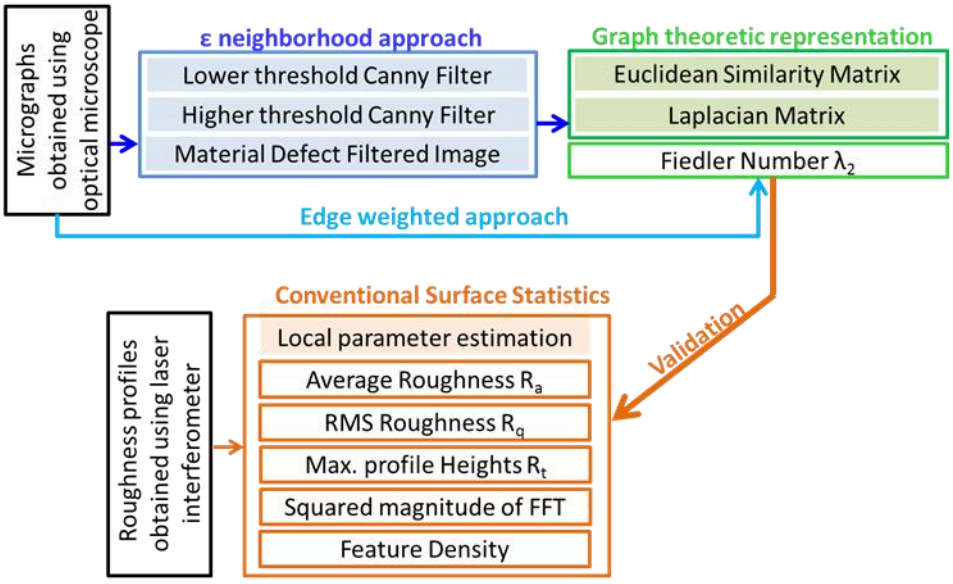


Figure 4-2: Summary of the research approach

Figure 4-2 summarizes the graph theoretic approach for nanoscale surface morphology quantification. Two methods are used for graph-based representation of surfaces from optical micrographs of the surface.

The first, called the *ε neighborhood approach* (Sec. 4.4.3), involves converting the gray scale micrographs into binary images using an edge detection technique (e.g., Canny filter) [177]. We note that the ε neighborhood representation is not loss-less, because it is contingent on (heuristically determined) filter parameters, and hence it is applied for cases where quantification of surface defects, such as nano-scratches is more pertinent.

Therefore, a second graph representation method, called *edge weighted approach* (Sec. 4.4.4) is also developed. This approach uses (gray scale) micrograph

images without the need for a filtering step and thus preserves surface texture information. Consequentially, the edge weighted graph representation is more suitable for assessing changes in surface texture.

We demonstrate the graph theoretic approach with artificially generated surfaces (Sec. 4.5). Using simulated surfaces we show that the graph topological theoretic invariant Fiedler number (λ_2) can be useful for quantifying different types of surface characteristics, and can overcome some of the drawbacks associated with traditional measurements. For reasons of completeness, we also investigate certain limiting scenarios (Sec. 4.5.6) where the Fiedler number may fail to capture surface morphology variations.

We validate the approach in Sec. 4.6 with CMP processed Cu wafers, where the graph-based topological invariant Fiedler number [178-180] serves as a discriminant for surface morphology variations. Subsequently, the trends observed in Fiedler number (λ_2) over different CMP intervals are verified against locally measured surface parameters, such as Sa and Sq (Sec. 4.6). The Fiedler number correlates with high fidelity when compared to conventional surface statistics; the correlation coefficient (ρ) estimated between $\sim 80 - 99$ %.

We note, these conventional measurements are scale-limited (they are measured over specific scales for verification), and require time intensive profile mapping and post-processing. The mathematical concepts for realizing a graph theoretic representation of surface morphology from optical micrographs are elucidated in Sec. 4.4.

4.4 Graph-theoretic Representation of Surface Morphology Variations

As noted previously (in Sec. 4.2), our current understanding of surface morphology appears to be restricted by the nature of statistics used for quantification [6]. Given the summits and valleys on a surface, we can estimate their statistical characteristics (Sa, Sq), higher moments of their probability distribution (skewness Ssk, kurtosis Sku, etc.), and spatial frequency (power spectral density) [8]. However, as it was recognized by Thomas [29], a description of the surface morphology in terms of the *connectedness* of summits and valleys is not forthcoming from these quantifiers, particularly (quoting Thomas) *"…how summits can join together to form ridges, or how valleys link up. (It) is as if we know the average area of a valley, but had no way of finding the easiest pass to the next valley…"* [29].

The essence of Thomas' argument can be rephrased in the following manner. Suppose we have complete areal measurements for a surface $\mathcal{I}$, i.e., we have mapped the peaks and valleys of the surface (e.g., using a laser interferometer). *How should we proceed to quantify the surface from these measurements?* The most accessible method would be to extract statistical moments, such as arithmetic mean (Sa), standard deviation (Sq), skewness (Ssq), kurtosis (Sku). However, by characterizing the surface using these global distribution parameters, we have occluded how the hills and valleys are connected.

In other words, the surface is not a single probability distribution, but a mixture of several distributions, i.e., the surface $\mathcal{I}$ is a convolution of several features $\mathcal{v}_k$ each occurring over a subspace $t_k\ (x, y)$, and having a probability distribution p_k. With this notation, we can mathematically represent $\mathcal{I}$ in the following form,

$$\begin{gathered}\mathcal{I} = \sum_{k=1}^{L} \mathcal{v}_k | t_k(x, y)\ \forall\ x, y \in \mathcal{I} \\ \mathrm{k} = \{1 \cdots L\}, \\ \mathcal{v}_k | t_k(x, y) \sim \mathrm{p}_k.\end{gathered} \tag{4-1}$$

The above equations interpret the surface $\mathcal{I}$ as a *stationary random field*. By characterizing the mixture distribution $\mathcal{I}$ with global distribution parameters, such as mean (Sa), standard deviation (Sq), etc., we implicitly assume that the random field $\mathcal{I}$ (i.e., the surface) has only one underlying distribution – one of the main impediments of statistical metrology, as pointed out by Thomas [29]. In contrast, the graph theoretic approach describes the surface morphology in terms of the *connectedness* among different features of the surface over varying scales. Using a graph-based approach allows us not only to represent the surface in terms of connectivity, but also enables quantification of topological characteristics (of the surface).

The main contribution of this work is in the graph theoretic representation and subsequent quantification of surfaces, based on treating such surfaces as a stationary mixture of random fields consisting of various heterogeneous features (e.g., pits, ridges, scratches; See Eqn. (4-1)) which allows topological invariants,

such as Fiedler number (λ_2) to effectively quantify and track variations in surface morphology.

The method has the following two key phases:

i. Representation of a surface as a network graph.

ii. Quantification of the topological aspects of the network graph (obtained for a surface).

The realization of these two aspects is formalized mathematically in the forthcoming section.

4.4.1 Representation of Surface Morphology as a Network Graph

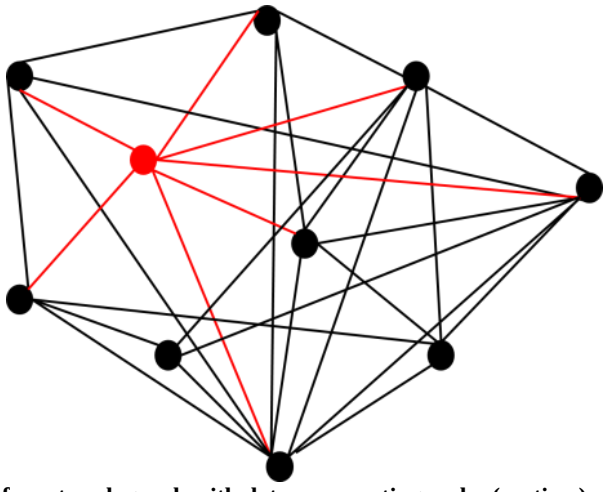


Figure 4-3: Illustration of a network graph with dots representing nodes (vertices) and lines as edges.

We now describe the method used to transform an image ($\mathcal{I}$) of a surface into a graph network ($G \equiv (V, E)$), i.e., achieve a mapping, $\mathcal{I} \mapsto G \equiv (V, E)$.

Let $\mathcal{I}^{\mathrm{M}\times\mathrm{N}}$ be the matrix representation of an optical micrograph (image taken with an optical microscope) of the surface acquired with a resolution of $\mathrm{M} \times \mathrm{N}$ pixels. The objective is to represent this surface as a network graph. For this

purpose, each row of $\mathcal{I}$ is considered to be the vertex (or node) of the *undirected* graph $G \equiv (V, E)$ with nodes (vertices) V and edges E [178-181] (an undirected graph is that in which the edges are bereft of any directional constraint, i.e., an edge is a two-way street). For example, in Figure 4-3 each dot represents a node and each line an edge, with movement possible in either direction along an edge. From the M rows of $\mathcal{I}$ we construct vectors $F_{(I)}$, $I = \{1,2, \cdots \mathrm{M}\}$ that are essentially row vectors of image pixel values[15].

We compute pairwise comparison metrics $\mathcal{w}_{ij}$ between $F_{(I)}$ using a kernel function Ω, and subsequently apply a threshold function Θ that converts $\mathcal{w}_{ij}$ into binary form. The above steps determine the topology of edge connections of the network $G \equiv (V, E)$ and are mathematically expressed in the following manner,

$$\mathcal{w}_{ij} = \Omega\big(F_i, F_j\big) \; \forall \; i, j \in I; \tag{4-2}$$

$$w_{ij} = \Theta\big(\mathcal{w}_{ij}\big), w_{ij} = \{0,1\}. \tag{4-3}$$

Using w_{ij} we construct the *similarity matrix* S (also called the adjacency matrix in graph theoretic context),

$$S \stackrel{\text{def}}{=} \big[w_{ij}\big]. \tag{4-4}$$

Since the graph is undirected, S is a (binary) symmetric matrix [182]. It enshrines the edge connections between nodes, i.e., if an edge exists between any two nodes i and j then w_{ij} = 1, else w_{ij} = 0. Essentially, the matrix S represents the *pruned* graph, such that only nodes satisfying the threshold condition set in

[15] Row-wise pixel comparisons are done in order to minimize computation load. Individual pixel comparisons will require $\frac{\mathrm{N}^2}{2} \times \frac{\mathrm{M}^2}{2}$ evaluations, only $\frac{\mathrm{M}^2}{2}$ computations are required with row-wise comparisons. Similar results are obtained using column vectors.

Eqn. (4-3) are connected with an edge. We note that these row-wise (column-wise) pixel comparisons are critical for facilitating analysis of homogeneous ultraprecision surfaces as they allow inclusion of features up to a length of N pixels (~ 200 μm when examining the CMP images in Sec. 4.6) as opposed to only local (< 0.20 μm, ~ 1 pixel) scales. Consequently, the computation load reduces from a bi-quadratic $\mathcal{O}(\mathrm{N}^2 \times \mathrm{M}^2)$ to quadratic $\mathcal{O}(\mathrm{M}^2)$ complexity. Furthermore, if individual pixel comparisons were to be computed, the resulting graph representation would then be replete with several nodes and edges, and consequently many of the large-scale variations in surface morphology would then be occluded.

Thus, by comparing the pixel rows (columns) of a surface, as opposed to individual pixels, we accomplish two objectives:

i. Measurements are over a global scale, as opposed to small local areas.

ii. Significant reduction in computation effort.

Thus, the graph $G \equiv (V, E)$ is a convolution of various morphological aspects of the surface, such as nano-scratches, pits, and ridges. As a result, the graph G captures the multi-scale aspects of surface morphology without the need to sift through different length (or areal) scales as typically required in wavelet decomposition and fractal mapping. At the end of this phase we have transformed the image $\mathcal{I}$ into a graph $G \equiv (V, E)$, i.e., $\mathcal{I} \mapsto G \equiv (V, E)$. Therefore, (strictly for the

purpose of analogy – we have not proved this duality claim mathematically) $G \equiv (V, E)$ can be considered a dual form of $\mathcal{I}$.

4.4.2 Quantification of Graph Network Topology

Once a surface has been represented in a graph theoretic form, we now proceed to quantify the topology of the resulting network $G \equiv (V, E)$. For this purpose, we first compute the diagonal *degree matrix* $\mathcal{D}$. The degree d_i of a node i is a count of the number of edges that are incident upon (connected to) that node. An isolated node will have degree 0. The degree d_i for node i is expressed as,

$$d_i = \sum_{j=1}^{j=\mathrm{M}} w_{ij} \;\; \forall \; i, j \in I \tag{4-5}$$

for the graph G, we formulate the degree $\mathcal{D}$ as a diagonal matrix,

$$\mathcal{D} \stackrel{\text{def}}{=} \begin{bmatrix} d_1 & \cdots 0 \cdots & 0 \\ 0 & \ldots d_k \ldots & 0 \\ 0 & \cdots 0 \cdots & d_{\mathrm{M}} \end{bmatrix}, \tag{4-6}$$

additionally, the *volume* $\mathcal{V}$ of the graph is given by,

$$\mathcal{V}(G) \stackrel{\text{def}}{=} \mathrm{tr}(\mathcal{D}), \tag{4-7}$$

with the degree $\mathcal{D}$ and similarity matrix S, we define the *combinatorial Laplacian* matrix [178, 180, 181, 183] as follows,

$$L \stackrel{\text{def}}{=} \mathcal{D} - S, \tag{4-8}$$

using the combinatorial Laplacian L and degree $\mathcal{D}$ matrices, the *normalized Laplacian* $\mathcal{L}$ [179, 180] is expressed as,

$$\mathcal{L} = \mathcal{D}^{-\frac{1}{2}} \times L \times \mathcal{D}^{-\frac{1}{2}}, \tag{4-9}$$

with the eigen spectrum for L and $\mathcal{L}$ computed,

$$Lu = \mu^* u$$
$$\mathcal{L}v = \lambda^* v. \tag{4-10}$$

We note, L and $\mathcal{L}$ have the following key properties:

i. They are singular, i.e.,$|L| = 0, |\mathcal{L}| = 0$, and diagonally dominant, i.e., $a_{ij} < a_{ii}\ \forall\ i \neq j$. This implies, they do not have full rank and are <u>not</u> invertible.

ii. They are symmetric M matrices (off-diagonal elements are negative or zero, $a_{ij} \leq 0\ \forall\ i \neq j$), also called *Stieltjes* matrices.

iii. They are positive semi-definite, i.e., $L, \mathcal{L} \succcurlyeq 0$.

iv. Therefore, their eigenvalues μ^*, λ^* are non-negative [179, 180],

$$0 = \mu_1 \leq \mu_2 \leq \mu_k \ldots \leq \mu_M$$
$$0 = \lambda_1 \leq \lambda_2 \leq \lambda_k \ldots \leq \lambda_M. \tag{4-11}$$

Additionally, it has been shown [184, 185] that the eigenvalues of $L, \mathcal{L}$ *interlace*,

$$0 = \lambda_1 = \mu_1 \leq \lambda_2 \leq \mu_2 \leq \lambda_3 \leq \mu_3 \ldots \lambda_k \leq \mu_k \leq \lambda_{k+1} \leq \mu_{k+1} \ldots \leq \mu_M \tag{4-12}$$

Both the combinatorial L and normalized Laplacian $\mathcal{L}$ share similar properties [179, 180]. For example, Butler [180] shows that λ^* and μ^* are related,

$$\frac{\mu_i}{\max(d_i)} \leq \lambda_i \leq \frac{\mu_i}{\min(d_i)} \tag{4-13}$$

In choosing between the combinatorial (L) and normalized Laplacian ($\mathcal{L}$) matrices, the latter (normalized Laplacian, $\mathcal{L}$) is preferred in recent applications of graph theory, especially in the area of image processing [182]. This is because the

eigenvalues (λ^*) of $\mathcal{L}$ are bounded between 0 and 2, i.e., $0 \leq \lambda^* \leq 2$ (the equality holds only for special cases), whereas μ^* is dependent on the maximum degree (see Eqn.(4-8) and (4-9)) [179, 180].

Therefore, λ^* is scaled based on the degree matrix $\mathcal{D}$. This property is pertinent from an application standpoint, because as a result of scaling based on $\mathcal{D}$, the effect of graph volume $\mathcal{V}(G)$, and thereby the image size $\mathcal{I}$ on λ^* are mitigated. This means, the eigen spectra of λ^* are not contingent on the size of the image. We also posit, based on empirical observations that λ^* is more robust to variations in brightness and contrast of surface micrographs compared to μ^*. Therefore, we primarily use the eigen spectrum λ^* of the normalized Laplacian $\mathcal{L}$ for quantification of surface morphology variations.

The smallest non-zero eigenvalue (λ_2) of λ^* is called the *Fiedler number*, and the corresponding vector (v_2) the *Fiedler vector* [178-180, 183]. The value of λ_2 equals 0 *if and only if* the graph is disconnected, i.e., there is an isolated node in the graph $G \equiv (V, E)$ that cannot be reached (this node will have degree = 0) [183]. As a result, the corresponding (to the isolated node) row in the similarity matrix S will have all elements equal to zero. Consequentially, S will degenerate into a defective matrix (i.e., it will not have a full set of eigen vectors).

Some of the well-known properties of $\mathcal{L}$ and its eigen spectra are listed on the next page [179-181, 185-188]. Subsequently, we outline the mathematics which shows that the Fiedler number (λ_2) is a useful measure for quantifying the *connectivity* of a graph network $G \equiv (V, E)$.

Some Properties of the Normalized Laplacian Matrix ($\mathcal{L}$)

$$\mathcal{L} = \mathcal{D}^{-\frac{1}{2}}(\mathcal{D} - \mathrm{S})\mathcal{D}^{-\frac{1}{2}} = \mathcal{D}^{-\frac{1}{2}}L\mathcal{D}^{-\frac{1}{2}}$$

1. Every graph has a unique Laplacian. The converse is not true.
2. $\mathcal{L}$ is a real symmetric matrix ($\mathcal{L} = \mathcal{L}^{\mathrm{T}}$).
3. $\mathcal{L}$ is positive semi-definite ($\mathcal{L} \succcurlyeq 0$).
4. All eigenvalues are non-negative ($\lambda_{\mathrm{i}} \geq 0$).
5. All eigenvalues are less than or equal to 2; the equality holds if and only if the graph is bipartite ($\lambda^{*} \leq 2$).
6. The first eigenvector is an identity vector $\mathcal{v}_1 = \mathrm{e} = 1$, the first eigenvalue is zero ($\lambda_1 = 0$).
7. The multiplicity of $\lambda_1 (= 0)$ as an eigenvalue is equal to the number of connected components in the graph.
8. The second eigenvalue is greater than zero if the graph is connected ($\lambda_2 > 0$) and zero ($\lambda_2 = 0$) if and only if the graph is disconnected.
9. The second eigenvalue is greater than 1 ($\lambda_2 > 1$) if an only if the graph is a complete graph (i.e., every node is connected to all other nodes).
10. The second eigenvalue is 1 ($\lambda_2 = 1$) if and only if the graph is bipartite.
11. The eigenvectors of $\mathcal{L}$ are orthonormal.

$$\mathcal{v}_1 \perp \mathcal{v}_2 \cdots \perp \cdots \mathcal{v}_n; \langle \mathcal{v}_i, \mathcal{v}_j \rangle = 0; \langle \mathcal{v}_i, \mathcal{v}_i \rangle = 1$$

this implies, the Gramian is an identity matrix $\mathrm{G}(\mathcal{L}) = \mathrm{I}$

hence, the transpose of the eigenvector spectrum is its inverse ($\mathcal{v}^T = \mathcal{v}^{-1}$).

A note for the reader

Graph theory is a well-developed (beginning with Euler's graph theoretic resolution (a negative result) of the *Bridges of Konigsberg Problem* in 1736) and extensive science. It appears (to this student at least) that there are two main areas of thrust in graph theory, (i) studies from a topological/combinatorial perspective, and (ii) the examination of the algebraic properties of graph networks (termed, *algebraic graph theory*). The former has been of interest to Industrial Engineers mainly from a logistic/OR background for some time, the latter, not as much (however, note the work of Hall [189]). This work is from the latter branch.

The novice in the area of algebraic graph theory must be prepared to face a rather steep mathematical learning curve, beginning with theorems in linear algebra (particularly those on eigen spaces). A graduate level textbook in linear algebra (e.g., [190]) may be useful in this regard. The following references in the area of algebraic graph theory were frequently referred by this student:

- The lecture notes by Spielman [186].
- Papers by Mohar [181], von Luxburg [182], and Spielman [187, 191].
- The master's thesis by Newman [188], and the doctoral dissertation by Butler [180].
- SIAM lecture series by F.R.K. Chung [179] (one of the definitive and highly cited references in this area).

The Fiedler number as a measure of connectivity of a graph network

Cheeger constant ($\mathrm{h}(G)$, also called the isoperimetric number or Cheeger number), and the Fiedler number (λ_2) are fundamental to the study of topological characteristics of graph networks [179-181]. They are often used to quantify the connectivity of a graph network. The larger the Cheeger number ($\mathrm{h}(G)$), the more resilient the network is to edge failures (viz. the severing of edges between nodes) [181]. This is evident from the following definition of $\mathrm{h}(G)$ [179].

Referring to the network $G \equiv (V, E)$ shown in Figure 4-3, consider a subset of nodes A (or sub-graph) of the graph $A \subset V(G)$ (e.g., the node colored red in Figure 4-3), it follows $V(G) - A = \bar{A}$ is the set of all nodes other than A in G (all the nodes colored black in Figure 4-3). With this notation $\mathrm{h}(G)$ is written [179] as,

$$\mathrm{h}(G) \stackrel{\text{def}}{=} \min\left\{\frac{|\partial(A)|}{\min\{\mathcal{V}(A), \mathcal{V}(\bar{A})\}}\right\} \tag{4-14}$$

$$\partial(A) \equiv \{(x, y) \forall x, y \in E(G) | x \in A,\ y \in \bar{A}\},$$

where, $\partial(A)$ denotes all edges connecting A with $\bar{A}$ (e.g., $|\partial(A)| = 6$ for the edges colored red in Figure 4-3 when the red colored node is the only node in the sub-graph A). Also, note that the denominator $\mathcal{V}$ is not the same as V, $\mathcal{V}(G)$ refers to the volume of the graph (Eqn. (4-7)), whereas $V(G)$ is the set of nodes or vertices in the graph. Cheeger number ($\mathrm{h}(G)$) is therefore a measure of *edge connectivity* of the network.

In other words, $\mathrm{h}(G)$ is a quantifier of the graph in terms of the most sparsely connected node, the severing of whose connections (edges) will break the graph in two (i.e., isolate a node). It is evident that when the density of *edge connections* is high (as in Figure 4-3), $\mathrm{h}(G)$ will be larger. This assertion was first proved by Fiedler [178], based on the smallest non-zero eigenvalue μ_2, and edge connectivity $\mathrm{e(G)}$, viz. analogous to $\mathrm{h}(G)$, obtained from the combinatorial Laplacian L for a graph with n nodes. Fiedler proved the following bound on μ_2,

$$\mu_2 \geq 2\mathrm{e(G)}\left(1 - \cos\left(\frac{\pi}{n}\right)\right). \tag{4-15}$$

The normalized Fiedler number λ_2 is related to $\mathrm{h}(G)$, by the following inequality [179], as shown by Chung [179].

$$\frac{[\mathrm{h}(G)]^2}{2} \leq \lambda_2 \leq 2\mathrm{h}(G) \tag{4-16}$$

From these bounds, it is evident that relatively high $\mathrm{h}(G)$ values (and as a consequence high λ_2 values) imply many edges must be removed in order to break the graph into disjointed chunks [188]. In contrast, a graph with low $\mathrm{h}(G)$ (and λ_2) is relatively easy to disrupt. By using the Fiedler number (λ_2) we measure the ease (the number of edges to be removed) by which a node can be isolated in a graph G, i.e., a measure of *the weakest links*.

We now rationalize the following concerns in a more mathematically rigorous manner:

i. How the Fiedler number (λ_2) could be used in a practical scenario.

ii. Why λ_2 is a good quantifier of surface morphology.

The Fiedler vector as an efficient means to partition a graph

Our approach, which uses the Fiedler number (λ_2) for measuring surface morphology, is inspired from the domain of graph-based image segmentation. The main objective in image segmentation (and clustering) is to partition an image into different parts (pixels) based on certain common features of interest, such as texture, brightness, contrast, etc. This is important in various applications, e.g., machine vision, medical imaging, object tracking – primarily areas in which the output signal/measurement is an image as opposed to a point statistic.

Consequently, algorithms that can accurately and efficiently partition images are an active area of research. Shi and Malik [192] describe a graph theoretic approach that uses the Fiedler vector (v_2) to partition images. In their paper, Shi and Malik [192] represent images in graph theoretic form and mathematically prove that the Fiedler vector (v_2) is an efficient basis for segmenting an image.

Though, our objective is similar in scope to the image segmentation problem of Shi and Malik [192], we wish to quantify the morphology of a (nanoscale) surface using graph theory. Notwithstanding this distinction, since segmenting a graph would first require mathematically (implicitly) quantifying the network topology, we can use the concepts described by Shi and Malik [192] to distinguish surfaces (from their optical images) based on the topology of their graph representations, instead of attempting to segment them (which is a more evolved problem).

However, certain modifications (described in the forthcoming sections; Sec. 4.4.3, Sec. 4.4.4) to the method described by Shi and Malik [192] are required to

make the graph theoretic approach tractable to our application area of nano-surface morphology quantification. Before we proceed, it is pertinent to acknowledge the work of Hall [189] (which is also referenced by Shi and Malik). In an article published over four decades ago (1970), Hall [189] mathematically outlines the properties of the Laplacian matrix (Hall however terms it the *disconnection* matrix). Hall's work is noteworthy in the following aspects:

i. Hall shows that properties of a graph can be mathematically quantified using the eigenvalues and vectors of the combinatorial Laplacian matrix L (termed *disconnection* matrix by Hall). Indeed, he uses an Euclidean measure to construct the adjacency matrix of the graph (which Hall terms as the *connection* matrix). The Euclidean measurement used by Hall translates to mapping the dissimilarity between nodes – a concept analogous to the similarity matrix S from equation Eqn. (4-4).

ii. Hall proves that the smallest non-zero eigenvector is an optimal solution to a very close (quadratic) form of the *mincut* problem.

iii. Hall applies the eigenvectors of the Laplacian for solving clustering problems, with several illustrated examples, e.g., on sequencing processes, clustering animals, placing electronic components in a grid. With these examples, Hall shows that the graph theoretic Laplacian eigen characteristics are perhaps the most efficient means to achieve a partition in r-dimension space.

We now describe the essential outlines of the proof given by Shi and Malik to show that the Laplacian eigenvector v_2 (Fiedler vector) is an efficient partition for a graph $G \equiv (V, E)$. Using the terminology established in the previous section, a graph $G \equiv (V, E)$ (e.g., Figure 4-3) is *cut* when we have been able to remove edges, such that, a node is isolated. The *cut* can be mathematically described as splitting the graph G into two parts (subgraphs), A and $\overline{A}$, by removing edges with weight w. This process is mathematically expressed as,

$$cut(A, \overline{A}) = \sum_{\mathrm{x} \in A,\, \mathrm{y} \in \overline{A}} w(\mathrm{x}, \mathrm{y}) \tag{4-17}$$

The objective (wrt image segmentation) is to minimize the *cut*, that is, find the RHS of Eqn. (4-14) in terms of the edge weights w, such that, the LHS has the least magnitude. This describes the essence of the graph cutting problem; viz. $\min_{w(\mathrm{x},\mathrm{y})} cut\left(A, \overline{A}\right)$, or as it is known, the *mincut* problem (we recognize that this is symbiotic to the definition of $\mathrm{h}(G)$ in Eqn. (4-14)).

However, it is observed that the *mincut* problem often degenerates, and produces a solution where only *one* node forms a cluster. This implies, instead of partitioning the graph into well-populated node clusters, the *mincut* criterion often converges to a solution in which there is only one node in a cluster. Therefore several image clusters, bereft of any useful common feature are formed – an inefficient outcome.

Therefore, as an alternative, a more constrained criterion is applied by Shi and Malik [192]. This criterion penalizes outcomes to the *mincut* problem with

sparse node clusters. They term this criterion as the *normalized cut* criterion ($Ncut$), and define it as follows,

$$Ncut(A,\overline{A}) = \frac{cut(A,\overline{A})}{assoc(A,V)} + \frac{cut(A,\overline{A})}{assoc(\overline{A},V)} \tag{4-18}$$

where,

$$assoc(\mathrm{A},\mathrm{V}) = \sum_{\substack{\mathrm{x}\in\mathrm{A},\\ \mathrm{y}\in\mathrm{V}(G)}} w(\mathrm{x},\mathrm{y}) \tag{4-19}$$

However, the solution to the $minNcut$ is NP-complete [192] (it is in essence a graph coloring problem, which is known to be NP-complete), and therefore not tractable. Herein lies one of the major accomplishments of Shi and Malik's paper [192]; they show that **the $\boldsymbol{minNcut}$ problem has an efficient discrete solution in the Rayleigh Quotient of the Laplacian matrix** i.e.,

$$\min\left(Ncut(A,\overline{A})\right) = \min_{\mathrm{x}} \frac{\mathrm{x}^{\mathrm{T}}\mathcal{L}\mathrm{x}}{\mathrm{x}^{\mathrm{T}}\mathrm{x}} \forall \mathrm{x} \perp e \tag{4-20}$$

From the properties of the Laplacian matrix $\mathcal{L}$ and its eigenvectors listed in Sec. 4.4.2 we notice the following:

i. the first eigenvector of the normalized Laplacian $\mathcal{L}$ is a unit vector $v_1 = \mathrm{e} = 1$, and

ii. the eigenvectors (of $\mathcal{L}$) form orthonormal basis, i.e., $v_1 \perp v_2 \cdots \perp \cdots v_n$; $\langle v_i, v_j \rangle = 0$; $\langle v_i, v_i \rangle = 1$.

These properties ensure the normality condition of Eqn. (4-20) is met. Consequently, on using the Courant-Fischer theorem (which gives min and max

bounds on the Rayleigh Quotient), Shi and Malik arrive at the following solution to the $Ncut$ problem,

$$\arg\min_{x_j} \frac{{x_j}^T \mathcal{L} x_j}{{x_j}^T x_j} = \mathcal{v}_2$$
$$\min_{x_j} \frac{{x_j}^T \mathcal{L} x_j}{{x_j}^T x_j} = \lambda_2 \tag{4-21}$$

Therefore, the **Fiedler vector ($\mathcal{v}_2$) solves the $minNcut$ problem, with Fiedler Number (λ_2) as the minimum attained**. This mathematically proves that Fiedler Number (λ_2) is not merely a statistic, but a *topological invariant*, which can therefore be useful as a discriminant for quantifying network of a graph network $G \equiv (V, E)$, and as an extension the image $\mathcal{I}$, since $\mathcal{I} \mapsto G \equiv (V, E)$. This is of important consequence to our application, because, a mathematically tenable link has been established between the Fiedler number (λ_2) and the surface morphology, via the graph $G \equiv (V, E)$ representation of the optical image $\mathcal{I}$ of the surface.

This is pertinent from our application standpoint, because using the Fiedler number, the morphological characteristics of a surface (from its optical image, $\mathcal{I}$) can be quantified.

The Kirchoff index as a measure of graph connectivity

Although, the Fiedler number (λ_2) and vector $\mathcal{v}_2$ are predominantly used in this work for quantifying surface morphology, it is natural to inquire whether the

higher order eigenvalues and vectors of the normalized Laplacian are of any consequence. In order to answer this question, we note the existence of another graph invariant called the *Kirchoff index*, which is estimated using the entire eigen spectrum of the Laplacian matrix [193].

The Kirchoff distance (K_f) for a graph with n nodes is written as,

$$K_f = n \sum_{k=2}^{k=n} \frac{1}{\mu_k} \tag{4-22}$$

where, μ_k are the non-zero eigenvalues of the Laplacian matrix L. Recently, Chen and Zhang [193] extended the metric to the normalized Laplacian $\mathcal{L}$,

$$\mathcal{K}_f = 2m \sum_{k=2}^{k=n} \frac{1}{\lambda_k} \tag{4-23}$$

with, m as the number of edges of an (undirected) graph, and λ_k the non-zero eigenvalue of the normalized Laplacian $\mathcal{L}$. It is easy to see,

$$\mathcal{K}_f = \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} S_{ij} \sum_{k=2}^{k=n} \frac{1}{\lambda_k} \tag{4-24}$$

Chen and Zhang [193] prove a relationship between K_f and $\mathcal{K}_f$, and show that these indices are evocative of topological characteristics of the network. In keeping with our terminology, we will call the former (K_f) as the *combinatorial* Kirchoff index, and the latter ($\mathcal{K}_f$) as the *normalized* Kirchoff index. Without delving into the mathematical details, we note it has been proved that the Kirchoff indices are also topological invariants of a network graph.

However, estimating the Kirchoff indices is computationally expensive, because the complete eigen spectrum (μ^*, λ^*) is required. Hence, we resort to the Kirchoff indices in only one instance in Sec. 4.5.6, which is a limiting case where the Fiedler number (λ_2) fails to discriminate between different surface morphologies. With this remark, we now detail the two complimentary approaches developed for representing nano-surface morphology in graph theoretic form.

4.4.3 ε Neighborhood Graph Representation

For this graph theoretic representation, the optical micrograph image ($\mathcal{I}$) of a surface is first converted into a binary image (using image filtering techniques, such as the Canny edge detection algorithm [177]). Then, the kernel function Ω (see Eqn. (4-2)) is defined [182] as follows,

$$\omega_{ij} = \Omega(F_i, F_j) = \left\|F_i - F_j\right\|^2 \forall\, i, j \in I,\, I = \{1,2,\cdots \mathrm{M}\} \tag{4-25}$$

we make the following observation regarding the bounds of ω_{ij},

$$\lim_{|F_i - F_j| \to 0} \omega_{ij} = 0$$

$$\lim_{|F_i - F_j| \to \infty} \omega_{ij} = \infty$$

$$\omega_{ij} = [0, \infty). \tag{4-26}$$

Subsequently, the threshold $\Theta(\omega_{ij})$ is applied,

$$\Theta\left(w_{ij}\right) = w_{ij} = \begin{cases} 1, & w_{ij} \geq r \\ 0, & w_{ij} < r \end{cases}$$

$$r = \frac{\sum_{i=1}^{\mathrm{M}} \sum_{j=1}^{\mathrm{M}} w_{ij}}{\mathrm{M}^2} \tag{4-27}$$

$$S \stackrel{\text{def}}{=} \left[w_{ij}\right].$$

These steps are further explained on the basis of Figure 4-4. We first compute the pairwise edge distance based on Eqn. (4-25) (Figure 4-4(a)), then edges with distance $< r$ are identified (colored blue in Figure 4-4(b)), and subsequently pruned (Figure 4-4(c)) using the threshold function in Eqn. (4-27).

Thereby, only those nodes, which are (topologically) farther away from each other are connected by an edge. This means, *dissimilar*[16] nodes are connected with an edge.

[16] The term similarity matrix is therefore a misnomer, in reality, S represents connections among dissimilar nodes.

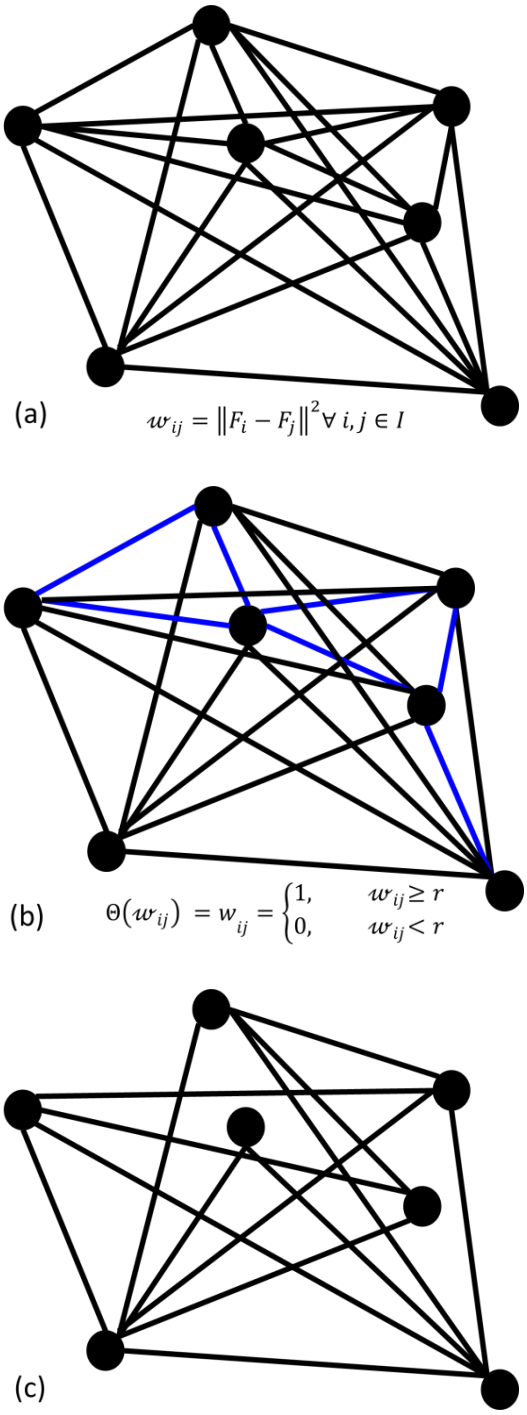


Figure 4-4: ε neighborhood graph representation.

(a) a well-connected representative graph (b) vertex distances identified based on the ε neighborhood (c) edges pruned using the distance threshold function (i.e., remove edges that are closer than ε).

4.4.4 Edge Weighted Graph Representation

In the previous graph representation, the nodes were connected on the basis of their topological closeness (measured using the Euclidean distance). In contrast, the edge weighted graph representation approach, the nodes are connected on the basis of their similarity (actually, dissimilarity) of texture. We compute the pairwise difference in texture between pixel rows with the kernel and threshold functions defined as follows [182],

$$\mathcal{w}_{ij} = \Omega(F_i, F_j) = e^{-\left(\frac{\|F_i - F_j\|}{\sigma_{\mathcal{I}}}\right)^2} \; \forall\, i, j \in I,\, I = \{1,2,\cdots \mathrm{M}\} \tag{4-28}$$

where $\sigma_{\mathcal{I}}$ is the standard deviation of (gray scale) pixel values of the image $\mathcal{I}$. Consequently, $\mathcal{w}_{ij}$ is bounded,

$$\begin{gathered} \lim_{|F_i - F_j| \to 0} \mathcal{w}_{ij} = 1 \\ \lim_{|F_i - F_j| \to \infty} \mathcal{w}_{ij} = 0 \\ \mathcal{w}_{ij} = (0, 1]. \end{gathered} \tag{4-29}$$

On applying threshold $\Theta(\mathcal{w}_{ij})$, $\mathcal{w}_{ij}$ is converted into binary form

$$\begin{gathered} \Theta(\mathcal{w}_{ij}) = w_{ij} = \begin{cases} 0, & \mathcal{w}_{ij} \geq r \\ 1, & \mathcal{w}_{ij} < r \end{cases} \\ r = \frac{\sum_{i=1}^{\mathrm{M}} \sum_{j=1}^{\mathrm{M}} \mathcal{w}_{ij}}{\mathrm{M}^2} \\ S \stackrel{\text{def}}{=} [w_{ij}]. \end{gathered} \tag{4-30}$$

The kernel function in Eqn. (4-28) is a Gaussian radial basis function, with the dispersion parameter set as the standard deviation of the image. These steps

are explained based on Figure 4-5. Figure 4-5(a) shows a well-connected graph, with different vertex colors representative of various node textures. The edges connecting nodes with similar textures are identified, (e.g., in Figure 4-5(b), a blue node connecting to a blue node), and is subsequently pruned (Figure 4-5(c)) using the threshold function of Eqn. (4-27). That is, edges are connected only between nodes with distinct textures. Since the nodes are connected based on texture *dissimilarity*, we use this approach (Sec. 4.6.3) to gain an insight into the rugosity aspects of a surface.

The salient features of the two approaches used for graph representation of micrograph images are summarized in Table 4-1. We now test the graph theoretic approach with artificially generated surfaces.

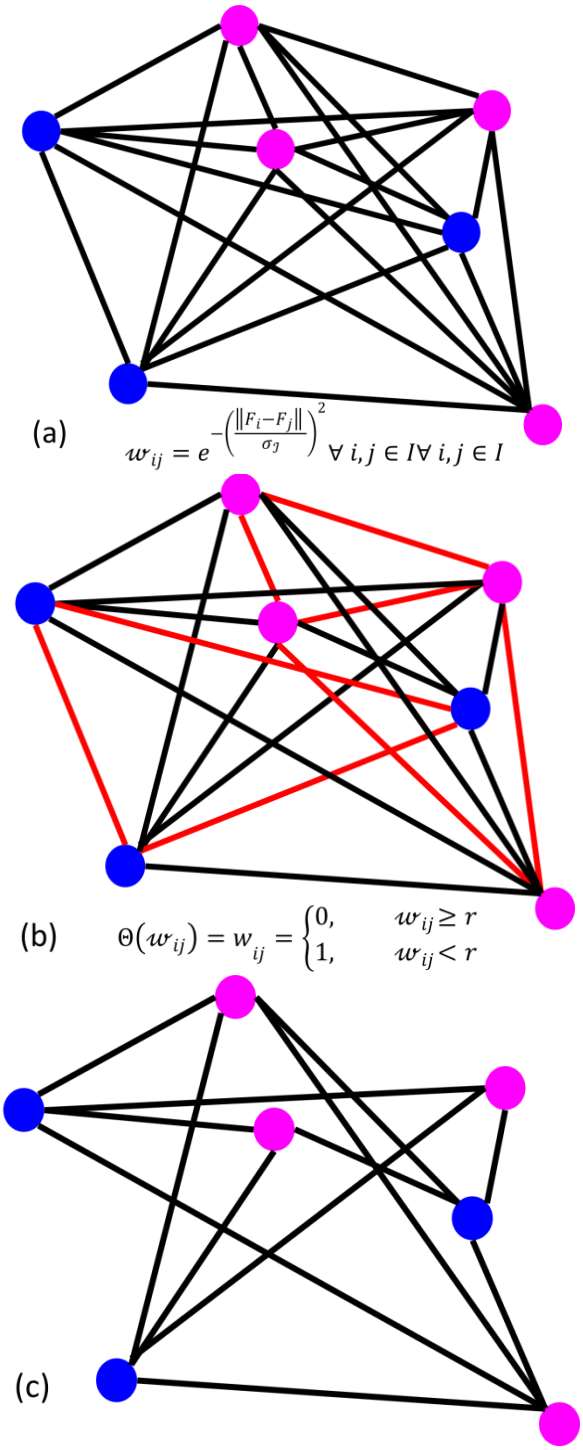


Figure 4-5: Edge weighted graph representation.

(a) a well-connected representative graph, with various colored nodes (analogous to different surface textures) (b) Edges connecting similar nodes are identified based on the radial basis function (c) edges connecting similar nodes are pruned using the threshold function.

Table 4-1: Salient aspects of the two approaches used for graph theoretic representation of surface micrographs

APPROACH	WEIGHT ESTIMATES	REMARKS
ε NEIGHBORHOOD GRAPH Nodes are connected with an edge only if they are farther than a Euclidean distance ε	$w_{ij} = \|F_i - F_j\|^2$ $w_{ij} = [0, \infty)$ $\Theta(w_{ij}) = w_{ij} = \begin{cases} 1, & w_{ij} \geq r \\ 0, & w_{ij} < r \end{cases}$ $r = \frac{\sum_{i=1}^{M} \sum_{j=1}^{M} w_{ij}}{M^2}$	• Uses binary images, requiring intermediate filtering steps • Filter parameters are set heuristically • Loss of surface texture information • Compares homogeneity of surface features detected (such as scratches) between pixels • Bias due to artifacts, such as material impurities are minimized
EDGE WEIGHTED GRAPH Nodes are connected with an edge only if their weights are less than a set threshold	$w_{ij} = e^{-\left(\frac{\|F_i - F_j\|}{\sigma_J}\right)^2}$ $w_{ij} = (0, 1]$ $\lim_{F_i - F_j \to 0} w_{ij} = 1$ $\Theta(w_{ij}) = w_{ij} = \begin{cases} 0, & w_{ij} \geq r \\ 1, & w_{ij} < r \end{cases}$ $r = \frac{\sum_{i=1}^{M} \sum_{j=1}^{M} w_{ij}}{M^2}$	• Uses gray scale images directly (no filtering step is required) • No heuristically set parameters required • Compares surface texture between pixels • Afflicted by bias caused due to presence of artifacts

4.5 Studies with Simulated Surfaces

In this section, we demonstrate the graph theoretic approach using artificially generated (computer simulated) surfaces. Six scenarios are tested in order to examine the suitability of using the Fiedler number as a discriminant under different criteria. In these tests, the Fiedler number is compared with well-established statistical quantifiers, such as defect density (i.e., the percentage of defects, as in undesirable features present on a surface). These simulation-based studies also uncover instances where the Fiedler number may not be a good discriminant of surface morphology. Thus, in addition to situations where the approach is applied amicably, we identify scenarios where the approach may fail (e.g., Sec. 4.5.6, which serves as a counter-example).

We note that for a majority of the cases tested, the artificially generated surfaces are binary (i.e., black and white, 0s and 1s, respectively). The white colored features (1s) will represent various surface morphologies, such as pits, scratches, inclusions, etc. In addition, unless otherwise stated, the ε neighborhood approach (Sec. 4.4.3), which is well suited for binary-type surfaces, will be used in these case studies.

4.5.1 Case 1: Effect of Morphology Type

Case 1(a): Differences in surface morphology under stationary process conditions.

It may be contended that a count of the number of surface features detected from a binary representation of the micrograph (readily obtained using image edge detection algorithms [177]) can also suitably estimate the surface morphology. Though this might be true for some cases, the feature count (density) may not be a universally appropriate measure for representing the surface topology. To demonstrate this assertion is the objective of this case study.

Aim of the study

Demonstrate that the graph theoretic invariant Fiedler number can capture differences in surface morphology.

Three types of surface morphologies with feature density, i.e., percentage of white colored features on the surface, identically set at 3% (Figure 4-6) are patterned. A surface with Type 1 defects shown in Figure 4-6(a) is similar to features typically observed on semiconductor wafers immediately following lapping. The wafer surface depicts nanoscale pits (see also Figure 4-53(a) and (c)) which require about 12 − 15 minutes of subsequent CMP operations. We artificially generate pits (white spots), the locations of which are obtained by sampling from a discrete uniform distribution.

For example, given a 1000 x 1000 pixel surface with no defects (black colored surface, in practice a 1000 x 1000 matrix of zeros), a random coordinate (sampled from a discrete uniform distribution with bounds [1 1000], $\{\lceil\mathcal{U}(1,1000)\rceil, \lceil\mathcal{U}(1,1000)\rceil\}$) is selected. At this coordinate we create a point defect (i.e., change the matrix element at the identified location to a 1). We continue in this manner, taking care that sampled locations do not repeat (sampling without replacement), until 3% of the total surface (i.e., 30,000 points out of the total ($1000 \times 1000 =$) 10^6 possible locations) has been converted into defects (white spots).

A surface with Type 2 defect is depicted in Figure 4-6(b). This is similar to a surface with randomly distributed nano-scratches. Plurality of such scratches are typically evident under sub-optimal CMP conditions, such as when processing with high polishing load (> 20 N), using worn out polishing pads, or when the polishing slurry is contaminated with particulate residue. When detected early, these scratches can (typically) be eliminated within 3 – 6 minutes of CMP with optimally adjusted parameters (unless the scratches are very deep).

We simulate a surface with (uniformly distributed) scratches between 20 and 200 pixels in length (discrete uniform distribution; $\lceil\mathcal{U}(20,200)\rceil$) at random locations ($\{\lceil\mathcal{U}(1,1000)\rceil, \lceil\mathcal{U}(1,1000)\rceil\}$) inclined at ± 45° to the horizontal. The starting locations for the scratches are sampled from a discrete uniform distribution, in a manner identical to the procedure described for the previous case. We

note that three parameters are simultaneously manipulated for this type of surface viz.:

i. the location of the scratch, which is randomly sampled (without replacement) from the distribution $\lceil \mathcal{U}(1,1000) \rceil, \lceil \mathcal{U}(1,1000) \rceil$,

ii. the slope of the scratch, alternated as 1 or -1, i.e., if a scratch has a positive slope, the subsequent scratch will have a negative slope, and

iii. the length of the scratch sampled from a uniform random distribution $\lceil \mathcal{U}(20,200) \rceil$.

Once the starting location has been identified, a scratch with the appropriate slope (± 1), and length is generated. This procedure is recursively implemented until 3% (i.e., 30,000 points out of the total (1000 × 1000 =) 10^6) possible locations) of the total surface is covered in scratches.

Figure 4-6(c) shows a surface with Type 3 defect. Such a surface typifies chemical corrosion and particulate embedding [19] defects observed during CMP. Compared to prior cases, a large contiguous portions of the wafer remain unaffected. In our experimental tests, such defects were remedied in < 3 minutes of CMP. The procedure for simulating Type 3 surfaces is largely identical to the Type 1 surface. But, instead of a point defect (as in Type 1), for Type 3 surfaces the neighboring 20 to 40 pixels (i.e., the area of a defect is 20 px. sq. to 40 px. sq, sampled from a discrete uniform distribution) in a square loci around a random location are also converted into defects (the matrix elements are changed to 1, note that there is a chance of overlap between features). Once again, the total area

occupied by defects is restricted to 3%. The characteristics of the three types of surfaces studied are summarized in Table 4-2 below.

Table 4-2: Morphology details for the three types of simulated defects. The total density is maintained identical (3%) for all three types.

MORPHOLOGY TYPE	SHAPE	LOCATION	SIZE	ORIENTATION
TYPE 1	Point defects	Uniformly sampled $\{\lceil\mathcal{U}(1,1000)\rceil, \lceil\mathcal{U}(1,1000)\rceil\}$	1	NA
TYPE 2	Scratch	Uniformly sampled $\{\lceil\mathcal{U}(1,1000)\rceil, \lceil\mathcal{U}(1,1000)\rceil\}$	$\lceil\mathcal{U}(20,200)\rceil$	Slope ± 1
TYPE 3	Square blobs	Uniformly sampled $\{\lceil\mathcal{U}(1,1000)\rceil, \lceil\mathcal{U}(1,1000)\rceil\}$	$\lceil\mathcal{U}(20,40)\rceil$	NA

Fifty two different instances (replicates) are simulated for each of the above surface types (total 156 data points). We used the ε neighborhood graph representation for these simulated cases because they are binary images. First, the weights w_{ij} are estimated using Eqn. (4-25), and subsequently the threshold from Eqn. (4-27) is applied. Thereafter, steps from Eqn. (4-4) thru (4-10) are implemented.

Plots of the first two non-zero Laplacian ($\mathcal{L}$) eigenvectors, v_2 and v_3, are shown in Figure 4-6(d)-(f). Here we notice a significant difference in the eigenvector trajectories. The corresponding Fiedler statistics are plotted in Figure 4-7. Though the defect density (percentage of white colored areas) were identical for these three cases (3%), the mean Fiedler numbers are significantly different, as seen from Figure 4-7 and Table 4-3. They range from close to 0.81 to 0.66 for Type 1 vs. Type 3 respectively. An ANOVA study revealed a statistically significant overall and pairwise (p-val. < 0.01) difference in the mean Fiedler number for the three cases illustrated.

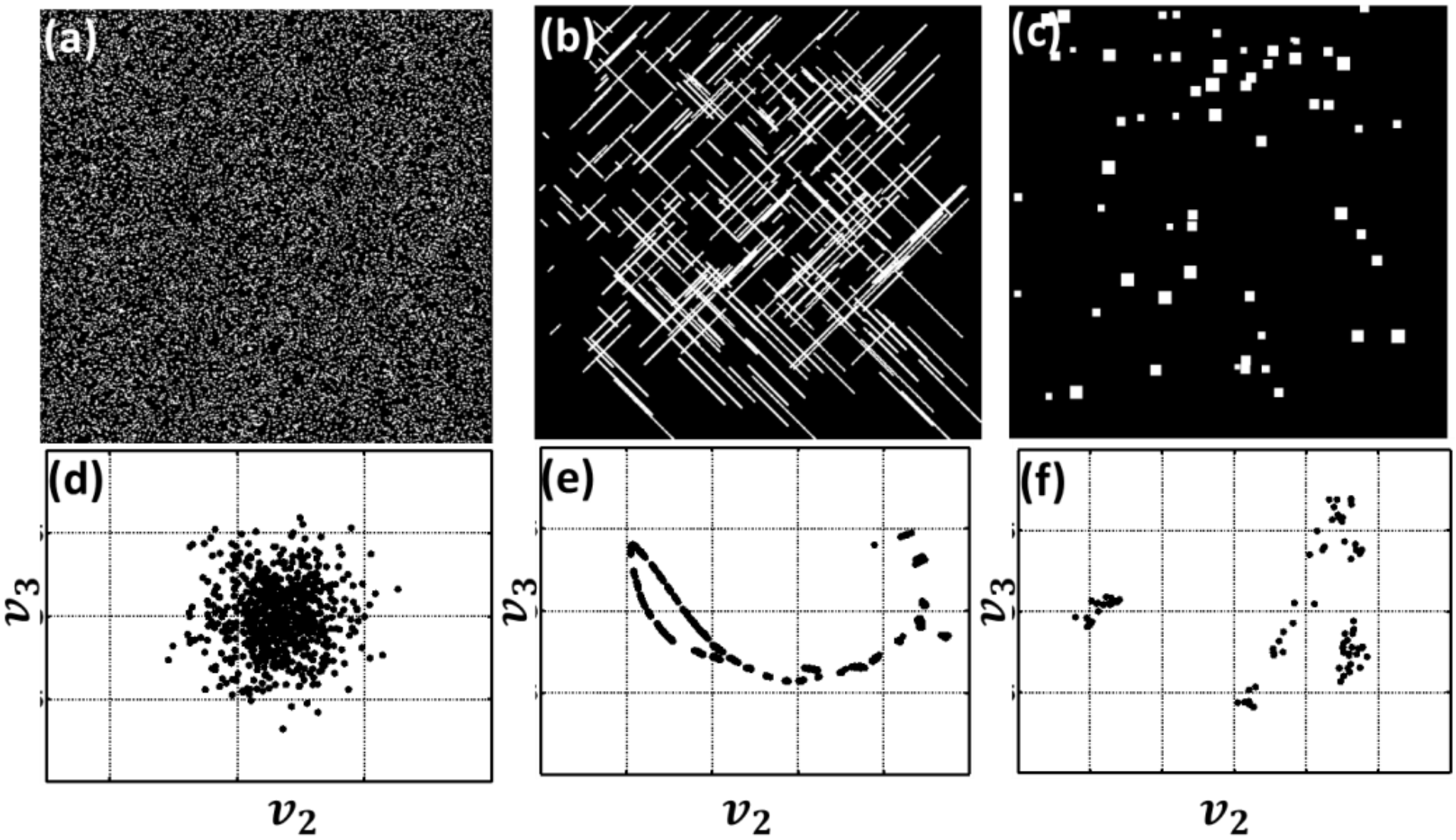


Figure 4-6: Different types of simulated surface morphologies.

(a) – (c): Three types of simulated topographies, white colored details represent wafer features. The density of wafer features (white spots) is maintained at 3%. (d) – (f): The eigenvectors map corresponding to topographies (a) – (c).

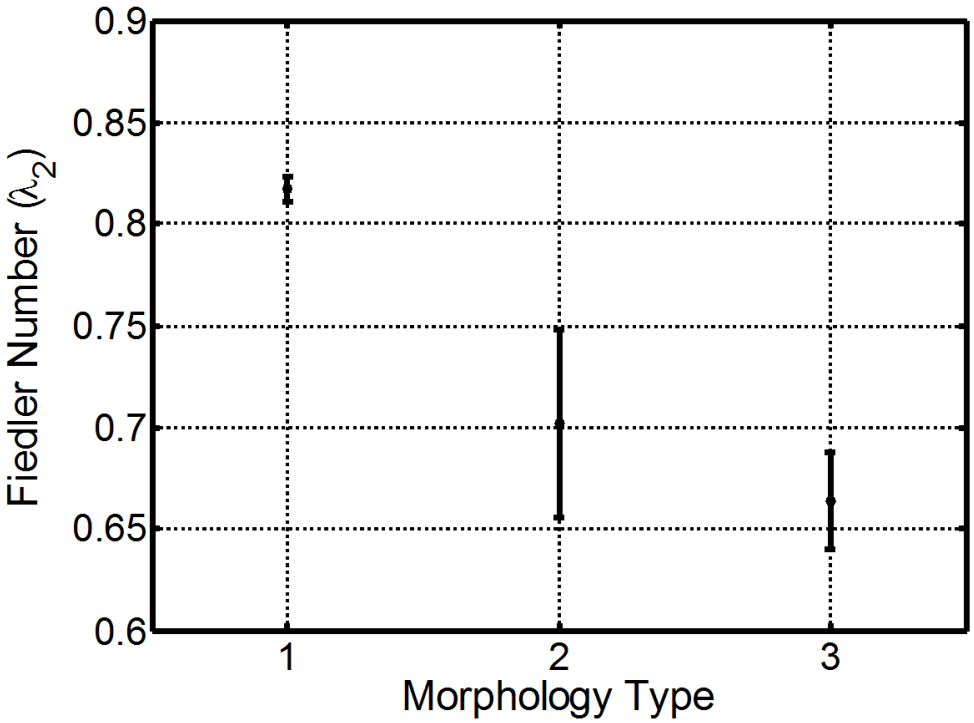


Figure 4-7: Interval plot of Fiedler number estimated for the three simulated defect types. The bars represent the ±1σ limits.

Table 4-3: Some descriptive statistics of the Fiedler number measured for the three simulated defect types (52 data points for each defect type)

DEFECT (MORPHOLOGY) TYPE	MEAN	STD. DEV.	RANGE	IQR
TYPE 1	0.8172	0.0061	0.0248	0.0084
TYPE 2	0.7019	0.0461	0.1743	0.0715
TYPE 3	0.6641	0.0237	0.1052	0.029
Number of data points		52 (for each feature type)		
Pooled standard deviation		0.0301		
Standard error		0.0042		

We observe (from Figure 4-7 and Table 4-3) that the Fiedler number is high (0.81) for Type 1 representations, this is explained based on the high degree of connectivity between nodes (the graph network is shown Figure 4-8(a)). In Type 1 (Figure 4-6(a)), the pairwise distances w_{ij} are largely dissimilar (pixel rows look different from each other). In other words, for Type 1 surfaces, w_{ij} = 1 (Eqn.

(4-3)) for a large number of nodes (vertices); consequently the similarity matrix S (Eqn. (4-4)), as well as, the degree matrix $\mathcal{D}$ (Eqn. (4-5)) are not sparse for this case. This manifests in the rich interconnectivity between nodes seen in Figure 4-8(a).

In contrast, for Type 3 (Figure 4-6(c)) the network is well pruned (w_{ij} = 0 for many connections, therefore S is sparse), because $\mathcal{w}_{ij}$ (note the difference between w_{ij} and $\mathcal{w}_{ij}$, see Eqn.(4-2) and (4-3)) does not vary significantly across pixel rows (only some pixel rows are different from others, the surface as a whole is homogenous). As a result, we notice from Figure 4-8(c) that the nodes of the graph are not as well connected compared to Type 1 (notice particularly, the sparsely connected nodes in the south-west region). Hence, it is relatively easy to isolate a node for Type 3 surfaces in comparison to Type 1 surfaces. Therefore, from the definition of $h(G)$ in Eqn. (4-16) – (4-14), one expects the Fiedler number to be higher for Type 1 relative to Type 3.

This study intends to illustrate that the Fiedler number is sensitive to variations in surface morphology that are not captured using statistical parameters, such as defect density. However, the converse is not presumed, i.e., the type of surface morphology may not be surmised from the Fiedler number if we do not have *a priori* knowledge of the surface morphology. In other words, the simulated surface morphologies may have the same Fiedler number if the defect densities were allowed to vary.

Inferences from the study

1. Defect count (density) is not a good quantifier for surface morphology.
2. Fiedler number could be a more appropriate statistic for assessing surface morphology.

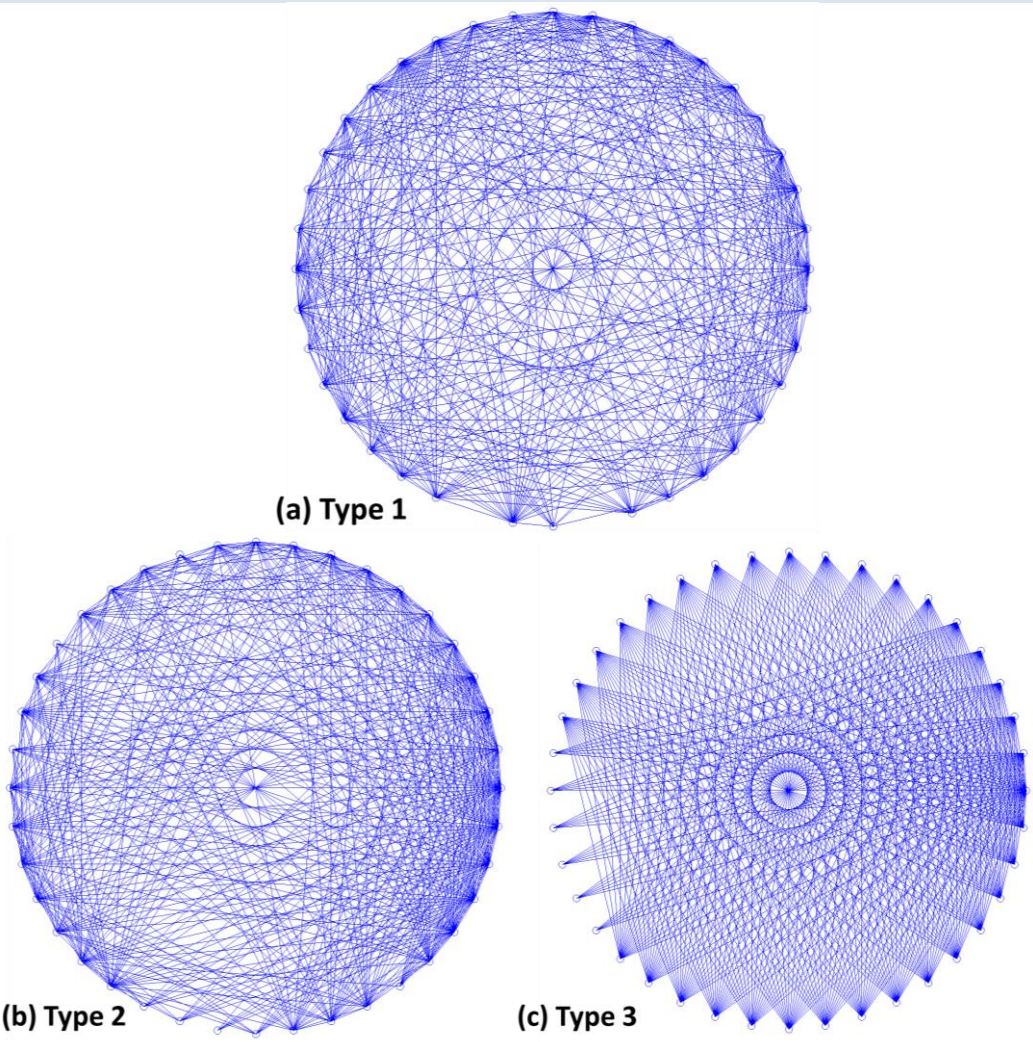


Figure 4-8: The typical graph network for the three defect types studied

(a) Type 1 (pits), (b) Type 2 (scratch), and (c) Type 3 (corrosion). These plots are essentially the similarity matrix S in graphical form; here we show the connections among 40 nodes of the graph $G \equiv (V, E)$. The nodes are located on the circumference of a unit circle.

Application to a practical scenario

The implications from this study are readily applied to some practical scenarios, such as differentiating the microstructure of materials, speckle defects on textiles, ink splatter patterns, etc. For example, we illustrate a case where the graph theoretic Fiedler number for differentiating the microstructure of cast iron. Micrographs of two types of cast iron microstructure (courtesy Metals Handbook Vol. 9 [194]), namely, lamellar and nodular matrix are shown in Figure 4-9, along with the Fiedler number (λ_2). The mean Fielder number assessed for the two different types of microstructures are statistically different at the 10% level of significance.

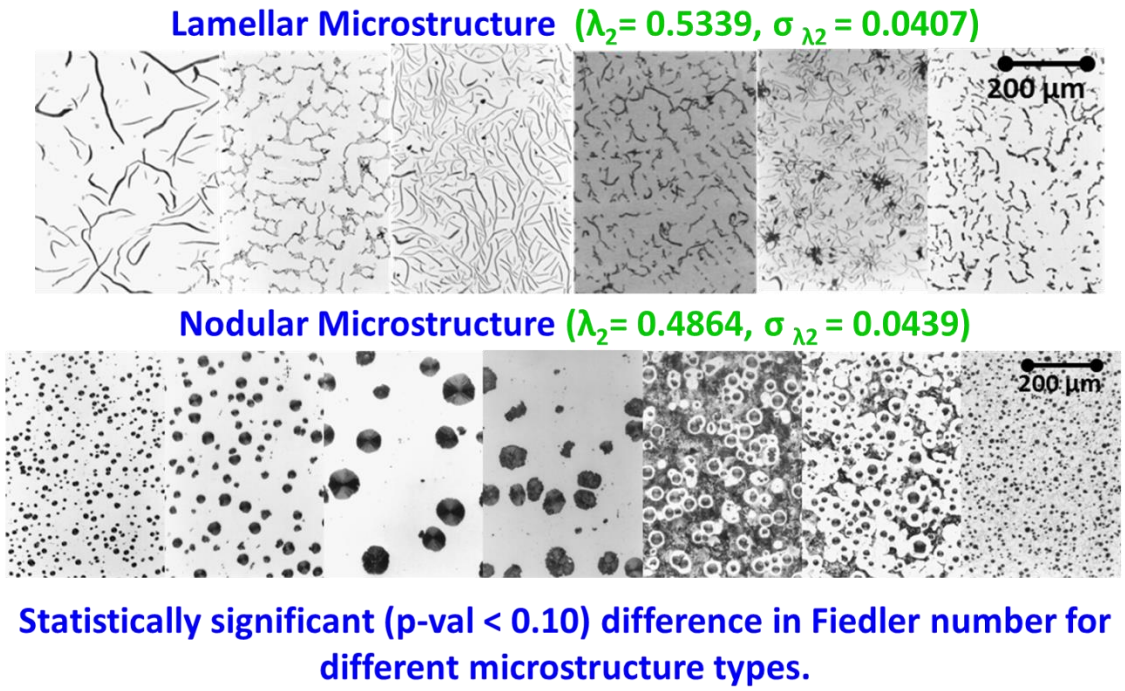


Figure 4-9: Fiedler number for two different types of cast iron microstructures.

Case 1(b): Detection of changes in surface morphology from an evolving process

Continuing our study using simulated surfaces with different morphologies (while maintaining the defect density constant at 3%), we now test the scenario where the morphology changes during operation (similar to the situation when the process conditions evolve to a sub-optimal state). We wish to quantify the delay in detection of a change in surface morphology using the Fiedler number as a discriminant.

Aim of the study

1. Demonstrate that the graph theoretic invariant Fiedler number can capture evolving surface morphology.
2. To quantify the average run length (ARL) statistics when Fiedler number is used in a SPC setting for detection of changes in surface morphology.

As a practical analogy, consider the Type 3 surface of the previous study which may be representative of an acceptable CMP surface. As we continue to polish, over time the process conditions will deteriorate (e.g., the polishing pad may become glazed, the slurry contaminated, etc.), and as a consequence the wafer surface quality may be affected, e.g., scratches may appear on the surface (Type 2 surface).

If we consider the Type 3 surface as 'normal', and a Type 2 surface with scratches is termed 'defective' (see Figure 4-10), *how many defective wafers will be produced, before we realize that the process is malfunctioning (i.e., the process*

state has changed), if the Fiedler number (λ_2) is used as a discriminant? In a physical sense, suppose that we measure the surface morphology in terms of the Fiedler number (λ_2, using an automated inspection device), of a semiconductor wafer as soon as it has finished a CMP step, then, how quickly could we detect an out of control process condition (right side surface in Figure 4-10).

In other words, we wish to quantify the average run lengths (ARL) under normal (ARL_0), and defective conditions (ARL_1), in an SPC scenario using the Fiedler number (λ_2) as the key process output variable (a large ARL_0 and small ARL_1 is desired).

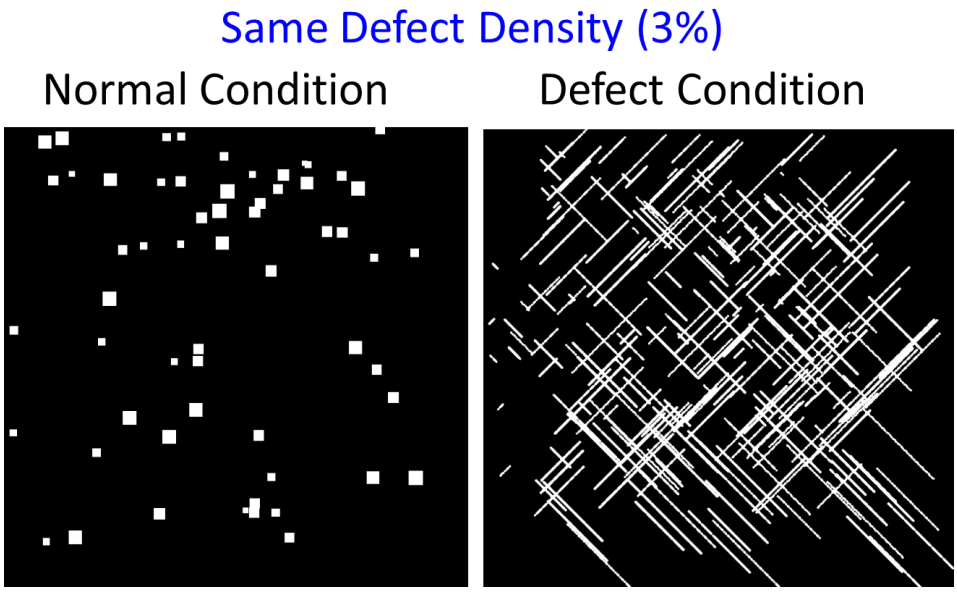


Figure 4-10: Simulated change in surface morphology.

Left: A representative surface produced under normal processing conditions.
Right: A representative surface produced when the process is malfunctioning.

Accordingly, we generated artifical surfaces as shown in Figure 4-10. The Type 3 surface morphology is considered as the normal condition, whereas the Type 2 surface is representative of the defective condition. First, 2337 instances

(about 12 hours of PC time) of Type 3 surface are generated, and their Fiedler number (λ_2) is computed. Subsequently, control limits for a 'runs' SPC chart (X-chart) are calculated, using which the number of points out of control during the normal condition can be observed, and the ARL_0 estimated (viz. an estimate of the Type 1 error). There were 5 points observed out of 2337, (Figure 4-11(a)), i.e., on an average one point in ~ 467 points. The Type I (false alarm) error is therefore 0.2139% (~ 2.85σ on the standard normal).

Thereafter, 1000 instances of the defect condition (right side Figure 4-10) are simulated. Using the control limits from the normal condition (after the delete and revise procedure), we can estimate the ARL for the defect condition (i.e., ARL_1). We found that 213 points were out of control out of 1000 (Figure 4-11(a)). This translates to an ARL_1 of 4.6.[17] Therefore, on an average, 4 wafers will be processed before we discover that the process is malfunctioning. In all, as indicated by Figure 4-11(b), there is close to 1.5σ shift (1.46 σ) going from the normal condition (stable condition) to out of control state.

Pertinently, if the defect (or feature) percentages (feature density is fixed at 3%) were to be used as a discriminant, one would not be able to recognize any change in the process state. This study demonstrates that the Fiedler number can overcome some of the lacunae associated with traditional statistical quantifiers, such as defect density.

[17] As one expects this can be reduced to about 2.13 if the AT&T runs rules are used, but one (necessarily) sacrifices the ARL_0 which reduces to ~ 65.

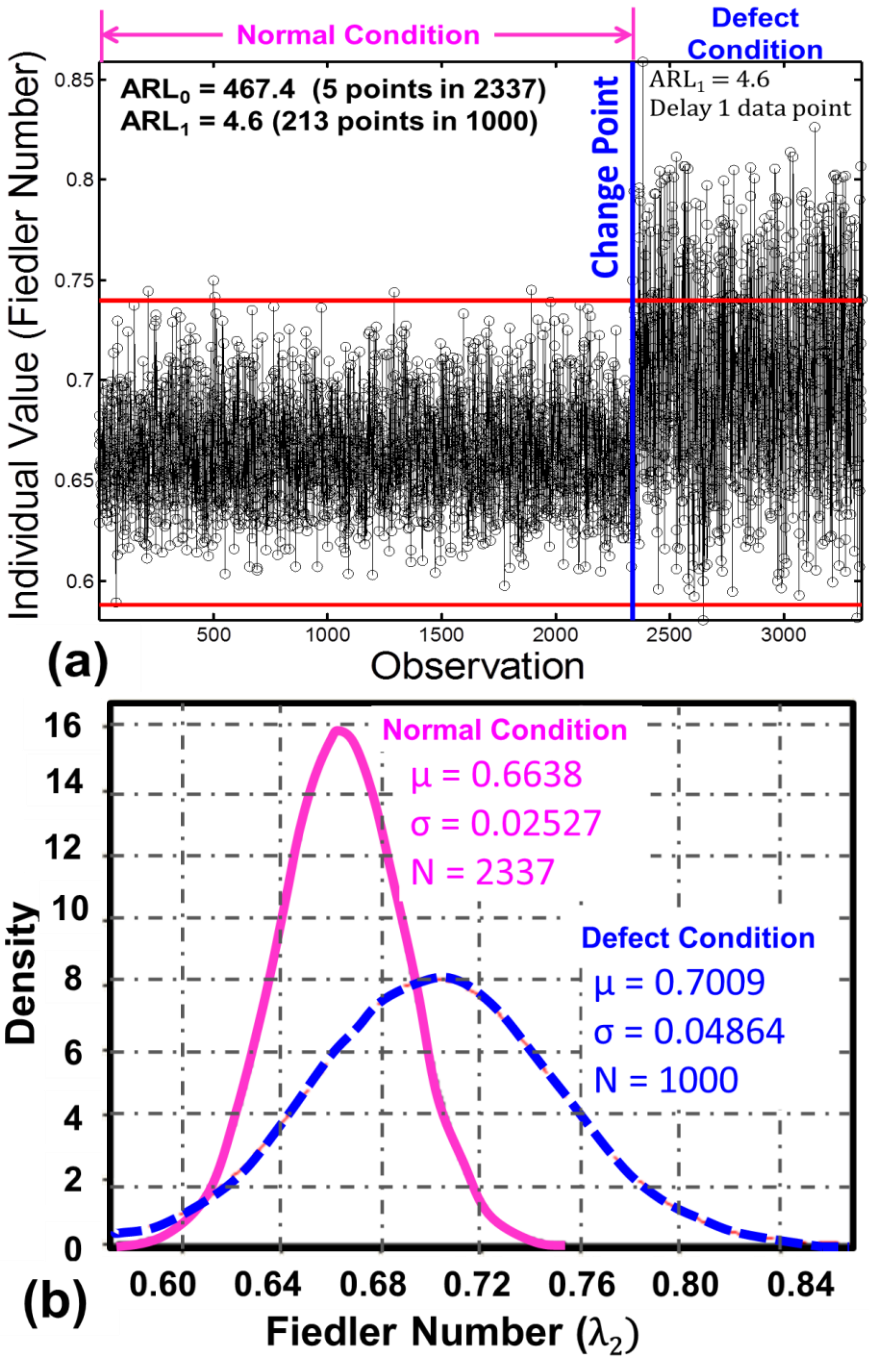


Figure 4-11: Using the Fiedler number as a discriminant to detect process drifts.

(a) ARL_0 and ARL_1 calculated from the X-chart control limits. (b): The distribution of the Fiedler number for the normal and defect conditions, indicating that there is shift of approximately 1.5 standard deviations (from the normal condition) in the Fiedler number.

Inferences from the study

The Fiedler number can be useful for capturing changing surface morphology, the ARL_0 and ARL_1 limits are estimated at 467 points and 4.6 units respectively.

4.5.2 Case 2: Effect of Defect Density

This study tests the sensitivity of the Fielder number (λ_2) relative to changes in defect (feature) density (ρ). In the previous case, the Fiedler number (λ_2) was used as a discriminant for different types of surface morphologies with the feature density (ρ) maintained constant at 3%. In this case the converse is tested, the morphology type is maintained constant as the defect density is varied.

Aim of the study

Demonstrate that the graph theoretic invariant Fiedler number can capture the changes in defect density.

In previous cases, the defects were randomly located (sampled from an uniform distribution), while the number of defects occurring over an area was a fixed parameter. Whereas, in this study, we consider the following two scenarios:

i. The defects (or features) have a deterministic arrival rate, but are located randomly (sampled from a uniform distribution).
ii. The defects have a stochastic (Poisson distributed) arrival rate apart from being located randomly (sampled from a uniform distribution).

We note the key difference in the sampling procedure applied between the Poisson and uniformly distributed cases. In the second case, the arrival rate (mean

number of the defects) is a random (Poisson) variable. While in the first case, the density of defects is a fixed parameter.

Contrary to usual practice, we discuss the more complicated Poisson case first. This is because certain observations are more amenable to an explanation due to well-known properties of the Poisson distribution, particularly, the mean and variation for a Poisson distribution are identical (i.e., $\mu = \sigma^2 = \lambda$).

Case 2(a): Poisson distributed defects

A spatial Poisson process is used to sample features on a 1000 × 1000 pixel area (V). This area is further subdivided into smaller 100 × 100 pixel areas (sub-spaces, v_k), $k = \{1 \dots 100\}$. For each v_k, and a set Poisson arrival rate (λ_i), the number of defects (n_k) for each v_k is estimated. (We note that in this case the number of defects (n_k) in the subspace v_k is a random variable. These preliminary conditions are formalized as,

$$\begin{gathered} V \equiv v_1 \cup v_2 \cup v_k \cdots v_N \\ v_k \subset \mathbb{R}^2 \\ v_1 \cap v_2 \cap \cdots v_N = \emptyset \\ n_k | v_k \sim \text{Pois}(\lambda_i). \end{gathered} \tag{4-31}$$

In effect, v_k are independent, non-overlapping, sub-spaces, within which Poisson sampled defects (n_k) occur (satisfying IID conditions). The defects (n_k) in each subspace v_k, i.e.,$(n_k|v_k)$ are (uniformly) randomly assigned.

In other words, for every $v_{k,}$ in V, ($v_k \subset V$), we obtain the number of defects n_k sampled from a Poisson distribution. Then, for each v_k we decide the location of the sampled defects (n_k) (i.e., the coordinate $(X_k, Y_k) = \{(x_k, y_k)_1, (x_k, y_k)_2, \cdots (x_k, y_k)_{n_k}\}$ of each n_k in v_k).

At these locations (X_k, Y_k) we create a defect (i.e., change the element $(x_k, y_k)_i$ of 1000 × 1000 matrix to 1). This gives us one instance of a Poisson generated surface $V^{1|\lambda_i}$ for a set Poisson arrival rate λ_i. We generate 50 instances of V for each λ_i, i.e., $V^{\lambda_i} = \{V^{1|\lambda_i} \ldots V^{50|\lambda_i}\}$ for each λ_i tested. The Fielder number λ_2 for each instance ($V^{j|\lambda_i}$) is subsequently computed. Representative simulated surfaces for different Poisson arrival rates are shown in Figure 4-12.

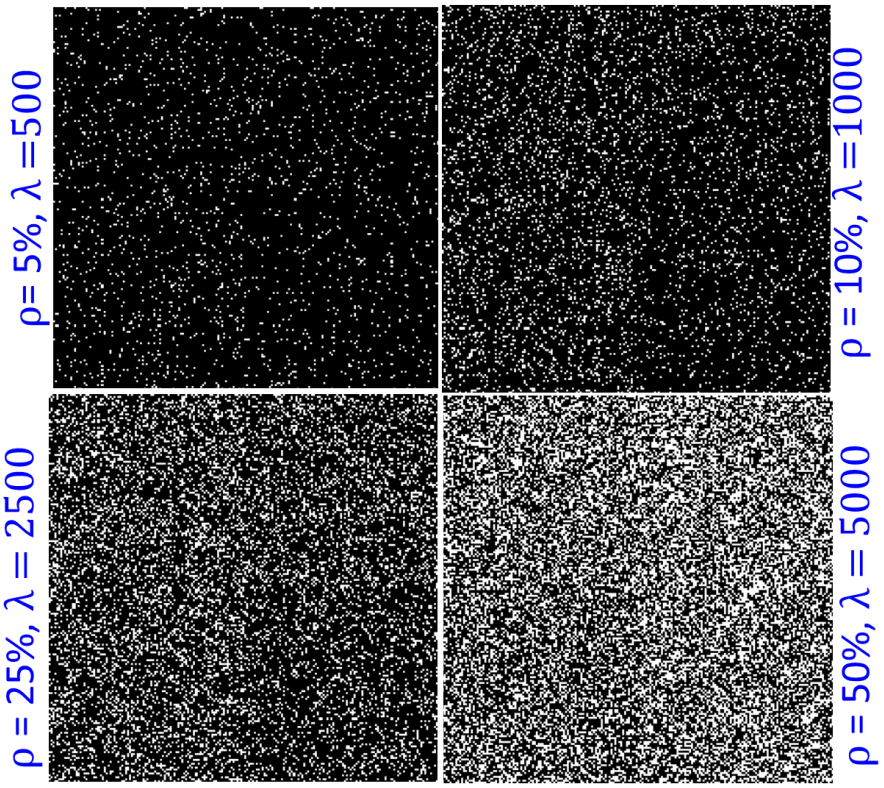


Figure 4-12: Typical surfaces obtained on locating defects (pits) sampled from a Poisson distribution

Furthermore, in Figure 4-13, the mean Fielder number ($\bar{\lambda}_2$, along with maximum, minimum, and standard deviation of the Fiedler number) for $V^{\boldsymbol{\lambda_i}}$ is graphed against the Poisson arrival rate λ_i. The abscissa also depicts the average defect density (ρ) as a percentage.

The defect density (ρ) can be calculated as follows: Suppose, $\lambda_i = 50$, which implies, there are 50 defects (on average) per v_k. Hence, there are 50 × 100 = 5000 defects (ρ, on average on the surface V). Therefore, the mean percentage of defects in V is, $\rho\% = \frac{5000}{10^3 \times 10^3} \times 10^2$ = 0.5%.

We note four regions of interest in Figure 4-13:

- Region 1: In the region of ρ < 1% (λ < 100), the Fielder number is not observed to be sensitive to changes in the feature density. This is because the simulated image is sparse for low arrival rates. In addition, as the mean of the Poisson distribution is the same as its variance, at small λ the surface generated is largely homogeneous. Consequently, the pixel rows are not substantially different from others, and therefore the similarity matrix *S* has few connections.
- Region 2: In the region between 1% < ρ < 100% (100 < λ < 10,000), the mean Fiedler number ($\bar{\lambda}_2$) increases with increasing Poisson arrival rate λ. This is explained based on the increasing variation in the surface. Since the variance of the Poisson distribution is identical to the mean, the surface becomes progressively inhomogeneous with increasing λ.

This is evident on observing the simulated surfaces shown in Figure 4-12 (which all belong to this range of λ). For instance, we notice that the bottom right of the surface for ρ = 50% (λ = 1000) in (Figure 4-12) is distinctly different (it has significantly fewer defects) from the rest of the surface. A network graph representative of this region is shown Figure 4-14 (top), in which the rich interconnection between nodes is apparent. This well-connected network has a relatively high λ_2 value (0.82).

- Region 3: In the region between 100% < ρ < 750 % (10,000 < λ < 75,000) the mean Fiedler number ($\bar{\lambda}_2$) starts decreasing, this is explained on the basis that the region to the right of the ρ = 100% in Figure 4-13 is essentially a complement of the region to the left ($\rho <$ 100%, recall that the surface is a binary (1 and 0) matrix).

  That is, the defects (1's) now dominate the rest of the surface (0's); instead of a black surface with white colored defects, we have the opposite case. As λ increases, the black areas ('good' areas) decrease; this is also evident from the network graph for a surface in this region shown in Figure 4-14 (bottom left), where the node connections are sparser in comparison to the network of Region 2 (Figure 4-14 (top)).

- Region 4: In the region of ρ > 750% (λ > 75,000) we observe a complex behavior in $\bar{\lambda}_2$, the explanation for which requires further understanding of graph theoretic concepts.

After the surface is almost saturated with defects (tail end of region 3), we briefly enter a state where the network has the characteristics of a bipartite graph [185]. As we will explain in more depth in Sec. 4.5.6, the Fielder number of a bipartite graph equals 1 ($\lambda_2 = 1$) [185]. In Figure 4-13 the surfaces generated in the $750\% < \rho < 1200\%$ region evolve to such a bipartite graph state. A representative bipartite graph from this region is shown in Figure 4-14 (bottom right). A peculiar aspect of this network is that none of the nodes from the left hand portion of the graph are connected to their immediate neighbors.

Furthermore, these left hand side nodes all have three edges, and connect to three particular nodes in the north-east region. Such graph networks typify a bipartite structure, and present a limiting (pathological) case for our approach (Sec. 4.5.6).

Also, in this region some (of the total 50) realizations can converge to a bipartite state ($\lambda_2 = 1$), while others are completely covered with defects (that is, there are no features to detect, consequently $\lambda_2 = 0$). For example, at $\rho = 1250\%$ we observed that 47 out of 50 realizations had $\lambda_2 = 1$, and the rest (i.e., three realizations) had $\lambda_2 = 0$, therefore the mean $\bar{\lambda}_2 = 0.94$.

In comparison, at $\rho = 1500\%$, about 60% of the realizations (31 out of 50) realizations are observed to have $\lambda_2 = 0$, while the rest (19 out of 50) are bipartite ($\lambda_2 = 1$), which implies $\bar{\lambda}_2 = 0.38$ (average of nineteen 1's and thirty

one 0's). This also the explains the relatively large variation observed in $\bar{\lambda}_2$ between 1200% < ρ < 1500%.

Finally, in the region of ρ > 1500 %, there are no features to detect as the surface has been completely 'whited out' with defects, therefore, $\bar{\lambda}_2 = 0$.

One might consider Region 2 (1% < ρ < 100%, and the complimentary Region 3 by induction) to be practically relevant. The tails of Figure 4-13 (i.e., Region 1 and 4) are largely of academic interest.

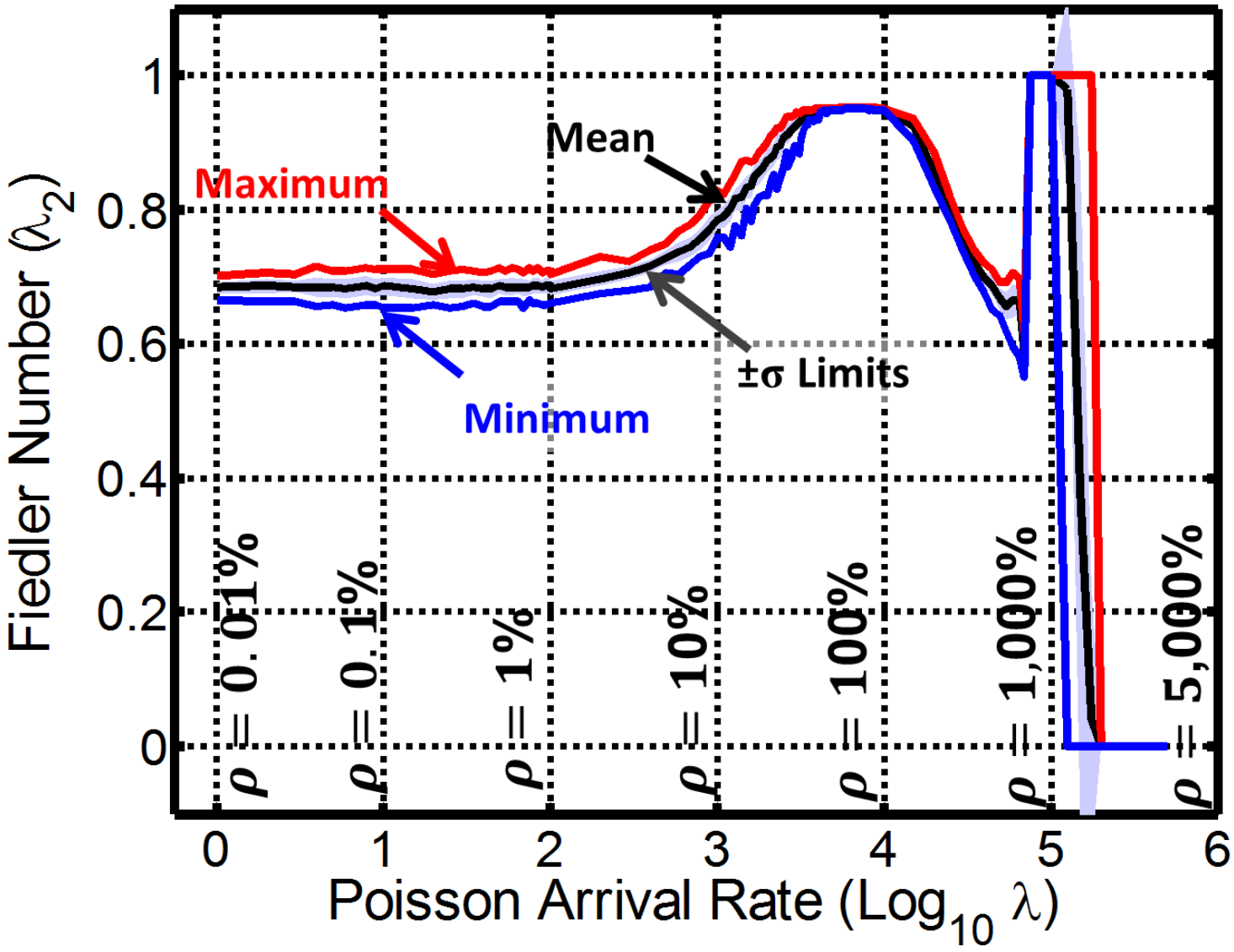


Figure 4-13: The Fiedler number for Poisson disturbed defects with arrival rate λ.
Each measurement point (162 points in the interval λ=1 to λ=500000), has 50 simulated instances.

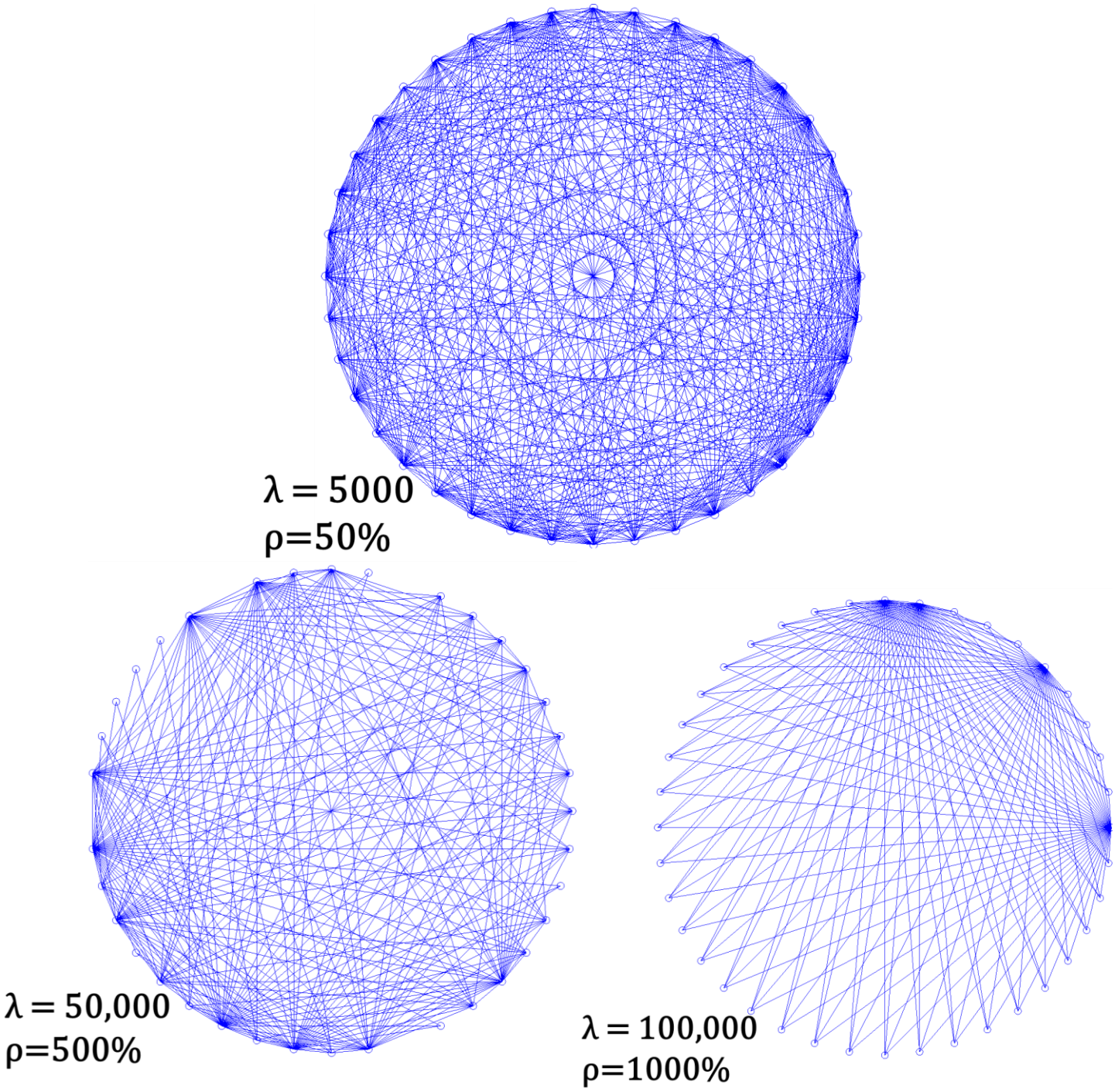


Figure 4-14: Some typical network graphs (forty nodes) at different Poisson arrival rates (λ).

For λ=5,000 the graph shows fairly robust connectivity, which is progressively depleted as λ increases to 50,000. Thereafter, close to λ=100,000 the network begins to approach the bipartite structure.

Case 2(b): Uniformly distributed defects

In this study, uniformly distributed defects with different defect densities (ρ) are generated, and the surface is simulated in a manner identical to the Type 3 surface of Sec. 4.5.1. We generate 50 replicates for each level of defect density (ρ), which were tested between 1% and 75% in intervals of 0.5%, hence there are 150 levels tested for ρ (the behavior at the right tail is similar to the Poisson case). We did not test beyond ρ > 75% due to computational constraints[18]. We note that ρ is a deterministic parameter, and hence in retrospect, this study is a special case of the previous Poisson case. Some typical surfaces simulated for this study are shown in Figure 4-15.

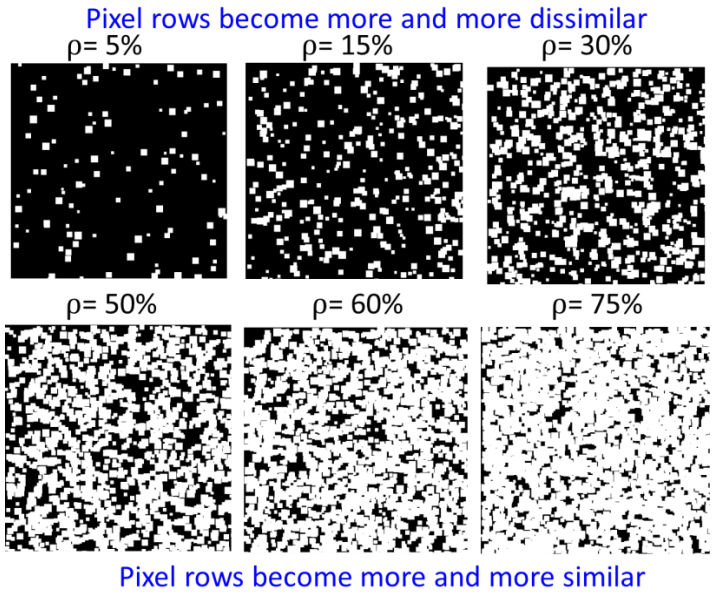


Figure 4-15: Typical surfaces obtained on simulating defect locations sampled from a uniform distribution.

[18] Because the locations for the defects are sampled without replacement, searching for a 'free' location is computationally expensive at large ρ.

Subsequently, the Fielder number (λ_2) is computed for each instance (there are 50 instances for each λ). Figure 4-16 shows the mean Fiedler number ($\bar{\lambda}_2$) plotted against ρ. At the ρ =1% level, $\bar{\lambda}_2$ depicts complex behavior, this is due to the graph network being extremely sparse and approaching the bipartite state (λ_2 = 1). The reasons for this behavior are identical to those tendered for Region 4 of the previous Poisson distributed case.

Beyond $\rho > 1\%$, the $\bar{\lambda}_2$ increases (almost linearly in the range of 3% to 10%) with ρ, and finally peaks at $\rho = 50\%$. Thereafter, $\bar{\lambda}_2$ begins to decline. This is because above the 50% defect density the surfaces generated are complementary to the surfaces for $\rho < 50\%$. This behavior is evident from examination of Figure 4-15.

The behavior of $\bar{\lambda}_2$ in the region of $3\% < \rho < 11\%$ is studied in greater detail (these ranges of ρ were chosen to reflect the typical defect ranges observed in a CMP process, see Sec. 4.6). The mean Fiedler number ($\bar{\lambda}_2$) and one standard deviation interval over different ρ (for $\rho < 11\%$) are plotted in Figure 4-17, to which the Tukey's pairwise comparison test was subsequently applied. The Tukey's pair wise comparison results are shown in Figure 4-18.

From Figure 4-18, we notice that the Fiedler number (λ_2) was a statistically significant discriminant, and was capable of detecting differences in mean defect (feature) density for a majority of the cases. For $\rho < 5.5\%$, a drift of 0.5% is detected using Tukey's distance statistic, this increases to about 1.5% and beyond, for $\rho > 6\%$. Thus, from a quality monitoring standpoint, the Fiedler number can

be considered as a sensitive parameter to detect changes in defect density of a given surface. The following inferences are drawn from this study:

Inferences from the study

1. The Fiedler number (λ_2) increases as defect density (ρ) increases, for both Poisson and uniformly distributed defects. However, this holds only within certain bounds, and is contingent on the condition that the surface is not dominated by defects. Within these constraints, λ_2 can be a good discriminant for surface quality.
2. A threshold for defect density (ρ) exists, beyond which λ_2 begins to decrease (instead of increasing). The threshold is reached when the defects are dominant over the rest of the surface. The Fiedler number (λ_2) is a good discriminant in this region because the surfaces are complimentary to those obtained below the threshold defect density (as in, inference 1 above).
3. When the defect distribution becomes either too large or infinitesimal, the network graph can approach a bipartite state ($\lambda_2 = 1$), or becomes disconnected ($\lambda_2 = 0$). The Fiedler number (λ_2) is **not** a good discriminant under such conditions.

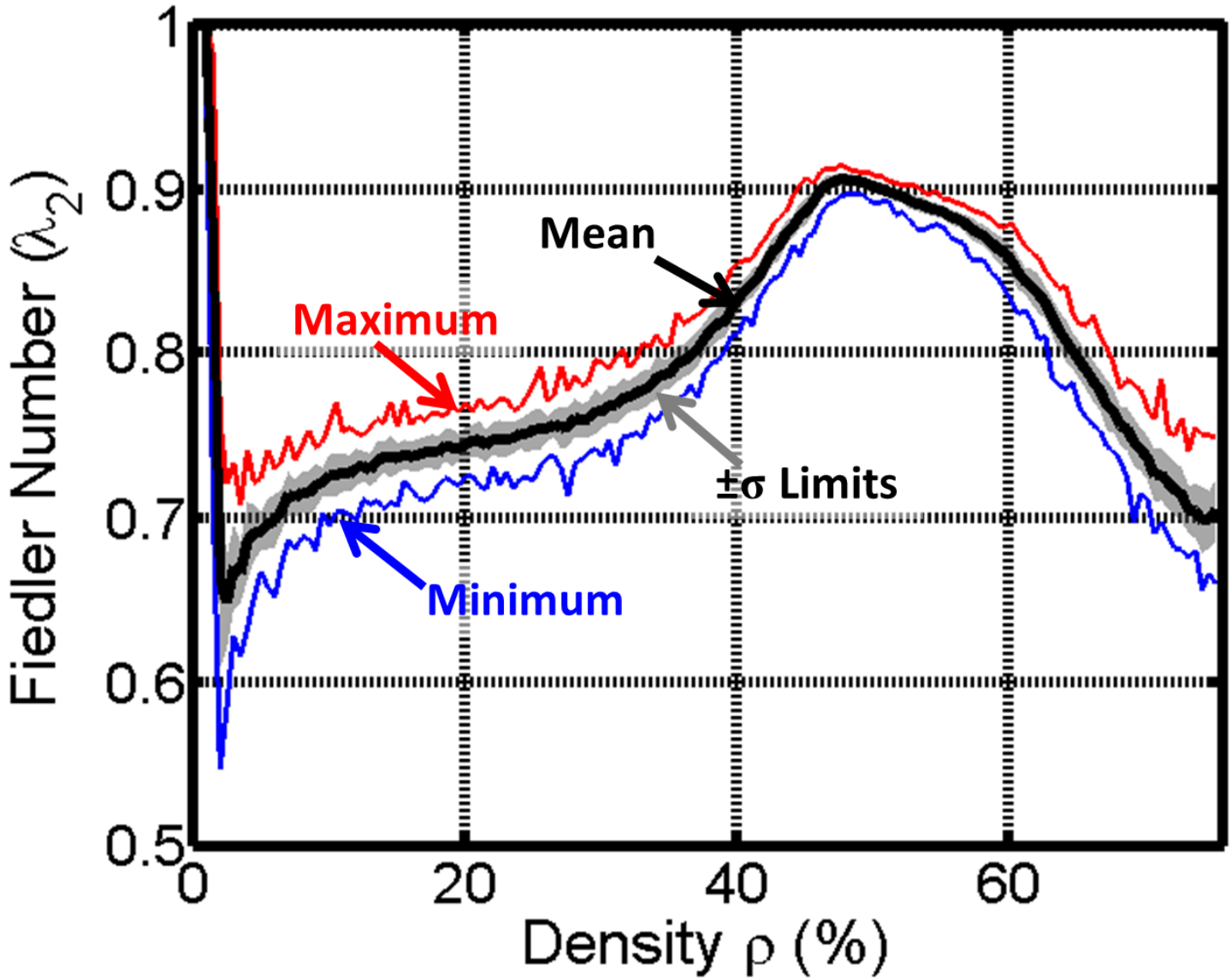


Figure 4-16: The Fiedler number for uniformly disturbed defects.

Each measurement point (in the interval ρ = 0% to λ=75% in steps of 0.5%), has 50 simulated instances.

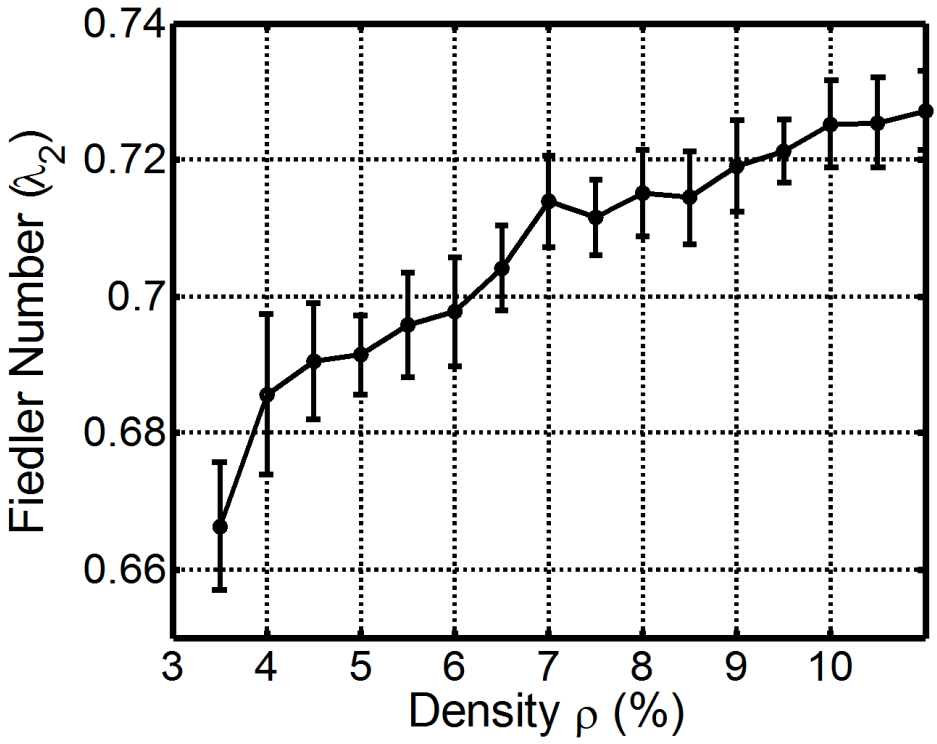


Figure 4-17: The Fiedler number for the defect density range ρ between 3.5% to 11% (typical CMP range).

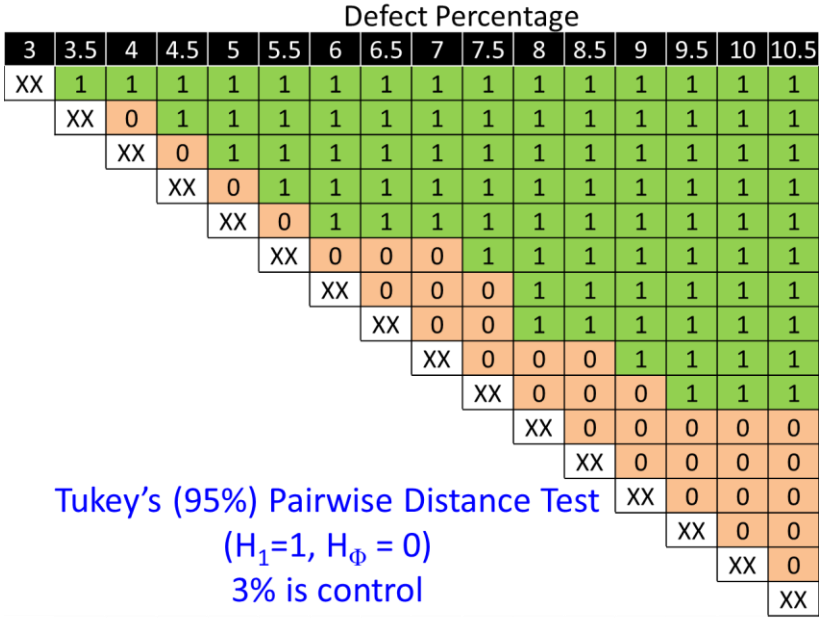


Figure 4-18: Tukey's pairwise comparison test applied for uniformly distributed defects for the (defect) density range 3% < ρ < 11%

4.5.3 Case 3: Effect of Feature size

We now study the behavior of the Fiedler number as the size of feature morphology changes, for a fixed constant defect density (3%). For this purpose, we artificially generate 1000 pixel × 1000 pixel images with the procedure used for simulating the Type 3 surfaces described in Sec. 4.5.1. Instead of maintaining the feature size at 20 sq. pixels as in the Type 3 defect in Case 1 (Sec. 4.5.1), we now vary the size (α) of the defect.

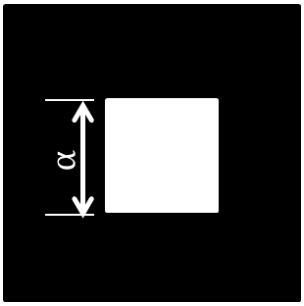


Figure 4-19: The morphology of the defect (white square) used for studying the effect of feature size on Fiedler number.

Aim of the study

Demonstrate that the graph theoretic invariant Fiedler number can capture differences in feature size of the morphology.

A sample defect is shown in Figure 4-19. The side (α) of the white colored square (the defect) shown in Figure 4-19 is varied. The defect size takes the following (dyadic sequence) levels; α =1, 2, 4, 8, 16, 32, with 50 instances simulated for each α value. We note that the features locations are sampled (without replacement) from a discrete uniform distribution$\{\lceil\mathcal{U}(1{,}1000)\rceil, \lceil\mathcal{U}(1{,}1000)\rceil\})$, with the defect density (ρ) maintained constant at 3%.

The study is conducted in two phases:

1. (Case 3(a)) The defect size is deterministic, i.e., the defects have a mean size α, with variation $\sigma^2 = 0$. Representative cases for this scenario are shown in Figure 4-21.
2. (Case 3(b)) For a fixed level of α, the mean defect size is a (discretized) uniform random variable in the range of [α, 2α]. A defect is therefore a (discrete) random variate with size $\alpha + \lceil \mathcal{U}(0, \alpha) \rceil$. The variation ($\sigma^2$) in defect size is therefore, $\frac{\alpha^2}{12}$. Typical realizations for this scenario are shown in Figure 4-24.

Case 3(a): Change in size (α) of the defect with zero variation ($\sigma^2 = 0$)

The mean Fiedler number ($\bar{\lambda}_2$, mean of 50 instances) for different levels of α is shown in Figure 4-20. This reveals a decreasing trend in $\bar{\lambda}_2$ as α increases. This behavior is explained on the observation (see Figure 4-21) that with increasing α, and with ρ fixed at 3%, the similarity matrix *S* becomes sparser (there are fewer countable defects on the surface as α increases). Consequently, the surface becomes relatively 'cleaner' at high α (compare the surface for α =1 with α = 6 in Figure 4-21), with only a few pixel rows being significantly different than the rest of the surface for the latter case. Therefore, given the progressive sparseness of the similarity matrix *S* with increase in α, one expects $\bar{\lambda}_2$ to reduce likewise.

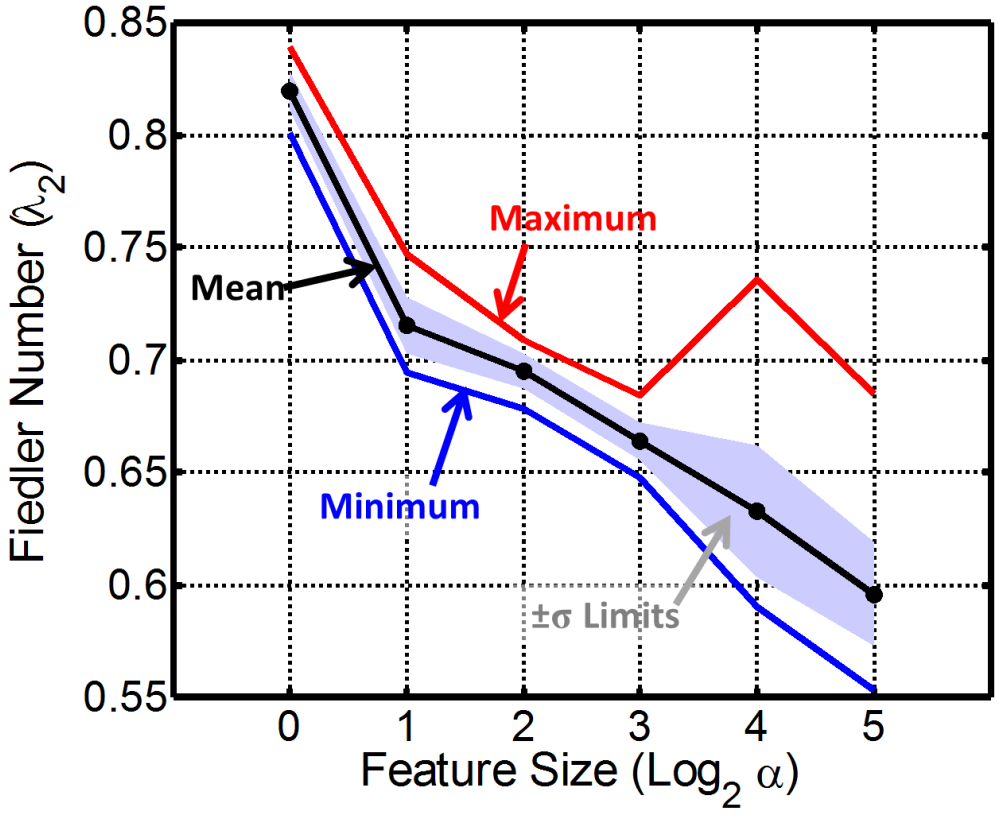


Figure 4-20: Change in mean Fiedler number ($\bar{\lambda}_2$) with change in feature size (α).

This inference becomes further evident on studying the network graphs, typical cases of which are shown in Figure 4-22. For example, at α = 6, we notice that some nodes of the graph have only one edge connecting them to the rest of the graph, while the weakest links for α = 3 have at least two edges. This implies, that the network for α = 6 is much easier to disrupt (in comparison to a network for α = 3).

Further analysis of the Fiedler number (see Table 4-4) revealed a statistically significant difference in mean Fiedler number ($\bar{\lambda}_2$) for the feature sizes (α) studied in this case. At 5% level of significance, all pairwise differences using Tukey's (multiple comparison) test were significant.

Table 4-4: Some descriptive statistics of the Fiedler number measured for the simulated defect types with different sizes (α, with zero variation; 50 data points for each defect type). Defect density is maintained at 3% for each α level.

Feature Size	Mean	Std. dev.	Range	IQR
α = 1	0.8200	0.0089	0.0380	0.0113
α = 2	0.7156	0.0127	0.0524	0.0141
α = 4	0.6951	0.0076	0.0300	0.0114
α = 8	0.6643	0.0083	0.0366	0.0122
α = 16	0.6329	0.0295	0.1458	0.0276
α = 32	0.5960	0.0232	0.1318	0.0305

Number of data points	50 (for each α)
Pooled standard deviation	0.0172
Standard error	0.0024

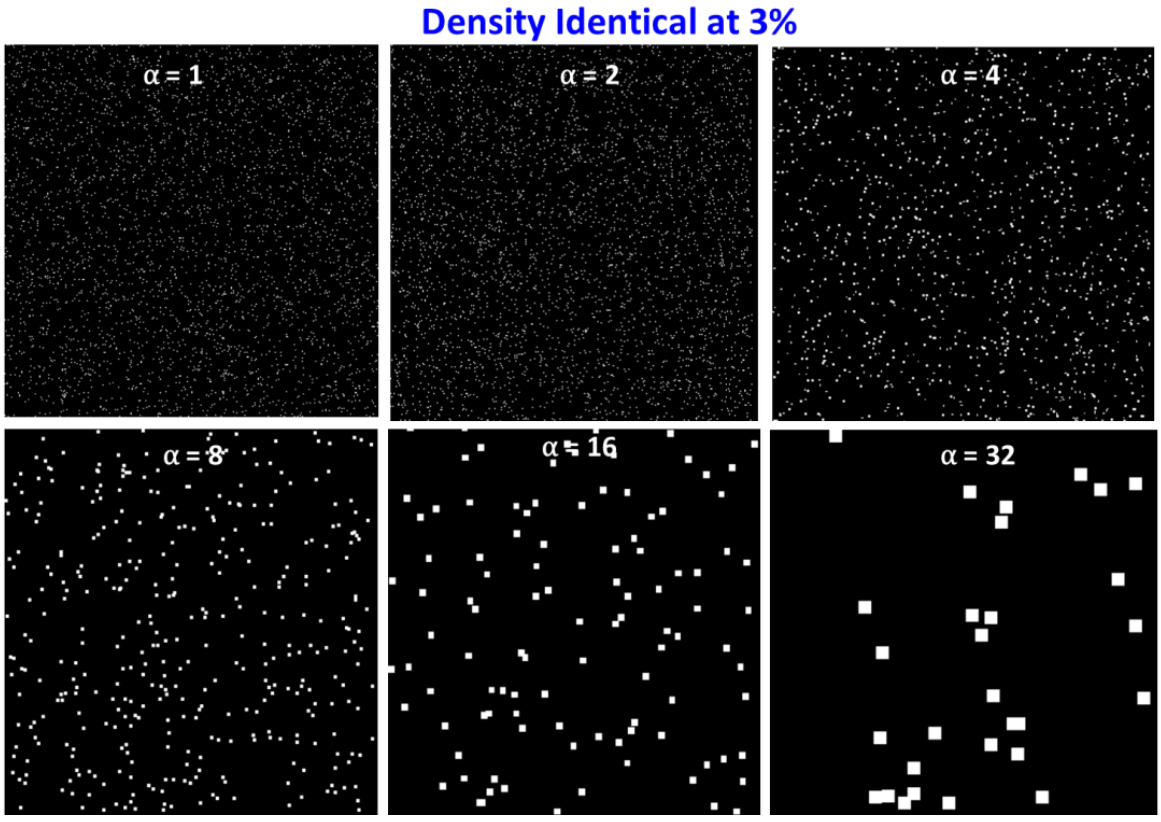


Figure 4-21: Typical surfaces for different defect sizes (α) for the zero variation scenario.

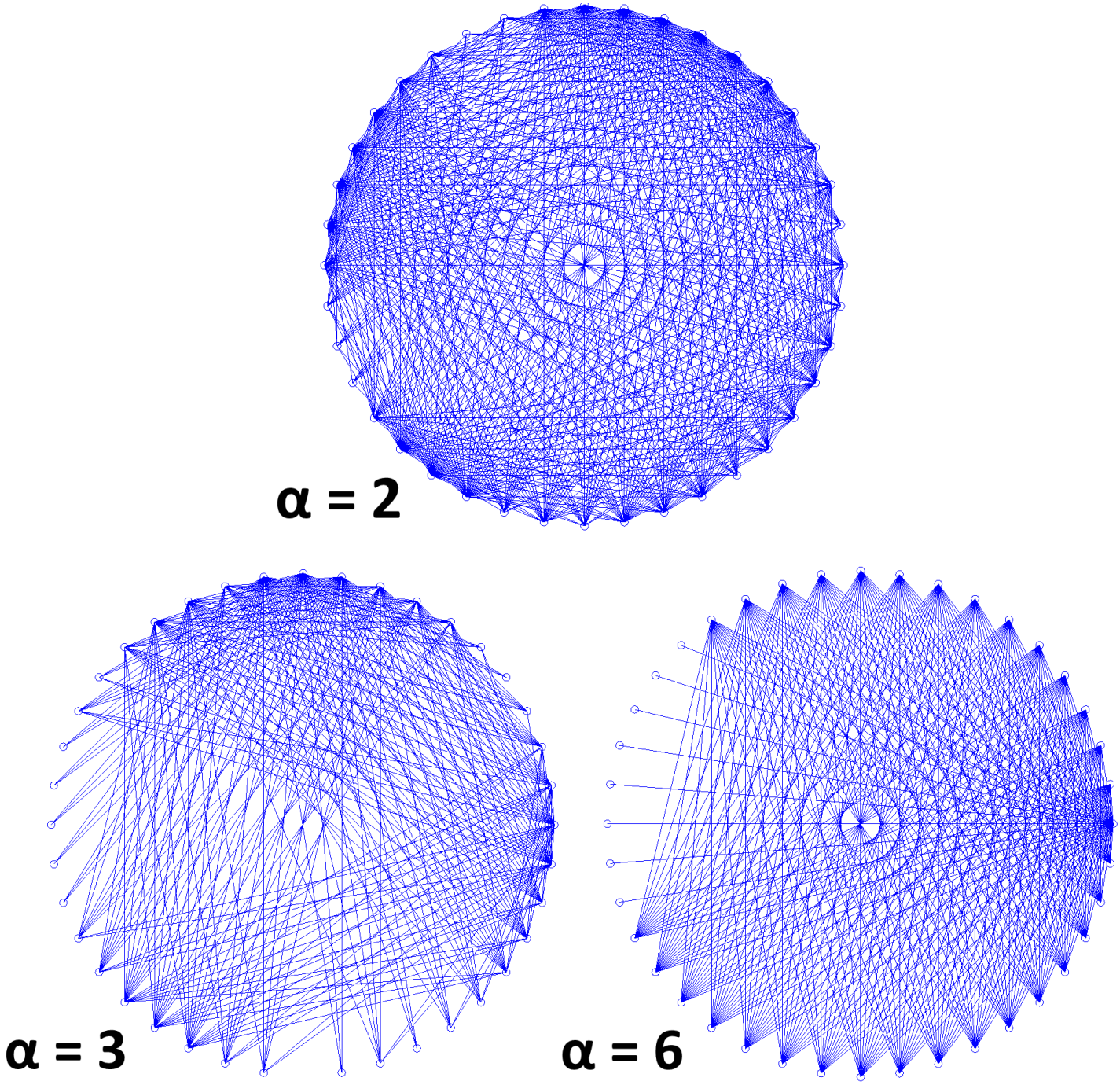


Figure 4-22: Some typical network graphs (forty nodes) for different feature sizes (α).

For α =2 the graph network shows robust connectivity, which reduces as α increases to 6.

Case 3(b): Change in size (α) of the defect with variation ($\sigma^2 \neq 0$)

Compared to the previous case where the mean Fiedler number $\bar{\lambda}_2$ showed a (statistically significant) decreasing trend with increase in defect size (α); in this case where the size of the defect (α) is a (discrete uniform) random variable, the behavior of $\bar{\lambda}_2$ with change in α is not comparatively tractable (see Figure 4-23). As we can observe from Figure 4-23, for the range α = 2 to α = 8, it appears that the Fiedler number fails to register any change (this was verified statistically, see also Table 4-5). An explanation for this behavior can be tendered on the basis of Figure 4-24, which shows the typical surfaces generated for different (mean) levels of feature (defect) size (α).

For instance, examination of the surface for α = 16 reveals that the defects vary significantly in size (this is to be expected, since by the very nature of the simulation we have set the variation, $\sigma^2 = \frac{\alpha^2}{12}$, i.e., larger the defect size, the larger is the variation). That is, we have confounded (by choice) two parameters, namely, mean size of the defect, and variation (σ^2) in the size of the defect α. Therefore, instead of the surface becoming more homogenous with increase in α (as observed in Case 3(a), Sec. 4.5.3), a more complex behavior ($\bar{\lambda}_2$) is observed for this case (Figure 4-23).

With increasing α, the variation in feature size would tend to make the surface more inhomogeneous (implying, larger $\bar{\lambda}_2$), while the increase in the mean size tends to make the surface more homogenous. These opposing trends combine

to make $\bar{\lambda}_2$ almost level for certain ranges of α (α = 2 to α = 8). Stated imprecisely, the increase in variation (σ^2) in α 'pulls' $\bar{\lambda}_2$ 'up', while the increase in mean α 'pushes' $\bar{\lambda}_2$ down, as a result of which the response is flat in some ranges. However, the tails of Figure 4-23 show distinct behaviors. For small α (e.g., α = 1) the effect of mean size is dominant, while for large α (e.g., α = 32) the variation becomes more prominent.

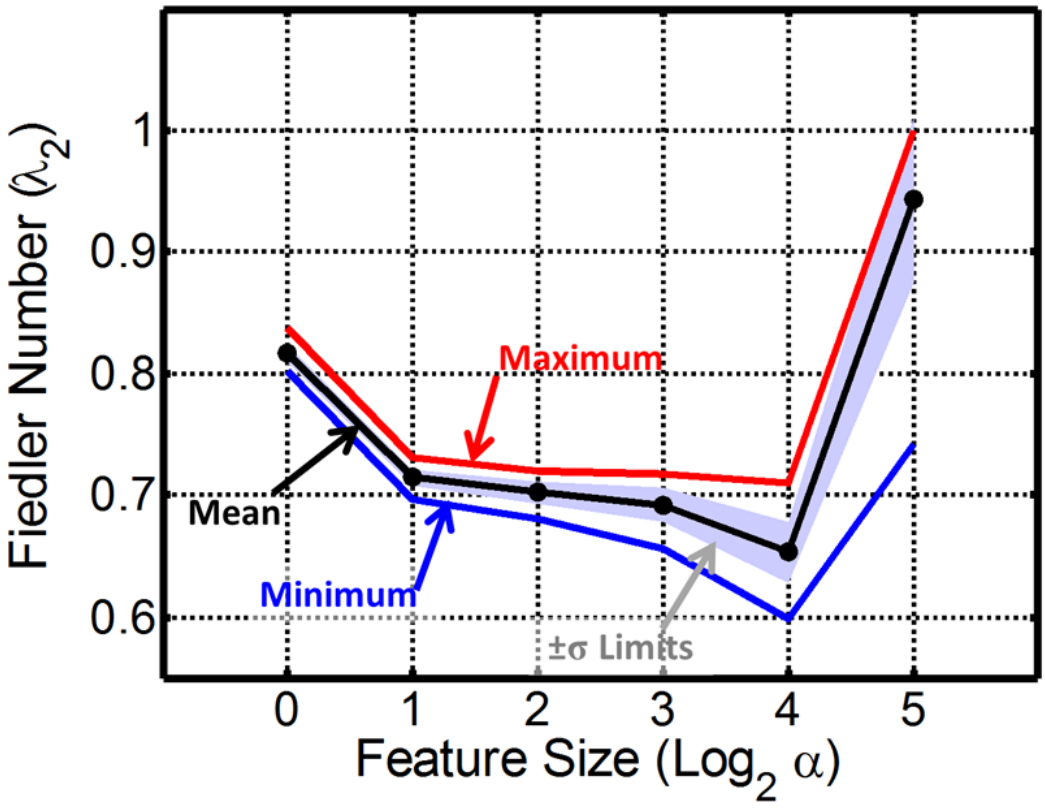


Figure 4-23: Change in mean Fiedler number ($\bar{\lambda}_2$) with mean feature size (α), where α is a uniform random variable .

Table 4-5: Some descriptive statistics of the Fiedler number measured for the simulated defect types with different sizes (α, with (discrete) uniform random variation; 50 data points for each defect type). Defect density is maintained at 3% for each α level

FEATURE SIZE	MEAN	STD. DEV.	RANGE	IQR
$\alpha = 1 + \mathcal{U}(0,1)$	0.817	0.0075	0.0346	0.0113
$\alpha = 2 + \mathcal{U}(0,2)$	0.7145	0.0073	0.0342	0.0103
$\alpha = 4 + \mathcal{U}(0,4)$	0.7024	0.0092	0.0397	0.0103
$\alpha = 8 + \mathcal{U}(0,8)$	0.6923	0.0136	0.0621	0.0201
$\alpha = 16 + \mathcal{U}(0,16)$	0.6536	0.0251	0.1106	0.0316
$\alpha = 32 + \mathcal{U}(0,32)$	0.9429	0.0687	0.2582	0.0911
Number of data points		50 (for each α)		
Pooled standard deviation		0.0309		
Standard error		0.0043		

Though it may be argued that this is a biased scenario, with the mean and variation effects of feature size (purposely) confounded, it is nonetheless of practical importance. Though we can easily separate the two effects when simulating artificial surfaces, we may not have such liberty with experimentally acquired surfaces. One observes that large dimension parts often have much wider tolerances (variance) compared to smaller dimension parts. We point the reader to Table 1-2 [4] in Chapter 1 as evidence in support of this position. Given such confounding and the insensitivity of λ_2 (in scenarios such as this), we wish to seek a better discriminant than the Fiedler number.

Thus far, we have not examined the Fiedler vector v_2 (which is used extensively in image segmentation applications [192]). The Fiedler vectors v_2 for different α from this scenario are shown in Figure 4-25. We observe from Figure 4-25 that the Fiedler vector (v_2) may contain information related to the spatial characteristics of the surface. For example, we perceive that as α increases, v_2 tends to cross the zero line less often. Subsequently, we extracted two statistical

characteristics, termed as, *zero crossing* and *persistence* to study the behavior of the eigenvector v_2.

The *zero crossing* statistic measures the number of instances v_2 crosses zero, i.e., the number of inflexions (or changes in sign) the eigenvector undergoes. The *persistence* statistic, is complementary to *zero crossing*, and measures for how long (in terms of data points) the eigenvector maintains its current position either above or below the center, (i.e., the run length before v_2 changes sign). The zero crossing and persistence statistics for different α levels are presented in Table 4-6, plotted in Figure 4-26. From where (Table 4-6 and Figure 4-26) it is apparent that there is a trend in these statistical measurements with respect to α, namely, the number of zero crossings decreases with α. This corroborates that the initial visual observation that v_2 seems to become less volatile with increase in α while the opposite is true for the persistence statistic. This study leads us to infer the following:

Inferences from the study

1. The Fiedler number can distinguish changes in size of the morphology.
2. The Fiedler vector can also be a valuable tool for distinguishing morphological changes.

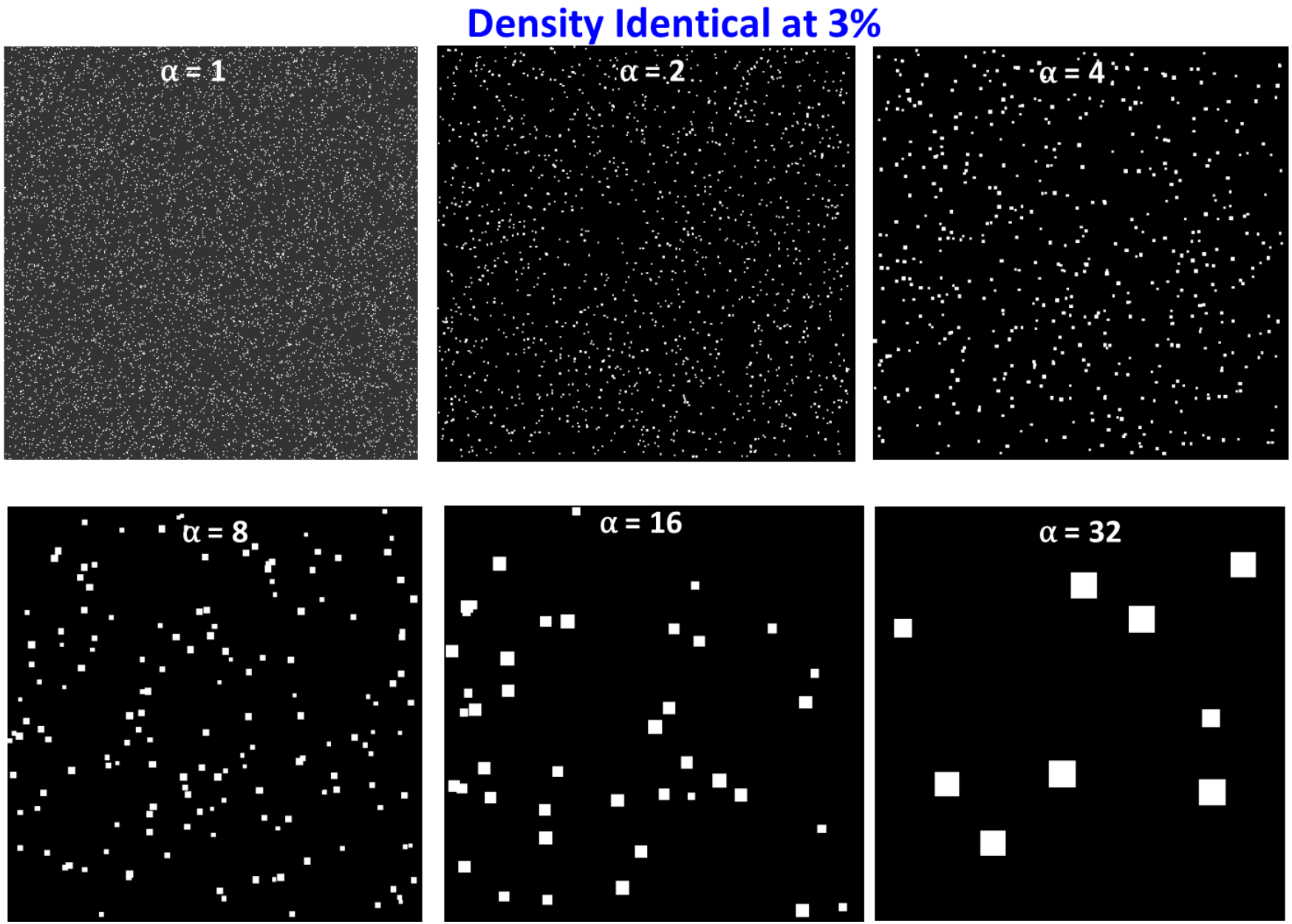


Figure 4-24: Typical surfaces for different mean defect sizes (α) with variation in defect size.

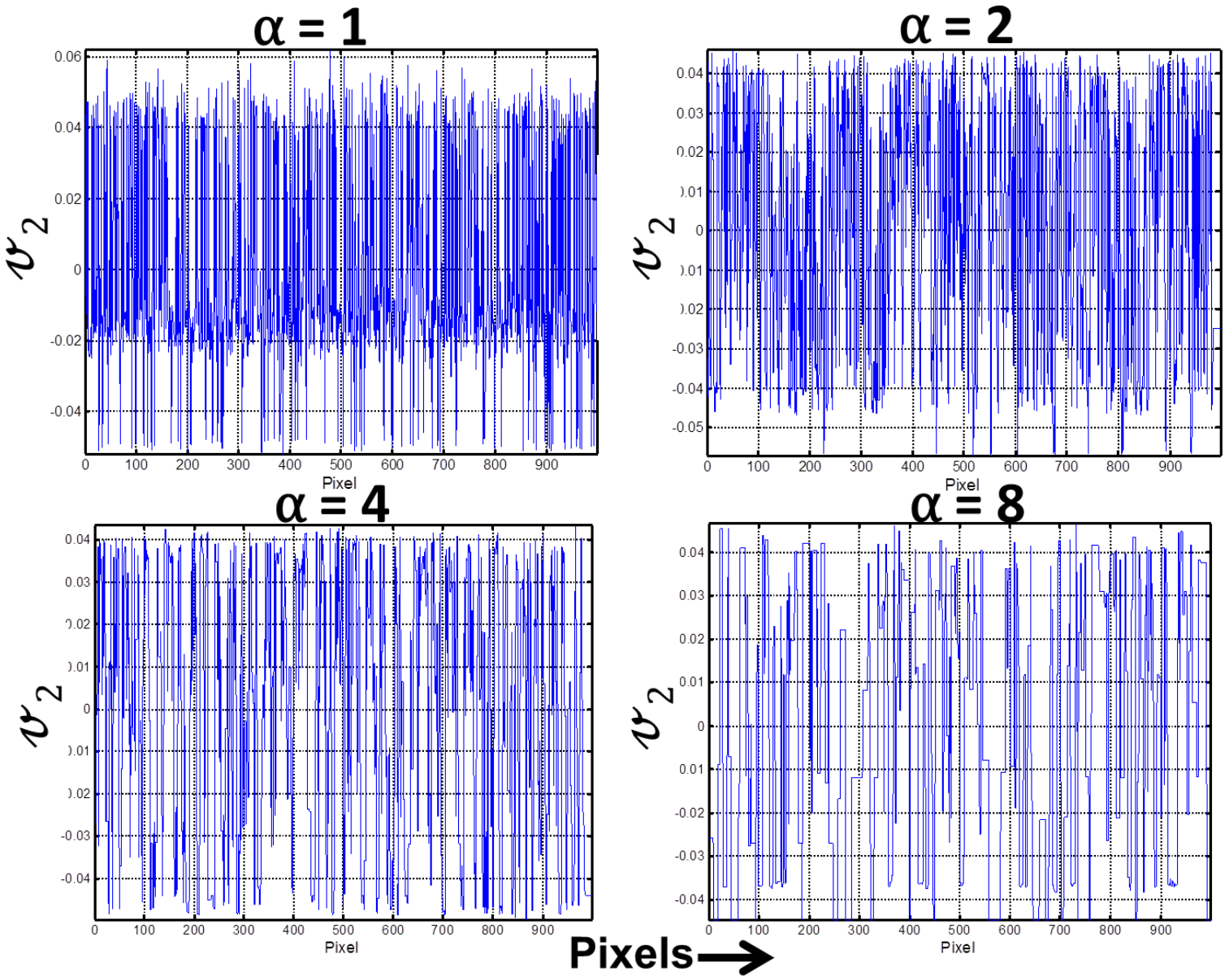


Figure 4-25: Representative Fiedler vectors (v_2) for different levels of α.
We observe that as α increases, v_2 becomes less volatile (there are fewer changes of sign)

Table 4-6: Zero-crossing and persistence statistics measured at different levels of α

Feature Size	Zero Crossings	Persistence
$\alpha = 1 + \mathcal{U}(0,1)$	429	570
$\alpha = 2 + \mathcal{U}(0,2)$	362	637
$\alpha = 4 + \mathcal{U}(0,4)$	188	811
$\alpha = 8 + \mathcal{U}(0,8)$	110	889
$\alpha = 16 + \mathcal{U}(0,16)$	54	945

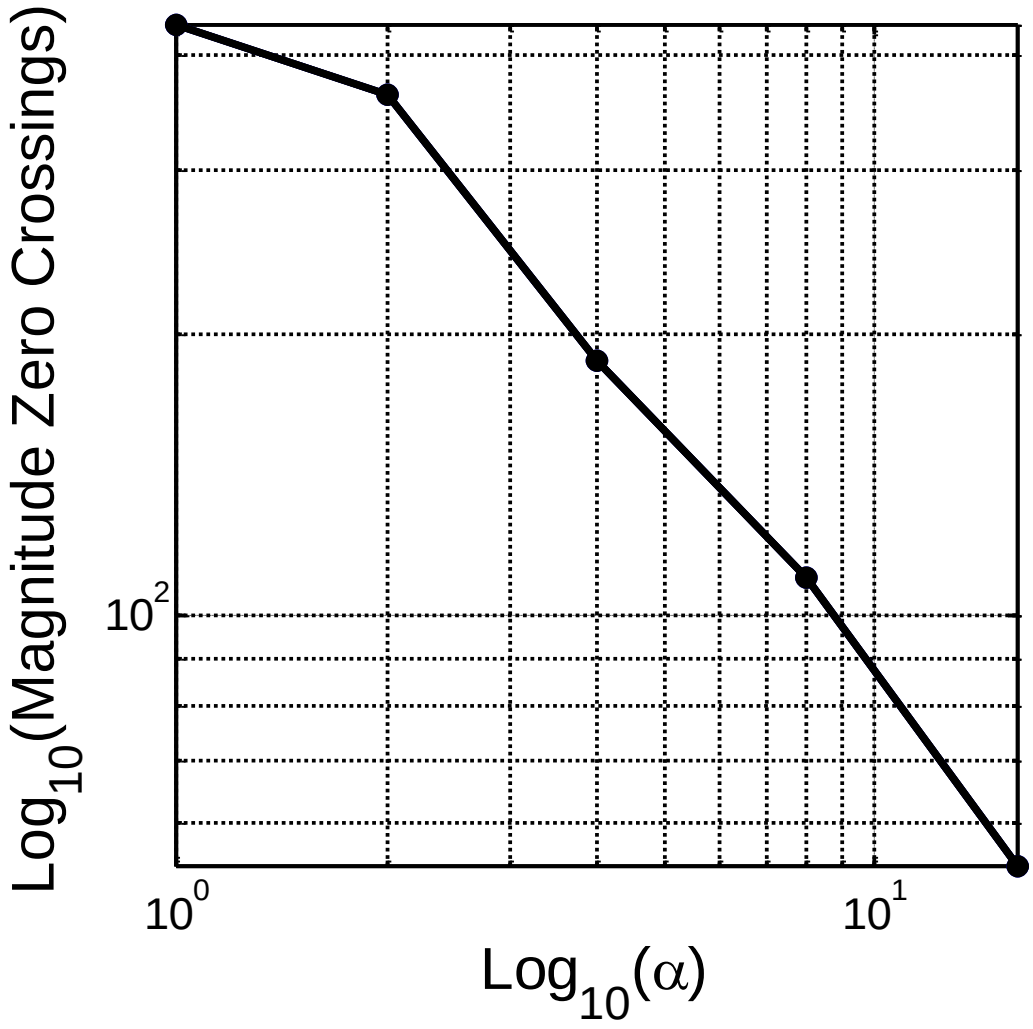


Figure 4-26: Trend of zero crossings vs. feature size (α).

As α increases the number of zero crossings decrease almost linearly (on the log scale)

4.5.4 Case 4: Effect of image properties

Case 4(a): Effect of image size

From a practical standpoint, it can be claimed that the Fiedler number is a parameter that is influenced by the size (resolution) of the image analyzed, and is therefore an unreliable key process output variable for quantifying surface morphology[19]. It is well known that image quality is easily affected by environmental and instrumental factors. Also, sometimes due to optical errors (e.g., from improper focus) some areas of the surface may be occluded. This study addresses these foregoing concerns.

Aim of the study

Demonstrate that the graph theoretic invariant Fiedler number is robust to image size.

In this study, we quantify the sensitivity of the Fiedler number (λ_2) to variation in image size. For this purpose, we generate artificial surfaces (Figure 4-27), of resolution 900 pixel × 900 pixel with multiple heterogeneous features (four in total, see Table 4-7). These features have distinct characteristics with respect to their texture[20] (i.e., hue of the feature), and size. They are uniformly distributed on the surface using the procedure described previously for Case 1 (Sec. 4.5.1) with identical density (3%, for each feature). For example, Figure 4-27 shows a typical

[19] This concern can be countered mathematically as per the reasons tendered in Sec. 4.4.1. We recall that λ_2 is obtained from the normalized Laplacian $\mathcal{L}$, which is scaled based on the degree $\mathcal{D}$, and therefore independent of the image size.

[20] The textures are in gray scale; 1 is white, and 0 is black.

simulated surface, on which the four features listed in Table 4-7 are present. Pertinently, the overall image is a gray scale image (background hue 0.8) as opposed to the binary images used in the studies thus far. Hence, we apply the edge weighted approach detailed in Sec. 4.4.4.

Table 4-7: Scale and texture parameters for the different features simulated for studying the influence of image scale on the Fiedler number.

Size	Texture
$1 + \mathcal{U}(0,1)$	0.40
$2 + \mathcal{U}(0,2)$	0.48
$3 + \mathcal{U}(0,3)$	0.5
$5 + \mathcal{U}(0,5)$	0.55

We generated 138 images of the type shown in Figure 4-27. For each simulated image generated, we measure the Fiedler number at different areal scales, starting from 100 pixel × 100 pixel area (L_1) around the center (shown with a red bordered box in Figure 4-27), 200 pixel × 200 pixel (L_2, green border), and so on, until the whole image is covered (900 pixel × 900 pixel (L_9)).

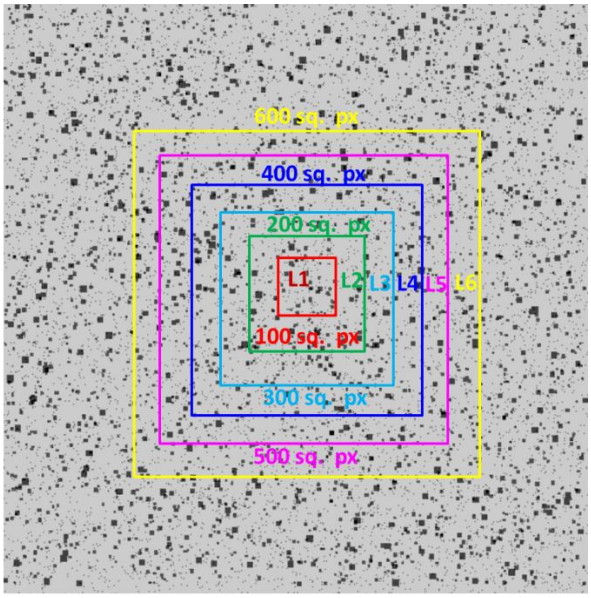


Figure 4-27: Gray-scale surface generated with heterogeneous defects to study the effect of image size on the Fiedler number.

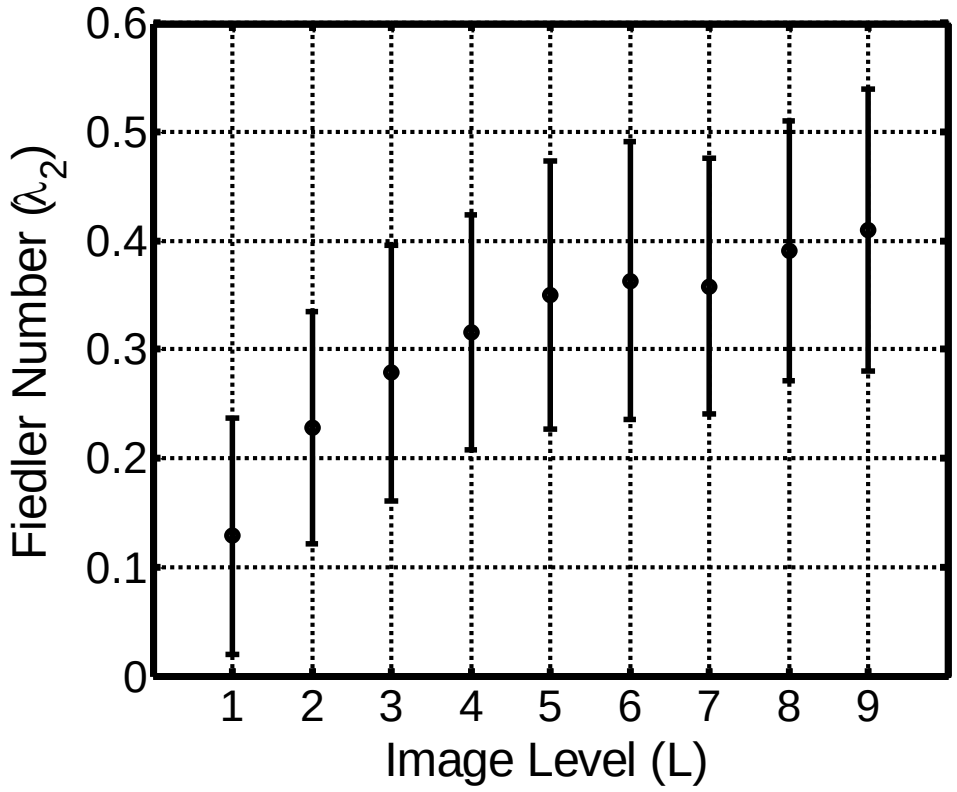


Figure 4-28: Fiedler number measurements at different image levels, the bar is ±1σ long.

The corresponding mean Fiedler number ($\bar{\lambda}_2$) for different image levels (i.e., $(\bar{\lambda}_2|L_k)\ \forall\ k,\ k = \{1,2,\ldots 9\}$) is presented in Figure 4-28 (see also Table 4-8), from which we perceive that above the scale of 500 pixel × 500 pixel (L_5) and beyond, the mean Fiedler number ($\bar{\lambda}_2$) seems to have converged. This observation is further verified using Tukey's pairwise multiple comparison test for different image levels (Figure 4-29), which reveals that the difference in $\bar{\lambda}_2$ for image levels at L_5 and above are (barring one instance) statistically insignificant (α = 5%).

Table 4-8: Descriptive statistics for the Fiedler number (λ_2) at different image levels.

IMAGE LEVEL	PIXEL SIZES	MEAN	STD. DEV	RANGE	IQR
L1	100 sq. pixels	0.1285	0.1091	0.4599	0.1577
L2	200 sq. pixels	0.2281	0.1068	0.5025	0.1498
L3	300 sq. pixels	0.2786	0.1178	0.5813	0.1711
L4	400 sq. pixels	0.3158	0.1076	0.4585	0.1459
L5	500 sq. pixels	0.3503	0.1233	0.5964	0.2062
L6	600 sq. pixels	0.3631	0.1277	0.5525	0.2185
L7	700 sq. pixels	0.3580	0.1175	0.5185	0.1765
L8	800 sq. pixels	0.3904	0.1194	0.6058	0.1693
L9	900 sq. pixels	0.4100	0.1297	0.5367	0.2059
Number of data points		138			
Pooled standard deviation		0.1179			
Standard error		0.0100			

This result has an important practical implication – the Fiedler number (λ_2) is independent of image size beyond a certain threshold, which is approximately 0.3 times the area of the image obtained under normal conditions. From a practical standpoint, this study indicates that the Fiedler number (λ_2) for a homogeneous surface (such as a CMP processed surface) is independent of the image size if the area analyzed is *sufficiently* large (at least greater than third the usual image area).

Hence, even under conditions where an image obtained may be occluded due to optical errors, as long as more than a third of the area of the image is useable, the Fiedler number obtained from such an image will not vary significantly from that of an optimal image. An implication of this is evident from Figure 4-30, which shows the SEM micrograph of an alumina substrate. The original image is about 580 px. × 580 px. Samples were taken at different location on the image, totally 5 samples were taken for each smaller image level. After about 300 px. ×

300 px. image size (about fourth of the total image area), the Fiedler number converges statistically.

Inferences from the study

The Fiedler number (λ_2) is found to be robust to changes in image size. The difference in Fiedler number is statistically insignificant (at 5% level of confidence) for portions that are greater than approximately one third of the area of the original image.

1	2	3	4	5	6	7	8	9
X	1	1	1	1	1	1	1	1
	X	1	1	1	1	1	1	1
		X	0	1	1	1	1	1
			X	0	1	0	1	0
				X	0	0	0	0
					X	0	0	1
						X	0	0
							X	0
								X

Tukey's (95%) Pairwise Distance Test
(H_1=1, H_Φ = 0)

Figure 4-29: Tukey's pairwise comparison test applied to the mean Fiedler number measured at different image levels.

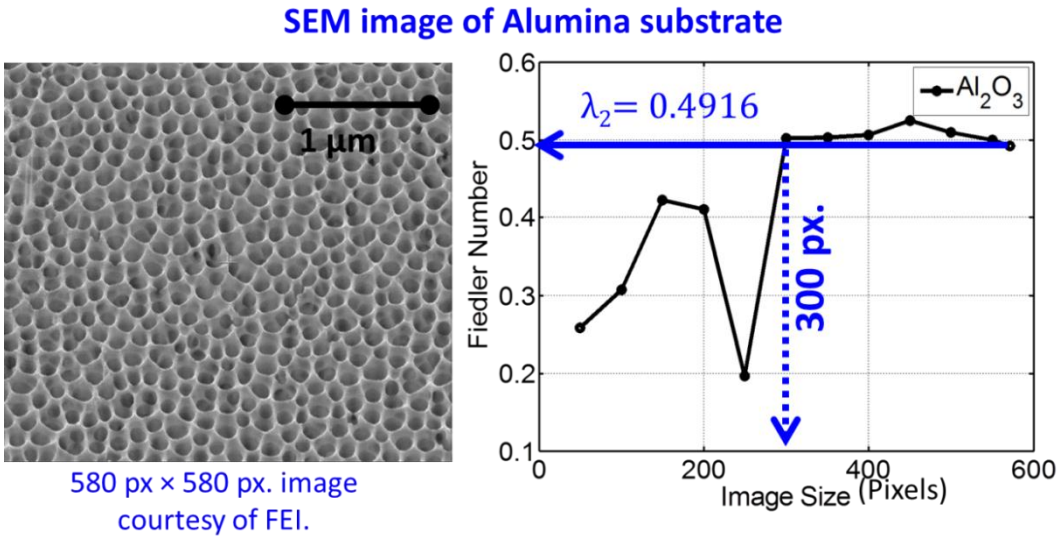


Figure 4-30: Left: Alumina substrate sample (courtesy FEI). Right: Mean Fiedler number at various image sizes.

Case 4(b): Effect of direction of measurement

It is contended that the direction of measurement can also affect the Fiedler number of a surface. As a practical analogy, consider the surface quality of a lapped wafer measured with a stylus-based surface roughness profiler. It is oftentimes noted that the surface roughness measurement will defer (assuming that the surface is homogenous) depending on the direction in which the readings are taken.

This study aims to quantify the difference in Fiedler number due to the direction in which a surface is measured. Thus far, we have used pixel rows for constructing the graph nodes (see Eqn.(4-2)); in this study, we will investigate whether there is a significant difference if pixel columns were instead used (for a homogenous surface).

Aim of the study

Quantify the difference in Fiedler number (λ_2) due to row-wise versus column-wise pixel computations (contingent on the surface being homogenous).

The difference in surface characteristics due to direction of measurement is often inconsequential in surfaces bereft of certain features contingent on the manufacturing process, e.g., polishing/lapping. For instance, for a well-polished CMP surface, it is not easy to identify the axial and radial directions of wafer rotation from the micrograph image.

However, in surfaces generated using processes, such as turning where tool feed marks are evident on the surface; the roughness measurement will depend on the direction in which the stylus of the profiler scans the surface. Under such conditions, it is standard practice to take surface roughness measurements cutting across, i.e., perpendicular to the feed marks. In this study, we use the three types of surfaces simulated in Sec. 4.5.1, Case 1(a). For each type of surface morphology (Figure 4-6) we generate 21 instances, and compute the Fiedler number (λ_2) using both pixel rows and columns for nodes. The results are graphically depicted in Figure 4-31 below, where we notice that the difference between row and column pixel computations of the Fiedler number is negligible.

This assertion is corroborated using a two sample paired t-test. The results from the statistical analysis are presented in Table 4-9, where we notice that the p-value >> 10% for pairwise comparisons between Fiedler number computed with pixel rows and pixel columns. Based on this evidence, we can conclude that given

a homogenous surface, the difference between row-wise and column-wise pixel computations of the Fiedler number are not statistically significant.

Inferences from the study

The Fiedler number (λ_2) is found to be robust to direction of measurement for a homogeneous surface, i.e., the difference between Fiedler number computed using row-wise and column-wise comparisons is not statistically significant.

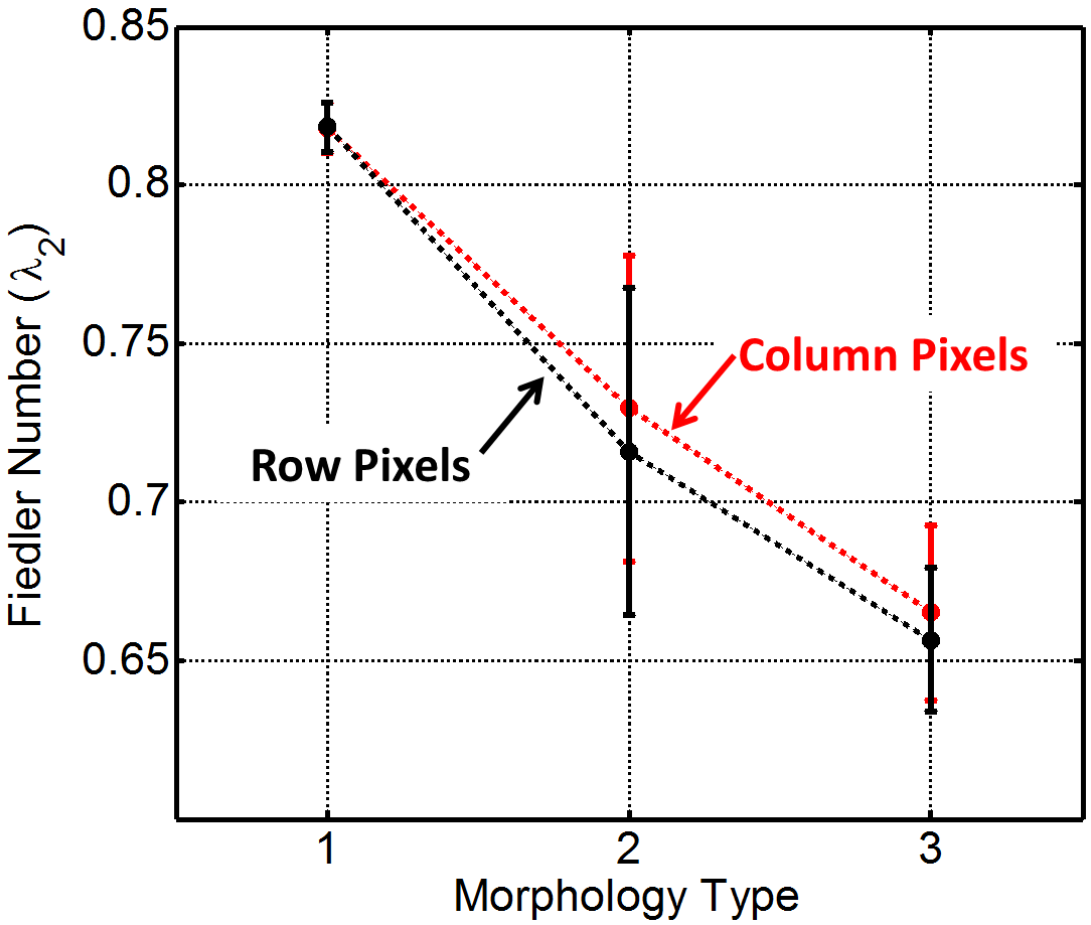


Figure 4-31: Fiedler number for three different surface morphologies taken in the row-wise and column-wise pixel directions.

Table 4-9: Paired t-test results for row-wise and column-wise Fiedler number estimates.

Morphology type	Mean		Std. dev.		Range		IQR	
	Row	Col.	Row	Col.	Row	Col.	Row	Col.
Type 1	0.8184	0.8179	0.0076	0.0079	0.0311	0.0270	0.0103	0.0095
Type 2	0.7156	0.7296	0.0516	0.0483	0.1799	0.1864	0.0908	0.0575
Type 3	0.6565	0.6650	0.0224	0.0273	0.0818	0.0970	0.0266	0.0386

Paired t-test
Difference in Fiedler number between row-wise and column-wise measurement

Morphology type	p-value	t-stat	Confidence interval		Pooled standard deviation
Type 1	0.6874	0.4502	-0.0017	0.0027	0.0048
Type 2	0.2827	-1.1041	-0.0404	0.0124	0.0581
Type 3	0.2492	0.0581	-0.0234	0.0064	0.0328

4.5.5 Case 5: Surfaces with deterministic (non-random) patterns

Case 5(a): Different morphologies

We now illustrate a case in which the features occur in a repeating pattern as opposed to a random manner. Some examples of such surfaces are, patterned semiconductors, nanostructured substrates, micro-channels, aligned carbon nano-tubes, etc. [33]. Two types of binary surfaces of resolution 1000 pixel × 1000 pixel, with feature densities (i.e., proportion of white color patterns) identically maintained at 19% are artificially generated (Figure 4-32). Since the surfaces are completely deterministic (there are no randomly determined locations, densities, or sizes), the need for replicates is precluded.

Aim of the study

Demonstrate that the graph theoretic invariant Fiedler number can capture morphology of deterministic patterned surfaces.

Although, the feature density is the same for both patterns, the Fiedler numbers (λ_2) are significantly different. For pattern A (left side of Figure 4-32) λ_2 is estimated at 1 (this surface has a bipartite graph structure, a scenario we will describe in detail in the forthcoming section (Sec. 4.5.6)). While for pattern B (right side of Figure 4-32), the Fiedler number is 0.89. Thus, the Fiedler number can discern topological differences on patterned surfaces that cannot be captured using statistical parameters such as feature density.

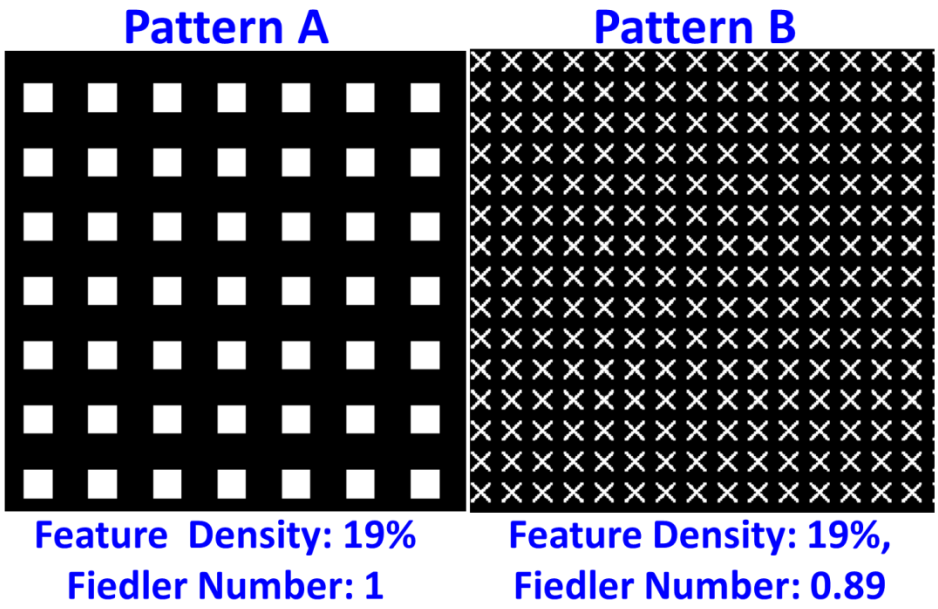


Figure 4-32: Two types of simulated structured surfaces.
Pattern A: Box patterns, Pattern B: X-type pattern.

As an illustrative example using real images, consider the SEM micrographs for various inorganic and organic samples shown in Figure 4-33. It is observed that the Fiedler number is sensitive to the different patterned morphologies.

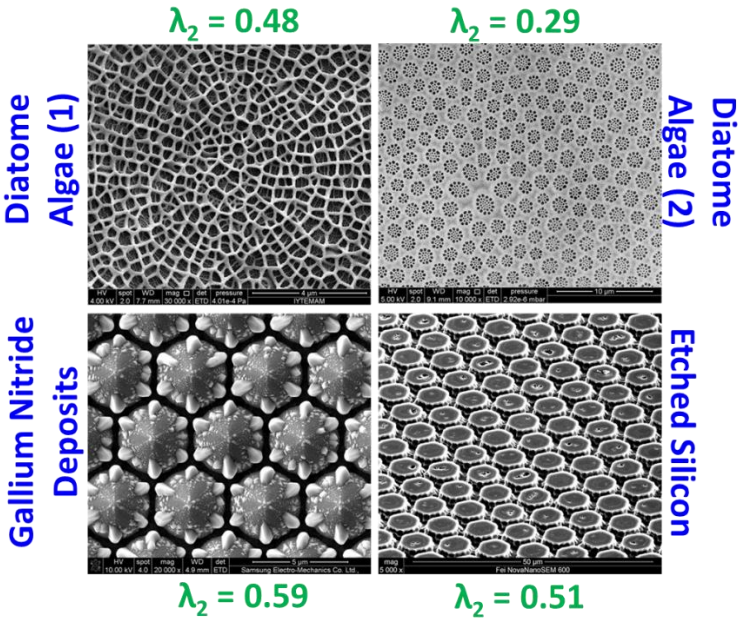


Figure 4-33: SEM micrographs of different organic and inorganic samples (courtesy of FEI).

Case 5(b): Effect of pattern defects

This ability of the Fiedler number (λ_2) to discern differences in surface topology can be valuable from a quality monitoring perspective. Consider Figure 4-34(a), which shows a surface with two types of features embedded. As a practical analogy, we might consider the bottom half of Figure 4-34(a) as a process related defect. For instance, the surface in Figure 4-34(a) could be a nanolithography-processed sample, the bottom half of which has been exposed (or etched) imperfectly due to anomalies with the UV beam traverse system. We estimate the feature density and Fiedler number statistics for different levels (percentage) of defects. In this case, 10% defective would mean 100 rows out of 1000 of the surface (the image resolution is 1000 pixel × 1000 pixel), have been defectively etched.

The resulting feature density (proportion of white colored elements) and Fiedler number plotted against defect density in Figure 4-34(b) and Figure 4-34(c) respectively, reveal that the Fiedler number is more sensitive to the occurrence of defects in comparison to the feature density statistic.

From Figure 4-34, we observe that the feature density:

i. has no discernible trend (it is almost identical for 20% and 40% defect levels), therefore the opportunity to correlate surface quality and feature density are limited, and

ii. has large variation, therefore only statistically tenuous inferences with respect to the surface quality (if any) can be made.

In contrast, the Fiedler number has an unequivocal trend with little variation (except close to the 60% level, because the defects become more dominant over the rest of the surface). In other words, it is possible using an SPC approach that the occurrence of process defects can be detected with smaller Type II (failing to detect) error when the Fielder number is used as a key process output variable compared to popular statistical parameters, such as feature density, which are routinely used for quantification of such binary surfaces.

Inferences from the study

1. The graph theoretic invariant Fiedler number can capture morphology of deterministic patterned surfaces.
2. The Fiedler number is more sensitive to changes in surface morphology in terms of pattern related mistakes compared to conventional parameters, such as defect density.

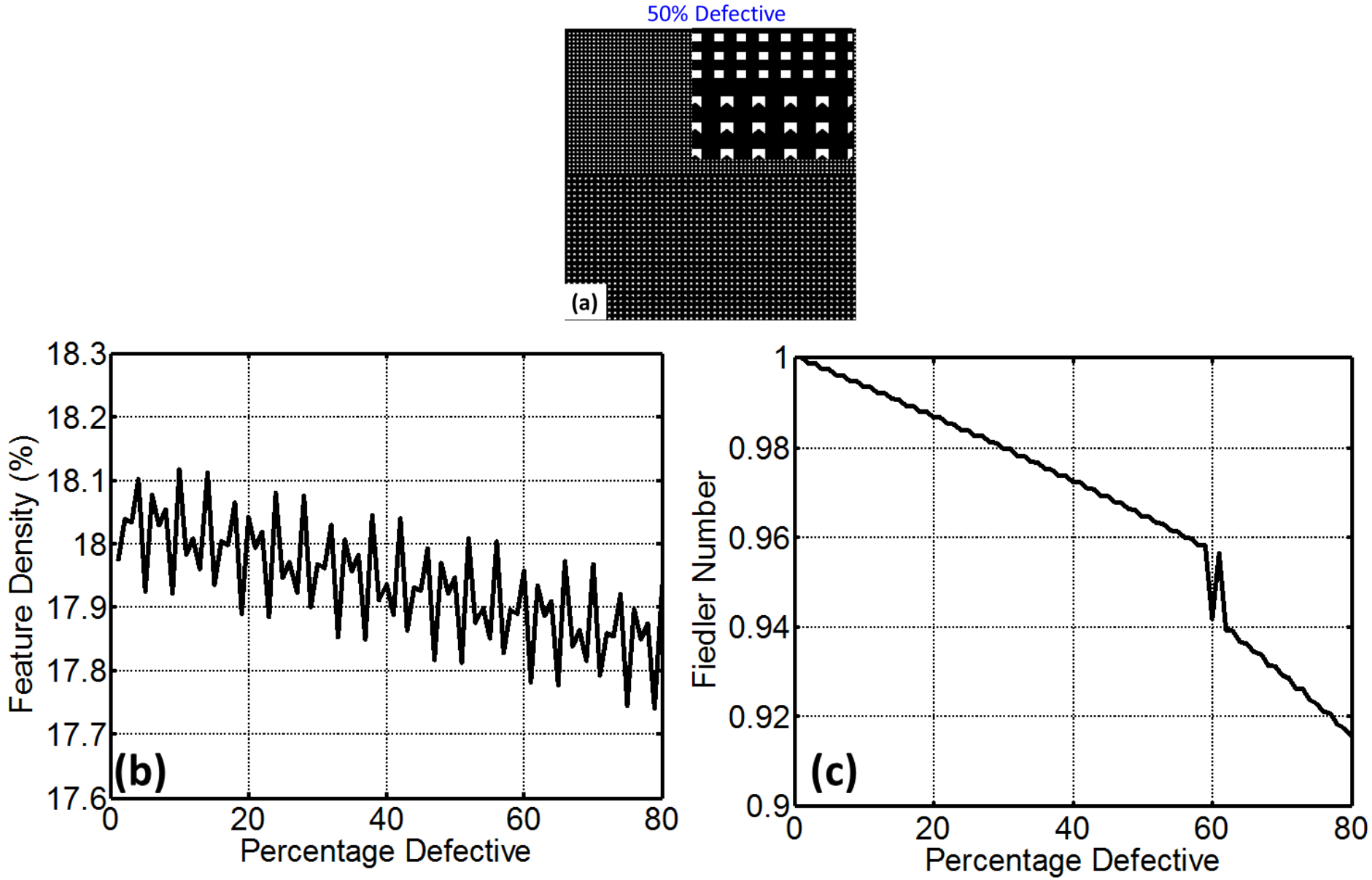


Figure 4-34: Effect of defective patterning.

(a) Pattern A with bottom half incorrectly patterned (50% defective), (b): Behavior of feature density vs. percentage defective, (c): Behavior of Fiedler number vs. percentage defective.

4.5.6 Case 6: A Limiting case – deterministic repeating patterns mapping to bipartite network graphs

If you make a theory, for example, and advertise it, or put it out, then you must also put down all the facts that disagree with it, as well as those that agree with it.
− Richard Feynman, *"Cargo-Cult Science,"* speech at Caltech (1974)

Keeping with the above injunction by Feynman, we present a limiting case for the graph theoretic approach, where the Fiedler number λ_2 is not sensitive to changes in surface morphology. In this case, we study surfaces that have a deterministic repeating morphology of a kind that leads to a special type of graph network, namely, a bipartite graph. We briefly alluded to such graphs in Sec. 4.5.2 and Sec. 4.5.5, where we noticed that under certain circumstances the Fiedler number equals 1 ($\lambda_2 = 1$).

Such scenarios present a limiting case for the graph theoretic approach in quantifying surface morphology, because under bipartite graph conditions the Fiedler number becomes equal to one ($\lambda_2 = 1$) regardless of the morphology of the surface [179, 180, 185]. We will demonstrate that this is an inescapable mathematical consequence[21] [179, 180, 185] (the proverbial brick wall), and not a chance simulation-based outcome.

Therefore, the Fiedler number **cannot** be used as a discriminant for surface morphology when the graph network representation (of the surface) becomes

[21] This result has been proved in literature, we are only demonstrating the mathematical implications, and not proving a theorem.

bipartite, although, there are ways around the problem, which we will explore towards the end of this section.

Aim of the Study

1. Explore limiting cases for the graph theoretic approach, using the bipartite graph as a demonstrative example, where $\lambda_2 = 1$.
2. Suggest modifications to the approach whereby the above limitation can be overcome.

Procedure

The surfaces simulated (Figure 4-35) in this study are strictly deterministic. In that, the sizes of the features, their locations, and densities do not vary. Moreover, the features occur at regular intervals, separated by a specific distance. Nine different such surfaces (total area 1024 pixel × 1024 pixel) are investigated, six of which are shown in Figure 4-35. These surfaces are labeled β_1 through β_9. The subscript is of significance, and denotes the distance between the features (as usual colored white and 1 pixel × 1 pixel in area).

For instances, 6 pixel × 6 pixel portions of two example surfaces β_1 and β_2 (of the nine simulated) are shown in Figure 4-36, from which we notice that the Manhattan distance between neighboring features is 2^n where, n is the subscript of the test image. Therefore, the Manhattan distance between features for image β_1 is equal to 2 ($d(\beta_1) = 2^1 = 2$), for β_2 the Manhattan is equal to 4, ($d(\beta_2) = 2^2 = 4$), and so on.

We notice that the choice of feature distance is a dyadic sequence. The total area of each image β_n is fixed at 1024 pixel × 1024 pixel. This is done so that the total image is exactly divisible by the feature distance, which eliminates any eventuality of 'edge effects' from the image size. This precaution will be consequential in computation of the isoperimetric number (Cheeger's number, $h(G)$, see Sec. 4.4.2) for the various images.

We analyze the structure of the network graph, alongside the examination the various matrices (similarity S, degree $\mathcal{D}$, combinatorial Laplacian L, and normalized Laplacian $\mathcal{L}$), along with the eigen spectra μ^* and λ^* (of L and $\mathcal{L}$, respectively) for two cases β_1 and β_2. From these we generalize the results to all instances through β_9. In addition, since the images are binary in nature, we will use the ε neighborhood approach for analysis (Sec. 4.4.3) of the surfaces (β_n, $n = \{1,2 \dots ,9\}$) in this study.

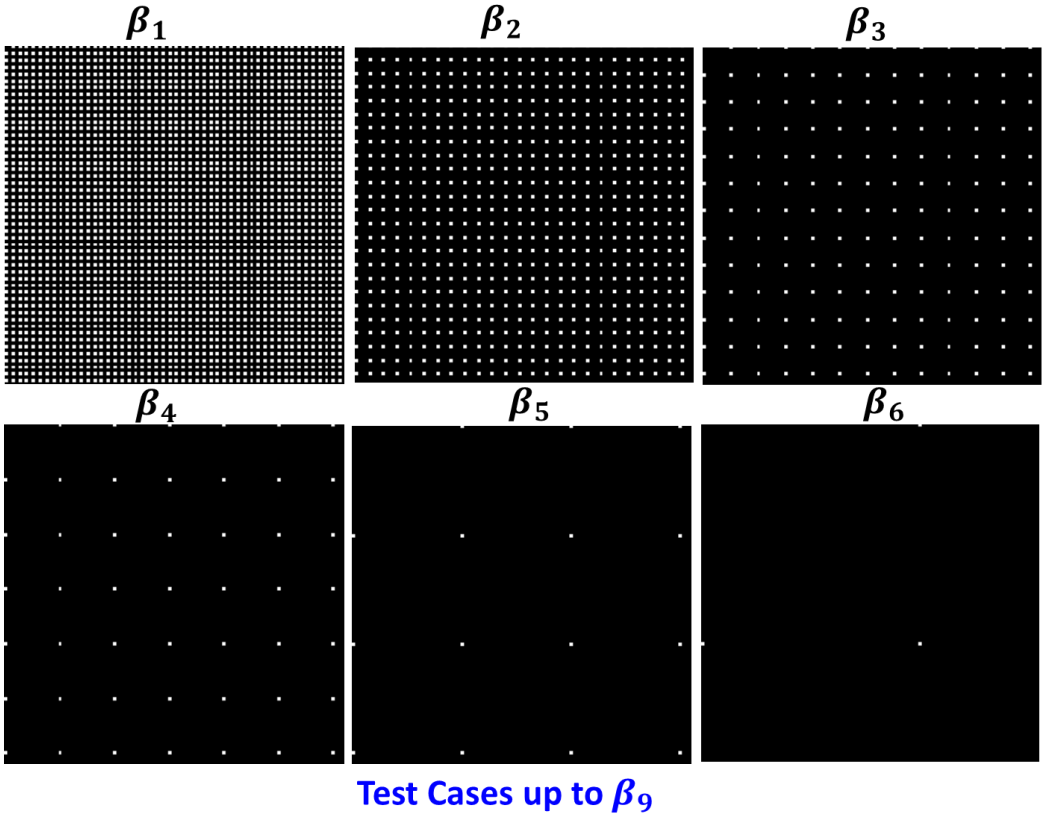


Figure 4-35: A 100 pixel × 100 pixel section of the surfaces (total 1024 pixel × 1024 pixel) simulated for testing bipartite graph structures.

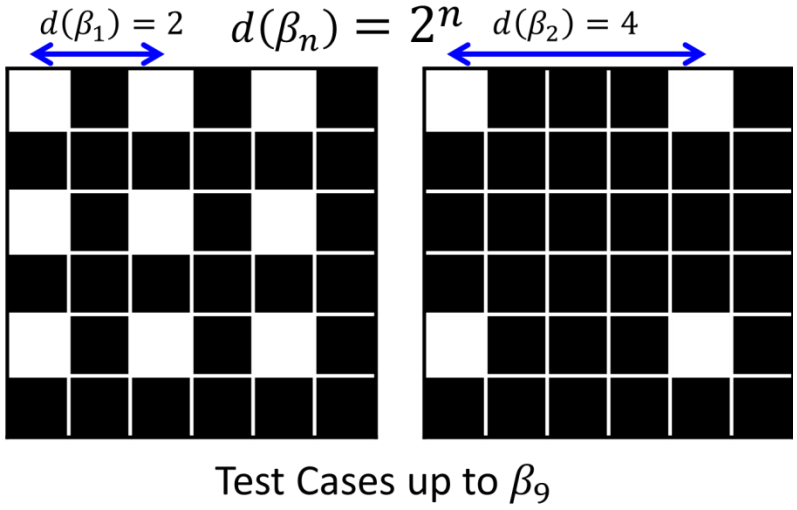


Figure 4-36: Two cases β_1, β_2 of the surfaces simulated for this study.

These are 6 pixel × 6 pixel sections of the total 1024 pixel × 1024 pixel area of the surface. The Manhattan distance between features (white colored) is β^n.

Analysis of surface type β_1

A 6 pixel × 6 pixel section of β_1 (out of 1024 pixel × 1024 pixel total area) is shown in Figure 4-36. As in previous cases, the white colored pixels (features) are coded 1, and the black colored pixels are coded 0 (see Figure 4-37). We recognize from Figure 4-37 that the alternate rows of the image β_1 are identical, i.e., row 1 is identical to rows 3 and 5; while row 2 is identical to rows 4 and 6.

The Euclidean norm between row 1 and row 2 is 1.732, and between row 1 and 3 is 0. Indeed, all (pairwise) distinct pixel rows will be 1.732 in Euclidean measure, whereas the Euclidean distance between identical pixel rows is zero. The average of all row-wise Euclidean distances (taken pairwise) is estimated at 0.8660. We also recall, that in graph theoretic mapping $\mathcal{I} \mapsto G \equiv (V, E)$ used in this work, each row is taken to represent a node of the graph $G \equiv (V, E)$.

Consequently, on applying the threshold function in Eqn. (4-27), if any (pairwise) pixel rows have a Euclidean norm greater than 0.8660, then the corresponding nodes are connected with an edge. Which implies, there is an edge connecting node 1 and node 2 (which correspond to row 1 and 2, respectively), but no edge connections exist between node 1 and node 3. The resulting similarity matrix S for the image β_1 is shown in Figure 4-37.

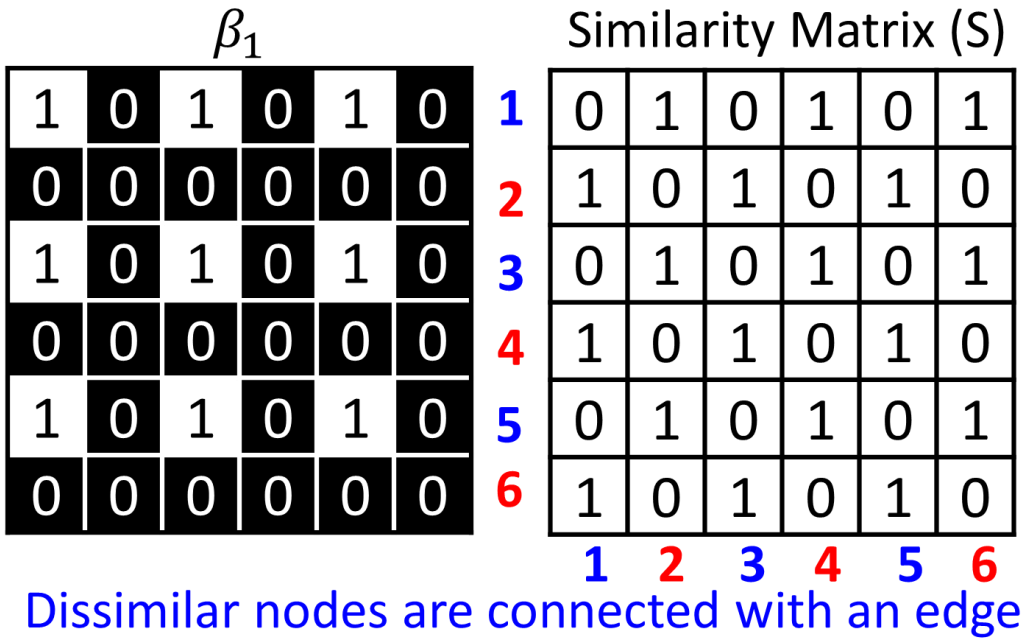


Figure 4-37: Coding for image β_1, and the resulting similarity matrix S.

Examination of the similarity matrix S (Figure 4-37) can be useful in a visual context. Since nodes 1 and 2 are connected (we observe, $S(1,2) = 1$), while row 1 and 3 are not ($S(1,3) = 0$), and so on, we can draw the graph network in Figure 4-38. This graph is a typical bipartite graph (also called 2-colorable). All nodes (in the graph) can be assigned to either of two sets, for didactic reasons we will call these sets *Red* and *Blue*, with the nodes in each set colored likewise. For instance, the nodes 1, 3 and 5 are grouped in the Blue set, while nodes 2, 4 and 6 are grouped in the Red set (Figure 4-38).

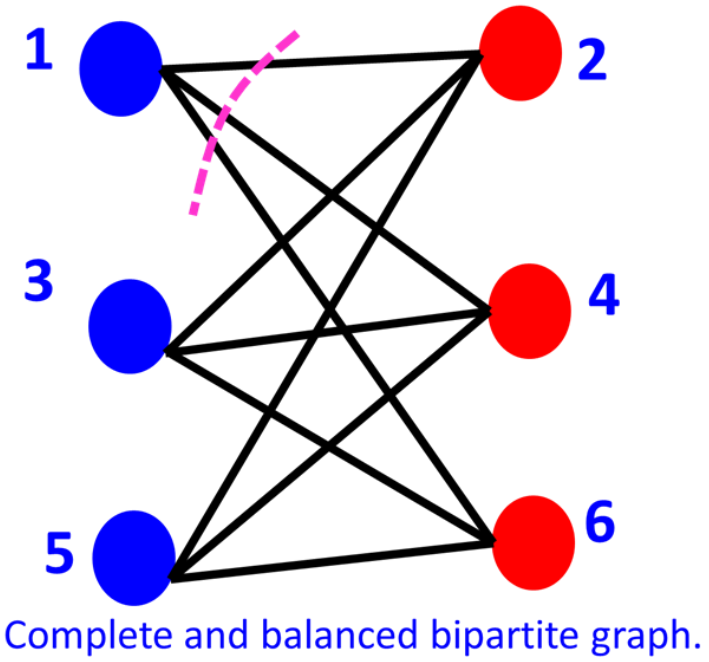


Figure 4-38: Graph network for image β_1,

Notice the graph is 2-colorable or bipartite, i.e., only two 'colors' (i.e., the nodes are either colored red or blue) are sufficient to separate the graph into two subsets. All blue colored nodes are connected to all the red colored nodes.

Using the similarity matrix S in Figure 4-37, we can draw the node connections, as shown in Figure 4-38. For example, the Blue node 1 is connected to the Red nodes 2, 4, and 6, (see the first row of the similarity matrix S, which is also color indexed). Continuing in this manner, we notice that edges exist only between nodes in differently colored sets; nodes in the same set are not connected – Blue nodes are only connected to Red nodes. This means, the graph can be exactly partitioned into two sub-graphs using two colors, hence the term *2-colorable*. The network graph for β_1 is shown in Figure 4-41, and as expected, the bipartite structure is evident.

We summarize the following salient characteristics from the graph for $\beta_{1,}$ referring to Figure 4-38:

i. Nodes are assigned to either of exactly two sets, in our case these are 'colored' Red and Blue. The Euclidean distance between nodes in the same set is equal to zero.

ii. Nodes in the same set are not connected. Edge connections exist only between nodes in different sets; i.e., Blue nodes only connect with Red nodes, and not among themselves.

iii. A node from one set connects to all nodes in the other. A Blue node is connected to all Red nodes.

iv. Both (Red and Blue) sub-graphs have exactly the same number of elements (for the small section considered, three elements each).

Properties (i) and (ii) imply the graph network is bipartite. Properties (iii) and (iv) ensure that the bipartite graph is **complete and balanced**, respectively.

An additional observation can be made from Figure 4-38; the minimum edges to be severed to isolate a node for the network in Figure 4-38 is exactly equal to 3 (note the pink line in Figure 4-38). That is, the Cheeger number equals 3 ($\mathrm{h}(G) = 3$, see Sec. 4.4.2). These implications can be generalized to the entire 1024 pixel × 1024 pixel area of β_1.

Since every alternate row of β_1 is identical, each set, Blue and Red, will have exactly 512 elements. Taking an element (viz. a node) in a set, e.g., a node *k* from set Blue, we know that node *k* will be connected to all 512 nodes in set Red. The

same reasoning can be extended to all nodes in the graph. Every node will be connected to 512 other nodes (from a differently colored set). Thus in order to separate any node from the graph, exactly 512 edges will have to be severed. In other words, the Cheeger number for β_1 (taking the entire 1024 × 1024 area), equals 512, therefore for β_1, $\mathrm{h}(G) = 512$.

Degree Matrix D

3	0	0	0	0	0
0	3	0	0	0	0
0	0	3	0	0	0
0	0	0	3	0	0
0	0	0	0	3	0
0	0	0	0	0	3

Combinatorial Laplacian
L = D − S

3	-1	0	-1	0	-1
-1	3	-1	0	-1	0
0	-1	3	-1	0	-1
-1	0	-1	3	-1	0
0	-1	0	-1	3	-1
-1	0	-1	0	-1	3

Figure 4-39: The degree matrix and combinatorial Laplacian matrix for image β_1

The degree matrix $\mathcal{D}$, and combinatorial Laplacian L for β_1 are shown in Figure 4-39 (for the 6 pixel × 6 pixel section only), from the latter we can extract the non-normalized eigenvalues μ^*. These eigenvalues of L (μ^*) are [0, 3, 3, 3, 3, 6], arranged in the ascending order. We notice that $\mu_1 = 0$, non-normalized Fiedler number $\mu_2 = 3$, and $\mu_6 = 6$. The first eigenvalue equals zero, the second equals 3 (note also $\mathrm{h}(G) = 3$), and the last eigenvalue μ_6 equals 6 (i.e., $2\mathrm{h}(G)$). These are in keeping with well-known properties of bipartite graphs [181, 188]. We will explain the reason for this eigen sequence in further depth following the analysis

for β_2. If we extend the analysis to entire the 1024 pixel × 1024 pixel area, it is easy to see that, $\mu_1 = 0$, $\mu_2 = 512$, and $\mu_{1024} = 1024$.

Normalized Laplacian

$$\mathcal{L} = \mathrm{D}^{-\frac{1}{2}}(\mathrm{D} - \mathrm{S})\mathrm{D}^{-\frac{1}{2}}$$

1	$-\frac{1}{3}$	0	$-\frac{1}{3}$	0	$-\frac{1}{3}$
$-\frac{1}{3}$	1	$-\frac{1}{3}$	0	$-\frac{1}{3}$	0
0	$-\frac{1}{3}$	1	$-\frac{1}{3}$	0	$-\frac{1}{3}$
$-\frac{1}{3}$	0	$-\frac{1}{3}$	1	$-\frac{1}{3}$	0
0	$-\frac{1}{3}$	0	$-\frac{1}{3}$	1	$-\frac{1}{3}$
$-\frac{1}{3}$	0	$-\frac{1}{3}$	0	$-\frac{1}{3}$	1

Figure 4-40: The normalized laplacian $\mathcal{L}$ for β_1.

Using the degree matrix $\mathcal{D}$ and the combinatorial Laplacian L, we obtain the normalized Laplacian $\mathcal{L} = \mathcal{D}^{-\frac{1}{2}} \times L \times \mathcal{D}^{-\frac{1}{2}}$ (see Eqn. (4-9)), shown in Figure 4-40. The eigen spectrum λ^* is [0, 1, 1, 1, 1, 2]. We notice, $\lambda_1 = 0$, the Fiedler number $\lambda_2 = 1$, and $\lambda_6 = 2$. As expected [179, 180, 185], $0 \leq \lambda^* \leq 2$, with $\lambda_2 = 1$ (because the graph is bipartite), and $\lambda_6 = 2$ for the same reason. This eigen spectrum when the analysis is extended to the entire 1024 pixel × 1024 area is $\lambda_1 = 0$ (occurring once); $\lambda_2 = 1$ (repeated 1022 times); and $\lambda_{N=1024} = 2$ (occurring once).

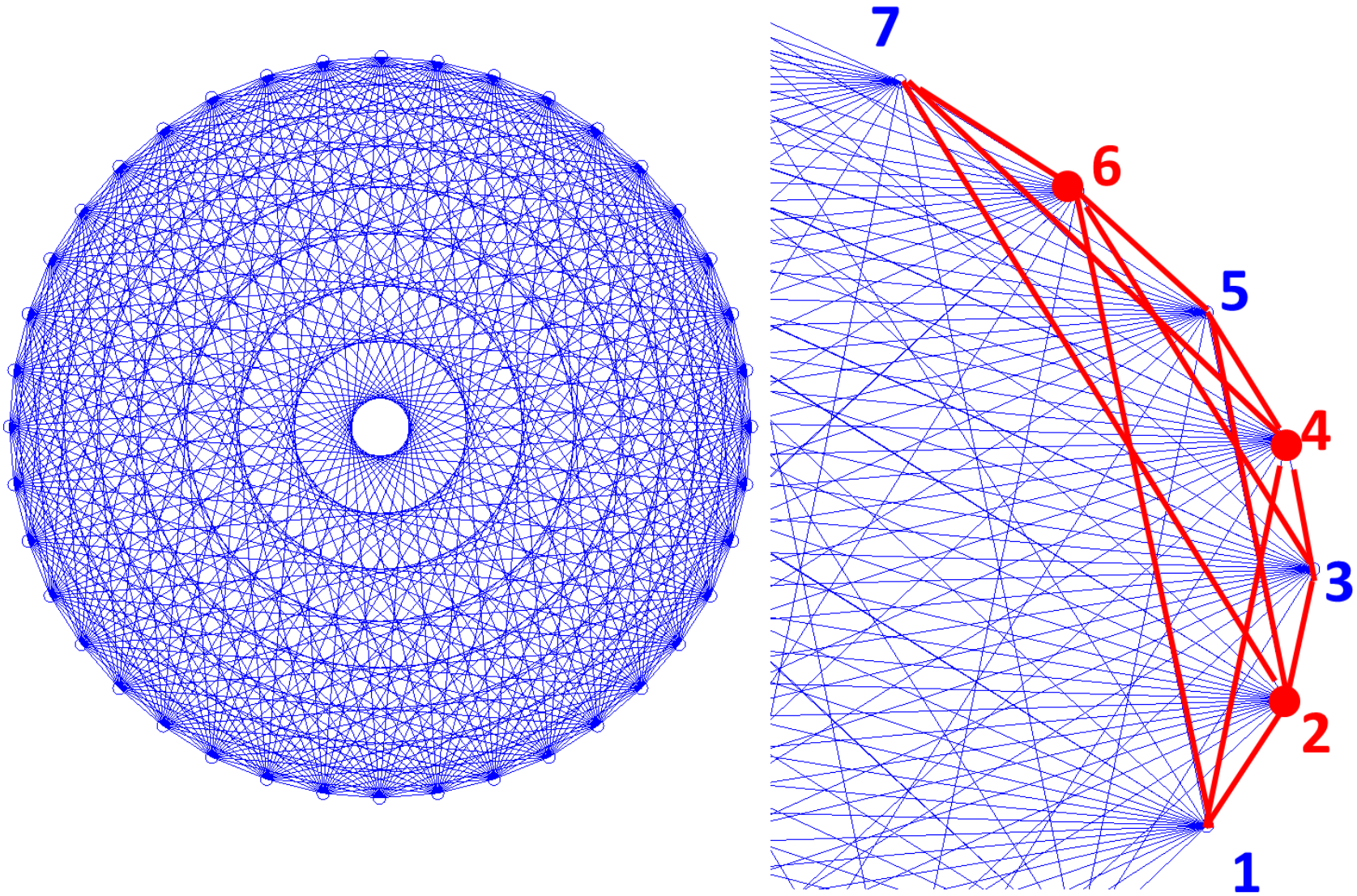


Figure 4-41: Network graph (40 nodes only) for image type β_1 (left), and zoomed-in portion showing the bipartite structure.

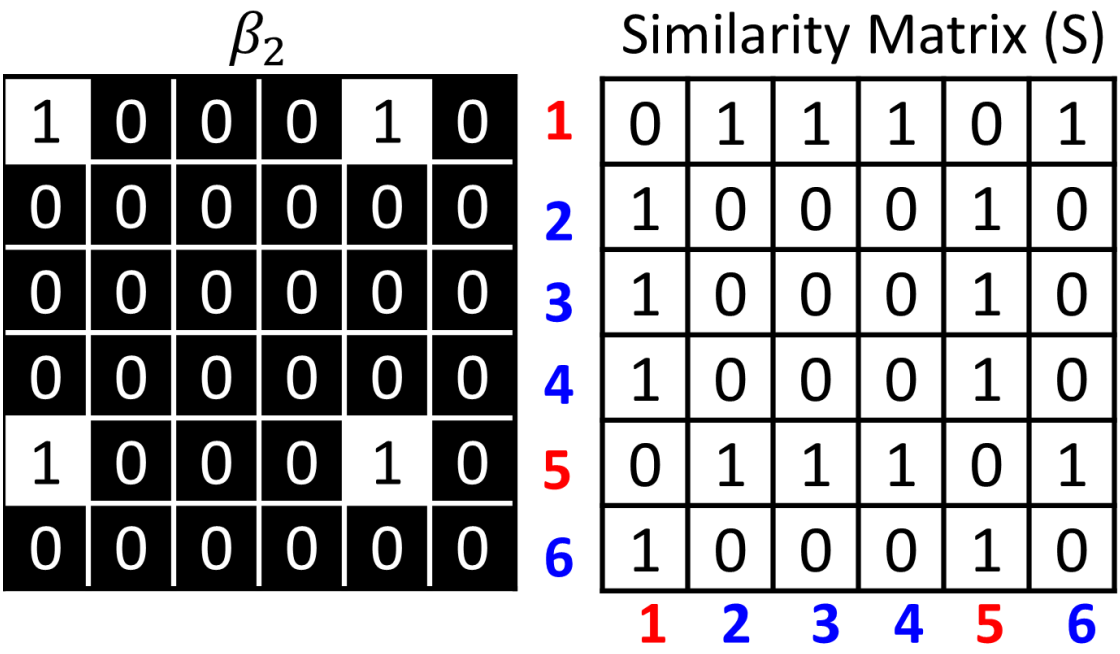


Figure 4-42: Coding for image β_2, and the resulting similarity matrix S.

Analysis of surface type β_2

Continuing with our analysis with surface type β_2, and following similar reasoning as before; we code the image and obtain the similarity matrix (for a 6 pixel × 6 pixel region) as shown in Figure 4-42. We notice that rows 1 and 5 of the image β_2 are identical (so are rows 2, 3, 4, and 6). The Euclidean norm between distinct rows is estimated at 1.4142, and the overall average of all pairwise distances is 0.7071; therefore the threshold r equals 0.7071 in Eqn. (4-27).

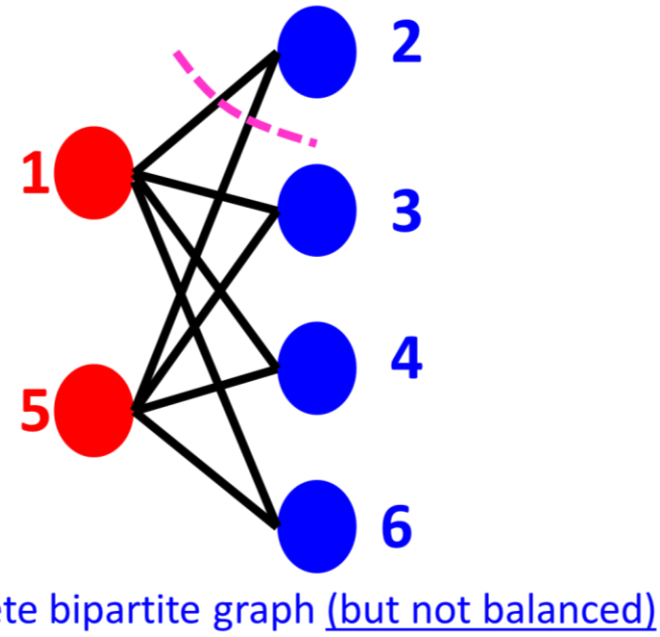


Figure 4-43: Graph network for image β_2.

Notice the graph is 2-colorable or bipartite, i.e., only two 'colors' (i.e., the nodes are either colored red or blue) are sufficient to separate the graph into two subsets. All blue colored nodes are connected to all the red colored nodes.

As result, the similarity matrix S takes the form shown in Figure 4-42, with the graph network shown in Figure 4-43. Using the set notation established previously for the analysis of β_1, we assign nodes 1 and 5 to the Red set (because node 1 and 5 are identical), and likewise nodes 2, 3, 4, and 6 to the Blue set. Subsequently, the following salient features are evident from the network graph for β_2:

i. As before, nodes belong to either of exactly two sets, 'colored' Red and Blue. The Euclidean distance between nodes in the same set equals zero.

ii. Nodes in the same set are not connected. Edge connections exist only between nodes in different sets; i.e., Blue nodes only connect with Red nodes and not among themselves.

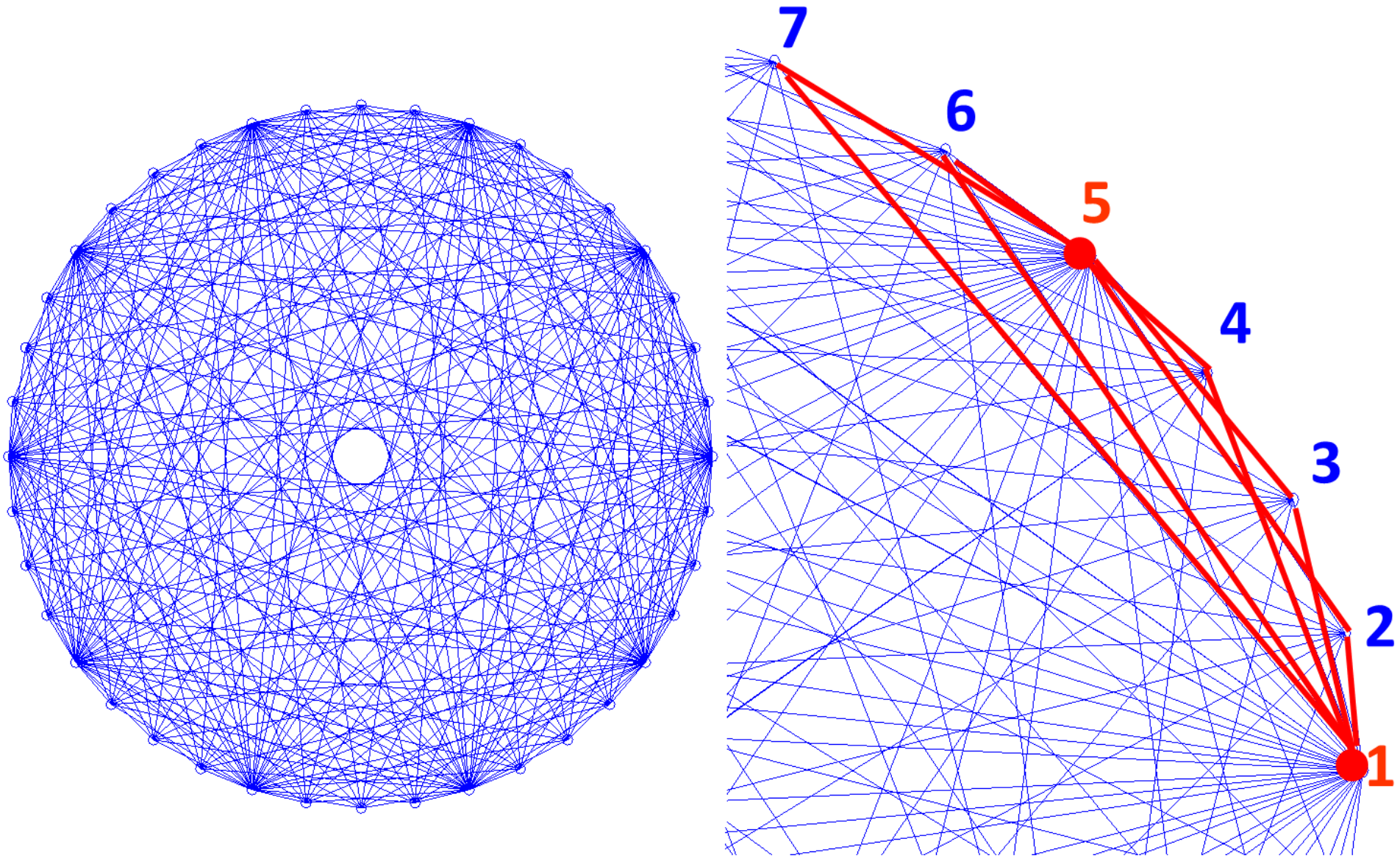


Figure 4-44: Network graph (40 nodes only) for image type β_2 (left), and zoomed-in portion showing the bipartite structure

iii. A node from one set connects to all nodes in the other. A Blue node is connected to all Red nodes.

iv. Both (Red and Blue) sub-graphs do not have the same number of elements. Therefore, the bipartite graph is not balanced.

These conditions imply that the graph is a complete (but not balanced) bipartite graph. The network graph for β_2 is shown in Figure 4-44, from which the bipartite structure is evident. We observe from Figure 4-43 that the Red set has less number of nodes.

The following additional interesting characteristics are noted for β_2 on referring to Figure 4-45:

v. Since every fourth row of β_2 is identical (as opposed to every alternate row for β_1), one of the sets (or sub-graphs) has fewer number of nodes. We will color the set with fewer nodes as Red. Nodes $\{1,5,9,...\}$ belong to the Red set, and the rest of the nodes, $\{2,3,4,6,7,8 ... \}$ are assigned to the Blue set.

vi. Due to the bipartite graph for β_2 being complete but not balanced, the nodes in the two sets have different degrees. Whereas, in case of β_1 all nodes had the same degree (equal to 3), the degree of a node for β_2 depends on the set (Red or Blue) in which a particular node is assigned. For nodes in the Red set (i.e., set with fewer nodes), the degree is equal to 4 (for the 6 pixel × 6 pixel region), while for nodes in

the Blue set the node degree is equal to 2. Hence, the maximum degree (d_{max}) is equal to the cardinality of the set with the larger number of nodes, while the minimum degree (d_{min}) is equal to the cardinality of the set with smaller number of nodes. These observations are annotated in Figure 4-45.

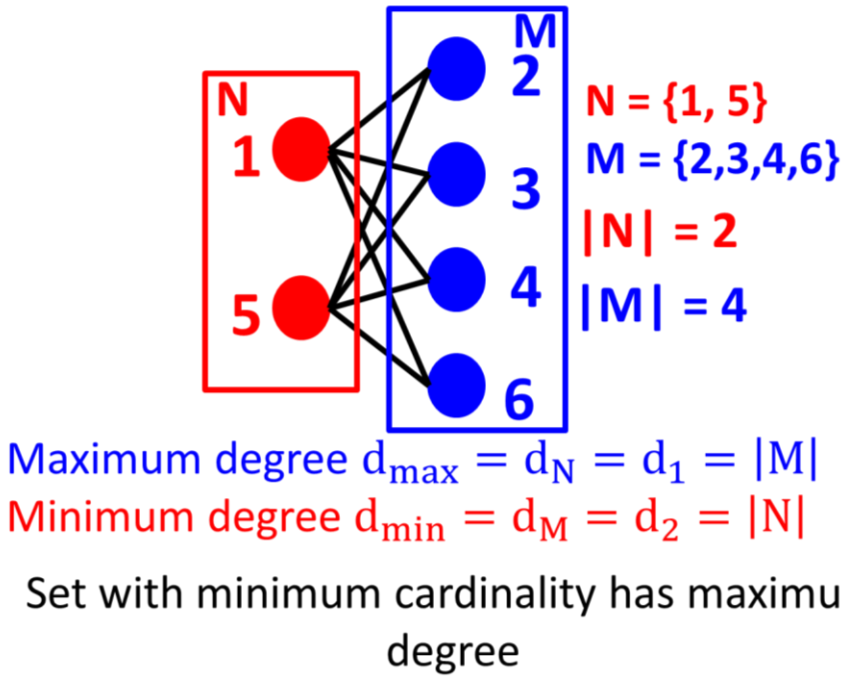


Figure 4-45: Additional observations based on node degree and set coloring for image β_2.

vii. Since the graph is a complete bipartite graph, we can estimate the Cheeger number ($h(G)$) based on the minimum degree (d_{min}). For the 6 pixel × 6 pixel region, the minimum degree (d_{min}) is equal to 2 (nodes in the Blue set). Therefore, $h(G) = d_{min} = 2$ for this case, as the severing of 2 edges from any node in the Blue set is sufficient to isolate the node (break the graph).

viii. On generalizing these observations for the entire 1024 pixel × 1024 pixel area, the Red set will contain 256 nodes (every 4th row in 1024 rows is identical), whereas the Blue has 768 (= 1024 – 256) nodes. The nodes in the Red set have the maximum degree, viz. 768 (each node in the Red set is connected to all the nodes in the Blue set), the nodes in the Blue set have the minimum degree (256). Therefore, for β_2, $\mathrm{h}(G) = 256$.

Degree Matrix D

1	4	0	0	0	0	0
2	0	2	0	0	0	0
3	0	0	2	0	0	0
4	0	0	0	2	0	0
5	0	0	0	0	4	0
6	0	0	0	0	0	2

Combinatorial Laplacian
$\mathrm{L} = \mathrm{D} - \mathrm{S}$

4	-1	-1	-1	0	-1
-1	2	0	0	-1	0
-1	0	2	0	-1	0
-1	0	0	2	-1	0
0	-1	-1	-1	4	-1
-1	0	0	0	-1	2

Figure 4-46: The degree matrix and combinatorial Laplacian matrix for β_2.

Next, we compute the degree matrix $\mathcal{D}$ and the combinatorial Laplacian L for the 6 pixel × 6 pixel region of β_2, are shown in Figure 4-46. From which we notice that the degree for nodes 1 and 5 (first and fifth row of $\mathcal{D}$ in Figure 4-46) are equal to 4 (= d_{max}), while the degree for the rest of the nodes is equal to 2 (= d_{min}) agreeing with our previous observations.

The (non-normalized) eigenvalues μ^* computed from L are [0, 2, 2, 2, 4, 6] arranged in the ascending order. Implying, $\mu_1 = 0$; Fiedler number $\mu_2 = \mathrm{h}(G) = \mathrm{d}_{\min} = 2$; $\mu_3 = \mathrm{d}_{\max} = 4$; and $\mu_{\mathrm{N}=6} = \mathrm{d}_{\min} + \mathrm{d}_{\max} = 6$. Based on these non-normalized eigenvalues (μ^*) the following additional observations can be made:

ix. As expected, the Cheeger number $\mathrm{h}(G)$ is equal to the minimum degree $\mathrm{d}_{\min}$.

x. The minimum degree ($\mathrm{d}_{\min}$) appears $|\mathrm{M}| - 1$ times as an eigenvalue in the eigen spectrum μ^* of L. Where, $|\mathrm{M}|$ is the cardinality of the set with more nodes, i.e., Blue set.

xi. The maximum degree ($\mathrm{d}_{\max}$) appears $|\mathrm{N}| - 1$ times as an eigenvalue in the eigen spectrum μ^* of L. Where $|\mathrm{N}|$ is the cardinality of the set with fewer nodes, i.e., Red set.

xii. These are results are readily extensible to the entire 1024 pixel × 1024 pixel area of β_2. It can be verified that $\mu_1 = 0$ (occurring once); Fiedler number $\mu_2 = \mathrm{h}(G) = \mathrm{d}_{\min} = 256$ (repeated 767 times); $\mu_3 = \mathrm{d}_{\max} = 1024 - 256 = 768$ (repeated 255 times); and $\mu_{\mathrm{N}=1024} = \mathrm{d}_{\min} + \mathrm{d}_{\max} = 1024$ (occurring once). The Cheeger number $\mathrm{h}(G) = \frac{1024}{2^2}$ for β_2.

xiii. These results can be induced over all images β_1 through β_9 because all the images studied form complete bipartite graphs. They have the same structure as β_2, and share the same properties. We note that although β_1 is a complete and balanced bipartite graph, these observa-

tions hold nonetheless (for β_1), because, β_1 is a special symmetric case of β_2 through β_9.

xiv. The Cheeger number $h(G)$ for all cases β_1 through β_9 (and beyond) is $\left\lceil \frac{K}{2^n} \right\rceil$, where n is the subscript of the corresponding surface β, and K is the size of the image in pixels along one dimension. For this study K = 1024; the relationship implicitly assumes the image is square (see Table 4-10).

The network graphs for some of the cases tested in this study are shown in Figure 4-47. It is evident on examination of results presented in Table 4-10 along with the various network graphs of Figure 4-47 that as the distance between features increases, the network becomes less well connected (while maintaining the bipartite structure).

Table 4-10: Various properties for images β_1 through β_9.

n	2^n	$h(g)=\left\lceil \frac{K}{2^n} \right\rceil$ (Calculated) K = 1024	Fiedler Number (Non-normalized) computed μ_2	Fiedler Number Normalized computed λ_2
1	2	512	512	1
2	4	256	256	1
3	8	128	128	1
4	16	64	64	1
5	32	32	32	1
6	64	16	16	1
7	128	8	8	1
8	256	4	4	1
9	512	2	2	1

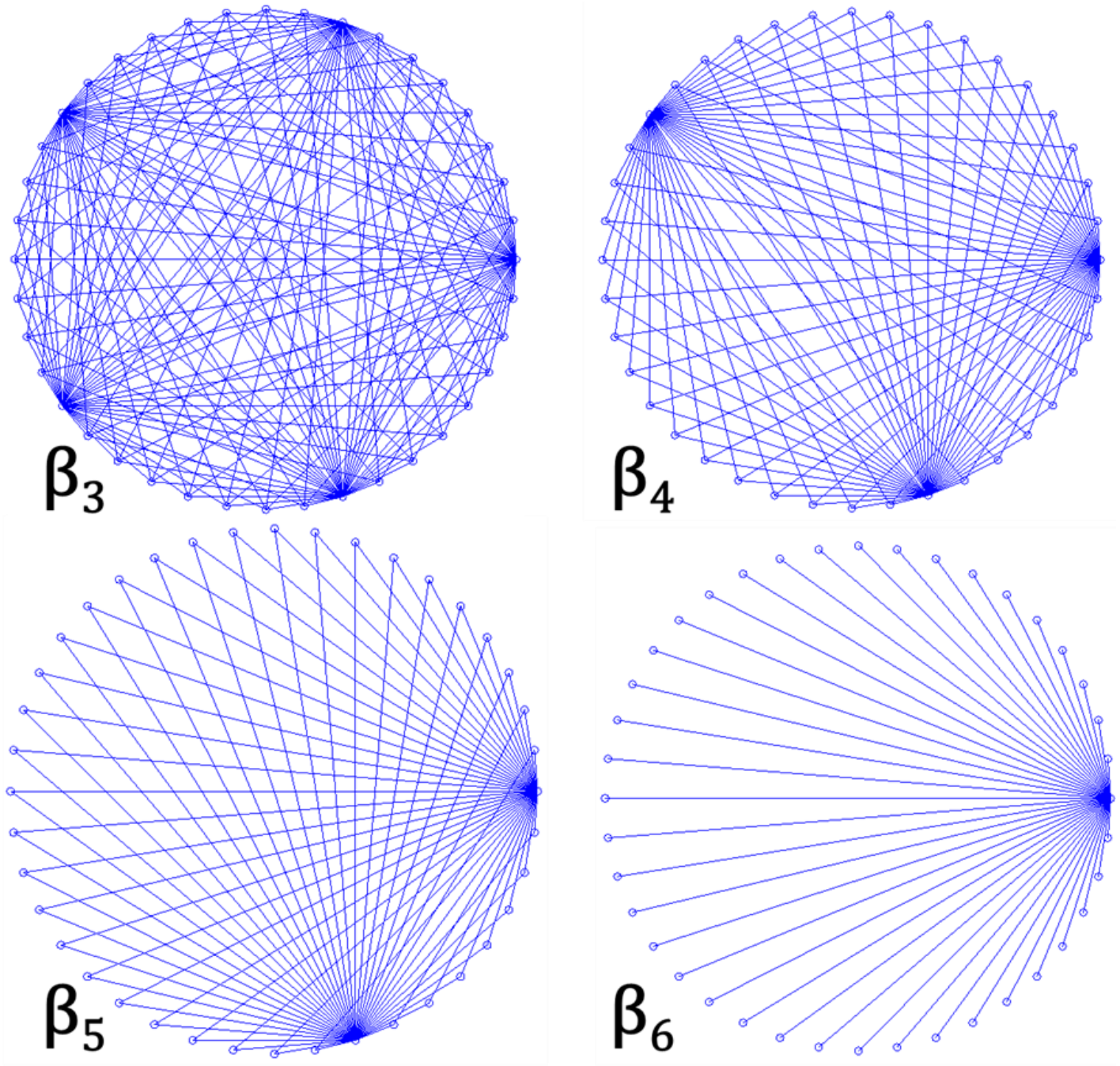


Figure 4-47: The network graphs for some tested cases.

The network graph is representative of 40 nodes. Notice the decrease in connectivity as the distance between features increases.

Normalized Laplacian

$$\mathcal{L} = \mathrm{D}^{-\frac{1}{2}}(\mathrm{D} - \mathrm{S})\mathrm{D}^{-\frac{1}{2}}$$

1	$-\frac{1}{\sqrt{d_1 d_2 = 8}}$	$-\frac{1}{\sqrt{8}}$	$-\frac{1}{\sqrt{8}}$	0	$-\frac{1}{\sqrt{8}}$
$-\frac{1}{\sqrt{8}}$	1	0	0	$-\frac{1}{\sqrt{8}}$	0
$-\frac{1}{\sqrt{8}}$	0	1	0	$-\frac{1}{\sqrt{8}}$	0
$-\frac{1}{\sqrt{8}}$	0	0	1	$-\frac{1}{\sqrt{8}}$	0
0	$-\frac{1}{\sqrt{8}}$	$-\frac{1}{\sqrt{8}}$	$-\frac{1}{\sqrt{8}}$	1	$-\frac{1}{\sqrt{8}}$
$-\frac{1}{\sqrt{8}}$	0	0	0	$-\frac{1}{\sqrt{8}}$	1

Figure 4-48: Normalized Laplacian matrix $\mathcal{L}$ for β_2

The normalized Laplacian $\mathcal{L}$ for β_2 is shown in Figure 4-48. The eigen spectrum λ^* is [0, 1, 1, 1, 1, 2]. As before, $\lambda_1 = 0$, the Fiedler number $\lambda_2 = 1$, and $\lambda_6 = 2$. Additionally, the non-zero elements of $\mathcal{L}$ (barring the diagonal, which is 1) are $-\frac{1}{\sqrt{d_{min} \times d_{max}}}$ which for image β_2 are $-\frac{1}{\sqrt{2\times 4}} = -\frac{1}{\sqrt{8}}$. Additionally, the non-diagonal, non-zero elements correspond to the non-zero elements of the similarity matrix S (see Figure 4-42, therefore, $\mathcal{L} = -\frac{1}{\sqrt{d_{min} \times d_{max}}} S + \mathrm{I}$).

The above relationship is easily generalized for all such situations.

$$\mathcal{L} = \frac{-1}{\sqrt{h(G) \times d_{max}}} S + \mathrm{I}$$

$$\mathcal{L} = \frac{-1}{\sqrt{\left\lceil \frac{K}{2^n} \right\rceil \times \left(K - \left\lceil \frac{K}{2^n} \right\rceil\right)}} S + \mathrm{I} \tag{4-32}$$

Overcoming the bipartite graph limitation

This study confirms that the deterministic repeating pattern of types β_1 through β_9 have a bipartite graph structure. The Fiedler number λ_2 for such types of surfaces equals 1 (see Table 4-10). Therefore, in such scenarios, the Fiedler number (λ_2) is unable to discriminate between different surface morphologies. Given that deterministic repeating patterns represent an important class of nano-surfaces, it is vital to resolve this limitation. There are three possible ways to overcome this impediment:

i. **Using the non-normalized Fiedler number μ_2**: In case the image size remains fixed, we can estimate the non-normalized Fiedler number μ_2, which we observe Table 4-10 decreases in inverse proportion to the distance β^n between features.

ii. **Using the normalized Fiedler vector v_2:** This preposition is based on our experience from Sec. 4.5.4, where the features extracted from v_2 proved valuable for distinguishing the surface morphology. Accordingly, we plot the Fiedler vector for the three test patterns β_2, β_4, and β_6 in Figure 4-49.

From Figure 4-49, we notice that the peaks in v_2 appear to be modulated at the distance between features for different image types. Thus, instead of using the Fiedler number λ_2, which is identically 1 for all cases tested (Table 4-10), we could instead use the Fiedler vector v_2 as a discriminant in cases where the graph network becomes bipartite.

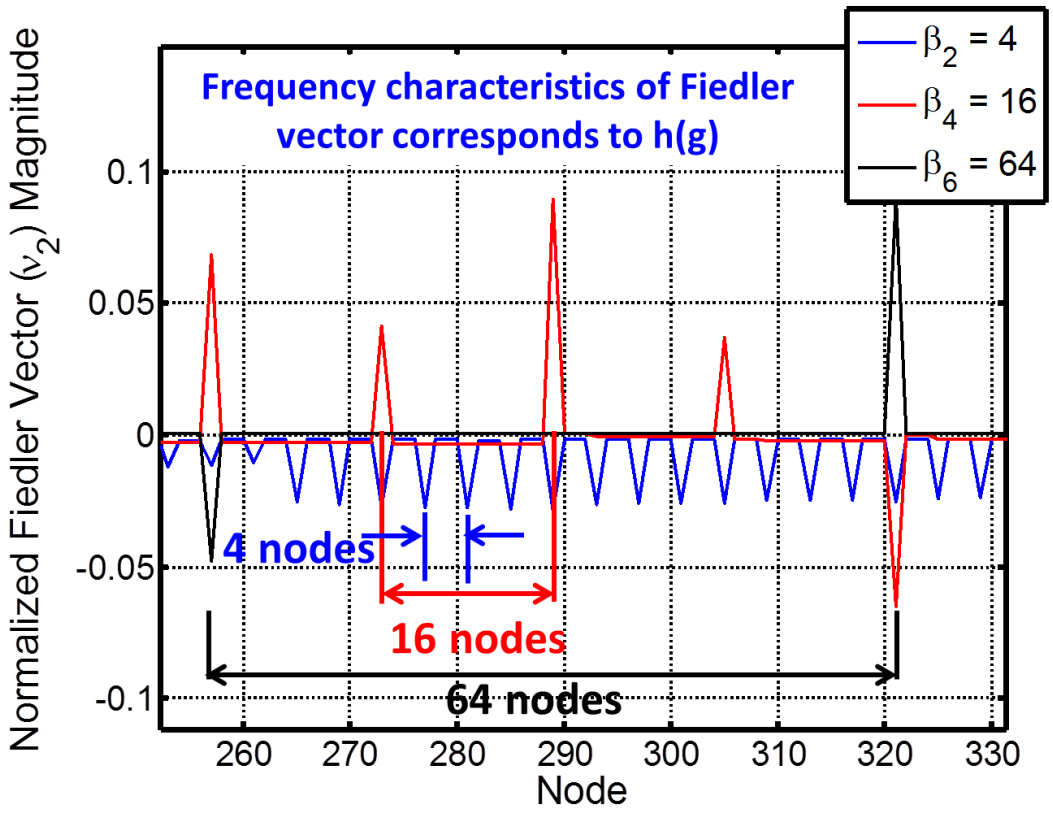


Figure 4-49: The Fiedler vector for three test patterns β_2, β_4, and β_6.

iii. **Using the Kirchoff indices K_f and $\mathcal{K}_f$** : In Sec. 4.4.2, we described the combinatorial and normalized Kirchoff indices. We now estimate these for surfaces β_1 through β_9, these are shown in Table 4-11. Since it is known that the combinatorial Kirchoff index is dependent on the degree of the graph, we will instead devote our attention to the normalized Kirchoff index $\mathcal{K}_f$.

From Table 4-11, we note that unlike the Fiedler number λ_2, the normalized Kirchoff index $\mathcal{K}_f$ seems to be sensitive to the change in surface morphologies. A concave-like trend is observed on plotting $\mathcal{K}_f$ vs. β_n, as shown

in Figure 4-50, which implies that $\mathcal{K}_f$ can be a useful discriminant for the bi-partite graph.

Table 4-11: Kirchoff indices for images β_1 through β_9.

n	2^n	Combinatorial Kirchoff Index $K_f \times 10^8$	Normalized Kirchoff Index $\mathcal{K}_f \times 10^6$
1	2	536.1	1.047
2	4	402.1	1.309
3	8	234.5	1.637
4	16	125.8	1.851
5	32	649.8	1.970
6	64	33.01	2.033
7	128	16.64	2.064
8	256	8.350	2.080
9	512	4.182	2.088

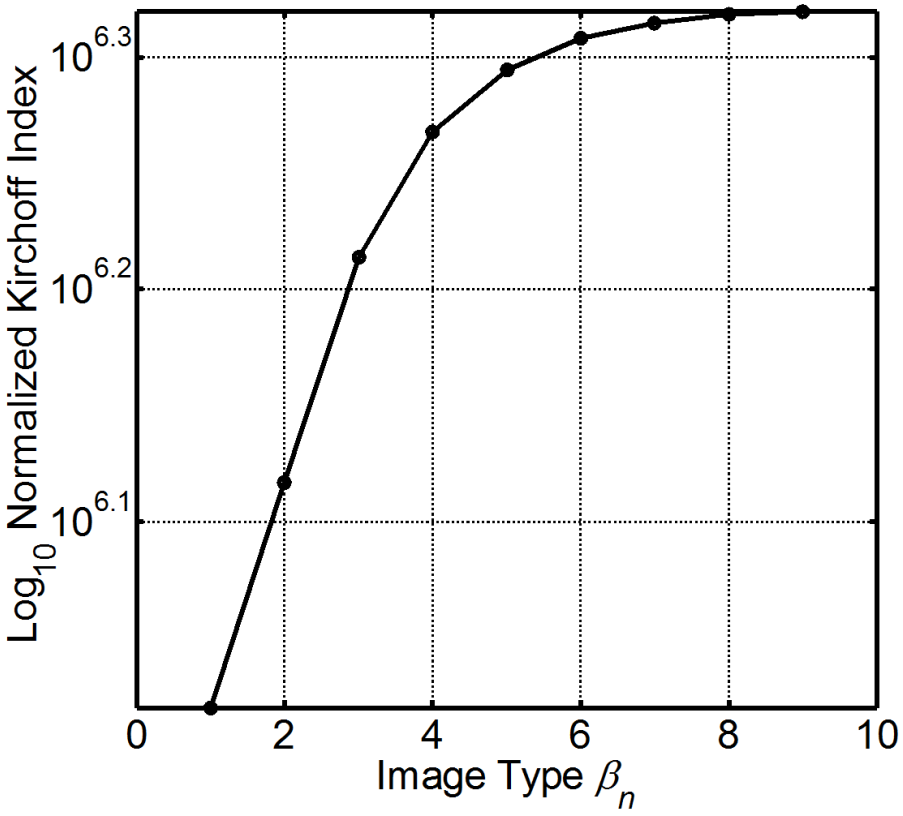


Figure 4-50: Normalized Kirchoff index (log scale) vs. image type.

Inferences from the study

1. The graph theoretic invariant Fiedler number λ_2 fails as a discriminant to capture differences (because $\lambda_2 = 1$, identically) in surface morphology for deterministic patterned surfaces that have an underlying bipartite graph network.
2. Instead, in such scenarios (where a bipartite graph is formed) the Fiedler vector v_2 is more sensitive to changes in surface morphology.
3. If the size of the image is maintained fixed, then the non-normalized Fiedler number μ_2 can also be used as a discriminant for surface morphology for bipartite graphs representations.
4. If computation overhead is not a concern, the normalized Kirkoff index $\mathcal{K}_f$ could be used as a discriminant for bipartite-like surfaces.

4.6 Verification with Experimentally Acquired CMP Surfaces

We now apply the approach described in Sec. 4.4 for characterizing CMP polished wafers. Blanket copper (Cu) C14500 series wafers (dia. 40.625 mm) are polished on a bench top CMP setup (Buehler Automet 250). We note that this particular copper alloy has Tellurium (Te) added as an alloying element to enhance machinability, which limits the surface finish that can be achieved [62].

The wafers are initially lapped to within 10 nm (Sa) using 0.3 μm alumina abrasives. The lapped wafers are subsequently CMP processed in three intervals of 3 minutes each (total 9 minutes of CMP), using alkaline colloidal silica slurry (70 nm particle size, 10 pH). The polished samples are cleaned with DI water, and a coating of benzotriazole (BTA) and isopropyl alcohol solution is applied to prevent corrosion. Near-specular surface quality (see Figure 3-3) was obtained using optimal processing conditions (identified *a priori).* After each polishing interval (which lasts 3 minutes) the wafer surface morphology is captured using a laser interferometer (MicroXAM) with an integrated optical camera.

We obtain the surface profile and corresponding optical images at six randomly selected positions on the wafer. We also confirmed these experimental results with a different wafer and obtained additional surface readings at two randomly chosen areas for each polishing interval.

Figure 4-51: Blanket copper wafers after 9 minutes of CMP with 70 nm colloidal silica slurry.

4.6.1 A Review of Wafer Metrology Approaches in Semiconductor Manufacturing

Tools currently used in nanotechnology research labs offer atomic-level resolution for characterizing* nanoscale *surfaces but are barely adequate to meet the requirements of high-volume nanomanufacturing. *For example, an Atomic Force Micro-scope (AFM) gives nearly atomic-level surface resolution, albeit at a very slow rate; it would be impossible to use AFM to characterize surfaces in commercial-scale high-rate operations.*

– S. Bukkapatnam, *et al.*, "Nanomanufacturing systems: opportunities for industrial engineers," *IIE Transactions,* (**44)**7, pp. 492-495, 2012/07/01 2012.

Surface metrology for semiconductor wafers is complicated due to the presence of patterned device structures (which may not be periodic) [9, 20, 195, 196], which affect the transmission and reflection of light waves incident on the surface. Furthermore, contact metrology is not preferred in semiconductor metrology due to the possibility of surface damage.

Stokowoski and Vaez-Iravani [170] provide a historic perspective on the evolution of semiconductor wafer inspection technology from standalone bright field and dark field systems to the hybrid systems predominant currently [197, 198].

In industrial CMP scenarios, the approach used for wafer metrology is contingent on throughput and yield considerations [170], and is typically a multi-step process [170, 195]. Conventional optical microscopy is preferred due to cost and time advantages; it is mainly used for detecting catastrophic surface defects, such as scratches, pitting, and delamination [199-201]. In many optical inspection systems [199, 200, 202, 203] imperfections are treated in a statistical process control (SPC) context, with event counts relating to wafer defects.

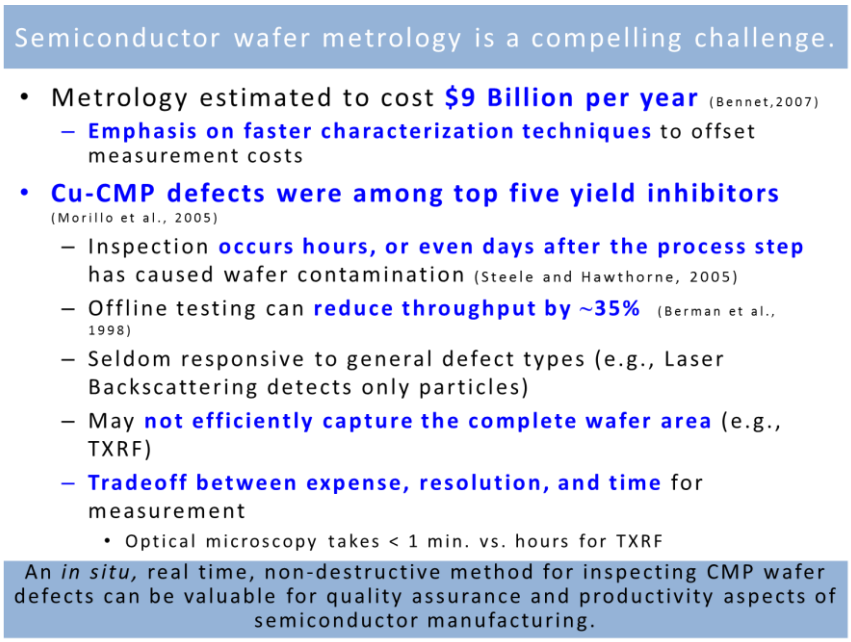
However, as we demonstrated in several instances (Sec. 4.5), such defect count approaches may not be appropriate for evaluating surface quality, and are often inadequate quantifiers of surface morphology.

Automated optical inspection methods usually employ some form of post-processing algorithms to detect and classify defects. Neural networks [204, 205], image filtering [198, 203, 206], and pixel comparison algorithms [195, 196, 200, 207] are some of the post-process defect classification methods that are incorporated into newer optical inspection systems. Although, high wafer examination rates almost matching production speeds (> 100 wafers per hour) are reported using optical systems [26, 170], they are confined to preliminary inspection stages, because, evaluation of key process output variables, namely, within wafer non-uniformity (WIWNU), step height, scratch depth, etc., is not possible with conventional microscopy.

Dimensional parameters in semiconductor wafers are measured using laser and glancing light metrology methods [206, 208-211]. Energy-based metrology

[212-214], e.g., scanning and transmission electron microscopy (SEM, TEM) are used for detection of particulate and near-surface/sub-surface imperfections, chiefly slurry residue, device fracture, peeling, contact failures, etc. Since the time required for SEM/TEM metrology is relatively long, wafers are inspected in small batches.

Characterization of chemical and atomistic level defects are carried out using mass spectrometry approaches, namely, μ-Raman, secondary ion mass spectrometry (SIMS), vapor phase decomposition (VPD), total X-ray fluorescence. These may require more than one production shift, and are sometimes destructive [26]. Therefore, such methods are employed over sparse sample sizes. Steele, *et al.*[26] summarize some of the salient aspects of methods widely used for industrial wafer metrology (reproduced in Table 4-12). These challenges are summarized below in Graphic XI.



Graphic XI: Challenges for semiconductor wafer metrology.

Table 4-12: Salient aspects of some of the wafer metrology methods used in the semiconductor industry (after Steele, *et al.*[26])

ANALYTICAL METHOD	ADVANTAGES	DRAWBACKS
OPTICAL MICROSCOPY	• In-line • Inexpensive • Non-contact/Non destructive	• Nano-particle and chemical residue detection is limited
LASER BACKSCATTERING	• In-line • Inexpensive • Can detect small residue particles • Non-contact/Non-destructive	• Chemical residue detection not possible
VAPOR PHASE DECOMPOSITION INDUCTIVELY COUPLED MASS SPECTROMETRY (VPD ICP-MS)	• Extreme sensitivity to a wide range of defects/contaminants	• Off-line • Only for bare silicon • Destructive • Complex • No topography image (only spectra)
TOTAL REFLECTION X-RAY FLUORESCENCE (TXRF)	• Very sensitive • Nondestructive in most cases • Some topography mapping capability	• Off-line • Limited wafer area covered • Un-patterned wafers only
SECONDARY ION MASS SPECTROMETRY (SIMS)	• Extreme sensitivity to a wide range of defects/contaminants • Sub-surface can be mapped	• Off-line • Expensive • Destructive

4.6.2 ε Neighborhood Graph Theoretic Approach for CMP Wafer Morphology Quantification

We note that the Te particles present in the Cu metal-matrix significantly biases the surface measurement readings obtained from the laser interferometer and can therefore be construed as extraneous artifacts having little bearing on the actual surface finish. We used a two-step Canny image filtering technique [177] for eliminating the regions with Te from the optical image ($\mathcal{I}$, Figure 4-52(a)) and therefore minimizing the bias.

The Canny filter is a well-established and predominant edge detection approach used in image processing applications. In the first step, the Canny filter parameters were set such that the resulting image consists mainly of regions with Te inclusions (indicated with yellow arrows in Figure 4-52(b)). The application of the Canny filter converts the grayscale image $\mathcal{I}$ into a binary image [177]. In the second step, the filter is set to capture both the prominent features on the surface, as well as, Te inclusions as shown in Figure 4-52(c) (yellow and green arrows).

We term the image from step 1 and stage 2 as $\mathcal{I}_H$, and $\mathcal{I}_L$, respectively. The difference, $\mathcal{I}_D = \mathcal{I}_L - \mathcal{I}_H$, in Figure 4-52(d) is a binary image largely consisting of relevant surface features (green arrows only) with Te rich regions removed. The Canny filter parameters are the *only* heuristically adjusted parameters in the entire approach. They are estimated only once using a representative image and remain fixed for all subsequent images (for all experimental CMP surfaces obtained in this study). A similar two stage approach using Gaussian filtered binary images

for defect detection in patterned wafers has been documented by Nichani, *et al.*[206].

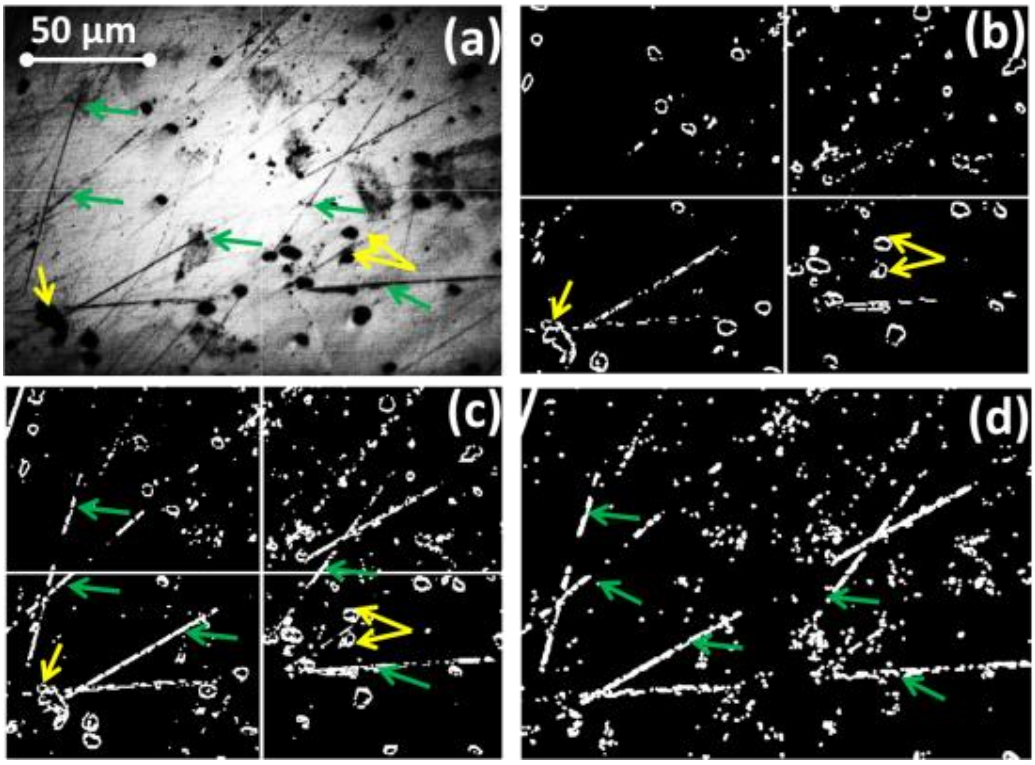


Figure 4-52: Canny filtering applied to CMP surfaces.

(a): Representative optical image. (b): Binary image obtained after the first Canny filter step capturing mostly the Te affected regions ($\mathcal{I}_H$). (c) Binary image obtained after the second Canny filter step capturing both Te regions and surface features ($\mathcal{I}_L$). (d): The difference image with the Te areas largely discounted ($\mathcal{I}_D$).

Next, we estimated the Fiedler number for CMP polished Cu test wafers (after the Canny filtering procedure). Representative results shown in Figure 4-53(a)-(b) depict the optical micrographs for a lapped Cu wafer (before CMP), and after 9 minutes of CMP processing, respectively. The corresponding Canny filtered images using the procedure described above are shown in Figure 4-53(c) and (d). Their graph eigenvector maps (v_3 vs. v_2) are plotted in Figure 4-53(e)-(f).

It is evident that as the wafer surface is polished over time, the number of distinct clusters in the eigenvector map reduces from over 100 for the lapped surface to < 20 after 9 minutes of CMP. Concurrently, the Fiedler number reduces from 0.85 to 0.65 (Figure 4-56 (a)).We also note that there is a less significant decrease in Fiedler number between 3 minutes and 6 minutes of polishing compared to the last 9 minutes. The corresponding network graphs (with 40 nodes out of 1131 total) presented in Figure 4-54, shows a gradual reduction in network connectivity with progressive CMP stages.

Additionally, the trajectories of eigenvectors (v_2, v_3, and v_4) corresponding to the three smallest non-zero eigenvalues (λ_2, λ_3, and λ_4, respectively) of the normalized Laplacian matrix $\mathcal{L}$ for different processing stages are mapped in Figure 4-55. The eigenvector trajectories show a gradual reduction in the number of distinct node clusters with CMP (consistent with Figure 4-53(e)-(f), and Figure 4-55). A Tukey's pairwise comparison test based on Fiedler numbers at different stages confirms this observation: the p-value is < 0.01 for pairwise difference between as lapped and all other stages; the same is true for comparisons with the 9 minute CMP stage. This is however not the case (p-val. > 0.01) between 3 minute and 6 minute stages.

Another salient aspect of this result is noted in the similarity of Type 1 (Figure 4-6 (a)) defects illustrated in Sec. 4.5.1 with the lapped (before CMP) case in Figure 4-53(c). The Fiedler numbers are nearly identical for both cases (0.81, compare Table 4-3 and Table 4-13) for such a type of topography. Also, the

topography in Figure 4-53(c) appears to be a close combination of defect Types 2 and 3. Consequently, the Fiedler numbers (0.66) estimated from CMP stages represented by Figure 4-53(c) are closer to the respective simulated cases (0.64 to 0.69). This further strengthens the assertion that the Fiedler number can capture changes in surface topology as opposed to mere density of detected features.

We also tested the approach with binary images obtained by applying only a one-step Canny filter. In this case only the lower threshold images ($\mathcal{I}_L$) are used, which leaves the Te rich areas as is, but minimizes the eventuality of losing valuable surface feature information due to excessive filtering. The results are summarized in Figure 4-56(b) and Table 4-13, from where it is indicative that an adequately designed binary filter can also be effective for surface characterization. We have also utilized different kinds of edge detection filters, such as Prewitt, Sobel, etc., with comparable results.

As noted earlier, the ε neighborhood representation essentially captures the topography of nanoscale imperfections. Hence, it is better suited for situations where defect density and distribution are to be quantified.

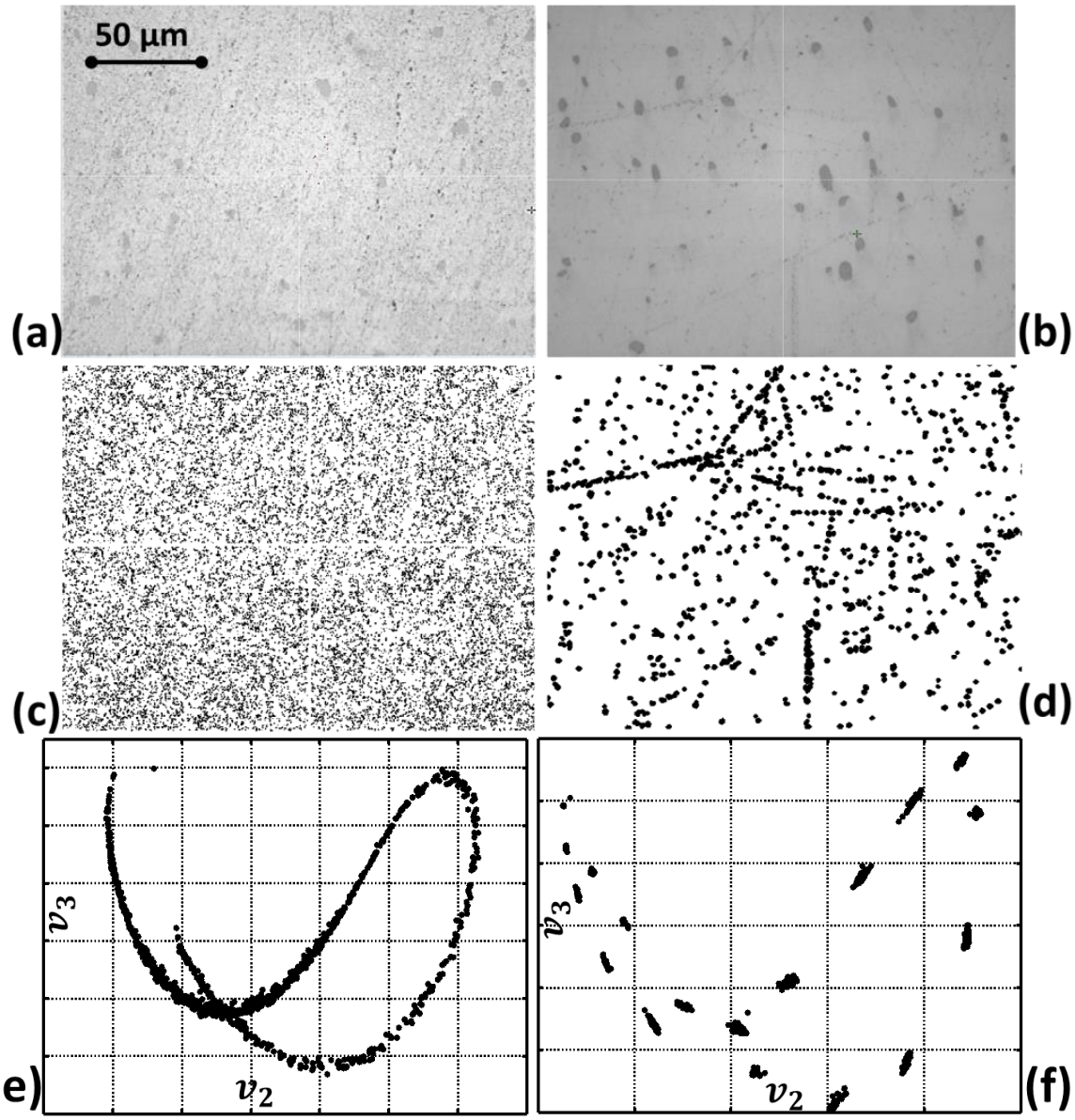


Figure 4-53: ε neighborhood graph representation applied to CMP surfaces.

(a)–(b): Representative optical images for as received and 9 minutes CMP polished wafer, (c) – (d) The binary Canny filtered images corresponding to (a) and (b) respectively, and (e) – (f): The respective maps of the two smallest eigenvectors (v_3 vs. v_2).

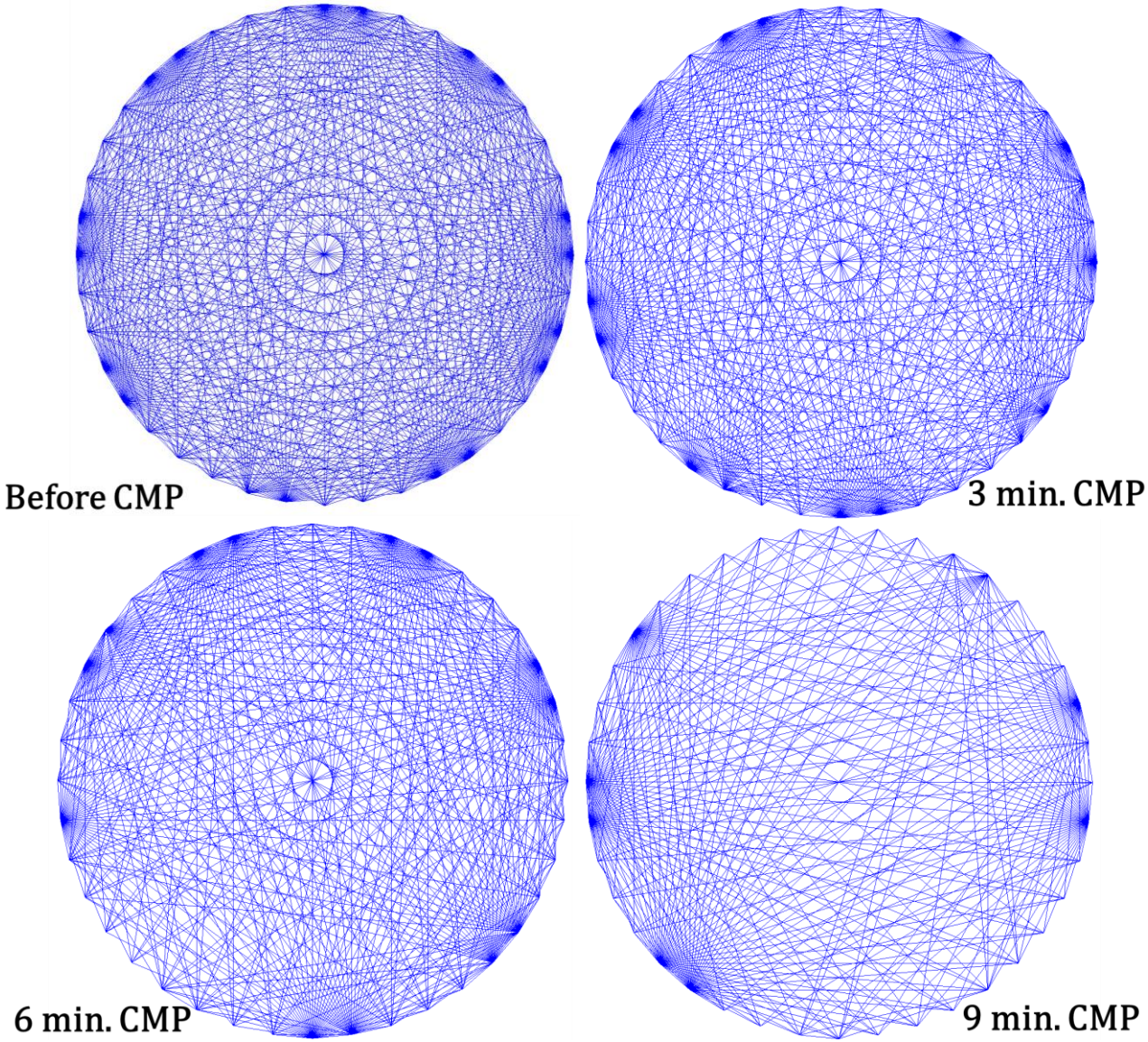
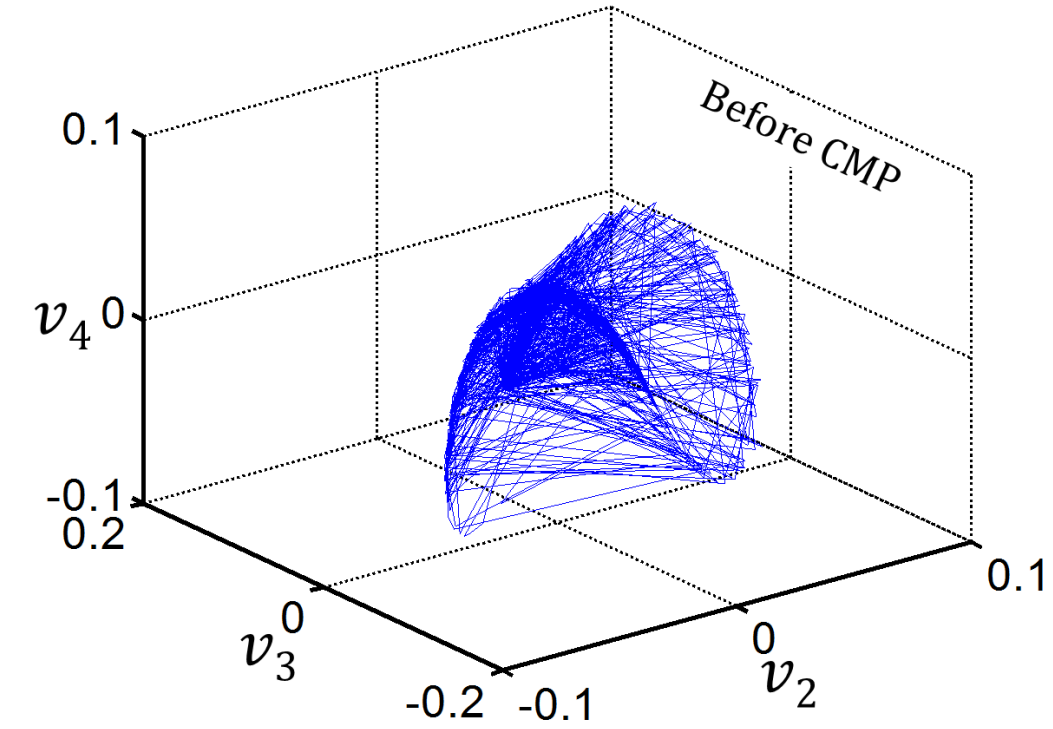
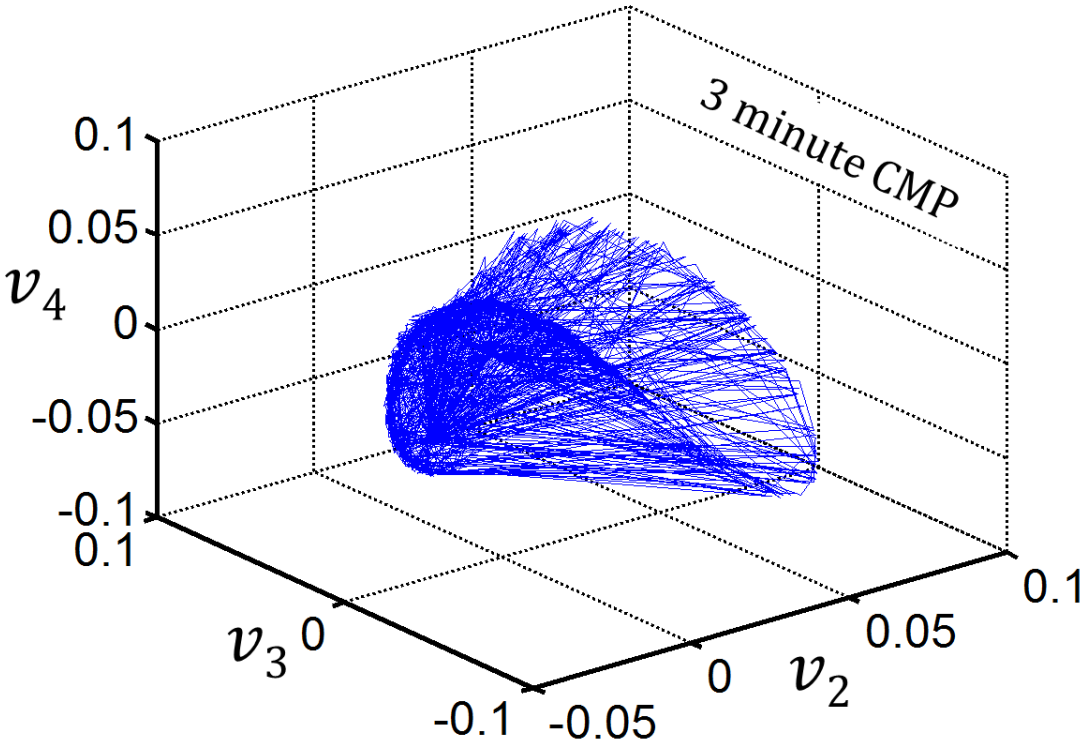


Figure 4-54: Graph networks (40 nodes only) for different stages of CMP obtained using the ε neighborhood approach. Notice the gradual reduction in connectivity with CMP processing time.


Before CMP
0.1
v_4 0
-0.1
0.2
v_3 0
-0.2
-0.1
0
v_2
0.1


3 minute CMP
0.1
0.05
v_4 0
-0.05
-0.1
0.1
v_3 0
-0.1
-0.05
0
v_2
0.05
0.1

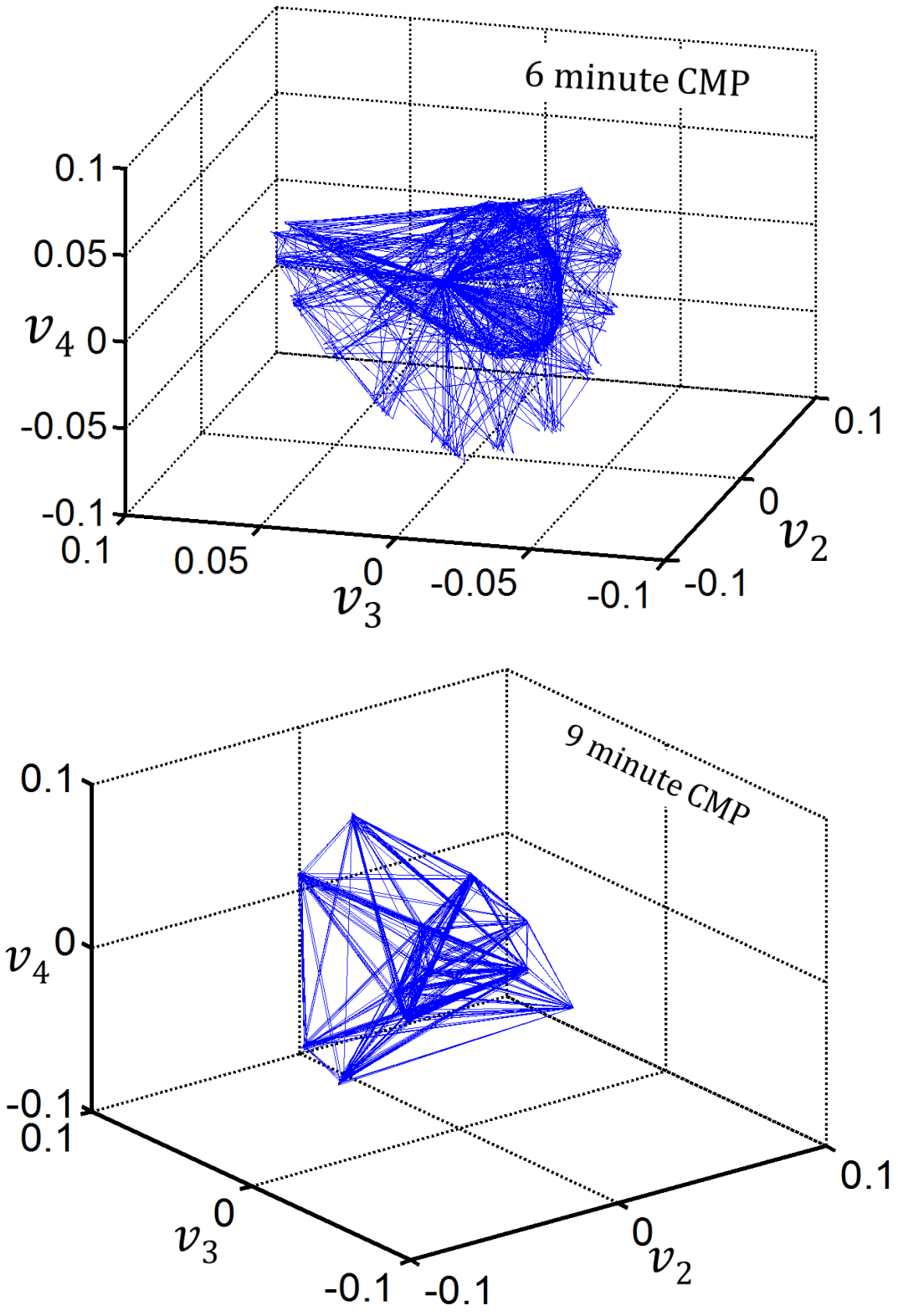


Figure 4-55: The trajectories of three smallest eigenvectors.
v_2, v_3, and v_4, are obtained from the normalized Laplacian $\mathcal{L}$ using the ε neighborhood representation for various CMP stages.

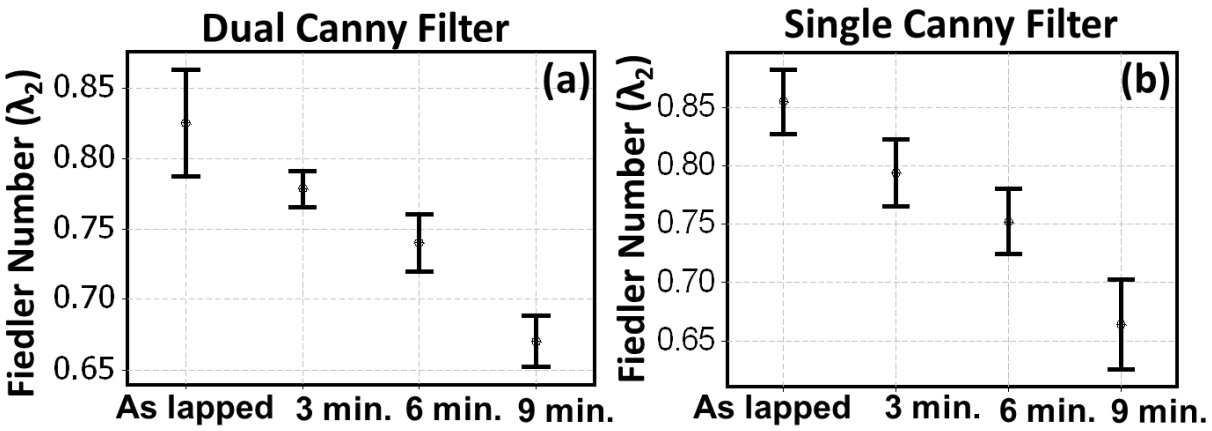


Figure 4-56: Evolution of Fiedler number with polishing time.

(a) Using two Canny filtering steps to remove Te inclusions, and (b) using one Canny filter (Te inclusions not removed).

4.6.3 Edge Weighted Graph Theoretic Approach for CMP Wafer Morphology Quantification

The approach presented in the previous section used filtered images for graph representation. Selecting the correct filter parameters can be challenging, and filtering may lead to loss of valuable information. For example, in the previous case there was a complete loss of image texture information (due to Canny filtering).

Therefore, in this study, we eliminate the filtering step and use the (gray scale) optical micrograph. For this purpose, we use the edge weighted approach outlined in Sec. 4.4.4, that compares the texture differences in the surface image based on a Gaussian radial basis function instead of Euclidean distances from filtered binary images.

The optical micrographs are acquired as 1131 × 1451 (1.64 mega pixel) gray-scale tagged image files (TIFF format). The TIFF format is popular in scientific

image processing applications due to its high fidelity and resolution in comparison to other formats. The image obtained is a matrix of integers between 0 and 255 that indicates pixel contrast, with zero indicating black and 255 white. We use Eqn. (4-28) and Eqn. (4-30) for calculating the pairwise pixel row differences.

In this case, we note that heuristic filter threshold parameters are precluded. Figure 4-57(a)-(b) shows representative (graphically rendered) similarity matrices for an as lapped and 9 minutes polished CMP wafer, respectively. These correspond to the surfaces in Figure 4-53(a)-(b). The Fiedler number estimated for different stages is depicted in Figure 4-57(c). We note that the mean Fiedler number reduces from close to 0.75 for lapped wafers to 0.33 for 9 minutes CMP processed wafers. This implies the surfaces have become relatively homogenous due to CMP. This is consistent with the change in connectivity of the graph networks (with 40 nodes out of 1131 nodes shown) for various CMP stages shown in Figure 4-58.

The results from the edge weighted graph seem to agree well with the ε neighborhood approach, these are juxtaposed in Table 4-13. Tukey's multiple comparison procedure indicates significant difference in $\bar{\lambda}_2$ between stages (p-val < 0.01) for all comparisons paired with as lapped and 9 minutes CMP stages, but is not statistically significant (p-val > 0.01) for comparisons between 3 minute and 6 minute stages.

Table 4-13: Fiedler number estimates across CMP stages using different approaches. Each estimate includes 8 points, two of which are obtained from confirmation runs.

Polishing Stage	ε neighborhood approach Dual Canny Filtered Image		ε neighborhood approach Single Canny Filtered Image		Edge weighted approach Gray Scale Image	
	$\bar{\lambda}_2$	σ_{λ_2}	$\bar{\lambda}_2$	σ_{λ_2}	$\bar{\lambda}_2$	σ_{λ_2}
As lapped	0.81	0.036	0.84	0.038	0.75	0.066
3 minutes	0.76	0.024	0.78	0.034	0.46	0.152
6 minutes	0.72	0.034	0.74	0.357	0.49	0.151
9 minutes	0.66	0.017	0.66	0.032	0.33	0.148

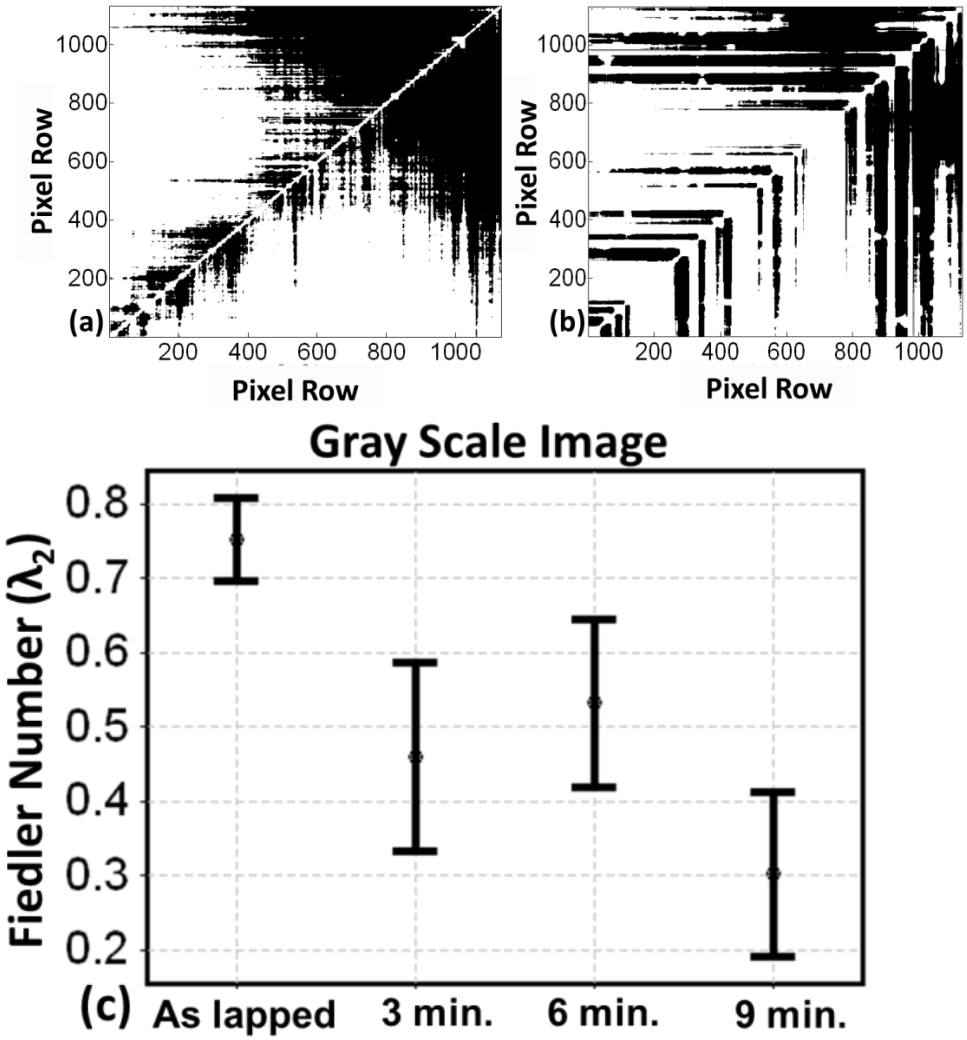


Figure 4-57: Edge weighted graph representation applied to CMP processed surfaces.

(a) – (b) Similarity matrices for lapped and 9 minutes CMP processed wafer surfaces, these correspond to micrographs depicted in Figure 4-56 (a) and (b), respectively. (c) Fiedler number vs. polishing time directly estimated from gray scale images.

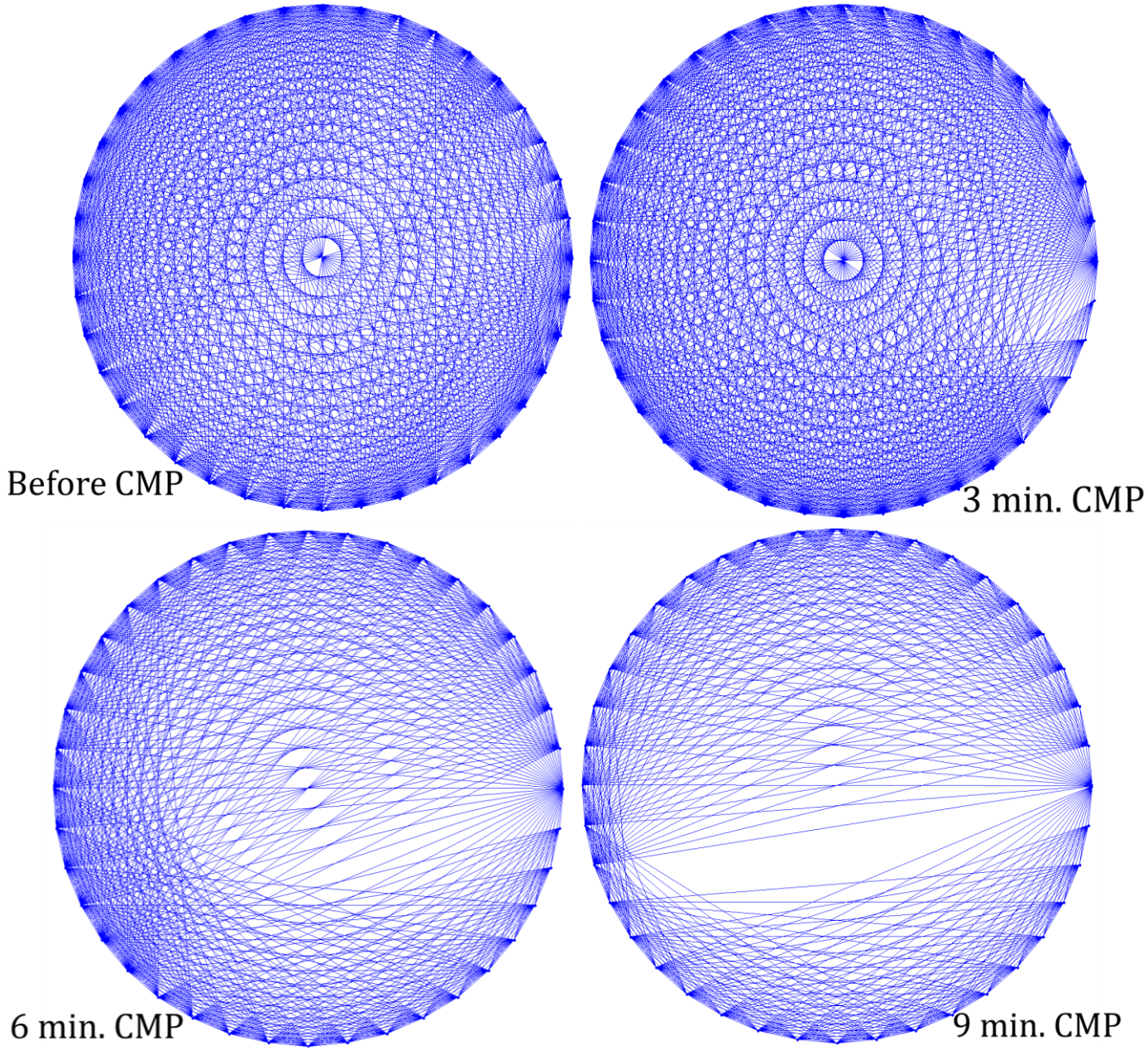


Figure 4-58: Graph networks (40 nodes only) for different stages of CMP obtained using the edge weighted representation.

Notice the gradual reduction in connectivity with CMP processing time.

4.6.4 Verification with Local Surface Roughness Measurements

We proceeded to verify the graph theoretic results using locally measured conventional statistics, such as Sa, feature counts (which are obtained from binary filtered optical images), and 2D Fourier transform energy (sum-of-square magnitude). We considered the 3D wafer topography W obtained using a laser interferometer (MicroXAM), and compared its corresponding optical micrograph. By juxtaposition, we note the coordinates of Te inclusions, $S \equiv \{S_1, S_2 \cdots S_n\}$.

Subsequently, conventional surface roughness statistics, such as arithmetic mean (Sa), root mean square (Sq), peak-to-valley depth (S_z), etc., were evaluated at 20-30 randomly generated non-overlapping square shaped areas measuring 35 μm × 35 μm from the set $R \equiv W - S$. This measurement scale was determined heuristically. Since we have obtained 3D wafer topography readings at 6 random wafer locations after every 3 minutes of CMP, there are a total of 100 – 200 measurements available at every polishing interval. These local surface roughness statistics are shown in Figure 4-59 and subsequently tabulated in Table 4-14.

An interesting aspect of these results is noted from Figure 4-59, where both feature count and conventional surface roughness parameters, such as Sa are observed to decrease with CMP over time, implying that surface quality improves both in terms of defect density and texture. However, this might not always be the case, as imperfections, such as nano-scratches and rugosity, preside over fundamentally different length scales (frequency) and may therefore occur independently of each other (a similar analogy is the difference between surface roughness

and waviness) [8, 57]. In this context, the two graph approaches, ε neighborhood and edge weighted representations, can be applied in a complimentary fashion, the former for characterizing nanoscale imperfections, and the latter for quantification of surface texture morphology.

The trend of locally measured surface features with CMP stages corroborates well with the Fiedler number (Table 4-15). It is noteworthy that the correlation coefficient (ρ) ranged from ~ 80% to 99% for all comparisons between conventional parameters and Fiedler number as shown in Table 4-15. These results are a further confirmation in favor of the Fiedler number as a statistic for surface characterization.

Moreover, conventional statistics are often cumbersome to obtain, sensitive to artifacts, such as material inclusions, and scale dependent. Also, 3D surface profiles require significant post-processing to eliminate systemic bias, and determining the appropriate sampling lengths can be computationally demanding. These challenges can be overcome by using the presented approach.

Table 4-14: Conventional statistics measured at different stages of CMP (100-200 measurements for each stage)

Quantifier	**Feature Count (%)**		**Sa (nm)**		**Sq (nm)**		**S_z (nm)**		**Energy ($\times 10^{-5}$)**	
Stage	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
As lapped	5.06	0.49	9.35	1.46	11.33	1.53	39.82	4.63	1.73	0.62
3 minutes	2.65	0.56	8.96	1.78	10.59	1.76	32.92	2.99	1.55	0.51
6 minutes	1.41	0.36	7.93	0.44	9.47	0.48	32.42	2.12	0.93	0.04
9 minutes	0.47	0.09	7.34	0.27	8.76	0.33	29.23	1.97	0.87	0.05

Table 4-15: Correlation coefficients ρ for various conventional parameters (Table 4-14) when compared with the mean Fiedler number (Table 4-13) obtained across CMP stages

Correlation Coefficient ρ			
Parameter	$\overline{\lambda_2}$ from ε neighborhood approach using single Canny filtered image	$\overline{\lambda_2}$ from ε neighborhood approach using dual Canny filtered image	$\overline{\lambda_2}$ from edge weighted approach gray scale image
Sa (nm)	> 0.99	> 0.99	0.81
Sq (nm)	> 0.99	> 0.99	0.87
St (nm)	0.92	0.92	0.98
Energy	0.93	0.94	0.77
Feature count (%)	0.94	0.94	0.93

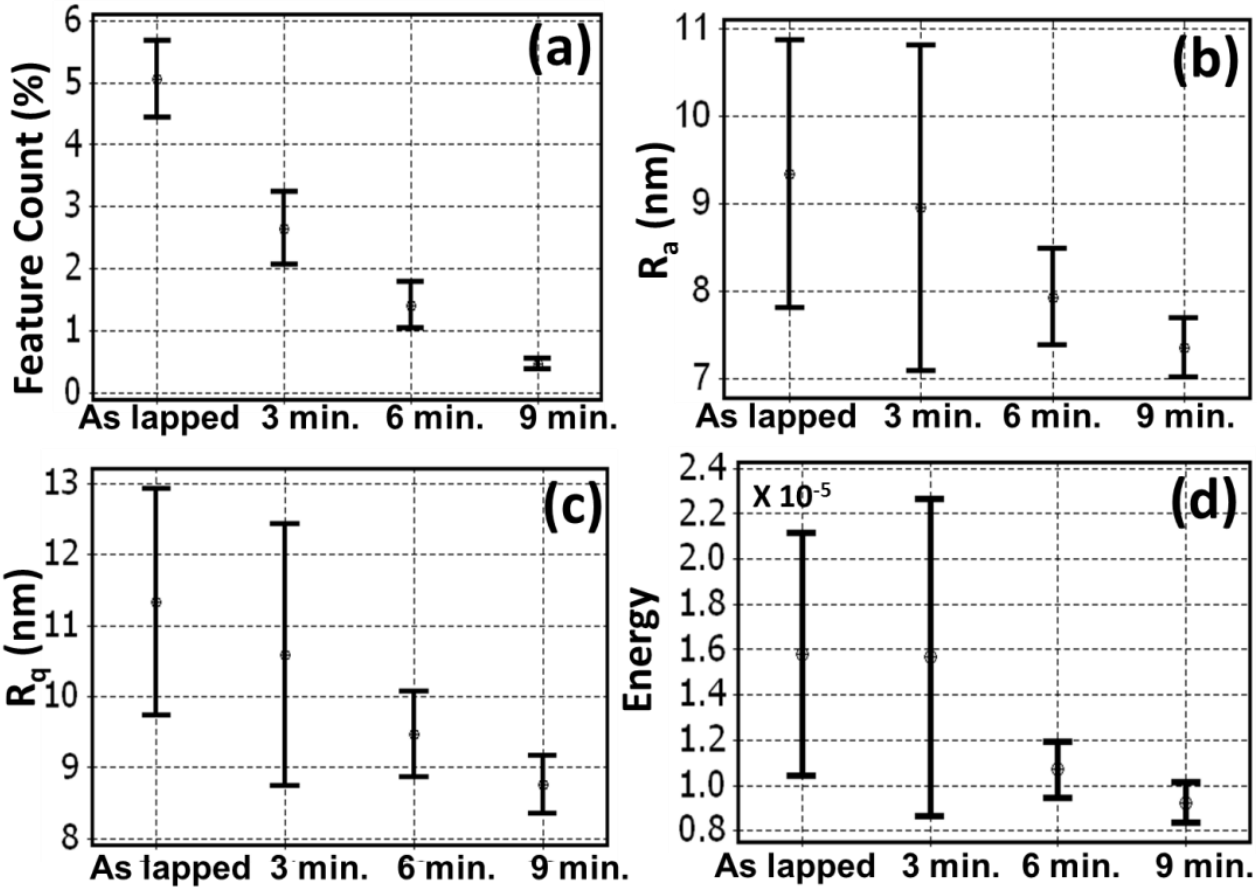


Figure 4-59: Verification with conventional roughness parameters.

(a)-(d): Conventional statistics vs. CMP stage, (a) percentage feature count, (b) arithmetic mean roughness (areal Ra, i.e., Sa), (c) root mean squared roughness (areal Rq, i.e., Sq), and (d) Sum of square of Fourier transform magnitude (energy).

4.6.5 The nature of the defect distribution for CMP processed wafers

In this section, we investigate the nature of the defect (count) distribution for CMP processed wafers. For reasons of consistency with the presented graph theoretic approach, we compute the number of defects that occur per pixel row of the Canny filtered image $\mathcal{I}_D$ (i.e., the difference image $\mathcal{I}_L - \mathcal{I}_H$ from the Canny filtering process described in Sec. 4.6.2). Doing so is analogous to the row-wise pixel comparisons (Eqn. (4-2)) used for representing CMP surfaces as network graphs.

Aim of the study

Investigate the nature of the defect count distribution for CMP processed surfaces.

The empirical probability distributions of defect counts (taken row-wise) are shown in Figure 4-60 (each distribution is approximated from close to 10,000 data points for every CMP stage). From Figure 4-60(a), we observe that the mean defect count reduces significantly with progressive CMP stages. In addition, we can discern that the shape of the defect count distribution changes (Figure 4-60(b)) with CMP over time. A Gaussian-like distribution is evident for the initial (before CMP) stage, which evolves into a heavy-tailed (right skewed) shape after 9 minutes of CMP (see also Table 4-16).

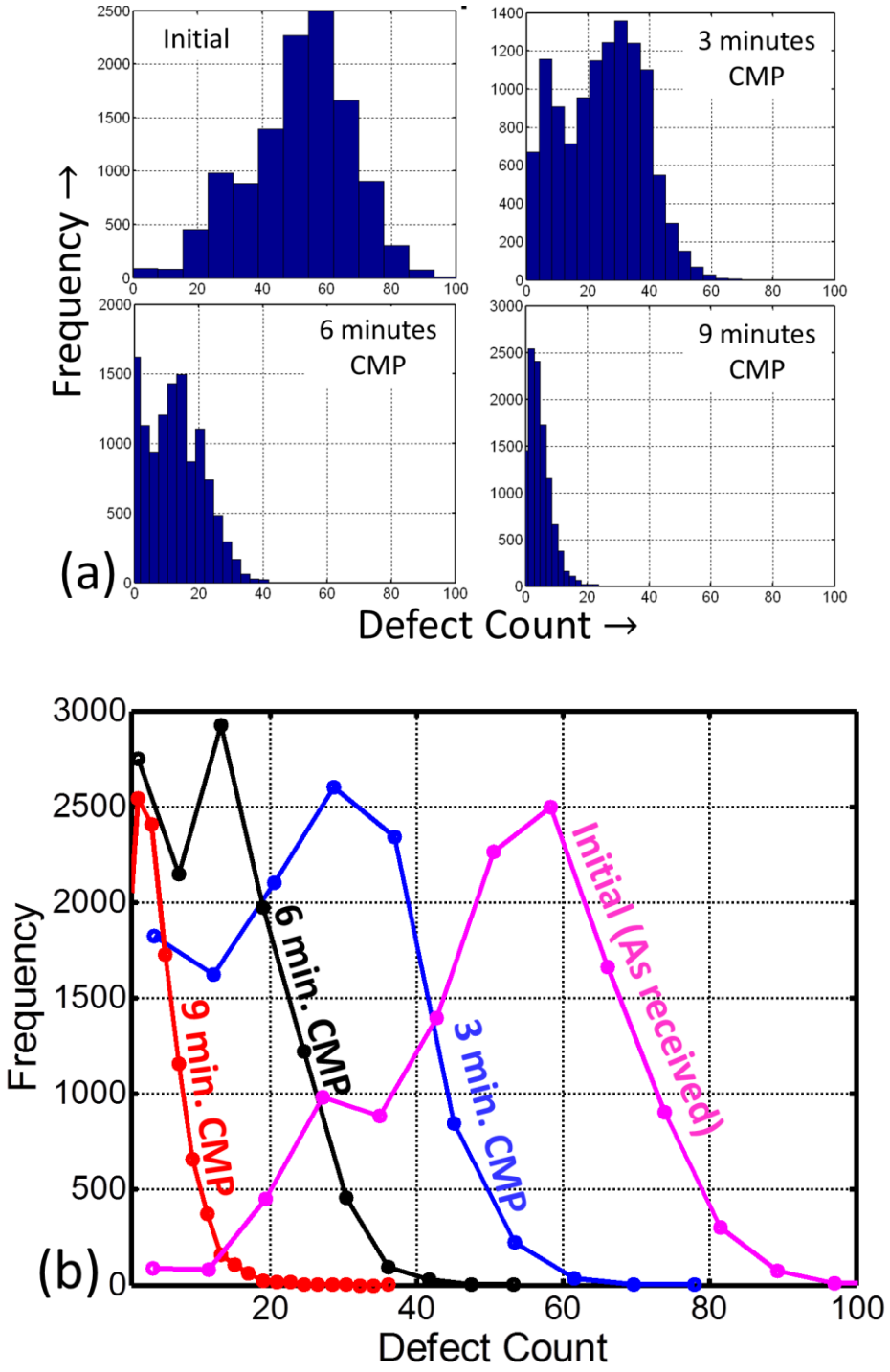


Figure 4-60: Empirical (~10,000 data points each) distribution of defect counts for different CMP stages.

(a): Histograms of defect counts for different CMP stages. Note the change in location and shape of the distribution with CMP stages.
(b) Empirical frequency distribution for different CMP stages. Note the steady evolution towards a more heavy-tailed shape with progressive CMP stages.

Further continuing with our study, we now fit theoretical probability distributions to the defect count data. We note that the simplest (and well-known) distributions, which fit reasonably well (by visual observation), are preferred over more complex (but statistically better fitting) candidates. For instance, referring to the probability distribution plots shown in Figure 4-61, the extreme value distribution (not shown in Figure 4-61) was assessed statistically (using the Anderson-Darling statistic) to be a better fit than the Gaussian distribution for the 6 minute CMP stage[22]. However, for reasons of simplicity we chose the Gaussian distribution, which although departs significantly at the tails, fits the majority of the data amicably.

The different distributions and their parameters used in this study are shown in Table 4-16. Given the complex nature of the data, three candidate distributions are evaluated for the 9 minute CMP stage, namely the gamma, negative binomial, and Weibull distributions. For this particular CMP stage (9 minute), we observe that the negative binomial and gamma distribution both (visually) seem to fit the data reasonably. This is pertinent because the negative binomial converges to a Poisson distribution as the parameter $r \rightarrow \infty$ [215, 216][23]. This implication helps to justify the practical suitability of some of the simulated cases presented in Sec.

[22]Although, continuous distributions ,such as the Gaussian, gamma, and Weibull would not be typically used for fitting discrete integer data , we make an exception based on the large number of data points available, this allows us to treat the defect distribution as continuous and real valued.

[23] Also, the negative binomial distribution can be approximated using a gamma-Poisson mixture function, i.e., a mixture of Poisson distributions convoluted over a gamma distribution; see the www.wikipedia.org page on the negative binomial distribution.
(http://en.wikipedia.org/wiki/Negative_binomial)

4.5, especially, Sec. 4.5.2, Case 2(a) where the defects were simulated from a spatial Poisson process.

Inferences from the study

1. The distribution of defect counts over a CMP processed wafer is not stationary, and changes over time with CMP.
2. The defect counts are approximately Gaussian distributed for the as-lapped (before CMP) stage, but progressively become more heavy-tailed (pronounced right skew) with CMP.
3. The wafer defect count distribution after the final CMP stage (9 minutes CMP), is well represented with three candidate distributions, namely, negative binomial, gamma, and Weibull.
4. Since the negative binomial is known to converge to the Poisson distribution, some of the simulated case studies (e.g., Sec. 4.5.2, Case 2(a)) can be evocative of practical scenarios.

Table 4-16: Candidate distributions for different CMP stages.

Stage	Candidate Distribution	Parameters	Mean	Var.	Mean	Var.
			Theoretical		Empirical Estimate	
Initial (as lapped)	Gaussian	mean: 51.8 variance: 247.63	51.8	247.63	51.4	264.95
3 minutes	Gaussian	mean:24.7 variance: 167.31	24.7	167.31	24.5	170.94
6 minutes	Gaussian	mean:13.7 variance: 66.68	13.7	66.68	12.5	76.51
9 minutes	Gamma	shape: 2.03 scale: 2.511	5.1	12.81	4.4	15.20
9 minutes	Negative Binomial	r = 3.4 p = 0.4060	5.1	12.56		
9 minutes	Weibull	shape: 1.45 scale: 5.66	5.1	12.81		

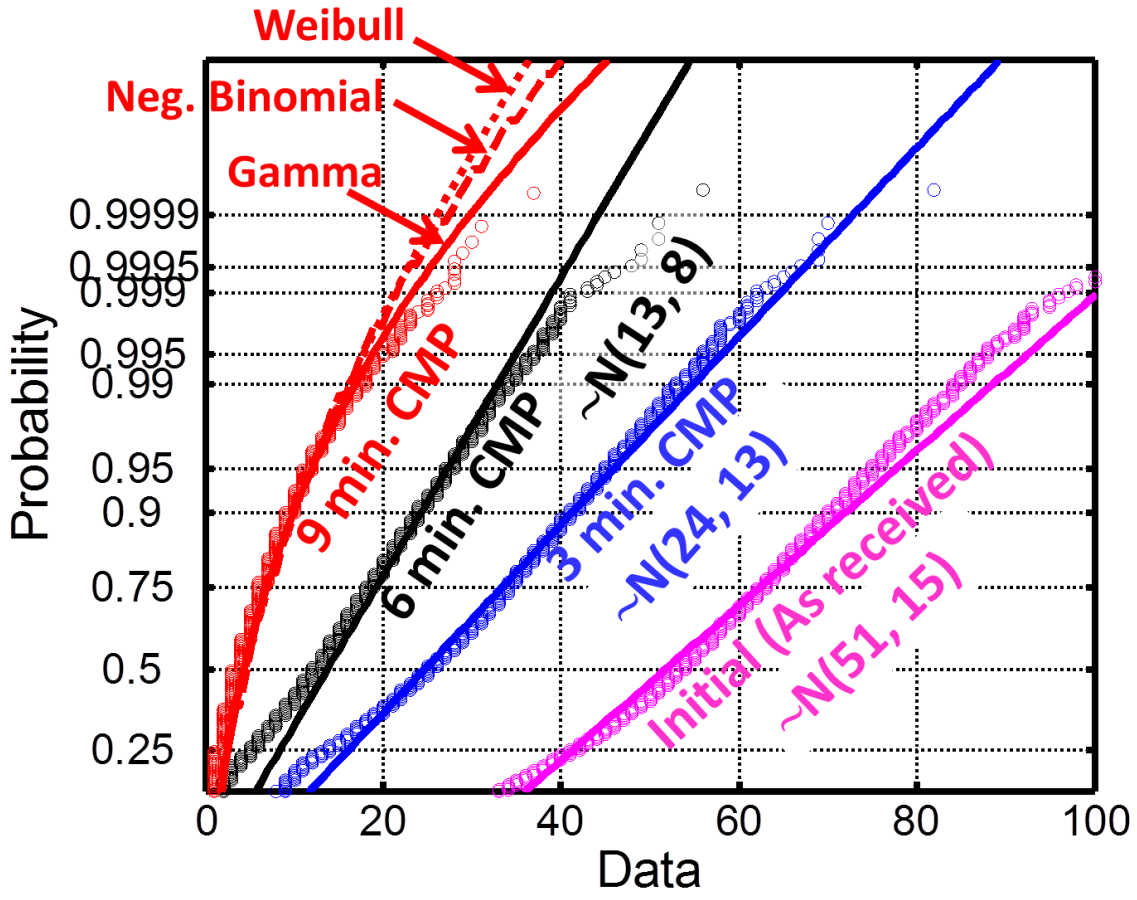


Figure 4-61: Theoretical distributions fitted to the data over different CMP stages (~ 10,000 data points per stage).

4.7 Summary

We have presented an approach invoking algebraic graph theoretic concepts as a means to quantify evolving nano-surface morphology. The approach is demonstrated to be capable of capturing surface morphology variations in near-specular CMP processed Cu wafers using non-contact optical micrographs and therefore overcomes much of the lacunae (e.g., scale dependency and bias) encountered in nanoscale surface metrology. Since reticent off-line characterization methods, such as interferometry are precluded, the approach can be valuable for in-line assessment of wafer surface integrity so that timely corrective action may be taken to minimize yield losses in semiconductor CMP. Specific contributions are as follows:

1. Two parallel methods to represent optical micrograph images as graph networks are detailed. The first graph representation called *ε neighborhood approach* uses binary filtered micrographs to quantify nanoscale wafer imperfections, such as scratches and pits. The second called *edge weighted approach* directly uses gray scale micrographs and captures textures aspects of the surface.
2. The Fiedler number was assessed to be a better quantifier of surface morphology in comparison to conventional defect count measurement. Three different defect distributions, with identical defect counts and bearing a close resemblance to actual CMP surfaces, were simulated. The Fiedler numbers estimated for the three cases were significantly different (p-val <

0.01), indicating that the Fiedler number is more responsive to changes in surface morphology.

3. The approach was verified against experimentally acquired CMP wafer surface micrographs and topography scans (obtained using a laser interferometer). Both methods of graph representation were used. The estimated Fiedler numbers were compared with locally sampled (35 μm × 35 μm) conventional surface characterization parameters, namely; Sa, Sq, S_t, Energy (sum of squares of FFT magnitude), and percentage defect count. The correlation coefficient ρ between conventional parameters and Fiedler number was estimated to be in the range of ~ 80 – 99% for all combinations tested.
4. It was observed that as the surface quality improves, the mean Fiedler number generally decreases and is significantly different (p-val < 0.01) than at the start of CMP. For lapped (before CMP) wafer surfaces, the Fiedler number is in the range of 0.75 to 0.85. After 9 minutes of CMP the Fiedler number decreases to ~ 0.65 when evaluated using the *ε neighborhood approach*. Using the *edge weighted approach* the Fiedler number after 9 minutes of CMP showed a more significant decrease to ~ 0.3.

5 Future work

Future Directions for Further Application of the Graph-theoretic Approach

The following directions for further research have emerged concerning the graph theoretic approach for surface morphology inspections:

1. Using higher order Laplacian eigenvectors $\mathcal{v}^*$, instead of only the second eigenvector $\mathcal{v}_2$ for quantification of surface morphology.
2. Applying the graph theoretic approach for process diagnosis (i.e., classification of process state), instead of restricting the approach to process monitoring.
3. Expanding the scope in terms of the quantifiers, and modifying the approach, such that the nodes are constructed based on the peculiarities of the process.

5.1.1 Using Eigenvectors k > 2

In response to the contention that the higher order eigenvectors $\mathcal{v}_k$, k > 2 may contain useful information, but are not being used in this work, we tender a brief comment. In particular, we recognize that the eigenvectors $\mathcal{v}^*$ might contain more information than merely a scalar invariant, namely, λ_2. The eigenvectors, for example, can be useful for image classification and monitoring purposes in a SPC scenario.

For instance, using the first three non-zero eigenvectors, v_2, v_3, and v_4, one can construct control charts as described by Huang, *et al.* [217] (the approach involves transforming the eigenvectors into Hilbert space), and subsequently monitor the surface morphology in an SPC setting.

5.1.2 Using the Graph Theoretic Approach for Process Diagnosis

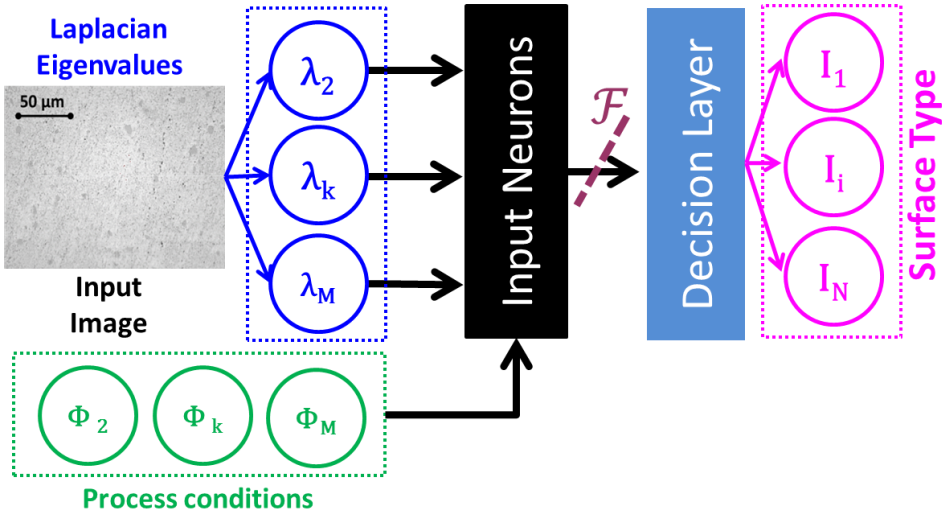


Figure 5-1: A neural classifier of the process state using Laplacian eigenvalues of the obtained surface.

Another direction for future research is to use the graph theoretic approach for process diagnosis, as opposed to only measurement of surface morphologies. For example, using the Laplacian eigenvalues λ^* from the image and process conditions Φ a neural network can be trained that classifies the process state in terms of the surface type (Figure 5-1). In essence, we classify the process outcomes into different clusters.

By doing so, if the surface is classified in a different cluster than the one expected (given the process conditions), we can re-trace the path (using the neural

network) to identify which process condition might have drifted and therefore caused an *out of control* condition. In effect, we can create a dictionary of process conditions that can lead to a deterioration of surface finish.

This can be valuable from a process diagnosis point of view, because, using such an *a priori* knowledge of the effect of specific process conditions on the surface morphology, an operator can quickly troubleshoot the process before the losses become catastrophic.

5.1.3 Modifying the Approach using Differently Shaped Nodes and Other Graph-theoretic Quantifiers

Thus far, the approach has been applied to surfaces that are homogenous, i.e., the features on the surface are not orientated in any particular direction. This constraint needs to be relaxed for application to manufacturing processes, such as UPM where there are prominent material flow signatures.

For example, the bottom-half of Figure 5-2 shows an Al 6061 surface processed using our UPM setup. The process has a directional component, given the feed marks observed on the surface. Experts have suggested that aspects of the surface, such as form, texture, bearing volume, etc., which are critical to functional performance of the component, are contingent on the direction and areal scale of measurement [6, 56]. Some of these aspects are depicted on the top row of Figure 5-2 (after Jiang , *et al.* [56]).

By adapting the node shape used for graph theoretic quantification (See Eqn. (4-2)), e.g., instead of the pixel rows, one could consider diagonal pixels, or pixel

boxes, and by careful correlation with empirical observations, it might be possible to forward function specific graph theoretic quantification of surface morphology.

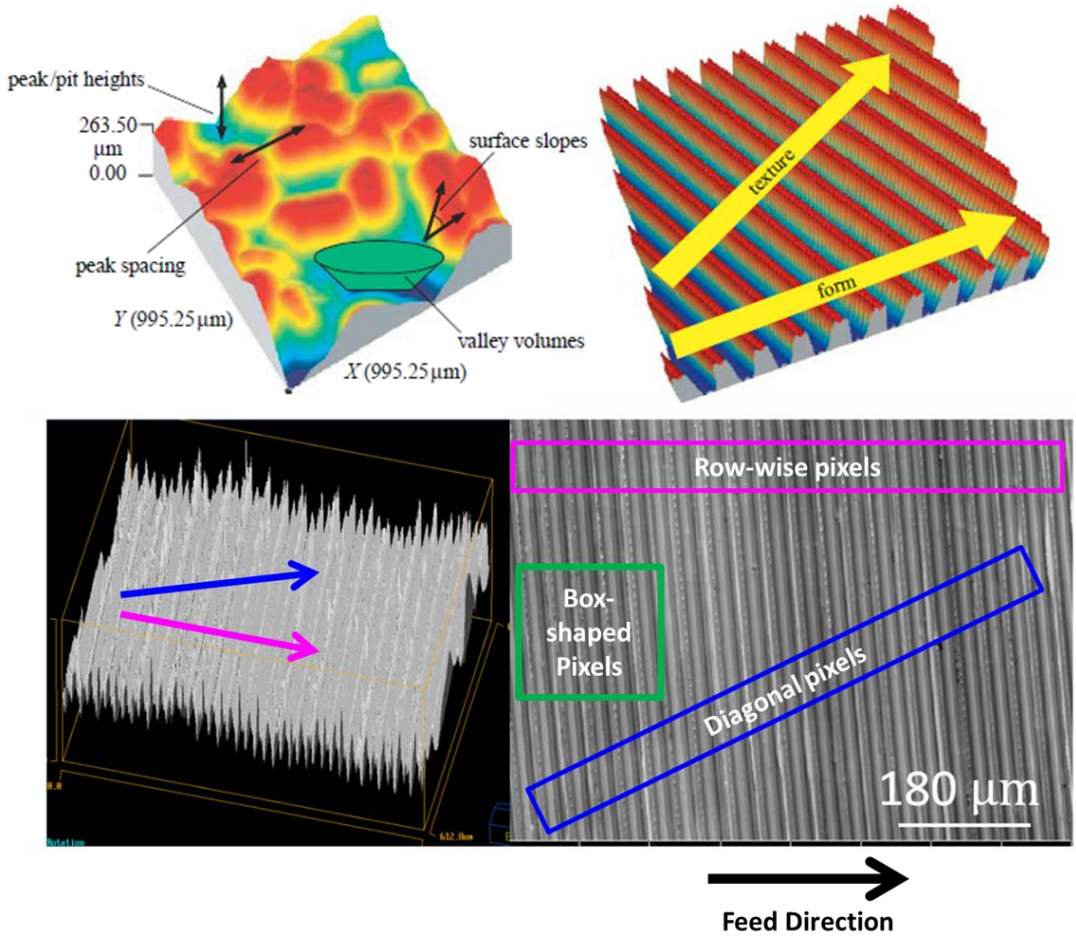


Figure 5-2: Top: Scale-dependent attributes of a surface (after Jiang, *et al.* [56]), and Bottom: A sample UPM surface showing prominent feed marks.

Additionally, apart from a single quantifier (Fiedler number), various other invariants of the graph network, such as mean distance, diameter, ratio of eigen-values, graph coloring parameters, etc., can be used to elucidate the different physical aspects [193, 218, 219].

6 Conclusions

In this chapter we summarize the intellectual merit and contributions of this research. The objectives of this research are reiterated below:

Research Objectives

1. To forward approaches for monitoring and identification of incipient anomalies in nanomanufacturing process, specifically UPM and CMP, based on *in situ* sensing techniques.
2. To suggest methods for surface quality assessment of nanomanufactured ultraprecision components that can be incorporated *in line* and are capable of capturing subtle nanoscale surface morphology variations.

The research satisfying these objectives is divisible into three components:

1. Identification and monitoring of surface morphology variations in ultraprecision machining (UPM) process using *in situ* sensors (detailed in Chapter 2).
2. The integration of a physical process model and wireless sensor data for prediction of process anomalies in chemical mechanical planarization (CMP) (detailed in Chapter 3).
3. The quantification of surface morphology in ultraprecision processes using a graph theoretic approach (detailed in Chapter 4).

6.1 Real-time Identification of Incipient Surface Morphology Variations in Ultraprecision Machining (UPM) Process

This part of the research (detailed in Chapter 2) achieves the first objective in context of the UPM process with the goal stated under.

Goal 1
(Satisfying Objective 1)

Develop an approach that can quantify ultraprecision surface morphologies in a rapid and parsimonious manner.

Intellectual merit

Monitoring of surface quality in ultraprecision machining has been largely pursued on a post-process basis [86, 220]. That is, the surface is characterized *after* the machining operations have been completed. Hence, a defective product is not often detected until it has reached the inspection stage, which compromises process yield.

Sensor-based monitoring in UPM has been investigated extensively [53, 77]. These prior efforts are mostly concerned with investigation of specific phenomena in UPM (e.g., effect of depth of cut, type of cutting regime), as opposed to prediction of incipient process anomalies. As a result, most of the research in sensor-based monitoring in UPM relies upon traditional signal processing techniques, such as frequency spectrum analysis, statistical feature extraction, and data-driven modeling. These methods are not sufficient for predicting the onset of UPM process defects in a timely manner, because:

i. The analysis is mostly *post hoc*, and is therefore reticent. For example, one has to wait close to 2-10 seconds for the signals to be gathered and analyzed before a process drift is detected [17, 77]. It is too late by then, in a practical sense, to take any corrective action. For instance, with typical (conservative) cutting conditions of 4 mm/min feed rate and 1200 RPM, in about five seconds the tool has traversed (linearly) more than three hundred micrometers, and the spindle has made one hundred revolutions.

ii. Current analytical approaches use only one type of sensor for process monitoring; multi-phenomena sensing has not been pursued extensively.

The novelty of this research lies in addressing the above two aspects in UPM process. It provides the mathematical framework for integrating a class of neural networks known as recurrent neural networks (RNN) that have the ability to capture complex signal patterns and accurately predict the process state [79], with responsive Bayesian particle filtering (PF) techniques [81]. By fusing these two approaches, we not only predict surface related defects *in situ* during the process, but also use data from multiple sensors.

This work provides the intellectual foundation for future research in the following aspects:

i. *In situ* and in real-time defect prediction in UPM.

ii. Extends the state-of-art in sensor-based monitoring by fusing two powerful signal analysis techniques, namely, recurrent neural networks and Bayesian particle filtering.

iii. A method for simultaneous multi-phenomena sensing using several signals (7) acquired from different types of sources (vibration, force, and acoustic emission).

iv. Provides a means to analyze non-stationary signal patterns, and overcome computationally intensive retraining of neural networks.

v. Demonstrates a technique to detect surface morphology changes in UPM in a timely manner.

Contributions

Specific contributions of this work are as follows:

i. A diamond turning setup was instrumented with multiple *in situ* sensors for UPM process monitoring.

ii. Near-specular surface finish was obtained on Al 6061 workpieces using conditions identified from empirical tests.

iii. A recurrent predictor neural network - particle filter (RPNN-PF) approach was devised for predicting UPM process anomalies using information from multiple sensors.

iv. Surface morphology variations in UPM were detected within 15 ms of their inception using the RPNN-PF approach.

6.2 Process-Machine Interaction (PMI) Model-based Monitoring of the Chemical Mechanical Planarization (CMP) Process using Wireless Vibration Sensors

This work is described in Chapter 3 (see also Appendices I, II) and is intended toward achieving the first objective in context of CMP. This work accomplishes the following two goals (continuing from the previous section):

Goal 2
(Satisfying Objective 1)

Identify and statistically quantify those aspects of MEMS wireless sensor signal components that are relevant to the CMP process dynamics, and therefore useful from a real-time monitoring perspective.

Goal 3
(Satisfying Objective 1)

Explain based on process dynamics and physical phenomena, the reason for certain vibration sensor patterns observed in CMP, and thereby facilitate real-time process prognosis.

Intellectual merit

In this work, we integrated (sensor) data-driven process monitoring methods with a physical model of the CMP process. Due to this outcome, we showed that process drifts can be identified from sensor data patterns, and predicted *a priori* from the process dynamics. In other words, we not only identify that there is a change in the process state, but also, (i) *why* there is a change, and (ii) more pertinently, *when* such a change in signal patterns is likely to occur.

Consider for example, the glazing of the polishing pad during CMP [153, 165]. This process anomaly can damage the semiconductor wafer surface, inhibit the surface quality, and consequentially affect the process yield. Sensor data alone may not be sufficient to identify the onset of pad glazing, given the complex nature of signal patterns in CMP.

This is because the signal patterns indicating pad glazing may not be obvious, and are often dominated by other sources, such as from spindle errors, extraneous environmental conditions, variations in polishing load, and the characteristic electromagnetic response of the sensor. In other words, the signals are contaminated, and information relevant to the process state is occluded by the complex and broadband characteristics of the signal patterns.

By using a hierarchical multi-scale physical model to describe different aspects of the process (the detailed formulation of the model is documented in Appendix III), we can identify *beforehand* the physical source of signal patterns. This allows us to focus solely on signal patterns contingent to certain phenomena, such as pad glazing, and thereby anticipate process drifts. The physical model essentially indicates what to look for in the signal patterns when certain types of process drifts are imminent.

The integration of a nonlinear differential equation model incorporating multi-scale aspects of the process with wirelessly acquired vibration signals for predicting anomalies in a CMP process is the novel intellectual outcome of this work.

This completes the loop between physics-based modeling and data-driven monitoring in CMP, an aspect that has not been addressed thus far in the literature.

Contributions

The following are some of the practical contributions from this work:

i. A CMP setup was integrated with miniature MEMS sensors of different types (vibration and sound). This enabled sustained wireless data acquisition at moderately high sampling rates (> 750 Hz).

ii. The complex non-stationary and nonlinear CMP vibration signal patterns were studied using contemporary analysis tools (recurrence quantification analysis, signal dimensions, and chaotic embedding), and their behavior correlated with process conditions.

iii. A nonlinear differential equation model of CMP process dynamics was invoked for explaining some of the complex characteristics of experimentally acquired vibration signals in CMP.

iv. The PMI-model generated solutions when used in conjunction with experimentally acquired vibration signals were able to capture the CMP process dynamics under evolving scenarios (e.g., changes in polishing load, pad glazing, spindle speed) with high fidelity ($R^2 > 75\%$).

Thus, this work enables *in situ* monitoring of CMP process using sensor data, which is closely monitored using a physical model.

6.3 A Graph Theoretic Approach for Quantification of Ultraprecision Surface Morphology

This work detailed in Chapter 4, achieves the second objective for quantification of nanoscale surface morphologies, and has the following goal.

Goal 4
(Satisfying Objective 2)

Develop an approach that can quantify ultraprecision surface morphologies in a rapid and parsimonious manner.

Intellectual merit

Approaches for quantification of surface morphology rely largely on statistical parameters [8, 54]. Experts in this area have identified the inadequacy of characterizing surfaces, especially nanoscale surfaces, using statistical parameters [5, 9, 29, 56, 57]. However, instead of exploring fundamentally novel quantification approaches, more statistical parameters are continually added to the extensive set already available [171].

In this research, instead of trying to fit a surface topography to a statistical distribution (which is the essence of statistical metrology), and subsequently extract all manner of moments from the resulting distribution, we try to unearth the underlying geodesic relationship from the topography of a surface using graph theoretic techniques. The difference is analogous to characterizing a city by measuring the average height of its buildings (statistical approach) vs. analyzing the density and connections of its streets (graph theoretic approach).

Such a graph theoretic quantification of surface morphology has not been attempted thus far. The methods developed to represent a surface in graph theoretic form, and subsequently quantify its morphology with graph invariants are the novel outcomes of this research.

The graph theoretic representation succeeds in mapping the topological relationships of a surface, and is therefore distinct from the statistical distribution mapping approach of conventional metrology. We have shown with various simulated and experimentally acquired surfaces that graph theoretic invariants, such as Fiedler number are more responsive to subtle changes in surface morphology (particularly for nanoscale surfaces), compared to traditional statistics-based quantifiers. The intellectual merit of this work is justified on the following grounds:

i. Methods for representing surfaces in graph theoretic form from optical images were developed.

ii. The graph theoretic invariant Fiedler number (λ_2) was tested as a means to detect surface morphology changes. Using both simulated, and empirically obtained surfaces, graph theoretic invariants were shown to be responsive discriminants of surface morphology variations that were not captured using traditional statistical measurements.

Contributions

The graph theoretic approach for quantifying nano-surface morphology offers several practical benefits:

i. Since surfaces can be quantified using optical micrograph images, reticent surface mapping techniques are precluded. Because optical microscopy is quicker than profile mapping techniques (SEM, AFM, and laser interferometry), the approach is easily adapted for *in-line* surface quality monitoring. This is valuable from a process control standpoint, because graph theoretic invariants (e.g., Fielder number (λ_2)) can be used as a means to identify process anomalies in a timely manner, and therefore expensive yield losses can be minimized.

ii. Since optical imaging is non-contact and non-destructive, the possibility of damage to the specimen is avoided. For instance, if a diamond/tungsten tipped stylus is used to measure a mirror finished aluminum part, a scratch will be inevitably generated on the surface, i.e., the process of measuring the surface can affect the functional integrity of the component.

iii. Nanoscale surfaces have heterogeneous features occurring at multiple scales. The measurement of multi-scale characteristics using wavelet and fractal approaches is computationally expensive. The graph theoretic representation is a significantly faster approach and does not require profile mapping at different scales.

iv. We have shown using near-specular CMP processed surfaces that the graph theoretic invariant Fiedler number (λ_2) captured visibly prominent surface morphology changes that were not identified using statistical quantifiers, such as arithmetic mean roughness (Ra), root mean square roughness (Rq), skewness (Rsq) and kurtosis (Rku).

6.4 Closure

This research presented approaches for real-time *in situ* monitoring and inspection of surface morphology in two precision nanomanufacturing applications, ultraprecision machining (UPM) and chemical mechanical planarization (CMP). The key outcomes from this research are listed herewith:

1. An ultraprecision diamond turning machine is instrumented with three miniature accelerometers, a 3-axis piezoelectric dynamometer, and an acoustic emission (AE) sensor for process monitoring. The machine tool is used for face turning aluminum 6061 discs to a surface finish (Sa) in the range of 15-25 nm. A real-time monitoring approach, based on predicting and updating the process states from sensor signals (using advanced neural networks and Bayesian analysis) is reported for detecting the incipient surface variations in UPM. Experimental investigations show that variations in surface characteristics during UPM can be detected within 15 ms of their inception using the RPNN-PF approach, as opposed to 30 ms or higher with the conventional statistical change detection methods tested.

2. A Buehler (model Automet® 250) bench top CMP machine was instrumented with miniature MEMS 3-axis accelerometer (Analog Devices ADXL 335) and audio sensors (Analog Devices ADMP 401). The CMP setup is used for finishing blanket copper workpieces to a surface finish of Ra ~ 5 nm. While the sensor signals are sensitive to variations in the CMP process, the extraneous noise prevents the direct use of raw signal patterns for early detection of defects. Consequently, instead of primarily monitoring the raw sensor signal patters, we isolated signal features that are indicative of process state from those that are mere artifacts, and thus potentially valuable for process monitoring.
3. A deterministic process-machine interaction (PMI) model that can associate different complex time-frequency patterns to specific CMP process mechanisms was invoked. The PMI model captures the effects of the non-uniform structural properties of the polishing pad, pad asperities, and machine kinematics on CMP dynamics using a deterministic two degree of freedom nonlinear differential equation.
4. The deterministic PMI model were used to explain not just the physical sources of various time-frequency patterns observed in the measured vibration signals, but also their variations with CMP process conditions. The features extracted from experimental CMP vibration data, such as power spectral density over the 115 − 120 Hz band, and nonlinear recurrence measures were statistically significant estimators (R^2 ~ 75%) of process parameter settings.

5. We developed an algebraic graph theoretic approach for surface morphology quantification from optical micrographs of ultraprecision components, such as CMP processed copper (Cu) wafers polished to near-specular finish. The approach is based on treating various heterogeneous features (e.g., pits, ridges, scratches) as a stationary mixture of random fields, allowing for compact graphical representation of the surface morphology. Two complementary methods to realize graph representations and subsequently overcome some of the lacunae associated with conventional statistics-based surface metrology quantifiers were reported. Our experimental findings established close correlations ($\rho \sim 80\% - 99\%$) between the graph topological invariant Fiedler number and conventional surface roughness statistics measured over local scales.
6. The graph theoretic approach, by primarily employing optical microscopy, eschews relatively reticent profile mapping techniques, such as laser interferometry and atomic force microscopy. Thus, by invoking the algebraic graph theoretic topological invariant Fiedler number of the graph reconstructed from the images, surface morphology of ultraprecision components, such as semiconductor wafers can be assessed from optical micrographs and therefore enable timely, non-contact, and *in situ* metrology of such surfaces.

References

[1] N. Ikawa, R. Donaldson, R. Komanduri, W. Konig, P. McKeown, T. Moriwaki, and I. Stowers, "Ultraprecision metal cutting — the past, the present and the future," *CIRP Annals - Manufacturing Technology,* (**40)**2, pp. 587-594, 1991.

[2] N. Taniguchi, "Current status in, and future trends of, ultraprecision machining and ultrafine materials processing," *CIRP Annals - Manufacturing Technology,* (**32)**2, pp. 573-582, 1983.

[3] R. Komanduri, D. A. Lucca, and Y. Tani, "Technological advances in fine abrasive processes," *CIRP Annals - Manufacturing Technology,* (**46)**2, pp. 545-596, 1997.

[4] N. Taniguchi, *Nanotechnology: Integrated processing systems for ultra-precision and ultra-fine products*. Oxford, UK.: Oxford University Press, 1996.

[5] L. De Chiffre, H. Kunzmann, G. Peggs, and D. Lucca, "Surfaces in precision engineering, microengineering and nanotechnology," *CIRP Annals-Manufacturing Technology,* (**52)**2, pp. 561-577, 2003.

[6] X. Jiang, "Precision surface measurement," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* (**370)**1973, pp. 4089-4114, 2012.

[7] V. Venkatesh, F. Fang, and W. Chee, "On-mirror surfaces obtained with and without polishing," *CIRP Annals - Manufacturing Technology,* (**46)**1, pp. 505-508, 1997.

[8] L. De Chiffre, P. Lonardo, H. Trumpold, D. Lucca, G. Goch, C. Brown, J. Raja, and H. N. Hansen, "Quantitative characterisation of surface texture," *CIRP Annals-Manufacturing Technology,* (**49)**2, pp. 635-652, 2000.

[9] P. M. Lonardo, D. A. Lucca, and L. De Chiffre, "Emerging trends in surface metrology," *CIRP Annals - Manufacturing Technology,* (**51)**2, pp. 701-723, 2002.

[10] E. DeGarmo, J. T. Black, and R. Kosher, *Material and processing in manufacturing*. New York, NY: John Wiley and Sons, 2003.

[11] M. Weck, S. Fischer, and M. Vos, "Fabrication of microcomponents using ultraprecision machine tools," *Nanotechnology*, (**8)**3, p. 145, 1997.

[12] P. McKeown, "The role of precision engineering in manufacturing of the future," *CIRP Annals - Manufacturing Technology*, (**36)**2, pp. 495-501, 1987.

[13] J. B. Bryan, "Design and construction of an ultraprecision 84 inch diamond turning machine," *Precision Engineering*, (**1)**1, pp. 13-17, 1979.

[14] J. Klingmann. (2001) The world's most accurate lathe. *Science and Technology Review*. 12-14.

[15] J. Roblee. (11/09/2011). *Factors affecting surface finish in diamond turning*. Available: http://www.precitech.com/pressroom/FactorsAffectingSurfFinish.pdf

[16] R. Dahlgren and M. C. Gerchman. (1987, 11/09/2011). *The use of aluminum alloy castings as diamond machining substrates*. Available: http://www.precitech.com/pressroom/The%20Use%20of%20Aluminum%20Alloy%20Castings%20as%20Diamond%20Machining%20Substrates.pdf

[17] C. F. Cheung and W. Lee, *Surface generation in ultra-precision diamond turning: Modelling and practices*. Suffolk, UK: Professional Engineering Publishing Limited, 2003.

[18] I. Stowers, R. Komanduri, and E. Baird, "Review of precision surface generating processes and their potential application to the fabrication of large optical components," in *Proceedings of the SPIE*, 1988, pp. 62-73.

[19] P. B. Zantye, A. Kumar, and A. K. Sikder, "Chemical mechanical planarization for microelectronics applications," *Materials Science and Engineering - R: Reports*, (**45)**3-6, pp. 89-220, 2004.

[20] R. Doering and Y. Nishi, *Handbook of semiconductor manufacturing technology*. Boca Raton, FL: CRC Press, 2007.

[21] P. Van Zant, *Microchip fabrication*. New York, NY: McGraw-Hill, 2004.

[22] J. Moyne, E. Del Castillo, and A. M. Hurwitz, *Run-to-run control in semiconductor manufacturing*. Boca Raton, FL: CRC Press, 2001.

[23] J. D. Morillo, T. Houghton, J. M. Bauer, R. Smith, and R. Shay, "Edge and bevel automated defect inspection for 300 mm production wafers in manufacturing," in

Advanced Semiconductor Manufacturing Conference and Workshop, 2005 IEEE/SEMI, 2005, pp. 49-52.

[24] H. S. Bennett, "Will future measurement needs of the semiconductor industry be met?," *National Institute of Science and Technology, Journal of Research,* (**112)**1, p. 25, 2007.

[25] M. Berman, T. Bibby, and A. Smith, "Review of *in situ* & in-line detection for chemical mechanical planarization applications," *Semiconductor Fabtech,* (**8)**pp. 267-274, 1998.

[26] M. B. Steele and J. A. Hawthorne, "Semiconductor wafer inspection system," United States Patent 6,957,154 B2.

[27] S. Bukkapatnam, S. Kamarthi, Q. Huang, A. Zeid, and R. Komanduri, "Nanomanufacturing systems: Opportunities for industrial engineers," *IIE Transactions,* (**44)**7, pp. 492-495, 2012/07/01 2012.

[28] J.-C. Lu, S.-L. Jeng, and K. Wang, "A review of statistical methods for quality improvement and control in nanotechnology," *Journal of Quality Technology,* (**41)**2, p. 148, 2009.

[29] T. R. Thomas, "Trends in surface roughness," *International Journal of Machine Tools and Manufacture,* (**38)**5–6, pp. 405-411, 1998.

[30] C. Y. Poon and B. Bhushan, "Comparison of surface roughness measurements by stylus profiler, AFM and non-contact optical profiler," *Wear,* (**190)**1, pp. 76-88, 1995.

[31] R. P. Feynman, "There's plenty of room at the bottom," *Engineering and Science,* (**23)**5, pp. 22-36, 1960.

[32] "PBS Nova, Making Stuff: Smaller," ed, 2011.

[33] W. Ahmed and M. J. Jackson, *Emerging nanotechnologies for manufacturing*. Oxford, UK.: William Andrew, 2009.

[34] J. Corbett, P. A. McKeown, G. N. Peggs, and R. Whatmore, "Nanotechnology: International developments and emerging products," *CIRP Annals - Manufacturing Technology,* (**49)**2, pp. 523-545, 2000.

[35] "Societal implications of nanoscience and nanotechnology," National Science Foundation, 2001.

[36] "National nanotechnology initiative strategic plan," National Science and Technology Council, 2011.

[37] G. E. Moore, "Cramming more components onto integrated circuits (reprinted from electronics, volume 38, number 8, april 19, 1965, pp. 114)," *IEEE Solid-State Circuits Newsletter,* (**11)**5, pp. 33-35, 2006.

[38] (2011, 10/20/2012). *Wikepedia page on moore's law, www.wikipedia.org*.

[39] F. W. Taylor, "On the art of cutting metals," *Transactions of the ASME, Annual Meeting,* 1907.

[40] K. Blaedel and D. Thompson. (1998) Engineering precision into laboratory projects. *Science and Technology Review*. 12-21.

[41] W. J. Wills-Moren, H. Modjarrad, R. F. J. Read, and P. A. McKeown, "Some aspects of the design and development of a large high precision cnc diamond turning machine," *CIRP Annals - Manufacturing Technology,* (**31)**1, pp. 409-414, 1982.

[42] J. Steigerwald. (2009). *Chemical mechanical polish: The enabling technology*. Available: http://www.techdesignforums.com/eda/volumes/volume-6/march-2009/chemical-mechanical-polish-the-enabling-technology/

[43] G. Bylinsky. (2003, Closing in on perfection ultraprecision machine tools are putting manufacturers within nanometers of absolute accuracy http://money.Cnn.Com/magazines/fortune/fortune_archive/2003/06/23/344598/index.Htm. *Fortune Magazine*.

[44] G. Chapman, "Ultra-precision machining systems; an enabling technology for perfect surfaces," *Moore Nanotechnology Systems,* 2004.

[45] P. J. Davis. (1994) Artificial hip joints: Applying weapons expertise to medical technology *Engineering and Technology Review*. 16-21.

[46] D. C. Thompson, J. L. Chrislock, and L. E. Newton, "Development of an inexpensive, high-accuracy diamond turning machine," *Precision Engineering,* (**4)**2, pp. 73-77, 1982.

[47] C. J. Burrows, J. A. Holtzman, S. Faber, P. Y. Bely, H. Hasan, C. Lynds, and D. Schroeder, "The imaging performance of the hubble space telescope," *The Astrophysical Journal,* (**369)**pp. L21-L25, 1991.

[48] G. Byrne, D. Dornfeld, I. Inasaki, G. Ketteler, W. König, and R. Teti, "Tool condition monitoring (TCM)—the status of research and industrial application," *CIRP Annals - Manufacturing Technology,* (**44)**2, pp. 541-567, 1995.

[49] D.-E. Lee, I. Hwang, C. M. Valente, J. Oliveira, and D. A. Dornfeld, "Precision manufacturing process monitoring with acoustic emission," *International Journal of Machine Tools and Manufacture,* (**46)**2, pp. 176-188, 2006.

[50] R. Teti, K. Jemielniak, G. O'Donnell, and D. Dornfeld, "Advanced monitoring of machining operations," *Annals of the CIRP - Manufacturing Technology,* (**59)**2, pp. 717-739, 2010.

[51] D. Snr and E. Dimla, "Sensor signals for tool-wear monitoring in metal cutting operations—a review of methods," *International Journal of Machine Tools and Manufacture,* (**40)**8, pp. 1073-1098, 2000.

[52] A. G. Rehorn, J. Jiang, and P. E. Orban, "State-of-the-art methods and results in tool condition monitoring: A review," *The International Journal of Advanced Manufacturing Technology,* (**26)**7-8, pp. 693-710, 2005.

[53] D. E. Lee, I. Hwang, C. M. O. Valente, J. F. G. Oliveira, and D. A. Dornfeld, "Precision manufacturing process monitoring with acoustic emission," *International Journal of Machine Tools and Manufacture,* (**46)**2, pp. 176-188, 2006.

[54] J. Peters, J. B. Bryan, W. T. Estler, C. Evans, H. Kunzmann, D. A. Lucca, S. Sartori, H. Sato, E. G. Thwaite, P. Vanherck, R. J. Hocken, J. Peklenik, T. Pfeifer, H. Trumpold, and T. V. Vorburger, "Contribution of cirp to the development of metrology and surface quality evaluation during the last fifty years," *CIRP Annals - Manufacturing Technology,* (**50)**2, pp. 471-488, 2001.

[55] X. Jiang, P. Scott, and D. Whitehouse, "Wavelets and their applications for surface metrology," *CIRP Annals - Manufacturing Technology,* (**57)**1, pp. 555-558, 2008.

[56] X. Jiang, P. J. Scott, D. Whitehouse, and L. Blunt, "Paradigm shifts in surface metrology. Part 2 : The current shift," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science,* (**463)**2085, pp. 2071-2099, 2007.

[57] D. Whitehouse, "Surface metrology," *Measurement Science and Technology,* (**8)**p. 955, 1997.

[58] P. Rao, S. Bukkapatnam, O. Beyca, Z. J. Kong, and R. Komanduri, "Real-time identification of incipient surface morphology variations in ultra-precision machining process," *Transactions of the ASME, Journal of Manufacturing Science and Engineering,* (**In-Press)** 2013.

[59] P. Rao, M. B. Bhushan, S. Bukkapatnam, Z. Kong, S. Byalal, O. Beyca, A. Fields, and R. Komanduri, "Process-machine interaction (PMI) modeling and monitoring of chemical mechanical planarization (CMP) process using wireless vibration sensors," *IEEE Transactions in Semiconductor Manufacturing,* (**Under Review)** 2013.

[60] D. Dornfeld, S. Min, and Y. Takeuchi, "Recent advances in mechanical micromachining," *CIRP Annals - Manufacturing Technology,* (**55)**2, pp. 745-768, 2006.

[61] N. Ikawa, "Minimum thickness of cut in micromachining," *Nanotechnology,* (**3)**1, p. 6, 1992.

[62] S. Takasu, M. Masuda, T. Nishiguchi, and A. Kobayashi, "Influence of study vibration with small amplitude upon surface roughness in diamond machining," *CIRP Annals - Manufacturing Technology,* (**34)**1, pp. 463-467, 1985.

[63] T. Moriwaki, "Effect of cutting heat on machining accuracy in ultra-precision diamond turning," *CIRP Annals - Manufacturing Technology,* (**39)**1, pp. 81-84, 1990.

[64] X. Liu, R. E. DeVor, S. G. Kapoor, and K. F. Ehmann, "The mechanics of machining at the microscale: Assessment of the current state of the science," *Transactions of the ASME, Journal of Manufacturing Science and Engineering,* (**126)**4, pp. 666-678, 2004.

[65] R. G. Jasinevicius, J. G. Duduch, L. Montanari, and P. S. Pizani, "Phase transformation and residual stress probed by raman spectroscopy in diamond-turned single crystal silicon," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture,* (**222)**9, p. 9p, 2008.

[66] N. Ikawa, S. Shimada, H. Tanaka, and G. Ohmori, "An atomistic analysis of nanometric chip removal as affected by tool-work interaction in diamond turning," *CIRP Annals - Manufacturing Technology,* (**40)**1, pp. 551-554, // 1991.

[67] D. Lucca, R. L. Rohrer, and R. Komanduri, "Energy dissipation in the ultraprecision machining of copper," *Annals of the CIRP - Manufacturing Technology,* (**40)**1, pp. 69-72, 1991.

[68] D. A. Lucca, Y. W. Seo, and R. Komanduri, "Effect of tool edge geometry on energy dissipation in ultraprecision machining," *CIRP Annals - Manufacturing Technology,* (**42)**1, pp. 83-86, 1993.

[69] X. Liu, R. E. DeVor, and S. G. Kapoor, "An analytical model for the prediction of minimum chip thickness in micromachining," *Journal of Manufacturing Science and Engineering,* (**128)**2, pp. 474-481, 2005.

[70] S. Asai, Y. Taguchi, K. Horio, T. Kasai, and A. Kobayashi, "Measuring the very small cutting-edge radius for a diamond tool using a new kind of SEM having two detectors," *CIRP Annals - Manufacturing Technology,* (**39)**1, pp. 85-88, // 1990.

[71] C. Cheung and W. Lee, "Modelling and simulation of surface topography in ultra-precision diamond turning," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture,* (**214)**6, pp. 463-480, 2000.

[72] O. Abouelatta and J. Madl, "Surface roughness prediction based on cutting parameters and tool vibrations in turning operations," *Journal of Materials Processing Technology,* (**118)**1-3, pp. 269-277, 2001.

[73] C. Beggan, M. Woulfe, P. Young, and G. Byrne, "Using acoustic emission to predict surface quality," *The International Journal of Advanced Manufacturing Technology,* (**15)**10, pp. 737-742, 1999.

[74] M. Hayashi, H. Yoshioka, and H. Shinno, "An adaptive control of ultraprecision machining with an in-process micro-sensor," *Journal of Advanced Mechanical Design, Systems, and Manufacturing,* (**2)**3, pp. 322-331, 2008.

[75] H. Yoshioka, S. Matsumura, H. Hashizume, and H. Shinno, "Minimizing thermal deformation of aerostatic spindle system by temperature control of supply air," *JSME International Journal, Series C : Mechanical Systems, Machine Elements and Manufacturing,* (**49)**2, pp. 606-611, 2006.

[76] H. Shinno, H. Hashizume, and H. Sato, "In-process monitoring method for machining environment based on simultaneous multiphenomena sensing," *CIRP Annals - Manufacturing Technology,* (**46)**1, pp. 53-56, 1997.

[77] W. B. Lee, C. F. Cheung, and S. To, "Characteristics of microcutting force variation in ultraprecision diamond turning," *Materials and Manufacturing Processes,* (**16)**2, pp. 177-193, 2001.

[78] E. R. Marsh and A. J. Schaut, "Measurement and simulation of regenerative chatter in diamond turning," *Precision Engineering*, (**22)**4, pp. 252-257, 1998.

[79] L. C. Jain and L. R. Medsker, *Recurrent neural networks: Design and applications*. Boca Raton, FL: CRC Press, 2000.

[80] M. Han, J. Xi, S. Xu, and F. L. Yin, "Prediction of chaotic time series based on the recurrent predictor neural network," *IEEE Transactions on Signal Processing*, (**52)**12, pp. 3409-3416, 2004.

[81] A. Doucet, N. De Freitas, and N. Gordon, *Sequential monte carlo methods in practice*. New York, NY: Springer Verlag, 2001.

[82] T. Moriwaki and K. Okuda, "Machinability of copper in ultra-precision micro diamond cutting," *CIRP Annals - Manufacturing Technology*, (**38)**1, pp. 115-118, 1989.

[83] W. Prager, "The theory of plasticity: A survey of recent achievements," *Proceedings of the Institution of Mechanical Engineers*, (**169)**1955, pp. 41-57, 1955.

[84] M. Kong, W. Lee, C. Cheung, and S. To, "A study of materials swelling and recovery in single-point diamond turning of ductile materials," *Journal of Materials Processing Technology*, (**180)**1-3, pp. 210-215, 2006.

[85] C. F. Cheung and W. B. Lee, "A framework of a virtual machining and inspection system for diamond turning of precision optics," *Journal of Materials Processing Technology*, (**119)**1-3, pp. 27-40, 2001.

[86] K. C. Chan, C. F. Cheung, M. V. Ramesh, W. B. Lee, and S. To, "A theoretical and experimental investigation of surface generation in diamond turning of an al6061/sic_p metal matrix composite," *International Journal of Mechanical Sciences*, (**43)**9, pp. 2047-2068, 2001.

[87] H. Hocheng and M. L. Hsieh, "Signal analysis of surface roughness in diamond turning of lens molds," *International Journal of Machine Tools and Manufacture*, (**44)**15, pp. 1607-1618, 2004.

[88] S. T. S. Bukkapatnam, A. Lakhtakia, and S. R. T. Kumara, "Analysis of sensor signals shows turning on a lathe exhibits low-dimensional chaos," *Physical Review - E.*, (**52)**3, pp. 2375-2387, 1995.

[89] J. D. Hamilton, *Time series analysis*. Princeton, NJ: Princeton University Press, 1994.

[90] L. Li, J. S. A. Collins, and A. Y. Yi, "Optical effects of surface finish by ultraprecision single point diamond machining," *Transactions of the ASME, Journal of Manufacturing Science and Engineering,* (**132)**2, pp. 021002-9, 2010.

[91] C. F. Cheung, K. C. Chan, S. To, and W. B. Lee, "Effect of reinforcement in ultra-precision machining of al6061/sic metal matrix composites," *Scripta Materialia,* (**47)**2, pp. 77-82, 2002.

[92] H. Wang, S. To, C. Y. Chan, C. F. Cheung, and W. B. Lee, "A theoretical and experimental investigation of the tool-tip vibration and its influence upon surface generation in single-point diamond turning," *International Journal of Machine Tools and Manufacture,* (**50)**3, pp. 241-252, 2010.

[93] E. Paul, C. J. Evans, A. Mangamelli, M. L. McGlauflin, and R. S. Polvani, "Chemical aspects of tool wear in single point diamond turning," *Precision Engineering,* (**18)**1, pp. 4-19, 1996.

[94] M. T. Hagan, H. B. Demuth, and M. Beale, *Neural network design*. Boston, MA.: PWS Publishing, 1997.

[95] O. DeJesús and M. T. Hagan, "Backpropagation through time for a general class of recurrent network," in *11th International Joint Conference on Neural Networks*, Washington, D.C., 2001, pp. 2638-2642.

[96] Z. Kong, A. Oztekin, O. F. Beyca, U. Phatak, S. T. S. Bukkapatnam, and R. Komanduri, "Process performance prediction for chemical mechanical planarization (CMP) by integration of nonlinear bayesian analysis and statistical modeling," *IEEE Transactions on Semiconductor Manufacturing,* (**23)**2, pp. 12 -, 2010.

[97] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* (**24)**5, pp. 603-619, 2002.

[98] M. C. Shaw, "The chemico-physical role of the cutting fluid," *Metal Progress,* (**42)**pp. 85-91, 1942.

[99] E. Brinksmeier, D. A. Lucca, and A. Walter, "Chemical aspects of machining processes," *CIRP Annals - Manufacturing Technology,* (**53)**2, pp. 685-699, 2004.

[100] T. J. Burns and M. A. Davies, "Nonlinear dynamics model for chip segmentation in machining," *Physical Review Letters,* (**79)**3, pp. 447-450, Jul 1997.

[101] H. Kantz and T. Schreiber, *Nonlinear time series analysis*. Cambridge, UK: Cambridge University Press, 1997.

[102] S. P. Murarka, I. V. Verner, and R. J. Gutmann, *Copper - fundamental mechanisms for microelectronic applications*. New York, NY: John Wiley & Sons, 2000.

[103] U. Paik and J. G. Park, *Nanoparticle engineering for chemical-mechanical planarization: Fabrication of next-generation nanodevices*. Boca Raton, FL: CRC Press, 2009.

[104] S. Prasad, W. Loh, A. Kapoor, E. Chang, B. Stine, D. Boning, and J. Chung, "Statistical metrology for characterizing chemical mechanical planarization processes," *Microelectronic Engineering,* (**33)**1–4, pp. 231-240, 1997.

[105] D. S. Boning, W. P. Moyne, T. H. Smith, J. Moyne, R. Telfeyan, A. Hurwitz, S. Shellman, and J. Taylor, "Run by run control of chemical-mechanical polishing," *IEEE Transactions on Components,Packaging and Manufacturing Technology - C,* (**19)**4, pp. 419 - 430, Oct. 1996.

[106] E. Sachs, A. Hu, and A. Ingolfsson, "Run by run process control: Combining spc and feedback control," *IEEE Transactions on Semiconductor Manufacturing,* (**8)**1, pp. 26-43, 1995.

[107] A. Hu, D. He, S. Wong, P. Renteln, and E. Sachs, "Application of run by run controller to the chemical-mechanical planarization process - part 2," in *'Low-Cost Manufacturing Technologies for Tomorrow's Global Economy'. Proceedings 1994 IEMT Symposium*, 1994, pp. 371-378 vol.1.

[108] T. Bibby and K. Holland, "Endpoint detection in CMP," *Journal of Electronic Materials,* (**27)**pp. 1073-1081, 1998.

[109] H. Hocheng and Y. L. Huang, "A comprehensive review of endpoint detection in chemical mechanical planarisation for deep-submicron integrated circuits manufacturing," *International Journal of Materials and Product Technology,* (**18)**4, pp. 469-486, 2003.

[110] Z. Kong, A. Oztekin, O. F. Beyca, U. Phatak, S. Bukkapatnam, and R. Komanduri, "Process performance prediction for chemical mechanical planarization (CMP) by integration of nonlinear bayesian analysis and statistical

modeling," *IEEE Transactions on Semiconductor Manufacturing,* (**23)**2, pp. 316-327, 2010.

[111] Z. Kong, O. Beyca, S. Bukkapatnam, and R. Komanduri, "Nonlinear sequential bayesian analysis-based decision making for end-point detection of chemical mechanical planarization (CMP) processes," *IEEE Transactions on Semiconductor Manufacturing,* (**24)**4, pp. 523-532, 2011.

[112] H. Jeong, H. Kim, S. Lee, and D. Dornfeld, "Multi-sensor monitoring system in chemical mechanical planarization for correlations with process issues," *CIRP Annals - Manufacturing Technology,* (**55)**1, pp. 325-328, 2006.

[113] P. W. Carter and T. Werts, "A method for measuring frictional forces and shaft vibrations during chemical mechanical polishing," *Journal of The Electrochemical Society,* (**154)**1, pp. H60-H66, 2007.

[114] D. L. Hetherington, D. J. Stein, J. P. Lauffer, E. E. Wyckoff, and D. M. Shingledecker, "Analysis of *in-situ* vibration monitoring for end-point detection of chemical mechanical planarization processes," in *EUROPTO conference on In-Line Characterization, Yield Reliability, and Failure Analyses in Microelectronic Manufacturing*, Edinburgh, United Kingdom, 1999, pp. 89-101.

[115] H.-Y. Lee, "Development of the piezoelectric based polishing system for low stress chemical mechanical planarization (CMP)," M.S. Thesis, Mechanical Engineering, National Cheung Cheng University, Taiwan, 2005.

[116] S. Bukkapatnam, P. Rao, and R. Komanduri, "Experimental dynamics characterization and monitoring of material removal rate (MRR) in oxide chemical mechanical planarization (CMP) process," *International Journal of Machine Tools and Manufacture,* (**48)**12-13, pp. 1375-1386, 2008.

[117] U. Phatak, S. Bukkapatnam, Z. Kong, and R. Komanduri, "Sensor-based modeling of slurry chemistry effects on the material removal rate in copper-chemical mechanical planarization process," *International Journal of Machine Tools and Manufacture,* (**49)**2, pp. 171-181, 2009.

[118] A. Ohri, "Wireless sensor fusion approach for monitoring chemical mechanical planarization (CMP) process," M.S Thesis, Mechanical Engineering, Oklahoma State University, Stillwater, OK, 2010.

[119] P. W. Carter and J. P. Chamberlain, "CMP process involving frequency analysis-based monitoring," United States Patent 6,431,953 B1, Dated: Aug. 13, 2002.

[120] Y. Yamada, M. Kawakubo, O. Hirai, N. Konishi, S. Kurokawa, and T. Doi, "Tribological behavior of metal CMP and detection of process abnormality," *Journal of The Electrochemical Society*, (**155)**p. H569, 2008.

[121] T. Kojima, M. Miyajima, F. Akaboshi, T. Yogo, S. Ishimoto, and A. Okuda, "Application of CMP process monitor to cu polishing," *IEEE Transactions on Semiconductor Manufacturing*, (**13)**3, pp. 293-299, 2000.

[122] Y. Homma, K. Fukushima, S. Kondo, and N. Sakuma, "Effects of mechanical parameters on CMP characteristics analyzed by two-dimensional frictional-force measurement," *Journal of The Electrochemical Society*, (**150)**12, pp. G751-G757, 2003.

[123] J. Yi, "On the wafer/pad friction of chemical mechanical planarization (CMP) processes part 2: Experiments and applications," *IEEE Transactions on Semiconductor Manufacturing*, (**18)**3, pp. 371-383, 2005.

[124] X. Chi, G. Dongming, J. Zhuji, and K. Renke, "A signal processing method for the friction-based endpoint detection system of a CMP process," *Journal of Semiconductors*, (**31)**12, p. 126002, 2010.

[125] J. Tang, D. Dornfeld, S. K. Pangrle, and A. Dangca, "In-process detection of microscratching during CMP using acoustic emission sensing technology," *Journal of Electronic Materials*, (**27)**10, pp. 1099-1103, 1998.

[126] R. Ganesan, T. K. Das, A. K. Sikder, and A. Kumar, "Wavelet-based identification of delamination defect in CMP (cu-low k) using nonstationary acoustic emission signal," *IEEE Transactions on Semiconductor Manufacturing*, (**16)**4, pp. 677-685, 2003.

[127] H. Hocheng and Y.-L. Huang, "*In-situ* endpoint detection by pad temeperature in chemical mehanical polishing of copper overlay," *IEEE Transactions on Semiconductor Manufacturing*, (**17)**2, pp. 180-186, May 2004 2004.

[128] S. Park, S. Joo, Y. Kim, H. Jeong, and H. Kim, "Development of ae monitoring system for CMP process," in *IEEE International Conference on Planarization / CMP Technology (ICPT)*, Dresden, Germany, 2007, pp. 1-5.

[129] A. Sikder, F. Giglio, J. Wood, A. Kumar, and M. Anthony, "Optimization of tribological properties of silicon dioxide during the chemical mechanical planarization process," *Journal of Electronic Materials*, (**30)**12, pp. 1520-1526, 2001.

[130] M. Birang and A. Gleason, "Apparatus and method for *in-situ* endpoint detection for chemical mechanical polishing operations," United States Patent US 6676717 B1, Dated: Jan. 13, 2004.

[131] F. Mauersberger, P. J. Beckage, P. R. Besser, F. N. Hause, E. T. Ryan, W. S. Brennan, and J. A. Iacoponi, "Method and apparatus for detecting the endpoint of a chemical-mechanical polishing operation using optical techniques," United States Patent US 6,809,032 B1, Dated: Oct. 26, 2004.

[132] D. Zeidler, M. Plotner, and K. Drescher, "Endpoint detection method for CMP of copper," *Microelectronic Engineering,* (**50)**1-4, pp. 411-416, 2000.

[133] M. Meloni, "System and method for detecting CMP endpoint via direct chemical monitoring of reactions," United States Patent 6,287,171, Dated: Sep. 11, 2001.

[134] Y. L. Wang, C. Liu, M. S. Feng, and W. T. Tseng, "The exothermic reaction and temperature measurement for tungsten CMP technology and its application on endpoint detection," *Materials Chemistry and Physics,* (**52)**1, pp. 17-22, 1998.

[135] K. E. Mautz, "Method and apparatus for monitoring a polishing condition of a surface of a wafer in a polishing process," United States Patent 6,709,312 B2, Dated: Mar. 23, 2004.

[136] D. White, J. Melvin, and D. Boning, "Characterization and modeling of dynamic thermal behavior in CMP," *Journal of The Electrochemical Society,* (**150)**4, pp. G271-G278, 2003.

[137] E. Gupta, "Real-time estimation of material removal rate (MRR) in copper chemical mechanical planarization (CMP) using wireless temperature sensor," M.S. Thesis, Mechanical Engineering, Oklahoma State University, Stillwater, 2010.

[138] J. Luo and D. A. Dornfeld, "Material removal mechanism in chemical mechanical polishing :Theory and modeling," *IEEE Transactions on Semiconductor Manufacturing,* (**14)**2, pp. 112 -133, May, 2001 2001.

[139] C. Wang, P. Sherman, A. Chandra, and D. Dornfeld, "Pad surface roughness and slurry particle size distribution effects on material removal rate (MRR) in chemical mechanical planarization (CMP)," *CIRP Annals - Manufacturing Technology,* (**54)**1, pp. 309-312, 2005.

[140] C. Wang, P. Sherman, and A. Chandra, "A stochastic model for the effects of pad surface topography evolution on material removal rate decay in chemical-

mechanical planarization," *IEEE Transactions on Semiconductor Manufacturing,* (**18)**4, pp. 695-708, 2005.

[141] J.-Y. Lai, "Mechanics, mechanisms and modeling of the chemical mechanical polishing process," PhD Dissertation, Mechanical Engineering, MIT, Boston, 2001.

[142] H. Liang and G. Helen Xu, "Lubricating behavior in chemical–mechanical polishing of copper," *Scripta Materialia,* (**46)**5, pp. 343-347, 2002.

[143] H. Liang and D. Craven, *Tribology in chemical-mechanical planarization*. Boca Raton , FL: CRC Press, Taylor and Francis Group, 2005.

[144] J. Luo, "Integrated modeling of chemical mechanical planarization/polishing for integrated circuit fabrication: From particle scale to die and wafer scales," PhD Dissertation, Mechanical Engineering, University of California, Berkeley, 2003.

[145] A. Bastawros, A. Chandra, Y. Guo, and B. Yan, "Pad effects on material removal rate in chemical mechanical polishing," *Journal of Electronic Materials,* (**31)**10, pp. 1-10, 2002.

[146] G. Fu and A. Chandra, "A model for wafer scale variation of material removal rate in chemical mechanical polishing based on viscoelastic pad deformation," *Journal of Electronic Materials,* (**31)**10, pp. 1066 - 1073, 2002.

[147] A. Chandra, P. Karra, A. Bastawros, R. Biswas, P. Sherman, S. Armini, and D. Lucca, "Prediction of scratch generation in chemical mechanical planarization," *CIRP Annals - Manufacturing Technology,* (**57)**1, pp. 559-562, 2008.

[148] J. Luo and D. A. Dornfeld, "Effects of abrasive size distribution in chemical mechanical planarization: Modeling and verification," *IEEE Transactions on Semiconductor Manufacturing,* (**16)**3, pp. 469-476, 2003.

[149] L. Borucki, "Mathematical modeling of polishing rate decay in chemical mechanical polishing," *Journal of Engineering Mathematics,* (**43)**pp. 105-114, 2002.

[150] K. Qin, B. Moudgil, and C.-W. Park, "A chemical mechanical polishing model incorporating both the chemical and mechanical effects," *Thin Solid Films,* (**446)**2, pp. 277-286, 2004.

[151] D. G. Thakurta, C. L. Borst, D. W. Schwendeman, R. J. Gutmann, and W. N. Gill, "Pad porosity, compressibility and slurry delivery effects in chemical-

mechanical planarization: Modeling and experiments," *Thin Solid Films,* (**366)**1–2, pp. 181-190, 2000.

[152] C. F. Higgs III, S. H. Ng, L. Borucki, I. Yoon, and S. Danyluk, "A mixed-lubrication approach to predicting CMP fluid pressure modeling and experiments," *Journal of The Electrochemical Society,* (**152)**3, pp. G193-G198, 2005.

[153] R. Bajaj, M. Desai, R. Jairath, M. Stell, and R. Tolles, "Effect of polishing pad material properties on chemical mechanical polishing processes," in *Material Research Society Spring Meeting*, 1994, pp. 637-644.

[154] H. Lu, Y. Obeng, and K. A. Richardson, "Applicability of dynamic mechanical analysis for CMP polyurethane pad studies," *Materials Characterization,* (**49)**2, pp. 177-186, 2002.

[155] Z. Stavreva, D. Zeidler, M. Plotner, and K. Drescher, "Characteristics in chemical-mechanical polishing of copper: Comparison of polishing pads," *Applied Surface Science,* (**108)**1, pp. 39-44, 1997.

[156] H. Wang, X. Zhang, A. Kumar, and Q. Huang, "Nonlinear dynamics modeling of correlated functional process variables for condition monitoring in chemical–mechanical planarization," *IEEE Transactions in Semiconductor Manufacturing,* (**22)**1, 2009.

[157] R. Allen, C. Chen, T. Trikas, K. Lehman, R. Shinagawa, V. Bhaskaran, B. Stephenson, and D. Watts, "*In-situ* CMP copper endpoint control system," in *IEEE International Semiconductor Manufacturing Symposium,* 2001, pp. 391-394.

[158] J. Greenwood and J. Williamson, "Contact of nominally flat surfaces," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences,* (**295)**1442, p. 300, 1966.

[159] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports,* (**438)**5-6, pp. 237-329, 2007.

[160] R. Komanduri, P. Rao, P. Vinod, O. Beyca, S. Jagannath, A. Fields, Z. Kong, and S. Bukkapatnam. (2012, On the identification of physical sources of vibration sensor signal patterns in chemical mechanical planarization process. Available: http://www.okstate.edu/commsens/Report_8.pdf

[161] S. T. S. Bukkapatnam and C. Cheng, "Forecasting the evolution of nonlinear and nonstationary systems using recurrence-based local gaussian process models," *Physical Review E,* (**82)**5, p. 056206, 2010.

[162] A. Ahmed and E. P. Xing, *Tutorial: Dynamic non-parametric mixture models and the recurrent chinese restaurant process*: Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2007.

[163] S. T. S. Bukkapatnam and R. Palanna, "Experimental characterization of nonlinear dynamics underlying the cylindrical grinding process," *Transactions of the ASME, Journal of Manufacturing Science and Engineering,* (**126)**2, pp. 341-344, 2004.

[164] A. H. Nayfeh and B. Balachandran, *Applied nonlinear dynamics: Analytical, computational, and experimental methods*. Weinheim, Germany: Wiley-VCH Verlag, 2007.

[165] G. Byrne, B. Mullany, and P. Young, "The effect of pad wear on the chemical mechanical polishing of silicon wafers," *CIRP Annals - Manufacturing Technology,* (**48)**1, pp. 143-146, 1999.

[166] H. Lu, B. Fookes, Y. Obeng, S. Machinski, and K. A. Richardson, "Quantitative analysis of physical and chemical changes in CMP polyurethane pad surfaces," *Materials Characterization,* (**49)**1, pp. 35-44, 2002.

[167] S. Bukkapatnam and B. Clark, "Dynamic modeling and monitoring of contour crafting—an extrusion-based layered manufacturing process," *Transactions of the ASME, Journal of Manufacturing Science and Engineering,* (**129)**p. 135, 2007.

[168] International technology roadmap for seminconductors www.Itrs.Net [Online].

[169] I. Sherrington and E. H. Smith, "Modern measurement techniques in surface metrology: Part 2 optical instruments," *Wear,* (**125)**3, pp. 289-308, 8/1/ 1988.

[170] S. Stokowski and M. Vaez-Iravani, "Wafer inspection technology challenges for ulsi manufacturing," in *American Institute of Physics Conference*, 1998, p. 405.

[171] D. J. Whitehouse, "The parameter rash — is there a cure?," *Wear,* (**83)**1, pp. 75-78, 1982.

[172] J. C. Russ, "Fractal dimension measurement of engineering surfaces," *International Journal of Machine Tools and Manufacture,* (**38)**5–6, pp. 567-571, 1998.

[173] S. Ganti and B. Bhushan, "Generalized fractal analysis and its applications to engineering surfaces," *Wear,* (**180)**1–2, pp. 17-34, 1995.

[174] P. F. Chauvy, C. Madore, and D. Landolt, "Variable length scale analysis of surface topography: Characterization of titanium surfaces for biomedical applications," *Surface and Coatings Technology,* (**110)**1–2, pp. 48-56, 1998.

[175] A. Majumdar and B. Bhushan, "Role of fractal geometry in roughness characterization and contact mechanics of surfaces," *Journal of Tribology,* (**112)**p. 205, 1990.

[176] D. J. Whitehouse, "Fractal or fiction," *Wear,* (**249)**5–6, pp. 345-353, 2001.

[177] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 6, pp. 679-698, 1986.

[178] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Mathematical Journal,* (**23)**pp. 298-305, 1973.

[179] F. R. K. Chung, *Spectral graph theory*. Providence, RI: American Mathematical Society, 1997.

[180] S. Butler, "Eigen values and structures of graphs," PhD Dissertation, Department of Mathematics, University of California, San Diego, 2008.

[181] B. Mohar, "The laplacian spectrum of graphs," *Graph Theory, Combinatorics, and Applications,* (**2)**pp. 871-898, 1991.

[182] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing,* (**17)**4, pp. 395-416, 2007.

[183] M. Fiedler, "A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory," *Czechoslovak Mathematical Journal,* (**25)**100, pp. 619-633, 1975.

[184] W. H. Haemers, "Interlacing eigenvalues and graphs," *Linear Algebra and its Applications,* (**226)**pp. 593-616, 1995.

[185] G. Chen, G. Davis, F. Hall, Z. Li, K. Patel, and M. Stewart, "An interlacing result on normalized laplacians," *SIAM Journal on Discrete Mathematics*, (**18)**2, pp. 353-361, 2004.

[186] D. A. Spielman, "Spectral graph theory," *Lecture Notes, Yale University*, 2009.

[187] D. A. Spielman, "Algorithms, graph theory, and linear equations in laplacian matrices," in *Proceedings of the International Congress of Mathematicians*, 2010, pp. 2698-2722.

[188] M. W. Newman, "The laplacian spectrum of graphs," MS Thesis, Department of Mathematics, University of Manitoba, Winnipeg, Canada, 2000.

[189] K. M. Hall, "An r-dimensional quadratic placement algorithm," *Management Science*, (**17)**3, pp. 219-229, 1970.

[190] C. Meyer, *Matrix analysis and applied linear algebra book and solutions manual* vol. 2. Philadelphia, PA: SIAM, 2000.

[191] D. A. Spielman, "Spectral graph theory and its applications," in *48th Annual IEEE Symposium on Foundations of Computer Science*, 2007, pp. 29-38.

[192] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (**22)**8, pp. 888-905, 2000.

[193] H. Chen and F. Zhang, "Resistance distance and the normalized laplacian spectrum," *Discrete applied mathematics*, (**155)**5, pp. 654-661, 2007.

[194] *ASM metals handbook: Metallography and microstructures* vol. 9. Metals Park, OH: American Society of Metals, 2000.

[195] B. E. Dom, V. H. Brecher, R. Bonner, J. S. Batchelder, and R. S. Jaffe, "The p300: A system for automatic patterned wafer inspection," *Machine vision and applications*, (**1)**4, pp. 205-221, 1988.

[196] B. H. Khalaj, H. K. Aghajan, and T. Kailath, "Patterned wafer inspection by high resolution spectral estimation techniques," *Machine vision and applications*, (**7)**3, pp. 178-185, 1994.

[197] G. Bultman, A. Levy, K. A. Brown, M. Nikoonahad, D. Wack, and J. Fielden, "Methods and systems for determining a presence of macro and micro defects on a specimen," United States Patent 7460981, Dated: Dec 2, 2008.

[198] C. R. Fairley, T. Y. Fu, G. Perelman, and B. M. B. Tsai, "High throughput brightfield/darkfield wafer inspection system using advanced optical techniques," United States Patent 6,816,249, Dated: Nov. 9, 2004.

[199] K. W. Tobin Jr, S. S. Gleason, T. P. Karnowski, and H. Sari-Sarraf, "Automated defect spatial signature analysis for semiconductor manufacturing process," United States Patent 5,982,920, Dated: Nov. 9, 1999.

[200] N. G. Shankar and Z. W. Zhong, "Defect detection on semiconductor wafer surfaces," *Microelectronic Engineering,* (**77)**3–4, pp. 337-346, 2005.

[201] P. D. Kinney and N. P. Rao, "Optical inspection module and method for detecting particles and defects on substrates in integrated process tools," United States Patent 5,909,276.

[202] M. J. Davidson, "Defect inspection efficiency improvement with in-situ statistical analysis of defect data during inspection," United States Patent 6,635,872 B2.

[203] A. Hamamatsu, H. Shibuya, Y. Oshima, S. Maeda, H. Nishiyama, and M. Noguchi, "Statistical threshold method for semiconductor inspection," in *Asia-Pacific Conference on Non-Destructive Testing*, Auckland, New Zealand, 2006.

[204] F. L. Chen and S. F. Liu, "A neural-network approach to recognize defect spatial pattern in semiconductor fabrication," *IEEE Transactions on Semiconductor Manufacturing,* (**13)**3, pp. 366-373, 2000.

[205] S. Liu, F. Chen, and W. Lu, "Wafer bin map recognition using a neural network approach," *International journal of production research,* (**40)**10, pp. 2207-2223, 2002.

[206] S. Nichani and J. Scola, "Semiconductor device image inspection utilizing image subtraction and threshold imaging," United States Patent 5,949,901, Dated: Sep 7, 1999.

[207] P. B. Chou, A. R. Rao, M. C. Sturzenbecker, F. Y. Wu, and V. H. Brecher, "Automatic defect classification for semiconductor manufacturing," *Machine vision and applications,* (**9)**4, pp. 201-214, 1997.

[208] S. H. Hwang and T. Y. Fu, "Method and apparatus using interferometric metrology for high aspect ratio inspection," United States Patent 7,061,625, Dated: Jun 13, 2006.

[209] I. Moriyama, Y. Tanabe, I. Ishimaru, M. Noguchi, and K. Mizukami, "Surface inspection system with automatic discrimination of microscratches and particles on chemical mechanical polished wafers," in *IEEE International Symposium on Semiconductor Manufacturing* Santa Clara, CA, 1999, pp. 139-142.

[210] U. Ganesha, B. K. A. Ngoi, H. C. F. Goh, and M. N. Yusoff, "Defect detection in unpolished si wafers by digital shearography," *Measurement Science and Technology,* (**15)**1, p. 35, 2004.

[211] C. F. Bevis, M. Kirk, and M. Vaez-Iravani, "Systems for inspection of patterned or unpatterned wafers and other specimen," United States Patent 7,068,363, Dated: Jun 27, 2006.

[212] S. Biellak, S. E. Stokowski, and M. Vaez-Iravani, "Systems and methods for a wafer inspection system using multiple angles and multiple wavelength illumination," United States Patent 6,956,644, Dated: Oct 18, 2005.

[213] H. Shinada, Y. Yajima, H. Murakoshi, M. Hasegawa, M. Nozoe, A. Takafuji, K. Sugiyama, K. Kuroda, K. Umemura, and Y. Usami, "Patterned wafer inspection method and apparatus therefor," United States Patent 7,242,015, Dated: Jul 10, 2007.

[214] A. Ache and K. Wu, "Production implementation of state-of-the-art electron beam inspection," in *IEEE/SEMI Advanced Semiconductor Conference*, 2004, pp. 344-347.

[215] F. J. Anscombe, "The transformation of poisson, binomial and negative-binomial data," *Biometrika,* (**35)**3/4, pp. 246-254, 1948.

[216] M. H. Quenouille, "A relation between the logarithmic, poisson, and negative binomial series," *Biometrics,* (**5)**2, pp. 162-164, 1949.

[217] S. Huang, Z. J. Kong, and W. Huang, "High-dimensional process monitoring and change point detection using embedding distributions in reproducing kernel hilbert space (rkhs)," *Under Review (Manuscript Submitted),* 2013.

[218] B. Mohar, "Isoperimetric numbers of graphs," *Journal of Combinatorial Theory, Series B,* (**47)**3, pp. 274-291, 1989.

[219] B. Mohar, "Eigenvalues, diameter, and mean distance in graphs," *Graphs and combinatorics,* (**7)**1, pp. 53-64, 1991.

[220] C. Cheung, K. Chan, and W. Lee, "Surface characterization in ultra-precision machining of al/sic metal matrix composites using data dependent systems analysis," *Journal of Materials Processing Technology,* (**140)**1-3, pp. 141-146, 2003.

[221] D. J. Whitehouse, *Handbook of surface and nanometrology*. Boca Raton, FL: CRC press, 2011.

[222] D. Whitehouse, "Metrology at nanometre level," in *Proceedings of the First national conference in precision engineering, Chennai, India*, 2000, pp. 73-98.

[223] S. Beeby, G. Ensell, M. Kraft, and N. White, *MEMS mechanical sensors*. Boston, MA: Artech House, Inc, 2004.

[224] W. Gopel, J. Hesse, and J. Zemel, *Sensors: A comprehensive survey* vol. 7 (Mechanical Sensors): Wiley-VCH, Wienheim, 1994.

[225] J. Fraden, *Handbook of modern sensors: Physics, designs, and applications*: Springer Verlag, 2004.

[226] Y. Akiyama and O. Sugiura, "A study of 2sf beat phenomena in induction motors," in *IEEE Industry Applications Society Annual Meeting*, Houston, TX, 1992, pp. 100-106 vol.1.

[227] Y. Akiyama, "A study of 2sf beat phenomena in induction motors-found out of akiyama's wave," *Research Reports of Kanagawa B-17,* pp. 151-161, 1993.

[228] E. W. Summers, "Vibration in 2-pole induction motors related to slip frequency," *Transactions of the American Institute of Electrical Engineers, Part 3: Power Apparatus and Systems.,* (**74)**3, pp. 69-72, 1955.

[229] A. H. R. Streatfield, "Development of vibration monitoring techniques to aid in evaluating the mechanical performance of rotating machinery," *Proceedings of the Institution of Mechanical Engineers,* (**184)**318, pp. 209-217, 1969.

[230] R. Ueda, T. Sonoda, and S. Takata, "Experimental results and their simplified analysis on instability problems in pwm inverter induction motor drives," *IEEE Transactions on Industry Applications,* (**25)**1, pp. 86-95, 1989.

[231] Y. R. Jeng, P. Y. Huang, and W. C. Pan, "Tribological analysis of chemical mechanical planarization with partial asperity contact," *Journal of The Electrochemical Society,* (**150)**p. G630, 2003.

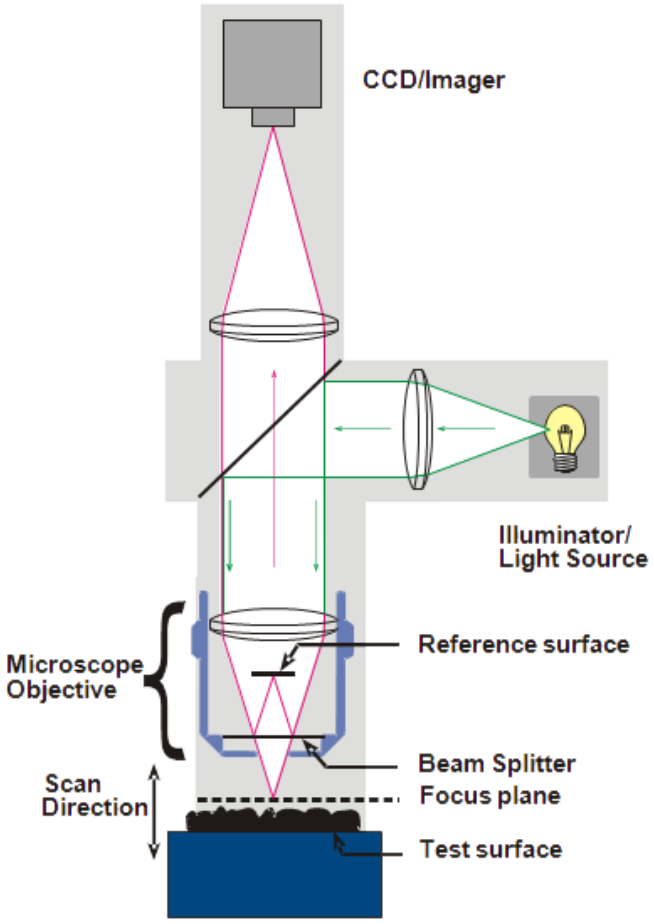
[232] N. Hoffmann and L. Gaul, "Effects of damping on mode-coupling instability in friction induced oscillations," *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik,* (**83)**8, pp. 524-534, 2003.

[233] Z. B. Hou and R. Komanduri, "On the mechanics of the grinding process – part i. Stochastic nature of the grinding process," *International Journal of Machine Tools and Manufacture,* (**43)**15, pp. 1579-1593, 2003.

Appendix I: Operating Principle of the MicroXAM Optical Interference Surface Profiling Microscope

The MicroXAM interferometer is primarily used in our studies for obtaining the surface profile of specular finished UPM and CMP surfaces in a non-contact manner. The operating principle of the instrument is based on profiling the surface topography of a specimen by analyzing optical interference patterns. Referring to Fig. 1, a collimated beam of light from an illuminator is split (using a beam splitter), one part of the beam is focused on the test specimen, and the second on a reference specimen [5, 6, 9, 56, 169, 221, 222]. On recombination of the two beams, an interference pattern is obtained. This interference pattern is analyzed (after imaging with a CCD camera) using proprietary computer software (called MapVue) to reconstruct the peaks and valleys on the test specimen.

In the particular instrument model used in our study, two possible optical modes of operation based on the wavelength of light (white and green light) are possible. Either mode is activated from the software environment. In our studies, we primarily use the green light (547.8 nm wavelength) mode. In addition, the instrument is equipped with two objective lens magnifications (Fig. 1), we use the higher 50x magnification for our measurements.



Objective Magnification	50x	10x
Numerical aperture	0.55	0.30
Measurement area (µm)	165 × 125	827 × 626
Spatial sampling (µm)	0.22 × 0.26	1.1 × 1.3
Optical resolution (µm, @ 550 nm)	0.50	0.92
Working distance (mm)	3.4	7.4
Depth of focus (µm, @550 nm)	1.16	3.89
Maximum surface slope	22.6°	4.8°

Fig. 1: Schematic diagram of the MicroXAM surface profiling interferometer (image courtesy KLA-Tencor).

Appendix II: Identification of Physical Sources of Vibration Sensor Signal Patterns in Chemical Mechanical Planarization (CMP) Process

We detail the battery of tests conducted to ascertain the physical sources of dominant vibration sensor signal patterns observed in chemical mechanical planarization (CMP) process. A Buehler (model Automet® 250) bench top CMP machine is instrumented with miniature MEMS 3-axis accelerometer (Analog Devices ADXL 335) and audio sensors (Analog Devices ADMP 401). The CMP setup is used for finishing blanket copper workpieces to a surface finish of Ra ~ 5 nm. While the sensor signals are sensitive to variations in the CMP process, the extraneous noise prevents the direct use of raw signal patterns for early detection of defects. Consequently, instead of primarily monitoring the raw sensor signal patters, we isolate signal features which are indicative of process state from those which are mere artifacts, and thus potentially valuable for process monitoring.

The frequency spectrum of typical MEMS vibration sensor signals acquired during the CMP process contains three dominant components, namely:

1. Component 1: A low frequency component in the 0.5 – 1 Hz region;
2. Component 2: Broadband frequency regions centered around 25 Hz and 50 Hz;
3. Component 3: A broadband frequency region at approximately 120 Hz.

Nine empirical tests are reported to ascertain the underlying physical cause of each of these components. The inferences are summarized herewith:

- Component 1 is shown to result from eccentricity errors in the polishing head (spindle);
- Component 2 is most likely a conjoined effect due to sensor characteristics, electromagnetic interference from machine elements, and structural vibration;
- Component 3 is observed to respond to changing downforce (polishing load) conditions, and process state, such as pad wear, and as such is useful for process monitoring applications.
- In addition, a −110 dB background (white) noise is evident throughout the frequency spectrum of CMP vibration signals; this is explained as originating from measurement errors and environmental factors.

Introduction

Chemical mechanical planarization (CMP) is considered a vital component of back-end-of-line (BEOL) processes in semiconductor manufacturing as it is the process of choice for obtaining both local and global planarity on a variety of semiconductor materials [19, 21]. Since CMP is often the final step before wafer test and packaging stages, wafer anomalies and defects resulting from CMP can lead to irrecoverable waste of labor, material resources, and potential revenue. It

has been estimated that significant damage to a 300 mm wafer may cause ~ $100,000 in losses [22]. Under such circumstances, conventional statistical process control (SPC) techniques are of limited utility. SPC techniques are typically incapable of distinguishing subtle inherent process drifts, such as gradual pad wear, slurry coagulation, slurry contamination, etc., from process noise [22, 105-107]. It is therefore desirable to ensure defect-free operation in CMP by employing real-time *in situ* sensor-based process monitoring approaches [22].

Toward this end, *in situ* process monitoring techniques have been tailored specifically for CMP applications. Current CMP monitoring techniques are predominantly based on piezoelectric (force, vibration, and friction sensors) [112-115], acoustic emission (AE) [53, 125, 126], laser [130, 131], electro-chemical [132], and thermo-optical [134-136] sensing elements. These sensing approaches are mainly geared towards detection of polishing endpoint and control of process inputs, such as downforce, slurry flow rate, pH, pad wear, etc. Though, they have been investigated as possible avenues for *in situ* defect detection during CMP, factors relating to cost and maintainability have stymied their large scale adaptability [108]. Apart from cost, these sensing systems require careful attention to calibration and location, e.g., *in situ* optical sensing systems for CMP wafer endpoint detection typically need specially designed polishing pads with optical filter windows [131, 135]. Also, the added bulk of these sensing systems constrains the scant real estate available on the equipment, and may superimpose undesirable

vibrations on the machine. Lastly, due to their high power consumption, these sensing systems are not amicable for non-intrusive wireless applications.

In our work, MEMS sensors akin to those embedded in smart phones and video game systems are used for CMP process monitoring [223]. Apart from being inexpensive, MEMS sensors: (i) are powered by low capacity sources, such as lithium polymer batteries, (ii) can be easily replaced, (iii) preclude amplification and expensive DAQ systems required in piezoelectric based systems, and (iv) can lend towards close proximity monitoring of the process owing to their miniature size. These factors combine to make MEMS sensors an attractive choice for wireless *in situ* monitoring applications. However, MEMS sensing systems present a new set of challenges:

i. **Low sampling rates**: In contrast to piezoelectric sensors, MEMS sensors are typically powered by 3.3 – 5V DC sources. This relatively low power input does not lend for sustained high sampling rates. Typical sampling rates of MEMS sensors are below 1 kHz, compared to over 50 kHz possible with a piezoelectric system.

ii. **Susceptibility to external noise**: Since on-board filters and DAQ cards are not used, the ability to mitigate extraneous noise sources is limited.

iii. **Impeded applicability for monitoring manufacturing processes**: Piezoelectric sensors for manufacturing applications are crafted with extreme precision [224, 225]. For example., a piezoelectric crystal is

cut and ground in particular crystal orientations that minimizes noise and cross talk between sensing axes. In contrast, MEMS sensors are largely designed for use in consumer electronics and automotive applications that do not demand such precision. Hence, significant variability may exist in response characteristics between sensors.

Before the developed sensor array can be used for defect detection purposes, it is imperative to study the nature of signals acquired from the process in order to understand the physical source of particular sensor signal characteristics. By doing so, we can begin to isolate those signal features which are indicative of process state and thus potentially valuable for process monitoring, from those signal patterns which may be mere artifacts. The results from the tests conducted to isolate such dominant signal features are summarized herewith.

Experimental Setup

CMP machine

Fig. 2 shows a Buehler (model Automet® 250) metallographic polishing apparatus instrumented with MEMS vibration and sound sensors. Round copper workpieces of diameter 40.6250 mm ± 0.1 mm (1.625 in.), and thickness ~ 12.5 mm ± 2 mm (0.5 in.) are polished on this apparatus. In order to accommodate machining operations for mounting the sensing system, C14500 series copper is used for polishing experiments. This particular copper series is 99.5% pure with tellurium (Te) as an alloying element. Te improves the machinability rating of

copper [62], but limits the surface finish that can be achieved. Defect free, mirror-like finish with Ra ~ 5 nm is reported in this work (see Fig. 3). The polishing machine shown in Fig. 2 comprises a polishing platen whose rotational velocity can be adjusted between 10 – 300 RPM in both clockwise and counterclockwise directions. A workpiece is located on the steel workpiece holder (workholder), and the spindle lowered on the platen. The spindle senses contact with the platen and adjusts to a set distance of ~ 2.5 mm (0.1 in.) above the platen. The spindle can be set to rotational speeds between 30 – 60 RPM (0.5 – 1 Hz) in intervals of 10 RPM, and allows motion only in one (clockwise) direction.

Downforce (polishing load) is applied by a pneumatic system. Compressed air is supplied to actuators, which impress upon the workpiece to the set load. The downforce can be adjusted between ~ 2 – 10 lb. (~ 1 – 4.5 kg) in integer steps. The applied downforce is continuously monitored by a microprocessor-based system, which adjusts the air pressure. In case of deficit from the set load, a system of solenoids let additional air into the pneumatic cylinder in which the actuators move freely. Valves are activated by a secondary solenoid to relieve air if an excess downforce is detected.

Polishing slurry is supplied by means of a peristaltic pump powered by a regulated DC source. A maximum slurry flow rate of 100 ml/min can be supplied. The spindle location can be adjusted by means of a knob to various offset positions with respect to the platen. This adjustment provides for an eccentric polishing action. Ideally, the offset position is set such that the workpiece is well within

the polishing pad circumference at its outer most point, and approximately the center of the workpiece coincides with the platen center at the innermost position.

The polishing pad is affixed by means of a pressure sensitive adhesive on a metallic plate having a magnetic backing. The plate with the polishing pad is held by magnetic means on the polishing platen. Stability of the polishing platen is maintained by a vacuum system, which couples the platen to a motor in the base of the machine.

Fig. 2: Buehler Automet® 250 experimental CMP polishing setup showing the sensor arrangement and various machine elements.

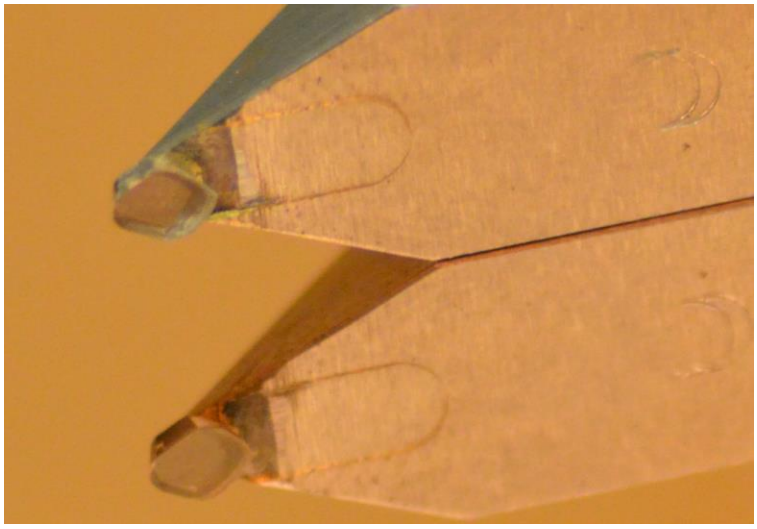


Fig. 3: Surface of copper workpiece obtained after CMP. Photograph of a single crystal diamond cutting tool (top) and its reflection on the polished surface (bottom).

Sensing system

The sensing system comprises of MEMS vibration and sound sensors (one each). The vibration sensor is a model ADXL 335 tri-axial accelerometer manufactured by Analog Devices (see Fig. 4). This vibration sensor is similar to the accelerometer used in the Nintendo Wii® video game controller (ADXL 330). The sound sensor is a model ADMP 401 MEMS microphone also manufactured by Analog Devices. These sensors are mounted on breakout boards supplied by SparkFun® Electronics. In this work, data gathered from the accelerometer is presented. The accelerometer is capable of measuring vibration between ± 3g, and has a maximum sampling rate of 1600 Hz for each axis at 3.3V DC power input. Data transmission takes place wirelessly. The sensing arrangement for gathering vibration data is shown in Fig. 4.

The sensors convey data to an XBee® radio transmitter (IEEE 802.15.4 Protocol RF module) supplied by Digi International. The transmitter wirelessly

communicates with a coupled receiver. The receiver is powered by USB attachment that connects to a laptop. The transmitter is fixed on a pinout board (also supplied by SparkFun Electronics). A 3.3V DC lithium polymer battery powers the (transmitter) pinout board. The power input is regulated to minimize voltage fluctuations and subsequently directed to both the transmitter and sensor. Power regulation is critical for amicable performance of the sensor-transmitter-receiver schema, since excessive variations in input potential can cause spurious fluctuations in sensor data.

Two transmitters are fixed by adhesive in a plastic enclosure, one each for the vibration and audio sensor. The power source for the transmitters are located diametrically opposite in a similar plastic enclosure (Fig. 2). These enclosures have aluminum covers to reduce electromagnetic interference. The transmitter and battery enclosures are fixed onto aluminum rests by means of industrial strength Velcro® connections. One face of these aluminum rests are shaped in a circular manner and bolted onto a shaft collar fastened to the spindle. The total weight of the transmitter pair is ~ 50 gm., which is equivalent with the battery weight. The sensors are located in a specially machined aluminum container to minimize electromagnetic noise. The sensor enclosure is sealed with an aluminum cover.

Since soldered connections are easily broken, Molex® connections are used to allow the slight twisting motion of the workpiece during polishing. The complete sensing system is fabricated to maintain equitable distribution of weight, because,

weight discrepancies on these attachments can cause the spindle to wobble – akin to an unbalanced shaft.

In this work we use the data gathered from the vibration sensors in tangential (V_X) and radial (V_Y) directions (in the X-Y plane with respect to the rotating spindle) because the applied downforce constrains the movement of the work-piece in the vertical direction. Therefore, the observed magnitude of the vibration sensor response in the vertical direction (V_Z) was diminished.

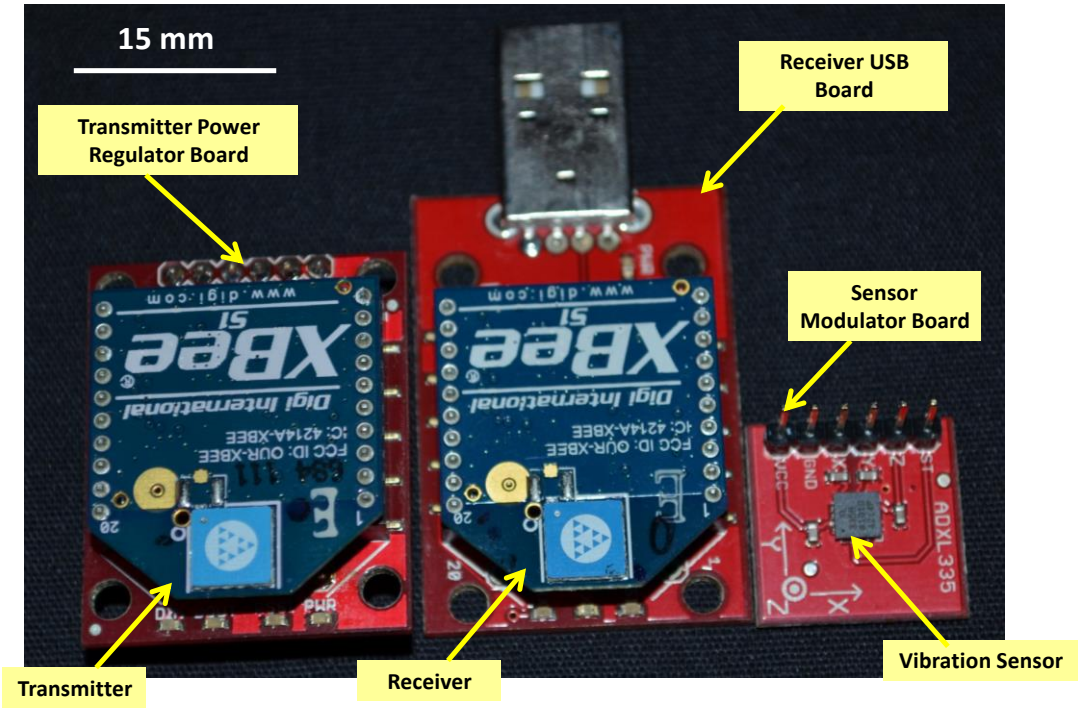


Fig. 4: Components of wireless sensor array mounted on the CMP setup for gathering vibration signals.

Experimental Procedure

MEMS sensor data is often contaminated by a high level of extraneous noise relating to; (i) sensor response characteristics, (ii) sensor placement, (iii) dynamic instability and geometrical inaccuracy of mechanical components, (iv) interfer-

ence from electromagnetic sources, such as motors and pumps, and (v) other significant environmental conditions, such as temperature and acoustic noise [225].

Therefore, in order to understand the process dynamics, it is necessary to identify the sources of prominent features observed from sensor signal data. In addition, by identifying those components of the sensor data that may be mere artifacts signal processing techniques focusing on noise reduction and multi-scale resolution can be applied with less likelihood of losing valuable information relating to the process.

The task of identifying sensor data patterns is oftentimes more conducive in the frequency domain as compared to the time domain. For example, Fig. 5 (a) shows a typical vibration time series data gathered during CMP polishing. From the time series, a prominent low frequency component with time period of ~ 2 seconds (0.5 Hz) is observed. The frequency spectrum (see Fig. 5 (b)), Fast Fourier Transform (FFT), of the signal reveals the following characteristics:

1. Low frequency component in the 0.5 – 1 Hz region. The 1 Hz component can be surmised as an integer multiple of the peak at 0.5 Hz.
2. Broadband frequency regions centered around 25 Hz and 50 Hz.
3. Broadband frequency region around 120 Hz.
4. A relatively small background (white) noise evident throughout the spectrum.

Given these signal characteristics, our objectives are as follows:

1. Isolate the cause of the foregoing signal characteristics.
2. Delineate characteristics that are pertinent to the process state from those which maybe artifacts.

In order to achieve these objectives, we conduct the various tests described in the forthcoming sections.

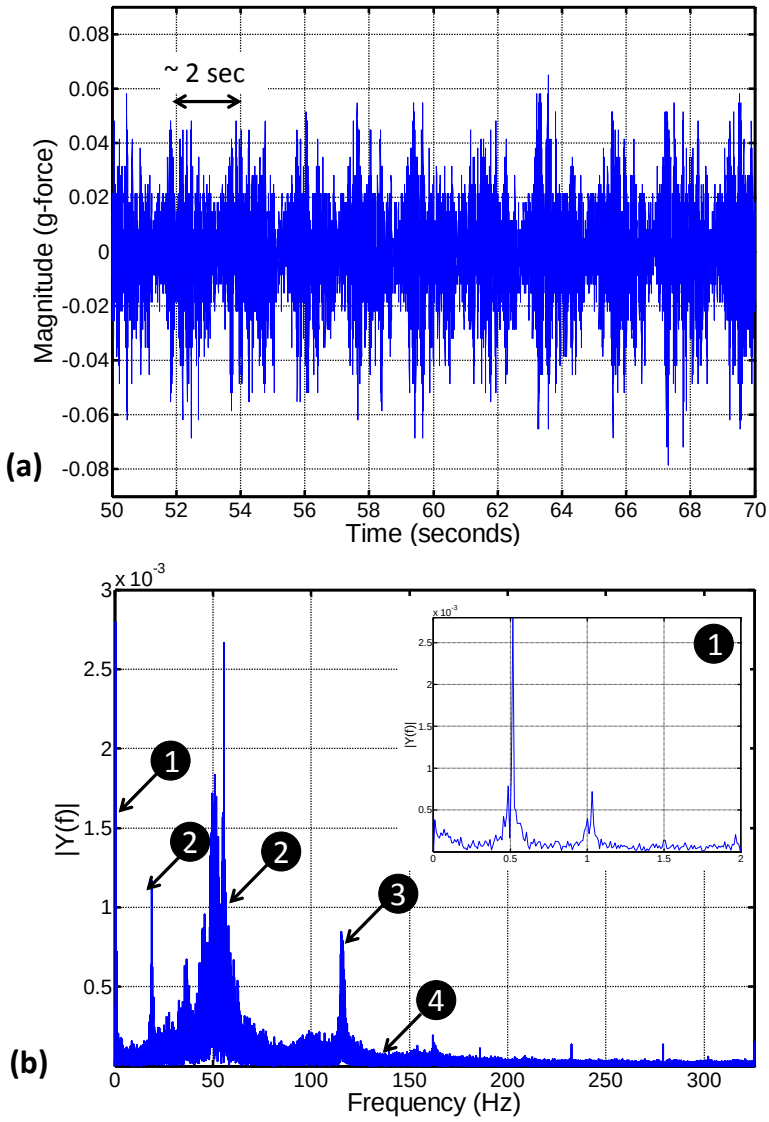


Fig. 5: Typical vibration sensor data in the tangential (V_X) direction obtained during CMP of copper workpieces under the following condition; 30 RPM spindle speed, 150 RPM platen speed, and 2 lb.

down force. (a): Time series plot showing ~ 20 seconds of the total (180 seconds) data, time period between prominent low frequency components ~ 2 seconds (b): FFT frequency spectrum of the data, showing presence of 4 main components, e.g., low frequency component (0.5 Hz – 1 Hz) seen in (a) is marked by 1. A zoomed in view of component 1 is shown in the inset.

Ambient response of sensor

The vibration sensor was mounted on a pneumatically stabilized vibration isolation table, and data collected at ~ 690 Hz sampling rate. The sensor data reveals the presence of a prominent broadband region around the 25 Hz and 50 Hz, and integer multiples thereafter (Fig. 6). This broadband region was ~ 4 – 5 times the white noise characteristic (component 4) of the sensor. The white noise was estimated at 5×10^{-5} units in the frequency spectrum for sensor data calibrated in g-force units (Fig. 6 (a)). In decibels, the base noise translates to ~ −110 dB mean, with the 25 Hz region peaking at ~ −80 dB (Fig. 6 (b)). Thus, the region marked by 2 in Fig. 5 (b), can be attributed to some extent as having originated from the sensor characteristics. Presence of resistors and capacitors on the sensor board can lead to such typical low frequency noise component termed *flicker* noise (see e.g., [225]). However, as comparison with Fig. 6 (b) shows, component 2 does not completely account for the energy observed during polishing. Thus additional contribution of the spindle motor (60V permanent magnet DC gear motor, modulated from AC mains), cannot be completely ruled out [225]. Interestingly, the white noise spectrum (component 4) seems to be agreement with Fig. 5 (b).

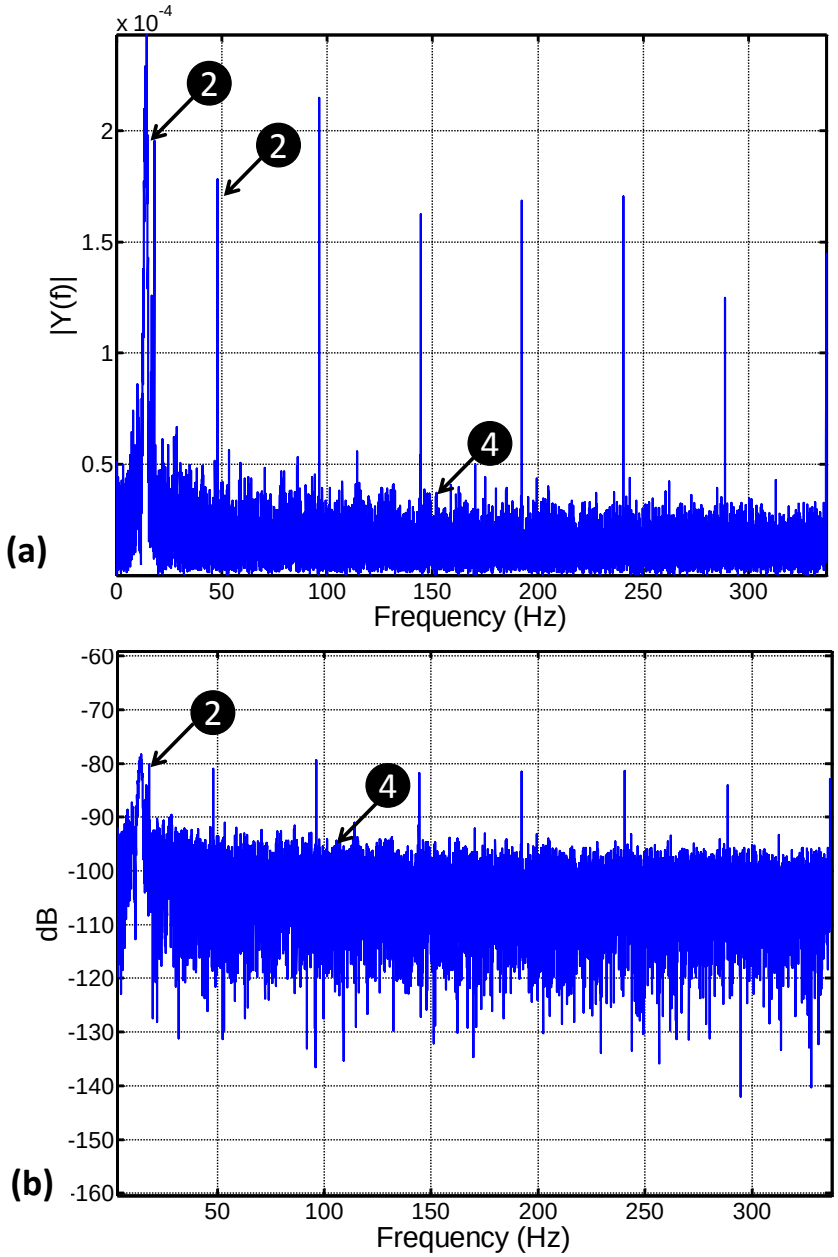


Fig. 6: FFT (a) and Power spectral density (PSD) (b) of ambient vibration sensor data gathered for ~ 30 seconds at ~690 Hz. Portions marked 2, partially accounts for corresponding components (Fig. 5 (b)). Portion marked 4 is the white noise component, and is in agreement with Fig. 5 (b).

Influence of load and spindle speed

Full factorial design of experiments tests (T 1) were conducted under different load and spindle speed conditions for both polishing and no-polishing cases. For the no-polishing case the workpiece was mounted on the spindle and polishing simulated without actual contact between the pad and the workpiece. These tests were in accordance with a 2 × 2 full factorial design of experiment (T 1) for a total of four treatment conditions, with load and spindle speed as main factors each at two levels of; 2 lb. and 8 lb., 30 RPM and 60 RPM, respectively. The platen speed was maintained constant at 150 RPM. Each treatment condition was replicated thrice. Vibration sensor data in the tangential (V_X) and radial direction (V_Y) was gathered for 3 min test runs, at ~ 690 Hz, yielding ~ 120,000 data points (for each interval, for each sensor direction). From this data set, the first and last 30 seconds of data was quarantined in order to eliminate presence of transient elements in the signal. Thus, ~ 80,000 data points are used for this analysis. We present the results from the V_X direction vibration sensor. The observations drawn from this data were subsequently corroborated with the V_Y direction data.

T 1: 2 x 2 full factorial design of experiment for Test 1

Treatment Condition	Process Variables	
	Downforce (load) lb.	Spindle (Head) Speed RPM
1	2	30
2	8	60
3	2	30
4	8	60
Fixed (constant) factors Platen Speed (RPM): 150 Pad type: Buehler microcloth Polishing time : 3 min Slurry flow rate: 25 ml/min		

Observation of the spectrum at ~ 120 Hz (component 3) shown in Fig. 7 reveals that during polishing the energy content in this component is almost twice as high (Fig. 7, (b1) and (b2)) compared to the no polishing state for the same load (Fig. 7, (a1) and (a2)). Secondly, for the same test condition i.e., polish or no-polish, as the load is increased from 2 lb. (Fig. 7 (a1) and (b1)) to 8 lb. (Fig. 7, (a2) and (b2)) the magnitude of frequency component 3 increases in the range of 50% – 75%. However, this apparent variation can lead to ambiguous results under lax test conditions. A more complete test to ascertain the effect of load on the 120 Hz region, which quantifies the variation in signal component 3 is presented in the forthcoming section.

Fig. 8 shows a comparison of signals between the 30 RPM vs. 60 RPM spindle rotation with all other factors maintained identical. As spindle speed is increased to 60 RPM (Fig. 8 (a2) and (b2)), for both the no polish (Fig. 8 (a1) and (a2)) vs. polish conditions (Fig. 8, (b1) and (b2)) we notice the following:

i. 50 Hz region (component 2) is ~ 1.5 – 2 times higher compared to 30 RPM conditions.

ii. 120 Hz region (component 3) is not significantly affected by change in spindle speed.

iii. 25 Hz region (component 2) is attenuated with the 50 Hz region.

iv. White Gaussian noise (component 4, see Fig. 5) does not change significantly for various conditions tested.

v. Several broadband regions between 50 Hz and 100 Hz which were insignificant in the 30 RPM case, particularly around the 75 Hz region, become prominent.

Notably, component 2 remained unaffected when load was varied (i.e., spindle speed maintained constant - compare left ((a1), (b1)) and right ((a2), (b2)) panels in Fig. 8) for both polish and no-polish conditions. However, a variation in spindle speed (i.e., load maintained constant) significantly alters both the shape and magnitude of component 2 (compare left ((a1), (b1)) and right ((a2), (b2)) panels in Fig. 8). This may lead to the conjecture that component 2 is likely a combination of aspects associated with the spindle machine elements and sensor characteristics. Given this association of component 2 with extraneous noise, it would not be conducive to use the same from a process monitoring standpoint.

The effect of spindle speed on component 1 which represents the low frequency component (0.5 – 1 Hz) region of the signal is apparent in Fig. 9. For the 30 RPM condition (Fig. 9, (a1) and (b1)) component 1 first appears at 0.5 Hz and

~ 1 Hz (the first integer multiple), at 60 RPM, (Fig. 9, (a2) and (b2)) this component is seen at 1 Hz and 2 Hz (the first integer multiple). Interestingly, the magnitude of component 1 appears to be higher during polishing (Fig. 9, bottom panel) compared to no-polishing state (Fig. 9 top panel). However, this observation is only affirmed in 9 of 12 total cases – thus being less reliable from process monitoring applications in comparison to component 3 as an indicator of applied down force. This test leads to the following inferences:

Referring to Fig. 7

i. Magnitude of the vibration signal in the 120 Hz region, i.e., component 3, is responsive to polish vs. no-polish states. The magnitude of component 3 is almost double during polish state in comparison to no-polish state.

ii. Increase in down force from 2 lb. to 8 lb. lead to increase of 50% – 75% in magnitude of component 3 in the frequency spectrum.

Referring to Fig. 8

iii. Component 2 is likely an artifact emerging from a combination of the characteristics of the sensor response, and mechanical, electromagnetic aspects of the spindle motor.

iv. Taken together (i) and (ii), help associate component 3 as a likely indicator of process dynamics responsive to variation in downforce.

Referring to Fig. 9

v. Low component frequency, i.e., component 1, appears irrespective of polishing vs. no polishing conditions.

vi. Component 1 is closely correlated with the spindle speed. The first peak appears at frequencies corresponding to the spindle RPM. At 30 RPM, the first peak is seen at 0.5 Hz. Similarly, at 60 RPM, the first peak appears at 1 Hz. This observation was confirmed with the spindle speed at different RPM settings.

vii. From (v) and (vi), component 1 can be conjectured to be a result of the spindle dynamics. However, the possibility of close resonance among machine elements leading to component 1 cannot be ruled out at this stage.

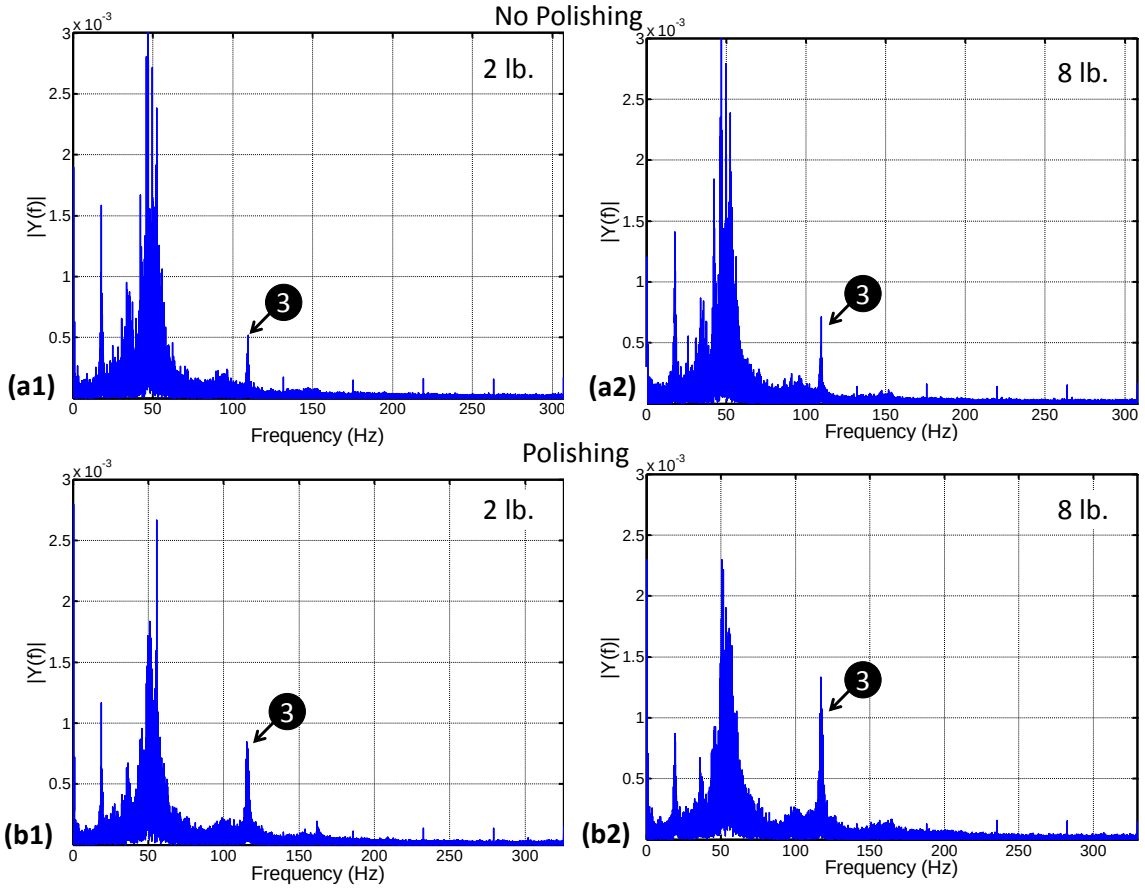


Fig. 7: Comparison of No Polishing (a, top panel) vs. polishing condition (b, bottom panel) wrt to different down force conditions for identical spindle speed of 30 RPM. 2 lb (left panel, marked suffix 1) and 8 lb. load (right panel, marked suffix 2). Magnitude of component 3, representing the 120 Hz region is markedly higher for polish vs. no polish condition (compare top and bottom panels). 8 lb. load leads to ~ 50% increase in magnitude of component 3 (compare left and right panels).

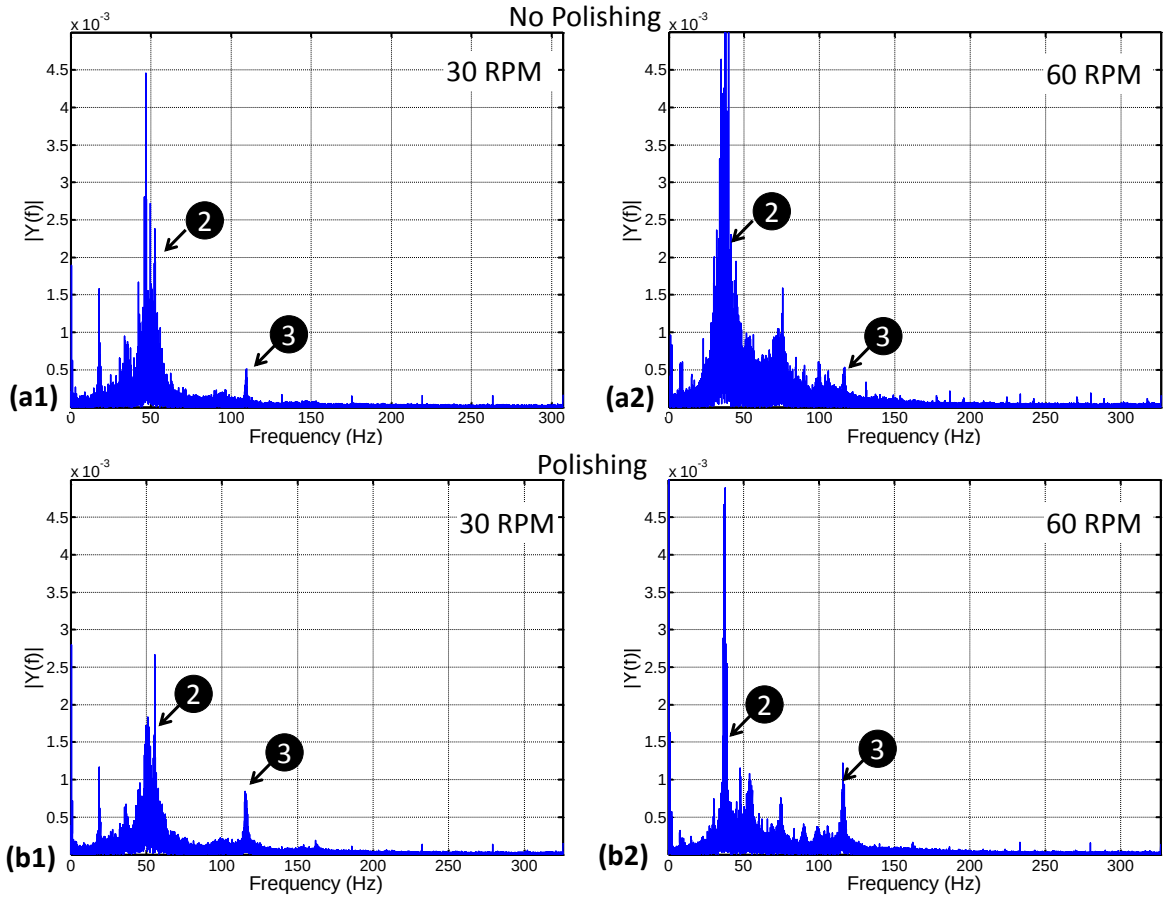


Fig. 8: Comparison of No Polishing (a, top panel) vs. polishing condition (b, bottom panel) wrt to different spindle speed conditions, with 30 RPM (left panel, marked suffix 1) and 60 RPM (right panel, marked suffix 2) for identical downforce of 2 lb. Magnitude of component 3, does not change significantly with RPM.

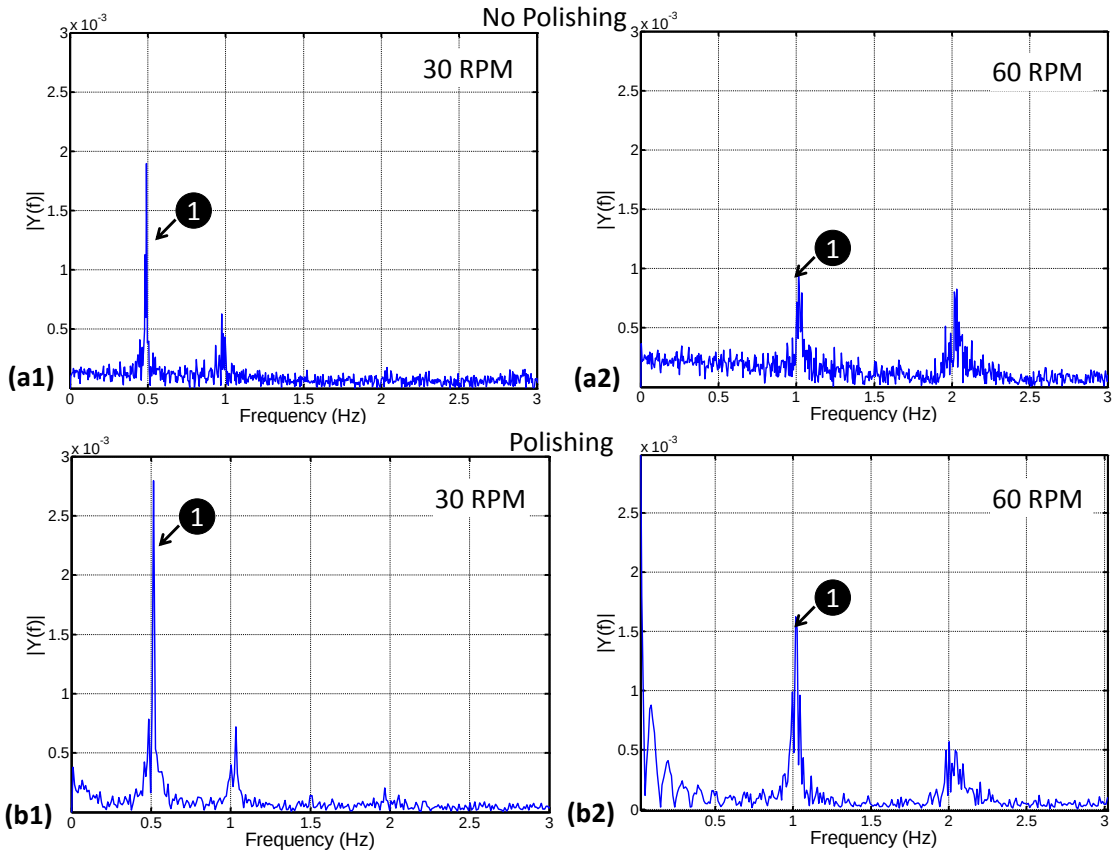


Fig. 9: Zoomed in portions of conditions corresponding to Fig. 8. Sharp peaks in the frequency spectrum appear corresponding to the spindle speed. No Polishing (a, top panel) vs. polishing condition (b, bottom panel) wrt to different spindle speed conditions, with 30 RPM (left panel, marked suffix 1) and 60 RPM (right panel, marked suffix 2) for identical downforce of 2 lb.

Influence of machine elements

In order to further ascertain the origin of component 2 (see Fig. 7), sensor data was gathered during the course of the various machine elements (see Fig. 10) being activated just before and after polishing. A linear motion (Fig. 10 (b)) is imparted to the spindle by a 12V DC motor before and after polishing. This motor engages with a gear and pinion assembly via a toothed belt drive, and drives the spindle in the vertical direction. This vertical motion allows the workpiece to make contact and disengage from the platen before and after polishing respectively. Rotational motion to the spindle is imparted by a different 60V permanent magnet DC (PMDC), 1/6 HP eccentric shaft gear motor. It must be noted that the two motors never operate in unison. The machine operates according to the following sequence:

a. Before polishing phase (Fig. 11)
   i. Wafer is mounted on the holder (Fig. 11 (a)).
   ii. Spindle is lowered onto the platen, during this phase the 12V motor is engaged (Fig. 11 (b)).
   iii. Machine senses contact between the work piece and base, spindle stops lowering and holds position (Fig. 11 (c)).
   iv. Air is supplied to pneumatic actuators to apply set load upon the workpiece (Fig. 11 (c)).
   v. Polishing begins after delay of ~ 1 seconds with slurry pump being activated concurrently (Fig. 11 (d)).

b. After polishing phase

i. Spindle stops rotating after set polishing time.

ii. 60V motor is disengaged; phase (iii) and (iv) begin after a delay of ~ 3 – 4 seconds.

iii. Pneumatic actuators rise and disengage from the workpiece.

iv. Spindle begins to raise up to default position by means of the 12V motor.

The effect of machine components on vibration patterns can be isolated by gathering data during the spindle lifting and lowering phases for polishing. Analysis of vibration data (V_Z, ~ 500 Hz sampling rate) during the lowering phase enables measurement of spindle vibration due to activation of the 12V DC motor engage phase. Similarly, vibration measurements taken during the disengage phase allows; (i) isolation of pneumatic actuator vibration pattern, (ii) estimating the ambient contribution of the 60V PMDC motor during the 3 – 4 seconds delay before the pneumatic actuators lose contact with the workpiece, and (iii) a second measurement of the 12V motor vibration characteristics.

Analysis of vibration sensor data during from the engage phase (lowering) of the spindle is shown in the top panel of Fig. 12 (a1 and a2). Bottom panel of Fig. 12 (b1 and b2) depicts the same during the disengage (lifting) phase. During the engage phase, the vibration dynamics of the 12V motor fails to show presence of dominant frequencies (Fig. 12 (a2)) – indicating that this phase is composed of Gaussian white noise. On the other hand, during the disengage phase the data reveals (Fig. 12 (b)) presence of dominant frequencies at integer multiples of 50

Hz. This is in accordance with the 50 Hz AC power supply specification for the motor. Furthermore, juxtaposition (Fig. 12 (b2)) of the frequency spectrum from the lifting phase and experimental data (e.g., Fig. 12 (a1)) obtained during polishing shows close agreement in the magnitude of the 50 Hz region (component 2). The prominent 200 Hz frequency region in evidence Fig. 12 (b2) is likely due to action of the pneumatic actuators. Also, comparison of Fig. 12 (a2) and Fig. 12 (b2) reveals that the overall contribution of the 12V motor (deemed white noise) is minuscule in comparison to that of the 60V motor. Also, we cannot perceive the presence of a dominant frequency in Fig. 12 (b2) in the 120 Hz region (component 3).

Hence, component 3 may be considered as a viable indicator of polishing state, and not a mere artifact. More pertinently, the results from this test can lead us to infer with greater confidence that the 50 Hz frequency region (component 2) most likely is a resultant combination of electromagnetic effects from the system – in particular, the 60V DC gear motor, and sensor noise characteristics Fig. 10).

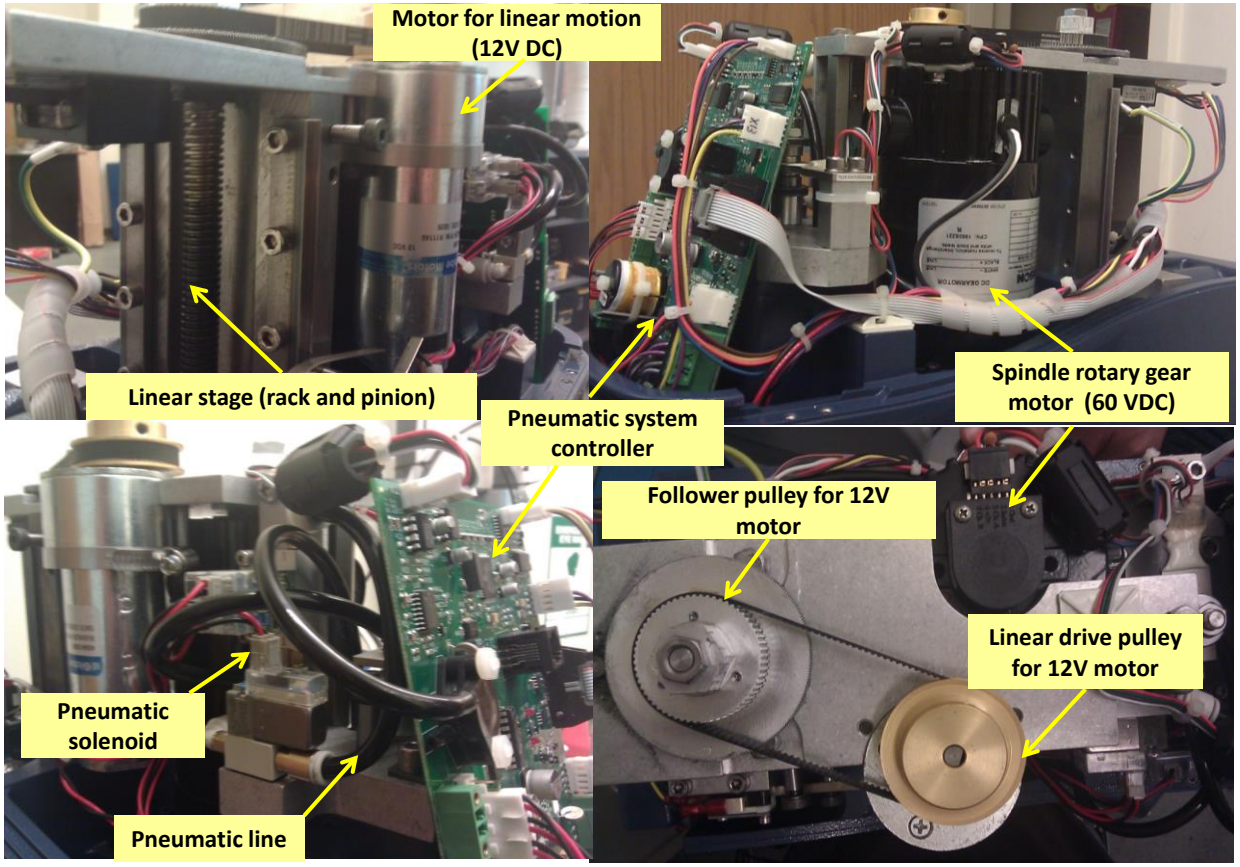


Fig. 10: Photographs showing the main components of the polishing machine: DC motors (12 VDC motor for the linear stage, and 60 V PMDC gear motor for spindle rotation), drive arrangement, pneumatic system for controlling downforce, and microcontroller.

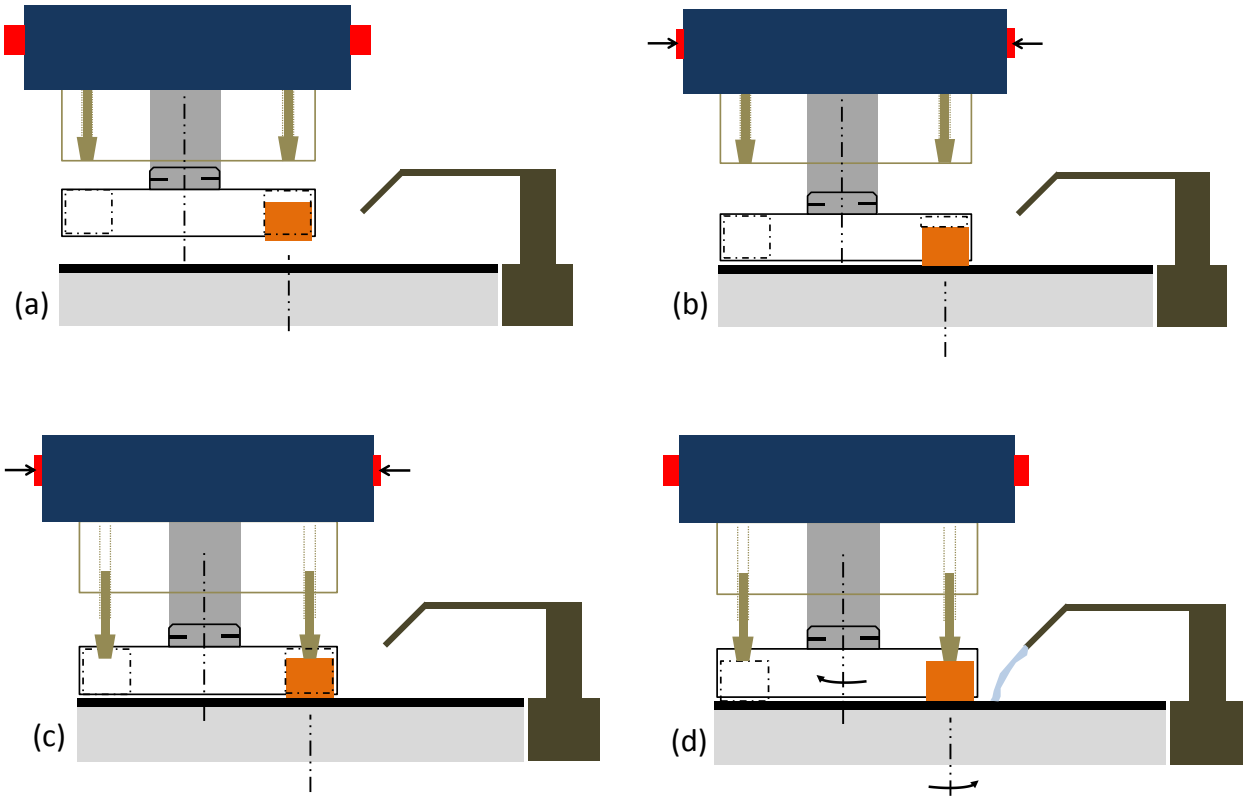


Fig. 11: Sequence of operations just prior to start of polishing (a) Workpiece is loaded on the spindle along with the sensor (not shown) (b) The spindle is lowered onto the platen by means of the linear drive motor along a gear and pinion rack (c) The pneumatic system is activated, and applies the set downforce, after a delay of ~1 seconds the polishing action (d) begins with slurry being dispensed synchronously. The sequence operates in reverse after the end of polishing.

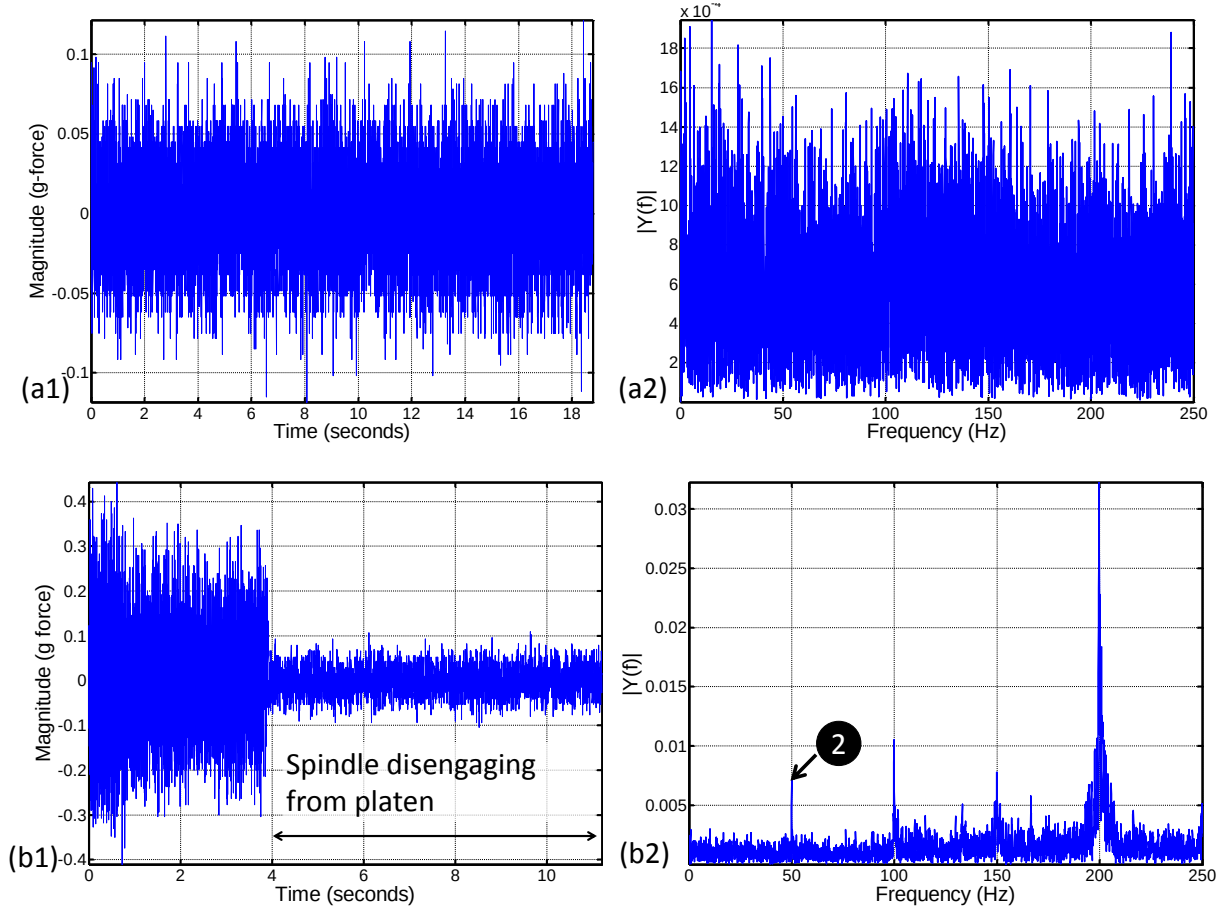


Fig. 12: Vibration sensor data gathered during the spindle engage, (top panel, marked a) and disengage phases (bottom panel, marked b). The time series for the two phases and their frequency spectrum are shown in the left (suffix 1) and right panels (suffix 2) respectively.

Tests with varying load conditions

As noted in the foregoing, the region around ~ 120 Hz (component 3) in the frequency spectrum was responsive to the different levels of downforce applied upon the workpiece. The following experimental procedure was followed to further study the effect of applied downforce on the sensor signal patterns. Three identical workpieces initially lapped to a surface finish Ra ~ 10 nm were polished on the CMP setup for one minute. Each workpiece was subjected to three different downforce conditions of 2 lb., 5 lb., and 8 lb. with spindle (head) and platen (base) speed maintained at 60 and 150 RPM respectively. In addition, each of the workpieces were polished thrice for a given load. A fresh polishing pad and silica slurry at 100 ml/min flow rate was used for each experiment. Thus, there are 9 data points for each load condition giving a total of 27 test points. Sensor data was gathered for the complete 1 minutes run of the experiment. Sampling rate was set at ~ 690 Hz for both the tangential (V_X) and radial direction (V_Y) vibration sensor channels. The first and last 5 seconds of the data are not used for the analysis to exclude transient signal features, thus in all ~ 35,000 vibration sensor data are obtained for each of the 27 test points.

Since the frequency spectrum of the vibration signals show marked broadband characteristics in the region of interest (component 3), a Gaussian weighted window with a mean centered at 117.5 Hz and standard deviation of 0.4 Hz was used to capture the energy content in this region. Though this could be considered a relatively broader window size, it minimizes the possibility of losing pertinent

information, albeit at the risk of being more inclusive of extraneous noise. The result from this analysis is shown in Fig. 13. The following inferences can be drawn from these analysis:

i. The energy content (sum of squares of amplitudes in the frequency spectrum) in the 115 – 120 Hz region is significantly different for the three loads applied (Fig. 13 (a) and (b)). A statistical ANOVA analysis of the data confirms this observation – the p-value was close to zero (<< 0.01), with R^2 ~ 80% in a linear regression setting. Tukey's pairwise comparison tests also revealed statistical significance at the 95% confidence level for all pairwise combinations tested.

ii. The energy content in the 115 – 120 Hz region, tends to increase with applied load, however the trend is not linear. For calibrated g-force units, the average increase in energy when load is increased from 2 lb. to 5 lb. is ~ 50%, while from 5 lb. to 8 lb. the increase is only ~ 25%.

iii. For the 2 lb. and 5 lb. downforce conditions, the energy values appear to be more tightly clustered in comparison to the 8 lb. load condition (Fig. 13 (a1) and (b1)). Close examination of the figure shows that the majority of the anomalous behavior stems from data from workpiece #3.

Inspection of workpieces showed that workpiece #3 has a tendency to fit more tightly compared to other workpieces in the holder. Presence of burrs and sharp notches from machining operations are known to cause the workpieces to

wedge. When a workpiece is wedged in the holder, the pneumatic actuators are hindered from applying the correct load. In cases of severe wedging damage to the platen can result. On the other hand, minor occurrence of workpiece wedging is harder to detect, since a susceptible workpiece can slide freely when initially loaded, but may get wedged due to agglomeration of slurry particles within the holder during polishing – as is most likely the case with workpiece # 3. When workpiece # 3 was censored and the data reanalyzed, a much tighter variation with lesser ambiguity is in evidence (Fig. 13 (a2) and (b2)).

Thus from this experiment, it can be concluded that the sensor signals are sensitive to the variation in downforce. The energy content in the 120 Hz region (component 3) is responsive to the applied load, and the difference in energy levels wrt to down force is statistically significant. In general, an ~ 75% increase in mean energy level is observed for a corresponding increase in applied load from 2 lb. to 8 lb.

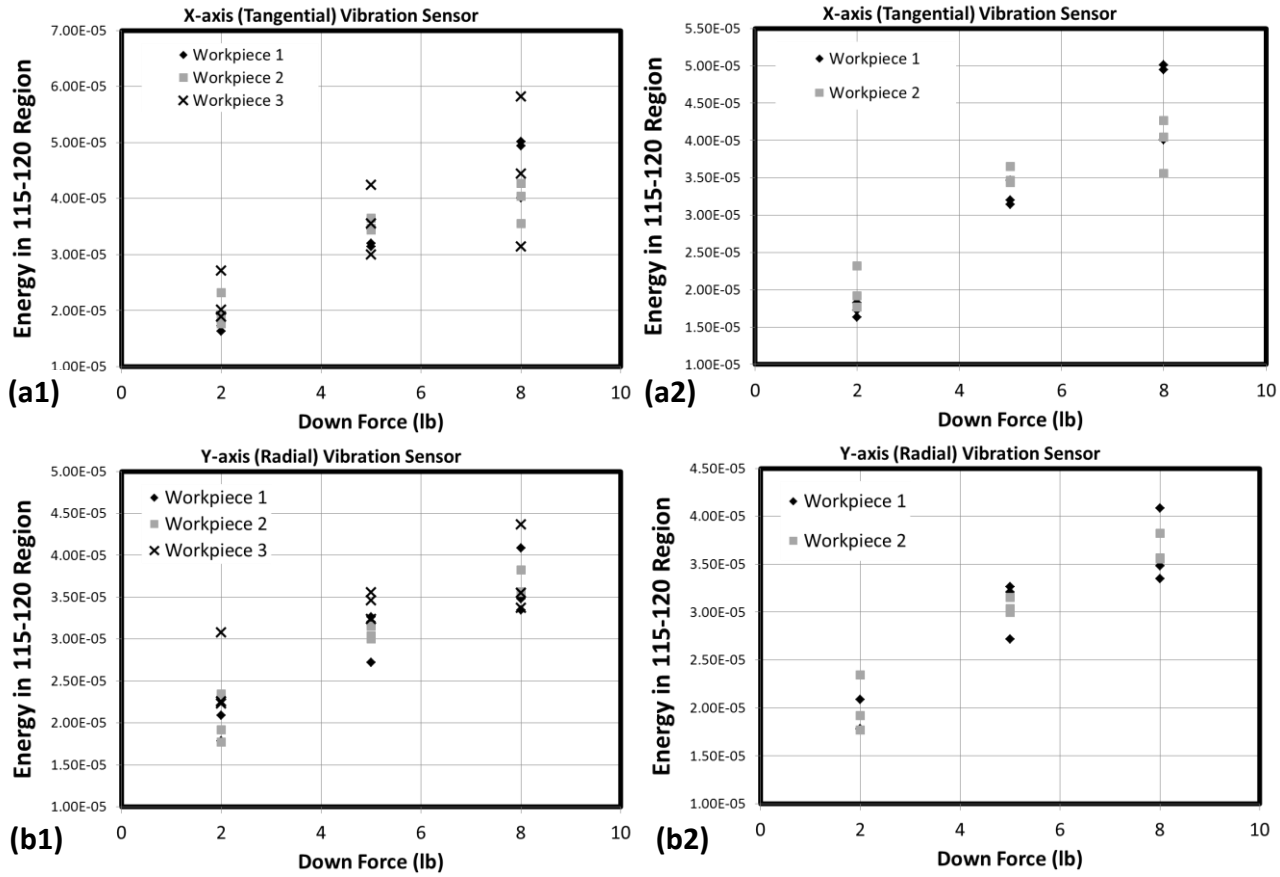


Fig. 13: Effect of applied down force (load) on energy content in the ~120 Hz region (component 3). Tangential sensor results shown in the top panel, radial sensor results shown in the bottom panel. Left panel shows data from all workpieces, right panel with workpiece 3 censored.

Modal analysis of machine structure

We previously Fig. 6 delineated the ambient behavior of the sensor. In the foregoing, the presence of a prominent broadband characteristics at ~ 25 Hz followed by sharp peaks at integer multiples was noticed. The region around 25 Hz was noted as a likely manifestation of sensor ambient response. Next, we hypothesized the effect of the motors and other machine elements associated with rotational and linear motion aspects of the spindle, namely; the 50 Hz frequency likely originating as a result of electromagnetic behavior of the machine components.

With this test, we intend to further confirm the above hypothesis by monitoring the vibration response characteristics (modal testing) of the polishing machine structure by subjecting it to impacts. In order to characterize the resonance behavior of the polishing machine, tests were conducted by exciting the machine structure under different conditions. These modal analysis tests are carried out according to the following schema:

1. Two different hammers are used to excite the column supporting the spindle head; (i) hard rubber mallet with a wood handle, and (ii) stainless steel hammer with a steel handle.
2. The workpiece holder is loaded under the following conditions (i) with no workpieces affixed, and (ii) with four workpieces, each weighing ~ 250 gm.

The MEMS vibration sensor is mounted and calibrated to gather data in the vertical direction (along the column). The sensor is fixed on the spindle head with double-sided adhesive tape. Sampling frequency ~ 690 Hz is maintained. In accordance to the orientation change of the sensor, the traditionally tangential direction (V_X) of the sensor now monitors vibration in the vertical (z-axis, machine coordinate) direction whilst the radial direction sensor (V_Y) now measures along the horizontal (y-axis, machine coordinate) direction in the X-Y plane.

The machine column is subjected to impacts at ~ 2 seconds intervals for a total of 25 seconds (~10 – 11 impacts). Each of these experiments is repeated twice. Typical results from this test are shown in Fig. 14 (y-direction sensor data for modal tests conducted with the rubber mallet, no workpieces mounted). We observe from Fig. 14 (a) that the resulting structural excitations are damped within a short time (~ 0.35 sec) attesting to the high relative stiffness of the machine assembly. Also, the low sampling frequency (~ 690 Hz) of the MEMS vibration sensors used for these tests do not afford the higher resolution typically possible with wired piezoelectric sensors, where the sampling frequency can be set to well over 10 kHz. However, as evidenced before (see Fig. 5), the various vibration signal components of interest are within 250 Hz, i.e., less than half the maximum sampling frequency of the MEMS sensors. Hence, the data acquired from our MEMS vibration sensors can serve as a reasonable estimate of the various active resonant frequencies in the machine structure.

In contrast to the ambient sensor characteristics (Fig. 6), the 25 Hz and 50 Hz regions (component 2) for these tests are of comparable magnitude (Fig. 14 (b)). In the ambient sensor test (Fig. 14 (a)), the 25 Hz frequency dwarfed the 50 Hz region. On the contrary, for this test the 50 Hz frequency is almost twice as large in magnitude compared to the ambient case. Secondly, component 3 is absent in the Fig. 14 (b), thus indicating that it (component 3) is unrelated to the modal characteristics of the machine. Also, we observed that the presence of 4 workpieces seems to damp the resonance behavior in the 50 Hz region by ~ 50% compared to the no work piece condition depicted, however the 25 Hz region is not significantly affected.

These tests characterize the vibration patterns emerging from resonance associated with the machine structure. Component 2 can therefore be considered as a combined manifestation of sensor ambient response (25 Hz), electromagnetic interference from machine components (50 Hz), and structural resonance (50 Hz).

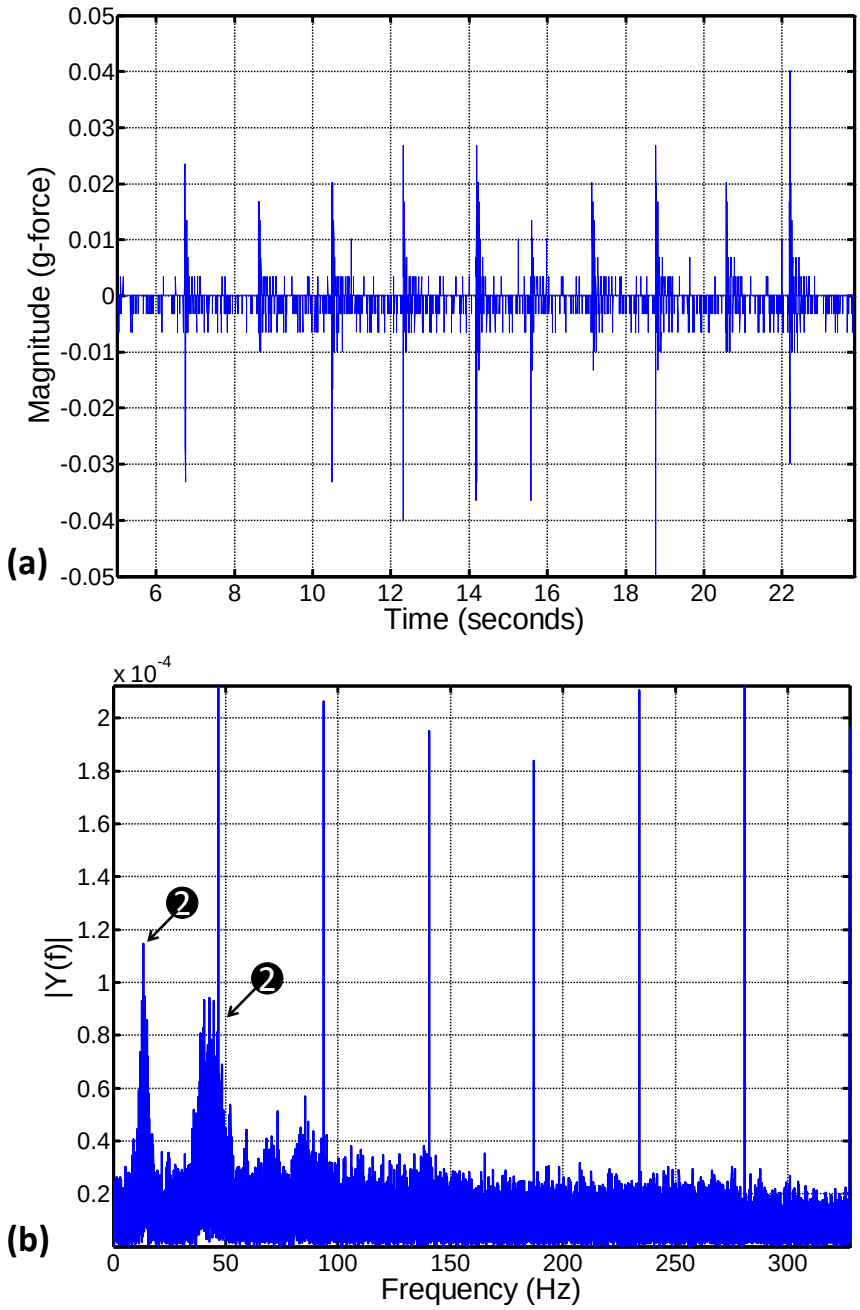


Fig. 14: Modal analysis results from striking with a rubber mallet when no workpieces are mounted, taken in the y-direction. The machine structure seems to depict a resonance in the region of component 2 (25, 50 Hz) and its multiples.

Influence of spindle start and stop positions

In previous sections we alluded to the possibility of low frequency (0.5 – 1 Hz, component 1) component arising out of mechanical and geometrical errors from the spindle. However, we have not thus far eliminated the possibility that two coupled frequencies might also cause such behavior. Low frequency modulated patterns are a common phenomenon associated with vibrations and acoustics caused when two signals or waveforms of very close frequencies or very far apart frequencies interact. The result of the interaction of the first kind is the addition of a low frequency component corresponding to the difference of the two interacting frequencies. The second form of interaction results in an amplitude modulation of the higher frequency signal by the lower frequency signal. In either case the resulting waveform shows a periodic low frequency oscillation.

Akiyama, *et al.* [226, 227] attributed low frequency patterns (called *beats* by the author) observed in induction motors to an eccentricity in the rotor. The frequency at which the beats occurred was found to be twice the slip frequency of the induction motor due to the interaction of the electromagnetic and mechanical systems of the motor. Tests were conducted on prepared versions of four commonly found eccentricities in rotors, namely:

(a) **Static eccentricity** due to axial misalignment of shaft and rotor centers,

(b) **Dynamic eccentricity** akin to whirling of a bent shaft,

(c) **Inclined eccentricity** from angular misalignment between shaft and rotor axis, and

(d) A combination of the first two types.

Observation of vibration signal data showed that the *beat* phenomenon occurred only in cases (b) and (d) with the latter showing lower vibration amplitude, leading to the conclusion that *beats* in induction motors are a result of dynamic eccentricity caused by a bent and axially misaligned shaft which leads to a whirling-like action. Similar low frequency modulated components are documented [228-230] to occur in rotating machinery, primarily due to eccentricity and imbalance in bearings.

Carter, *et al.* [119] attributed similar low frequency components observed in the friction force sensor signal in the CMP process, noting its correlation with the frequency of rotation of the carrier. Dornfeld, *et al.* [112] utilized piezoelectric and acoustic emission sensors for end point detection in CMP process. The spectrum of the vibration signal obtained from the piezolelectric sensors shows similar low frequency components. The frequency of these components varies with the head (carrier) RPM but there is no change in the magnitude, further confirming the hypothesis that the low frequency components are a result of some form of eccentricity in the head (carrier) rotor or shaft.

In this test, we monitor the effect of starting the spindle at measured positions on the corresponding vibration pattern. We also monitor the vibration signal when the spindle is suddenly stopped. For the first part of the test, the workholder is loaded with a workpiece and the spindle is manually rotated to marked positions. In all, 8 radial positions at intervals of 45° are designated as starting positions and

polishing commenced. Vibration sensor data in the tangential and radial are gathered at sampling rate of ~ 690 Hz. Polishing continues for ~ 1 minutes Each test is repeated thrice. Representative results, for the quick start test at positions of 0°, 180°, and 315° are shown in Fig. 15 (a), (b), and (c) respectively. Fig. 15 we notice that the starting point of the characteristic low frequency component (component 1) is dependent upon the starting position of the spindle. The highest amplitude position along the time series is observed at the 180° (Fig. 15 (b)) position. The 0° (Fig. 15 (a)) and 315° (Fig. 15 (c)) positions are similar implying that the spindle returns to the original state (displacement) after every rotation. Pertinently, this indicates that component 1 is likely caused by geometrical errors in the spindle. Similarly, when the spindle rotation is suddenly stopped, a corresponding abrupt termination of the low frequency waves was observed.

These observations help mitigate the possibility of structural resonance as a cause for component 1, since; partial waves are unlikely in case of periodic behavior originating from structural resonance.

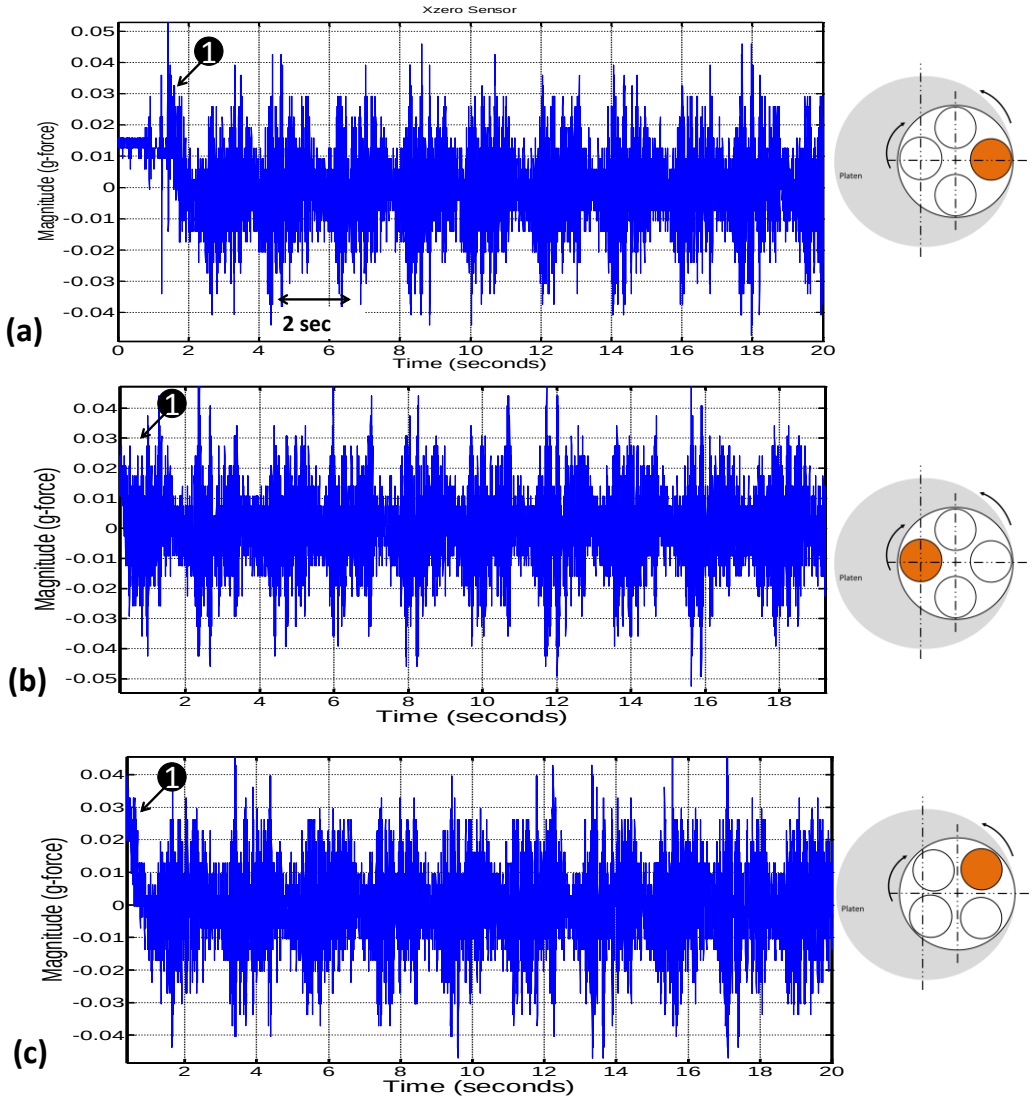


Fig. 15: Quick start operation of the spindle, beginning at (a) 0°, (b) 180°, and (c) 315° positions. The starting position and phase of component 1 (low frequency component corresponding to spindle speed) appear to be correlated. The 0° (a) and 315° (c) are similar. Component 1 attains it highest magnitude at 180° position.

Verification with data gathered from wired piezoelectric sensors

In the previous tests, we have not eliminated the presence of sensor attachments as possible cause of some of the characteristics observed from the vibration sensor data. The added weight from the sensing system, namely; the shaft collar, sensor rest, and sensor power attachments can cause imbalance in the spindle. Such an imbalance can lead to whirling of the spindle shaft, and thus influence the vibration patterns.

Toward this end, the MEMS wireless sensors and associated attachments are disassembled from the machine. Instead, piezoelectric vibration sensors are mounted on various positions (see Fig. 16) on the machine. The sensors shown in Fig. 16 are Kistler (model 8728A500) miniature piezoelectric accelerometers; data is gathered at a sampling rate of 10 kHz, and subsequently amplified. The sensors are mounted in the following positions on the machine; (i) the machine base, (ii) inside the spindle casing nearer to the spindle shaft, and (iii) casing of the 60V motor. Representative results are as follows:

1. The 5 kHz wide frequency spectrum for the polish vs. no polish condition is shown in Fig. 16, Prominent peaks are not seen over 500 Hz. This attests to the viability of wireless sensors with their characteristics low sampling rates being adequate for monitoring CMP process.
2. In the 0 – 500 Hz frequency range, the polish vs. no polish cases show results similar to those discussed earlier. During polishing the 120 Hz region (component 3) has much higher magnitude (Fig. 17 (b2)) com-

pared with the same during no-polish condition (Fig. 17 (a2)). This observation corroborates for component 3 as an indicator of applied down force.

3. Similar to previous observations, prominent peaks in the FFT corresponding to the spindle rotation speed are seen in the 0.5 – 1 Hz region (component 1) (Fig. 17 (a), and (b)).
4. The 50 Hz region (component 2) is present for both polish and no-polish conditions (Fig. 17 (a2), and (b2)).

This test confirms the following:

i. Component 1 (0.5 – 1 Hz) is most likely due to geometry and eccentricity errors from the spindle.

ii. Component 2 (25 Hz, 50 Hz) most likely originates from structural resonance and added electromagnetic effects associated with the machine elements, particularly the spindle motor, and thus can be considered as extraneous noise of little importance from a process monitoring standpoint.

iii. Component 3 (120 Hz) is sensitive to the polishing forces acting at the workpiece-pad interface.

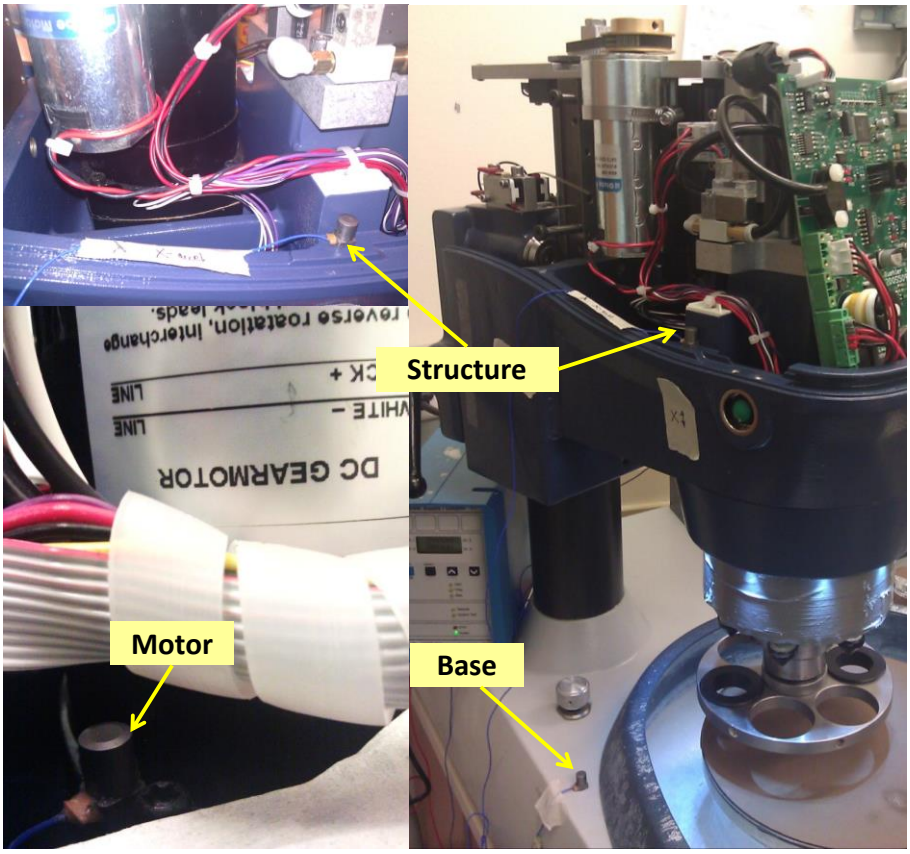


Fig. 16: Position of Kistler (model 8728A500) piezoelectric vibration sensors located on machine. MEMS sensor attachments are removed for tests conducted with piezoelectric sensors.

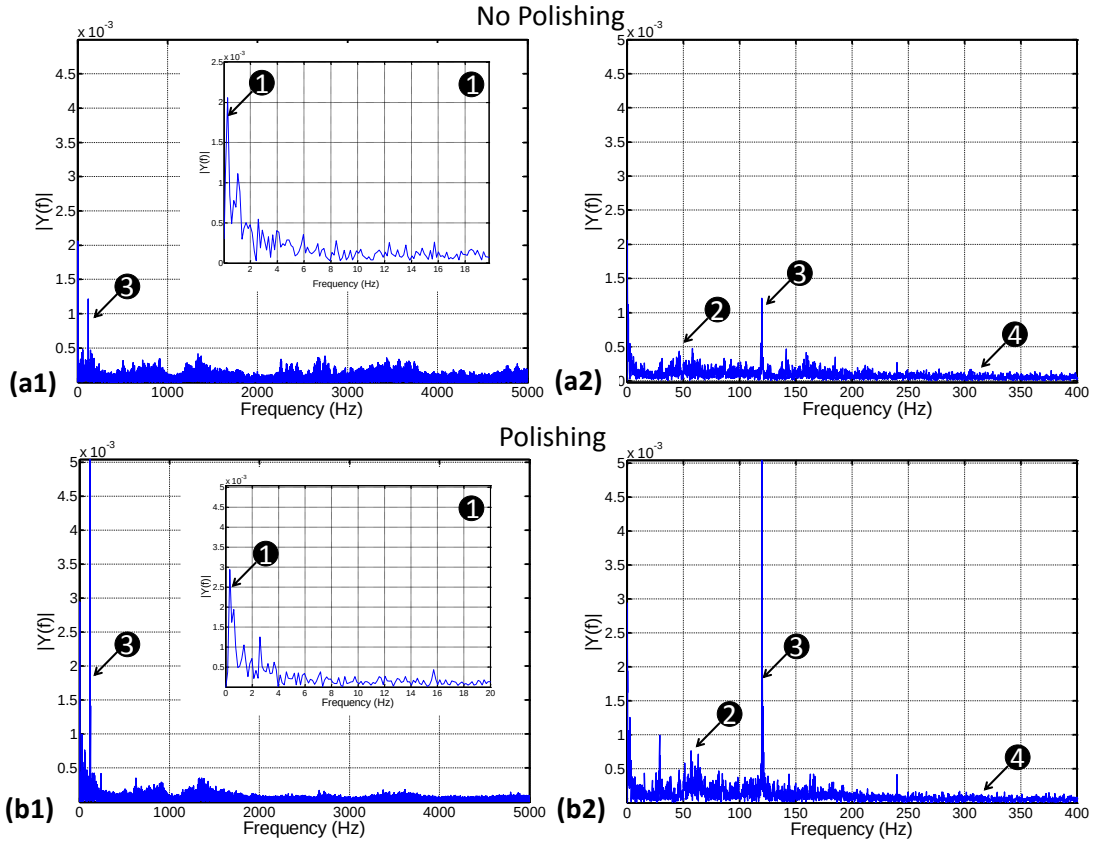


Fig. 17: Vibration signals obtained from piezoelectric sensor mounted near the motor (a) During no polish condition (b) During polishing at 5 lb. down force. Component (1) corresponding to spindle speed (40 RPM) is seen for both polishing (b1) and no-polish conditions (a1). Component 3 – 120 Hz region is ~ 4-5 times higher during polishing (b2) compared to no-polish condition (a2).

Influence of pneumatic system

Though, the possibility of component 1 arising from structural resonance effects has been somewhat mitigated, the likelihood of observed periodic air pressure oscillations from the pneumatic system as a cause (for component 1) has yet to be eliminated.

During polishing, air is periodically let into the cylinders of the pneumatic actuator in order to maintain the set down force. This action is regulated by the microcontroller board shown in Fig. 18. The microcontroller in turn activates a system of solenoids (set at normally closed position) that allow the entry of air into the actuator cylinder. The entry of air into the cylinder can subject the machine to impact loads causing possible resonating vibration in the machine. To investigate this possibility, vibration sensors were mounted on the spindle and solenoid (see Fig. 18). Vibration data was gathered during polishing at different loads and RPM conditions. The sampling rate for each sensor was maintained at ~ 690 Hz. Representative results are shown in Fig. 19.

The solenoid activation process is observed on disassembly of the protective casing around the spindle head. A click is heard, and indicator lights are visible each time the solenoid is activated. With activation of the solenoid the needle on the air pressure indicator drops momentarily in the range of 2 – 3 psi.

The solenoid vibration data depicts sharp peaks as shown Fig. 19. These peaks correspond to the characteristic clicking sound of the solenoid which was noted with a stop watch. Though, the solenoid activates at regular intervals, the

period between successive activations do not correspond with component 1 (see Fig. 6). Also, the solenoid activation rate is observed to vary with applied down force as opposed to spindle speed. For example, at 2 lb. load the solenoid clicks at ~ 2 seconds intervals (Fig. 19 (a)); while at 8 lb. load the interval is ~ 0.5 seconds (Fig. 19. (b)). This is likely, because, at high load greater pressure needs to be maintained inside the actuator, which in turn increases the tendency for air to escape from the actuator seals.

Since the solenoid activation is not observed to be in phase with the low frequency component corresponding to spindle speed (component 1), the possibility of the pneumatic system as cause for the same can be eliminated.

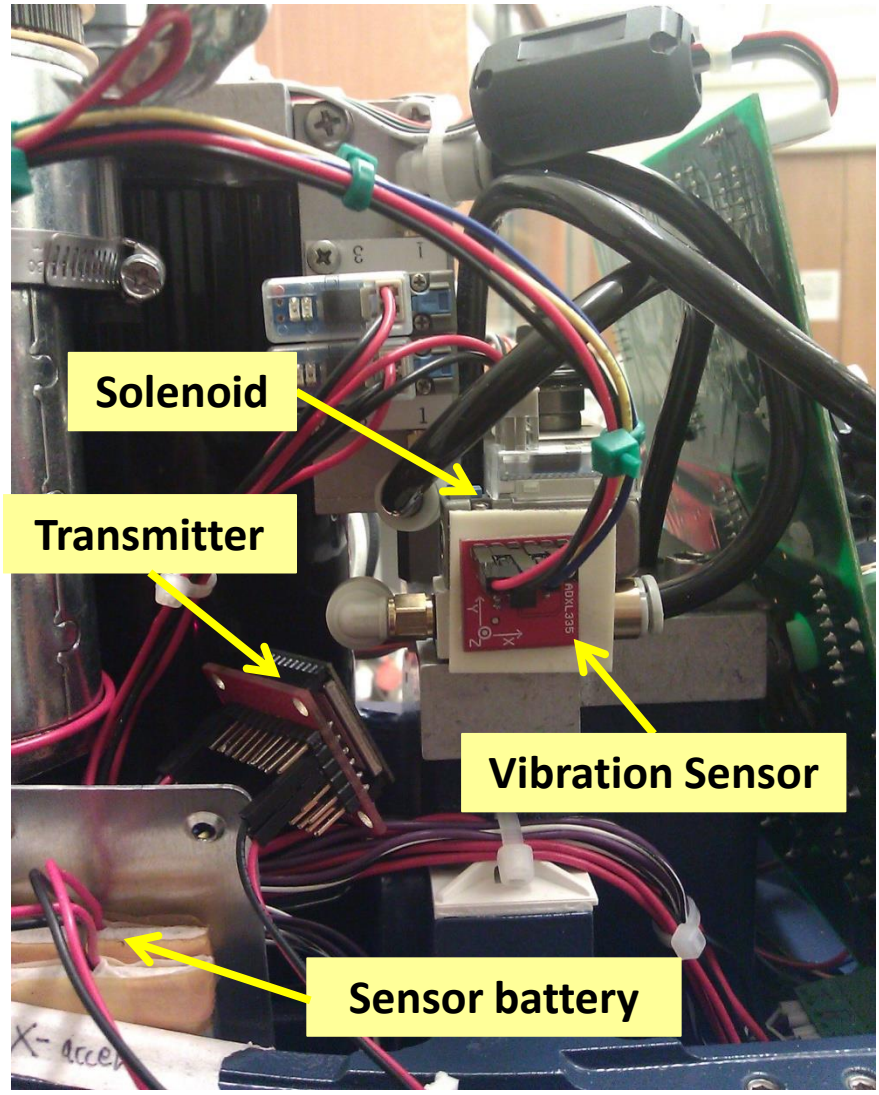


Fig. 18: MEMS vibration sensor mounted on pneumatic solenoid to monitor the behavior of the pneumatic system.

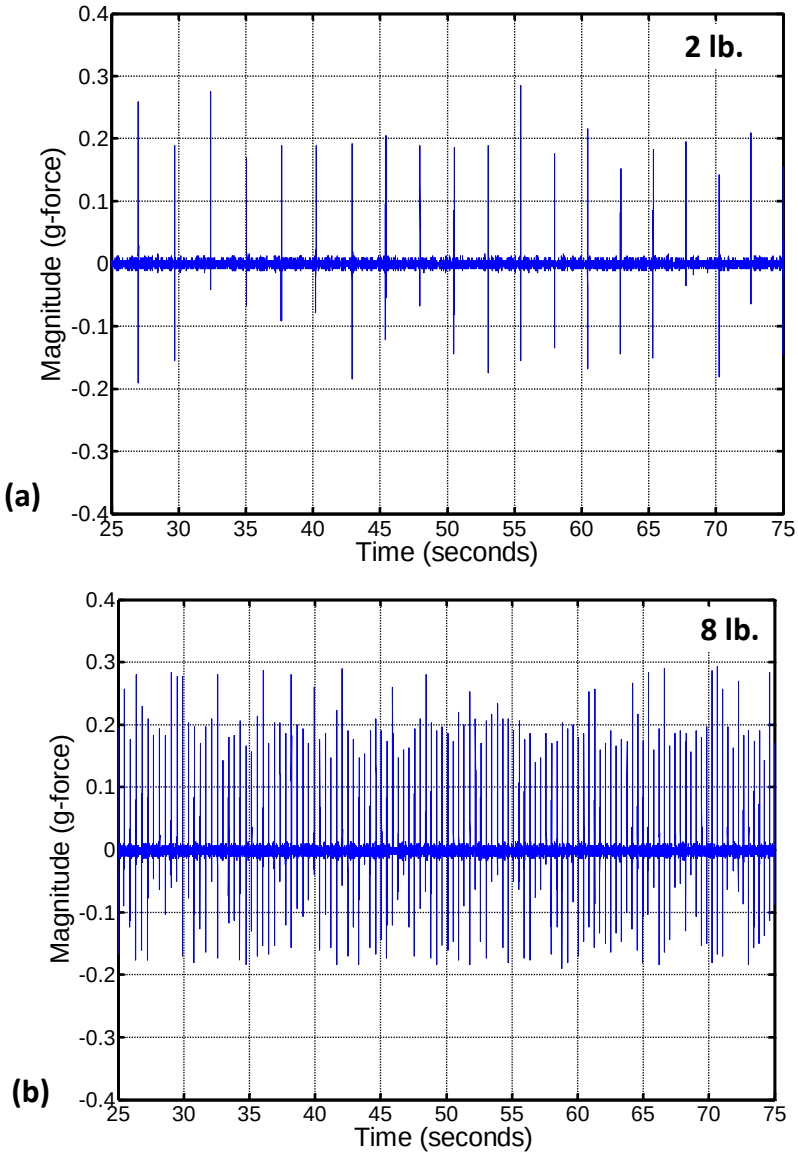


Fig. 19: The solenoid activation cycles for differing load conditions (a) 2 lb., and (b) 8 lb. for the same spindle speed of 60 RPM. The solenoid activation frequency does not seem to be correlated with spindle speed, but is more reflective of applied downforce.

Effect of pad wear

Lapped workpieces (Ra ~ 15 nm) were polished on the CMP setup in 3 min intervals with silica slurry. The polishing conditions were as follows; platen speed 150 RPM, head speed 60 RPM, and downforce 4 lb. Vibration sensor data was gathered during the experiments at sampling rate ~ 690 Hz, akin to tests detailed in the foregoing sections. After 3 minutes of CMP the average workpiece Ra improved to ~ 7 nm. Subsequently, pad wear was accelerated by soaking the pad in slurry for ~ 45 minutes. At the end of 12 minutes of CMP, glazing of the polishing pad is observed (Fig. 20 (a)). In the same interval, scratches were seen on the workpiece (Fig. 20 (b)), and Ra increased to ~ 22 nm. The FFT of the tangential direction (V_X) vibration sensor obtained after 3 minutes, and at the end of 12 minutes (when glazing of pad is observed) are compared in Fig. 21 (a) and (b) respectively. The magnitude of component 3 increases by ~ 30 – 40% at the end of 12 min (Fig. 21 (b)) of CMP, indicating the effect of pad wear on vibration data. Thus component 3 can be considered as being responsive to changes in process parameters (downforce), and process conditions, such as pad wear.

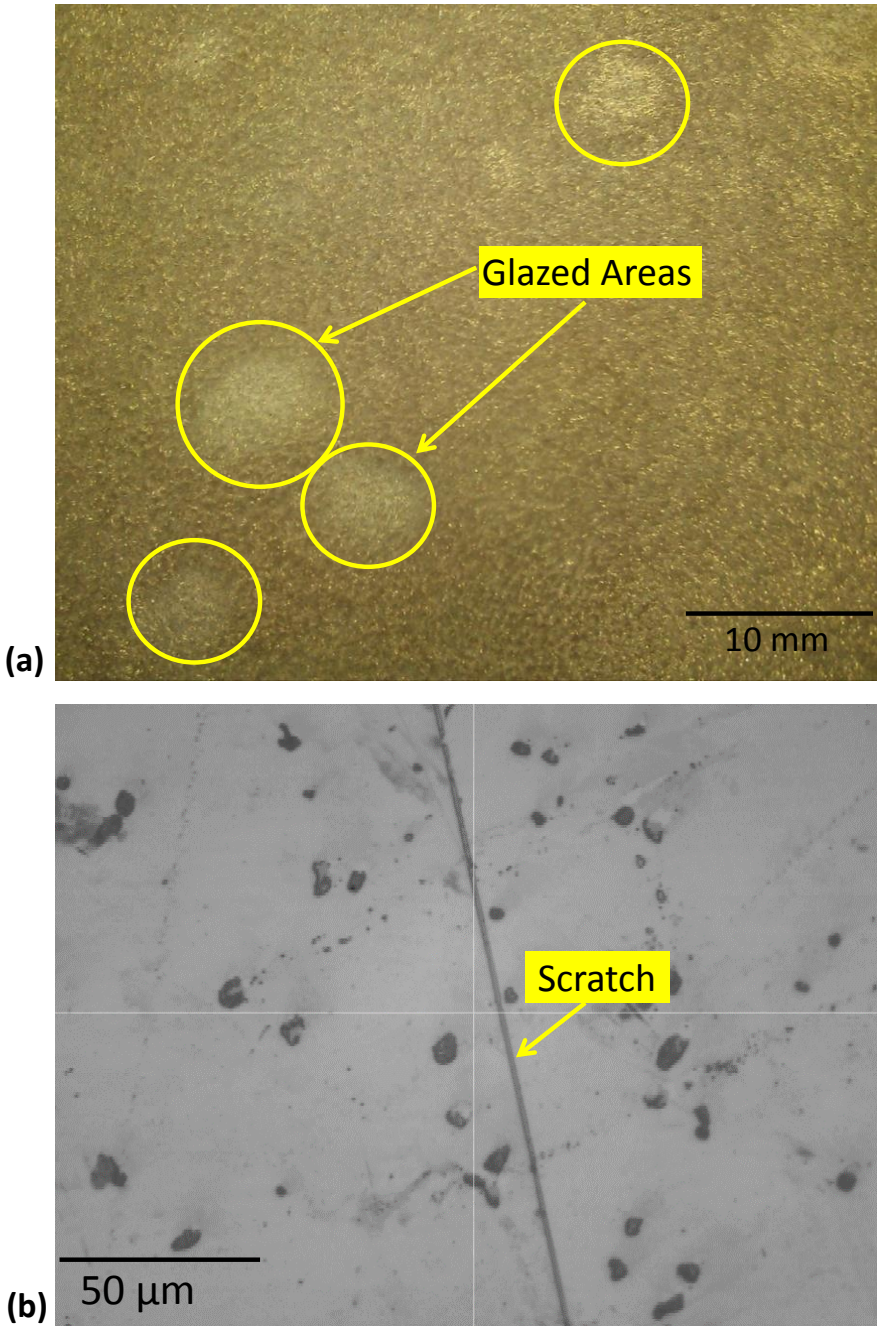


Fig. 20: (a) Glazed pad after 12 minutes of CMP (b) Scratches on workpiece observed at the end of 12 minutes of CMP.

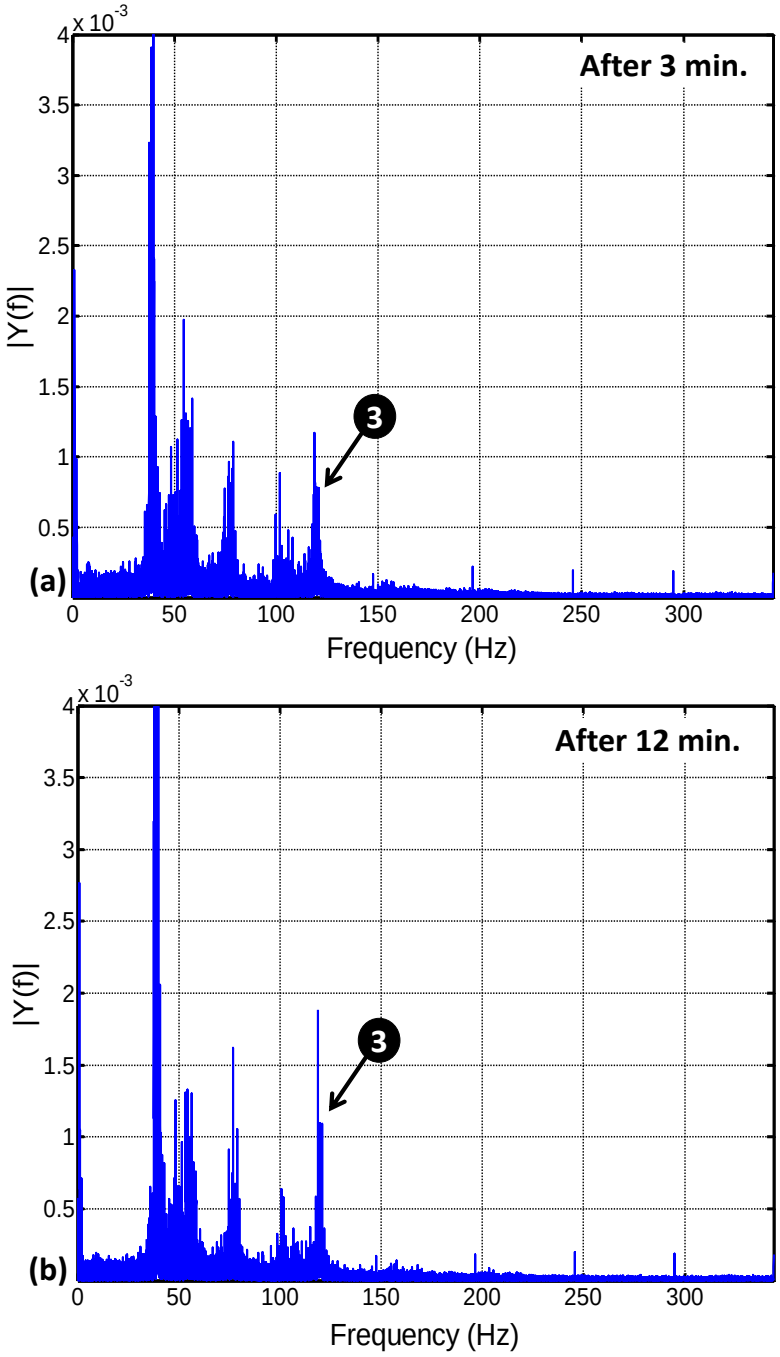


Fig. 21: FFT of experimental vibration data obtained for (a) new pad vs. (b) glazed pad

Summary

Various spectral components of vibration sensor signal patterns from CMP were identified, and their respective underlying physical sources isolated. In all, three dominant components are of interest:

1. Component 1 manifests in the low frequency region between 0.5 – 1 Hz. The first dominant peak in this region was observed to coincide with the spindle (head) speed. Pertinently, the appearance of high amplitude portions for this component was shown to be spatially related with the spindle location with respect to the platen. Tests (see T 2) conducted with regard to component 1 indicate that it is most likely a result of spindle shaft eccentricity.
2. Component 2 comprises of two broadband frequency regions centered around 25 Hz and 50 Hz. It is found to be likely from a conjoined effect of sensor ambient characteristics, electromagnetic interference from machine elements, and vibration from the machine structure (see T 3)
3. Component 3 is observed in the region in the vicinity of 120 Hz; typically centered at 117.5 Hz (T 4) provides a brief summary of the tests conducted to arrive at the following inferences:
    - The spectral energy content for component 3 almost doubles during polishing in comparison to situations where no contact is made between the workpiece and pad.

- A statistically significant relationship was observed between the spectral energy content with changing downforce. Higher energy content is observed at higher downforce conditions. However, this trend is not linear. The average increase in energy when load is increased from 2 lb. to 5 lb. is ~ 50%, while from 5 lb. to 8 lb. the increase is only ~ 25%.
- Component 3 was observed to capture subtle changes in process state, such as pad wear. Due to pad glazing the magnitude of FFT peaks was ~ 50% higher in comparison to conditions where a fresh pad is used.

T 2: Tests and corresponding observations relating to 0.5 Hz – 1Hz component (Component 1)

Source(s)	Test conducted	Observation(s)	Inference(s)
Polishing head (spindle) eccentricity	Full factor DoE with varying spindle speeds	Dominant peaks in the frequency spectrum corresponding to the spindle speed. E.g., for spindle speed set at 30 RPM, a prominent peak is observed at 0.5 Hz, and integer multiples thereon Component 1 is present during both no polish and polish states	Frequency of component 1 is identical to the set spindle speed.
	Quick stop and start tests with spindle set at various positions	The start point of component 1 is dependent upon the starting position of the spindle When the spindle rotation is suddenly stopped, a corresponding abrupt termination of component 1 is observed Highest amplitude of component 1 is observed at a particular radial position of the spindle head	Component 1 does not result from structural resonance, but is likely due to eccentricity errors.
	Piezoelectric vibration sensors at various locations on the machine. Fixtures used for MEMS sensors are removed	Dominant peaks in the frequency spectrum corresponding to the spindle speed.	Component 1 does not result from imbalance due to sensor fixtures used for MEMS sensors.
	Sensors mounted in the vicinity of the machine pneumatic system	Period between activation cycles of the pneumatic valves are dependent upon the applied downforce and not on set spindle speed	Component 1 does not result from activation of the pneumatic valves.

T 3: Tests and corresponding observations relating to 25 Hz, 50 Hz component (Component 2)

Source(s)	Test conducted	Observation(s)	Inference(s)
Ambient sensor noise, electromagnetic interference, and structural resonance	Response of sensor observed under ambient conditions	Presence of a prominent broadband region around 25 Hz and relatively less prominent peak at 50 Hz, and integer multiples thereafter	Presence of measurement noise due to sensor response characteristics. Measurement noise manifests in 25Hz region.
	Sensor data gathered in the course of various machine elements being activated prior and subsequent to polishing	No peaks observed in the 0.5–1 Hz region Dominant peaks in integer multiples of 50 Hz observed during the spindle disengage phase Broadband noise observed during the spindle engage phase just prior to commencement of polish	Machine elements, particularly 60V PMDC spindle motor causes electromagnetic interference manifesting in the 50 Hz broadband spectrum.
	Modal analysis of the machine structure. Structure is excited using different types of hammers	Dominant peaks corresponding to 25 Hz, 50 Hz, and integer multiples thereon. Both 25 Hz and 50 Hz region show prominent broadband characteristics. No peaks observed in 0.5 Hz – 1 Hz region, or 120 Hz region	25 Hz, 50 Hz region accentuated by natural vibration frequency of the machine structure. Structural vibration not a cause for 0.5 – 1 Hz or 120 Hz components.
	Piezoelectric vibration sensors at various locations on the machine. Fixtures used for MEMS sensors are removed	Dominant peaks corresponding to 25 Hz and 50 Hz Magnitude of 50 Hz region does not change significantly between polish vs. no polish states	Electromagnetic interference manifesting in the 50 Hz broadband spectrum.

T 4: Tests and corresponding observations relating to 120 Hz component (Component 3)

Source(s)	Test conducted	Observation(s)	Inference(s)
Process parameters, such as downforce, pad wear, friction at the pad-workpiece interface.	Full factor DoE comparing polish vs. no polish states	Energy in 120 Hz region increases by ~ 75% – 100% during polishing vs. no polishing state maintaining constant load and speed. Spindle speed does not seem to significantly affect energy in component 3.	Component 3 is responsive to polishing condition, and is not effected significantly by changing spindle speed.
	Full factor DoE with varying downforce	Energy in 120 Hz increases with increasing downforce, in a statistically significant manner (p-val < 0.1)	Component 3 is significantly correlated with the applied downforce.
	Polish vs. no polish experiments with piezoelectric sensors	Peak corresponding to 120 Hz is ~ 4 times higher during polishing vs. no polish conditions	Component 3 is not an artifact related with MEMS sensors, but may capture the both friction (due to polishing) and motor torque.
	Experiments simulating pad wear	Peak corresponding to 120 Hz is ~50% in worn pad condition compared to fresh pad	Increased friction at pad-workpiece interface due to worn pad is reflected in component 3

Appendix III: Formulation of the Deterministic Process-Machine Interaction (PMI) Model for CMP

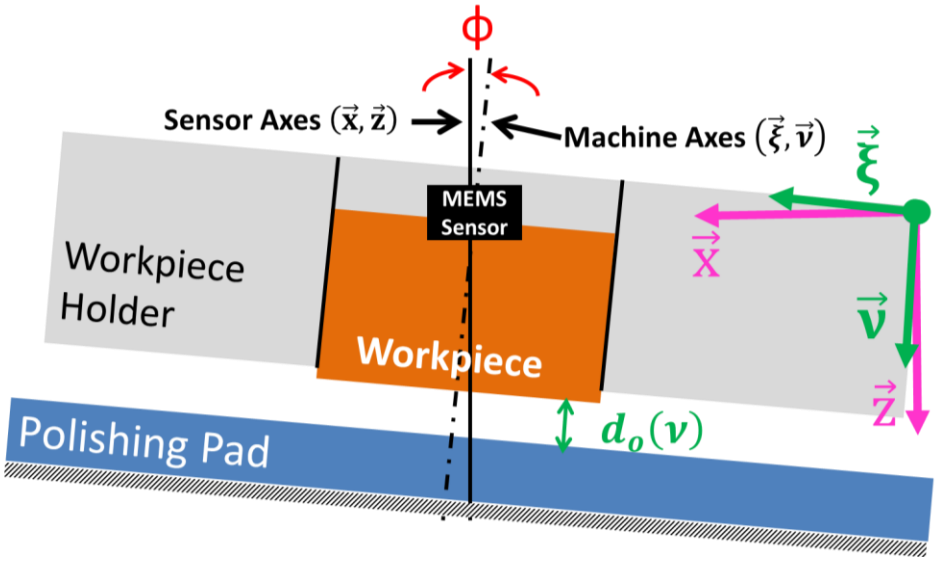


Fig. 22: Schematic representation of the CMP setup, showing an infinitesimal misalignment ϕ in the sensor and machine axes.

In Chapter 3, we used wireless MEMS vibration sensors (accelerometers) mounted in close proximity to the workpiece (cylindrical copper discs) to capture the CMP process dynamics. A schematic representation of the setup is shown in Fig. 22 above (see also Figure 3-3(d)). The sensors acquire vibration patterns along their tangential and normal axes denoted as $\vec{x}$ and $\vec{z}$, respectively. Under ideal circumstances, the sensor axes $(\vec{x}, \vec{z})$ can be assumed to align along the machine axes $(\vec{\xi}, \vec{\nu})$.

However, in practice it is often observed that the sensor axes $(\vec{x}, \vec{z})$ may not be precisely aligned with the machine axes $(\vec{\xi}, \vec{\nu})$. Due to this misalignment in the sensor and machine axes (viz. denoted as ϕ in Fig. 22), the sensor signal patterns acquired along either of the directions $\vec{x}$ and $\vec{z}$ captures the process dynamics from

both the normal and tangential directions with respect to the machine coordinates $(\vec{\xi}, \vec{v})$.

Therefore, in formulating the CMP process dynamics we assume an infinitesimal misalignment ϕ between the sensor and machine axes as shown in Fig. 22. Also in Fig. 22 the mean separation distance between the wafer and pad is represented as $d_0(v)$ (see Ref. [145, 149, 231]). The significance of the dynamic term $d_0(v)$ will become clear as we proceed through the PMI model formulation.

Physically, the effect of misalignment can also be elucidated from the perspective of mode-coupling phenomena encountered in mechanical processes where sliding friction (e.g., CMP) plays a prominent part (see Ref. [232]). Pertinently, the dynamics of sliding friction processes are often captured using 2 degree of freedom systems in the literature [232].

In this work, the overall dynamics of the CMP process dynamics is captured using a 2 degree of freedom, nonlinear differential equation lumped-mass model of the form,

$$\mathrm{m}'\ddot{x} + \mathrm{k}[x - d(x)] + \mathrm{c}\dot{x} = \mathcal{F}_{\mathrm{X}}(x) \tag{1}$$

where, $\ddot{x}$ is the acceleration (vibration), $\dot{x}$ is the velocity, and x the displacement along the direction $\vec{\mathrm{x}}$, and the RHS term, $\mathcal{F}_{\mathrm{X}}(x)$ is the tangential force acting along the axis $\vec{\mathrm{x}}$ (see Fig. 22). In the forthcoming, we will explain how the various terms in Eqn. (1) relate to the CMP process dynamics.

Formulation of CMP process dynamics as a 2 degree of freedom lumped-mass system

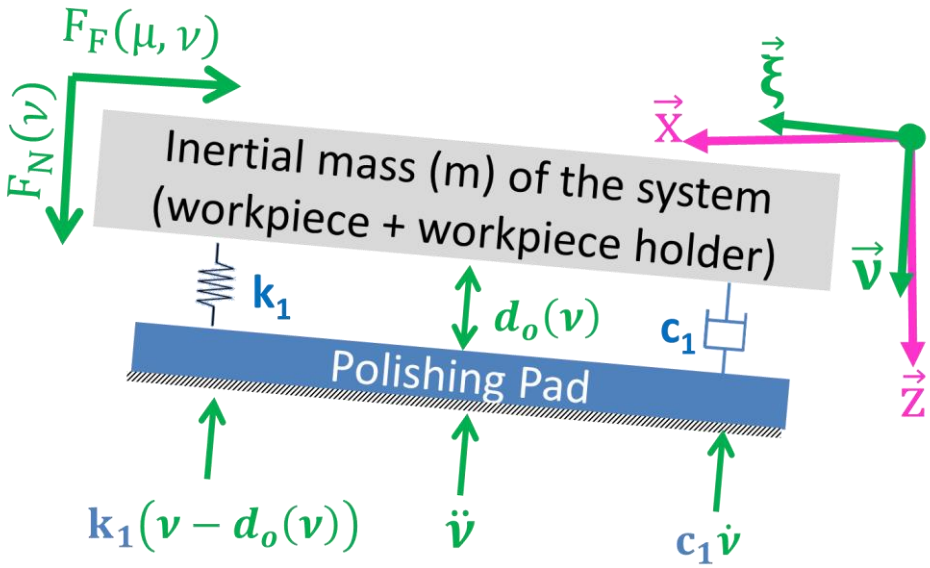


Fig. 23: Normal $F_N(\nu)$ and friction forces $F_F(\mu, \nu)$ in terms of the machine coordinates $(\vec{\xi}, \vec{\nu})$.

We first formulate the process dynamics along the vertical direction $\vec{\nu}$ of the machine axes $(\vec{\xi}, \vec{\nu})$. Using a lumped-mass model as shown in Fig. 23, the normal $F_N(\nu)$ and friction force $F_F(\mu, \nu)$acting along the machine axis $\vec{\nu}$ and $\vec{\xi}$, respectively, are written as (referring to Fig. 23),

$$F_N(\nu) = m\ddot{\nu} + k_1[\nu - d_o(\nu)] + c_1\dot{\nu}, \tag{2}$$

$$F_F(\mu, \nu) = \mu F_N(\nu) \tag{3}$$

In Eqn. (2), ν is the displacement, $\dot{\nu}$ is the velocity, and $\ddot{\nu}$ the acceleration along the vertical machine axis $\vec{\nu}$. Additionally, c_1 and k_1 represent the damping and stiffness coefficients, respectively, and m is the inertial mass of the system.

In Eqn.(3), μ is the kinetic coefficient of friction at the wafer-pad interface. Essentially, $F_F(\mu, \nu)$ is a scaled form of $F_N(\nu)$ contingent on μ. Such a relation-

ship connecting $F_N(v)$ and $F_F(\mu, v)$ is often considered in literature for sliding friction systems (see e.g., Hoffmann, *et al.* [232]).

Next, the forces along the sensor axes $(\vec{x}, \vec{z})$ are formulated similarly using a lumped-mass system representation. Accordingly, the force F_X along the sensor axis $\vec{x}$, can be written as (see Fig. 24),

$$F_X(x) = m\ddot{x} + k_2[x - d(x)] + c_2\dot{x} \tag{4}$$

where, $x = v \sin\phi + \xi \cos\phi$, and $d(x) = d_o(v) \sin\phi$.

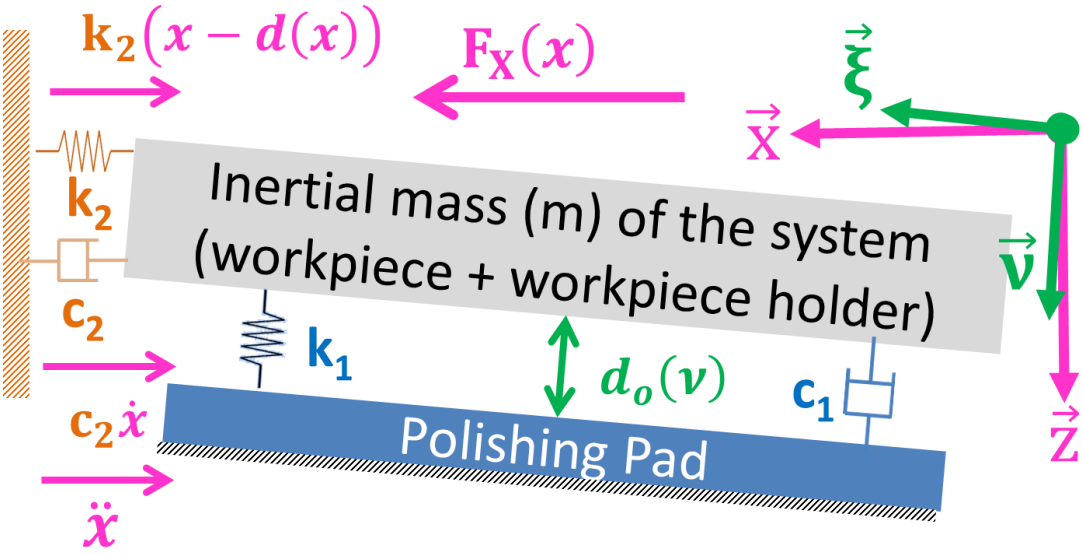


Fig. 24: Tangential force $F_X(x)$ represented in terms of the sensor coordinates $(\vec{x}, \vec{z})$.

From the above equation, it is apparent that

$$F_X(x) = m\ddot{x} + k_2[v \sin\phi + \xi \cos\phi - d_o(v) \sin\phi] + c_2\dot{x} \tag{5}$$

Subsequently, resolving $F_X(x)$ in the direction of the machine coordinates $(\vec{\xi}, \vec{v})$, $F_X(x)$ can be expressed in terms of as $F_N(v)$ and $F_F(\mu, v)$ as,

$$F_X(x) = F_F(\mu, v) \cos\phi + F_N(v) \sin\phi. \tag{6}$$

Using the above method, $F_Z(z)$ is similarly expressed as,

$$F_Z(z) = F_N(v) \cos\phi - F_F(\mu, v) \sin\phi. \tag{7}$$

Essentially, Eqn. (6) and (7) imply that the forces $F_X(x)$ and $F_z(z)$ along the sensor axes $(\vec{x}, \vec{z})$ are closely coupled with the normal $F_N(\nu)$ and friction $F_F(\mu, \nu)$ forces. On further simplification Eqn. (6) and (7) reduce to,

$$F_X(x) = F_N(\nu) \cdot [\sin\phi + \mu\cos\phi] \tag{8}$$

$$F_Z(z) = F_N(\nu) \cdot [\cos\phi - \mu\sin\phi] \tag{9}$$

From, Eqn. (5) and (8), we can write $F_X(x)$ as,

$$\frac{F_X(x)}{[\sin\phi + \mu\cos\phi]} = F_N(\nu) \tag{10}$$

$$\frac{m\ddot{x} + k_2[\nu\sin\phi + \xi\cos\phi - d_o(\nu)\sin\phi] + c_2\dot{x}}{[\sin\phi + \mu\cos\phi]} = F_N(\nu) \tag{11}$$

Treating ϕ and μ as constants, the denominator term in the above two equations can be set equal to K. Further recalling, $d(x) = d_o(\nu)\sin\phi$ and $x = \nu\sin\phi + \xi\cos\phi$, we can reconcile the LHS of (10) and (11) as,

$$\frac{F_X(x)}{K} = \frac{m\ddot{x} + k_2[x - d(x)] + c_2\dot{x}}{K} \tag{12}$$

on letting,

$$m' = \frac{m}{K},\ k = \frac{k_2}{K},\ c = \frac{c_2}{K},\ \text{and}\ \mathcal{F}_X(x) = \frac{F_X(x)}{K} \tag{13}$$

and thereby noting that $\mathcal{F}_X(x)$ is essentially a scaled form of $F_X(x)$ acting along $\vec{x}$. The CMP process dynamics can be captured using a 2 degree of freedom, nonlinear differential equation lumped-mass model of the form,

$$m'\ddot{x} + k[x - d(x)] + c\dot{x} = \mathcal{F}_X(x).$$

Where, $\ddot{x}$ is the vibration, $\dot{x}$ is the velocity, and x the displacement direction x. Also, c (= 2 N-s/m) and k (= 1000 N/m) represent the damping and stiffness

coefficients of the lumped mass system, with m' (=30 N) treated as the inertial mass of the system. With the RHS term $\mathcal{F}_X(x)$ in the above equation representing the tangential force along the axis $\vec{x}$.

From Eqn. (8), it is evident that the normal force $F_N(\nu)$ at the wafer-pad interface can be treated as the excitation force to drive the structure (Fig. 22). The essence of the subsequent model development is to delineate the multi-scale aspects of $F_N(\nu)$.

The normal force $F_N(\nu)$ is a nonlinear function of ν, as evident from Eqn. (16) through (19). Furthermore, $F_N(\nu)$ is expressed as (see Fig. 25),

$$F_N(\nu) = P_0 - \bar{p}(\nu) \cdot A_*(\nu) \tag{14}$$

where P_0 is the applied downforce, $\bar{p}(\nu)$ the effective load at the wafer-pad interface, and $A_*(\nu)$ the effective contact area at the wafer-pad interface [149, 158]. The dynamic RHS terms $\bar{p}(\nu)$ and $A_*(\nu)$ are obtained from Eqn. (16) and (17), respectively.

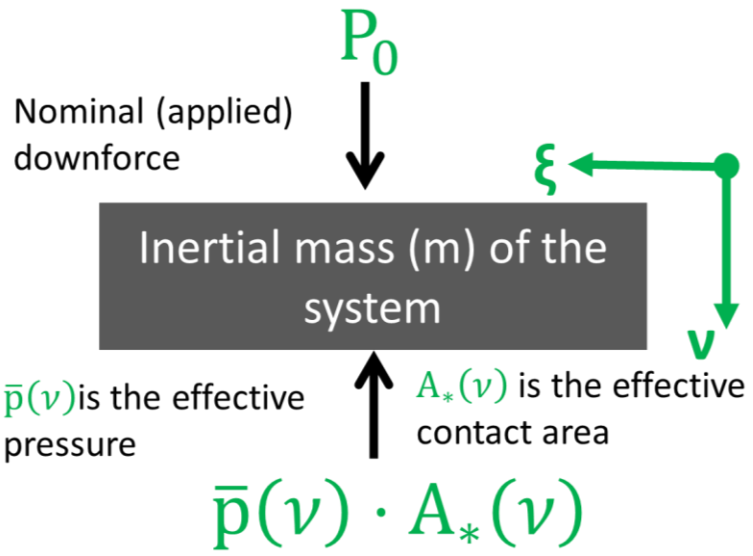


Fig. 25: Free body diagram showing the effective $\bar{p}(\nu)$ and nominal P_0 loads.

The expressions for $\bar{p}(\nu)$, and $A_*(\nu)$, are obtained based on consideration of process dynamics at the following scales:

- *Wafer-pad asperity interface mechanics:* The polishing action at wafer-pad interface is formulated according to Borucki's model [149], which is based on the Greenwood and Williamson approach [158]. Unlike previous approaches, we capture the effects of vibrations and bulk pad behavior on the wafer-pad separation distance.
- *Bulk pad structural dynamics:* During polishing the bulk structure of the pad is considered to cyclically compress and relax in response to the dynamic load $\bar{p}(\nu)$. This aspect affects the mean separation between the wafer and pad, as well as the effective structure stiffness.
- *Machine level kinematic effects:* Effects of geometrical inaccuracies of machine elements and eccentricity in spindle motion introduce low fre-

quency cyclic displacements that are assumed to independently superimpose on the vibrations generated due to wafer-pad interface and pad structure effects.

Wafer-pad asperity interface effects

The Greenwood-Williamson (GW) approach [158] is used to compute the real contact area and pressure in wafer-pad asperity models by Luo, *et al.* [138, 144, 148], Qin, *et al.* [150], and Borucki [149]. The model by Borucki [149] is used in this work, since it provides a closed-form solution for obtaining the mean-separation distance between wafer and pad. Essentially the GW model suggests a probabilistic distribution of pad asperities primarily to determine the area of contact at the interface. Most abrasive finishing processes consider such an approach to model uncertainty in the nature of contact [3, 140, 233].

In addition to the assumptions made by the GW model [158], we impose the following conditions:

i. The pad asperity probability density function $\psi(z)$ is stationary, and does not change with time – an assumption valid for at most a minute of polishing as shown by Borucki [149].
ii. The dynamics effects of slurry particles and hydrodynamic pressure distribution due to the slurry film is negligible.
iii. Pad deformations are purely elastic. The various constants used in the PMI model are identical to those used by Borucki [149], these are listed in T 5 below.

T 5:CMP process constants used in the PMI model

Constant	Value
Asperity density η_s	2 x 10^8 /m^2
Curvature of asperities κ_s	2 x 10^4 m^{-1}
2D Young's modulus of pad E_*	119 MPa
Asperity height distribution $\psi(z)$	μ_z = 0, σ_z = 15 x 10^{-6} m
Coefficient of sliding (dynamic) friction μ	0.75

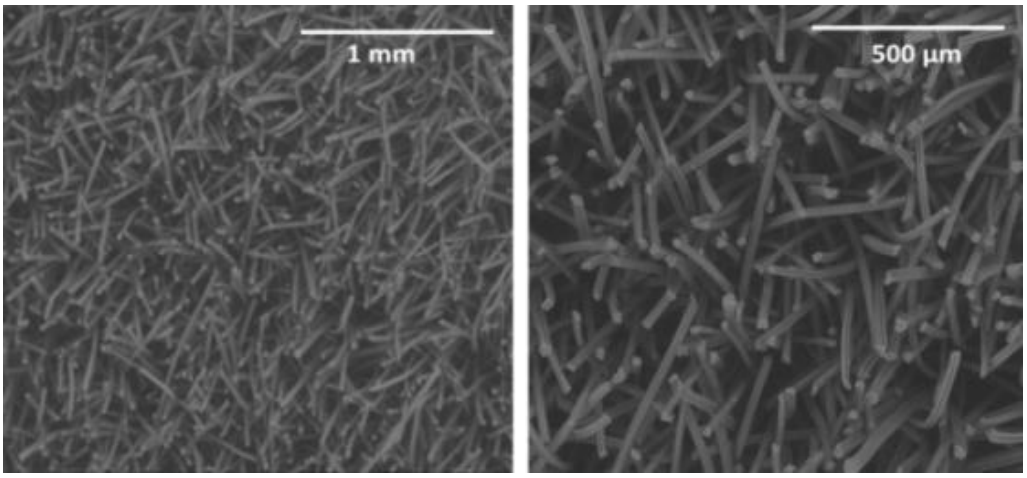


Fig. 26: SEM micrographs of the Buehler MicroCloth polishing pad.

Images are at different magnifications (left: 100X, right: 200X) depicting a dense collection of fibrous asperities (courtesy of Buehler).

The surface of a CMP pad may be visualized as being composed of a relatively dense network of fibers (Fig. 26) forming asperities. Considering the asperity heights to be Gaussian distributed with probability density function $\psi(z)$ (Fig. 27), the probability that an asperity of height z makes contact with wafer surface at a distance ν, where $(\nu > z)$, from the nominal plane (see Fig. 27) of the pad is given by,

$$\mathrm{P}(\nu > z) = \int_{\nu}^{\infty} \psi(z) \cdot \mathrm{d}z \tag{15}$$

using Hertzian contact formulation [158], the total expected effective load $\bar{\mathrm{p}}(\nu)$ over the wafer is given by,

$$\bar{p}(\nu) = \frac{4}{3} \cdot \frac{\eta_s \cdot A_{nom} \cdot E_*(\nu)}{\kappa_s^{1/2}} \int_{\nu}^{\infty} (z-\nu)^{3/2} \cdot \psi(z) \cdot dz \qquad (16)$$

where the effective (real) contact area $A_*(\nu)$ between the wafer and pad may be modeled as,

$$A_*(\nu) = A_{nom} \cdot \eta_s \int_{\nu}^{\infty} \frac{z-\nu}{\kappa_s} \cdot \psi(z) \cdot dz \qquad (17)$$

where $E_*(\nu)$ is the effective stiffness determined by the bulk structural properties of the pad, A_{nom} (= 13 x 10^{-4} m^2 for dia. 40.625 mm wafer used in our experiments) is the nominal or the projected area of contact between the wafer, κ_s (= 2 x 10^8 m^{-1}) is the curvature of the pad asperities, and η_s (= 2 x 10^8 /m^2) is the asperity density.

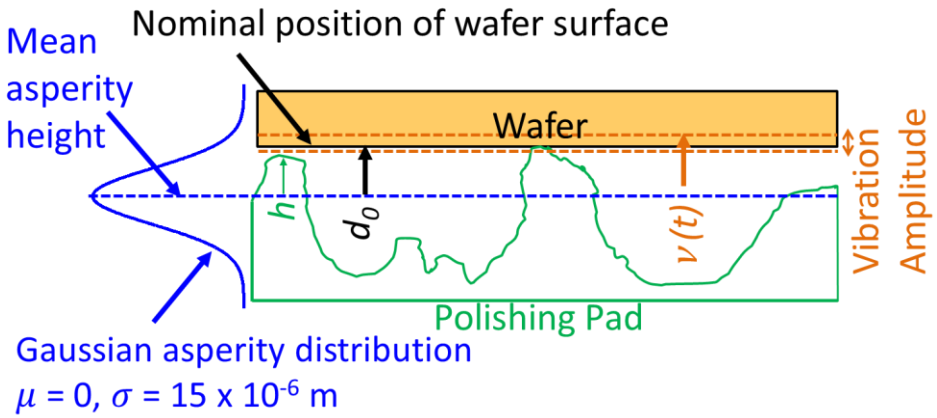


Fig. 27: The Greenwood-Williamson (GW) model applied to CMP.

GW model considers the polishing pad with Gaussian distributed asperities and the wafer to be relatively smooth.

Bulk pad structure effects

The polishing pad (Buehler MicroCloth) used in this work has a porous rayon top layer backed by a pressure sensitive adhesive layer (~ 2 mm thick). With respect to the CMP pads used in industry, such as SUBA and IC series [155], the

pad used in our study resembles the latter, since both MicroCloth and IC series have prominent asperities and are relatively rigid [149].

The MicroCloth pad is adhesively fixed to a stainless steel bimetallic plate, which in turn is held (magnetically) on the polishing platen. The polishing pad system, comprising of the MicroCloth pad along with the bimetallic plate is considered as contributing to the overall bulk pad dynamics. We found that applying the IC 1000 pad parameters [149] gives comparatively closer agreement with experimental data.

As the wafer is held down on the pad with an applied downforce P_0, the pad and wafer separation distance (ignoring viscoelastic behavior of the pad) is assumed to settle to a nominal distance $d_0(v)$ given by [149],

$$d_0(v) = \arg\min_{d \in v} \left| P_0 - \int_v^{\infty} (z-d)^{3/2} \cdot \psi(z) \cdot dz \right| \forall\, z, (z > v) \qquad (18)$$

The values of $d_0(v)$ varied in the range of 35 µm – 38 µm as P_0 is decreased from 8 lb. (35.6 N) to 4 lb. (17.8 N). These values are consistent with literature [231].

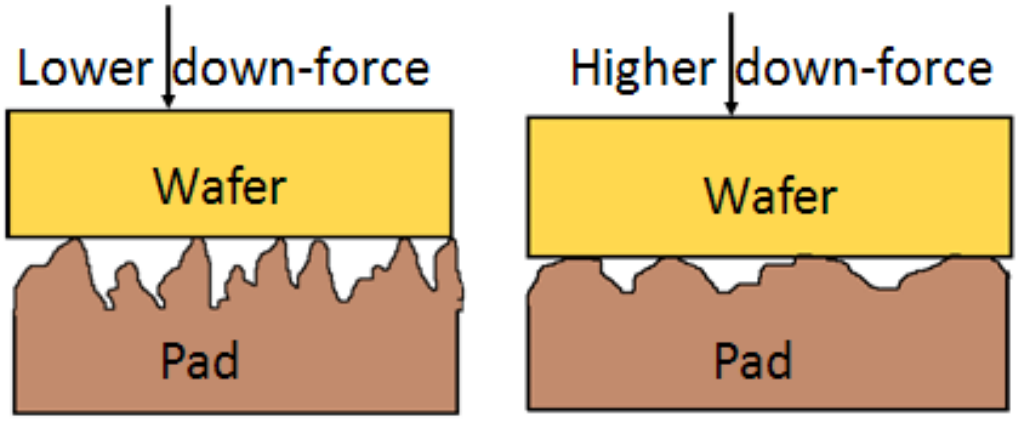


Fig. 28: Behavior of the bulk pad material at different load conditions.

As the downforce P_0 increases, the asperities compress (Fig. 28), and $E_*(\nu)$ – the effective pad stiffness increases. The pad structure parameters chosen were consistent with those of porous pads, such as MicroCloth and IC 1000. We use two sigmoid functions [94] to represent the nonlinear variation of $E_*(\nu)$ with displacement and also account for pad heterogeneity [154]. This behavior is consistent with observed polishing pad behavior documented by Bastawros, *et al.* [145]. At high P_0, an increase in $E_*(\nu)$ leads to increase in $\bar{p}(\nu)$ and reduced $d_0(\nu)$ (Eqn. (16) and (18)). With the asperities compressed, the effective contact area between the pad and wafer $A_*(\nu)$ also increases.

$$E_*(\nu) = \begin{cases} 1800 \frac{\text{kN}}{\text{m}} \ \forall \ \nu < \sigma_z \\ 1000 \frac{\text{kN}}{\text{m}} \forall \ \nu \geq \sigma_z \end{cases} \tag{19}$$

where σ_z is the standard deviation of the pad asperity distribution $\psi(z)$.

Machine kinematic effects

As shown in Fig. 29, a small angular error θ in the spindle axis translates into additional periodic displacement

$$x_m = 2\text{r} \cdot \tan(\theta) \cdot \sin\left(\frac{2\pi \text{N} t}{60}\right) \tag{20}$$

where r is the distance between the center of the wafer to the spindle axis, and N is the spindle speed.

Since the time and length scales over which the effects due to spindle error occur are much larger than those of the asperity-pad induced vibrations, we ignore

the effect of small time-scale asperity and pad-induced vibrations on the spindle error-induced displacements x_m.

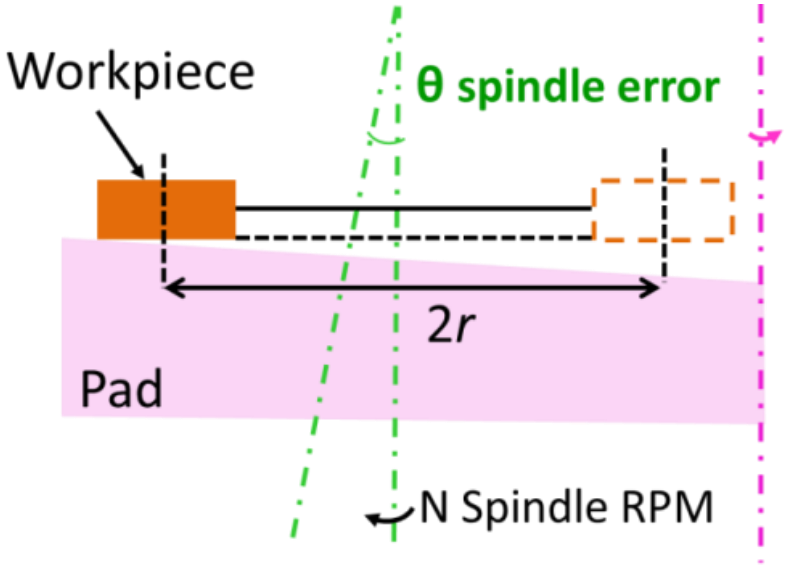


Fig. 29: Component representing spindle inaccuracy.

Consequently, the spindle motion introduced from kinematic effects $\ddot{x}_m$ is formulated as,

$$\ddot{x}_m + \left(\frac{2\pi}{t}\right)^2 x_m = 0 \tag{21}$$

In Eqn. (20) an inaccuracy (θ = 0.018°) in the vertical plane was determined through extensive experimental studies [160] to be inherent to the system, leading to the low frequency pattern ($\ddot{x}_m$) discussed in Sec. 3.1, represented in Eqn. (21).

Since the time and length scales over which the effects due to spindle error occur are much larger than those of the asperity-pad induced vibrations, we ignore the effect of small time-scale asperity and pad-induced vibrations on the spindle error-induced displacements x_m.

Appendix IV: Data Management

Since this research is funded via several NSF grants (CMMI 0700680, CMMI 0729552, CMMI 0927557, CMMI 1000978), a plan for data management was carefully devised for wide dissemination purposes. We now present a brief description of the data stored, as well as, the organization of the data files, images, and codes for wider dissemination[24].

Data from the UPM Process

Type of data acquired

- Piezoelectric force, vibration, and acoustic emission (AE) data are acquired in the UPM process.
- Cutting force and vibration data are collected along three axes, while the AE data is along one axis (see Figure 2-5); data from each type of sensor is stored in a separate folder.

Format of data and description of content

- The vibration and cutting force data is available in several ASCII files numbered sequentially (the data is divided into multiple files in order to avoid one unwieldy large file).

[24] The National Science Foundation (NSF) mandates that proposals submitted for funding consideration on or after January 18, 2011 must include a *supplementary document of no more than two pages describing a Data Management Plan* (see www.nsf.gov/eng/general/dmp.jsp).

- The files containing the cutting force and vibration data are preceded by 12 header lines. These are lines of text describing the acquisition date and time.
- Data is preserved in four columns; the first of four columns contains the time stamp at which a particular instance of data was acquired. The next three columns contain the vibration (or force) output corresponding feed, tangential and depth of cut directions, respectively.
- The AE data differs from the force and vibration data in only one aspect; the AE data has only two columns – the time stamp, and output (in mV).

Data from the CMP Process

Type of data acquired

- Wireless MEMS vibration data are acquired for the CMP process (see Figure 3-3(b)). The vibration data is along three axes: tangential, radial, and vertical for most cases.
- In some cases vibration along only the tangential and radial direction is acquired (the vertical direction vibration was not found to be sensitive to process conditions).
- The data is acquired at ~ 675 – 680 Hz.

Format of data and description of content

- The data is stored in ASCII format. There is one file per experimental condition; the files do not have any header lines.

- The tangential direction vibration data is stored in the first column, the radial direction vibration in the second column, and vertical direction vibration in the third.

Profile and Surface Roughness Data

Type of data acquired

- Surface profiles are acquired in four ways, by taking a screen shot of the 3D surface profile, screenshot of the 2D surface, profile stored in an array, and roughness values (Ra, Rq, Rz, etc.) noted by hand on paper.

Format of data and description of content

- The screenshots are saved as uncompressed .TIFF metafiles. These are readily converted into grey scale arrays using popular applications, such as MATLAB.
- The profiles are stored as .SDF files (scientific data format), viz. a widely used format for storing multi-dimensional arrays.

Storage and Accessibility

The data totaling close to 150 GB along with computer code (several hundred MATLAB scripts) currently resides in two hard-drives of the Pentium IV PC used by this student. This PC (barring some peripheral equipment and video cards) was acquired by the PIs for research purposes.

A back-up copy of the data is stored in an external flash memory hard-drive. These are located at the Advanced Technology Research Center (ATRC) labs under the purview and discretion of the dissertation adviser (Dr. S.T.S. Bukkapatnam). Shown below in Fig. 30 is the high-level folder organization of the data in context of this dissertation.

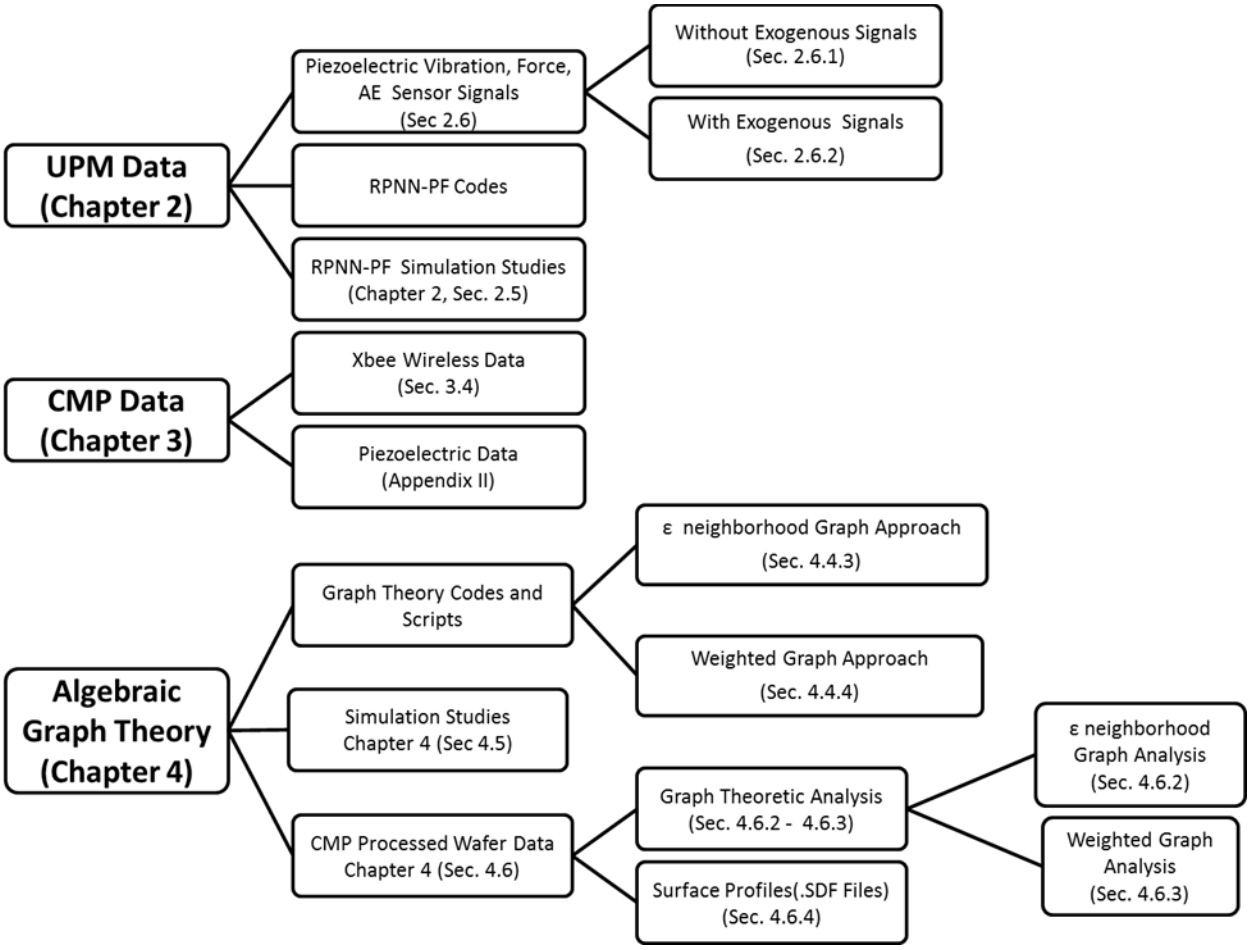


Fig. 30: High-level folder organization of the data used in this dissertation.

VITA

Prahalada Krishna Rao

Candidate for the Degree of

Doctor of Philosophy

Thesis: SENSOR-BASED MONITORING AND INSPECTION OF SURFACE MORPHOLOGY IN ULTRAPRECISION MANUFACTURING PROCESSES

Major Field: Industrial Engineering and Management

Biographical: Born on 5th August 1981, Nagpur, Maharashtra State, India

Education:
Completed the requirements for the Doctor of Philosophy in Industrial Engineering and Management at Oklahoma State University, Stillwater, Oklahoma in July, 2013.

Completed the requirements for the Master of Science in Industrial Engineering and Management at Oklahoma State University, Stillwater, Oklahoma in July, 2006.

Completed the requirements for the Bachelor of Engineering (First Class) in Production Engineering at Victoria Jubilee Technical Institute (VJTI), Bombay University, Mumbai, Maharashtra, India.

Experience:
1. Instructor, Oklahoma State University, IEM, 2012-2013
2. Instructor, Oklahoma State University, International Outreach, 2008-2013
3. Research Assistant, Oklahoma State University, 2008-2013
4. Teaching Assistant, Oklahoma State University, 2008-2011
5. Teaching Assistant, Pennsylvania State University, 2006-2008
6. Research Assistant, Oklahoma State University, 2004-2006
7. Trainee Process Engineer, Mahindra and Mahindra Automotives, 2003

Professional Memberships:
1. American Society for Quality (ASQ)
2. Institute for Operations Research and Management Science (INFORMS)
3. American Production and Inventory Control Society (APICS)
4. Institute of Industrial Engineers (IIE)
5. Alpha Pi Mu, Industrial Engineering Honor Society