Applications of raman spectroscopy for silicon stress characterization in integrated circuits

Colin Mcdonough
University at Albany, State University of New York, cm279188@albany.edu

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Nanoscience and Nanotechnology Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
APPLICATIONS OF RAMAN SPECTROSCOPY FOR SILICON STRESS CHARACTERIZATION IN INTEGRATED CIRCUITS

by

Colin J. McDonough

A Dissertation
Submitted to University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy

College of Nanoscale Science & Engineering
University at Albany
2011
APPLICATIONS OF RAMAN SPECTROSCOPY FOR SILICON STRESS CHARACTERIZATION IN INTEGRATED CIRCUITS

by

Colin J. McDonough

COPYRIGHT 2011
ABSTRACT

The introduction of mechanical stress in Si-based integrated circuits (ICs), whether desired or undesired, is intrinsic to IC fabrication. The origins are diverse and result from the numerous materials, geometries, and processes involved in fabrication. These stresses can lead to such effects as delamination, void formation and migration, and fracture, and can significantly affect device performance. As a result, stress development is a major concern for reliability, process control, and device design. It is necessary to investigate and characterize the origins and levels of the induced stresses. A more complete fundamental understanding of the evolution of stress in ICs and novel ways in which it can be characterized can lead to more effective strategies to mitigate or control stress development.

Equivalent scaling strategies such as strain-engineered MOSFET channels and 3D integration schemes are important for maintaining IC performance increases at current and future semiconductor technology nodes. Strained Si can affect carrier mobility, either negatively or positively, depending on direction and magnitude of strain and the majority carrier type. This is of benefit in strain engineering of the MOSFET channel region. However, in 3D architectures, the integration of Cu TSVs through the active region can induce thermo-mechanical residual stresses in the nearby Si, which could lead to undesirable performance variations. Therefore, it is important to have instrumentation available for characterizing the stresses in the MOSFET channel regions as well as in Si regions surrounding TSV structures in 3D-ICs.
Raman spectroscopy is an all-optical technique that is applicable for measuring stress in Si based on changes in the crystalline vibrational modes. Part of this dissertation covers work completed in evaluation of so-called tip-enhanced Raman spectroscopy (TERS) for sub-diffraction limited, localized stress characterization for FEOL applications. This includes characterization of apertureless near-field scanning optical microscope (NSOM) probe enhancement as well as spatial resolution and stress characterization on strained SOI mesa test structures. Additionally, this dissertation covers research in Si thermo-mechanical stress evolution near Cu TSVs for 3D-ICs. Residual stress measurements in Si were conducted using micro-Raman spectroscopy. These results were used to examine the evolution based on ex-situ annealing conditions, processing steps, and depth dependence. Correlation with AFM topography and FIB cross-sections is included.
ACKNOWLEDGEMENTS

It is with the most humble and sincere gratitude I acknowledge those that have contributed to the success of this work through constant support and continued assistance, both professionally and personally. Without their help this work would not have been possible.

First and foremost, I thank Dr. Robert Geer, my research advisor and committee chairman. Throughout my tenure as a doctoral candidate he has provided unending support, for which I am most grateful. Through thick and through thin, he has afforded me numerous opportunities to broaden my experiences by allowing me to work on various research projects, collaborate with colleagues in academia and industry, and present my work to key industry and academic audiences. These opportunities have contributed, in no small part, to my growth as a researcher and as a professional. It is only with his guidance, as mentor and colleague, that this work was made possible. I cannot thank him enough.

I would also like to thank my research committee members Dr. Alain Diebold, Dr. James Lloyd, Dr. Larry Smith, and Dr. Wei Wang for your support, encouragement, and sharing your vast knowledge of all things related to the semiconductor industry realm. Your help has contributed to the success of this work. Special thanks goes to Dr. Larry Smith for providing insight into possible experiments for understanding thermal evolution of stresses near Cu TSVs, as well as for providing the wafers studied in this dissertation.
It is with the utmost appreciation I acknowledge those that have provided direct support to my research. For all the assistance with installation, training, modification, alignment, and all other related hardware and software problems or questions, as well as willingness to answer and discuss other issues regarding our two spectrometers, I thank Dr. Richard Bormett of Renishaw and Dr. Fran Adar and Dr. Eunah Lee of HORIBA Jobin-Yvon. I acknowledge Dr. Hesham Taha of Nanonics Imaging, Ltd. for installation, training, and unending assistance and conversations about the MV2000 NSOM instrument used for TERS measurements. I especially acknowledge the lab support and assistance provided by Mr. Stephen Stewart of the College of Nanoscale Science & Engineering for instrument installations and modifications. Thanks also go to Dr. Thomas Murray, who provided countless hours of his time to assist with FIB cross-sections of TSV structures.

I am indebted to our collaborators, first of whom I acknowledge are Dr. Ehrenfried Zschech, Dr. Jochen Rinderknecht, and Dr. Michael Hecker for allowing me the opportunity to work alongside them at AMD Dresden, including their support and instruction, access to their labs, and useful discussions regarding Si stress measurements with TERS. Many thanks also go to Dr. Pulickel Ajayan at Rice University for allowing me unfettered access to his laboratory resources and for allowing me the opportunity to work alongside his research team, especially Dr. Shaijumon Mannikoth, Dr. Fung Suong Ou, and Mr. Sanketh Gowda, to develop and fabricate the Au-CNT hybrid nanowires used in this dissertation (Appendix A).

My acknowledgements would not be complete without thanking my classmates and colleagues, all of whom have provided me with immeasurable support in my graduate
studies and research: Mr. Jacob Atesang, who initially worked on TERS for Si stress characterization, as well as introducing me to graduate research; Ms. Gayathri Rao, my fellow Raman spectroscopy researcher, with whom I could discuss anything related to research or life; Mr. Yunfei Wang, who provided AFM assistance even though he had his plate full; Ms. Lay Wai Kong, who provided much food for thought regarding stress in Cu TSVs, as well as using her time to provide X-ray tomography maps of several of my annealed Cu TSV structures; and Mr. Benjamin Backes, who took the TSV FE modeling to a new level and provided me with any data requested, despite his busy schedule.

Many others deserve thanks, but to list them might well rival the length of this document. Therefore, I will forego such an endeavor and simply thank my close friends that supported me along the way: Ms. Mayrita Arrandale, Mr. Matthew Bresin, Ms. Jihan Capulong, Mr. Mark Knight, Mr. Christopher Miller, Mr. Viet Nguyen, and Dr. Phillip Rogers. Without their support and friendship this work would have been a great deal harder to accomplish.

Acknowledgements go to the Semiconductor Research Corporation and International SEMATECH for providing funding for this research. It goes without saying this work was only made possible with their support.

Final thanks go to my family, especially my parents. Their support and words of motivation kept me going this entire time. Words cannot express my gratitude.
# TABLE OF CONTENTS

Abstract..........................................................................................................................iii

Acknowledgements........................................................................................................v

PART ONE: INTRODUCTION AND MOTIVATION FOR RAMAN STRESS CHARACTERIZATION IN SILICON INTEGRATED CIRCUITS......................... 1

1.1. INTRODUCTION TO STRESS IN INTEGRATED CIRCUITS.................... 1
   1.1.1. Strain Engineering for MOSFET Carrier Enhancement.................... 2
   1.1.2. Three-Dimensional Silicon Integration and Through-Silicon-Via Induced Stress.............................................................. 4
   1.1.3. Motivation for Work Accomplished.............................................. 8

1.2. STRESS CHARACTERIZATION FOR SI INTEGRATED CIRCUITS: SURVEY OF TECHNIQUES FOR STRESS CHARACTERIZATION AND DETAILED EXPLANATION OF RAMAN SPECTROSCOPY.................... 12
   1.2.1. Introduction........................................................................................... 12
   1.2.2. Raman Spectroscopy........................................................................... 15
       1.2.2.1. Brief History of Raman Spectroscopy........................................ 16
       1.2.2.2. Overview of the Technique......................................................... 17
       1.2.2.3. Raman Effect in Single Crystals............................................... 19
       1.2.2.4. Silicon Stress Effects on Raman Shift....................................... 23
       1.2.2.5. Operational Aspects and Limitations of Raman Spectroscopy..... 28

PART TWO: TIP-ENHANCED RAMAN SPECTROSCOPY FOR SUBDIFFRACTION LIMITED RESOLUTION STRESS CHARACTERIZATION....30

2.1. INTRODUCTION AND BACKGROUND..................................................... 30
   2.1.1. Overcoming the Optical Diffraction Limit....................................... 30
   2.1.2. Aperture Near-field Scanning Optical Microscopy.......................... 32
   2.1.3. Apertureless NSOM and Tip-Enhanced Raman Spectroscopy......... 36
   2.1.4. Potential Advantages of Apertureless NSOM for FEOL TERS Stress Characterization............................................................ 40

2.2. TIP ENHANCED RAMAN SPECTROSCOPY EXPERIMENTAL SETUP... 42
   2.2.1. Experimental Objective................................................................. 42
   2.2.2. Instrumentation.................................................................................. 42
   2.2.3. Strained Silicon-on-Insulator Samples........................................... 45
   2.2.4. Finite Element Model for Stress Analysis in Mesa Structure......... 48

2.3. TERS STRESS RESULTS ON A STRAINED SILICON-ON-INSULATOR MESA............................................................. 50
   2.3.1. Characterization of Tip Enhancement.............................................. 50
   2.3.2. TERS on a Patterned Strained Silicon-On-Insulator Wafer............ 52

2.4. DISCUSSION OF TERS RESULTS......................................................... 58
   2.4.1. Effective Resolution of the TERS Probe....................................... 58
   2.4.2. Stress Measurements Compared with Finite Element Simulation.... 61

2.5. CONCLUSIONS......................................................................................... 63
PART THREE: SILICON STRESS EVOLUTION NEAR A THROUGH-SILICON-VIA USING MICRO-RAMAN MICROSCOPY

3.1. INTRODUCTION TO MECHANICAL STRESSES NEAR COPPER THROUGH-SILICON-VIAS ...........................................................................................................67

3.1.1. Thermally Induced Stresses in the Si/Cu-TSV System...........................................................................................................68

3.1.2. Plastic Deformation......................................................................................................................................................................70

3.1.3. Residual Stresses in Si Near Cu TSVs............................................................................................................................................74

3.2. EXPERIMENTAL APPROACH AND SETUP .................................................................................................................................78

3.2.1. Experimental Objective and Approach........................................................................................................................................78

3.2.2. Sample Details with Relevant Process History..............................................................................................................................79

3.2.3. Micro-Raman Scanning Details and Calibration.........................................................................................................................81

3.2.4. Micro-Raman Spectrometer Setup................................................................................................................................................83

3.2.5. Thermal Anneal Setup and In-Situ Raman Measurements...........................................................................................................84

3.2.6. Raman Depth Profile Measurements............................................................................................................................................88

3.3. RESULTS.................................................................................................................................................................................................90

3.3.1. Raman Characterization of TSV-induced Stress in Si: Geometric Effects..........................................................................................90

3.3.2. Raman Measurements on As Received Wafer A TSVs.........................................................................................................................93

3.3.3. Wafer A Raman Measurements on Annealed TSVs........................................................................................................................107

3.3.4. Raman Measurements on As Received Wafer B TSVs.........................................................................................................................117

3.3.5. Raman Measurements on Wafer C, Post-CMP..............................................................................................................................127

3.3.6. Raman Depth Profiling on As Received Wafers A & C.........................................................................................................................132

3.4. THERMO-MECHANICAL FINITE ELEMENT MODELING QUALIFICATION OF RAMAN MEASUREMENTS...........................................................................141

3.4.1. Finite Element Modeling of Cu TSVs in a Si Substrate.........................................................................................................................141

3.5. DISCUSSION OF RAMAN STRESS RESULTS ON TSV STRUCTURES.........................................................................................150

3.5.1. Origin of the Residual Stress Profile in Silicon..............................................................................................................................150

3.5.2. Discussion on Stress Superposition Effects.................................................................................................................................151

3.5.3. Stress Evolution Due to Cu Annealing..........................................................................................................................................153

3.5.3.1. Temperature Anneal Studies.....................................................................................................................................................153

3.5.3.2. Copper Self-Annealing..........................................................................................................................................................157

3.5.3.3. Process Induced Variation.....................................................................................................................................................158

3.5.4. Depth Profiling..................................................................................................................................................................................158

3.6. CONCLUSIONS......................................................................................................................................................................................163

PART FOUR: SUMMARY AND GENERAL REMARKS ..........................................................................................................................168

APPENDIX A: GOLD-CNT HYBRID NANOWIRE AS NOVEL TERS PROBE..............................................................................................................171

A.1. INTRODUCTION AND MOTIVATION................................................................................................................................................171

A.2. TEMPLATE FABRICATION OF HYBRID NANOWIRES.....................................................................................................................172

A.3. ATTACHMENT OF HYBRID NANOWIRES TO SHARPENED TUNGSTEN WIRES.............................................................................175

A.4. SUMMARY.............................................................................................................................................................................................179

BIBLIOGRAPHY........................................................................................................................................................................................181
PART ONE: INTRODUCTION TO RAMAN STRESS CHARACTERIZATION IN SILICON INTEGRATED CIRCUITS

1.1 INTRODUCTION AND MOTIVATION

Development of stresses in integrated circuits (ICs) is an unavoidable (although at times desirable) aspect of complementary metal-oxide-semiconductor (CMOS) processing. The diverse array of materials and their differing mechanical properties, as well as the finite CMOS geometries and deposition techniques used can lead to inherent stresses throughout an IC process flow, in both thin film structures as well as the substrate. The origins of stress related problems can be classified, non-exhaustively, into several groups [Hu, 1991]: thermally-induced stresses due to thermal expansion coefficient mismatches; intrinsic stresses due to film growth and film-edge induced stresses (e.g. SiO$_2$; SiN); localized stresses in embedded structural elements (e.g. metal lines; shallow trench isolation); stress due to thermal oxidation (e.g. viscoelastic flow of SiO$_2$ causes plastic deformation in substrate); and localized stresses due to material lattice mismatches (e.g. between epitaxial films and substrate; dopant incorporation). These are only general classifications for the purpose of understanding the diversity of stress origination in an IC process flow. As inferred from their sources, the problem of IC stresses will only become more acute with increasing device complexity (i.e. material sets and geometries). They can lead to such problems as crystalline defect formation, vacancy and void formation and migration (e.g. metal interconnects), delamination (e.g. thin films; metal lines), and fracture, to name a few. Any one of these could lead to device performance degradation and, in the worst case, complete failure. As a result, stress development is of great concern for IC reliability.
However, if carefully controlled, stress can be used to increase device performance. This will be discussed in the following section.

1.1.1 Strain Engineering for MOSFET Carrier Enhancement

One of the key technologies for continuation of metal-oxide-semiconductor field-effect transistor (MOSFET) scaling has been the introduction of strain (or alternately, stress) to the Si channel region to increase carrier mobility [M.L. Lee, 2005; Thompson, 2006; Dhar, 2007]. An example of a strained 67 nm channel MOSFET is displayed in Figure 1.1a. Biaxial and uniaxial strain along the channel plane has been shown to increase carrier mobility and is currently used in commercial devices. Take for example an n-type MOSFET where the majority carriers are electrons. In the presence of tensile strain, the degeneracy of the six-fold conduction band valleys (Figure 1.1b) is lifted so that the two out-of-plane valleys are lowered with respect to the 4 in-plane valleys. This leads to a decrease in intervalley phonon scattering. Furthermore, the out-of-plane valleys become preferentially filled with respect to the four in-plane valleys. The effective transverse electron mass is lowered in these valleys ($0.19m_e$ vs. $0.98m_e$), which leads to an increased in-plane electron mobility. Mobility enhancement due to intervalley phonon scattering suppression and reduced effective electron mass is found to be dependent upon the amount of stress (Figure 1.1c).
Introduction of tensile stress leads to breaking the six-fold degeneracy of the conduction band so that the two out-of-plane valleys become preferentially filled. Uniaxial stress applied along [110] leads to enhancement of in-plane NMOS mobility (red/blue) [DHAR, 2007]. The common methods of introducing strain are tensile and compressive SiN capping layers, epitaxial SiGe (tensile), tensile shallow trench isolation (i.e. compression on channel region), and strained silicon-on-insulator (tensile) [MIZUNO, 2000; THOMPSON, 2006; GÁMIZ, 2007]. All of these methods are scalable and compatible with current complementary MOS (CMOS) process techniques; therefore, strain is expected to remain a key component of current and future technology nodes [SVERDLOV, 2011; ITRS, 2009].
Whichever stress induction method is employed, it is highly desirable to be able to characterize the amount of stress in the channel. Therefore, it is important to have tools on hand that can measure and simulate such stresses to be able to pinpoint problematic design and processing issues. Front-end-of-line (FEOL) stress metrology can also be considered important for process control and characterization.

1.1.2 Three-Dimensional Silicon Integration and Through-Silicon-Via Induced Stress

Transistor scaling has long been the main performance driver for ICs. Performance enhancements were typically achieved by shrinking transistors, thus making them faster and consuming less power, and allowing for larger device densities, which increase computational power. Additional factors have augmented scaling to improve performance, reduce fabrication cost, or improve yield. Examples include introduction of larger wafer sizes and new materials, and more efficient processor design (e.g. multi-core). However, CPU clock speeds have essentially reached a limit over the past 5 years due to gate oxide scaling [KEAST, 2007]. Therefore, equivalent scaling has become a primary driver for continued performance increases. Equivalent scaling involves using new materials and methods to increase device performance without shrinking device sizes. For example, the use of metal gate/high-k dielectrics to replace SiO₂ allows for smaller gate oxide thicknesses and decreased (>100-fold [INTEL, 2007]) gate leakage current, which limits switching speed [CHAU, 2005]. Furthermore, the use of strained Si channels allows for increased carrier mobility, as discussed in the previous section. However, lithographic challenges for future technology nodes have become a major
hurdle. Implementation of extreme ultraviolet radiation (EUV) will help continue transistor size scaling, but is not yet ready for large-scale commercial use. Moreover, though it will allow for increased device density, current scaling issues will only be exacerbated at the smaller length scales, and bandwidth bottlenecks will continue to increase. Therefore, an alternative scaling paradigm that allows for equivalent scaling of current CMOS technology has emerged: three-dimensional integration (3D-i).

3D-i involves vertical integration to increase device density in a smaller form factor. This is achieved through a variety of advanced technologies, including, but not limited to: thinned Si wafers, through-silicon-vias (TSVs), and die-to-die (DtD) and die-to-wafer (DtW) bonding techniques (e.g. Cu-Cu bond pads; benzocyclobutene (BCB) adhesive bonding) [KNICKERBOCKER, 2008]. An illustration of such a 3D integrated system is seen in Figure 1.2.

![Diagram of 3D-i circuit](image)

Figure 1.2. Schematics illustrating a 3-tier stacked 3D-i circuit. Active device regions, TSVs, thinned Si substrates (1, 2, 3), and BCB regions are labeled. [With kind permission from Springer Science+Business Media: Three Dimensional System Integration; Chs.: TSV-Based 3D Integration, TSV Characterization and Modeling; 2011; pp. 22, 34; James Burns, Michelle Stucchi et al; Figs.: 2.9, 3.1.]
The benefits of 3D-i are not limited to increased device density. The use of TSVs, wafer thinning, and BCB bonding techniques allow for heterogeneous integration, one of the fundamental drivers for 3D-i [PAPANIKOLAOU, 2011; DE MUNCK, 2005]. Heterogeneous integration involves integrating multiple technologies and material sets on a single chip for increased functionality. For example, DtW integration of III-V photonic devices on SOI waveguides via BCB bonding has been demonstrated [STANKOVIĆ, 2010]. Furthermore, the density of TSVs can reach levels of 3-5 orders of magnitude higher than traditional off-chip wire bonding, which allows for large increases in bandwidth capability and decreased power consumption and signal delay [KNICKERBOCKER, 2008; PAPANIKOLAOU, 2011; CHEN, 2006; TU, 2008].

TSVs are one of the key interconnect technologies enabling 3D-i [KNICKERBOCKER, 2008]. The most prevalent metal used for TSVs is Cu. This choice is intuitive, considering its high conductivity, compatibility with current CMOS technology and processing, and the technological expertise developed since its introduction as the metal interconnect of choice in ICs. However, due to the large via sizes (currently 5×50 µm), there is the problem of thermally induced stresses in nearby active layer devices due to the large CTE mismatch between Cu and Si. Thermal cycling during the manufacturing process, with temperatures reaching up to 400 °C or higher, induce thermo-mechanical stresses in the nearby Si which can affect mobility of nearby devices. This would lead to non-uniform device performance, which would be a function of distance from Cu TSVs. Stress levels calculated via finite element analysis have shown stress levels can reach the order of several hundred MPa [KARMARKAR, 2009, 2010; SELVANAYAGAM, 2009a,b;
LU, 2010]. In this regime, the fractional change in carrier mobility, $\Delta \mu / \mu$, can be found by the piezoresistance constants of Si [THOMPSON, 2006]:

$$\frac{\Delta \mu}{\mu} = |\pi_{\parallel} \sigma_{\parallel} + \pi_{\perp} \sigma_{\perp}|$$  \hspace{1cm} (1.1)

where $\sigma_{\parallel}$ and $\sigma_{\perp}$ are the longitudinal and transverse stresses, and $\pi_{\parallel}$ and $\pi_{\perp}$ are the longitudinal and transverse piezoresistance coefficients expressed in $\text{Pa}^{-1}$. For Si, the corresponding coefficients are shown in Table 1.1. A 100 MPa stress, can induce up to 7% mobility change in Si. Therefore, it will be important to understand the thermally induced stress development throughout a 3D-i process flow to determine so-called Keep-Out-Zones that will determine minimum distances between TSVs and nearby MOSFET for maintaining acceptable mobility deviations.

<table>
<thead>
<tr>
<th></th>
<th>$\pi_{\parallel}$</th>
<th>$\pi_{\perp}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-MOSFET</td>
<td>$-3.16 \times 10^{-10}$</td>
<td>$-1.76 \times 10^{-10}$</td>
</tr>
<tr>
<td>p-MOSFET</td>
<td>$7.18 \times 10^{-10}$</td>
<td>$-6.63 \times 10^{-10}$</td>
</tr>
</tbody>
</table>

Table 1.1. Measured piezoresistance coefficients for NMOS and PMOS (001) Si with a [110] channel direction, as adapted from THOMPSON, et al. (2006).
1.1.3 Motivation for Work Accomplished

The overarching theme of this dissertation is Si stress characterization. As previously stated, stress development in ICs can occur at any stage of the manufacturing process from a variety of sources. Two critical areas of Si stress development in ICs are those that can affect MOSFET performance: 1) FEOL strain engineered Si channel for increasing carrier mobility and 2) Thermo-mechanical stress development near Cu TSVs for 3D integration. In both cases, stresses that develop in Si affect device performance; however, in the first case these stresses are desirable, whereas in the second case these stresses may be deleterious due to undesirable, non-uniform device performance variations. Therefore, characterization of Si stresses is important for understanding their evolution throughout the manufacturing process.

In order to sufficiently and effectively characterize stress generation (or relaxation) in the areas of interest, a non-destructive characterization technique is desirable.

FEOL metrology for Si stress characterization requires high-resolution due to the small channel dimensions (< 50 nm). As a result of this restriction, the current techniques of choice are transmission electron microscopy (TEM) based, as will be discussed in the following section. However, TEM based approaches are destructive in nature and require thin cross-sections, where possible stress relaxation in the sample may occur. Therefore, they often require additional analysis in the form of computer simulation to help take stress relaxation into account. Analysis time can be on the order of days or longer. A non-destructive approach would be desirable to mitigate the problems associated with the TEM based approaches. Such a non-destructive technique
may exist in the form of Raman spectroscopy. It is an all-optical characterization method that can measure Si phonon mode frequency shifts that occur in the presence of lattice strain (i.e. stress). Being an optical technique, resolution is limited by the diffraction limit. Therefore, to be able to measure stresses in channel regions with length scales below this limit, it can be paired with near-field optical microscopy (NSOM). This combined technique, called tip-enhanced Raman spectroscopy (TERS), relies on sub-wavelength light scattering from a small metallized tip to increase spatial resolution to the order of $< 10$ nm. Evaluation of TERS for high-resolution Si stress metrology will be the subject of Part Two of this dissertation.

For Si stress metrology associated with TSVs and 3D integration, the spatial length scales of interest are typically on the order of microns; therefore, nanometer resolution is not of critical importance. For example, stress variations in Si in the vicinity of Cu TSVs in a 3D-i manufacturing process extend over tens of microns. In this case, conventional micro-Raman spectroscopy can be employed for non-destructive Si stress characterization. To date, most analysis of thermo-mechanical stress development in and around Cu TSVs comes in the form of finite element (FE) modeling. Only a limited amount of physical characterization of Si stresses around Cu TSVs has been conducted. Therefore, Part Three of this dissertation focuses on micro-Raman spectroscopy analysis of thermally induced stresses in the Si surrounding Cu TSVs. This work involves measurements on various TSV arrays to understand the affect of stress superposition effects. It also involves measuring stresses on wafers that have undergone varying degrees of processing in a 3D-i process flow to gain insight into the evolution of stress as a result of process history. Measurements were conducted on samples that have
undergone various thermal anneal treatments to understand the temperature induced stress evolution. Raman measurements on annealed samples were then correlated with TSV topographies from atomic force microscope (AFM) measurements to gain insight into the Si stress evolution as a result of Cu protrusion. Moreover, cross-sections of the TSV arrays (one sample set) were prepared by focused-ion beam (FIB) for correlation of Si stress evolution with possible Cu grain growth and/or void formation. Finally, similar FIB cross-sections were prepared on samples from two different Cu TSV geometries to investigate the stress distribution along the depth of the TSV.

This dissertation is organized into parts, rather than chapters, for ease of navigation. The following section of Part One will provide a brief survey of common techniques used for Si stress characterization. Section 1.2 will also go into a detailed description of Raman spectroscopy and provide the theoretical basis for its application to Si stress characterization. This should provide the necessary understanding for the stress measurements conducted in the rest of this dissertation. Part Two encompasses evaluation of TERS for high-resolution Si stress characterization by using well-defined, patterned, strained silicon-on-insulator mesa structures. An introduction to NSOM and TERS will be given, followed by the experimental approach, results, and a discussion of the results. Part Three follows in a similar organization as Part Two, with the focus of this part being on the measurement of Si stresses surrounding Cu TSVs for various parameters, including TSV array size, process history, anneal conditions. An introduction will be given for stress development in Cu TSVs and the resultant residual stresses in the surrounding Si. Then the experimental approach will be given, followed by results, a short section that gives a qualitative comparison of the Raman measurements
with FE models, and finally a discussion of the results. Conclusions will be given in Parts Two and Three, respectively, with a general conclusion to the dissertation given in Part Four. Appendix A will outline NSOM probe development work done in conjunction with Part Two.
1.2 STRESS CHARACTERIZATION FOR SI INTEGRATED CIRCUITS: SURVEY OF TECHNIQUES FOR STRESS CHARACTERIZATION AND DETAILED EXPLANATION OF RAMAN SPECTROSCOPY

1.2.1 Introduction

There are several techniques currently used in the IC industry for measuring stress or strain in Si. The most common are x-ray diffraction (XRD), Raman spectroscopy, transmission electron microscopy (TEM) based diffraction techniques, such as convergent beam diffraction (CBED) and nano-beam diffraction (NBD), and TEM imaging techniques, such as high-resolution electron microscopy (HREM) and dark-field holography (HoloDark). Each has its own set of advantages and limitations. This is not meant to be a comprehensive review of the techniques, but rather a brief discussion of selected advantages and limitations of each.

XRD relies on the use of scattering of x-rays at specific angles, following Bragg’s law, to determine the atomic configuration in a crystal [BOWEN, 2006]. The positions and intensities of diffraction peaks give information on orientation and spacing of crystal planes. In the presence of an externally applied stress, the peak positions can be displaced. From these displacements it is possible to directly determine the amount and direction of the strains, and hence, stress. This technique provides excellent strain sensitivity, on the order of $10^{-4}$; however, it suffers from poor spatial resolution, $\sim 50 \mu m$ [BELYANSKY, 2009]. Therefore, for small areas, such as for MOSFET channel regions or TSV arrays, with size scales on the order of a few microns, XRD does not provide the necessary resolution.
NBD is an electron beam technique that can provide high spatial resolution. An electron beam illuminates a thinned sample, on the order of 100-300 nm, in the direction of the strain axis, [110], for devices [USUDA, 2005; VARTULI, 2007; ARMIGLIATO, 2008; BÉCHÉ, 2010]. Diffraction patterns with sharp spots are produced, similar to conventional electron diffraction. By comparing the lattice spacing in both strained and unstrained regions with a numerical fitting algorithm, a quantitative determination of the lattice strain can be found. While NBD produces superior spatial resolution, < 3 nm [BÉCHÉ, 2010], it is limited in its accuracy of lattice spacing determination to $10^{-3}$ [USUDA, 2005; ARMIGLIATO, 2008]. Furthermore, it requires time consuming sample preparation, and thinning of the sample can cause stress relaxation along certain crystal directions.

CBED, on the other hand, provides superior strain sensitivity, $10^{-4}$ [TODA, 2000] and high-resolution, ~ 10 nm [TODA, 2000; TOH, 2005]. In CBED, a convergent electron beam is incident on a thinned sample in a TEM. This gives rise to diffraction patterns in disks. Along certain directions the CBED pattern contains so-called higher order Laue zone (HOLZ) lines. These lines can be used to determine lattice parameters [JONES, 1977]. In the presence of strain, the HOLZ lines will be shifted. Strain can be determined from these shifts by fitting simulated CBED patterns with experimental data. This technique displays several limitations, one being broadening of the HOLZ lines in thin layers [TOH, 2005], such as would be found in the channel region or on sSOI. This decreases the strain sensitivity determination. Another is that the sample must be tilted off axis to prevent dynamical interactions between CBED patterns [ARMIGLIATO, 2008]; therefore, spatial resolution is typically reduced. Furthermore, HOLZ line splitting
occurs in high strain regions, such as in a MOSFET channel region, or at regions of high strain gradient, such as near interfaces [HOUDELLIER, 2006]. In the case of HOLZ splitting, heavy numerical algorithms are needed to accurately determine strains. Furthermore, sample thinning can lead to strain relaxation.

The TEM based imaging techniques, such as HREM and HoloDark both provide very high spatial resolution, on the order of 2-5 nm, as well as allow for imaging larger areas at once [HÜE, 2008; HÝTCH, 2008], but both suffer drawbacks regarding sample preparation, which is both time consuming and destructive in nature, and may also lead to strain relaxation. HREM provides a high-resolution image of the crystal lattice. By comparing the crystal lattice spacing between strained and unstrained regions provides a measurements of the strain. This presents a challenge in that an internal strain reference is needed for comparison [HÜE, 2008]. It has a strain sensitivity of $\sim 10^{-3}$. HoloDark, on the other hand, is obtained by using a bi-prism to interfere the diffracted beams from an adjacent strained region (channel) and unstrained region (substrate). The phase difference can be measured directly from the holographic fringes obtained. The strain information can then be extracted from the phase gradients [HÝTCH, 2008]. While this provides strain sensitivity on the order of CBED, $10^{-4}$, this technique is also destructive in nature due to sample preparation.

While the above techniques all have major benefits, they all present their own limitations. XRD has high strain sensitivity, but suffers from a large spot size; therefore, spatial resolution presents an obstacle. All the TEM techniques provide adequate strain sensitivity, but suffer major drawbacks in both time (i.e. numerical modeling, extensive in some cases) and, most importantly, sample thinning. The TEM techniques are all
destructive in nature, however sample thinning posses the risk of stress relaxation, whereby numerical simulations need to be carried out to evaluate the extent of such relaxation, increasing the analysis time on the order of days to weeks. Another technique that has the potential to incorporate all of the benefits of the above techniques is Raman spectroscopy. As will be described in the following sections, Raman microscopy has strain sensitivity comparable to the aforementioned techniques. Traditional micro-Raman allows for spatial resolutions down to ~300 nm, depending on optics; however, incorporating a near-field scanning optical microscope (NSOM) to the system, that resolution can be further increased to the scale of 10’s of nm, as will be discussed in Section 3.1 of this dissertation.

1.2.2 Raman Spectroscopy

Raman spectroscopy can be used as a method of measuring stresses at or near the surface of crystalline silicon. In a crystalline sample, the Raman effect is a result of photon energy loss or gain as a result of scattering from phonon modes associated with the lattice. Lattice perturbations due to strain results in a shift in the phonon mode energy, and thus a shift in Raman frequency. It is an all-optical technique, meaning it is non-contact and has a spatial resolution determined by the wavelength of light and the associated optics. However, the resolution is typically on the order of one to several microns for visible wavelengths focused by a conventional microscope. This limit can be pushed to nanometer resolution with the incorporation of a near-field scanning optical microscope (NSOM), as will be discussed in Part Two of this dissertation. Thus, Raman
spectroscopy has potential for probing local stresses in semiconductor materials in a non-destructive manner for various structures.

1.2.2.1 Brief History of Raman Spectroscopy

The Raman effect is so named after Sir C.V. Raman, who first discovered its existence in 1928. He proposed in his letter to Nature [RAMAN, 1928] that there should exist two types of scattering using ordinary light, one determined by the normal state of molecules and atoms, and a second effected by their fluctuations from the normal state. By using a telescope and a lens to converge sunlight on various liquids and gasses, and by implementing various color filters in the incident beam path and in the scattered beam paths, he was able to observe a shift in the color of light. Thus, he proved the existence of modified scattered radiation. This discovery led to his winning the Nobel Prize in 1930, only two years after its discovery.

Since then, progress in the technical instrumentation, including high power density laser sources, charge-coupled device (CCD) detectors, high numerical aperture objectives, low cutoff frequency edge and notch filters, high density gratings, and many others, has led to the advancement and more popular use of the technique in fields ranging from chemistry, biology, and physics. Development of the mathematical theory and experimental verification of the Raman effect in crystalline structures like Si, especially during the 1960s and ‘70s [LOUDON, 1964; GANESAN, 1970; CARDONA, 1975, 1982] has led to the potential application of Raman spectroscopy for semiconductor materials. In particular, silicon (Si) has been well studied with Raman spectroscopy, and
it is known that crystalline strain leads to shifts in the characteristic Si Raman peak [GANESAN, 1970; ANASTASSAKIS, 1970, 1990; DE WOLF, 1996]. The sensitivity is on the order of $10^{-3}$-$10^{-4}$, depending on the optics. This is comparable to some of the TEM techniques described previously. The Raman theory is well developed, and there is a direct correlation between the magnitudes of the Si Raman peak shift with the amount of strain present. However, before going into more details on this correlation, it is appropriate at this time to first discuss the operational aspects of Raman spectroscopy, including general instrumentation, then move onto the underlying theory behind the Raman effect in crystalline structures, in particular the diamond cubic class. Finally, the theory between stress and strain and how they relate to a shift in the Raman peak will be covered. These points will be necessary for the remainder of the dissertation, since they are fundamental to understanding and interpreting the results obtained.

1.2.2.2 Overview of the Technique

Raman spectroscopy involves the inelastic scattering of light, typically in the optical range for semiconductor materials. In the case of Raman scattering, photons from a monochromatic laser beam interact with the sample to either excite or de-excite crystal vibrations. The quantum of a crystal vibration is called a phonon with energy $\hbar q_j$ ($q_j$ is the phonon momentum, or wave vector). The changes in the phonon mode energies correspond to shifts in their corresponding frequencies. These are known as Stokes and anti-Stokes shifts in the cases of phonon excitation and de-excitation, respectively. By following conservation of energy ($\hbar \omega$), this process is written as [ESSER, 2000]:
where $\omega_s$ is the frequency of the scattered photon, $\omega_i$ is the frequency of the incident photon, and $\omega_j$ is the frequency of the phonon mode. The "-" corresponds to phonon creation, or excitation (Stokes process), and the "+" corresponds to phonon annihilation, or de-excitation (anti-Stokes process). Furthermore, conservation of momentum dictates the wave vector $q_j$ involved:

$$k_s = k_i \pm q_j$$  \hspace{1cm} (1.3)$$

where $k_s$ and $k_i$ are the scattered and incident photon wave vectors, respectively. In the visible spectrum, both the scattered and incident photon momentums are small compared to reciprocal lattice vectors; therefore, only phonon excitations with $q_j \sim 0$ participate in Raman processes [RUF, 1998]. In other words, only phonons near the Brillouin zone center are observed.

Typically, room temperature Raman measurements involve observation of the Stokes process. As can be inferred from above, anti-Stokes processes mean the phonon must already be in an excited state. The ratio of Stokes to anti-Stokes shift is given by [LONG, 1977; RUF, 1998]:

$$\frac{l_{Stokes}}{l_{anti-Stokes}} = e^{\frac{\hbar \omega_j}{kT}}$$  \hspace{1cm} (1.4)$$

where $k$ is Boltzmann’s constant, and $T$ is temperature. Therefore, at room temperature, the Stokes processes dominate. A majority of the scattered light from a monochromatic source is due to elastic, Rayleigh scattering, which has frequency $\omega_0$ equal to that of the
incident light. The Raman scattering cross-section is much smaller than the Rayleigh scattering cross-section, thus only ~1 Raman scattered photon is generated for every $10^6$ Rayleigh scattered photons. This means the elastically scattered light needs to be separated from the Raman scattered light for efficient detection.

A typical Raman spectrometer has several key components. The first is a monochromatic laser as the illumination source. Monochromatic light is used since the spectra measured are in terms of Raman shift with respect to the incident frequency $\omega_0$. For micrometer spatial resolution, a high numerical aperture (N.A.) objective should be used. In the scattering direction, a low frequency cutoff filter is used to separate out the Rayleigh scattered photons, allowing only the Stokes shift, Raman scattered light to pass through. If the anti-Stokes spectrum needs to be observed, a notch filter, which is a band-pass filter, can be used to selectively block out the laser frequency. The Raman light is then focused onto a grating, which spreads the light into its spectrum. The spectrum is then incident on a CCD detector, which is read out to a computer for analysis. Figures 2.3 and 3.4 illustrate the Raman spectrometer systems used in this dissertation.

1.2.2.3 Raman Effect in Single Crystals

Raman scattering can be explained in a classical manner. In effect, it is induced by the periodic modulation of the dielectric susceptibility $\chi_{\alpha\beta}(\omega)$. The dielectric susceptibility describes the interaction of the crystal in response to an electric field $\mathbf{E}(\omega)$. An electric field induces a dipole moment, $\mathbf{P}(\omega)$, given by [DE WOLF, 1996; ESSER, 2000]:
\[ P(\omega) = \varepsilon_0 \chi_{ij}(\omega) \mathbf{E}(\omega) = \varepsilon_0 \chi_{ij}(\omega) E_0 e^{i(k_i \cdot r - \omega t)} \]  

(1.5)

where \( \varepsilon_0 \) is the permittivity of free space. Modulation of the dielectric susceptibility occurs due to deformations caused by excitation of phonon modes. These phonon modes \( j \), with wave vector or crystal momentum \( q_j \) and frequency \( \omega_j \) are given by the following equation at position \( r \) [CARDONA, 1975; DE WOLF, 1996]:

\[ Q_j = A_j e^{(\pm i[q_j \cdot r - \omega_j t])} \]

(1.6)

where \( Q_j \) is the phonon amplitude in terms of a generalized coordinate, and \( A_j \) is a time-dependent amplitude. As mentioned, excitation of phonons can cause changes in the dielectric susceptibility. A phonon near the zone center produces a lattice deformation, which preserves translational symmetry [CARDONA, 1975]. Therefore, \( \chi_{\alpha\beta}(\omega) \) can be calculated as a function of the phonon normal coordinates \( Q_j \). \( \chi_{\alpha\beta}(\omega) \) can be expanded as a Taylor series with respect to \( Q_j \) as follows [KUZMAN, 1998; ESSER, 2000]:

\[ \chi_{\alpha\beta}(\omega_i, \omega_s) = \left( \chi_{\alpha\beta}(\omega_i) \right)_0 + \sum_j Q_j \left( \frac{\partial \chi_{\alpha\beta}(\omega_i)}{\partial q_j} \right)_0 + \sum_{j,j'} Q_j Q_{j'} \frac{1}{2} \left( \frac{\partial^2 \chi_{\alpha\beta}(\omega_i)}{\partial q_j \partial q_{j'}} \right)_0 + \cdots \]  

(1.7)

where the first term gives rise to Rayleigh scattering, the second to first-order Raman scattering, the third to second-order Raman scattering, etc. For simplicity, the third and higher order terms are ignored. So, for first-order Raman scattering, by substituting (1.7) and (1.6) into (1.5), the following is obtained:

\[ P(\omega_s) = \varepsilon_0 \chi_0 \cdot E_0 e^{i(k_i \cdot r - \omega t)} + \varepsilon_0 E_0 \sum_j \left( \frac{\partial \chi_{\alpha\beta}(\omega_i)}{\partial q_j} \right) A_j \cdot e^{-i(\omega_i \pm \omega_j) t} e^{i(k_i \pm q_j) r} \]  

(1.8)

From this it is obvious that the oscillating polarization, \( P(\omega_s) \), that gives rise to the scattered light wave has three distinct frequency components: \( \omega_i \), in the first term, for
Rayleigh scattering, and the $\omega_i + \omega_j$ and $\omega_i - \omega_j$, in the second term, for anti-Stokes and Stokes Raman scattering, respectively. Furthermore, it becomes clear that Raman scattering occurs only if $(\partial \chi_{\alpha \beta} / \partial Q_j)$ is non-zero. $(\partial \chi_{\alpha \beta} / \partial Q_j)$ is known as the Raman susceptibility [CARDONA, 1982]. It is an entity that is proportional to the so-called Raman tensor.

Raman tensors conveniently summarize selection rules of the Raman process. Selection rules determine which components of the Raman tensor are non-zero for a particular phonon mode, thus enabling one to know which phonon modes are Raman active. The selection rules result from symmetry considerations for the polarizations of incident and scattered fields. Contraction of the Raman tensors with the electric-field polarization vectors of the incident and scattered photons gives the scattering intensity [LOUDON, 1964; CARDONA, 1982]:

$$
I_s \sim |\mathbf{e}_i \cdot R_j \cdot \mathbf{e}_s|^2 
$$

(1.9)

where $R_j$ is the Raman tensor for phonon $j$, and $e_i$ and $e_s$ are the polarization vectors for the incident and scattered fields, respectively. Raman tensors are derived by group theory. Such an analysis has been carried out for all irreducible representations of the 32 crystallographic point groups [LOUDON, 1964; CARDONA, 1975]. For example, crystalline silicon, a diamond cubic structure with point group $O_h$ has the following Raman tensors for crystal directions $x = [100]$, $y = [010]$ and $z = [001]$:

$$
R_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & d \\ 0 & d & 0 \end{pmatrix}, \quad R_y = \begin{pmatrix} 0 & 0 & d \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad R_z = \begin{pmatrix} 0 & d & 0 \\ d & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} 
$$

(1.10)
where \( d \) is a constant related to the scattering efficiency. Table 1.2, obtained from [DE WOLF, 1996] shows which of the three modes can be observed with various incident and scattered field polarization directions, calculated from (1.9) for typical Si surfaces, (001) and (110) cleaved planes. It follows that from a backscattering configuration on (001) Si, only the longitudinal optical (LO) \( R_z \) mode is observed. The two transverse optical (TO), \( R_x \) and \( R_y \), modes are only observable from a (110) cleaved surface. Thus, polarization directions and optical configurations become important in determining which phonon modes are observed. (001) Si wafers are typical in the semiconductor industry; therefore, Raman measurements can be performed for characterization.

<table>
<thead>
<tr>
<th>Polarization</th>
<th>( e_i )</th>
<th>( e_s )</th>
<th>( R_x )</th>
<th>( R_y )</th>
<th>( R_z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backscattering from (001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(100)</td>
<td>(100)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(100)</td>
<td>(010)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>X</td>
</tr>
<tr>
<td>(1-10)</td>
<td>(1-10)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>X</td>
</tr>
<tr>
<td>(110)</td>
<td>(1-10)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Backscattering from (110)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1-10)</td>
<td>(001)</td>
<td>X</td>
<td>X</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(1-10)</td>
<td>(1-10)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>X</td>
</tr>
<tr>
<td>(001)</td>
<td>(001)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 1.2. Polarization selection rules for backscattering from (001) and (110) Si surfaces, as adapted from DE WOLF (1996).
1.2.2.4 Silicon Stress Effects on Raman Shift

In the absence of stress, the three Raman modes are threefold degenerate, with Raman shift $\omega_0 = 521$ cm$^{-1}$ [DE WOLF, 1996]. It has been shown [ANASTASSAKIS, 1970] that stresses in Si can lift the degeneracy of the $q \sim 0$ optical phonons and shift the characteristic, single-crystal Si Raman peak. The paper by GANESAN et al. (1970) was one of the first to address theoretically the relationship between mechanical stress or strain and the Raman mode shifts in crystals of the diamond structure. They developed a microscopic theory that explains the stress induced modifications in the first-order Raman spectra of such crystals. The theory assumes that atoms in a crystal subjected to some general forces execute harmonic vibrations about equilibrium positions that are shifted from the original equilibrium positions in the absence of such forces. Those generalized forces that cause the shift in equilibrium positions only affect the atomic vibrations through the effective harmonic force constants that determine the vibration frequencies [GANESAN, 1970]. This is called the quasi-harmonic approximation. Using the notation of [ANASTASSAKIS, 1970] for simplification, the dynamic equations for the triply degenerate $q \sim 0$ optical phonons in a diamond-like crystal under a finite strain, $\eta_{klm}$ is given by:

$$\ddot{u}_i = -\left[ K_{il}^{(0)} u_i + \sum_{klm} \frac{\partial K_{ik}}{\partial \eta_{lm}} \eta_{lm} u_k \right]$$  \hspace{1cm} (1.11)

where $u_i$ is the $i^{th}$ component of the relative displacement of the two atoms in a unit cell; $\bar{m}$ is the reduced mass of the atoms; $K_{il}^{(0)} = \bar{m} \omega_0^2$ is the effective spring constant of the phonon modes in the absence of strain; $(\partial K_{ik} / \partial \eta_{lm}) \eta_{lm}$ is the change in the spring constant due to applied strain $\eta_{lm}$; and $i, k, l, m$ designate $x, y, or z$. For very small
atomic displacements, typically less than a few percent in strain, the Lagrangian finite
strain tensor $\eta_{kl}$ reduces to the conventional strain tensor $\varepsilon_{kl}$. From the dynamical
equations, GANESAN et al. (1970) derived the following secular equation, whose
eigenvalues yield the frequencies of the optical phonons in the presence of strain
[ANASTASSAKIS, 1970; GANESAN, 1970]:

\[
\begin{vmatrix}
 p\varepsilon_{xx} + q(\varepsilon_{yy} + \varepsilon_{zz}) - \lambda & 2r\varepsilon_{xy} & 2r\varepsilon_{xz} \\
 2r\varepsilon_{xy} & p\varepsilon_{yy} + q(\varepsilon_{zz} + \varepsilon_{xx}) - \lambda & 2r\varepsilon_{yz} \\
 2r\varepsilon_{xz} & 2r\varepsilon_{yz} & p\varepsilon_{zz} + q(\varepsilon_{xx} + \varepsilon_{yy}) - \lambda
\end{vmatrix} = 0
\]

(1.12)

where $p$, $q$, and $r$ are so-called phonon deformation potentials, given by ANASTASSAKIS
(1990) as:

\[
p = m^{-1} \frac{\partial \chi_{ii}}{\partial \varepsilon_{ii}} = -1.85 \omega_0^2
\]

(1.13a)

\[
q = m^{-1} \frac{\partial \chi_{ij}}{\partial \varepsilon_{ij}} = -2.31 \omega_0^2
\]

(1.13b)

\[
r = m^{-1} \frac{\partial \chi_{ij}}{\partial \varepsilon_{ij}} = -0.71 \omega_0^2
\]

(1.13c)

where $\chi_{ij}$ is the dielectric susceptibility tensor, and $\omega_0$ is the unstrained, triply degenerate
$q \sim 0$ optical phonon mode frequency. The eigenvalues give the frequency shifts between
the strained and unstrained phonon modes $j$ as:

\[
\lambda_j = \omega_j^2 - \omega_0^2
\]

(1.14a)

\[
\Delta \omega = \omega_j - \omega_0 \approx \frac{\lambda_j}{2\omega_0}
\]

(1.14b)
The polarization direction of each phonon mode is given by the corresponding eigenvectors of the secular equation.

Solving for the expected frequency shift for a given mode requires prior knowledge of the stress or strain state. However, in the presence of a symmetrical stress distribution, the between the frequency shift and the stress reduces to a proportionality constant. Two examples will be given: a) uniaxial stress along [100]: $\sigma_{ij} = \sigma_{xx}$, and b) a biaxial stress along [100] and [010]: $\sigma_{ij} = \sigma_{xx} + \sigma_{yy}$, where $\sigma$ is the stress magnitude, and $\sigma_{ij}$ is the $ij^{th}$ component of the stress tensor.

The secular equation is solved in terms of strain; therefore, stresses must first be converted to strains. The relationship between stress and strain is represented in a generalized form of Hooke’s law by [SVERDLOV, 2011]:

$$\sigma_{ij} = C_{ijkl}\varepsilon_{kl}$$  \hspace{1cm} (1.15)

where $C_{ijkl}$ is the fourth-order elastic stiffness tensor of the material. For crystals of the diamond structure, the elastic stiffness tensor is given by three independent parameters: $c_{11}$, $c_{12}$, and $c_{44}$ [CLELAND, 2003]. Therefore, Hooke’s law for diamond-like crystals in tensor form is:

$$\begin{pmatrix}
\sigma_{xx} \\
\sigma_{yy} \\
\sigma_{zz} \\
\sigma_{yz} \\
\sigma_{xz} \\
\sigma_{xy}
\end{pmatrix} = \begin{pmatrix}
c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\
0 & c_{11} & c_{12} & 0 & 0 & 0 \\
0 & c_{12} & c_{11} & 0 & 0 & 0 \\
0 & 0 & 0 & c_{44} & 0 & 0 \\
0 & 0 & 0 & 0 & c_{44} & 0 \\
0 & 0 & 0 & 0 & 0 & c_{44}
\end{pmatrix} \begin{pmatrix}
\varepsilon_{xx} \\
\varepsilon_{yy} \\
\varepsilon_{zz} \\
2\varepsilon_{yz} \\
2\varepsilon_{xz} \\
2\varepsilon_{xy}
\end{pmatrix}$$  \hspace{1cm} (1.16)
where \( c_{11} = 166 \text{ GPa}, \) \( c_{12} = 64.0 \text{ GPa}, \) and \( c_{44} = 79.6 \text{ GPa} \) [SVERDLOV, 2011]. Likewise, strain can be calculated for a given stress by reversing Hooke’s law by using the inverse of the elastic stiffness tensor called the elastic compliance tensor \( S_{ijkl} \):

\[
\varepsilon_{ij} = S_{ijkl} \sigma_{kl}
\]

(1.17)

where the full tensor form is:

\[
\begin{pmatrix}
\varepsilon_{xx} \\
\varepsilon_{yy} \\
\varepsilon_{zz} \\
2 \varepsilon_{yz} \\
2 \varepsilon_{xz} \\
2 \varepsilon_{xy}
\end{pmatrix}
= 
\begin{pmatrix}
S_{11} & S_{12} & S_{12} & 0 & 0 & 0 \\
S_{12} & S_{11} & S_{12} & 0 & 0 & 0 \\
S_{12} & S_{12} & S_{11} & 0 & 0 & 0 \\
0 & 0 & 0 & S_{44} & 0 & 0 \\
0 & 0 & 0 & 0 & S_{44} & 0 \\
0 & 0 & 0 & 0 & 0 & S_{44}
\end{pmatrix}
\begin{pmatrix}
\sigma_{xx} \\
\sigma_{yy} \\
\sigma_{zz} \\
\sigma_{yz} \\
\sigma_{xz} \\
\sigma_{xy}
\end{pmatrix}
\]

(1.18)

The elastic compliance tensor elements \( s_{ij} \) can be calculated from the elastic stiffness tensor elements \( c_{ij} \) by the following:

\[
s_{11} = \frac{c_{11} + c_{12}}{c_{11}^2 + c_{11} c_{12} - 2 c_{12}^2}
\]

(1.19a)

\[
s_{12} = \frac{-c_{12}}{c_{11}^2 + c_{11} c_{12} - 2 c_{12}^2}
\]

(1.19b)

\[
s_{44} = \frac{1}{c_{44}}
\]

(1.19c)

so that \( s_{11} = 7.67 \times 10^{-12} \text{ Pa}^{-1} \), \( s_{12} = -2.13 \times 10^{-12} \text{ Pa}^{-1} \), and \( s_{44} = 1.26 \times 10^{-11} \text{ Pa}^{-1} \).

Now, back to the uniaxial case, where the stress tensor is given by \( \sigma_{ij} = \sigma_{xx} \), the strain components, therefore, are \( \varepsilon_{xx} = s_{11} \sigma_{xx} \), \( \varepsilon_{yy} = s_{12} \sigma_{yy} \) and \( \varepsilon_{zz} = s_{12} \sigma_{zz} \). From solving (1.12) and (1.14) by substituting in the strain tensor elements, the following phonon mode frequency shifts are obtained:
\[
\Delta \omega_x = \frac{\lambda_x}{2 \omega_0} = \frac{1}{2 \omega_0} (p s_{11} + q s_{12}) \sigma
\]

(1.20a)

\[
\Delta \omega_z = \Delta \omega_y = \frac{\lambda_y}{2 \omega_0} = \frac{1}{2 \omega_0} [p s_{12} + q (s_{11} + s_{12})] \sigma
\]

(1.20b)

where the threefold degeneracy is partially lifted so that there is one non-degenerate mode polarized along \(x\) and two degenerate modes polarized along \(y\) and \(z\). The corresponding eigenvectors are found to be \(\varepsilon_x = (1,0,0)\), \(\varepsilon_y = (0,1,0)\), and \(\varepsilon_z = (0,0,1)\). As shown in Table 1.2, backscattering from a (001) surface only allows observation of the \(z\) polarized phonon mode \(\Delta \omega_z\). Inserting the values for \(\omega_0, p, q, r, s_{11}, \) and \(s_{12}\) the relationship becomes:

\[
\Delta \omega_z (cm^{-1}) \cong -2 \times 10^{-3} \sigma \text{ (MPa)}
\]

(1.21)

Solving in a similar fashion for a biaxial stress \(\sigma_{ij} = \sigma_{xx} + \sigma_{yy}\), the \(z\)-polarized mode becomes:

\[
\Delta \omega_z (cm^{-1}) \cong -4 \times 10^{-3} \left(\frac{\sigma_{xx} + \sigma_{yy}}{2}\right) \text{ (MPa)}
\]

(1.22)

where the eigenvectors are the same as with the uniaxial stress case.

From these linear relationships, under the approximation of a symmetric stress distribution, it becomes clear that a Raman peak shift above the frequency of the unstrained phonon mode is the result of a compressive stress. Likewise, a shift in the Raman peak to a lower frequency is due to a tensile stress. The relationships between mechanical stress and Raman frequency shift are fundamental for stress characterization. These concepts will be utilized for the rest of this dissertation.
1.2.2.5 Operational Aspects and Limitations of Raman Spectroscopy

Several important considerations and limitations must be taken into account when using Raman spectroscopy for Si stress characterization. They include: calibration, environment stability (i.e. room temperature), sample heating, laser polarization and sample orientation, laser focusing, focal depth, and spatial resolution.

Calibration is a critical issue. Since stress measurements rely on quantifying shifts in the Si phonon modes, the unstressed Si peak must be taken as a reference point. Therefore, calibrating the system to the stress free Si peak is of critical importance, and this is typically accomplished by calibrating the spectrometer to a Si standard. Depending on the temperature stability of the room and the period of time the measurements are to be conducted, calibration might need to be carried out more often. However, calibration before a given set of experiments is typical and has been found to be appropriate for the collection times used in this dissertation.

Variations in room temperature over extended periods of time can lead to system drift, which can cause artificial Raman peak shifts. This goes hand-in-hand with calibration and could be a consideration if anomalies in peak shift occur at specific times over extended collection times.

As a rule of thumb, laser power is kept below 10 mW at the sample surface to minimize sample heating. Heating of the sample can cause shifts in the characteristic Si peak. For all measurements in this dissertation laser power at the samples were generally ~3-5 mW, as measured by a power meter.
It was demonstrated in Section 1.2.2.3 that polarization and crystallographic orientation of the sample is important in observation of specific phonon modes, as summarized for Si in Table 1.2. Therefore, it is important to keep this in mind when designing Raman stress experiments.

Light penetration depth in opaque materials is determined by the optical absorption coefficient. For example, 514.5 nm wavelength light penetrates into Si by ~770 nm [DE WOLF, 1996]. Therefore, the measured signal is a spatial average over this depth. If a more surface specific measurement is required, then decreasing the wavelength will be necessary; however, this may involve the need for additional optics and/or detector if moving into the UV range. In addition, the use of confocal Raman systems can significantly reduce the depth of focus. However, for opaque materials, the depth average becomes more complex due to refraction [EVERALL, 2000, 2007; BALDWIN, 2001; ADAR, 2010]. As a result, consideration of the Raman penetration depth is necessary when making statements about or analyzing stresses or shifts at different depths.

Finally, spatial resolution is limited by the diffraction limit of light. When determining whether Raman spectroscopy is appropriate for a given sample structure, this must be kept in mind. However, there is a technique for overcoming this limitation: by implementing a NSOM with a Raman spectrometer the resolution can reach the nanometer regime. So-called TERS will be discussed in Part Two of this dissertation.
PART TWO: TIP-ENHANCED RAMAN SPECTROSCOPY FOR SUB-DIFFRACTION LIMITED RESOLUTION STRESS CHARACTERIZATION

2.1 INTRODUCTION AND BACKGROUND

This section provides the background and theoretical basis for near-field scanning optical microscopy and its application to tip-enhanced Raman spectroscopy. It is meant as a cursory overview for understanding the experimental setup and procedure employed in this dissertation to evaluate TERS for sub-diffraction limited resolution stress characterization. For more exhaustive background and theoretical treatments of NSOM the author refers the reader to works such as OHTSU (1999), COURJON (2003), and NOVOTNY (2006).

2.1.1 Overcoming the Optical Diffraction Limit

ABBÉ (1873) showed that the best optical resolution possible with conventional optics is given by:

$$\Delta r_{\text{min}} = 0.61 \frac{\lambda}{\text{N.A.}}$$  \hspace{2cm} (2.1)

where $\Delta r_{\text{min}}$ is the minimum distance between two points resolvable as separate, also deemed the optical spatial resolution; $\lambda$ is the wavelength of the illuminating light; and N.A. is the numerical aperture of the lens, defined by:

$$\text{N.A.} = n \sin \theta$$  \hspace{2cm} (2.2)
where $n$ is the index of refraction of the medium between the lens and the sample ($n=1$ for air), and $\theta$ is the beam convergence angle, or half the angular spectrum. Equation (2.1) is called the *diffraction limit*, which arises due to the spreading (i.e. diffraction) of plane waves passing through the interstitial spaces of the sample and lens. Therefore, the minimum resolvable distance between two points is limited by this spreading due to diffraction.

A typical, high-resolution, dry objective has a N.A. of $\sim 0.9$. This essentially limits the optical resolution to $\sim \lambda/1.5$. For a wavelength of 532 nm, as is primarily used in the Raman measurements for this dissertation, the minimum theoretical resolution attainable would be $\sim 361$ nm. This is an ideal resolution limit; however, taking into account the beam profile and imperfections in optics, the actual resolution is typically greater than this value. Despite that fact, for micro-Raman spectroscopy applied to stress measurements, the resolution of $\sim 361$ nm is acceptable in regions where stress fields do not vary rapidly over micron length scales. However, for measuring highly localized stresses, such as in the channel region of MOSFETS with $< 32$ nm half-pitch, a minimum resolution $\lesssim 32$ nm would be required. From the diffraction limit, it is clear that this is not possible at optical wavelengths with conventional optics.

In 1928, after discussions with A. Einstein, E. Synge published his idea to overcome the diffraction limit [SYNGE, 1928; GREFFET, 1997]. The idea is simple, if a small aperture is brought into close proximity with a sample and illuminated from the backside, then the illuminated spot will not be diffraction limited. This illumination spot would be limited by the aperture size. Therefore, by scanning this small aperture over the surface, step-by-step, an image of the surface can be formed that has a resolution determined by
the aperture – not the wavelength – of the illumination source. This sort of system would require a very small aperture (diameter $< \lambda$) and a system that would allow for maintaining a distance between the aperture and sample on the order of nanometers.

The advent of the scanning probe microscope in 1982 [BINNIG, 1982] allowed for high precision distance control between a probe and sample, thus, providing a platform to realize Synge’s concept. Shortly after the introduction of the scanning probe microscope, two, independent groups demonstrated subwavelength image formation at optical frequencies by scanning a small aperture over a sample [POHL, 1984; LEWIS, 1984]. The subwavelength aperture was created by impinging a metal-coated, quartz tip into a sample until light leaked through. This was the birth of the near-field scanning optical microscope (NSOM).

2.1.2 Aperture Near-field Scanning Optical Microscopy

As mentioned previously, the initial concept for overcoming the diffraction limit of light for imaging was to use a very small aperture in close proximity to the surface. In modern day instruments, the most typical aperture probe is a tapered fiber with a metal coating near the apex, as illustrated in Figure 2.1. Light propagates through the fiber and becomes confined to the aperture, essentially acting as a very small illumination source. The metal coating, typically Al, is used to confine the light to the aperture and prevent leakage from the tapered sidewalls.
Optical confinement through a subwavelength aperture can be described by the angular spectrum representation of optical fields and Heisenberg’s uncertainty principle. Consider a plane wave, \( E = E_0 e^{i(k_z z - \omega t)} \), with wave vector \( k_z \) propagating along the \( z \)-direction and passing through an aperture, which is in the \( x-y \) plane. The two-dimensional electric field in the aperture plane is given as [Novotny, 2006]:

\[
E(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{E}(k_x, k_y; z = 0) e^{i(k_x x + k_y y + k_z z)} dk_x dk_y
\]  

(2.3)

where \( \hat{E}(k_x, k_y) \) is the Fourier transform of the field in the aperture plane, \( x, y, \) and \( z \) are spatial coordinates, and \( k_x, k_y, \) and \( k_z \) are the corresponding wave vector components. The dispersion relation is given as:
\[ k^2 = k_x^2 + k_y^2 + k_z^2 = \left( \frac{\omega}{c} \right)^2 n \]  

(2.4)

where \( n \) is the index of refraction of the medium, \( \omega \) and \( c \) are the angular frequency and speed of light, respectively. In a dielectric medium or air, \( n \) is a real and positive quantity, and \( k_z^2 \) can be either positive or negative. If positive, the waves diffracted through the aperture take on the form \( e^{ik_z z} \), which corresponds to an oscillatory, propagating wave as for the initial plane wave. If, however, \( k_z^2 \) is negative, then the diffracted wave takes on the form \( e^{-ik_z z} \), which describes an exponentially decaying (evanescent or non-propagating) wave in the \( z \)-direction. According to Heisenberg’s uncertainty principle [GIRARD, 1996]:

\[ \Delta s \Delta k_s \geq 2\pi \]  

(2.5)

where \( s = x, y \) and \( \Delta s \) is the radius of the aperture in the \( x-y \) plane. Therefore,

\[ \Delta k_z \geq 2\pi \Delta s^{-1} \]. For a subwavelength aperture:

\[ 2\pi \Delta s^{-1} > 2\pi \lambda^{-1} = \frac{\omega}{c} = k \]  

(2.6)

Substituting (2.6) back into (2.5), it is seen that \( \Delta k_z > k \). According to the dispersion relation this requires \( k^2 < k_x^2 + k_y^2 \), thus \( k_z^2 \) becomes negative, resulting in an evanescent field that decays away from the subwavelength aperture. When the aperture probe is brought into sufficient proximity to the sample (i.e. several nm), the oscillating evanescent field is able to locally interact with the sample and induce optical emission which can be detected back through the illuminating fiber, as illustrated in Figure 2.1. Thus, the evanescent field generated from a subwavelength aperture acts as a localized, illumination source.
There are two critical issues with aperture probes that limit its suitability for use as a high-resolution Raman technique for measuring stresses in FEOL devices. As already mentioned, an appropriate spatial resolution for optical (Raman) stress measurement of stress in the channel region for the current and emerging technology nodes would be < 32 nm. Light incident on a conducting metal has a finite penetration depth called the skin depth. Therefore, resolution of an aperture probe is decreased by the skin depth associated with the metal coating on the outside of the fiber. For a good metal coating, the skin depth is ~6-10 nm (on one side of the probe) at optical frequencies [NOVOTNY, 2006, Ch.5]. If the aperture was zero, the skin depth would still be ~15 nm. Furthermore, as the aperture size decreases, the signal-to-noise ratio decreases due to decreases in optical throughput (~10^{-4} for 100 nm aperture) [NOVOTNY, 2006, Ch.6]. As a result of these limitations, aperture probes are typically kept between ~ 50-100 nm.

Considering the fact that Raman scattering has a much lower scattering probability than elastically scattered light, as mentioned in Section 1.2, the optical loss with decreasing aperture size becomes a major hurdle for its application to high-resolution Raman imaging for stress characterization. Furthermore, the finite resolution limit due to the skin depth associated with the metal coating becomes troublesome when looking at channel regions \( \lesssim 32 \) nm.

One solution is associated with the fact that small metallic structures, with size \( \ll \lambda \), can generate localized, evanescent fields [INUYE, 2001]. Moreover, when these tips or particles are made with certain metals, significant enhancements of the locally scattered fields can occur, which increase the scattering intensity observed. This is particularly
beneficial to high-resolution Raman imaging, as will be discussed in the following section.

2.1.3 Apertureless NSOM and Tip-Enhanced Raman Spectroscopy

Obtaining optical resolution $< 50$ nm (ideally, $\lesssim 32$ nm) with high signal intensities is essential for high-resolution Raman spectroscopy as applied to MOSFET channel length size scales. Due to the limitations of aperture based NSOM approaches, a new NSOM approach was introduced using an apertureless probe. Lateral resolutions for an apertureless approach have been reported as low as 3 nm [SPECHT, 1992; ZENHAUSERN, 1994].

Apertureless NSOM makes use of strongly enhanced, localized electric fields produced around the vicinity of a sharp metallic tip under laser illumination. An illustration can be seen in Figure 2.2. Laser illumination incident on the probe, with radius $\ll \lambda$, can generate an evanescent field localized to the probe tip [INOUE, 2001]. Due to the subwavelength probe radius, the electric field lines become crowded, thereby confining a large amount of electromagnetic energy in a small area around the tip [NOVOTNY, 2006]. This so-called “lightning rod” effect is similar to that in antenna theory and can lead to field enhancement in the vicinity of the tip. Furthermore, it has been shown that apertureless probes made of or coated with noble metals (Au or Ag) can exhibit large enhancement factors, on the order of $10^3$-$10^7$, due to excitation of local surface plasmon-polaritons (LSPs) [FESTY, 2004]. As a result, placement of an illuminated apertureless probe in the vicinity of a Raman-active material can induce
locally enhanced Raman scattering. This is the basis for so-called tip-enhanced Raman Scattering (TERS) [Wessel, 1985; Novotny, 1998; Anderson, 2000; Hayazawa, 2000; Festy, 2004; Anderson, 2006].

![Diagram of apertureless probe](image)

Figure 2.2. Illustration an apertureless probe. Far-field illumination (a-top) scatters off a subwavelength metallized probe and generates a localized evanescent field. Upon interaction with a sample (a-bottom), the scattered field is collected in the far-field. Simulation (b) of an enhanced evanescent field localized in the vicinity of a metal probe. [With kind permission from John Wiley and Sons: Cançado et al. (2009)].

TERS is based upon excitation of LSPs in a noble metal probe. The probe can be a sharpened noble metal probe, a noble metal coated probe, or a dielectric probe with a noble metal particle at the apex. The optical properties of metals are described by the complex dielectric function $\varepsilon(\omega)$. In a bulk metal, the free electrons move in a
background of fixed positive ions. The free electron gas is, by definition, a plasma, with a dielectric function given in an electric field as [Kittel, 2005]:

\[ \epsilon(\omega) = 1 - \frac{ne^2}{\epsilon_0 m \omega^2} \]  

(2.7)

where \( n \) is the number of electrons, \( e \) the electron charge, \( \epsilon_0 \) the permittivity of free space, \( m \) the mass of the electron, and \( \omega \) the angular frequency of the electric field. The term \( \sqrt{\frac{ne^2}{\epsilon_0 m}} \) is referred to as the plasma frequency, \( \omega_p \), which describes the oscillation of the free electron charge density in a bulk material. Therefore, (2.7) becomes:

\[ \epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2} \]  

(2.8)

In a metallic nanoparticle, such as a sphere or a tip, however, a damping term must be introduced to account for electron collisions. Therefore, (2.8) is given as the Drude model [Pinchuk, 2004; Maier, 2007]:

\[ \epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\gamma \omega} \]  

(2.9)

where \( \gamma \) is the damping rate (i.e. electron collision rate). The Drude model holds for metals where the incident photon energies are below electronic band transition energies, or where electron excitations remain within the conduction band (intra-band). This is true for most metals, where those energies lie in the ultraviolet and higher energy regions of the electromagnetic spectrum. However, some noble metals, such as Au and Ag have interband transitions at energies as low as 1 eV, where bound electrons are excited to higher energy bands. To account for interband transitions, the dielectric function becomes [Pinchuk, 2004]:
\[
\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\gamma \omega} + \chi^{ib}(\omega)
\]  

(2.10)

where \(\chi^{ib}(\omega)\) is the complex interband transition susceptibility, which only contributes to the total dielectric function at frequencies above the interband transition edge. The effect of interband transitions in Au and Ag leads to strong plasmon resonances at optical wavelengths [PINCHUK, 2004].

For small Au or Ag particles or structures (\(d << \lambda\)), the electromagnetic modes that the particle can support become discretized, rather than continuous [LE RU, 2009]. These modes, called localized surface plasmon-polaritons (LSPs), can be excited by an incident wave of appropriate polarization and frequency. Due to the spatial confinement, the LSPs are essentially non-propagating. Furthermore, the curvature of the particle exerts an effective restoring force on the electron oscillations, so that resonance can arise, which leads to field amplification [MAIER, 2007]. Therefore, direct illumination of a Au or Ag apertureless NSOM probe with \(d << \lambda\) can locally excite LSPs to create a locally enhanced light source at the tip apex.

As mentioned, the important factors in exciting LSPs in a Au apertureless TERS probe, and hence the degree of enhancement, are polarization and frequency. The LSP resonance (LSPR) for a Au nanosphere is \(\sim 510\) nm and is independent of sphere radius for small spheres, with slight redshift for larger spheres [AIZPURUA, 2005]. Therefore, if the tip of a Au TERS probe is assumed to be a sphere, which would be true for a Au particle at the end of a dielectric pipette (i.e. the Nanonics setup for this dissertation), then a driving excitation source of \(\sim 510\) nm should excite the LSPs in the probe tip. Furthermore, the LSPR FWHM is typically on the order of 50 nm. Therefore,
enhancement due to LSP excitation can occur for a range of illuminating wavelengths, though with varying degrees of enhancement. The Ar$^+$ laser used in this dissertation has a wavelength of 514.5 nm, which is expected to be close to the LSPR peak.

With regards to polarization, p-polarization (E-field parallel to tip axis) has been shown to be necessary for most efficient LSP excitation [NOVOTNY, 1998; SUN, 2003a; DEMMING, 2005]. Therefore, the polarization in this dissertation is set to align with the tip axis.

One other important factor that can affect enhancement is probe-sample separation. It has been shown that the field enhancement depends exponentially on probe-sample separation [LARSEN, 2001; RICHARDS, 2003; FESTY, 2004; BOUHELIER, 2004; ANDERSON, 2006]. The closer the probe is to the sample, the stronger the field-enhancement. This is due to the exponential decay of the localized, evanescent field excited at the probe tip. For a separation distance of ~1-2nm, the field-enhancement should be maximized.

2.1.4 Potential Advantages of Apertureless NSOM for FEOL TERS Stress Characterization

As stated previously, aperture probes have several critical limitations that pose problems for stress characterization in current and future generation FEOL devices. Resolution limits of ~50 nm will not be sufficient for measurements where half-pitches are < 32 nm. Furthermore, severe decreases in optical throughput at such aperture sizes would not allow for efficient Raman collection times.
Aperture probes have the benefit of resolutions that are determined by the probe radius and near-field enhancement due to excitation of LSPs. Since the evanescent near-field generated at the tip is localized to the apex of the tip, the effective resolution is determined by the sharpness of the probe. Therefore, spatial resolution is dependent upon probe manufacturing. As mentioned before, resolutions of ~3 nm have been reported. Moreover, with weak Raman scattering cross-sections, the enhanced near-field due to LSPRs increases the obtained signal, allowing for more efficient signal collection with better signal-to-noise.

The remainder of Part Two of this dissertation will describe the TERS experimental setup for measuring localized stresses on patterned strained silicon-on-insulator samples. This will be followed by the obtained results including comparison of the observed stresses in the patterned structure with finite element simulations. Finally, the resolution and potential for TERS as a FEOL stress characterization technique will be discussed and evaluated.
2.2 TIP ENHANCED RAMAN SPECTROSCOPY EXPERIMENTAL SETUP

2.2.1 Experimental Objective

Tip-enhanced Raman spectroscopy (TERS) was used to characterize local stresses on well-defined, patterned strained silicon-on-insulator (sSOI) test structures and to evaluate its potential for high-resolution stress measurements for front-end-of-line (FEOL) devices. To further improve resolution and repeatability of an apertureless NSOM approach, small diameter (< 10 nm desired) and robust probe tips with reproducible near-field enhancements must be utilized. Thus, a novel NSOM probe is proposed and manufactured. The full details of that work can be found in Appendix A. This section will describe the experimental details of the commercial TERS system implemented as well as the sSOI samples used for characterizing the system’s potential for high-resolution applications.

2.2.2 Instrumentation

The TERS system employed in this work is a combination of two commercial systems. Nanonics Imaging, Ltd. has designed an NSOM that directly integrates onto a Renishaw Raman spectrometer. Figure 2.3 illustrates this setup schematically [ATESANG, 2006]. The Nanonics MultiView 2000 (MV2000) NSOM sits directly on a Leica DM/DL optical microscope stage allowing normal incidence illumination of the apertureless probe tip as well as the sample.
The main components of the setup, as seen in Figure 2.3, include: a Coherent Innova 300D Ar$^+$ laser with 514.5 nm wavelength; Renishaw RM-100 Raman spectrometer; and Nanonics MV2000 NSOM with an apertureless NSOM that provides near-field enhancement. The laser light is illuminated through a 50× long working distance objective (N.A. 0.5) at normal incidence with polarization along the axis of the apertureless NSOM probe (to generate enhancement [NOVOTNY, 1998]), which is bent at approximately 45° from the vertical. Backscattered light passes through an edge filter, which blocks the elastically scattered Rayleigh light, allowing only the Raman photons to pass through. The Raman light is then dispersed onto a Peltier cooled CCD detector by a 3000 line/mm diffraction grating. The effective spatial resolution with this setup is ~0.05 cm$^{-1}$, correlating to a biaxial stress resolution of ~22 MPa.

Figure 2.3. Schematic of the main components of the Renishaw RM-100 Raman spectrometer and Nanonics MV2000 NSOM.
Nanonics has developed a unique NSOM for TERS. It has independent $x$, $y$, and $z$ piezo scanning capabilities, with $\sim$ 1 nm resolution, for the sample (bottom of NSOM head) and the tip (top of NSOM head). This is possible with two separate controllers. The apertureless NSOM probe is attached to the top of the NSOM head and maintains feedback with the sample surface ($\sim$ 1-2 nm separation) through the use of a shear-force tuning fork based system. The tuning fork is tuned to its resonance frequency, and a constant height is maintained through feedback by detecting phase changes in the vibration of the tuning fork when the probe tip encounters topography. With TERS, maintaining constant $z$-height is carried out by adjustments of the $z$-piezo scanner where the tuning fork is mounted (top of NSOM head). The sample is scanned in $x$ and $y$ by the piezos below the sample (bottom of NSOM head).

The laser spot size is $\sim$ 1.5 $\mu$m in diameter. When illuminating a small apertureless NSOM probe there is a significant far-field background signal. Therefore, the signal detected in the presence of near-field enhancement at the probe tip is a combination of the far-field signal and the locally enhanced near-field signal. Therefore, it is necessary to eliminate the background signal to extract the near-field signal. The Nanonics TERS system allows for a differential scanning approach for separating the near-field spectrum from the far-field. In other words, it takes the difference between two spectra, one when the tip is in feedback and one when it is retracted, to eliminate the far-field contribution. The details of this approach are elaborated on in Section 2.3.

Apertureless NSOM probes were purchased from Nanonics. The probes consist of an Ag nanoparticle at the end of a tapered, glass micropipette, which is then coated with Au. A conventional SEM image of a representative tip is seen in Figure 2.4. The nominal tip
radii observed from these commercial tips are on the order of 150-250 nm, as seen in the inset. Micropipettes are attached to tuning forks, which are in turn attached to Nanonics specific mounts for electrical connection to the NSOM head.

Figure 2.4. SEM image of a representative Nanonics micropipette tip.

2.2.3 Strained Silicon-on-Insulator Samples

Strained SOI can be used for enhancing carrier mobility [SVERDLOV, 2011; HORSTMANN, 2005] in CMOS device channels. Therefore, it is relevant to investigate material sets with the end goal of employing TERS for device channel strain metrology. Furthermore, the structures are such that characterizing tip-enhancement is straightforward. Since near-field enhancement is a highly localized surface technique, only the strained Si layer should display an enhanced signal, while the Si substrate peak provides a constant baseline for comparison.
Blanket (001) sSOI wafers were provided by SOITEC. The stack structure is illustrated in Figure 2.5.

Figure 2.5. Left: Illustration of SOITEC stack cross-section. Right: SOITECH strained Si (sSi) and Buried Oxide (BOX) layer thicknesses.

A micro-Raman spectrum (far-field) was collected from the blanket wafer to characterize the stress of the strained device layer relative to the bulk Si (Figure 2.6). The fit parameters $y_0$, $x_c$, $w$, and $A$ correspond to the baseline, peak center, peak width, and peak height, respectively. Lorentzian fits are shown for both the strained SOI (514.65 cm$^{-1}$) and Si substrate (520.68 cm$^{-1}$) peaks. From the difference in peak centers, the a relative peak shift difference of 6.04 cm$^{-1}$ is obtained, corresponding to a biaxial stress [DE WOLF, 1996] of 2.63 GPa. This corresponds to an in-plane strain of ~0.79%, which is close to the SOITEC specified 0.75% and the XRD obtained in-plane strain of 0.746% on the same wafer [ATESANG, 2006].

A blanket sSOI wafer was processed at CNSE to obtain patterned structures with well-defined edges to investigate TERS and evaluate spatial profiling to determine its applicability for localized stress measurements. A contrast inversion photolithography
process was used in combination with a conventional interconnect via mask to pattern an array of thin (50 nm) Cr islands across the sSOI wafer [MCDONOUGH, 2007]. Wet etch with tetramethylammonium hydroxide (TMAH) resulted in removal of the unprotected

Figure 2.6. Micro-Raman spectrum of a blanket sSOI sample. Peaks are labeled as sSi and Si for strained Si and substrate layers, respectively.

strained Si layer. The buried oxide acted as a highly selective etch stop. Cr was then removed via a wet process. SEM and AFM images can be seen in Figure 2.7. The patterned mesa structures are ~ 2 µm in diameter, and display an octagonal shape as a result of the anisotropic TMAH wet etch. The AFM line profile in (b) shows the thickness of this particular mesa to be ~ 63 nm, the same specification provided by
SOITEC. Therefore, there is little to no thinning of the sSi mesas compared to the bulk sSi device layer.

Figure 2.7. Images of ~2 µm sSi mesa structure with (a) SEM and (b) AFM.

2.2.4 Finite Element Model for Stress Analysis in Mesa Structure

A simulation of the geometrically driven stress relaxation in a 2 µm diameter sSi mesa was conducted with COMSOL Multiphysics®. The geometry is displayed in Figure 2.8. A line profile of the simulated stress along the specified direction [100] is compared to the corresponding TERS-collected profile (Figure 2.13). A static analysis was conducted with the Structural Mechanics Module. The material properties were taken from the COMSOL MEMS materials library and are listed in Table 2.1. The mesa was given an initial biaxial stress of 2.63 GPa, as determined by micro-Raman measurements and subsequent peak analysis on blanket sSOI (Figure 2.5).
Figure 2.8. Geometry used for FE model (a) and corresponding mesh (b). Red indicates Si substrate, gray indicates SiO$_2$, and blue indicates sSi mesa. Material thicknesses are taken from SOITEC specifications (Figure 2.5—right).

<table>
<thead>
<tr>
<th>Material</th>
<th>Young’s Modulus E (GPa)</th>
<th>Poisson’s ratio $\nu$</th>
<th>Initial Biaxial Stress $\sigma_{xx}, \sigma_{yy}$ (GPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si (anisotropic)</td>
<td>Direction dependent</td>
<td>Direction dependent</td>
<td>—</td>
</tr>
<tr>
<td>sSi (isotropic)</td>
<td>170</td>
<td>0.28</td>
<td>2.63</td>
</tr>
<tr>
<td>SiO$_2$ (isotropic)</td>
<td>70</td>
<td>0.17</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 2.1. Material properties of sSOI mesa for FE model.
2.3 TERS STRESS RESULTS ON A STRAINED SILICON-ON-INSULATOR Mesa

2.3.1 Characterization of Tip Enhancement

Taking advantage of the differential scanning inherent with TERS requires near-field enhancement at a probe tip. Characterizing near-field enhancement for a given tip is necessary to identify its suitability for TERS mapping on the sSOI mesa. TERS probes fabricated by Nanonics were characterized for enhancement using a differential approach on blanket sSOI.

A specific protocol was employed to optimize near-field enhancement by positioning the probe within the laser beam. A probe mounted to the Nanonics MV2000 NSOM is tuned to its resonance frequency and approached to the surface until feedback is achieved. Once feedback is attained, the probe is translated through the laser spot in the x and y directions by controlling the probe positioning piezo-crystals [ATESANG, 2006]. Figure 2.9 illustrates this schematically. While simultaneously collecting Raman spectra, the probe is moved within the laser spot by incremental steps (on the order of nm) in the x and y directions, until the strained Si peak intensity is maximized. A maximum strained Si peak intensity results when near-field enhancement from the TERS probe is at its highest, thereby increasing localized scattering at the sample surface. With this obtained optimum position, characterization of the enhancement can be carried out using the differential approach.
Due to the large laser spot size, as compared to the TERS probe, the Raman spectrum collected with the probe at its optimum enhancement position contains both far-field and near-field components. It is necessary to subtract out the far-field component to determine the magnitude of near-field enhancement. To this end, the probe is retracted away from the surface, and a second Raman spectrum is collected, which consists entirely of far-field scattering. If near-field enhancement occurs when the probe is in feedback with the surface, then subtracting this far-field Raman spectrum from the feedback spectrum yields a differential spectrum comprising the near-field signal.
Figure 2.10. (a) Raman spectra for blanket sSOI used for characterizing the TERS probe used in these studies. The feedback spectrum contains both near-field and far-field components. The retracted spectrum consists entirely of far-field scattering. Illustrations show the tip position for both cases. (b) Differential spectrum between the feedback and retracted spectra.

The resulting characterization spectra on blanket sSOI for the probe used in the measurements for this study are displayed in Figure 2.10. The feedback spectra and retracted spectra refer to the position of the probe with respect to the sample (see illustrations in Figure 2.10a). The peaks centered at ~ 516 cm\(^{-1}\) and ~ 520 cm\(^{-1}\) are the strained Si (top layer) and bulk Si peaks (substrate), respectively. Note that the substrate peak is observed, since light at 514.5 nm wavelength is expected to have a penetration depth of 770 nm [De Wolf, 1996] in (001) Si. The thin, strained Si layer is not expected to have a significant impact on this penetration depth, so the far-field illumination will penetrate the strained Si and oxide layers (~ 200 nm) to the Si substrate. As expected for a probe displaying near-field enhancement, the strained Si peak in the feedback spectrum has a higher intensity than that of the retracted spectrum. The raw enhancement is
quantifiable as the strained Si peak intensity differential between feedback and retracted modes normalized to the intensity of the retracted probe mode:

\[
\%\text{enhancement} = \frac{I_{\text{feedback}} - I_{\text{retracted}}}{I_{\text{retracted}}} \tag{2.11}
\]

where \( I \) is the measured strained Si peak intensity, and feedback and retracted refer to the state of the tip with respect to the sample. Based on the strained Si peak intensities obtained from Lorentzian peak fitting, 53\% raw enhancement of the strained Si Raman intensity is observed for this probe.

### 2.3.2 TERS on a Patterned Strained Silicon-On-Insulator Wafer

Strained SOI mesas have been used to evaluate the improved spatial resolution potential offered by TERS and its applicability to localized stress measurements. An estimate of the confinement of the near-field signal in the direction normal to the surface will be determined using the patterned wafer. Spatial resolution in the x-y plane will be evaluated by TERS mapping across a strained SOI mesa. Lastly, the stress profile calculated from the same TERS mapping will be evaluated and compared with a FEM simulation of the mesa geometry.

The same measurement conducted in Figure 2.10 was repeated on a section of the patterned SOI wafer where the strained Si layer had been etched away. This region consisted only of the exposed, 140 nm buried oxide layer on top of the Si substrate. Again, two spectra were collected with the tip in feedback and with it retracted. A
differential signal was calculated from these two spectra. Figure 2.11 contains the results of this measurement. The red spot indicated on the SEM inset illustrates the region measured off the patterned strained SOI mesa. No bulk Si peak is observed in the differential signal (red curve), which indicates that the near-field signal generated by the enhanced, evanescent field at the probe tip is confined to a z-depth of 140 nm. This agrees with Figure 2.10b, where no bulk Si peak from the substrate is observed in the differential signal. Therefore, any near-field enhancement measured from the strained Si layer is expected to be confined to that layer.

Figure 2.11. Differential TERS spectra collected on a region of the patterned wafer where no strained Si remained.
To evaluate the spatial resolution and stress measurements on the patterned sSOI mesas, differential scanning was conducted using the same protocol used in probe enhancement characterization above. However, unlike a single point spectral measurement, TERS mapping mode collects spectra across one- or two-dimensional arrays of points that are selected on the substrate. At each point specified in the array both feedback and retracted mode spectra are collected in order to calculate a spatially resolved differential signal. This isolates the localized near-field signal, providing a higher resolution mapping of the structure than what can be obtained with far-field, micro-Raman microscopy. The position of the strained Si peak from the differential signal is used to calculate a relative shift, $\Delta \omega_{\text{Si}} [\text{cm}^{-1}]$, from the bulk Si peak, which is then used to calculate the biaxial stress in the mesa, as per the description in Section 1.2.

The probe characterized in the previous section was applied to a sSOI mesa to evaluate its spatial resolution. Figure 2.12 contains the Raman intensity profiles for a line across the center of the mesa structure. The intensities plotted were taken from Lorentzian peak fits of the spectra. The inset is the measured topographic map with an illustrative line of where the spectral profiles were collected. It is immediately evident that the differential signal (green) comprised entirely of scattering from the enhanced optical near-field localized to the probe tip, yields a dramatic improvement in spatial resolution compared to the far-field profile (black). It will be shown in Section 2.4 that the resolution obtained is on the order of the probe diameter, which is below the diffraction limit of the illumination source wavelength.
Figure 2.12. Plot of Feedback, Retracted, and Differential line profiles from a 2 µm sSOI mesa structure.

Local stress in the strained Si mesa was measured using the strained Si peak Raman shifts from the differential profile and validated with computer simulation. Peak center positions were obtained from Lorentzian peak fitting of the spectra. Those values were used in determining the amount of shift, $\Delta \omega_{\text{Si}}$ [cm$^{-1}$], between the strained Si and bulk Si peaks. $\Delta \omega_{\text{Si}}$ was then converted to stress using the biaxial stress approximation [DE WOLF, 1996], under the assumption that processing of the mesas did not significantly break the biaxial symmetry of the blanket strained Si layer. This is experimentally reasonable given the thickness of the sSi layer and the free-surface condition on the top and sides of the sSi mesa. The measured stress profile (red triangles)
is plotted in Figure 2.13 along with the corresponding differential (green) and AFM (magenta) profiles. A stress profile (red curve) calculated using FE modeling is in agreement with the measured stress profile.

Fig 2.13. Line profiles for simulated and measured stress, differential signal, and AFM topography for the 2 µm sSOI mesa.
2.4 DISCUSSION OF TERS RESULTS

2.4.1 Effective Resolution of the TERS Probe

Extension of the near-field in the direction normal to the surface was determined to be less than 140 nm, the thickness of the oxide layer. The oxide essentially acts as a buffer to increase the distance between the probe and the Si substrate. The fact that no enhancement of the bulk Si peak is seen proves that the evanescent field is confined to less than 140 nm, which is provided as an upper limit. Near-field enhancement has been shown to depend upon probe-sample separation distance (see Section 2.1). It has been shown that Raman enhancement is typically confined to regions of 40 nm and less, with highest enhancements for separation distances around 1-2 nm [HARTSCHUH, 2003]. A similar probe-sample separation study has not been carried out for these tips due to limitations in z-height manual control of the Nanonics system. However, it does not affect the results obtained, as the tip is maintained at a constant z-height of approximately 1 nm with the shear-force tuning mechanism feedback loop.

The effective lateral resolution of the NSOM probe has been determined by fitting a convolution of the Heaviside step function and a Gaussian, in the form of complementary error function, to the differential signal. The Heaviside step function provides a mathematical description of a vertical boundary [WEISSTEIN, 2007]. Deviations away from this case are due to the system response, which involves complex interactions between the laser source, probe, sample, and the detection system. As will be seen, a Gaussian distribution provides a good approximation of the system resolution. The
convolution of these two functions is of the form of a complementary error function [CASA SOFTWARE, 2006]. The mathematical description is given here:

$$erfc\left(\frac{c-x}{t}\right) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-\left(\frac{c-y}{t}\right)^2} dy$$

(2.12)

The parameter $c$ represents the position of the edge, and $t$ represents the width of the Gaussian. This function is fit to our differential signal using a least squares fitting algorithm in OriginPro 8.5. Figure 2.14 displays the results of fitting the complementary error function to the differential signal edge.

The resolution of the system is estimated by the width of the Gaussian used in the fit. This value is seen in the fit parameters of Figure 2.14 to be $\sim 210 \pm 80$ nm. Due to the differential TERS approach taken, most of the far-field effects and system response should have negligible influence on the resolution determined by the differential signal fit. Therefore, the dominant influence on the system resolution, and hence, the Gaussian width, should be the probe. Literature has shown that the lateral resolution of NSOM is approximately on the order of the probe radius [SUN, 2003b; FESTY, 2004]. Comparing the resolution obtained here from the differential signal edge fit with the SEM measurement in Figure 2.4 is trivial. It is found that the experimentally determined TERS resolution of $210 \pm 80$ nm is indeed on the order of the probe size ($\sim 200$ nm). This provides confirmation to what has been shown in literature, as well as supports the claim of a probe dominated system resolution in differential TERS.

Probe size dependent resolution opens up the possibility of FEOL device characterization. As has already been shown, the resolution capabilities are highly dependent upon the probe radius. This means fabrication of probes with radii on the
order of 10 nm would yield resolution on the same order. This scale size would be necessary for FEOL device characterization, which is currently at the 32 nm node.

![Figure 2.14](image)

**Figure 2.14.** Complementary error function fit of mesa edge for determining effective tip resolution.

Though the tips obtained from Nanonics were on the scale of 100-200 nm, there have been demonstrations of resolutions down to 20 nm in literature [SÁNCHEZ, 1999; HARTSCHUH, 2003]. However, probe manufacturing and repeatability still presents a major challenge for reproducible results, thus currently limiting large scale uses of the technique in a fab setting.
An alternative probe geometry with the potential for mass manufacturability has been proposed. The new probe design is a gold and carbon nanotube (Au-CNT) hybrid nanowire developed by Pulickel Ajayan’s group at RPI and Rice University [Ou, 2006]. Attaching one of these hybrid nanowires to the end of a robust tip, for example an electrochemically etched W wire, allows for its use as a TERS probe. Details of the fabrication process, hybrid nanowire attachment to a sharpened W wire, and SEM characterization can be seen in Appendix A. Work on determining the stability of these tips still needs to be ironed out before enhancement and resolution characterization studies can be conducted. However, the probe radii are on the order of 10-15 nm, and thousands can be produced at a time. These tips are a potentially viable option for highly repeatable, high resolution TERS measurements for FEOL applications.

2.4.2 Stress Measurements Compared with Finite Element Simulation

In order to be used effectively for measuring stresses in FEOL devices, TERS must satisfy two criteria. The first is it must have high resolution, on the order of 10’s of nm to measure stresses on the length scales of the devices. This potential has been discussed in the previous section. The second criterion is that it must provide accurate probing of localized stresses. Looking at Figure 2.13, it is observed that the localized stresses measured from the differential signal have been validated with FE modeling, see Section 2.2.4 for model description.

The resulting stress profile is geometrically driven due to processing of the mesas. In the FE model, a strained Si mesa was given an initial, biaxial tensile stress of 2.63
GPa, which was uniform throughout the volume of the mesa. This value was calculated from micro-Raman measurements on blanket sSOI, where $\Delta \omega_{\text{Si}}$ was measured as $-6.06 \pm 0.02$ cm$^{-1}$, and converted to stress using the biaxial approximation. After undergoing processing, the resulting mesa is observed to have a biaxial stress of $\sim 2.1$ GPa near the center, with significant process induced relaxation of up to 67% at the edges. Comparing the mesa stress with the initial blanket stress it is evident that the mesa has been relaxed by $\sim 21\%$ at the center. Considering the uniformity of the initial plane stress throughout the mesa in the FE model, this relaxation is determined to be purely geometrical in nature. A similar effect has been observed in literature, where bending at the edges of Si/SiGe/Si mesas on SiO$_2$ due to compression from the SiGe layer led to overall relaxation at the center of the mesa [CHIANG, 2005]. Furthermore, the relaxation observed coincides with $\sim 3$ nm change in the strained Si layer thickness. AFM shows the strained Si mesa height to be $\sim 60$ nm, while the SOITEC specification for the strained Si layer of the blanket wafer is $\sim 63$ nm. Based on the close correlation between the measurements and simulation, the thinning of the layer is not considered to have a significant role in the stress relaxation. Therefore, it is concluded that the local stress changes at the edges of the mesa as well as overall relaxation is geometrically driven.

Differential TERS has proven capable of measuring local stress changes across a strained SOI mesa. Not only does it provide resolution beyond the diffraction limit of light, but it also provides accurate, localized stress measurements.
2.5 CONCLUSIONS

A differential TERS scanning technique was employed to investigate efficacy of an apertureless NSOM approach for high-resolution stress characterization for potential application in FEOL stress metrology. Implementation was conducted using commercially available Raman (Renishaw) and apertureless NSOM (Nanonics) systems.

Strained SOI samples provided by SOITEC were characterized with micro-Raman measurements to confirm initial, biaxial tensile strain in the strained Si layer. Results were comparable to SOITEC specifications. After the strained Si layer was patterned into ~ 2 µm mesa structures, SEM and AFM were used to characterize the test structure topography and strained Si layer thickness.

Commercially available (Nanonics) apertureless NSOM probes were characterized for enhancement on blanket sSOI samples. The differential strained Si Raman peak intensity was used to define raw enhancement. A tip displaying ~ 50% raw enhancement was used for TERS profiling of a sSOI mesa. By implementing a Gaussian step function fit to the resulting differential intensity profile, the effective probe resolution was found to be ~ 210 nm. Though the resolution is still relatively comparable to UV Raman spectroscopy, it is below the diffraction limit for the 514.5 nm excitation wavelength. Furthermore, this resolution is comparable to the radius of the probe tip, as determined from SEM imaging. This provides confirmation of the probe size dependence on the near-field resolution. Therefore, manufacturing a probe with radius < 10 nm should result in < 10 nm optical resolution. Probe size dependent resolution is an important
result and supports the potential application of TERS for FEOL channel region stress metrology.

A stress profile was extracted from the differential, near-field TERS signal. The result shows geometric relaxation in the patterned structure, as compared to the blanket strained SOI layer. Relaxation at the edges can be substantial. Comparison of the near-field determined localized stresses with a static finite element model (COMSOL) yields good agreement. In addition, the relaxation at the mesa edges occurs over several hundred nanometers. TERS provides the ability to measure localized stress changes on a sub diffraction limited scale.

TERS has been implemented on patterned sSOI test structures for evaluating its resolution and localized stress measurement capability. Resolution and localized stress changes were found to be on the order of the size of the apertureless NSOM tip. FE modeling shows good agreement with the measured stress profile. Therefore, FEOL stress measurements using TERS is shown to be a viable technique.

FUTURE WORK

The main limitations lie with the strong near-field dependence of the apertureless probes. First and foremost is manufacturing a tip with diameters < 20 nm, for implementation on ½ pitch length scales of current and future semiconductor technology nodes. Fabrication of small diameter tips is not so much the problem, but rather the variability. Repeatable near-field enhancement requires probes with highly repeatable geometries. Variations in geometry and structure can lead to strong variations in the near-field signal. Therefore, a novel probe design has been demonstrated, with a
template fabrication method that can yield highly uniform probe geometries on a large scale (see Appendix A). Probe diameters are on the order of 20 nm, thus TERS resolution would be expected to be on this order of magnitude. Mechanical stability of these probes is still an issue that needs to be overcome before implementing a testing regime.

The differential TERS approach has proven to be successful; however, by its nature it takes over twice the amount of time to collect a profile as it would if near-field extraction would be accomplished without the need to retract the tip during scanning. It has been shown that the near-field scattered light near a small particle at the end of a quartz tip is partially depolarized [POBORCHII, 2005]. This opens the possibility of far-field polarization rejection, which has been shown to be successful in suppressing the far-field signal [POBORCHII, 2005; MEHTANI 2005]. Implementing such a technique would allow for collection times less than half of those for a differential scanning approach, making the technique more useful for characterizing multiple FEOL devices for a given time.

Recommendations for future work include the following:

- Making a mechanically robust apertureless NSOM probe using the novel hybrid-nanowire structures fabricated as in Appendix A. This involves an attachment scheme that results in stable feedback when the probe is approached to the sample surface. Once stability is reached, an approach similar to the one taken in this dissertation can be implemented on the sSOI mesa structures for comparison. Then, using structures with sizes on the order of the hybrid nanowire diameter (~ 20 nm) can be investigated to evaluate the resolution. Repeating with multiple tips will give
an idea of variability in the template fabrication method and its viability for large-scale, apertureless probe manufacturing.

- Investigate possible polarization rejection to eliminate the far-field contribution without the need for tip retraction. Polarization control is necessary for such an approach. This would include investigation with both commercially available tips as well as the hybrid-nanowire probe design to determine whether probe geometry has a large effect.

- Apply small-diameter, apertureless probe TERS approach to measure stress in the strained Si channel of a MOSFET. Correlate the measured stress with electrical characterization to determine an effective mobility, which can be compared to computer simulation.
3.1 INTRODUCTION TO MECHANICAL STRESSES NEAR COPPER THROUGH-SILICON-VIAS

The introduction of Cu TSVs into the Si substrate may lead to residual stresses in the surrounding Si. Thermo-mechanical stresses associated with Cu TSVs arise as a result of the mismatch of the coefficients of thermal expansion (CTE) between Cu and Si. As elaborated in Section 1.1, stresses in Si can lead to changes in device mobility, which in this case could cause undesirable variations in performance [OKO 2008a; KARMARKAR, 2009; MERCHA, 2010]. If sufficient stresses are present, TSV reliability may also become a concern [RANGANATHAN, 2008; LU, 2010; KARMARKAR, 2010; LADANI, 2010; RYU, 2011]. Delamination at the Cu liner or liner/Si interface is a potential issue, as is Cu extrusion (sometimes referred to as ‘popping’ or ‘pumping’). Therefore, it is necessary to understand the evolution of the residual stresses in Si around Cu TSVs throughout a 3D interconnect process flow. Far-field micro-Raman spectroscopy is well-suited to measure and characterize these stresses since it provides a spatial resolution on the order of 1 µm and is sensitive enough to measure stresses on the order of 10 MPa. It has been demonstrated as a viable technique for stress measurements in Si surrounding Cu TSVs [OKO, 2008a; TRIGG, 2010]. Furthermore, finite element modeling (FEM) has proven to be a useful and necessary tool for modeling stress evolution in both TSVs and the surrounding Si [OKO, 2008a; KARMARKAR, 2009, 2010; SELVANAYAGAM, 2009]. It is fast, reasonably accurate, and enables large-scale simulations. Combining both Raman measurements with FEM provides a robust
platform for characterizing TSV-induced stresses in Si and can yield valuable insight into emerging models for stress determination.

This dissertation focuses solely on Raman measurements of TSV-induced stresses in Si for several wafers that have undergone various processing steps in a 3D IC process flow. The wafer details will be given in Section 3.2, followed by the experimental results in Section 3.3. Cross-sections and AFM measurements were performed to observe possible void formation and grain growth in the Cu TSVs, as well as to characterize Cu protrusion as a result of thermally induced stresses in the TSV structures. The FEM results given in Section 3.4 are used to assist in interpreting the Raman measurements and provide insights into the relative symmetry of the TSV-induced stress in Si (e.g. biaxial, uniaxial etc...). A short discussion will follow presentation of those results. For a full description of the modeling, see the M.S. thesis of Benjamin Backes (Finite Element Modeling of Thermomechanical Stress in Silicon Induced by Copper Through-Silicon Vias (TSVs) for 3D Integrated Circuits, 2011). The first section will discuss the thermo-mechanical stress development in the Si-Cu TSV system as a result of the CTE mismatch. It is meant to provide the framework for understanding the origin of the residual stresses measured in Si, as well as possible mechanisms taking place in the Cu that could lead to stress evolution in the surrounding substrate.

3.1.1 Thermally Induced Stresses in the Si/Cu-TSV System

When a material is heated it will expand by an amount determined by its coefficient of thermal expansion (CTE), a material constant. Assuming a uniform temperature
distribution and a material suspended in free space, the degree of linear expansion is given by \( \alpha \Delta T \), where \( \alpha \) is the CTE and \( \Delta T \) is the change in temperature (Figure 3.1a). Upon cooling to its initial temperature, the material will return to its original state.

If that material is now bounded by another material at zero initial stress, with a different CTE, then, when a uniform temperature is applied to the system, each material will expand per its own CTE, as in Figure 3.1b. However, due to confinement, and the opposing expansion at that boundary, the two bodies will push against each other leading to nonzero stresses in both materials when thermal and mechanical equilibrium is reached. Those stresses are a result of the mismatch in CTEs. Assuming material A is an infinitely long cylinder inside an infinite matrix (material B) and a uniform temperature change, the stresses can be given in terms of a plane strain approximation [LU, 1991; LU, 2009]:

\[
\sigma_c^r = \sigma_c^\theta = -B_c \Delta \alpha \Delta T
\]

\[
\sigma_m^r = -\sigma_m^\theta = -\frac{B_m \Delta \alpha \Delta T}{2} \left( \frac{R}{r} \right)^2
\]

where \( c \) and \( m \) denote the cylinder and matrix, respectively; \( B \) is the biaxial elastic modulus of the material; \( r \) denotes the radial direction (or distance away from the cylinder in Eqn. (3.2)); \( \theta \) is the direction tangential to the cylinder; and \( R \) is the radius of the cylinder. From Eqn. (3.2) it becomes clear that the thermally induced stresses are at a maximum near the boundary due to the CTE mismatch, but dissipate at larger radial distances. For two elastic materials, removal of the temperature load returning the system to its initial temperature state will restore the zero stress state.
The behavior would be modified in the case that one or both materials were to undergo plastic deformation. Such deformations are discussed in the following section.

![Figure 3.1](image)

**Figure 3.1.** Schematic illustrating 2D thermal expansion of (a) an isolated body, and (b) a body bound by another material. Black arrows indicate expansion directions of material A, while red arrows indicate expansion directions of material B. A difference in expansion will lead to a build-up of stress in both materials. A uniform temperature distribution throughout the system is assumed.

### 3.1.2 Plastic Deformation

The response of a material to an applied load can be summarized in the form of a stress-strain curve. An example curve for an ideal material is seen in Figure 3.2a. The first part of the curve is the elastic response of the material. For low strains the stress-strain relation in this region follows Hooke’s law, as in Eqn. (1.14). The slope represents the stiffness of the material by way of a material modulus (e.g. Young’s modulus, \( E \)) given as the tensorial elastic constants, \( C_{ijkl} \), in Eqn. (1.14). Once a certain level of stress is reached, often referred to as the yield stress, \( \sigma_{ys} \), the material deviates from this elastic...
behavior and undergoes plastic deformation. Plastic flow occurs at this point, where the microstructure attempts to accommodate the additional strain in the form of atomic rearrangement and dislocation propagation. A body in the deformed state has a higher free energy, and therefore is thermodynamically unstable [KOVÁCS, 1973]. Therefore, the microstructural changes are a way for the material to accommodate the additional stresses in the form of relaxation. Since this rearrangement is driven by atoms trying to reach new equilibrium positions in the presence of these stresses, plastic deformation causes a certain amount of permanent strain after complete release of the external load, which is termed the residual strain. This is the point where the released load curve crosses back over the zero-strain axis in Figure 3.2a (dashed line).

In the ideal case, a material is plastically deformed such that strain continues to increase without the need for increasing stress. However, real metals require additional stress to maintain the increasing strain in the plastic regime, as illustrated in Figure 3.2b. The curves in plot (b) are taken from NICOLA (2006), where both experimental (dashed lines) and theoretical (solid lines) stress-strain responses for Cu films of varying thickness (1 µm, 1.9 µm, 4.2 µm) are displayed. As seen, once the yield stress is reached, plastic deformation occurs; however, the increase in strain occurs with a certain degree of additional stress. This behavior is known as strain hardening. Dislocations propagate and continue to build up in the material, thus becoming obstacles for additional dislocation propagation [NIX, 1998]. Therefore, to overcome the increased dislocation density, additional stress must be applied to cause further plastic flow, as represented by the slope of the stress-strain curve in the plastic regime of a real metal.
Note the non-abrupt transition from the elastic regime to the plastic regime. Before the yield stress is reached, the curve starts to show non-elastic behavior. This is attributed to crystal glide [KOVACS, 1973]. Since glide planes are randomly distributed, and there are 24 for fcc crystals (e.g. Cu), some are bound to be in orientations such that the applied internal stresses will create small slips before macroscopic plastic flow occurs.

Another plastic deformation mechanism, known as creep, may also be importance. It can contribute to non-elastic behavior below the yield stress. In plastic materials, irreversible deformation can begin as a function of the time the external load is applied.
Therefore, external stress loads below the yield stress, if applied for sufficient time, can begin to cause plastic deformation. However, creep occurs mostly at high temperatures where grain-boundary and surface diffusion play an important role in stress relaxation due to Cu migration. For example, Cu grain boundary diffusion is found to be a dominant plastic deformation mechanism in Cu damascene lines at high temperatures (>200 °C) [THOULESS, 1993; KOBRINSKY, 2001]. It also appears to be a dominant mechanism for Cu TSV plating chemistries, as reported recently by OKORO, et al. (2011). At lower temperatures, the dominant plastic deformation mechanism is expected to be dislocation motion.

A final consideration to irreversible Cu deformation is grain growth. It has been shown recently, that grain growth in Cu TSV electrodeposits can occur upon annealing [OKORO, 2008b, 2011]. However, this appears to be strongly dependent upon the electroplating chemistry. More important is grain growth that occurs at room temperature after plating. Electroplated Cu is unstable at room temperature, leading to recrystallization in the form of grain growth [BRONGERSMA, 2005]. This process is known as self-annealing and occurs via grain boundary annihilation. In Cu damascene lines, deposited grains are on the order of 50 nm and can grow to the order of microns [GEISLER, 2005]. It has been shown to occur within hours of electroplating [RITZDORF, 1998; LAGRANGE, 2000]. The room temperature Cu grain growth induced stress, \( \sigma_{gr} \), is given by [PAIK, 2004]:

\[
\sigma_{gr} = -E_{eff} \delta_{gb} \left( \frac{1}{D_1} - \frac{1}{D_0} \right)
\]  

(3.3)
where $E_{\text{eff}}$ is the effective elastic modulus for anisotropic Cu (estimated), $\delta_{gb}$ is the free grain boundary volume, and $D_0$ and $D_1$ are the grain sizes before and after growth, respectively. Therefore, it is determined that if the initial electroplated Cu stress is compressive, it becomes tensile, but if the initial stress is tensile, then it will become more tensile [Huang, 2010]. Typical stress values for as-deposited Cu after annealing are on the order of several hundred MPa [Keller, 1998; Geisler, 2005; Okoro, 2011]. However, this is highly dependent upon plating conditions, especially chemistry, as well as Cu thickness or via size.

After electroplating, Cu is expected to have a stress as determined by self-annealing due to grain growth. This is expected to be dependent upon plating chemistry. Furthermore, upon plastic deformation, Cu is expected to irreversibly deform, inducing stress changes that will affect the nearby Si.

3.1.3 Residual Stresses in Si Near Cu TSVs

Considering the system in Figure 3.1b, if material A is taken as Cu and material B as Si, thermal stresses will arise due to the large CTE mismatch. The corresponding CTEs are 16.7 ppm/°C and 2.3 ppm/°C, respectively [Gonzalez, 2005]. Due to the higher CTE of Cu, its expansion upon annealing will be constricted by the surrounding Si matrix, leading to a net compressive stress in the TSV. When those stresses reach the Cu yield stress, the TSV undergoes plastic deformation. From Eqn. (2.1), and assuming material properties based on Gonzalez (2005), the differential temperature at which Cu will reach its yield stress, ~172 MPa, assuming the mentioned CTE mismatches between
Cu and Si and an effective Young’s modulus of 117 MPa, is $\Delta T \approx 100$ °C. This is based on elastic behavior, so an induced temperature difference lower than ~100 °C is expected to induce a reversible, elastic deformation. However, most temperature processes that the wafer will see during production are at higher temperatures; therefore, plastic deformation must be taken into account. In addition, the initial Cu stress due to self-annealing must be accounted for in the analysis of the residual stresses in Cu. Since Cu will be expected to be in an intrinsic stressed state due to the irreversible deformations of self-annealing, it is expected that the surrounding Si will likewise be under residual stress as a result of the Cu deformation (excluding Cu delamination).

As a result, Eqn. (3.2), which is based on a single temperature step and does not include plastic deformation mechanisms must be modified. A modified equation that takes into account self-annealing as well as plastic deformation would be [KELLER, 1999]:

$$\sigma_{high} = \sigma_0 + \sigma(T) - \sigma_{relax}$$

(3.4)

where $\sigma_0$ represents the stress at room temperature (initially due to self-annealing), $\sigma(T)$ is essentially Eqn. (3.2) and represents the stress induced due to the CTE mismatch, and $\sigma_{relax}$ represents the relaxation due to plastic deformation. This accounts for the stress at high temperature; however, the stress at room temperature after a complete anneal step would be more accurately represented by:

$$\sigma_{residual} = \sigma_{high} + \sigma'(\Delta T) - \sigma'_{relax}$$

(3.5)

where $\sigma'(\Delta T)$ and $\sigma'_{relax}$ represent the CTE mismatch induced stress and plastic deformation stress that occur upon cooling. Solving for additional anneal steps would
involve an iterative approach, where the previous total stress would become the initial stress for the following anneal cycle. The solution of such an approach is non-trivial due to the complexities of the deformation mechanisms and irreversible processes that occur throughout each step of the anneal cycle.

Based on nano-indentation results on cross-section TSVs, it has been shown that the effective Young’s modulus of a Cu TSV may, in some cases, vary based on the anneal history [OKORO, 2008b, 2011]. Though, measurement errors show that variation may not be statistically significant. Therefore, in the case that the elastic modulus of Cu does not change significantly upon annealing, $\sigma'(\Delta T)$ would be approximately equal to $-\sigma(T)$ due to the fact that all variables would be constant except for the opposing direction of the temperature differential (see Eqn. (3.1)). In such a case, by combining (3.5) and (3.4), under the assumption that the Cu elastic modulus does not change significantly upon annealing, the residual stress can be rewritten as:

$$\sigma_{\text{residual}} = \sigma_0 - \sigma_{\text{relax}} - \sigma_{\text{relax}}'$$  \hspace{1cm} (3.6)

From this it becomes clear that the residual stress in Cu, and, hence, the surrounding Si, will depend predominantly upon irreversible plastic deformation that occurs in the Cu TSV. Though the stress due to the CTE mismatch is not included, it must be pointed out that the plastic deformation contributions are functions of the CTE mismatch induced stresses, $\sigma(T)$. It is important to note that Eqn. (3.5) is the general case, while Eqn. (3.6) is an approximation based on small changes in Cu elastic modulus due to grain growth upon additional thermal annealing, which would be dependent upon electroplating and Cu chemistries.
Finite element modeling has proven successful in simulating the residual stresses in Si surrounding 5 μm Cu TSVS after complete annealing cycles, as will be discussed in more detail in Section 3.4. Micro-Raman spectroscopy has been successful in directly measuring the stress states in Si for various annealing conditions and wafers that have undergone various degrees of processing. These results are presented in Section 3.3.
3.2 EXPERIMENTAL APPROACH AND SETUP

3.2.1 Experimental Objective and Approach

Micro-Raman spectroscopy is employed to measure residual stresses in near-surface Si surrounding Cu TSVs to characterize the stress field and its evolution with processing. A top-down scanning approach is taken to characterize stresses on wafers that have undergone varying degrees of processing to characterize process-induced effects based on thermal history. Since the residual stresses can be thermally induced (due to the large CTE mismatch between Cu and Si) understanding the stress evolution as a function of temperature is critical. Furthermore, Cu self-annealing is expected to play a role in the intrinsic TSV stress. Consequently, studies were performed with wafers that had seen a variety of post-Cu plate thermal annealing. In addition, atomic force microscope (AFM) topography profiles and focused ion beam (FIB) cross-sections of annealed samples were taken to investigate the nature and possible role of Cu protrusion from the TSV free-surface, void formation and migration, and, to a modest degree, Cu grain variation. Finally, a cross-sectional Raman mapping approach was investigated to characterize residual stresses in Si as a function of depth.

Finite element modeling undertaken in conjunction with Raman measurements [Backes, 2011] helps interpretation and qualification of the measurements. Full details will not be given in this dissertation; however, a relevant discussion of the results and comparison with measurements is given in Section 3.4.

This section describes the sample details, instrumentation, and experimental approaches used.
3.2.2 Sample Details with Relevant Process History

All wafers used in this thesis were provided by SEMATECH. Copper TSV test structures were incorporated into the process flow for the purpose of measuring Si stress profiles near well-defined and mechanically isolated structures with micro-Raman microscopy. The test structures include isolated and arrayed round Cu TSVs, which were laid out at relatively large separations (>100 µm) from nearby TSV structures. This avoided influence of stresses from structures not under investigation. The TSV arrangements measured in this dissertation are illustrated in Figure 3.3. Circles in the figure represent TSVs from a top-down view. The lines illustrate typical Raman scan directions for a majority of the measurements in this dissertation. Line scans were also collected perpendicular to those illustrated for each array size. In addition, scans were performed on isolated TSVs in directions 45° and 135° rotated from those illustrated. Isolated TSVs provide a fundamental understanding of the thermally induced residual stresses in the surrounding Si. Linear, 1x4 TSV arrays give insight into stress superposition effects along the array axis, while maintaining a stress behavior that resembles an isolated TSV perpendicular to the array axis. 2-dimensional, 8×8 TSV arrays provide insight into symmetric stress superposition effects on a larger scale. All arrays observed have a pitch to TSV diameter ratio of 2:1.
Experimental results presented in this dissertation were collected on four (001) Si wafers with notches oriented along <110>, designated as wafers A, B, C, and X. All four have different process histories. The following will describe the relevant process histories for each wafer.

Wafer X was used for preliminary micro-Raman measurements and did not contain the round, Cu TSV test structures already mentioned. Initial measurements were taken on square TSVs with nominal dimensions of 5×5×25 µm, with the last dimension being the depth. The relevant process history is: Si reactive ion etch (RIE); high aspect ratio process (HARP) SiO₂ liner; liner/barrier/Cu seed; Cu electrochemical deposition (ECD); 150 °C anneal; removal of Cu overburden via chemical mechanical planarization (CMP).
After CMP, the SiO\textsubscript{2} thickness on the top surface was \( \approx 1.4 \mu m \). The processes up through Cu plate are the same for each of the following wafers. The subsequent processes, however, are different.

Wafers A and B incorporated the round, Cu TSV test structures. All test structures from these wafers had nominal dimensions of 5 \( \mu m \) diameter by 25 \( \mu m \) depth. Wafer A underwent Cu plate, followed by CMP then a 150 °C anneal. Wafer B underwent Cu Plate, CMP, 350 °C furnace anneal, CMP, metal 1 (M1) oxide deposition [SiCN (nBlock) / tetraethyl orthosilicate (TEOS)] at 400 °C, resist stack deposition. Note that wafer A was only subjected to a 150 °C post-plate anneal. Wafer B was subjected to higher temperature steps through M1 processing. Comparisons of measurements taken from these wafers provide insight into stress evolution over multiple process steps.

Wafer C incorporated new, round Cu TSV dimensions with 5 \( \mu m \) diameter by 50 \( \mu m \) depth. This wafer was only processed through Cu plate and post-plate CMP to remove overburden. No thermal anneals were performed; therefore, it will provide insight into the intrinsic residual stress in Si due to Cu self-annealing.

3.2.3 Micro-Raman Scanning Details and Calibration

Several initial measurements involved 2D mapping of the stress around isolated Cu TSVs; however, non-linear drift was observed over the long collection times (>12 hrs.) Therefore, a majority of the measurements taken are line scans along [110] to decrease the scan time and minimize drift. Due to Si anisotropy, as will be seen in Section 3.3, stresses along <110> directions are higher than along <100> directions. Due to
symmetry, the highest stresses are found along directions in line with the center of the TSVs; therefore, line scans were performed along the TSV center axis. In this dissertation, line scans will also be referred to as profiles.

Experimentally acquired Raman scan profiles represent an average of 10 repeated line scans with error bars representing one standard deviation. The standard deviation for all measurements is ~0.02 cm$^{-1}$, which corresponds to a biaxial stress of ~9 MPa. This is on the order of the Raman spectrometer’s measurement error. The Raman shifts measured are close to the order of the error; therefore, calibration is critical.

A careful calibration approach is taken for all measurements. First, the spectrometer is calibrated to a Si standard provided by the Raman spectrometer manufacturer. This is performed before each use of Raman system. Before and after each line scan, a spectrum is collected in a Si region >40 µm from the test structures of interest. This region was confirmed to be stress free in advance by comparison of the Raman shift at that region with the Si standard. The Raman shift at this “stress-free region” obtained before and after each line scan was then used to account for any drift that may occur over the collection time (~20 minutes for a single line scan) by providing a baseline subtraction of the drift in that peak shift. In addition, this baseline subtraction approach allows for calculation of the Si Raman peak shift ($\Delta\omega_{\text{Si}}$) from the stress free Si position, which determines the amount of residual stress in the Si, as outlined in Section 1.2.
3.2.4 Micro-Raman Spectrometer Setup

The Raman spectrometer used for these studies is a HORIBA Jobin-Yvon LabRAM HR 800 UV system. A schematic can be found in Figure 3.4. Laser excitation sources with wavelengths of 325 nm (HeCd), 532 nm (Solid-State Diode Pumped), and 633 nm (HeNe) are available. This allows for measurements at different depths in Si, as outlined in Section 3.3.7. Besides the multi-wavelength measurements performed in that section, the rest of the measurements were performed with 532 nm excitation, which has a penetration depth of ~1.1 µm in Si, as calculated by the method in DE WOLF, 1996. Power levels were maintained below ~10 mW by utilizing neutral density (ND) filters to minimize laser-induced sample heating that could cause artificial shifting of Si Raman peaks. A 100× objective with a numerical aperture (N.A.) of 0.9 provides a nominal spot size of ~1 µm. The Raman scattered light is dispersed on a Peltier cooled, 1024x256 pixel CCD detector by a 2400 line/mm diffraction grating. After Lorentzian peak fitting, the experimental spectral resolution is found to be ~0.02 cm⁻¹.

Raman mapping is performed at room temperature by scanning the sample with a PI P-562 piezo-scanning stage with 1 nm step size resolution. Total scanning distances are 100 µm in x and y, and 20 µm in z. Due to the spot size of ~1 µm, and considering stress variations in the samples do not vary rapidly, the step size chosen for scans was 1 µm. This was consistent for all scans.
3.2.5 Thermal Anneal Setup and In-Situ Raman Measurements

Thermal annealing was performed with a Linkam THMS600 heating stage, shown in Figure 3.4b. This stage integrates directly onto the microscope stage and provides full control of annealing profiles. It has $x$ and $y$ positioning capability and a glass window on
top to allow for in-situ measurements. A 50× long-working distance (N.A. 0.5) objective is required to collect measurements with the Linkam stage in place. Since the piezo-stage does not integrate onto the system at the same time as the heating stage, line profiles are collected with the DuoScan feature. This feature consists of two mirrors in the optical path of the microscope that can deflect the laser beam in both $x$ and $y$ to allow for scanning in these directions. However, due to aberrations when the beam is deflected toward the outer edges of the objective lens, this feature is limited to scan lengths of ~35 µm. Therefore, it provides limited use in the case of line scans on the arrayed structures of interest. With the exception of one set of measurements, all measurements are taken at room temperature after annealing. Samples were allowed to cool for a sufficient amount of time (> 1 hr.) before performing Raman line scans.

Thermal anneal measurements were conducted on wafers A and C. Similar measurements were not undertaken on wafer B due to its complex thermal history. 1×4 linear Cu TSV arrays were selected to observe residual stress superposition changes over larger arrays, in part, to minimize collection time for efficiency. For all samples, the Raman profile collection methodology was the same: die were randomly selected for the specific wafer; as-received Raman line profiles were collected; each die was annealed at a different temperature over 50 °C increments, as specified in Table 3.1; temperatures were ramped at 50 °C/min; after sufficient cooling to room temperature, Raman line profiles were collected; post-anneal AFM mapping was conducted on a Veeco Dimension 3100 with the same tip.
<table>
<thead>
<tr>
<th>Wafer</th>
<th>Temperature Range (°C)</th>
<th>Time (minutes)</th>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>150 - 400</td>
<td>210</td>
<td>Air</td>
</tr>
<tr>
<td>A</td>
<td>150 - 400</td>
<td>25</td>
<td>Air</td>
</tr>
<tr>
<td>A</td>
<td>150 - 400</td>
<td>25</td>
<td>N₂</td>
</tr>
<tr>
<td>C</td>
<td>100 - 400</td>
<td>25</td>
<td>N₂</td>
</tr>
</tbody>
</table>

Table 3.1. Summary of the samples and conditions of the various anneal studies conducted. The temperature range was divided into 50 °C increments, with a separate die annealed at each temperature.

The first set of anneal measurements were conducted for six die randomly selected across wafer A. A temperature range of 150-400 °C was selected for annealing. This range was chosen since it covers both the low and high temperature ranges typically used for post-Cu plate anneals and/or subsequent processing. The temperatures between this range were in 50 °C increments (as with all annealing studies). These anneals were conducted in air. Since in-situ Raman measurements were taken at each corresponding anneal temperature, and each collection of 10 line scans takes ~3.5 hours, this was the nominal anneal time. Raman profiles for these samples were conducted both at anneal temperature and room temperature (pre/post anneal).

After both Raman measurement and AFM mapping was concluded, these samples were FIB cross-sectioned to observe possible void formation and grain growth. The samples were first cleaved and mechanically polished to within 10 μm of the 1x4 arrays. This facilitated quick cross sectioning by FIB. FIB milling, and subsequent ion and electron beam imaging, was conducted by Dr. Thomas Murray (CNSE) on a FEI Nova 600 Nanolab dual-beam system.
A second set of annealing and Raman measurements was conducted for six more die randomly selected from wafer A. Annealing was performed exactly the same as with the first set, with the only exception being a shorter anneal time of 25 minutes, excluding ramp time. This time was selected to compare shorter anneal times but also to imitate a more realistic time in a fabrication setting.

A final set of anneals on wafer A were conducted in N\textsubscript{2}. Concern of Cu protrusion heights resulting from Cu oxidation during the anneals resulted in a repeat of the 25 minute anneals just described with the notable exception of annealing in an inert environment to limit or eliminate Cu oxidation. A N\textsubscript{2} line was connected to the Linkam stage via an available gas inlet. The gas outlet was left open so that N\textsubscript{2} could escape. Before each die was annealed, N\textsubscript{2} was allowed to flow through the Linkam heating stage chamber for 30 minutes at a rate of \(~ 20\) cc/min. With an estimated chamber volume of 234 cc, this would effectively purge the O\textsubscript{2} within 12 minutes. The 30 minute purge time was chosen to be conservative. Subsequent to this purge step, annealing was performed.

The last set of anneal studies were performed on wafer C (5×50 µm Cu TSVs with no post-Cu plate anneal). Seven random die were chosen from this wafer for annealing. The temperature range was extended to 100-400 °C. The addition of one sample annealed at a lower temperature (100 °C) was to investigate the effect of lower temperatures on TSVs not subjected to a post-Cu plate anneal. All annealing was performed in N\textsubscript{2} with the same experimental details as the previous set of samples (wafer A—N\textsubscript{2} anneals).
3.2.6 **Raman Depth Profile Measurements**

Depth profiling was carried out via a cross-sectional Raman measurement approach. Cross-section preparation was conducted by cleaving specified die and mechanically polishing to a distance ~ 10 µm from 1×4 Cu TSV arrays. Dr. Thomas Murray at CNSE performed the cross-sections on the FEI Nova 600 Nanolab dual-beam system, as noted above. Two as-received die were randomly selected from each of wafers A (5×25 µm) and C (5×50 µm). For each wafer, one die was cross-sectioned along a <110> direction through the center of the 1×4 array (on the array axis). The second die was cross-sectioned along a <110> direction ~ 2-4 µm from the 1×4 array. This effectively leaves a 2-4 µm buffer layer of Si. Since the stresses measured are highest closer to the arrays, the Si buffer layer should display smaller stresses.

Raman line scans were performed using 532 nm wavelength excitation with the same experimental parameters as the top-down Raman measurements. The cross-section plane is {110} and is positioned facing up to allow line profiles along that plane. The geometry is seen in Figure 3.34. As seen in Table 1.2, backscattering from a {110} plane only allows access to the z-polarized phonon. So this setup allows correlation between stresses measured near the top of the TSV to the top-down measurements. Two line profiles are collected for each die along [001], one is taken between the center two TSVs and the other is taken ~ 3 µm from edge of the outer TSV, as illustrated by the black arrows in Figure 3.34. Due to stress superposition effects, there may be differences in the corresponding profiles.
To calibrate to a stress free Si region in the cross-sectional geometry, FIB milling was taken at a depth $\sim 2 \times$ the depth of the TSVs, respectively. These regions were confirmed to be stress free by comparison to a Si standard. The same calibration and stress free baseline protocol was utilized for these scans as with the top-down measurements. The FIB cuts left some topography, which was noticeable by eye. Initial measurements showed noticeable intensity variations along these topography changes. Employing the AutoFocus feature on the Raman system mitigated these variations. The AutoFocus feature scans the sample in the $z$ direction over a specified distance until a maximum Raman intensity is obtained. It does this at each point in the line profile before collecting a spectrum.
3.3 RESULTS

3.3.1 Raman Characterization of TSV-induced Stress in Si: Geometric Effects

Initial investigations of TSV-induced stress in Si included comparative surveys of TSV shapes (square vs. round) and etch-induced stress in Si (pre-fill). These measurements were used to harden data acquisition and drift-correction protocols as well as elucidate various features of TSV-induced stress fields in Si.

Preliminary Raman mapping measurements of square TSVs were conducted on 5 µm × 5µm Cu TSVs from wafer X. The collected Si spectra were fitted with Lorentzian peak functions to extract the Si Raman peak positions. These Raman peak positions were then subtracted from the relaxed Si peak position (taken at a distance of 40 µm away from the TSV of interest) to determine the Si Raman peak shift and the amount of stress present, as detailed in Section 1.2. Figure 3.5a displays a map of the Si Raman peak shift surrounding a 5 µm square Cu TSV. Positive Raman shifts (green, yellow, red), which represent compressive stresses, are observed within ~2 µm of the Cu TSV. Negative Raman shifts (blue), representing tensile stresses, are observed at greater distances (> 2 µm) from the TSV faces. A four-fold symmetrical Si Raman peak distribution around the TSV mirrors the TSV shape, but may also be influenced by the cubic anisotropy in the Si elastic stiffness matrix. (See Section 3.4) An overall slope in the average Si Raman peak position is observed along the y-axes in Figure 3.5, which is due to system drift over the long collection time (~8 hrs).

Based on symmetry of the Si Raman peak distribution, as well as the TSV, a biaxial stress profile is assumed. This assumption has been validated (Section 3.3.2). With a
biaxial stress approximation, the Si Raman peak can be converted to stress using the linear conversion [OKORO, 2008a]:

\[
\sigma_{x-y} \ (MPa) = -434 \times \Delta \omega_{Si} \ (cm^{-1})
\] (3.7)

where \( \Delta \omega_{Si} \) is the measured Si Raman peak shift, relative to unstressed Si. The conversion to stress is plotted in Figure 3.5b. Note that the sign represents direction of the stress, where negative values represent compressive stresses, and positive values represent tensile stresses after the linear conversion. The largest compressive stresses measured are \( \sim 150 \) MPa. The largest tensile stresses measured are \( \sim 65 \) MPa, but extend farther away from the Cu TSV.

Due to high stress concentrations at the corners of square TSVs, a round TSV geometry is better suited to minimize the stress field in nearby Si. Due to these and other reasons a round TSV geometry was adopted for TSV test structures. As a result, the remainder of the Raman data presented in this dissertation was collected from round Cu TSVs.

Figure 3.6 displays a Si Raman peak shift map and corresponding, linear biaxial stress map from a 5 \( \mu m \) diameter, round Cu TSV (wafer A). A region of compressive stress is observed in the Si within \( \sim 1-2 \mu m \) of the Cu TSV. In regions further away (>2\( \mu m \)), the four-fold symmetric tensile stress distribution is observed. The presence of the four-fold symmetric stress field here cannot be attributed to the TSV geometry. Consequently, it arises solely from the Si anisotropy. Tensile stress relaxation occurs at larger distances (>15 \( \mu m \)) from the TSV in the <110> directions. The diminished
compressive stress region in the <100> directions is consistent with the orientation
dependence of Si Young’s modulus and is discussed in more detail in Section 3.4.

Figure 3.5. (a) Si Raman peak shift map surrounding a 5×5 μm square TSV. (b) Linear conversion of Si Raman peak shift to biaxial stress displays a four-fold symmetrical biaxial stress distribution with high compressive stresses concentrated near the corners of the TSV.

Figure 3.6. (a) Si Raman peak shift map surrounding an isolated, 5×25 μm Cu TSV. (b) Linear conversion of Si Raman peak displays a uniform distribution of stress. Some compression of the Si is measured near the empty TSV.
Raman measurements were conducted on unfilled, 5×25 µm TSVs following RIE etch and prior to liner deposition. The resulting Si Raman peak map and corresponding stress map are shown in Figure 3.7. A fairly uniform stress distribution is observed, disregarding system drift, with some slight compression in the Si in regions > 2 µm from the TSV edge. This may be attributable to geometrically driven stress localization at the edges of the TSV. The four-fold symmetric Si stress distribution observed near the Cu-filled TSVs is not present for an unfilled TSV, which leads to the conclusion that four-fold symmetrical residual stresses in Si result directly from the introduction of Cu in the TSV.

![Figure 3.7](image.png)

Figure 3.7. (a) Si Raman peak shift map surrounding an unfilled, 5×25 µm via. (b) Linear conversion of Si Raman peak displays a uniform distribution of stress. Some compression of the Si is measured near the empty TSV.

3.3.2. Raman Measurements on As Received Wafer A TSVs

A. Isolated TSVs
Raman measurements of TSV-induced stress in Si near isolated TSV structures were carried out. The first set of measurements conducted was on isolated, round 5×25 µm Cu TSVs on die sections from wafer A (as-received). All measurements consisted of linear scans, taken across the centers of the TSVs. Stress concentrations were found to be highest in the regions along the axis passing through the centers of the TSVs in the <110> directions. Thus, line scans along those directions across the centers permitted probing of the highest stresses induced by the TSV. Utilizing line scans also decreased the collection times required for accurate mapping, thus mitigating system drift issues similar to those encountered during acquisition of the 2D Raman maps in Section 3.3.1.

As noted above, a biaxial approximation was used to convert measured Raman shifts to stress in MPa. To validate the use of this approximation a finite element (FE) model was constructed to simulate TSV-induced stress in Si following a 350 °C thermal anneal cycle of an isolated 5×25 µm Cu TSV. The FE simulation yielded the full stress tensor at each point along the path corresponding to the Raman line profile. Figure 3.8 shows a comparison between the biaxial approximation (blue line) and the 3 eigenvalue solutions to the full secular equation (green, yellow and red lines). Based on the nearly matching profiles (within the Raman system error of ~0.02 cm⁻¹) of the secular solution (Eigenvalue 2) with the biaxial approximation, it is assumed that the stress distribution in Si surrounding a round, Cu TSV is biaxially symmetric. Eigenvalue 2 is found to be z-polarized based on the corresponding eigenvector. Due to the similarities, the linear conversion for Si Raman peak shift to stress based on the biaxial approximation is used for all remaining Raman profiles, so that the y-axis will be in terms of stress (MPa) for ease of discussion. Similar agreement was seen for line profiles along 1×4 TSV arrays.
Preliminary Raman measurements were conducted on isolated, round, 5×25 µm Cu TSVs on wafer A. Figure 3.9 illustrates four line profiles collected along the <110> and <100> directions. The dashed line represents completely relaxed, or stress-free, Si. Any features below that line represent tensile stresses, while features above it represent compressive stresses. A compressive stress region is observed in the Si within a region of ~ 2 µm of the TSV edge, beyond which a compressive to tensile crossover point is reached. Moving further away from the TSV edge, the Si is in a tensile stress state, which relaxes further away from the TSV edge (~ 15 µm). The profiles along the <100> directions (black) display a smaller maximum tensile stress than the tensile stresses in the <110> directions. Just as was seen in the 2D map of Figure 3.6, the difference in stress magnitude is due to the anisotropic nature of crystalline Si.
It is worthwhile to note at this time that the regions and general shape of this profile are typical of all the profiles observed. They are qualitatively similar to the Raman measurements conducted at IMEC [OKORO, 2008a]. All profiles are collected along the <110> directions, since these directions display the highest stresses. Variations in the magnitudes of the maximum tensile stresses will become the main points of discussion for comparison of the stress evolution between different profiles. Furthermore, all error bars are $1\sigma$ for the average of ten repeated profile measurements, unless otherwise noted.

Figure 3.9. Multi-directional Raman shift profiles across an isolated $5\times25$ µm TSV qualify the role of Si anisotropy on the residual stress profiles.
Si Raman shift measurements were conducted on five die across wafer A. The resulting profiles are displayed in Figure 3.11a, with the relative die positions on the wafer shown in the inset. The corresponding profiles all appear to fall within 10 MPa of each other, which is the approximate error of the Si Raman peak shift measurements (±10 MPa).

In order to quantitatively characterize the isolated TSV profiles, four metrics were chosen. They are illustrated schematically in Figure 3.10. (Note that the stress and the Raman shift profiles are inversely related. For ease of inspection, the stress axis is inverted in the subsequent plots. In this manner, the similarity between the experimentally acquired Raman data and corresponding converted stress is maximized.) The first metric is the maximum tensile stress ($\sigma_{\text{maxT}}$). The variation of this quantity gives an idea of the stress evolution of the Si as a result of the Cu plastic deformation. A plot of the maximum tensile stress for each die is shown in Figure 3.11b. The dashed line represents the average tensile stress for the five die (-15 ± 4 MPa). All die fall within a 95% confidence interval of the average maximum tensile stress. Therefore, die-to-die variation based on this metric is shown to be insignificant.

The other three metrics chosen were in terms of distances away from the TSV edge: 1) the compressive-tensile cross-over point, 2) maximum tensile stress point ($x_{\text{maxT}}$), and 3) the 50% tensile relaxation point. These points give an idea of changes in relaxation distances, which, in correlation with the maximum tensile stress magnitude, can be key metrics for determination of minimal separation distances between TSVs and devices (keep-away-zones). Plot (c) of Figure 3.11 displays these three metrics. Dashed lines
represent the average distances away from the TSV edge for the five die. The mean distances of the crossover point,

![Diagram](image)

Figure 3.10. Schematic illustration of points of interest in Si Raman shift profiles near a TSV. $\sigma$ denotes stress. $\Delta \omega$ denotes Si Raman shift.

maximum tensile stress point, and 50% reduction in the maximum tensile stress point are $1.4 \pm 0.2 \, \mu m$, $3.2 \pm 0.5 \, \mu m$, and $6.6 \pm 0.5 \, \mu m$, respectively. Considering the errors associated with each measurement, all die fall within a 95% confidence interval of the mean for each metric. Therefore, the die-to-die stress variation across wafer A for isolated, round $5 \times 25 \, \mu m$ Cu TSVs is determined to be statistically insignificant.
Fig 3.11. (a) Stress profiles of isolated, round, 5×25 μm Cu TSVs for 5 die on wafer A. Die-to-die variation is within the error of the measurements. Key metrics for quantitative comparison of die-to-die variation for isolated Cu TSVs include (b) maximum tensile stress and (c) measurements of 3 key points for determining the extend of the stress field away from the TSV edge.

B. Linear, 1×4 TSV Arrays

To evaluate the impact of closely spaced TSVs (1:1 pitch ratio), Raman profile measurements were continued on the same die as in the previous section for linear, round
$5 \times 25 \ \mu m$ Cu TSV arrays. The resulting profiles are displayed in Figure 3.12a. Outside of the arrays, the Si stress profiles resemble what is observed for the isolated TSVs in the previous section. However, between TSVs (inside the array) significant stress superposition effects are observed. Plot (b) shows the maximum tensile stress points on each side of the TSVs. The maximum tensile stress is observed to be at the center of the array. The profiles appear to have mirror symmetry based on this center point. Comparing the maximum tensile stress points for each die, it is determined that the die-to-die variation is not significant (all within one standard deviation). The important point, though, is that nearby TSVs can cause significant increases in stress due to overlapping stress fields.

Raman profiles were collected in the [-110] direction for the first two TSVs of the $1 \times 4$ arrays for dies 1 and 2 of wafer A (see inset of Figure 3.13). This direction is perpendicular to the direction of the array; therefore, there are no nearby TSVs in the [-110] direction. The results are plotted in Figure 3.13. These stress profiles mimic the profiles observed for an isolated, Cu TSV of the same geometry. Thus, it is concluded that nearby TSVs along the [110] axis do not have a significant impact on the stress profiles of neighboring TSVs along the [-110] axis.
Figure 3.12. (a) 1×4 TSV array (a) stress profiles and (b) maximum tensile stress points along [110] for wafer A.
Figure 3.13. Stress profiles across individual TSVs in a 1×4 TSV array along the [-110].

One concern with the measurements is whether the scans extend a sufficient distance away from the array in the [110] direction to safely assume that the end of the profile is at a stress-free position ~15 μm from the nearest TSV edge. To confirm this, extended scans were performed that measured the Si Raman peak shift profiles an extra 30 μm. Figure 3.14 displays the results, where the blue curve is the stress profile collected with a normal Raman scan, and the black curve is the extended Raman line scan. The profiles sufficiently overlap; therefore, the extended line profile confirms that the baseline assumption used for the typical Raman profile collected for the 1×4 TSV arrays is accurate. This is as expected, considering that the profile outside the 1×4 TSV array is approximately the same as that of the isolated TSV case.
Figure 3.14. Extended Raman scans on a 1×4 TSV array from wafer A were collected to confirm an accurate stress-free position.

C. 2-Dimensional, 8×8 TSV Arrays

With the linear arrays, it was seen that the superposition of stresses occurred between neighboring TSVs. Also, in the [-110] direction the stress profile mimicked that of an isolated TSV. If the number of TSVs increases along both [110] and [-110], then stress superposition effects should be observed along [-110], and the magnitude of the overall tensile stress should increase with the number of TSVs.
Raman measurements were taken on 2-dimensional, 8×8, round 5×25 µm Cu TSV arrays on the five die from wafer A used in the previous two sections. Profiles collected in the 8×8 array were along the [110] and [-110] directions through the rows and columns at the center of the array. The logic behind this was that the stress profiles would be similar along each row and profile, with only the magnitudes varying depending on position within the array. In other words, the tensile stress build up is expected to be highest near the center of the array.

Stress profiles along [110] can be seen in Figure 3.15a, with the plot of the maximum tensile points in (b). Examining the maximum tensile stress points (b), it is noted that all points are within ±10 MPa of each other (one standard deviation), leading to the conclusion that die-to-die variation of stress surrounding the round 5×25 µm Cu TSVs in an 8×8 2-dimensional arrays is not significant. It is also observed that the stress magnitudes are highest near the center of the array. Furthermore, the overall tensile stress magnitudes are higher than those observed in the 1×4 TSV arrays, approximately 48 MPa vs. 38 MPa, respectively. Therefore, the amount of stress present due to superposition effects is directly related to the number of TSVs in the array.
Figure 3.15. As received (a) stress profiles and (b) maximum tensile stress points along [110] for five die across wafer A.
Figure 3.16 shows a comparison of two profiles taken in either the [110] or [-110] directions near the center of an 8×8 TSV array. Despite some variation in the tensile stresses seen at the left hand side of the plot, the stress profile appears to be consistent. This implies symmetry in the <110> directions, which is expected based on the measurements for isolated, 5×25 μm Cu TSVs.

Figure 3.16. Comparison of scans along [110] and [-110] directions near the center of an 8×8 array (wafer A).
3.3.3. Wafer A Raman Measurements on Annealed TSVs

In order to understand the effects of annealing conditions on the evolution of the residual stress in Si, Raman profiles were collected on die from wafer A. Due to its minimal die-to-die variation, as observed in the previous section, this wafer provides a consistent platform from which to make correlations between changes in stress due to thermal annealing. Experimental details for the following measurements are found in Section 3.2.

The first set of annealing measurements involved measuring the stress profiles for each die at the anneal temperature, as well as at room temperature after the samples cooled. Since ten profiles are collected and averaged for each measurement at elevated temperature, the total time for the anneal was 3.5 hrs. To perform the measurements at the annealing temperature, a 50× long working distance objective was required in conjunction with the spectrometer’s DuoScan feature and the Linkam heating stage.

The post-anneal profiles at room temperature are shown in Figure 3.17. When compared with pre-anneal measurements in Figure 3.12, it is observed that annealing at temperatures up to 200 °C does not cause any significant change to the stress profiles. However, at anneal temperatures of 250 °C and above, there is significant tensile relaxation in the Si between the TSVs. The profiles outside the 1×4 array do not appear to have significant variation until an anneal temperature of 350 °C, where some relaxation in the tensile stress is observed. At 300 and 350 °C, the Si between the TSVs is in a net compressive state. The trend appears to be one of increasing Si compression between the TSVs with increasing anneal temperature. Only the 400 °C anneal does not
Figure 3.17. (a) Post-Anneal (air, 3.5 hrs) (a) stress and (b) AFM profiles along [110]. AFM profiles display significant Cu protrusion from the top of the TSVs.
appear to follow this trend. It may simply be due to differences in the Cu grain structure. However, the onset of compression in the Si coincides with a significant increase in the Cu TSV protrusion from the free surface above the oxide layer, as seen in Figure 3.17b. At annealing temperatures of 350 and 400 °C, Cu protrusion reaches over 1 µm above the oxide surface.

Figure 3.18 displays profiles collected with the Raman system’s DuoScan feature at the anneal temperature for each die. One obvious point to note is the lack of a compressive region next to the Cu TSV. There is variation in the magnitudes, however no obvious trend emerges with respect to temperature. The stress profiles behave more like what is observed for a single temperature ramp, rather than a complete annealing cycle. This is qualified in plots (b) and (c). Plot (b) displays the measured Raman profiles at a 250 °C anneal (red circles) and after a complete cycle back to room temperature (blue triangles). It would be assumed that at high temperatures the Si would be under a compressive stress, which contradicts the profile (red circles) in (b). It has already been shown that before annealing, the intrinsic TSV induced stress profile is tensile away from the TSV. Therefore, an additional 250 °C anneal would be expected to decrease that tensile stress, but not necessarily cause it to become compressive. This is qualitatively supported with FE simulations, as seen in plot (c), where a single temperature ramp (red line) mimics the profile in (b) at 250 °C, while the full anneal cycle (blue line) mimics what is seen during room temperature measurements.
Figure 3.18. Stress profiles (a) at anneal temperature. Comparison of (b) Raman profiles taken at 250 °C anneal (red circles) and after cooling to room temperature (blue triangles). FE simulations (c) of a single temperature ramp at 50 °C (red line) and a full 350°C temperature cycle (blue line).
Six more die from wafer A were chosen for further anneal measurements. Anneal times were decreased to 25 minutes with the idea to target a more realistic annealing time scale. Raman profile measurements were repeated at room temperature before and after annealing. Post-anneal measurements were conducted after the die had time to cool. All other experimental conditions were the same as the previous study, with the only change being the anneal time. Figure 3.19a displays the obtained stress profiles, while (b) displays the post-anneal AFM line profiles through the center of the array. Note in (a), the black dots represent an average of the pre-anneal profiles for all die (note the small error bars, which confirm profile similarity). A similar trend as the previous anneal study is observed. Relaxation of the tensile stress in the Si between the TSVs is observed with increasing anneal temperature. However, the onset of this relaxation does not occur until ~250 °C. Once again, this onset of compression corresponds to a significant increase in the Cu protrusion above the TSV free surface, as seen in Figure 3.19b. The stress evolution for these shorter anneals appears to be more monotonic with anneal temperature than the longer anneal times. A further observation is that at the higher temperatures (300-400 °C) the Si between the TSVs displays some tensile stress. These results point out the importance of annealing times on the TSV stress evolution and resultant residual stresses in Si.

Cross-sections of all 1×4 TSV arrays that were annealed for the 25 min. measurements can be seen in Figure 3.20. Only the die that underwent a 350 °C anneal displayed visible voiding in the Cu (e). No significant void formation or migration is observed for any of the die. Furthermore, the grain structures showed no significant
differences between die. This implies void formation and Cu grain growth did not play a significant role in the observed residual stress profiles in Si for these samples.

The large Cu protrusion heights seen in the previous studies were believed to have been partly due to Cu oxidation. It is well known that Cu oxidizes in the presence of air at elevated temperatures [WHITE, 1942; CAMPBELL, 1947]. Therefore, to mitigate any oxidation during annealing, all further anneals were conducted in an inert environment (N\textsubscript{2}).

Both pre- and post-anneal profiles were collected for six more die from wafer A. All experimental conditions were the same as with the previous 25 minute anneals, with the only change being the anneal environment: air to N\textsubscript{2}. Pre-anneal data is displayed in Figure 3.21. No significant die-to-die variation was observed. Post-anneal profiles are observed in Figure 3.22a. A similar trend as with the 25 minute anneals in air is observed, whereby the Si compression between the TSVs increases monotonically with anneal temperature. Compression of the Si between the TSVs also occurs at the same temperature (250 °C). However, the compression at 350 °C is not as significant as the previous study. One notable difference is the significantly reduced Cu protrusion, as seen in Figure 3.22b. The largest Cu TSV protrusion is observed for the 400 °C anneal; however it is still less than half (under 500 nm) that of the anneals conducted in air. Thus, it can be concluded that the larger amount of Cu protrusion observed in the annealing studies conducted in an ambient environment was mainly due to Cu oxidation.
Figure 3.19. (a) Post-anneal (air, 25 min.) (a) stress and (b) AFM profiles along [110]. Inset is an optical image of array at 400 °C displaying possible Cu oxidation.
Figure 3.20. FIB cross-sections of round, 5×25 μm Cu 1×4 TSV arrays annealed for 3.5 hrs. Different features are labeled in (a). Images (a)-(f) correspond to anneal temperatures 150-400 °C, respectively. Note that all TSVs are completely filled and no significant void formation is observed. Voids observed after a 350 °C anneal are circled in (e). The severe taper of the TSVs in (f) are due to over-milling in the FIB at a slight angle.
Figure 3.21. Wafer A pre-anneal stress profiles along [110]. No significant die-to-die variation observed.
Figure 3.22. (a) Post-anneal ($N_2$, 25 min.) (a) stress and (b) AFM profiles along [110]. AFM topography displays a significantly reduced Cu protrusion than air anneals.
3.3.4 Raman Measurements on As Received Wafer B TSVs

A. Isolated TSVs

The role of additional fabrication steps, including higher processing temperatures, may induce further stress evolution in Cu TSVs, resulting in changes to the residual stress in the surrounding Si. Raman measurements were performed on wafer B, which had undergone additional high-temperature processing compared to wafer A. The full wafer details can be found in Section 3.2. However, the important aspects of the processing conditions beyond what wafer A had seen included a 350 °C post-Cu CMP furnace anneal and M1 oxide deposition at 400 °C. Therefore, wafer B had undergone additional processing with multiple high temperature anneals. This provided an important comparison for the stress evolution as a function of process history.

As with wafer A, five die were chosen from wafer B, and Raman profiles were collected along the [110] direction to get an idea of the die-to-die variability. The averages of ten profiles for each die are plotted in Figure 3.23a, where the inset illustrates the relative wafer locations of the die. The corresponding profiles all appear to fall within ±10 MPa of each other, which is on the error of the measurements. Based on this observation, the die-to-die variation appears to be insignificant.

For a more quantitative approach to compare the die, the same metrics used in Figure 3.11b&c for wafer A are used for the isolated TSVs of wafer B. The maximum tensile stress points are seen in Figure 3.23b, while the positions from the TSV edge of the compressive-to-tensile crossover point, maximum tensile stress point, and 50% reduction
in tensile stress point are plotted in Figure 3.23c. The dashed lines represent averages for all points across the five die.

All of the maximum tensile stress points, except for die 1, fall within a 95% confidence interval of the mean (-10 ± 3 MPa). However, considering measurement error (~ ±8 MPa), die 1 cannot be considered significantly different. In Figure 3.23c, the averages of the compressive-to-tensile crossover point, the maximum tensile stress point, and the 50% reduction of the maximum tensile stress points are 2.2 ± 0.3 µm, 4.3 ± 1.2 µm, and 8.3 ± 1.2 µm, respectively. Considering measurement error, all die lie within a 95% confidence interval for the corresponding metrics. Though the magnitudes of the stresses, and the region of compression around the isolated, round 5×25 µm Cu TSVs are not significantly different, there is some variation in the distances of the tensile region away from the TSV edge. However, the die-to-die variations are still not statistically significant.

Comparison plots of the key metrics for wafers A and B are seen in Figure 3.24. The first plot in (a) is a comparison of the average maximum tensile stress on either side of an isolated, round 5×25 µm Cu TSV in the [110] direction. A maximum tensile stress of -10 ± 3 MPa for wafer B is lower than that of wafer A (-15 ± 4 MPa). However, the difference of the average maximum tensile stresses are not statistically significant by way of a two-sample t-test with a significance level $\alpha = 0.05$ [MONTGOMERY, 2005]. Plot (b) compares the distances of the three metrics relative to the TSV edge. In a likewise manner as the maximum tensile stress, a two-sample t-test with $\alpha = 0.05$ was conducted for each metric. The differences in both the crossover points and the 50% reduced maximum tensile points between wafers A and B are considered statistically significant.
Figure 3.23. (a) Stress profiles of isolated, round, $5 \times 25$ $\mu$m Cu TSVs for 5 die on wafer B. Die-to-die variation is within the error of the measurements. Key metrics for quantitative comparison of die-to-die variation for isolated Cu TSVs include (b) maximum tensile stress and (c) measurements of 3 key points for determining the extend of the stress field away from the TSV edge.
However, the difference between the maximum tensile point distance between wafers A and B is not considered statistically significant. Therefore, the main difference that can be considered statistically significant is the extension of the stress field away from the TSV after an increased number of processing steps and higher temperatures (wafer B). Another observation noted is the larger amount of variation in the metrics on wafer B. Both of these observations have important implications for determining keep-out-zones and for minimizing stress superposition effects.

Figure 3.24. Comparison of average (a) maximum tensile stress and (b) crossover, max tensile, and 50% relaxed tensile points for isolated Cu TSVs from 5 die on wafers A and B.

B. Linear, 1×4 TSV Arrays

As with wafer A, Raman profiles were measured along the [110] direction for the five die from wafer B. The results plotted in Figure 3.25a clearly display superposition effects. This result confirms what was seen with wafer A. However, where in wafer A, the variability of the profiles fell within approximately 1σ, the profiles for the 1x4 arrays of wafer B display greater variations. This is more clearly illustrated in Figure 3.25b,
where a maximum variability on the order of ~ 35 MPa is observed for the maximum tensile stress points between TSVs. This confirms the die-to-die variability of the position of the maximum tensile and 50% reduced tensile stresses observed for the isolated TSVs on this wafer. The variability, however, appears to be amplified due to stress superposition effects.

Figure 3.26 plots the average maximum tensile stress between each TSV along [110] for the five die across wafers A and B. On average, the tensile stress appears to be greater for wafer B, which has undergone further processing than wafer A. Though, due to the large error bars, this cannot be considered statistically significant. However, the most important observation is the increase in variability of the stresses, indicated by the error bars. This is in agreement with the results of the isolate TSV measurements. An increased variability implies that the Cu, undergoing additional plastic deformation, is not in a mechanically stable state after the first post-CMP anneal (wafer A).

Confirmation of the stress-free baseline procedure for wafer B was conducted by an extended Raman scan for the 1×4 TSV array, as was done in wafer A. The results are plotted in Figure 3.27. The blue curve is the stress profile for a typical 5×25 μm 1×4 TSV array, and the the black line is the extended stress profile. Overlap in the profiles confirms that the last point in the typical Raman scan (blue) is indeed at a relaxed position.
Figure 3.25. 1×4 TSV array (a) stress profiles and (b) maximum tensile stress points along [110] (wafer B).
Figure 3.26. Maximum tensile peak comparisons in 1×4 Cu TSV arrays for wafer A vs. B.
Figure 3.27. Extended Raman scans on a $1 \times 4$ TSV array from wafer B were collected to confirm an accurate stress-free position.

**B. 2-Dimensional, 8×8 TSV Arrays**

Raman line profiles collected near the center of an $8 \times 8$ TSV array along the [110] direction for five die across wafer B are plotted in Figure 3.28a. Overall, the stress magnitudes are higher than seen in the $1 \times 4$ TSV arrays, as expected due to increased superposition effects due to larger numbers of TSVs. By examining the maximum tensile stress points in (b), large variations, up to approximately 35 MPa, are observed. The variations appear to increase closer to the center of the array. This evidence, in addition to initial observation of the variability between $1 \times 4$ TSV arrays on wafer B, leads to the
conclusion that superposition effects do have an impact on the amount of stress variation for numerous high temperature processes. It can be said that the effects of numerous TSVs in an array fashion can induce high stresses in nearby Si due to superpositioning, and if the Cu is not mechanically stable (work hardened), then the additional Cu deformation due to repeated high temperature processes can induce large amounts of stress variation throughout the array, which could have detrimental effects on large scale performance consistency.

A comparison of the average maximum tensile stresses between TSVs in a 2D 8×8 array for five die across wafers A and B is seen in Figure 3.29. As already mentioned, the large die-to-die variation is observed in the form of the error bars of the wafer B data. However, it is also noted that the average tensile stress for the 8×8 array is significantly lower than wafer A. Reduced average tensile stress for the larger arrays could be a result of unknown relaxation mechanisms. For example, tensile stress reduction was seen in the case of annealed, uncapped Cu TSVs due to Cu protrusion from the free surface (see Section 3.3.3). Stress mitigation on the wafer B 8×8 arrays may have implications for possible relaxation mechanisms for larger arrays undergoing higher temperature processing.
Figure 3.28. 8×8 Cu TSV array (a) stress profiles and (b) maximum tensile points along [110] for five die across wafer B.
3.3.5 Raman Measurements on Wafer C, Post-CMP

To observe whether any residual stress was present in Si surrounding Cu TSVs before any thermal anneal steps were conducted, Raman measurements were performed on wafer C. The complete wafer details can be seen in Section 3.2. The importance of this wafer is that it has only undergone processing up through Cu plate and a post-place CMP. Therefore, any stresses present in the as-received samples would not be thermomechanically driven. The results for the Raman line scans of the as-received profiles of seven, 1×4 Cu TSV arrays (each a separate die on wafer C) are plotted in Figure 3.30a.
The presence of a stress profile indicates that initial residual stresses in the Si are not thermo-mechanically driven, but rather a result of Cu plastic deformation in the form of self-annealing. Measured stresses are on the order of magnitude for as-received wafers A and B 1×4 Cu TSV profiles. However, direct comparisons of the magnitudes cannot be considered without taking into account the differences in Cu TSV volume. Despite that, the profiles display qualitative agreement with a thermo-mechanically driven, Cu deformation induced, residual stress profile.

Die-to-die variation appears to be significant on wafer C. Figure 3.30b shows the maximum tensile stress points for the various die. The stresses between the Cu TSVs vary from a maximum of 25 MPa to less than 5 MPa. Large stress variations for Cu that has not undergone a post-CMP anneal would indicate that the residual stress profiles due to self-annealing are highly dependent upon Cu structure, which can be related to the Cu plating process. Furthermore, it indicates that a post-CMP anneal acts as a mechanical stabilizer, perhaps through work hardening, for the Cu TSVs, such as in the cases of wafers A and B.

The stress evolution of non-thermally annealed Cu is studies on wafer C. Seven die were annealed at various temperatures in N₂ for 25 minutes each (full details in Section 3.2). After the samples cool to room temperature, Raman measurements are performed. The resulting profiles are seen in Figure 3.31a. As with the anneal studies performed on wafer A, a trend of increasing Si compression is observed with increasing anneal temperature. However, in the case of the previously non-thermally annealed 5×50 µm, compression of the Si reaches up to 80 MPa. This is more evident in Figure 3.32a, which illustrates the stress “minima” between the TSVs. Significant compression is
Figure 3.30. (a) $1\times4$ TSV array (a) stress profiles and (b) maximum tensile stress points along [110] (wafer C).
clearly observed. This compression coincides with Cu protrusion from the tops of the TSVs, as evidenced in Figure 3.32b. However, the amount of protrusion is at most 400 nm, which is on the same order of Cu protrusion from similar anneals performed on wafer A.

Figure 3.31. Post-anneal (N₂, 25 min.) stress profiles along [110] for 5×50 µm 1×4 Cu TSV arrays.
Figure 3.32. Post-anneal (a) maximum tensile points and (b) corresponding AFM profiles along [110] for 5×50 µm 1×4 Cu TSV arrays.
3.3.6 Raman Depth Profiling on As Received Wafers A & C

All profiles measured up to this point have been top-down Raman profiles. Considering the optical penetration depth of 532 nm wavelength excitation is approximately 1.1 µm in Si, a top-down approach works well for probing the stress regions of the device layer at early stages of a 3D process flow. However, at later stages where the device layer is essentially buried (i.e. after bonding), a top-down optical approach is not feasible for probing stress evolution at that depth. Doing so would require much larger wavelengths (IR) of light, including the necessary optics and detection system for such a setup, in order to probe deeper into the Si. Furthermore, to characterize the stress fields along a TSV as a function of depth would require multiple wavelengths of light to probe various the various depths, where the penetration depths of various wavelengths can be calculated as in DE WOLF (1996) based on the absorption coefficient of Si. Since the profiles obtained would consist of Si Raman peak shifts which are averaged over the entire probing region of the source wavelength, extracting stress information versus depth require extensive modeling [SPARKS, 1992; ATKINSON, 2001], which is non-trivial.

Preliminary Raman stress profiles collected with wavelengths of 325, 532, and 633 nm excitation were conducted to see how the stress profiles change over different depths. The results are displayed in Figure 3.33. The inset shows the penetration depths for various laser wavelengths. What is observed is that the tensile stress magnitude is greatest near the surface (~10 nm with 325 nm wavelength). The tensile stress decreases with increasing depth, as observed with the profiles collected with 532 and 633 nm wavelength, which have ~1.1 and ~3.2 µm, respectively. Though this provides a
qualitative understanding of the depth and stress variation, there is need for extensive modeling to extract exact stress information for specific depths.

Figure 3.33. Stress profiles collected with 325 (violet), 532 (green), and 633 (red) nm wavelength excitation sources. Tensile stress appears to decrease with increasing depth.

An easier solution, which is practical for characterization purposes, is a cross-sectional Raman approach, whereby a FIB cross-section of the TSVs is prepared. Figure 3.34 illustrates this approach. Cross-section Raman scans allow the use of a single wavelength to scan along the length of TSV, providing stress information as a function of the depth along [001]. Furthermore, the results from this approach are essentially top-
down, meaning there is no need for extensive modeling to extract the sought after stress information.

Cross-sections were prepared for a total of four die. Two of which were from wafer A, and two were from wafer C. Samples from both wafers were chosen to get an idea of the stress variation as a function of TSV depth for the two TSV sizes (5×25 μm vs. 5×50 μm, respectively). The two as-received die taken from each wafer include one cross-sectioned along [110] through the center of a 1×4 Cu TSV array, while the other is cross-sectioned along [110] ~2-4 μm from the array edge, leaving a 2-4 μm Si region on top of the 1×4 Cu TSV array. Cross-sections through the center of the array are expected to display the highest amounts of stress (as observed in the top-down Raman measurements), while measurements through a Si buffer layer are expected to display a more relaxed stress profile.

Figure 3.35 displays SEM images of the cross-sections for all prepared samples. The wafer A Si buffer layer sample in (a) displays a flat surface, and from the position of the
TSVs it can be seen that the Si buffer layer is ~ 2 µm away from the TSV array. The image in (b) shows a cross-section along [110] through the center of the 1×4 Cu TSV array. Notice that for this die, the Cu TSVs were not completely filled. At least one had a significant void near the bottom center, while the other three were only ~ 70% filled. Other cross-sections on 1×4 Cu TSV arrays from wafer A showed the TSVs to be completely filled with Cu, with no significant void formation (as seen in Figure 3.20). Images (c) and (d), likewise, display similar cross-sections for wafer C. While the Si buffer layer for this sample appears to be slightly thicker, ~ 3-4 µm, it is still observed to be a flat surface. The cross-section through the center of the 5×50 µm 1×4 Cu TSV array shows that the vias are completely filled. In fact, there appears to be some Cu bulging at the sidewalls starting several microns below the oxide/Si interface. Whether this is due to the via etch or Cu plating is as yet undetermined.

Raman measurements were performed in a top-down configuration, with the scan direction along the [001] Si axis. Raman measurements were taken at a region ~ 40 µm below the TSVs to confirm that this region is stress free. All Si Raman peak shifts were taken in reference to this region to calibrate to zero-stress. On each sample, two Raman shift profiles were collected. One along [001] that passed between the center two TSVs, and one along [001] ~ 3µm from the outer TSV edge. These positions should approximately coincide with the maximum tensile stress points observed in top-down Raman measurements.

Figure 3.36 displays the obtained Raman profiles for the cross-sectioned samples. The stress profiles for both the TSV and the Si buffer layer cross-sections (a) display a similar trend. At the oxide/Si interface there appears to be a high tensile stress, while
moving deeper into the Si the tensile stress relaxes within ~ 2 µm before becoming compressive. This agrees qualitatively with the multi-wavelength measurements in Figure 3.33. Though the 633 nm wavelength would be expected to measure a compressive stress, based on its penetration depth (~ 3.2 µm), it also averages over the larger tensile stress region, thus displaying a tensile stress away from the TSV (Figure 3.33—red), and demonstrating the need for modeling for exact stress extraction.

The stress relaxes deeper in the Si, until near the bottom of the TSV the Si is back to a relaxed state, which continues deeper into the Si. The magnitude of stress measured in the Si buffer layer is less than that measured at the 1×4 array cross-section. This is expected, since it has been observed in the top-down Raman measurements that the stress becomes more relaxed radially away from the TSV axis.

Figure 3.37a displays stress depth profiles at the center and edge of the TSV array, corresponding to the blue curves in Figure 3.36a. Comparison between the FE model and the experimental measurements yields qualitative agreement. The large tensile stresses near the top of the TSV (larger at the center of the array, as expected) shifts to compression ~ 5 µm from the top. This compression quickly relaxes within another 5 µm and then abruptly increases again at the bottom of the TSV before a tensile jump at the bottom of the TSV. In the experimental profiles, the compression relaxes more slowly, and there is no significant stresses, neither compressive nor tensile, observed at the bottom of the TSV. Though the FE model indicates stresses present at the bottom of the TSV, which are not observed experimentally, there is qualitative agreement with the stress depth profiles for 5×25 µm, 1×4 Cu TSVs.
Figure 3.35. SEM images of FIB cross-sections along [110] through a Si buffer layer (a) and 1×4 TSV array (b) from wafer A, as well as through a Si buffer layer (c) and 1×4 Cu TSV array (d) for wafer C.

Raman measurements of the Si buffer layer scan for the 5×50 µm 1×4 Cu TSV array, Figure 3.36b (black curve), display a similar trend as seen in the 5×25 µm 1×4 Cu TSV array, Figure 3.36a (black curve). The magnitude of the stress is even around the same level: ~ 20 MPa at the peak tensile stress point. The main difference is the distance
before the Si stress becomes relaxed. Where in the 5×25 µm TSV case the relaxation distance is ~ 25 µm, it is ~ 40 µm in the 5×50 µm TSV case. Thus, the relaxation distance approximately scales with the TSV depth.

Looking at the cross-section through the 5×50 µm 1×4 Cu TSV center along [110], Figure 3.36b (blue curve), a different trend is observed than in the other cross-section profiles. Starting from a compressive stress at the oxide/Si interface, rather than becoming tensile at a further depth, the compression spikes at ~ 5 µm from the interface, then begins to relax over a distance of ~ 50 µm. The tensile nature of the stress along the central part of the TSV has qualitative similarities with FE modeling for a 5×50 µm 1×4 Cu TSV array, as in Figure 3.37b. In the FE model, after a jump to compression within ~ 10 µm of the top of the TSV, the stresses observed are tensile with an increasing magnitude until ~ 70% of the depth of the TSV where it begins to relax until complete relaxation is observed corresponding to the bottom of the TSV, before a tensile jump is seen, which is similar to what is observed in the FE model for the 5×25 µm Cu TSV geometry. This tensile behavior is not observed experimentally, rather than having a compressive jump ~ 10 µm from the top of the TSV array, the Si remains under a tensile stress, though there is a corresponding jump in stress that results in a decrease of the tensile stress. From there the tensile stress over the depth of the TSV relaxes at a consistent rate until the bottom of the TSV is reached. The origin of this behavior is not fully understood and requires further investigation. However, the qualitative correlation between the Raman depth measurements and the FE simulated depth stress profiles confirm that a cross-section approach is promising for measuring the residual stresses in Si at various depths of a TSV.
Figure 3.36. Stress profiles along [001] for cross-sections through the center of the TSVs and through a Si buffer layer display similar stress profiles (a) for a 5×25 μm 1×4 Cu TSV array, where a lower stress magnitude is observed when a Si buffer layer is present. Stress profiles along [001] for cross-sections through the center of the TSVs and through a Si buffer layer display differing stress profiles (b) for a 5×50 μm 1×4 Cu TSV array, where the measurements through the center of the TSV array unveil a highly compressive stress.
Figure 3.37. FEM generated stress profiles at the center (black) and edge (red) of (a) 5×25 µm and (b) 5×50 µm 1×4 Cu TSV arrays. (Courtesy B. Backes)
3.4 THERMO-MECHANICAL FINITE ELEMENT MODELING QUALIFICATION OF RAMAN MEASUREMENTS

Thermo-mechanical modeling of Cu TSVs in a Si substrate using FEM has proven to be effective in determining the residual stresses in Cu and Si for both reliability [OKORO, 2007; BENTZ, 2005; ZHANG, 2005; LU, 2010; RYU, 2011] and effects on mobility [OKORO, 2008a; KARMARKAR, 2009, 2010; SELVANAYAGAM, 2009b]. OKORO et al. (2008a) used micro-Raman spectroscopy to qualify their models to determine the best parameters and plastic properties for Cu TSVs. Following those results, a FE model was established using COMSOL Multiphysics® that relied on a plastic yield in Cu following a strain-hardening behavior. Complete studies to determine the effect of Cu material constants on the residual stresses in Si were conducted, as well as the effect of different Cu geometries and liner damping effects. The full results can be seen in BACKES (2011). This section provides a brief overview of the FE models employed and their importance for qualifying Raman measurements and providing insight into their interpretation.

3.4.1 Finite Element Modeling of Cu TSVs in a Si Substrate

Finite element modeling is a computer-aided mathematical technique for obtaining approximate numerical solutions to the partial differential equations that predict the response of the physical system subjected to external influences [BURNETT, 1987]. In other words, it is a technique that allows for a piecewise approximation of a continuous function [DAVIES, 1980]. This is accomplished by dividing a given system geometry into small polygons, called finite elements. The complete set of finite elements for a given geometry is called a mesh. The vertices where finite elements in the mesh meet are
called nodes. At each node, the partial differential equations that describe the physics of
the problem at hand are solved in an iterative fashion within certain upper and lower
bounds until convergence within a specified limit is reached. Thus, solutions to the
equations at each node in the mesh provide a complete approximation to the continuous
function for the behavior of the specified geometry.

The FE method was originally developed by engineers to solve structural mechanics
problems [DAVIES, 1980]. It has an extensive history with a well-developed theory for
solving problems of a mechanical nature [BURNETT, 1987; DAVIES, 1980; SMITH, 1998;
CHEN, 2005]. It is essentially based on calculating virtual displacements of the nodes as
a result of applied loads. COMSOL solves for mechanical equilibrium using the
principal of virtual work, which states that the work done by internal strains equals the
work done by external loads [COMSOL]:

\[ \delta W = W_i - W_e = 0 \]  

(3.8)

where \( W \) is the total stored energy, and \( W_i \) and \( W_e \) are the virtual work from the internal
strains and external loads, respectively. Therefore, given certain initial conditions and
specified boundaries, the solutions are solved iteratively until the difference in total
stored energy is \( \sim 0 \), within a specified convergence tolerance.

Figure 3.38 shows the geometry used for solving an isolated TSV. A quarter section
of the full geometry is used with symmetry planes to account for the stress of the
complete TSV. Using a smaller subsection of the overall structure geometry by
exploiting the system symmetry allows for less computational, but still provides accurate
solutions. The material regions include both a Si substrate (30×30×50 µm) and 5×25 µm,
round Cu TSV with a 1.2 μm SiO₂ layer on top. The material parameters used are listed in Table 3.2. They were chosen based on the parameters used in OKORO (2008a). $E$ is Young’s modulus, which describes the elastic response of a material to an applied load, or stiffness; $\nu$ is Poisson’s ratio, which describes the proportion of transverse strain to longitudinal strain for an applied external load; $Y_0$ is the yield stress, which describes the transition point from a reversible, elastic process to an irreversible, plastic process; CTE is the coefficient of thermal expansion, which describes the expansion of a material as a function of temperature; and $h$ is the tangent modulus, which describes the linear plastic response of a material as a function of plastic strain. It is used in the isotropic strain-hardening yield criterion, which provides an approximation for the average plastic response of the Cu:

$$Y(\varepsilon^p) = Y_0 + h\varepsilon^p$$  \hspace{1cm} (3.9)

where $Y(\varepsilon^p)$ is the yield stress as a function of plastic strain. This Cu strain hardening function was shown to provide the best fit between simulations and experimental results with Raman spectroscopy [OKORO, 2008a]. This is confirmed by the results below.

<table>
<thead>
<tr>
<th>Material</th>
<th>Young’s Modulus E (GPa)</th>
<th>Poisson’s ratio $\nu$</th>
<th>CTE (ppm/°C)</th>
<th>Yield Stress $\sigma_y$ (MPa)</th>
<th>Tangent Yield Stress h (MPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si (anisotropic)</td>
<td>Direction dependent</td>
<td>Direction dependent</td>
<td>2.3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Cu (elastoplastic)</td>
<td>117</td>
<td>0.3</td>
<td>16.7</td>
<td>172.3</td>
<td>517</td>
</tr>
<tr>
<td>SiO₂ (isotropic)</td>
<td>75</td>
<td>0.17</td>
<td>0.5</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 3.2. Material properties were chosen based on OKORO (2008a).
Figure 3.38. COMSOL quarter model geometry (top down view in upper right) for an isolated TSV with corresponding mesh. Mesh elements are smaller closest to the TSV where geometry variations are more abrupt. [BACKES, 2011]

The model above was solved using the aforementioned parameters for the specified geometry in COMSOL using the Thermal-Structure Interaction feature within the Structural Mechanics Module. It involves the interaction between the elasto-plastic response with heat transfer for a given temperature gradient. A top down view of the biaxial stress surrounding the isolated TSV is seen in Figure 3.39. The anisotropy of Si leading to the four-fold symmetric tensile stress distribution in the radial direction seen in
Figure 3.6 is clearly observed in the FE simulation. Note that this plot is rotated by 45 degrees due to the fact that the x and y directions in this model were aligned to the <100> directions rather than the <110> directions. However, this only affects the presentation. The results clearly qualify the observed Raman measured, 2D stress distribution surrounding an isolated 5×25 µm Cu TSV.

Figure 3.39. Top down view of a one-quarter isolated, 5x25 µm Cu TSV displaying anisotropy in the Si, corresponding to the four-fold symmetry observed in Figure 3.6.

Figure 3.40 shows a comparison of the simulation result with the experimentally obtained Raman shift profile (wafer A) for an isolated Cu TSV along the [110] direction.
The simulation used an effective anneal cycle from room temperature to 350 °C back to room temperature. As a reminder, wafer A has only undergone a 150 °C after Cu CMP.

The simulation qualitatively agrees with the Raman profile obtained. The only discrepancy is that the effective anneal temperature used to obtain the simulation result was much higher than the actual anneal temperature that wafer A saw. Thus, the Raman results confirm that the residual stress profiles measured are due to complete anneal cycle where the Cu has undergone plastic deformation. Additional simulations where the Cu was given different initial stress and strain conditions displayed markedly different

Figure 3.40. Comparison of Raman measurement (wafer A) with a FE simulation (350 °C effective anneal) for an isolated, round 5×25 µm Cu TSV along [110].
residual stress profiles than what was obtained experimentally [BACKES, 2011]. This provides further evidence that the resulting stress profiles are best described by a complete thermal anneal cycle. In addition, the higher effective anneal temperature that is used in the simulation to obtain a similar profile as obtained by Raman spectroscopy implies that the Cu in the real TSV has undergone additional, irreversible deformation (i.e. self-annealing), as mentioned in Section 3.3.7. More discussion on this will follow in Section 3.5.

An additional comparison between simulation and experiment for a 1×4, 5×25 µm Cu TSV array was conducted. The associated model geometry and the comparison plot are displayed in Figure 3.41 and Figure 3.42, respectively. The x and y axes correspond to <110> directions in this model to take into account the rotation with respect to the sample orientation. Again, there is qualitative agreement between the experimentally obtained Raman shift profile and that obtained via FE simulation. However, there is a difference between the actual anneal temperature for wafer A and the effective anneal temperature used in the simulation. The effective anneal temperature essentially determines the magnitude of the stresses. As already mentioned, the temperature difference correlates to additional plastic deformation of the Cu, attributed to self-annealing.

The FE simulation qualitatively confirms the Raman shift profiles observed experimentally. Residual stresses in Si are proven to result from plastic deformation of the Cu TSVs undergoing full anneal cycling. Differences in the effective anneal temperatures used to obtain similar magnitudes of stress as the Raman profiles are attributed to self-annealing. In other words, wafer A, which has undergone a 150 °C anneal, appears to have undergone a higher temperature process, according to simulation.
This points to additional plastic deformation in the Cu than what a 150 °C anneal would cause. There will be further discussion on this point in the following section.

Figure 3.41. COMSOL quarter model geometry (top down view in upper right) for a 1×4 TSV array with corresponding mesh. Mesh elements are smaller closest to the TSV where geometry variations are more abrupt. [BACKES, 2011]
Figure 3.42. A 1×4, round 5×25 µm Cu TSV array (a) FE model geometry and (b) comparison between Raman measurement (wafer A) with simulation (350 ºC effective anneal) along [110].
3.5 DISCUSSION OF RAMAN STRESS RESULTS ON TSV STRUCTURES

3.5.1 Origin of the Residual Stress Profile in Silicon

The characteristic Raman profiles obtained are consistent with those previously shown in literature [OKORO, 2008a]. In the Si immediately surrounding the Cu TSV is a compressive region, while farther away, Si is shown to be tensile. These regions have also been observed in FE studies [GONZALEZ, 2005; KAWA, 2010] and are attributed to the tangential compression of Si around the Cu and radial tension farther away. A schematic is shown in Figure 3.43, which illustrates this behavior. At elevated temperatures, Cu, with a higher CTE than Si, becomes compressed due to the constraint of the surrounding substrate. The resultant stress between two materials with different CTEs is calculated from (Eqns. 3.1, 3.2). Upon subsequent cooling to room temperature, after an irreversible plastic deformation, Cu and Si contract per their corresponding CTE’s. Thus, Cu, contracting more with temperature, undergoes a radial tensile strain, assuming no TSV sidewall delamination, due to the Si substrate restraining the Cu contraction. Therefore, the Si, is also subject to an inward stress. This leads to a residual tensile region in the Si, as observed. Furthermore, since the Si undergoes a tensile deformation in a radial direction, due to the circular geometry, the Si closest to the Cu will be contracted tangential to the TSV. This is represented schematically in Figure 3.43b. Since both compressive and tensile stresses, albeit different directions, are perturbations of the crystal lattice and by definition the phonon modes, they can be observed by Raman scattering. Similar behavior is observed for the square TSVs, where the Si compression is highest near the TSV corners, which can be described as boundaries of extreme curvature.
3.5.2 Discussion on Stress Superposition Effects

When independent stress fields are brought together, superposition can occur [CLELAND, 2003]. This behavior is observed in an array of TSVs when they are brought sufficiently close together [LU, 2009]. In the case of a linear array of 1×4 TSVs with a 1:1 pitch ratio, the stress fields along the array axis, [110], are observed to interfere constructively (e.g. Figure 3.12). Due to symmetry, the maximum tensile stress is observed to be approximately midway between TSVs, with the center of the array displaying the highest maximum tensile stress in the array. However, in the direction perpendicular to the array axis, [-110], where there are no nearby structures, there are no observed superposition effects (e.g. Figure 3.13). Therefore, the profiles along [1-10] are comparable to that of an isolated via. Furthermore, the profile along [110] at the outside of the array shows a tensile stress magnitude comparable to an isolated TSV (Figure 3.9).
This is expected, once again, due to the lack of any nearby TSVs outside the array to induce superpositioning of stresses.

Likewise, superposition effects are observed with the 8×8 TSV arrays (e.g. Figure 3.15). The magnitude of the tensile stresses between adjacent TSVs is increased compared to that of the 1×4 arrays (~ 45 vs. ~ 35 MPa, respectively). Since the 8×8 array has a larger number of adjacent TSVs, it is expected that tensile stresses should be increased near the center of the array due to superposition effects. Furthermore, Raman stress profiles collected in the [110] and [1-10] directions (Figure 3.16) have comparable stress magnitudes due to the 2-dimensional symmetry of the array.

These results have important implications for designing and placing TSV structures in a device. As seen, the density and spacing of TSVs can severely impact the amount of residual stress seen in Si near Cu TSV structures. Placing the TSVs too close together along <110> directions could lead to significant increases in stress in the Si, thereby impacting device performance. In addition, the increased stresses could impact the Cu, causing delamination or aiding void formation and propagation, which could lead to failure. However, due to the Si anisotropy, placing TSVs as nearest neighbors along the <100> directions could limit the stress superposition effects, allowing closer placement. This was demonstrated in [LU, 2009]. Therefore, the understanding of process induced thermal-stress evolution, as well as the magnitudes and directions of said stresses, within TSV arrays is important for determining acceptable tolerances for TSV pitches.
3.5.3 Stress Evolution Due to Cu Annealing

Temperature anneal studies on wafers A and C, as well as comparisons of measurements performed on as received wafers A and B, provide insight into the evolution and nature of the residual stresses in Si as well as the Cu TSV. In this section, the effects of annealing conditions, such as time, environment, and temperature on Cu protrusion and stress evolution will be discussed; process-induced stress variation by comparing wafer A and B will be discussed; and self-annealing as evidenced by wafer C will be discussed.

3.5.3.1 Temperature Anneal Studies

Three temperature anneal studies were conducted on wafer A. The first was a range of anneals done on separate samples in an air environment, where each sample was annealed at a different temperature for 3.5 hrs. The second study included the exact same parameters, but with an anneal time for each sample lowered to 25 min. A third study was conducted similar to the second study, only the anneal environment was nitrogen rather than air.

First, a note on the comparisons of the stress profiles: for all anneal studies, the stress profiles changed most dramatically in between adjacent TSVs, indicating that stress superposition plays a dominant role in the stress evolution. The stress profiles on the outside of the 1×4 Cu TSV arrays remained fairly consistent, with minimal variations in magnitude. At higher temperatures (≥300 °C), the point of maximum tensile stress was observed to move away from the outer TSV in the array; however, this was less than ~1
μm and may not be significant. In any case, all profile comparisons are in terms of the relative magnitudes of the maximum tensile stress points between adjacent TSVs.

Anneal time appears to effect the stress development in Si between Cu TSVs. Comparing Figures 3.17 (3.5 hr anneal time) and 3.20 (25 min anneal time), a significant difference in the stress profiles, as well as the Cu protrusion, is observed. In Figure 3.19 (25 min anneal time), the stress profiles do not appear to change until a temperature of 300 °C. This coincides with a significant increase in the Cu protrusion, as seen in the corresponding AFM profile. The longer, 3.5 hr anneal in air, Figure 3.17, begins to show a shift in the stress profile at a temperature of 250 °C. This, likewise, corresponds to the onset of significant Cu protrusion from the free surfaces of the TSVs. Furthermore, the shorter anneal time displays a monotonic trend of tensile relaxation with increasing anneal temperature, which also corresponds monotonically with Cu protrusion. This hints at Cu protrusion as a significant source of relaxation in the Cu, and hence, a significant relaxation in the surrounding Si. The longer anneal displays a similar trend, though with a deviation at 400 °C. However, at the higher temperatures (>300 °C), the Si between the TSVs is left in a completely compressive state. This implies that as the tensile stress becomes fully relaxed due to the Cu protrusion, there is a transition to compression in the Cu TSV, which results in a residual compression in the surrounding Si.

The only experimental difference between these two sets of measurements was anneal time. Therefore, the earlier onset of tensile relaxation and the resulting differences between stress profiles and Cu protrusion heights for higher temperature anneals is attributed to creep. At temperatures above 300 °C, a grain-boundary diffusion
mechanism, Coble creep, is a significant source of relaxation [THOULESS, 1993; KOBINSKY, 2001; OKORO, 2011]. Evidence of this mechanism can be found in void formation along grain boundaries [OKORO, 2011]. The TSV sample shown in Figure 3.20e, after a 350 °C anneal, was the only sample to show significant void formation. These voids were all observed to be along grain boundaries, which supports this argument. However, no significant void formation was observed for any of the other cross-sections (Figure 3.20a-d, f). Therefore, though Coble creep may play a role in stress relaxation at longer anneal times, void formation does not appear to be a significant source of stress relaxation in these samples. The relaxation is attributed mainly to the Cu protrusion. However, the mechanism for the protrusion has not been verified. Since it has been observed to occur at temperatures >250 °C, it could be attributed to Cu grain-boundary diffusion, though the lack of visible voids in all except the 350 °C annealed sample does not lend support to such an argument. Further investigation is required.

The very high Cu protrusion heights (>1 µm) for both sample sets annealed in air are attributed to Cu oxidation. Comparing Figures 3.19b (25 min anneal in air) and 3.22b (25 min anneal in N₂), the AFM profiles display a significantly lower Cu protrusion height for the N₂ annealed samples: ~400 nm compared to ~1 µm. However, the stress magnitudes as well as the trend of tensile relaxation with increasing temperature are similar. This supports the fact that the much larger protrusion heights seen for anneals performed in air are partly due to oxidation. Though oxidation can be attributed to the larger protrusion heights, it does not appear to have a significant effect on the stress magnitudes for an anneal time of 25 minutes.
Figure 3.31 displays the obtained stress profiles for 5×50 μm Cu TSVs in a 1×4 array after annealing samples ranging from 100-400 °C in N$_2$ for 25 minutes. The trend of tensile relaxation with increased anneal temperature, and in fact, compression in Si at temperatures >250 °C between adjacent TSVs is observed. Cu protrusion accompanies the Si compression. The compression observed in the Si is significantly increased, up to ~75 MPa at the midway point between adjacent TSVs, over what has been observed in the 5×25 μm anneal studies. This is attributed to the fact that wafer C had not seen any previous anneals. A thermal anneal step post-Cu plate has been shown to stabilize the Cu microstructure due to grain growth [OKORO, 2011]; therefore, the large compression in the Si can be attributed to grain growth during annealing. Grain growth has been shown to cause larger tensile stresses in Cu [GEISLER, 2005; HUANG, 2010]. This would correspond to an increased radial tension in the surrounding Si. However, it has been shown that radial tensile stresses in Si around a TSV are proportional to tangential compressive forces [RYU, 2011]. Considering the distance between the adjacent TSVs is 5 μm, and the maximum tensile stress is half that distance, it is plausible that the tangential compressive stress field is high enough to extend further away from the TSVs to cause an overall compressive force in the Si midway between vias.

The as-received measurements for wafer C are shown in Figure 3.30. It displays stress profiles for non-annealed 5×50 μm 1×4 Cu TSVs that resemble the characteristic profiles of a wafer that has undergone thermal annealing (wafer A & B). This indicates that before any annealing takes place, the Cu undergoes a process of irreversible deformation that resembles that of an anneal cycle where the Cu undergoes plastic deformation. This is attributed to self-annealing, which has been known to occur for
Furthermore, due to the die-to-die variation seen for the as received samples of wafer C, it is concluded that the self-anneal process is not completely uniform and is most likely dependent upon Cu plate conditions that effect the deposited grain structure.

3.5.3.2 Copper Self-Annealing

As already mentioned, there appears to be significant residual stress in Si between adjacent TSVs in a 1×4 array before any thermal processing is conducted. The origin of the as-received, residual stress profiles in Si on wafer C is attributed to Cu self-annealing. This wafer had been processed up to post-Cu plate CMP. However, no thermal anneals were conducted that would induce plastic deformation in the Cu. The as-received stress profiles, as seen in Figure 3.30 show significant tensile stresses in the Si, comparable to that of as-received wafers A and B. This implies that Cu undergoes non-thermally induced, irreversible deformation, which results in the initial residual stresses in Si. Moreover, the fact that self-annealing results in a similar profile as a thermal anneal step, as seen in the FE models (Figure 3.42), means that it can essentially be modeled as an effective anneal. This is confirmed by the fact that the FE model required a higher effective anneal (350 °C) to match the actual anneal temperature (150 °C) that wafer A experienced.

Significant Cu self-annealing has been observed to occur within hours of electrochemical deposition of Cu [RITZDORF, 1998; LAGRANGE, 2000], and is attributed to microstructure stabilization via grain growth [GEISLER, 2005; OKORO, 2011].
Considering the time the wafer was exposed to ambient conditions was on the order of weeks, it is evident that self-annealing would have occurred in these samples.

3.5.3.3 Process Induced Variation

Based on the trend of tensile relaxation with increasing anneal temperature, it would be expected that the stresses observed between adjacent TSVs in a 1×4 array from wafer B (350 + 400 °C anneals) would be less tensile than those in wafer A (150 °C anneal). However, this is not the case. Comparing Figures 3.12 (wafer A) and 3.25 (wafer B), the maximum tensile stress magnitudes are comparable, on the order of 40 MPa. The only difference is significant die-to-die variation observed on wafer B. This variation is also observed for the 8×8 arrays (Figure 3.28) on wafer B. The fact that the stress magnitudes of the profiles from die on wafer B are comparable to those of wafer A implies that the Cu microstructure is stable enough to resist additional, irreversible deformation leading to changes in the residual stress profiles in Si, and that Cu protrusion was the reason for the significant stress relaxation observed in the anneal studies. However, the increased die-to-die variation observed in wafer B for the TSV arrays is of concern. Though it may be related to process variations, it requires further investigation to understand whether it is a result of other underlying microstructural changes in the Cu.

3.5.4 Depth Profiling

Top-down Raman measurements give important insight into the stresses induced at the Si surface. However, to measure the stresses along the depth of the TSV, or at the
device layer after wafer bonding, an approach is needed to probe those regions. Cross-sectioning was chosen due to ease of implementation, as mentioned in Section 3.3.6. Based on the back-scattering geometry from a (110) plane, only the z-polarized Raman phonon is observed (Table 1.1). Thus, there should be a correlation between the stresses measured in a cross-section fashion and those from a top-down approach.

One concern is elastic stress relaxation at the surface due to the FIB cross-section preparation. This may be true locally at the surface; however, it is expected to be insignificant based on a sample thickness on the order of millimeters. Estimation of minimal relaxation is based on TEM sample preparation estimates saying that for thick samples, > 200-600 nm, elastic stress relaxation should be sufficiently minimized [TODA, 2000; USUDA, 2005]. Though this may be a liberal estimate, considering the sample thickness being several orders of magnitude higher than a typical TEM sample, it is decidedly appropriate. Furthermore, relaxation that does occur should be localized to the surface, considered < 200 nm based on the same TEM preparation sample preparation estimates. Since the optical probe depth is on the order of 770 nm, it is expected that the stresses measured over the probe depth constitute accurate measurements of the stresses in the actual structure, with insignificant relaxation.

Cross-section profiles on the 5×25 μm 1×4 Cu TSV arrays along [110] are seen in Figure 3.36a. By way of review, the blue and black curves correspond to the samples that were cross-sectioned through the center of the array axis and through the Si substrate ~ 2-3 μm away from the array, respectively. The open and closed circles correspond to scans taken along [001] ~ 2-4 μm outside the array (edge) and at the midway point between the middle two TSVs (center), respectively. Within the top ~ 2 μm of the
TSV/oxide interface, the stress observed in Si is tensile. Moving away from the top, the trend is toward compression in the Si, where ~ 2.5μm from the TSV top, the residual stress in Si crosses over to a compressed state, which becomes maximum ~ 5μm from the top of the TSV. It then gradually relaxes over the depth of the TSV. In the case of the 5×25 μm cross-section through the center of the array, the Si becomes completely relaxed at ~ 15 μm from the top. This roughly corresponds to the depth of the Cu TSV for this sample, considering there is an incomplete Cu fill, as seen in Figure 3.35b.

The sample cross-sectioned through the Si region before the TSV array becomes relaxed ~ 25 μm from the TSV top (Figure 3.36a (black)), which corresponds to the expected TSV depth. Thus, the sample cross-sectioned through the Si region is expected to have a complete Cu fill, as has been observed in the cross-sections from the same wafer (Figure 3.20). It is also noted that the tensile stress observed near the top of the TSV for the cross-section through the center of the array is higher than that through the Si region away from the array. This agrees with what is seen in the top-down measurements, where the residual stress distribution tends to relax away from the TSVs. Furthermore, comparing the center to edge profiles for the respective cross-sections, a higher tensile stress near the top is observed at the center of the array. Once again, this is expected due to stress superposition effects discussed in Section 3.5.2. Moreover, comparing the relaxation in the tensile stress moving away from the Si surface, this qualitatively agrees with the measurements taken with multiple wavelengths (Figure 3.33). However, the multi-wavelength measurements are averaged over significant depths (e.g. ~ 3.2 μm for 633 nm wavelength); therefore, quantitative correlations cannot
be made without modeling to extract depth specific stress magnitudes, as mentioned in Section 3.3.6.

Likewise, Figure 3.36b displays stress profiles for similar cross-sections of 1×4 Cu TSV arrays for wafer C. The difference being that wafer C has deeper vias (5×50 µm compared to 5×25 µm on wafer A). The profiles through the Si region display similar trends as with the 5×25 µm TSVs, with the notable exception being the relaxation distance is greater, corresponding to the longer TSV depth. However, the sample with cross-section through the center of the TSV array axis shows a markedly different trend. Like the other profiles, the stress in Si near the top of the array is tensile. However, rather than trending toward compression right away, it becomes more tensile to a maximum value ~ 5 µm from the top of the TSV. It is notable that this maximum tensile point corresponds with the maximum compressive stress observed in the sample cross-sectioned through the Si region (black curve). After this point, the blue curve begins to show a consistent tensile relaxation over the depth of the TSV (~ 50 µm).

Comparing the experimental stress profiles to the obtained FE depth stress profiles in Figure 3.37, a fairly qualitative agreement is witnessed. In the case of the 5×25 µm samples in (a), the tensile stress near the Si surface becomes relaxed and moves to a compressive condition. This qualitatively agrees with the experimental data. There is divergence from there, where the experimentally obtained stress profiles show a consistent relaxation until the bottom of the TSV. The FE model, on the other hand, shows a quick relaxation after the first instance of compression, which then becomes tensile at the bottom of the TSV. The differing trend seen in the 5×50 µm TSV array for the center profile does have some qualitative agreement with the corresponding FE
model. After the initial tensile relaxation in the FE model, the stress becomes more tensile, displaying a parabolic type shape, before relaxing near the bottom of the TSV. Though the behavior of the stress profile differs from the experiment, the fact that a tensile stress is observed in both cases is clear and can only be said to confirm the measurements. However, further investigation is warranted to understand the origin and nature of the stresses observed along the depth of the TSV.

Depth profile results obtained with Micro-Raman spectroscopy yield qualitative agreement with top-down Raman. Furthermore, a maximum tensile stress near the Si/SiO₂ interface with relaxation over the top ~ 2 µm of the Si surface agrees with what was observed in the multi-wavelength, top-down Raman measurements. FE depth modeling yields somewhat qualitative agreement with the Raman measurements. Further investigation is warranted to clarify the differences observed.
3.6 CONCLUSIONS

Micro-Raman spectroscopy was used to determine residual stress levels in Si surrounding Cu TSVs for various array sizes, anneal conditions, and process history and comprises a relatively comprehensive study of stress evolution near Cu TSVs as a function of the aforementioned variables. Additional analysis by AFM topography and FIB cross-sections lend additional information on the corresponding Cu TSV deformation evolution. Residual stress is attributed to plastic yielding of Cu due to thermal stresses that arise as a result of the CTE mismatch between Cu and Si or from deformation within the Cu due to defect motion or grain boundary evolution during ‘self-annealing’.

Raman stress profiles were first conducted on as-received wafers to understand the stress profiles surrounding isolated and arrayed Cu TSVs. The profiles obtained show the characteristic tangential compression close to the TSV with radial tension farther away from the TSV before relaxation at distances ~ 15 µm away. These results agree with previous results [OKORO, 2008a] and are a result of Cu plastic deformation. Multi-directional line scans and 2D Raman maps show a four-fold symmetric stress profile with higher stresses along the [110] directions due to Si anisotropy. Furthermore, profiles collected on 1×4 and 8×8 Cu TSV arrays showed stress superposition effects between adjacent TSVs, where the amount of stress addition is proportional to the number of TSVs in an array.

Thermo-mechanical finite element analysis for both isolated and 1×4 Cu TSVs is in qualitative agreement with the Raman stress profiles. To reach stress levels comparable to a wafer that has seen a 150 °C post-CMP anneal, the FE model assumes an effective
anneal temperature of 350 °C. The difference between model and actual anneal temperatures is attributed to self-annealing of the Cu. Self-annealing results in a residual stress profile that resembles that of an effective anneal cycle. This has been confirmed with stress profiles obtained on a wafer that has undergone no thermal anneal treatment (wafer C). The non-annealed wafer yielded residual Si stress profiles near $1 \times 4$ Cu TSVs that resembled those of a wafer that had undergone thermal treatment. This confirms the fact that the Cu undergoes irreversible deformation at room temperature after electroplating. Therefore, it can be modeled with an effective anneal cycle, which results in plastic deformation. The result is that the simulated anneal cycle for a sample that has undergone annealing is modeled with a higher temperature than what was actually performed on the wafer.

Ex-situ temperature anneals show that the tensile stress in Si near TSVs relaxes monotonically with anneal temperature. At sufficient temperatures, the stress can become compressive. However, this trend has been shown to correlate with protrusion of Cu from the free surface of the TSVs. In addition, Cu oxidation can complicate the correlation between Cu protrusion and stress relaxation; therefore, it is recommended that anneal studies be conducted in an inert environment. Annealing time is expected to induce creep phenomena in the Cu TSVs, which may have been the cause of void formation for the wafer that underwent a 350 °C anneal (Figure 3.20e); however, the effect on the stress evolution was not confirmed to be significant.

Comparison of stress profiles for isolated and arrayed Cu TSVs between wafers that had undergone different amounts of processing showed that stress variation became significant for a longer process history. Wafer B, which had seen further processing than
wafer A, including higher temperature processes, showed similar stress magnitudes as wafer A. However, the die-to-die variation, as well as variation between adjacent TSVs was significantly increased. This could prove to be a concern for reliability and yield. However, it also implies that higher temperatures do not induce significant stress magnitude changes. This was partially confirmed with the ex-situ annealing studies, where significant changes only occurred in conjunction with Cu pumping.

Cross-sections of 1×4 Cu TSV arrays for both 5×25 μm and 5×50 μm TSVs were used to collect Raman profiles along the depth of the TSVs. The results qualitatively agreed with FE models for similar depth profiles. This is the first time a direct measurement of the stress as a function of TSV depth has been performed. It shows that the stress is highest near the top of the TSVs. Relaxation occurs over the depth of the TSV. However, the differences in the stress behavior for the different TSV lengths (25 vs. 50 μm) showed that analysis of the stress evolution over the TSV depth is not necessarily trivial. Nonetheless, this protocol provides a path from which to move forward for understanding these differences.

The measurements performed provide insight into the process-induced, thermo-mechanical stress evolution in Si as a result of Cu plastic deformation. Superposition effects, annealing conditions, and process history all appear to have an effect on the resultant residual Si stresses. Therefore, they should be taken into account when designing process flows and determining TSV array placement near devices.
FUTURE WORK

A number of possible pathways exist for extending Raman stress measurements on TSV structures. Some of the recommendations for future work include the following:

• More complete study to determine correlation between Cu grain growth with residual Si stress. This would include measurements for an as-received wafer that has undergone no thermal treatment, where measurements are conducted at various times after Cu plate and CMP to see the stress evolution due to self-annealing. Raman measurements conducted immediately after Cu plate and CMP would give an idea of the intrinsic Cu stress that can be separated from self-annealing. FIB cross-sections performed after each measurement will give an idea of Cu grain size to be correlated to each stress measurement. EBSD may be beneficial in determining grain sizes and orientations.

• Expanding ex-situ annealing of Cu TSVs with multiple die annealed at similar temperatures. Correlating Raman stress measurements with FIB cross-sections for multiple samples undergoing similar anneal conditions will provide better statistics on residual stress in Si and correlations with Cu grain growth and possible void formation. AFM topography will be included to get an idea of Cu protrusion variability.

• Repeating ex-situ anneals with TSVs that have been capped with an oxide. Since significant Si stress reduction occurs with increasing Cu protrusion, this trend may be altered if Cu protrusion is inhibited. Cross-sections are recommended to see the grain structure and void formation evolution in the TSVs. These may become more significant in the absence of Cu protrusion as a relaxation mechanism.
• Perform stress depth measurements for all cross-sectioned samples to see the residual Si stress evolution. The cross-section approach has been demonstrated to be viable for determining stresses along the entire length of a Cu TSV. Correlating all above measurements with depth profiles will provide a more complete picture of the 3D stress evolution in Si near Cu TSVs.

• Determine stresses in Si near the device layer for bonded wafers and wafers that have undergone processing beyond what has been seen in this dissertation. For bonded wafers, cross-section measurements will be the only way to access the appropriate region.

• Perform stress measurements on wafers that have functioning active devices processed at various distances and along various crystalline orientations around isolated and arrayed Cu TSVs. These sorts of test structures will allow for direct correlation between the residual Si stress and electrical performance of a MOSFET. Correlating stress with electrical performance is, in effect, the most important measurement that can be conducted. It will allow for FE model qualification as well as aid in determining Keep-Out-Zones for Cu TSVs.
PART FOUR: SUMMARY AND GENERAL REMARKS

The focus of this dissertation was the application of Raman spectroscopy for Si IC stress characterization. For FEOL applications, measuring stress in the Si channel region of strained MOSFETs is desirable. This is likely to hold true for future technology nodes, since MOSFET strain engineering is currently one of the most important equivalent scaling methods available. In addition, 3D integration is poised to substantially expand its role in IC processing. The increased device density in a smaller form factor, increased performance, and greater functionality all lend to its overall scaling benefits. However, due to the integration of Cu TSVs through the device regions, thermo-mechanical stress development due to the CTE mismatch between Cu and the surrounding Si can cause variation in performance of nearby devices. Therefore, it becomes apparent that understanding the stress evolution in the Si surrounding Cu TSVs is important for design and process control.

Part One of this dissertation provided a general introduction and overview of the document. Part Two of the dissertation was an evaluation of high-resolution Raman spectroscopy (TERS) for potential use for FEOL stress characterization. Currently, TEM based techniques are used for FEOL channel region stress characterization. However, due to the sample destruction and possible stress relaxation, in addition to extensive modeling and sample preparation required for such techniques, a quick and non-destructive technique is desirable. Traditional Raman spectroscopy is limited in spatial resolution due to the diffraction of light. However, integration of an apertureless NSOM with a Raman spectrometer setup provides the potential of sub-10 nm spatial resolution. So-called TERS relies on enhanced near-field scattering from a noble metal coated probe.
in the vicinity of the sample. The spatial resolution is approximately on the order of the probe diameter. Therefore, by decreasing the probe diameter, the spatial resolution can be increased. Commercial tips were investigated in this dissertation, providing spatial resolution on the order of 200 nm (the approximate diameter of the probe). Though the spatial resolution is still greater than what can be used for current FEOL devices, it is below the diffraction limit of the incident light. Moreover, sub-diffraction limited, localized stress variations were measured on strained SOI mesa test structures. It is expected that decreasing the probe diameter will increase the spatial resolution. To this end, a novel Au-CNT hybrid nanowire TERS probe is proposed and fabrication is demonstrated, with probe diameters on the order of 20 nm (see Appendix A).

Part Three of this dissertation focused on monitoring the residual Si stress evolution near Cu TSVs with micro-Raman spectroscopy. Residual Si stress is thermally driven, due to the CTE mismatch between the TSV and substrate, which leads to plastic deformation of the Cu. Measurements were conducted at various annealing temperatures to observe the effect of temperature on the surrounding Si. For increased annealing temperature, the residual Si stress is found to become more relaxed. However, the relaxation appears to occur due to Cu protrusion from the TSV free surface. Measurement of stresses on wafers that underwent different processing conditions showed similar magnitudes of residual Si stress. The main difference being increased die-to-die variation for the wafer that had a longer process history, including higher temperature steps. In addition, Raman measurements on non-annealed samples showed conclusively that Cu self-annealing is the origin of a significant amount of initial residual stress in Si. Finally, a cross-section Raman technique is demonstrated for application to
Si stress measurements over the depth of the TSVs. This approach will be useful after wafer bonding, where the active region is buried beyond optical penetration depths in Si. Measurement correlation with FE modeling will be indispensible for explaining the stress origins and evolution, as well as for creating predictive numerical models.

Raman spectroscopy is a quick and non-destructive technique for measuring stresses in Si. It is highly useful for BEOL applications, especially for monitoring residual stresses in Si for a 3D process flow. With a few modifications, Raman spectroscopy can be used for high-resolution, sub-diffraction limited scanning for FEOL applications. This dissertation introduced these concepts and further demonstrated the applicability and utility of Raman spectroscopy for Si IC stress characterization.
A.1 INTRODUCTION AND MOTIVATION

As mentioned in Part Two, one of the main limitations of apertureless NSOM is the probe. Variability in tips can lead to strong variations in near-field enhancement. Therefore, a highly repeatable manufacturing approach is desirable. However, further restrictions are imposed. Probes must be of small diameter (< 30 nm for current FEOL devices), provide strong near-field enhancement, and be mechanically robust.

A potential probe that can fulfill all the above requirements is a Au-CNT hybrid nanowire. Au-CNT hybrid nanowires have been fabricated by LUO et al. (2002) and OU et al. (2006). This hybrid nanowire consists of two connected segments, one a Au nanorod and the other a multi-walled carbon nanotube (MWCNT). Fabrication is accomplished with a template fabrication method, which can provide up to $10^{11}$ nanowires per square cm with highly uniform geometries [MARTIN, 1994; HURST, 2006]. In addition to large-scale, uniform fabrication, the hybrid nanowires will provide high spatial resolution, high-aspect ratio scanning, and mechanical robustness due to the CNT portion. These benefits have been previously demonstrated for CNTs attached to AFM probes [DAI, 1996; WONG, 1998a,b; HAFNER, 2001; ZHAO, 2008]. Finally, Au nanorods display strong plasmon resonances, and by controlling the Au segment length the plasmon resonance can be tuned to different wavelengths of excitation [MÜHLSCHLEGEL, 2005; NOVOTNY, 2007; KAPPELER, 2007; BRYANT, 2008]. Therefore, Au-CNT hybrid nanowires fabricated by a template method will provide uniform and robust wavelength-tunable, high-resolution apertureless NSOM probes on a large-scale.
Fabrication of hybrid nanowires using an AAO template method will be described, followed by SEM images of the resulting wires, and lastly methods used for attaching the hybrid nanowires to sharpened tungsten (W) wires for use in the Nanonics MV2000 NSOM system.

A.2 TEMPLATE FABRICATION OF HYBRID NANOWIRES

Template fabrication is performed using an anodized aluminum oxide (AAO) template. MARTIN (1994) pioneered this nanoscale fabrication method. AAO templates consist of highly ordered pores with uniform diameters. The essence of nanowire fabrication using AAO templates is filling the pores with the desired material, and then dissolving the AAO template so that the nanowires can be extracted.

To fabricate Au-CNT hybrid nanowires, Au is first electrochemically deposited, and then CNTs are grown on top. The processing steps follow the steps outlined by OU et al. (2006) and were done in collaboration with Dr. Pulickel Ajayan at Rice University (see Figure A.1 for illustration of steps). Both commercially available (Whatman Anodisc 25) and homemade AAO templates (anodization of aluminum foil) were used for fabrication. Pore diameters were on the order of 200 and 20 nm for the commercial and homemade templates, respectively. The small diameters of the hybrid nanowires are on the appropriate length scales for high-resolution TERS.
Figure A.1. Schematic of Au-CNT hybrid nanowire template fabrication method (courtesy of F.S. Ou). An AAO template is back-coated with Ag, Au is electrodeposited, and then CNTs are formed via CVD. Finally, the Ag backcoating is removed (HNO₃), followed by dissolving the AAO template (NaOH).

The first step in Au-CNT hybrid nanowire fabrication is electrodeposition of Au into the AAO template. Before Au can be deposited, the backside of the AAO template is coated with Ag, via thermal evaporation (Denton Vacuum BTT-1V), to serve as the working electrode for deposition. The backside Ag layer acts as the cathode for electrodeposition, where the positively charged metal ions in the electrolyte will be reduced to yield metal deposits in the AAO pores (\(\text{Au}^+ + e^- \rightarrow \text{Au}\)). After formation of the backside cathode layer, an Autolab PGSTAT potentiostat system is used to deposit a thin layer of Cu (electrolyte: CuSO₄; ~90 s deposition) inside the AAO template pores, to improve adhesion of the Au. Next, Au is deposited using the same potentiostat system and a proprietary Au electrolyte solution. Varying the deposition time controls the length of the Au wire segment. For example, a 5 min. electrodeposition results in ~200 nm Au segments.
After the Au wire segment is formed, chemical vapor deposition (CVD) is used to form multi-walled carbon nanotubes (MWCNTs) inside the pores of the AAO template. Since Au fills the bottom of the pores, the MWCNTs will grow in the remaining top portion. CVD occurs via pyrolysis of acetylene at 650 ºC. A 1 hr deposition results in filling of the AAO template, yielding hybrid nanowires ~40 µm in length. After CNT formation, there is an amorphous layer of carbon on the top surface of the AAO template. This is removed by plasma etching (Harrick Plasma Cleaner PDC-32G).

The backside Ag and Cu layers were selectively dissolved in HNO₃ to facilitate release of the hybrid nanowires from the AAO template. The hybrid nanowires were then released by dissolving the template in 3M NaOH. The solution was cleaned by rinsing with DI water and centrifuging. This was repeated 4-5 times to ensure complete removal of the template and NaOH. The hybrid nanowires were then dispersed in ethanol and sonicated to separate bundles of the wires. The dispersion was drop cast onto segments of Si wafer for SEM characterization.

Figure A.2 shows SEM images for Au-CNT hybrid nanowires fabricated by the commercial (a) and homemade (b) AAO templates, respectively. In image (a), the hybrid nanowires are seen in a bundle. This was due to suboptimal sonication. The ~2.5 µm Au segments are a result of 1 hr electrodeposition. Diameters of the wires are ~200 nm, corresponding to the commercial template pore size. The diameters of the nanowires fabricated with a homemade AAO template are on the order of 20 nm, corresponding to the template pore size. The Au segments on the smaller diameter wires in this case are <500 nm, a result of 30 min electrodeposition. When comparing the deposition times and corresponding Au segment lengths, one must also take into account the differences in
pore diameter. Therefore, optimization of Au segment lengths must include deposition
time as well as pore diameter. However, as a proof-of-concept, fabrication of Au-CNT
hybrid nanowires with diameters <30 nm is demonstrated to be successful. Therefore,
high spatial resolution TERS is expected to become a reality with these nanowires, if they
are successfully integrated onto a Nanonics probe mount.

![SEM images of Au-CNT hybrid nanowires fabricated using (a) commercial and (b) homemade AAO templates. Diameters are ~200 and 20 nm, respectively.]

Figure A.2. SEM images of Au-CNT hybrid nanowires fabricated using (a) commercial and (b) homemade AAO templates. Diameters are ~200 and 20 nm, respectively.

A.3 ATTACHMENT OF HYBRID NANOWIRES TO SHARPENED TUNGSTEN WIRES

The approach taken is to attach the hybrid nanowires to a sharpened W wire, which is
then attached to a Nanonics tuning fork mount. Sharpened W wires have been
successfully used with a Nanonics NSOM system previously [ATESANG, 2006].
Therefore, the main challenge is attaching the hybrid nanowires to a sharpened W wire.

Two methods were explored: micromanipulation and dielectrophoresis.
Micromanipulation involves mechanical placement of the nanowires with x, y, z micro-
translation stages. Dielectrophoresis involves electric field induced forces that cause the nanowires to align and move toward the sharpened W tip.

A 3-axis micromanipulator (Narishige) was used to attach an Au-CNT hybrid nanowire to a sharpened W wire. Hybrid nanowires fabricated with the commercial templates (diameter ~200 nm) were used since this technique requires visual feedback. The experimental setup is seen in Figure A.3a. First, a pulled glass rod picks up a hybrid nanowire via static charge. A second glass tip is coated with UV curable epoxy. Using this tip, the epoxy is applied to the sharpened W wire. The first tip with the nanowire is then used to position the hybrid nanowire for attachment to the epoxy coated W wire. After attachment, the sharpened W wire is cured for 1 hr under UV illumination (366 nm). Figure A.3b is an SEM image of an Au-CNT hybrid nanowire attached to a sharpened W wire with UV curable epoxy.

Figure A.3. Schematic (a) of the micromanipulator setup and SEM image (b) of a hybrid nanowire (diameter ~200 nm) attached to a sharpened W wire with UV curable epoxy.
Micromanipulation is a workable technique; however, it has several main drawbacks. The most general problem, which is related to the other drawbacks, is that attachment is highly dependent upon the operator. Picking up nanowires from a substrate, coating the UV curable adhesive, and attaching the nanowire to the W wire is very time consuming (~1-2 tips per day) and is more up to chance and skill than anything else. Breaking the glass tips is fairly common and lends to the difficulty and time of this approach. Therefore, a quicker and more repeatable procedure is desired.

Dielectrophoresis is an approach that has been successful in attaching CNTs to AFM probes [Park, 2005; H.W. Lee, 2005; Wei, 2008; Hao, 2008]. The dielectrophoretic force results from an asymmetric electric field between a sharp and flat electrode, as illustrated schematically in Figure A.4a. When an electric field is incident on a polarizable object, such as a CNT, a dipole moment is induced. This dipole moment causes the CNT to align with the asymmetric electric field. Furthermore, due to inhomogeneity of the field, there is an imbalance of forces acting on the CNT, thus causing it to move along the electric field gradient. The direction of the force depends on the relative polarizability of the CNT to the medium in which it is suspended, in other words, the relative permittivity. The induced dielectrophoretic force on the CNT is described by the following equation [Dimaki, 2004]:

\[
F_{DEP} = \Gamma \cdot \varepsilon_m Re\{K_f\} \nabla |E|^2
\]  

(A.1)

where \( \Gamma \) is a factor dependent upon geometry, \( \varepsilon_m \) is the real part of the permittivity of the suspending medium, \( E \) is the electric field, and \( K_f \) is a term that depends upon the
complex permittivity of the CNT and the suspending medium. For a cylindrical object, $\Gamma$ is given by:

$$\Gamma = \frac{\pi}{6} r^2 l$$  \hspace{1cm} (A.2)$$

where $r$ is the radius of the cylinder and $l$ is the length. For an elongated object aligned along the axis, $K_f$ is given by:

$$K_f = \frac{\varepsilon_p^* - \varepsilon_m}{\varepsilon_m^*} \hspace{1cm} (A.3a)$$

$$\varepsilon^* = \varepsilon - i \frac{\sigma}{\omega} \hspace{1cm} (A.3b)$$

where $p$ and $m$ denote the particle (CNT) and medium (ethanol), respectively; $\sigma$ is conductivity; and $\omega$ is angular frequency of the applied electric field. Therefore, the dielectrophoretic force is dependent on the frequency of the electric field and the relative permittivity of the hybrid nanowire to the suspending medium. For frequencies in the MHz range and lower, the dielectrophoretic force is always in the direction of the sharp electrode (W wire) [Kim, 2005]. A 10 V, 2 MHz electric field was used for the dielectrophoresis results obtained in this dissertation.

Tip-sample separation distance is important, and 10 $\mu$m has been determined to be the upper limit [H.W. Lee, 2005; Kim, 2005] for successful attachment of CNTs to AFM probes. To establish this separation, an oscilloscope was connected in parallel to the tip and sample to monitor the AC signal, while the W tip was moved toward the aluminum electrode. Vertical movement was performed using a 3-axis translation stage (differential screws provide down to 10 nm translation step sizes). When the tip contacts the aluminum electrode, the AC signal shows a short, and the W wire is backed away by 10
µm. The tip is maintained in the field for several minutes before being shut-off. The result of a single Au-CNT hybrid nanowire attached to the end of a W wire is seen in Figure A.4b.

Figure A.4. Schematic (a) of the dielectrophoresis setup and SEM image (b) of a hybrid nanowire (diameter ~20 nm) attached to a sharpened W wire. Asymmetric electric field lines cause nanowire to orient parallel to W wire. Dielectrophoretic force is in the direction of the W tip.

A.4 SUMMARY

A novel Au-CNT hybrid nanowire TERS probe is proposed and feasibility of system integration via attachment to sharpened W wires is demonstrated. The combination of a CNT and Au nanorod segment is expected to provide a high-aspect ratio, mechanically robust, and high-resolution probe with tunable near-field enhancement. Using an AAO template method provides highly uniform, large-scale fabrication with controllable geometries. This approach has been demonstrated with both commercial and homemade AAO templates.
Integration of Au-CNT hybrid nanowires requires an approach of attaching to a sharpened probe that can be readily attached to a Nanonics tuning fork mount. Two approaches are demonstrated: micromanipulation and dielectrophoresis. Both have proven successful; however, micromanipulation is time consuming and repeatability may be an issue. However, dielectrophoresis provides attachment in a time frame of minutes. Furthermore, it is highly repeatable and could prove useful for large-scale TERS probe manufacturing.

FUTURE WORK

Testing of the novel probe design requires AFM stability testing with the Nanonics system and characterization of near-field enhancement. Stability can be tested once a fully integrated hybrid-nanowire probe is attached to a Nanonics mount. If stable feedback is obtained between the probe and sample, then mechanical characterization via force-distance curves and topography scans become possible.

Near-field enhancement can be tested using the blanket sSOI and strained SOI mesas used in Part Two. Comparison of sSOI mesa profiles between the hybrid nanowire probes and commercial probes (Nanonics) will provide useful information on enhancement and spatial resolution and probe geometry. Due to the much smaller probe diameters (~20 nm), the Au-CNT hybrid nanowires can be tested on smaller structures, such as Si nanowires (SiNWs) on highly ordered pyrolitic graphite (HOPG). This sample system will allow for isolation of the Si peak, and the FWHM of a transverse scan across the SiNW will provide an idea of spatial resolution for this novel TERS probe.
BIBLIOGRAPHY


http://www.casaxps.com/release/release2312_/CasaXPS.HLP/SpectrumProcessing/EdgeMeasurements.htm


De Munck, K., et al. (2005): “3D interconnect technology for space applications,” *ESA Round Table on Micro/Nano Technologies for Space Location*,
https://escies.org/GetFile?rsrcid=1706


Intel (2007): “High-k and Metal Gate Research,”
http://www.intel.com/technology/silicon/high-k.htm

International Technology Roadmap for Semiconductors (ITRS) (2009),


http://vmsstreamer1.fnal.gov/VMS_Site_03/Lectures/Colloquium/presentations/070228Keast.pdf


