Ultra-thin bulk silicon thinning of advanced microprocessors & graphics processors fabricated in Sub-20 nm technology

Teo, Wee Siang

2020

Teo, W. S. (2020). Ultra-thin bulk silicon thinning of advanced microprocessors & graphics processors fabricated in Sub-20 nm technology. Master's thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/144132>

<https://doi.org/10.32657/10356/144132>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 16:57:24 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**ULTRA-THIN BULK SILICON THINNING OF ADVANCED
MICROPROCESSORS & GRAPHICS PROCESSORS
FABRICATED IN SUB-20 NM TECHNOLOGY**

TEO WEE SIANG

SCHOOL OF MATERIALS SCIENCE AND ENGINEERING

2020

**ULTRA-THIN BULK SILICON THINNING OF ADVANCED
MICROPROCESSORS & GRAPHICS PROCESSORS
FABRICATED IN SUB-20 NM TECHNOLOGY**

TEO WEE SIANG

SCHOOL OF MATERIALS SCIENCE AND ENGINEERING

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Master of Engineering

2020

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

30 June 2020

.....
Date

.....
Teo Wee Siang

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

30 June 2020

.....
Date

.....
Gan Chee Lip

Authorship Attribution Statement

This thesis contains material from 1 paper accepted at a conference in which I am listed as an author.

Chapter 4 is published as W.S. Teo, M.S. Wei, V. Narang, C.L. Gan, C. Richardson, and G. Liechty, “Targeted Silicon Ultra-thinning by Contour Milling for Advanced Fault Isolation,” in Proceedings of the 45th International Symposium for Testing and Failure Analysis, 2019.

The contributions of the co-authors are as follows:

- M.S. Wei suggested the initial project direction and scope.
- I designed the study and performed all the laboratory work at the Advanced Micro Devices (Singapore) Device Analysis Laboratory. I also analysed the data and prepared the manuscript draft.
- The manuscript was revised by V. Narang and C. L. Gan.
- C. Richardson and G. Liechty provided the software with which the experiments were conducted.

30 June 2020

.....
Date

.....
Teo Wee Siang

Abstract

The continual advancement of integrated circuits has seen the relentless scaling of minimum dimensions for semiconductor devices. Consequently, the burgeoning complexity of process technology and devices fabricated has necessitated advancements in device characterisation all through the failure analysis workflow. In particular, initial fault isolation techniques exploit infrared laser beams to probe the circuitry through silicon thinned mechanically to 50 μm . Moving down to sub-20 nm process nodes, the infrared based lasers are challenged to resolve minute features as it hits its resolution limits. There is thus a need to use shorter wavelengths beam probes in the visible spectrum and beyond. However, visible light's higher absorption coefficient in silicon necessitates silicon to be ultra-thinned to under 10 μm to reduce signal attenuation.

With existing contour milling techniques limited by its precision when ultra-thinning silicon, experiments were then conducted to understand key challenges in terms of the sample behaviour during thinning. Thereafter, a modified workflow was proposed to account for sample relaxation during ultra-thinning and tested on actual devices. With ultra-thinned samples prepared, challenges relating to the device's thermal stability were then subsequently encountered. In overcoming these difficulties, targeted 2D and 3D thinning techniques were then explored and validated on actual devices prepared.

In anticipation of further advanced probing techniques necessitating sub-micron remaining silicon thicknesses, focused ion beam milling techniques were then explored as a means to further localise ultrathin silicon to the area of interest. As a means for precise material removal, the milling rate was first characterised on different systems before the resulting milling finish was studied as a function of the beam parameters. The proposed workflow was then examined with its application on actual thinned devices. Process challenges and limitations were then discussed in relation to its application.

Lay Summary

Advancements within the semiconductor industry has seen the integration of increasingly numerous and miniaturized transistors into modern electronics. As a staple of product development, failure analysis has been key to identifying process limitations and enabling said advancements. However, the decreasing scale with which devices are fabricated necessitates greater precision in the methods which defects are identified with. One solution lies in the ultra-thinning of silicon on a device to under 10 μm so that more precise lasers may be used for probing.

While existing methods of thinning have been employed to thin silicon to 50 μm thickness, its precision is limited and presents a high risk of device destruction when used for ultra-thinning. In this work, the limitations behind this process are studied and subsequently improved upon to successfully ultra-thin devices. However, devices are prone to overheating and destruction when minimal silicon remains after ultra-thinning to dissipate heat during testing. To circumvent these challenges, the thinning process is further modified so as to localise the ultra-thin silicon only to the area of interest and improve heat dissipation. Samples are then prepared with the proposed changes to test its efficacy.

With the anticipated need for greater precision in defect localization, precise methods for further silicon ultra-thinning are then studied using focused ion beam tools. As a means to controllably thin silicon, its milling rate and impact on the sample under varying beam conditions is then studied by experimentation. With knowledge of its milling characteristics, a workflow is then proposed and studied by milling on actual ultra-thinned devices. Challenges and limitations faced during the process are then discussed in relation to the intended application and instrument.

Acknowledgements

This project is carried out under the industrial postgraduate program sponsored by the Economic Development Board (EDB) with Advanced Micro Devices (AMD), Singapore. The author would like to thank the following people who help in contributing to the success of this project:

- Mr. Vinod Narang (AMD), Manager Physical Failure Analysis, for providing invaluable guidance and encouragement throughout the project.
- Professor Gan Chee Lip (NTU), Project Supervisor, for providing the opportunity and guidance to the conduct of the project
- Colleagues of Device Analysis Lab (AMD), for support throughout the project and for help with collection of data
- Staff of Temasek Lab (NTU), for sharing their experiences and providing invaluable advice in the conduct of the experiments.

Table of Contents

Abstract	i
Lay Summary	ii
Acknowledgements	iii
Table of Contents	v
Table Captions	ix
Figure Captions	xi
Abbreviations	xxi
 Chapter 1 Introduction.....	 1
1.1 Background	1
1.2 Motivation	2
1.3 Objectives and Scope	3
1.4 Expected Outcomes/Challenges	3
 Chapter 2 Literature Review	 5
2.1 SIL Imaging for EFI.....	5
2.2 Device Warpage	8
2.3 CNC Contour Milling	11
2.4 FIB Silicon Trenching for Higher Resolution Probing	14
References.....	23

Chapter 3	Contour Milling	27
3.1	Ultra-thinning of Sample by Global Contour Milling	27
3.1.1	Experimental Setup and Process	27
3.1.2	Results and Discussion	29
3.2	Characterization of Sample Warpage Behavior During Thinning	35
3.2.1	Experimental Setup and Process	35
3.2.2	Results and Discussion	36
3.3	Ultra-thinning with Incorporation of Stress Relaxation Step	40
3.3.1	Experimental Setup and Process	40
3.3.2	Results and Discussion	41
3.4	Thermal Stability of Ultra-thinned Devices During Dynamic Testing	46
3.4.1	Experimental Setup and Process	46
3.4.2	Results and Discussion	46
3.5	Conclusion	49
Chapter 4	Targeted Milling	51
4.1	Targeted 2D Milling	52
4.1.1	Experimental Setup and Process	53
4.1.2	Results and discussion	53
4.2	Targeted Contour Milling	58
4.2.1	Experimental Setup and Process	59
4.2.2	Results and Discussion	60
4.3	Polishing Non-Uniformity After Targeted Milling	70
4.3.1	Experimental Setup and Process	70
4.3.2	Results and Discussion	70
4.4	Conclusion	73

Chapter 5	Localised FIB Milling	75
5.1	Ga FIB Milling Rate Characterisation	76
5.1.1	Experimental Setup and Process	76
5.1.2	Results and Discussion	78
5.2	Xe PFIB Milling Rate Characterisation	84
5.2.1	Experimental Setup and Process	85
5.2.2	Results and Discussion	86
5.3	Xe PFIB Milling Finish Characterisation	101
5.3.1	Experimental Setup and Process	102
5.3.2	Results and Discussion	103
5.4	Localised FIB Milling on Ultra-thinned Samples	123
5.4.1	Experimental Setup and Process	123
5.4.2	Results and Discussion	126
5.4.3	Comparison Between Different Milling Processes	137
5.5	Conclusion	140
Chapter 6	Conclusion and Recommendations	141
6.1	Conclusion	141
6.2	Recommendations for Future Work	142

Table Captions

Table 2.1 Material properties of chip and package (substrate) [10].

Table 2.2 Damage layer thickness induced by FIB milling [26-29].

Table 3.1 Table of RST values through sequential polishing steps from left to right.

Table 4.1 Table of RST values through sequential polishing steps from left to right and difference between global and local RST.

Table 4.2 Table of RST values through sequential polishing steps from left to right and difference between global and local RST.

Table 5.1 Table of milling parameters and results.

Table 5.2 Table of milling parameters used to mill 200 μm square windows over 10 mins.

Table 5.3 Table of lamella characteristics.

Table 5.4 Comparison between mechanical and FIB milling.

Figure Captions

Figure 2.1 Test setup schematic and close up view of laser focusing on device under test (DUT) with hemispherical SIL [4].

Figure 2.2 Absorption coefficient and penetration depth in silicon at varying wavelengths [4].

Figure 2.3 a) Typical flip chip cross-sectional schematic and b) plan view of an actual device.

Figure 2.4 a) Physical profile map of actual sample and b) its schematic cross-section.

Figure 2.5 a) Polisher for parallel lapping and b) CNC milling system [13].

Figure 2.6 RST profiles by parallel lapping vs contour milling.

Figure 2.7 Process flow for global contour milling.

Figure 2.8 Cross section and plan view schematic of microprocessor 3D integration [17].

Figure 2.9 Cross-sectional schematic of finFET [20].

Figure 2.10 FIB images of trench at a) n-well (left) with smaller opening at b) ≈ 355 nm and c) ≈ 250 nm from STI [21].

Figure 2.11 Spot size and brightness vs beam current (left) and spot size with mill rate comparison between Ga FIB and Xe PFIB (right) [22].

Figure 2.12 Schematic illustration of the collision cascade for a 30 kV Ga^+ ion impinging on a crystal, with associated volume of damage as well as projected range R_p and lateral range R_l [25].

Figure 2.13 SRIM plots for 0° incidence angle ion bombardment in Si at 30 kV. [28].

Figure 2.14 HRTEM images of sidewall amorphisation in Si by Ga^+ and Xe^+ beam milling at varying accelerating voltage [27].

Figure 2.15 Typical $200\text{ }\mu\text{m} \times 200\text{ }\mu\text{m}$ trench etched into silicon with pitted surroundings [34].

Figure 3.1 Modified workflow for global ultra-thinning.

Figure 3.2 a) Initial physical contour map with visualization and b) RST map with visualization after 2D milling.

Figure 3.3 a) RST map with visualization after 3D milling and b) final RST map with visualization after polishing.

Figure 3.4 a) Initial physical contour map with visualization and b) RST map with visualization after 2D milling.

Figure 3.5 a) RST map with visualization after 3D milling and b) final RST map with visualization after overmilling.

Figure 3.6 Optical images of a) overall die, b) over-milled top left corner, and c) over-milled bottom right corner.

Figure 3.7 a) Initial physical profile map and b) RST map of die at full thickness.

Figure 3.8 RST maps collected at a) 700 μm , b) 40 μm , and c) difference between the two.

Figure 3.9 Warpage as measured by RST range vs RST for 202 mm^2 device.

Figure 3.10 Warpage as measured by RST range vs RST for 319 mm^2 device.

Figure 3.11 Modified workflow for ultra-thinning while incorporating a stress relaxation step.

Figure 3.12 a) RST map at 300 μm after 2D milling, b) physical map after 2D milling, c) physical map after overnight stress relaxation, and d) difference map showing overnight change.

Figure 3.13 a) RST map at 40 μm after 2D milling, b) physical map after 2D milling, c) physical map after overnight stress relaxation, and d) difference map showing overnight change.

Figure 3.14 a) Final RST map, b) whole die optical image, c) whole device image, and d) zoomed in optical image.

Figure 3.15 a) Final RST map, b) Whole die optical image.

Figure 3.16 a) Final RST map and visualization, b) angled optical image of burnt edge, c) magnified angle optical image, and d) magnified plan view optical image.

Figure 3.17 Cross-sectional SEM image of typical metal melts in devices failing to shorts.

Figure 4.1 Targeted 2D milling after a) global and b) local levelling.

Figure 4.2 a) Physical contour map of globally levelled sample and b) sample locally levelled about AOI boxed up in map.

Figure 4.3 RST maps collected after a) contour milling and b) polishing.

Figure 4.4 Optical image of thinned device with thinnest section visible and RST labelled in μm .

Figure 4.5 Laser scanning microscope image taken with 1340 nm laser at a) 5x, b) 10x, c) 20x, and d) 50x magnification.

Figure 4.6 FIB cross-sectional image of thinned sample at thinnest.

Figure 4.7 Targeted milling by RST map manipulation during contour map generation.

Figure 4.8 a) Original RST map and b) edited map with values increased in window.

Figure 4.9 Physical contour map after levelling a) globally, and b) locally about AOI.

Figure 4.10 a) Original RST map after milling and b) edited RST map with increased thickness at AOI.

Figure 4.11 Physical profiles after a) 10 mins, b) 1 hour, and c) overnight.

Figure 4.12 a) Physical profile map, b) edited RST map, and c) generated contour map.

Figure 4.13 a) RST map post-milling, optical images of device milled b) partially, and c) completely.

Figure 4.14 a) Final RST map and b) optical image of ultra-thinned device with AOI visible.

Figure 4.15 Large area die image taken with IR camera at 1x magnification.

Figure 4.16 Laser scanning microscope image taken with 1340 nm laser at a) 5x, b) 10x, c) 20x, and d) 50x magnification.

Figure 4.17 RST maps after a) 3D milling, b) 6 μm polish, c) final 1 μm polish, and d) overall polished RST.

Figure 5.1 Illustration of FIB interface with *in-situ* ion beam milling and electron beam imaging.

Figure 5.2 a) Initial surface SE image, b) milled window surface image, and c) trench cross section.

Figure 5.3 Beam spot sizes with associated beam tails at varying beam currents.

Figure 5.4 a) Etched trench SE image at tilt, b) left edge cross section, and c) right edge cross section.

Figure 5.5 SE images of a) trench with GIS in, b) remnant shadow, c) etched surface, and d) surrounding surface.

Figure 5.6 SE images of a) milled windows, b) 10 min window milled trench, and c) trench cross section.

Figure 5.7 SE images of a) milled windows, b) milled trench across 10 min window, and c) trench cross section.

Figure 5.8 Associated spot sizes and beam tails of optimised (left) and unoptimised (right) beams.

Figure 5.9 Optical images of various windows milled across different currents and time.

Figure 5.10 SE images of a) 200 μm and b) 700 μm windows and its cross sections.

Figure 5.11 SE images of a) window milled at 225 nA and b) its cross section.

Figure 5.12 SE images of a windows milled at a) 1 μA with c) its cross section, and at b) 2 μA with d) its cross section.

Figure 5.13 Cross-sectional image of windows milled at 2 μA for a) 10 mins, b) 30 mins, and c) 60 mins.

Figure 5.14 SE images of a) 200 μm window etched over 10 mins at 146 nA, b) surface pitting, and c) its cross section.

Figure 5.15 SE images of a) 300 μm window etched over 23 mins at 146 nA, b) surface pitting, and c) its cross section.

Figure 5.16 SE images of a) 400 μm window etched over 40 mins at 146 nA, b) surface pitting, and c) its cross section.

Figure 5.17 SE images of a) GIS nozzle configuration and b) etch window with inset of higher magnification images at various sections over 400 μm etched window.

Figure 5.18 Plot of mill depth versus dose at different beam energy.

Figure 5.19 a) Sample 1 SE image of window at 52° tilt and b) deposited marker.

Figure 5.20 TEM image of sample 1 at various magnification.

Figure 5.21 STEM images and EDS maps of sample 1.

Figure 5.22 a) Sample 2 SE image of window at 52° tilt and b) deposited marker.

Figure 5.23 TEM image of sample 2 at various magnification.

Figure 5.24 STEM images and EDS maps of sample 2.

Figure 5.25 TEM image of sample 3 at various magnification.

Figure 5.26 STEM images and EDS maps of sample 3.

Figure 5.27 a) Sample 4 SE image of window at 52° tilt and b) deposited marker.

Figure 5.28 TEM image of sample 4 at various magnification.

Figure 5.29 STEM images and EDS maps of sample 4.

Figure 5.30 a) Sample 5 SE image of window at 52° tilt and b) deposited marker.

Figure 5.31 TEM image of sample 5 at various magnification.

Figure 5.32 STEM images and EDS maps of sample 5.

Figure 5.33 Control sample's a) TEM image, b) STEM image and C) associated EDS map.

Figure 5.34 a) Sample 6 SE image of window at 52° tilt and b) deposited marker.

Figure 5.35 TEM image of sample 6 at various magnification.

Figure 5.36 STEM images and EDS maps of sample 6.

Figure 5.37 a) Sample 7 SE image of window at 52° tilt and b) deposited marker.

Figure 5.38 TEM image of sample 7 at various magnification.

Figure 5.39 STEM images and EDS maps of sample 7.

Figure 5.340 Plot of amorphisation layer thickness and implantation depth versus beam energy.

Figure 5.41 Flowchart of process to locally ultra-thin silicon to 500 nm by mechanical and FIB milling.

Figure 5.42 Flowchart of process to locally ultra-thin silicon to STI by mechanical and FIB milling.

Figure 5.43 Maps of a) initial contour profile, b) edited RST profile after milling, c) generated contour profile, and d) final RST profile.

Figure 5.44 a) SE image of milled window beside marker, b) low mag image of window, and c) high mag image under brighter illumination.

Figure 5.45 a) RST map across window, b) measurement at center, and c) physical representation.

Figure 5.46 LSM image taken at a) 5x, b) 10x, c) 20x, and d) 50x magnification.

Figure 5.47 a) Final RST map, b) RST visualization, and c) optical image of ultra-thinned die.

Figure 5.48 SE images of a) thin silicon approaching STI, b) STI exposed as observed at 52° , c) STI exposed as observed at 0° , and d) STI imaged at higher magnification.

Figure 5.49 SE images of a) window with silicon remaining, b) STI partially exposed, c) STI intact at top left corner, and d) STI over-etched in bottom right corner of window.

Figure 5.50 a) SE image of both windows, and LSM images of b) both windows at 20x, c) milled window at 50x, and d) etched window at 50x.

Abbreviations

AOI	Area of Interest
BEOL	Back End of Line
CMP	Chemical Mechanical Polishing
CNC	Computer Numerically Controlled
CTE	Coefficient of Thermal Expansion
C4	Controlled Collapse Chip Connection
EDS	Energy Dispersive Spectroscopy
EFI	Electrical Fault Isolation
FA	Failure Analysis
FEOL	Front End of Line
FIB	Focused Ion Beam
FinFET	Fin Field Effect Transistor
Ga	Gallium
GaAs	Gallium Arsenide
GAE	Gas Assisted Etching
GIS	Gas Injection System
HAADF	High Angle Annular Dark Field
HRTEM	High resolution transmission electron microscopy
IC	Integrated Circuit
IO	Input-Output
IR	Infrared
LSM	Laser Scanning Microscope
LVP	Laser Voltage Probing
NA	Numerical Aperture
PEM	Photon Emission Microscopy
PFA	Physical Failure Analysis
PFIB	Plasma Focused Ion Beam
RST	Remaining Silicon Thickness

SEM	Scanning Electron Microscope
STEM	Scanning Transmission Electron Microscopy
Si	Silicon
SIL	Solid Immersion Lens
SRIM	Stopping and Range of Ions in Matter
STI	Shallow Trench Isolation
TEM	Transmission Electron Microscope
VLP	Visible Light Probing
Xe	Xenon
XeF ₂	Xenon Difluoride

Chapter 1

Introduction

1.1 Background

Since the advent of microelectronics in the 1960s, time has seen the relentless scaling of minimum dimension for semiconductor devices to sustain continual performance and efficiency gains. As a result, greater number of increasingly complex and minute transistors are being crammed into integrated circuits (ICs). Noting this, advancements in device characterisation all through the failure analysis (FA) workflow becomes necessary in product development. From early product qualification and reliability testing to later yield debug and field returns, FA feedbacks weaknesses all along the way in design, fabrication, and testing for corrective action.

Much like searching for a needle in a haystack, the objective of failure analysis is to identify root cause defects leading to failure throughout the billions of transistors on the device. Broadly, after identifying a failing chip for device-level FA, it is subjected to electrical fault isolation (EFI) to narrow down suspects through non-destructive means before physical failure analysis (PFA) is conducted to pinpoint and visualize defects through destructive means. For any successful root-cause analysis, it is critical that the defect is reasonably isolated during EFI for detailed PFA through the many interconnect layers in the back-end-of-line (BEOL) as well as transistors in the front-end-of-line (FEOL).

Modern EFI techniques largely involve optical techniques of either emission microscopy or laser probing of the flip-chip devices through the silicon backside to isolate defective circuitry. According to Rayleigh's criterion, the resolution of these techniques is thus contingent on the numerical aperture (NA) of the objective lens and wavelength of light used. While the development and use of higher NA solid immersion lenses (SIL) has

extended the practicality of such techniques for increasing scaled devices, it too is reaching its physical limits. Thus, in tuning the other knob to improve the resolution, there has been great thrusts into developing techniques for use with shorter wavelengths of light in the visible region. However, the use of shorter wavelengths diminishes the key advantage of operating in infrared (IR), its optical transparency in silicon. To overcome challenges associated with attenuation of visible light in silicon, there is thus a need for the ultra-thinning of bulk silicon for advanced microprocessors fabricated in sub-20 nm technology.

1.2 Motivation

Sample preparation techniques in failure analysis has always played a key role in visualizing defects on increasingly miniaturized devices. While the introduction of the SIL to optical EFI techniques has made bulk silicon thinning necessary, the use of higher NA SILs for higher resolutions requires further thinning down to 50 μm . Thinning onto target thicknesses at the same scale as device warpage, traditionally used parallel lapping techniques were then made obsolete as computer numerically controlled (CNC) contour milling came into prevalence for preparing samples within $\pm 5 \mu\text{m}$ of variation in its remaining silicon thickness (RST) throughout the die. With the shift to visible light for probing, light attenuation has then set an upper bound on the RST at about 5 μm while preserving mechanical and electrical integrity of the device for further testing.

Thinning onto RSTs at under the tolerances of the CNC milling process thus presents high risk of overmilling and failure due to the limits of the machine's precision as well as device relaxation characteristics while thinning. Additionally, thinning of the devices onto the realm of micrometers also brings about challenges in thermal stability as there is less silicon to dissipate heat off the circuitry and runs a risk of destroying the part. This is unacceptable in a high-stake environment where successful analysis of the often one and only sample drives important engineering and business decisions.

In meeting the requirements for ultra-thin sample preparation, there is a need for further understanding of the device characteristics while thinning as well as alternatives for greater

precision in silicon removal. To this end, the use of focused ion beam (FIB) as well as plasma focused ion beam (PFIB) tools are studied for controlled, local silicon etches. This report thus describes and validates the design of a process that delivers a reliable method for the ultra-thinning of flip chip devices to enable visible light probing that extends the resolution limits for failure analysis applications.

1.3 Objective and Scope

The main objective of this project is to establish a method for ultra-thinning of bulk silicon to 5 μm thickness and below while preserving device integrity for visible light probing (VLP) using compatible high NA SILs. This objective can be achieved by the following:

1. Understanding the mechanical characteristics of the device when thinning from bulk silicon
2. Experimentation with various thinning processes, such as contour milling and plasma focused ion beam milling
3. Evaluation of the various processes and suitability for SIL imaging
4. Optimization of various process parameters and assessment of the processes

1.4 Expected Outcomes/Challenges

Current methods of sample preparation largely rely on well-established CNC contour milling techniques for thinning down to 50 μm with $\pm 5 \mu\text{m}$ tolerances. Targeting thicknesses under 5 μm run into high risks of overmilling due to machining precision, polishing variation as well as device relaxation with ultra-thinned silicon on a warped organic substrate. On the other hand, chemical methods of silicon removal afford the precision necessary but often lacks in throughput. Thus, there is a need to examine and tweak existing methods of sample preparation by physical means as well as to harness complementary chemical techniques for precise and local removal of silicon.

Chapter 2

Literature Review

This chapter aims to establish the need for ultra-thinning in the failure analysis workflow by introducing electrical fault isolation techniques that includes visible light probing (VLP). Techniques for silicon thinning such as CNC contour milling and FIB milling are also discussed for a deeper understanding of their merits and constraints. Taken together, these will then form the basis upon which the project will be conducted.

2.1 SIL Imaging for EFI

At the start of any device-level failure analysis is EFI to isolate possible defects prior to destructive testing and inspection. The flip chip device is set on a tester and test patterns are run to simulate failing conditions as several optical techniques are used to study the failure from the backside through silicon.

Anomalous photon emission emitted from failing regions can then be collected and analyzed via photon emission microscopy (PEM) [1] or an IR laser can be used to probe for abnormal transistor activity via laser voltage probing (LVP) [2]. Since the inception of such techniques, the devices studied have shrunk from micron scale to state-of-the-art sub-20 nm minimum dimensions and the technique has seen continual advancements to extend its resolution limits. With respect to Rayleigh's criterion for resolution as follows:

$$d = \frac{0.61\lambda}{NA} = \frac{0.61\lambda}{n \sin\alpha} \quad (1)$$

- d : Minimum resolvable distance
 λ : Wavelength of light source used
 NA : Numerical aperture of lens
 n : Refractive index of medium
 α : Half angle of acceptance for objective lens

A higher resolution or lower minimum resolvable distance (d) can thus be achieved by either employing lenses with higher NA or light with shorter wavelength (λ) than IR currently exploited for its optical transparency in silicon. To this end, the use of high NA SIL in a configuration as detailed in Figure 2.1 has reduced the minimum resolvable dimensions for laser imaging and probing [3] as well as collection efficiency for emission techniques [4].

Figure 2.1: Test setup schematic and close up view of laser focusing on device under test (DUT) with hemispherical SIL [4].

With hemispherical / central SILs (h-SIL), the truncated spherical lenses are landed onto the silicon backside so as to focus light onto a point at a focal length equivalent to the

combined thickness of the SIL and the silicon. Thus, the silicon on the backside of the device needs to be thinned to specification at tight RST tolerances or the laser may not be brought to focus. Consequently, the probing beam spot size will widen and a corresponding loss in resolution results. In such a configuration, the $\sin \alpha$ term approaches 1 as the focal point is positioned at the center of the hemisphere. With the region of interest positioned at the focal point, the resolution thus scales down directly with an increase in the refractive index of the lens used. Using a 1064 nm laser in conjunction with a silicon h-SIL, resolution enhancements down to 230 nm was thus achieved [5]. Notably, any imperfections such as scratches on the surface will then introduce airgaps at the interface that scatter the signal away with deleterious effects on imaging.

With modern SILs using high refractive index materials such as Si and GaAs to obtain high NA of 3.3, the resolution is then limited by the wavelength of light used. Using visible wavelength of 663 nm as compared to 1150 nm, an approximate 2 times enhancement in resolution is observed [6-7]. However, while infrared wavelengths of light were initially chosen for its transparency in silicon, shorter wavelengths of light face high absorption and thus signal attenuation as seen in Figure 2.2. With a penetration depth of 5 μm in the visible spectrum under 700 nm [8], VLP thus requires ultra-thin RSTs to reduce signal attenuation.

Figure 2.2: Absorption coefficient and penetration depth in silicon at varying wavelengths [8].

2.2 Device Warpage

Modern microprocessors are usually interfaced by controlled collapse chip connection (C4) or otherwise known as flip chip for the manner in which the device is connected onto the substrate. A typical flip chip microprocessor is depicted in Figure 2.3 a) along with its cross-sectional schematic in Figure 2.3 b). For such devices, the active circuitry composed of the transistors and interconnect metallization are formed on the bottom face of the bulk silicon and interfaced to the rest of the system through the substrate by solder balls also known as C4 bumps.

Figure 2.3: a) Typical flip chip cross-sectional schematic and b) plan view of an actual device.

In performing EFI, access is necessary for optical probing of the transistors sandwiched between the bulk silicon and substrate. While the multitude of metallization, C4 bumps,

underfill and substrate constitute the frontside, only bulk silicon is found on the backside. As discussed earlier, silicon is relatively transparent in IR wavelengths and less so in the visible spectrum. Thus, for optical probing to be performed through the backside, the silicon must be thinned down for high resolution SIL imaging.

While thinning silicon from the backside of the device would be a straightforward process if the device is flat as idealized in the schematic above, real devices are often warped through chip-package interactions. This is attributed to the different materials and thus coefficients of thermal expansion (CTE) between the constituents of the chip bound together. According to literature, bulk silicon has a CTE of approximately 2.8 ppm/°C, the composite organic substrate can have a CTE of ~16 ppm/°C [9] as seen in Table 2.1. After the silicon die is fabricated and bumped with solder balls, it is set over the substrate patterned with corresponding contacts and heated up to ~230°C for reflow and connection. During such a process, large thermal stresses result in the chip due to CTE mismatch, reaching a maximum near the die edges and presenting reliability risks with solder bump failure. Thus, the next step involves underfill encapsulation with epoxy based encapsulants dispensed between the solder balls and cured at ~160°C till solidification [10].

Table 2.1: Material properties of chip and package (substrate) [10].

	Modulus E (GPa)	CTE (ppm/°C)
Underfill	9	22
Substrate	15.5	15 (in-plane) 51 (out-of-plane)
Die	112.7	2.8

After cooling, the underfill reinforces the solder bumps by coupling the die to the substrate and channeling stresses across the entire die. Owing to these multiple thermal steps during manufacture and later during use, high thermal stress is thus encountered at the die-underfill interface deformation and manifested as warpage on both the package and die [11]. For an approximately 22 mm x 14 mm device, thermal stresses and difference in shrinkage due to CTE mismatch can result in warpages of about 60 μm as illustrated in Figure 2.4.

Figure 2.4: a) Physical profile map of actual sample and b) its schematic cross-section.

Notably, some material parameters influencing warpage include the constituent's CTE, elastic modulus as well as size and thickness. While most of these remain a constant through further processing, silicon thinning for SIL imaging reduces the die thickness on the substrate. It has also been shown that as the substrate to die thickness ratio increases, changes in bending strength resulted in the net reduction of warpage in the device [12]. When ultra-thinning, the substrate thickness remains at 1250 μm while the die thickness is reduced from an initial 780 μm to 5 μm , accounting for a change in die to substrate thickness ratio from 1.6 initially to 250 when thinned to 5 μm . we can then expect relaxation effects resulting in a net reduction of the die curvature as the silicon is thinned. With device warpage inducing surface non-planarity on the silicon, there is thus a need for contour milling techniques for uniform removal of silicon from the backside.

2.3 CNC Contour Milling

Traditionally, backside silicon thinning has been performed with polishers as shown in Figure 2.5 a) by parallel lapping for material removal in 2D fashion. This method of thinning requires the device to be pressed against a rotating platen lined with abrasive diamond films to remove material uniformly in a planar fashion. However, this cannot be used to prepare highly warped samples because targeting thicknesses under the warpage at any areas other than the center would destroy the sample due to the curvature of the die. Thus, CNC milling systems capable of 3D milling for uniform silicon removal, has come into prevalence for backside sample preparation [14-16].

Figure 2.5 a) Polisher for parallel lapping and b) CNC milling system [13].

At its fundamentals, these techniques remove material by milling while accounting for the warpage inherent to the devices. This can be achieved by either stage tilting to level the tool with the surface in adaptive grinding [14] or by profiling the surface to obtain a 3D contour map that is fed for 3D contour milling [15]. As illustrated in Figure 2.6, the remaining silicon profile from parallel lapping will be plano-concave with thicker edges imparted by device warpage. On the other hand, 3D techniques such as contour milling will result in whole die uniform RST profiles that is necessary for SIL imaging.

Figure 2.6 RST profiles by parallel lapping vs contour milling.

While thinning down, changes in relative thicknesses and thus bending strength of the die affects the equilibrium of the system locked under thermal stress by the cured underfill. Thus, as strain is relieved from the silicon when partially removed by milling, stress relaxation results in changes to the warpage profile to equalize with that at its new equilibrium. Empirically, the relaxation effects are negligible down to a critical thickness of about 100 μm RST. Under 100 μm , milling with profiles collected at full thickness will result in RST non-uniformity as the physical profile of the device is relaxed. Thus, when targeting 50 μm RST, contour milling processes would incorporate a profile correction step to account for device relaxation. Alternatively, adaptive milling techniques modulate milling rates by *in-situ* measurements of the RST at each point for uniform removal.

In practice, CNC milling systems such as the Allied X-Prep™ Precision Milling system involve four basic steps for uniform silicon removal and will be explored for the scope of this project. As shown in the flowchart in Figure 2.7, the sample is first mounted on a glass carrier with wax and physically profiled with a ruby tipped probe. A contour map listing the relative elevation at each point of the die is stored in a grid map of the surface for subsequent milling. Secondly, a coarse diamond grinding disc is used for quick 2D grinding of bulk silicon from full thicknesses of ~ 770 μm down to ~ 80 μm . As it is carried out only in 2D with positional force feedback, the results are similar to those obtained from parallel lapping with a resulting plano-concave RST profile. Thirdly, the initial collected contour map is used to mill the warped profile back to the die surface to achieve uniform RST. An interferometer-based RST measurement tool is also used to ascertain the RST profile along the way. This RST profile is then used to correct the map used for milling in a proprietary

X-Correct™ step, accounting for relaxation effects anticipated under 100 μm . Finally, polishing is performed on the surface using a cloth pad charged with progressively finer grits of diamond paste. Executed in floating force feedback, this removes little material off the die evenly as the surface is brought to a mirror finish for intimate SIL contact.

Figure 2.7 Process flow for global contour milling.

Since its introduction in 2013, the process has proven reliable in preparing samples to the requisite 50 μm RSTs at $\pm 5 \mu\text{m}$ tolerances for SIL imaging. However, since then, the advancement of SIL technology as mentioned earlier has necessitated thinning to lower RSTs for VLP. Furthermore, the demand for higher performances from advanced microprocessors have driven larger dies and more complex 3D integration that complicate the thinning process [17]. With devices featuring 3D integration such as stacked dies as shown in Figure 2.8 and larger die areas, greater device warpage and relaxation characteristics are thus anticipated. In adapting the current contour milling techniques, study of the process and larger devices while thinning is thus necessary to achieve ultra-thinned devices for VLP.

Figure 2.8 Cross section and plan view schematic of microprocessor 3D integration [17].

2.4 FIB Silicon Trenching for Higher Resolution Probing

To extend the resolution of future probing techniques, the wavelengths of the probes used will have to decrease along with thinner RSTs beyond 5 μm . Alternatively, as laser techniques are ultimately limited by the optical limits of resolution, efforts have also been dedicated to electron beam probing techniques with its potential for sub-100 nm lateral resolutions [18]. However, the limited penetration depth of the electron beam similarly necessitates the silicon to be thinned down to sub-micron or even to the shallow trench isolation (STI) level for interaction with the circuitry and optimal detection of the secondary electron signals [19]. As illustrated in Figure 2.9, the fin Field Effect Transistor (finFET) schematic shows that at the STI level, little silicon remains between the fin base and the active part of the fin wrapped by the transistor gate.

Figure 2.9 Cross-sectional schematic of finFET [20].

The STI thus serves as a useful end-pointing feature for thin silicon when thinning from the backside through bulk silicon. For these purposes, FIB techniques similar to those used for circuit edit are usually employed to locally thin silicon to the requisite thickness as seen in Figure 2.10 [21].

Figure 2.10 FIB images of trench at a) n-well (left) with smaller opening at b) ≈ 355 nm and c) ≈ 250 nm from STI [21].

To prepare a device for FIB circuit edit, the bulk silicon is first mechanically pre-thinned to under $100\ \mu\text{m}$ using methods described in the previous section. Subsequently, the sample is navigated to the area of interest (AOI) in the FIB using an IR camera to identify circuit features. A window is then defined over the AOI for ion beam milling with predefined dosages and milling conditions. During milling, ions are accelerated up to 30 kV at specified currents to sputter material off the surface over a period of time. In all, the total charge of ions delivered is then quantified by the product of beam current (nA) and time (s) to give the total charge and thus quantity of singly charged ions. For the purpose of studying milling rates, the dosage is also often expressed in area-normalized quantities by dividing the total charge of ions delivered over the milling area as $\text{nC}/\mu\text{m}^2$ as defined in Equation (2). The sputter rate in turn, is then usually given in $\mu\text{m}^3/\text{nC}$ as the division of the mill depth by the dosage delivered.

$$Dosage \left(\frac{nC}{\mu m^2} \right) = \frac{\text{beam current (nA)} * \text{time (s)}}{\text{length (um)} * \text{width(um)}} \quad (2)$$

In traditional dual-beam FIB systems, the ion column is paired with an electron column as in a scanning electron microscope (SEM). This then allows both ion beam milling and *in-situ* high resolution electron beam imaging. In particular, gallium ions are often used and provisioned from a liquid metal ion source, accounting for beam currents between 1 pA up to 100 nA. However, for backside probing, large windows of up 160000 μm^2 may be necessary to cover all suspect sections. From literature, a 30 keV/10 nA gallium beam sputters silicon from normal incidence with a sputter rate of approximately 2.7 $\mu m^3/s$, so trenching such a window to even 1 μm deep would take a prohibitively long 16 h. Consequently, the range of beam currents used to open smaller windows for circuit edit purposes may be inadequate for milling larger windows for backside probing within reasonable time frames. Further, the use of gallium liquid metal ion sources at high currents approaching 100 nA disproportionately increases the solid angle of acceptance into the column and thus the spherical aberration of the beam. As a result, larger beam spot sizes as well as lower image side brightness and thus beam performance can be expected at higher beam currents as seen in Figure 2.11 [22].

Figure 2.11 Spot size and brightness vs beam current (left) and spot size with mill rate comparison between Ga FIB and Xe PFIB (right) [22].

To address limitations in milling rates and deteriorating beam performance at high currents, plasma FIBs with inductive coupled plasma sources for high currents has thus come into prominence [23]. By using an RF antenna to generate plasma inductively from noble gases such as xenon, much higher ion current densities can be extracted over smaller beam spot sizes. On top of that, experimental results as well as stopping and range of ions in matter modelling have similarly shown between 30 - 50% increase in sputter rates of silicon by Xe^+ ($0.37 \mu\text{m}^3/\text{nC}$) over Ga^+ ($0.27 \mu\text{m}^3/\text{nC}$) at 30 kV beam energies [24]. By combining the higher sputter yield and current densities of the xenon ions, these plasma FIB systems are then able to achieve higher milling rates while maintaining workable spot sizes and resolution as shown in Figure 2.11. However aside from milling efficacy, another key consideration in FIB processing and indeed, the overall failure analysis process is the introduction of artifacts onto the device. This can come as changes in electrical performance [21] or unintended physical alterations to chip as discussed below.

Regardless of the type of FIBs used, damage is a concern due to the intrinsic nature of the ion beam's interaction with matter. When the ion beam impinges upon the surface, the ions lose kinetic energy as they interact with the sample atoms in a collision cascade before coming to rest [25]. As illustrated in Figure 2.12, this results in two effects on the crystalline sample. Firstly, atoms within the collision cascade volume are displaced from their positions and thus amorphised to an extent more pronounced in the beam path than laterally. Secondly, the primary ion is then implanted within the sample at a projected range that is dependent on the initial beam energy.

Figure 2.12 Schematic illustration of the collision cascade for a 30 kV Ga^+ ion impinging on a crystal, with associated volume of damage as well as projected range R_p and lateral range R_l [25].

With results summarized from literature in Table 2.2, it has been shown that when milling a trench, the damage induced by the ion beam at normal incidence on the bottom wall is generally double that introduced by a glancing angle (4.5°) beam on the side wall. This is explained by face-to-face collisions between the incident gallium ions and the silicon atoms resulting in almost complete momentum transfer. With a critical density of silicon atoms recoiled by the incident beam, amorphisation then occurs with a depth equivalent to its energy-dependent projection range and straggle. This depth is then calculated to be 58 nm via Monte Carlo simulation and shows good agreement with experimental results [26]. Experimentally, this is also manifested in the amorphisation damage formed as two layers upon 30 kV ion beam milling. The upper 36 nm layer is characterised by high concentration of gallium implanted and corroborates with the penetration depth of gallium. The 56 nm

deep lower layer without gallium is then the result of recoiled silicon atoms previously discussed. Studies involving lower energy milling similarly corroborates with lesser amorphisation.

Table 2.2: Damage layer thickness induced by FIB milling [26-29].

Ion	Damage layer thickness by Xe ⁺ and Ga ⁺ milling (nm)			
	Beam Energy (kV)			
	2	5	10	30
Ga				
Side-wall	3	7	13	22 - 28
Bottom-wall	-	-	23	56
Xe				
Side-wall	2	4-5	-	13-18
Bottom-wall	-	14	-	54

Based on the same model, calculations based on gallium ions incident at 4.5° to model sidewall damage also yields values close to that experimentally determined at 28 nm. On the other hand, due to differences in ion-solid interactions, milling with heavier ions is expected to yield lesser side wall damage than with lighter ions of the same energy [27]. This is reflected in the smaller simulated longitudinal range of 24 nm for 0° incident Xe⁺ versus 28 nm of Ga⁺ as illustrated in Figure 2.13 a stopping and range of ions in matter (SRIM) plot [28].

Figure 2.13 SRIM plots for 0° incidence angle ion bombardment in Si at 30 kV. [28].

Through experimental results presented in other studies, the lateral sidewall damage induced by heavier Xe ion beam milling is also shown to be lesser than that of Ga ion milling by up to 25% [29]. In the case of glancing angle milling, the use of lower beam energies similarly yields lesser amorphisation as illustrated in Figure 2.14. Here, high resolution transmission electron microscopy (HRTEM) images taken on samples milled with a 5 kV Xe beam shows a smaller 4.17 nm amorphisation layer than with a 30 kV Xe beam showing a 12.9 nm amorphisation layer. Taken together, this suggests that to reduce the degree of amorphisation in silicon after FIB milling, either the use of low energy ion beam milling throughout or at the end as a polish step would be necessary.

Figure 2.14 HRTEM images of sidewall amorphisation in Si by Ga^+ and Xe^+ beam milling at varying accelerating voltage [27]

As previously discussed, the amorphisation of the sample during FIB milling is caused by the mechanical bombardment of the crystal lattice by the incident ions. When the incident ion comes to rest within the sample, it is then implanted near the milling surface. With Ga beam systems, the implanted ions can form deleterious intermetallic compounds with the

sample. When applied to the silicon backside for ultra-thinning, the implanted metallic gallium ions can also adversely impact circuit operations and optical characteristics [30]. To maintain a pristine surface free of implanted ions for further backside probing, it is thus necessary to utilise non-mechanical methods of material removal through chemical etching. In dedicated circuit edit FIBs, a gas precursor is commonly introduced through a gas injection system (GIS) nozzle during milling to enhance etch rates and selectivity [31]. For silicon gas assisted etching (GAE), halogen gas precursors such as Xenon Difluoride (XeF_2) is often used to enhance etch rates to up to 100 times of that without [32,33]. According to the following overall reaction, the silicon reacts with fluorine atoms dissociated from XeF_2 by the gallium beam to form SiF_4 , a volatile product that desorbs and pumps out from the system.

Due to the dependence on the gas precursor for etching, it is apparent that the etching characteristics and rate enhancements are dependent on the local XeF_2 gas pressure on top of the other milling conditions. As such, silicon trenching operations typically involve moderate $200\text{ }\mu\text{m} \times 200\text{ }\mu\text{m}$ trenches milled under specialized coaxial nozzles dispensing XeF_2 for etching uniformity [34]. A point of consideration during GAE is thus the avoidance of gas depletion by balancing the sputter action against the etching rate during milling. This can be achieved through the etchant by either increasing the gas flow or lowering the nozzle towards the milling surface to increase local gas pressures. On the other hand, the sputter rate can be reduced by either using lower beam currents, shorter dwell times, larger pixel spacing or delay time between raster. This would then allow adequate time for gas precursors to be adsorbed completely onto the surface before being sputtered by the ion beam. Should the milling process be performed under gas depletion, selectivity and rate enhancements afforded by GAE will diminish along with other benefits. When performed under optimal conditions as seen in Figure 2.15, it can be observed that while the surrounding areas are heavily pitted due to the spontaneous reactivity of the gas with silicon, the XeF_2 milled surface is only slightly roughened.

Figure 2.15 Typical 200 μm x 200 μm trench etched into silicon with pitted surroundings [34].

Taken together, the use of a FIB for silicon trenching promises to be a controlled and precise method for localised ultra-thinning to enable shorter wavelength visible light probing or electron beam probing. By careful characterisation of the milling rates across different individual Ga and Xe FIBs, measured quantities of silicon can be trenched where necessary. Furthermore, to avoid introducing artifacts and altering circuit performance, it is then necessary to reduce the inevitable FIB induced sample damage during processing. This should then be achieved by a combination of usage of lower beam energies and gas etchants.

References

- [1] C. Boit, "Fundamentals of photon emission (pem) in silicon-electroluminescence for analysis of electronic circuit and device functionality", *Microelectronics Failure Analysis: Desk Reference*, pp. 356-368, 2004.
- [2] U. Kindereit, G. Woods, J. Tian, U. Kerst, R. Leihkauf, and C. Boit, "Quantitative investigation of laser beam modulation in electrically active devices as used in laser voltage probing", *IEEE Transactions on Device and Materials Reliability*, vol. 7, no. 1, pp. 19-30, 2007.
- [3] K. A. Serrels, E. Ramsay, P. A. Dalgarno, B. Gerardot, J. A. O'Connor, R. H. Hadfield, Richard J. Warburton, and D. T. Reid "Solid immersion lens applications for nanophotonic devices," *Journal of Nanophotonics*, vol. 2, no. 1, p. 021854, 2008.
- [4] S. H. Goh, A. C. T. Quah, C. J. R. Sheppard, C. M. Chua, L. S. Koh, and J. C. H. Phang, "Effect of Refractive Solid Immersion Lens parameters on the enhancement of laser induced fault localization techniques," in 15th International Symposium on the Physical and Failure Analysis of Integrated Circuits, Singapore, 2008, pp. 1-6.
- [5] S. B. Ippolito, B. B. Goldberg, and M. S. Unlu, "High spatial resolution subsurface microscopy." *Appl. Phys. Lett.*, vol. 78, pp. 4071-4073, 2001.
- [6] C. Boit, H. Lohrke, P. Scholz, A. Beyreuther, U. Kerst, and Y. Iwaki, "Contactless Visible Light Probing for Nanoscale ICs through 10 μm Bulk Silicon." in 35th Annual NANO Testing Symposium, Toyonaka, 2015, pp. 215–221.
- [7] H. Lohrke, P. Scholz, A. Beyreuther, U. Ganesh, E. Uhlmann, S. Kühne, M. Jagodzinski, , Y. Iwaki, C. Boit, R. Chivas, and S. Silverman, "Contactless Fault Isolation for FinFET Technologies with Visible Light and GaP SIL." in 42nd International Symposium for Testing and Failure Analysis, Fort Worth, 2016, pp. 19-26.
- [8] M. A. Green, and M. J. Keevers, "Optical properties of intrinsic silicon at 300 K", *Progress in Photovoltaics: Research and Applications*, vol. 3, no. 3, pp. 189 - 192, 1995.
- [9] G. Wang, P. S. Ho, and S. Groothuis, "Chip-packaging interaction: a critical concern for Cu/low k packaging." *Microelectronics Reliability*, vol. 45, pp.1079-1093, 2005.

- [10] X. Dai, and P. S. Ho, "Thermo-mechanical deformation of underfilled flip-chip packaging," in 21st International Electronics Manufacturing Technology Symposium, Austin, 1997, pp. 326-333.
- [11] H. Bongtae, "Thermal stresses in microelectronics subassemblies: Quantitative characterization using photomechanics methods." *Journal of Thermal Stresses*, vol. 26, no. 6, pp. 583-613, 2002.
- [12] V. S. Milena, "Thermally induced deformations in die-substrate assembly". *Theor. Appl. Mech.*, vol. 35, no. 1–3, pp. 305–322, 2008.
- [13] Allied High Tech Products Inc. "State-of-the-Art Equipment Design & Manufacturing", 2019. [Online]. Available: <https://www.alliedhightech.com/equipment> [Accessed: 2019, Jun. 10].
- [14] R. Chivas, S. Silverman, and M. DiBattista, "Adaptive Grinding and Polishing of Silicon Integrated Circuits to Ultra-Thin Remaining Thickness", in 41th International Symposium for Testing and Failure Analysis, Portland, 2015, pp. 465-470.
- [15] C. Richardson, G. Liechty, C. Smith, and M. Karow "Contoured Device Sample Preparation Technique for $\pm 5\mu\text{m}$ Remaining Silicon Thicknesses That Meets Solid Immersion Lens Requirements" *Microelectronics Reliability*, vol. 53, no. 9-11, pp. 1434-1438, 2013.
- [16] M. S. Wei, H. B. Chong, S. H. Lim, and C. Richardson, "Sample Preparation for High Numerical Aperture Solid Immersion Lens Laser Imaging" in 40th International Symposium for Testing and Failure Analysis, Houston, 2014, pp. 392-396.
- [17] Advanced Micro Devices Inc. "High-Bandwidth Memory (HBM) reinventing memory technology", 2015. [Online]. Available: <https://www.amd.com/Documents/High-Bandwidth-Memory-HBM.pdf>. [Accessed: 2019, Jun. 14].
- [18] R. Schlangen, R. Leihkauf, U. Kerst, C. Boit, K. Wilsher, T. Lundquist, and et al., "Backside E-Beam Probing on Nano Scale Devices", in IEEE International Test Conference, Santa Clara, 2007, pp. 23.2 1-9.
- [19] J. Vickers, S. Somani, B. Freeman, P. Carlesonm L. Tůma, M. Unčovský, and P. Hlavenka, "Failure analysis of FinFET circuitry at GHz speeds using voltage-contrast and stroboscopic techniques on a scanning electron microscope", in 45th International Symposium for Testing and Failure Analysis, Portland, 2019, pp. 197-203.

- [20] Intel Inc. "Intel's Revolutionary 22 nm Transistor Technology", 2011. [Online]. Available:<https://www.intel.com/content/www/us/en/silicon-innovations/revolutionary-22nm-transistor-technology-presentation.html> [Accessed: 2020, Jun. 16].
- [21] R. Schlangen, "Development and evaluation of novel chip modification and analysis techniques based on backside focused ion beam preparation." Technischen Universität Berlin, Ph. D. Dissertation, 2010.
- [22] N. S. Smith, D. E. Kinion, P. P. Tesch, and R. W. Boswell, "A High Brightness Plasma Source for Focused Ion Beam Applications" *Microsc. Microanal.*, vol. 13, no. S02, pp. 180-181, 2007.
- [23] N. S. Smith, W. P. Skoczylas, S. M. Kellogg, D. E. Kinion, P. P. Tesch, O. Sutherland, A. Aaneslannd, and R. W. Boswell, "High brightness inductively coupled plasma source for high current focused ion beam applications" *J. Vac. Sci. Technol. B*, vol. 24, no. 6, p. 2902, 2006.
- [24] T.L. Burnett, R. Kelley, B. Winiarski, L. Contreras, M. Daly, A. Gholinia, M. G. Burke, and P. J. Withers, "Large volume serial section tomography by Xe Plasma FIB dual beam microscopy" *Ultramicroscopy*, vol.161, pp. 119-129, 2016.
- [25] C. Volkert, and A. Minor, "Focused Ion Beam Microscopy and Micromachining" *MRS Bulletin*, vol. 32(5), pp. 389-399, 2007.
- [26] S. Rubanov, and P. R. Munroe, "FIB - induced damage in silicon." *Journal of Microscopy*, vol. 214, no. 3, pp. 213-221, 2004.
- [27] R. Kelley, K. Song, B. Van Leer, D. Wall, and L. Kwakman, "Xe FIB Milling and Measurement of Amorphous Silicon Damage." *Microsc. Microanal.*, vol. 19(S2), pp. 862-863, 2013.
- [28] L. Giannuzzi, and N. Smith, "TEM Specimen Preparation with Plasma FIB Xe Ions." *Microsc. Microanal.*, vol. 17(S2), pp. 646-647, 2011.
- [29] L. Kwakman, M. Straw, G. Coustillier, M. Sentisl, J. Beyersdorfer, J. Schischka, F. Naumann, and F. Altmann, "Sample Preparation Strategies for Fast and Effective Failure Analysis of 3D Devices" in 39th International Symposium for Testing and Failure Analysis, San Jose, 2013, pp. 17-26.
- [30] M. J. Theisen, and T. G. Brown, "Optical Properties of Gallium Implanted Silicon," in *Frontiers in Optics, OSA Technical Digest (online)*, 2012, p. FTu4A.3

- [31] C. Rue, S. Herschbein, and C. Scudato, "Backside Circuit Edit on Full-Thickness Silicon Devices." in 34th International Symposium for Testing and Failure Analysis, Portland, 2008, pp. 141-150.
- [32] D. Xia, J. Notte, L. Stern, and B. Goetze, "Enhancement of XeF₂-assisted gallium ion beam etching of silicon layer and endpoint detection from backside in circuit editing." *J. Vacuum Sci. Technology B*, vol. 33, no. 6, pp. 06F501, 2015.
- [33] N. Bassom, and T. Mai, "Modeling and Optimization XeF₂- Enhanced FIB Milling of Silicon," in 25th International Symposium for Testing and Failure Analysis, Santa Clara, 1999, pp. 255-261.
- [34] G. P. Salazar, R. J. Shul, S. N. Ball, M. J. Rye, and B. S. Phillips, "A Quantitative Method for Measuring Remaining Silicon Thickness During XeF₂ FIB Trenching for Backside Circuit Operations," in 43rd International Symposium for Testing and Failure Analysis, Pasadena, 2017, pp. 246-250.

Chapter 3

Contour Milling

In this chapter, the contour milling workflow was applied to different samples across the whole device for ultra-thinning. Experiments were designed to characterise the sample behavior during the contour milling workflow process. Limitations of the existing workflow for ultra-thinning were then examined and improvements made. Changes made to the process were then validated.

3.1 Ultra-thinning of Sample by Global Contour Milling

Experiments were conducted in two parts to study the contour milling process for ultra-thinning. First, a moderately sized device was prepared to ultrathin with a target thickness of 10 μm using a modified workflow as detailed in Figure 3.1. Subsequently, a larger device was then prepared with the workflow to test the robustness of the workflow. After which, its limitations were evaluated, and further modifications were put forth in the workflow.

3.1.1 Experimental Setup and Process

First, a modified workflow for contour milling as shown in Figure 3.1 was applied with a 10 μm target thickness to study the process limitations. A 15.7 mm x 12.9 mm (202 mm²) sample was first mounted onto a glass slide with PELCO[®] Quickstick 135 wax heated to 120°C on a hotplate. Subsequently, the glass slide was set onto the sample jig, secured by screwing in the restraints and set on the stage in the X-Prep milling system. The sample on the jig was then levelled across the die using a ruby tipped probe, with the stage tilted iteratively and relative elevations probed across three die corners.

Figure 3.1 Modified workflow for global ultra-thinning.

A physical contour map was then collected across the die with the probe in a grid map defined by the die dimensions and milling tool diameter. The map was then exported into an excel file to visualize the warpage profile of the device for reference. To facilitate rapid removal of bulk silicon from its full thickness of 780 μm down to 50 μm, a diamond grinding disc tool was used for 2D grinding of the surface in a “flat-top” step. An RST map was then profiled in a grid map using the measurement tool, with points determined based on the intended milling tool size and device dimensions.

As the silicon profile becomes plano-concave with the RST non-uniform after 2D grinding, a convex profile equivalent to the inverse of the RST map needs to be reintroduced to the physical flat surface to achieve uniform RST across die. By assuming a flat surface after 2D grinding, a flat profile was generated similar to that physically profiled but with all points at the same elevation. Using the X-Correct macro for profile change, this flat physical profile was then correlated with the measured RST map to generate a contour map that was uploaded for 3D milling. This way, mild device relaxation would be accounted for as the system was instructed to mill according to the current RST profile. 3D milling was then performed incrementally with RST measurements taken until 25 μm. The rough milled die surface was then polished to a mirror finish for SIL contact and imaging. This was performed using diamond pastes with polishing cloth mounted tools running across the

surface in floating force mode. Running with progressively finer grits of diamond and collecting RST maps after every step, this was repeated until the target 10 μm thickness. The same was also repeated with a larger 22.3 mm x 14.3 mm (319 mm^2) sample to test the robustness of the workflow. Process limitations were then discussed and further tweaked.

3.1.2 Results and Discussion

After physically profiling the first 15.7 mm x 12.9 mm (202 mm^2) sample, the contour map was downloaded and visualized on an excel spreadsheet as illustrated in Figure 3.2 a). On the grid map, each point represents a single z-reading taken by the probe with reference to the center which was set as 0 μm . A convex profile with a moderate warpage of 37.5 μm was then observed between the highest and lowest point on the device. Because the highest part is at the center, it is coded blue and transitions to red at lower sections around the edges to show a convex profile as in the visualization. With IR-based fault isolation techniques requiring 50 μm RST on thinned samples, 2D milling alone would thus suffice as thinning the extremes to the requisite thickness would still leave 12.5 μm RST at the center. However, when working with larger samples with larger warpage, this would not be possible without overgrinding the sample. Also, when thinning to thicknesses under the warpage, contour milling would be necessary to avoid overgrinding and to achieve better RST uniformity.

Figure 3.2 a) Initial physical contour map with visualization and b) RST map with visualization after 2D milling.

Subsequently, a 2D ‘flattop’ milling step was performed until the device was at 55 μm at the thinnest and verified with the measurement tool. From the collected RST map in Figure 3.2 b), an RST range of 36.6 μm is observed between the thickest 90.0 μm by the corner and the thinnest 53.5 μm at the center. As the silicon is of uniform thickness originally and bound to the substrate in a warped profile, 2D milling introduced non-uniformity equivalent to the initial warpage after planarizing. Between the initial 37.5 μm warpage and the post flattop milling 36.6 μm RST variation, it can also be noted that there was little stress relaxation involved as the profile remained unchanged. Thus, by assuming a flat physical surface, a contour map similar to that initially profiled can be generated from the RST map collected.

Milling with the generated profile, a uniform RST profile was then obtained to within a range of 7.8 μm between the thickest and thinnest part of the die as shown in Figure 3.3 a). The die thickness statistics through the subsequent processes was then recorded in Table 3.1. After polishing, the RST map in Figure 3.3 b) shows a global average of 10.3 μm and a smaller range of 4.5 μm . Looking at the differences in profile pre and post polishing, it is noted that the device became thicker around the edges in comparison to the center than before and became more uniform. This is explained by the lower overlap of the polishing

tool around the edge and thus slower polishing rate. The improved uniformity is the result of the polishing tool smoothening the milled profile with an average 13.8 μm thickness of silicon polished off. However, as the post polishing RST map shows, silicon was polished more slowly about the edges. This would then aggravate RST non-uniformity should polishing be extended. To achieve greater uniformity, it is thus necessary to select milling parameters to minimise milling induced damage, polishing required and thus non-uniformity.

Figure 3.3 a) RST map with visualization after 3D milling and b) final RST map with visualization after polishing.

In all, this modified workflow resulted in a global average of 10.3 μm with a range of 4.5 μm , having a maximum RST at 13.1 μm and a minimum at 8.6 μm . With a low range of 4.5 μm and thus high uniformity, this implies that the sample can be safely thinned further to 4.5 μm without over polishing any part of the device. With this device, even in the worst-case scenario where the area of interest (AOI) is located at the thickest part of 13.1 μm , to further thin it down to 10.0 μm will still leave the thinnest part at a safe 5.5 μm , assuming uniform polish rates. Thus, it can be noted that for global contour milling, a good uniformity and RST range to achieve would be under the target thickness. It is then apparent that the RST range requirements are stricter with lower target thickness to avoid over polishing the device.

Table 3.1: Table of RST values through sequential polishing steps from left to right.

	After 3D	15µm polish	9µm polish	6µm polish	1µm polish	0.5µm polish	CMP	Overall polish
Global RST (µm)								
Average	24.1	21.0	16.8	13.1	11.1	10.5	10.3	13.8
Max	27.2	23.3	19.3	15.8	14.1	13.6	13.1	14.0
Min	19.3	16.1	12.1	10.3	9.1	8.6	8.6	10.7
Range	7.8	7.2	7.2	5.5	5.1	4.9	4.5	3.4
RST removed	-	3.0	4.2	3.7	2.0	0.7	0.2	13.8

A larger 22.3 mm x 14.3 mm (319 mm²) sample was then taken through the same workflow. On this sample, a larger warpage of 61.3 µm was observed from the collected physical profile as shown in Figure 3.4 a) as compared to 37.5 µm with the smaller device. Subsequent 2D milling then formed a similar plano-concave RST profile in Figure 3.4 b). Between the thickest 108 µm RST found at the edge and the thinnest 51.6 µm at the center, a 56.3 µm of silicon variation is measured. With the initial RST profile and the milled physical profile assumed to be flat, this variation can be explained by and approximated to the device warpage. Between the device at full thickness and at 51.6 µm, there has thus been a reduction in warpage of 5 µm with the thinned silicon now less convex. Relative to the smaller device, there is thus a larger degree of profile change when a larger device is thinned.

*all values in x, y, z (depth) given in µm

Figure 3.4 a) Initial physical contour map with visualization and b) RST map with visualization after 2D milling.

With the thinnest now at 51.6 μm , a profile was then generated and uploaded for milling to reintroduce the contour. With a target 56.3 μm of warpage, the current thickness variation can then be smoothened out for uniform RST. However, after milling, subsequent measurements found an odd convex RST profile. As shown in Figure 3.5 a), the RST was found to be thinnest at 18.0 μm by the edge, 30.8 μm near the center with an average of 27.4 μm . Two anomalies are of note in this contour reintroduction step intended for minimal material removal at the center reference. The first lies in the sample overmilling with reference to the center. While the extent of overmilling due to machine variation is typically under 5 μm , 21.6 μm of excess silicon was removed in this step. The second relates to the non-uniform RST profile obtained after milling with a range of 12.8 μm between 30.8 μm and 18.0 μm found on the resulting device. Compared to the 7.8 μm obtained with the smaller device, milling with the generated profile should already account for the 5 μm change in warpage, and thus should yield better silicon uniformity. Further milling to thin the center down to 25 μm for polishing then similarly resulted in overmilling with the center at 13.3 μm instead as shown in Figure 3.5 b).

Figure 3.5 a) RST map with visualization after 3D milling and b) final RST map with visualization after overmilling.

Optically, it was observed that the circuitry underneath has been exposed at the corners as in Figure 3.6 a). It can then be gathered that there is no silicon remaining at the corners and that erroneous RST values reported on the RST map were instead from the estimated range fed in for measurement. Regardless, both RST maps shows that the peripheries were abnormally thin compared to the center. This large silicon non-uniformity was then manifested in the final over-milled device. As observed in the optical image shown in Figure 3.6 b) and c), the edges' circuitry was exposed first as all the silicon was milled off from the thinnest edges. With the circuitry exposed, the electrical integrity was thus compromised and the whole device was destroyed with no further testing possible.

Figure 3.6 Optical images of a) overall die, b) over-milled top left corner, and c) over-milled bottom right corner.

While the above can be attributed to process variation during milling or measurement, many larger samples prepared the same way with similar observations suggests otherwise. The same trend thus points towards aggravated sample relaxation behavior causing profile change during milling. The profile changes can then possibly explain the larger silicon non-uniformity after milling and over-milling. Regardless, this current workflow was thus found to be inadequate for ultra-thinning of larger samples reliably. To this end, the next section will then characterise the sample warpage behavior that hampers ultra-thinning.

3.2 Characterisation of Sample Warpage Behavior During Thinning

The convex warpage profile of the device on the package accounts for a difference in elevation for otherwise equal thicknesses of silicon on the die. The initial levelling performed on the mounted sample then ensures that the highest part of the convex die lies at the center while the four edges lie at roughly the same height. Subsequently in the 2D “flat-top” milling step, bulk silicon is removed in a 2D fashion by positional feedback without regard for the warpage. Assuming ideal or near ideal tool efficiency in material removal using fresh grinding discs or metal bonded diamond tools, the resulting milling surface can be assumed to be flat with a plano-concave remaining silicon profile.

When fully flattened after removing material at least equivalent to the warpage of the die, the silicon would then be thicker at the edges than at the center. This RST difference between the thickest edge and thinnest center as measured from the collected thickness map is then the result of die curvature and can be taken as the warpage. Thus, we can study the warpage behavior by the variation in RST range as we thin down the sample by 2D ‘flat-top’ milling.

3.2.1 Experimental Setup and Process

To study sample warpage across different die sizes behavior during thinning, two devices were used, with a small one at 15.7 mm x 12.9 mm (202 mm²) as before and a larger one at 22.3 mm x 14.3 mm (319 mm²). Each device was prepared by wax mounting, followed

by levelling to optimize planarity before 2D milling to obtain a flat-top profile. The ~ 770 μm thick samples were then progressively thinned down by 2D milling to 50 μm with thickness maps collected in a grid map along the way. The variation of warpage measured in this way was used to study stress relaxation resulting from profile change during milling.

3.2.2 Results and Discussion

A 15.7 mm x 12.9 mm (202 mm²) device was first prepared by wax mounting and levelling. The die perimeter was then defined and a tool diameter of 1.5 mm was input for the generation of a square grid map with 21 data points in both axes. This grid map then guides physical profiling, RST map collection and subsequent contour milling. During physical profiling, z measurements were referenced to the center of the convex die with the depth of the other regions reflected in the physical map. In this way, the warpage was measured at 38 μm as the difference in depth between the highest and lowest parts of the die in Figure 3.7 a). Using the measurement tool, an RST map collected across the die also reflects an average full silicon thickness of 772 μm in Figure 3.7 b).

*all values in x, y, z (depth) given in μm

Figure 3.7 a) Initial physical profile map and b) RST map of die at full thickness.

Upon 2D milling to 700 μm at its thinnest, a concave RST profile was observed as shown in Figure 3.8 a) and well within expectations due to 2D grinding of the initial convex profile.

Notably, discontinuous thickness readings were speckled throughout the map. This is attributed to a rough, pitted surface finish obtained after coarse grinding and inaccuracies from measurements collected off a rough surface at high thicknesses. The range of RST values was measured at 43.6 μm and correlates well to the initial warpage of 38 μm after accounting for the variations mentioned.

With further 2D milling, a similar concave RST profile was observed down to 40 μm at its thinnest as seen in Figure 3.8 b). However, at 40 μm , a smaller RST range was observed at 29 μm with a smoother RST profile measured across the die. The smoothing of the profile is due to the improved accuracy of the Fast Fourier Transform estimates used during measurement at smaller thicknesses and range.

Figure 3.8 RST maps collected at a) 700 μm , b) 40 μm , and c) difference between the two.

Noting the longer time taken for the iterative milling and RST measurements, the change in warpage can be attributed to device relaxation as the silicon was thinned and its bending strength thus reduced. Comparing the above two RST maps collected at 700 μm and 40 μm in Figure 3.8 c), it is also observed that the amount of material removed ranged from 669 μm near the edge to 655 μm near the middle. This is despite material removal only being done through 2D milling performed at positional mode feedback where there should be uniform material removal. Noting the roughly concave profile of the comparison map with more material milled at the edge than center, this suggests that a change in warpage profile occurred during milling.

Furthermore, when thinned and mapped at various RSTs, a trend as plotted in Figure 3.9 can be observed in its variation and thus warpage. While the warpage was maintained at approximately 45 μm above 100 μm RST, a sharp decline to about 35 μm was observed thereafter under 100 μm . This suggests that while the warpage profile was maintained above 100 μm , thinning further quickly relaxed the die and resulted in a less convex profile and thus smaller warpage. Without accounting for this stress relaxation and thus profile change of 10 μm , further milling will then be unable to achieve a uniform RST profile.

Figure 3.9 Warpage as measured by RST range vs RST for 202 mm^2 device.

On the larger 319 mm² device, a larger warpage of 69 μm was measured by physical profiling. This is well within expectations for a larger device with higher thermal stresses. Upon flattop milling and RST mapping, an approximate 72 μm variation was measured within the die with 500 μm at its thinnest, correlating well with the initial profiled warpage. As with the smaller sample, a similar decline in warpage was observed when thinned as shown in Figure 3.10. Here however, the warpage gradually decreased from 78 μm at 400 μm to 70 μm at 200 μm followed by a dip to 45 μm under 100 μm . The greater warpage and extent of profile change over the smaller die thus indicate the greater stress and relaxation encountered by larger dies. As reported by other studies, simulations have shown that warpage is a function of not just the package and die size but also the thickness of the individual components [12]. In particular, increasing the substrate-to-die thickness ratio by decreasing the die thickness through thinning can reduce the warpage as observed here experimentally.

Figure 3.10 Warpage as measured by RST range vs RST for 319 mm² device.

The change in bending strength of silicon can thus be significant under about 150 μm to induce profile change. For thinning under 150 μm and indeed ultra-thinning, there must be caution paid to the effects of stress relaxation. This can include additional X-Correct[™] profile correction steps close to the target to account for relaxation or the incorporation of stress relaxation steps close to target thickness before final milling.

3.3 Ultra-thinning With Incorporation of Stress Relaxation Step

While the original workflow's profile correction (X-Correct TM) step accounted for mild device relaxation when thinning to 50 μm , it was inadequate when larger stress relaxation effects were involved. This was encountered during ultra-thinning due to changes in bending strength of the silicon being more significant with thinner silicon and larger devices. While it has been shown in the previous section that thinning results in dynamic profile changes, its effects are negligible as 2D milling is only intended for rapid material removal.

3D contour milling on the other hand, requires that an accurate map be uploaded and that the profile does not change to achieve uniform RSTs after milling. However, further profile changes due to stress relaxation during 3D milling complicates the process. To mitigate this in larger devices, a stress relaxation step was thus incorporated after the 2D step for the device warpage to equilibrate before 3D milling and profile correction (Figure 3.11).

3.3.1 Experimental Setup and Process

First, there is a need to study the effect and extent of static stress relaxation on samples of different sizes thinned to different thicknesses. This was first achieved by 2D milling a 22.3 mm x 14.3 mm (319 mm²) sample to 300 μm and 40 μm RST before physically profiling the flattop surface to obtain an initial frame of reference. The sample was then left overnight to relax under ambient conditions before it was physically profiled again to check for changes. To account for profile changes while thinning, a stress relaxation step thus needs to be incorporated in the workflow prior to 3D milling.

Figure 3.11 Modified workflow for ultra-thinning while incorporating a stress relaxation step.

To target a 5 μm target thickness with larger devices, a stress relaxation step then becomes necessary in a modified workflow as illustrated in Figure 3.11. A large 319 mm² sample was first wax mounted onto a glass slide and secured onto the sample jig. The sample was then levelled before bulk material was removed by 2D milling till the sample was about 40 μm at its thinnest. An RST map was then collected and the sample left overnight for the sample to equilibrate to its new profile after stress relaxation. Subsequently, the sample was profiled again and corrected with the collected RST map to generate a contour map accounting for stress relaxation. The generated map was then used for 3D milling till 20 μm at its thinnest before polishing. Finally, the sample was polished in 2D floating force mode using polishing pads with progressively finer grits of diamond paste until the target thickness of 5 μm. A whole die RST map was then taken with the visible light spectrometer on the measurement tool to verify the final thickness across the sample.

There was then a need to validate this workflow for ultra-thinning of yet larger samples while accounting for greater extent of warpage and stress relaxation. To this end, the process was also repeated with the largest device available for experimentation at 507 mm².

3.3.2 Results and Discussion

On the first 22.3 mm x 14.3 mm (319 mm²) sample, an initial warpage of 66 μm was

measured from physical profiling. Following 2D flat-top milling to 300 μm RST, a similar warpage value of 68 μm was found from the difference between the thickest and thinnest sections in a concave RST profile collected as Figure 3.12 a). By physically profiling this flattop sample, a flat surface with a slight concavity within 6 μm of variation in elevation was also observed in the physical map shown in Figure 3.12 b). When left overnight, the physical profile remained within 6 μm of variation in elevation with a slight change in shape as seen in Figure 3.12 c). When the difference between the physical maps was taken in Figure 3.12 d), only a slight concave of 4 μm was observed over the time span. Thus, it is deduced that stress relaxation effects at 300 μm RST are negligible due to the bulk of remaining silicon resisting bending into its equilibrium profile with lower warpage.

Figure 3.12 a) RST map at 300 μm after 2D milling, b) physical map after 2D milling, c) physical map after overnight stress relaxation, and d) difference map showing overnight change.

A second sample was then prepared the same way but to 40 μm at the center rather than 300 μm RST after 2D milling. As with the first sample, a similar warpage of 70 μm was physically profiled before it was directly milled down to target thickness and an RST map was collected. From the RST map collected, a similar warpage of 67 μm was observed in Figure 3.13 a). As the warpage approximates that initially collected, it is possibly explained by the lack of stress relaxation effects within the short span of time between the sample being thinned and mapped. A physical profile of the flattop surface immediately after then showed a flat surface to within 3 μm of height variation in Figure 3.13 b). After leaving it overnight for stress relaxation and profiling again, a concaved physical profile is now

observed in Figure 3.13 c). Without any physical intervention, it is observed that the edges have risen an approximate 21 μm relative to the center when compared to the map collected before. Shown in Figure 3.13 d), the 21 μm change in profile was accounted for by static stress relaxation as the sample equilibrated to its new warpage profile with its reduced bending strength. When compared to the change in warpage as the sample was progressively thinned, it is found to be similar at 28 μm after accounting for the higher initial warpage of 73 μm as opposed to 66 μm in this sample. From these results and those from progressively thinning, it is believed that stress relaxation is the result of a change in bending strength of the silicon as it is thinned and is a time dependent process.

Figure 3.13 a) RST map at 40 μm after 2D milling, b) physical map after 2D milling, c) physical map after overnight stress relaxation, and d) difference map showing overnight change.

Taking a sample through the above workflow and accounting for stress relaxation, the sample was then taken to ultra-thin with an overall average thickness of 3.1 μm and range of 8.8 μm . In actual failure analysis cases, low-resolution fault isolation techniques are usually capable of narrowing down suspects to within an area of 1 mm^2 with a slightly larger area necessary for navigation. Taking a 4.3 mm x 2.7 mm area at the center as boxed up in Figure 3.14 a), an average of 1.2 μm RST with a range of 0.8 μm was achieved, meeting the requirements for SILs for use with VLP. When optically imaged under bright illumination, Figure 3.14 b) and d) shows the underlying circuitry becoming visible due to the higher optical transmission afforded by ultra-thinned silicon. As observed from the whole die image in Figure 3.14 c), this was achieved while preserving the other surface

components including the package stiffener and chip capacitors. When testing for marginal fails on these devices, the ability to thin down the sample without damaging the chip capacitors would be invaluable in preserving electrical characteristics.

Figure 3.14 a) Final RST map, b) whole die optical image, c) whole device image, and d) zoomed in optical image.

Lastly, the largest device available at 26.3 mm x 19.3 mm (507 mm²) was also taken through the process for validation. On this device, a large warpage of 88 µm was measured after 2D milling to 43 µm at the center as shown in Figure 3.15 a). After contour milling and polishing to ultra-thin, an overall average thickness of 3.4 µm and range of 7.8 µm was then obtained. Furthermore, within a 4.3 mm x 2.7 mm boxed up area at the center, a local average of 0.8 µm thickness and range of 0.35 µm was achieved, meeting the specifications necessary for VLP. Also, of note is the correspondence of the final RST map and that of the

optical image in Figure 3.15b). Because the RST profile is in the shape of a fish with the thinnest regions in the middle, the image is similarly brighter at those parts. With thinner remaining silicon, the circuitry becomes clearer due to higher optical transmission and thus reflectance of light against the circuitry there. By integrating a stress relaxation steps into the workflow, it is then possible to achieve uniform ultra-thin RST profiles by contour milling even with large device sizes.

Figure 3.15 a) Final RST map, b) Whole die optical image.

3.4 Thermal Stability of Ultra-thinned Devices During Dynamic Testing

During EFI, the devices are powered up and running as test patterns are input to stimulate defectivity and emissions for optical analysis with the SILs. As with many high-performance microprocessors, the energy consumed and heat generated during operation is high. However, due to the need for optical access by the SILs, heatsinks typically mounted to contact the die for heat dissipation can no longer be used and spray cooling is used instead. Without adequate cooling or silicon for heat dissipation, devices under test may then overheat, burn, and be destroyed. Thus, thermal stability of the samples becomes a great concern for the implementation of VLP especially on ultra-thinned units. To examine the thermal stability of ultra-thinned units, a sample was prepared, electrically tested, and evaluated.

3.4.1 Experimental Setup and Process

To study the thermal stability of ultra-thinned units with respect to its RST, an ultra-thinned sample was prepared for testing with localised polishing. A 22.3 mm x 14.3 mm (319 mm²) sample was used and taken through the modified workflow for ultra-thinning. Bulk silicon as first thinned by 2D milling to 60 µm and left to relax overnight. It was then physically profiled and its RST map collected for X-Correction to generate a contour map for milling. It was then 3D milled to 25 µm for polishing by iterative milling and RST mapping. Polishing was performed selectively by marking out progressively smaller regions with successive polishing steps and finer grits of diamond pastes. In the final passes, only the bottom right corner was polished until its target thickness of 5 µm with 1 µm grit diamond paste. The device was then put through dynamic electric testing till failure and inspected.

3.4.2 Results and Discussion

Through 3D milling and selective polishing with progressively smaller areas, an RST profile as seen in the Figure 3.16 was observed. In the obtained profile, a higher degree of RST non-uniformity is observed with thicker silicon near the three corners and edges while

an approximately 5 μm region was obtained on the corner polished selectively. Furthermore, a graded profile ranging from about 5 μm on the bottom right edge to 25 μm on the bottom left edge can be observed.

Figure 3.16 a) Final RST map and visualization, b) angled optical image of burnt edge, c) magnified angle optical image, and d) magnified plan view optical image.

The device was then dismantled from the glass slide and remnant wax was washed off with acetone. Next, the device was put through a static open-short test and passed, showing that the electrical integrity was intact after the thinning process. Subsequently, dynamic functional tests were run on the device till failure with cooling solutions typically employed. During testing, an array of nozzles was used to spray coolant onto the device for cooling. However, the test was stopped almost as soon as it started as the current drawn from the device exceeded safety limits. Upon optical inspection, it was observed that the device had cracked at the silicon backside at the thinnest part near the corner. This is indicative that the device was burnt, resulting in metal melt within the circuitry that caused gross electrical

shorting so no further tests are possible. Furthermore, as seen in Figure 3.16 c), imaging with the camera tilted shows that the silicon within the vicinity had buckled up with the C4 bump pattern visible on the silicon backside. This is characteristic of metal melt as the heat generated causes the copper metallization to melt, pool together and expand to form an uneven surface as depicted in Figure 3.17. Meanwhile, the heat resulted in the now film-like silicon layer to conform onto the underlying C4 bump topology in the vicinity.

Figure 3.17 Cross-sectional SEM image of typical metal melts in devices failing to shorts.

As the burnt area was observed only at the thinnest section of the device, it can be gathered that the cause of the failure is the lack of silicon for heat dissipation. During dynamic testing, the heat generated by the active circuitry is usually dissipated into the bulk silicon and then to the coolant, albeit at a limited rate. Thus, the bulk silicon dissipates heat from the circuitry during operation and ensures the device does not overheat and burn. However, as the device was ultra-thinned, a lower heat capacity resulted in higher operating temperatures that easily lead to metal melting and device destruction. Compounding the problem of thinned silicon is the fact that analog input-output (IO) circuitries that draw higher current usually line the periphery of the device. The net heat generated on the device during testing is then concentrated along the edges with additional thermal runaway risk. Thus, the thermal stability of globally ultra-thinned device becomes a concern that needs to be addressed by not just external cooling but also ultra-thinned silicon itself as the limiting factor to cooling.

3.5 Conclusion

In this chapter, the global contour milling technique was studied through its application to ultra-thin two devices with different sizes. With the larger of the two destroyed during the process, subsequent experiments were conducted to characterise the physical behaviour of the sample through thinning. Noting significant warpage changes on large samples through milling, sample relaxation was then identified as a key challenge. Taking relaxation effects of large samples into account, appropriate modifications were made to the workflow and the technique was successfully adapted for global ultra-thinning. However, subsequent electrical testing with the prepared samples highlighted challenges associated with the poor thermal stability of globally ultra-thinned samples during testing. The following chapter will then explore targeted milling techniques to localise ultra-thin silicon to the area of interest for improved thermal stability.

Chapter 4

Targeted Milling

In the previous chapter, the thermal challenges associated with testing ultra-thinned devices were identified, thus further changes to the workflow becomes necessary. The deterioration in thermal stability of ultra-thin devices is rooted in the lack of silicon for heat dissipation on the chip during testing. Thus, there is a need to maintain thicker silicon on the device while obtaining the requisite thickness at the area of interest for SIL imaging.

To do so, silicon thinning can simply be localised by performing the entire workflow over a smaller defined region on the device. However, doing so complicates device cooling as trenching the AOI is detrimental for the flow of coolants to the surrounding silicon for heat dissipation. Thus, targeted ultra-thinning by 2D milling and contour milling becomes necessary to localize ultra-thinning while maintaining a smooth topology over the AOI for cooling. The two methods of targeted ultra-thinning will be examined therein.

*This section is published substantially as “Targeted Silicon Ultra-thinning by Contour Milling for Advanced Fault isolation” in Proceedings of the 45th International Symposium for Testing and Failure Analysis, 2019. Reprinted with permission of ASM International. All rights reserved. www.asminternational.org.

4.1 Targeted 2D Milling

In a typical workflow, the sample is always globally levelled across the entire die surface to ensure that the highest point of the convex profile is in the center as illustrated in Figure 4.1 a). Subsequent 2D milling then results in thinner silicon at the center than the edge due to the inherent sample warpage. However, this can be exploited by sample levelling to target a specified AOI to achieve targeted ultra-thinning using only 2D milling and polishing steps. Shown in Figure 4.1 b), the sample can be tilted so the surface is normal to the milling tool at the AOI. The AOI is then the highest point of the sample and subsequent 2D thinning ensures it is maintained at the thinnest.

Figure 4.1: Targeted 2D milling after a) global and b) local levelling.

While the same can be performed with lapping techniques that are much quicker with less sophisticated tools, milling allows thinning to be performed only in a predefined region. This means that the surrounding surface components such as stiffeners and chip capacitors can be left intact as only the die parameter is demarcated for milling. Meanwhile, parallel lapping thins surface components at the same level as the thinned silicon, not to mention the imprecision involved.

4.1.1 Experimental Setup and Process

To test the process of targeted 2D milling, a 22.3 mm x 14.3 mm (319 mm²) sample was ultra-thinned with a specified AOI. First, a 1.9 mm x 2.4 mm window including several points on the grid map was specified as the AOI near the center-right section of the device. The sample was then mounted and levelled globally by specifying the entire die area. Subsequently, a physical profile map was taken in a grid map for an initial reference. The sample was then levelled locally about the AOI by specifying only the periphery of the specified window for levelling. Following, a physical profile map was then collected to ascertain the current physical profile. The device was then milled down to 10 µm at the thinnest before it was polished with progressively finer diamond pastes to a final 1 µm target.

4.1.2 Results and Discussion

Taking a sample through the normal workflow by levelling globally, a convex profile is observed through the physical contour map collected. From Figure 4.2 a), a difference of 66 µm in warpage was measured throughout the whole device while an average of 11 µm with a range of 9 µm was measured within the specified AOI window. The implications of this is that if only 2D milling was performed, the subsequent RST non-uniformity would be about 9 µm and that the center would be over milled when the AOI was at about 11 µm. This is a consequence of the plano-concave RST profile obtained after 2D flat-top milling that necessitates contour milling for low target thicknesses.

Subsequent levelling about the AOI to under 1 µm tolerance then produces an offset convex profile centered about the specified AOI as seen in Figure 4.2 b). Within the AOI, a minimum of 0 µm with a range of 0.6 µm now indicates that the AOI is now the highest section of the device and that 2D thinning can achieve greater RST uniformity within the window. Because the device is now tilted about the AOI specified on the right section of the device, a greater slant is now observed on the left section with a relative depth of 119 µm.

Figure 4.2: a) Physical contour map of globally levelled sample and b) sample locally levelled about AOI boxed up in map.

After 2D milling down to 11 μm at its thinnest, an RST map was collected off the device in Figure 4.3 a). A similar offset concave RST profile was then observed centering about the AOI. The lowest thickness of 10.6 μm was then measured with a range of 1.5 μm within the window. Approximating the initial differences in elevation within the window before thinning, 2D milling effectively transferred the difference in elevation down to that in RST non-uniformity with the thickest found at 119 μm by the edge. Should the differences be transferred as is, the edge should be expected at about $119 \mu\text{m} + 11 \mu\text{m} = 130 \mu\text{m}$. However, the actual thickness removed from the edge was higher than that at the AOI even with a 2D process. This is likely caused by sample relaxation when ultra-thinned, thus a less convex profile results in the edge being thinned more.

Upon complete polishing, a final RST map was then taken with a visible light spectrometer on the measurement tool in Figure 4.3 b). A global average of 27.6 μm with a range of 98.1 μm between 99.1 μm and 1.0 μm as well as a local average of 2.2 μm with a range of 2.2 μm between 3.4 μm and 1.2 μm was measured. While the local RST variation barely meets SIL specs, imaging within a smaller window should be adequate while maintaining surrounding thick silicon for heat dissipation. The RST uniformity measured within the

window while higher than the initial difference, it is much lower than that obtained by levelling globally.

Figure 4.3: RST maps collected after a) contour milling and b) polishing.

Imaged optically under illumination, the circuitry at the thinnest also becomes visible as seen in Figure 4.4. Notably though, the thinnest section of the device visibly shifted slightly right of the specified AOI with a measured $1.0\ \mu\text{m}$ at the thinnest. The shifting of the thinnest section out of the specified AOI can be attributed to non-uniform polishing as well as sample relaxation and could be a cause for concern during routine preparation.

Figure 4.4: Optical image of thinned device with thinnest section visible and RST labelled in μm .

When imaged with a 1340 nm laser using a laser scanning microscope, Newton rings are observed as concentric rings with alternating bright and dark fringes radiating from the AOI. Formed by interference between signals reflected off the top surface and the bottom surface of silicon, the contrast is thus subjected to local conditions relating to the path difference. The approximately 1.2 mm central disc with the same contrast can then be assumed to be under a quarter wavelength of RST uniformity. At higher magnification, the underlying C4 bump topology can also be observed showing local variation in RST non-uniformity due to residual stress exerted by the bumps.

Figure 4.5: Laser scanning microscope image taken with 1340 nm laser at a) 5x, b) 10x, c) 20x, and d) 50x magnification.

Taking a FIB cross-sectional image at the thinnest part as shown in Figure 4.6, the final RST of the device at the thinnest was physically measured to approximate that measured by the spectrometer. Also notable through this cross-sectional image is the scale of the ultra-thinned silicon compared to that of the active circuitry below. With the expanse of transistors and interconnects, the deterioration of thermal stability when ultra-thinned can be better understood.

Figure 4.6: FIB cross-sectional image of thinned sample at thinnest.

In conclusion, targeted ultra-thinning by 2D milling proves to be useful in localising thinning of silicon with adequate uniformity within a small window for imaging. However, the possibility of the targeted area shifting due to sample relaxation and polishing non-uniformity as well as limitations in local uniformity achieved necessitates other more precise methods of target silicon ultra-thinning.

4.2 Targeted Contour Milling

Noting the challenges with thinning a specified AOI deterministically through 2D milling alone, 3D methods were explored. As originally implemented to achieve uniform RST after thinning, the same methods can be exploited to target AOIs specifically with greater local uniformity. Within the 3D contouring workflow, the X-Correct™ step serves to correct the milling profile to account for device relaxation. This is executed with an excel macro correlating the contour map used for milling and the measured RST map. From point to

point on the grid map, the milling depth is increased with thicker silicon measured on the same spot and vice versa. The final map is then smoothened and uploaded to the milling system to mill a uniform RST profile. To achieve the requisite thickness at the area of interest while thickening other parts, the profile can thus be manipulated by modifying the RST map as illustrated in Figure 4.7 for targeted milling.

Figure 4.7: Targeted milling by RST map manipulation during contour map generation.

4.2.1 Experimental Setup and Process

To validate the process, a 22.3 mm x 14.3 mm (319 mm²) sample was taken through the proposed workflow using modified RST maps to target a specified AOI for ultra-thinning. First, an AOI including several points in a window on the grid map were arbitrarily designated for targeted thinning. The sample was mounted and levelled about this area with the convex bow tilted so that the AOI was flat relative to the tool face. Subsequently, it was 2D milled down to 40 µm RST at its thinnest before being left to relax. The sample was then physically profiled and the RST map was collected as Figure 4.8 a). Within the RST map visualized in excel, the thickness values were then increased by approximately the warpage value within the designated AOI as seen in Figure 4.8 b). With the physically profiled map capturing the relaxed profile and this edited map specifying an area to

preferentially thin, a contour map was then generated with the X-Correct™ macro and uploaded for milling. The sample was then similarly polished to target thickness and mirror finish.

Figure 4.8: a) Original RST map and b) edited map with values increased in window.

4.2.2 Results and Discussion

First, a 22.3 mm x 14.3 mm (319 mm²) sample was levelled globally across the die and physically profiled with the tool in a grid map. From the collected physical profile in figure 4.9 a), a convex warpage of 68 μm was measured between the highest point near the center and the lowest points near the corners. A 1.9 mm x 2.4 mm window was then specified within the grid map near the center-right region as the AOI for targeted milling. The depth relative to the highest die center region was measured to be between 7.5 μm and 14.9 μm with an 8.8 μm range. After levelling about the designated window, a tilted convex physical profile was observed with the AOI surface normal to the tool. As observed in Figure 4.9 b), the AOI is boxed up with the smallest depth and range of 2.1 μm found within.

Figure 4.9: Physical contour map after levelling a) globally, and b) locally about AOI.

After 2D flat-top milling till $43\text{ }\mu\text{m}$ at the thinnest, a concave RST profile centered at the AOI was observed in Figure 4.10 a). Within the AOI window, the RST varied between $46.1\text{ }\mu\text{m}$ and $43\text{ }\mu\text{m}$ with a range of $3\text{ }\mu\text{m}$ that approximates the relative physical height difference prior. Due to the inherent device warpage, it was then observed that the RST thickened progressively when moving away from the AOI on the flat-top surface. On this RST map, the thickness values within the AOI window were then increased by $50\text{ }\mu\text{m}$ to give Figure 4.10 b) so it reflected between $96.1\text{ }\mu\text{m}$ and $93.1\text{ }\mu\text{m}$ within. This edited RST map was then used for later profile generation as the physical RST profile was left unchanged for subsequent relaxation and profiling.

Figure 4.10: a) Original RST map after milling and b) edited map with increased thickness at AOI.

After being left to relax, the flat-top device was then physically profiled at three intervals of 10 min, 1 hour and overnight. At 10 minutes, a slight concave physical profile was observed in Figure 4.11 a) with a height range of 4.2 μm owing to mild stress relaxation. The 10-minute interval after 2D milling was chosen as the physical profiling step is time consuming with a probe brought down to contact the sample at every point of the grid map. After an hour, the same was repeated and a more distinct concave profile was observed in Figure 4.11 b) with the height range now at 11.9 μm . Finally, the sample was left for a period of 16 hours overnight to ensure complete relaxation before profiling. As with the prior profiles, a slightly more concaved profile with 14.4 μm depth was then observed in Figure 4.11 c). When compared to a delta of 7.7 μm between the 10 min and 1 hour profiles, the smaller 2.5 μm change overnight suggests that the relaxation happens quickly at the start and tapers off to equilibrium after a longer period.

Figure 4.11: Physical profiles after a) 10 mins, b) 1 hour, and c) overnight.

With this relaxed physical profile and the edited RST profile, a 3D profile was then generated for targeted milling by profile correction. The physical profile in Figure 4.12 a)

now shows a slight concave with 14 μm of elevation from the edges to the center. Meanwhile, the RST map in Figure 4.12 b) shows an exaggerated concave profile with the thinnest center differing from the edge by 100 μm . Offsetting the slight concave physical profile with a severe concave RST profile while accounting for the increased RST values, a map was then generated for milling and smoothened. Shown in Figure 4.12 c), this contour map is convex and centered about the AOI to account for the RST variation around it. Furthermore, a slight dip relative to its vicinity was observed at where the RST values were increased, indicating that the AOI should be milled preferentially. Under normal circumstances, the correction step involving physical map and a measured RST map corrects sections with thicker RST on the physical map to achieve uniform RST. However, doing so with an edited map instructs the system to mill the AOI thinner than that necessary to achieve uniform RST.

Figure 4.12: a) Physical profile map, b) edited RST map, and c) generated contour map.

After this smoothed convex profile was uploaded to the milling system, it was then milled

progressively into the flat-top surface. As observed in the optical images in Figure 4.13, the device was then milled from the deeper parts near the edge towards the AOI. On the partially milled device in Figure 4.13 b), the reflective surface results from 3D milling with a diamond tool while the circular rough region resulted from prior 2D milling using a diamond grinding disc. Upon complete contour milling, a reflective surface was obtained throughout the device with a circular impression at the AOI as shown in Figure 4.13 b). The impression mirrors the dip in the uploaded profile for milling and represents preferential milling of that region. On the measured RST map in Figure 4.13 a), it is also confirmed that the designated 1.9 mm x 2.4 mm AOI window was milled to between 20.1 and 18.7 μm with a range of 1.5 μm . When compared to the 3 μm range obtained within the window after 2D milling, the uniformity obtained by contour milling has thus improved over 2D milling alone. Also, the AOI was still found to be the thinnest sections on the entire device with the thickest part thicker by 21.3 μm at 40.0 μm . While 3D contour milling typically produces a uniform RST profile to within $\pm 5 \mu\text{m}$ of variation, the thinnest region is usually spread throughout the die depending on sample and milling behavior. After subsequent polishing, the thinnest regions then approximate that obtained after milling. It is thus beneficial to have the AOI at the thinnest prior to polishing to prevent over-polishing, especially when ultra-thinning to low thicknesses of 1 μm or under.

Figure 4.13: a) RST map post-milling; optical images of device milled b) partially, and c) completely.

The sample was then polished with progressively finer grits of diamond pastes, from 15 μm to 0.5 μm to smoothen out surface damages induced by the milling process. RST maps were taken after every step and the results are summarized in Table 4.1. Key readings are tabulated globally across the die and locally in the AOI window across each step. Also, the difference between the local and global thickness are also calculated along with the material removed per step. Following the diamond polish steps, a final chemical mechanical polishing step was also performed with colloidal silica without significant material removal.

On the global RST, an average of 21.2 μm was removed throughout the die until a final average thickness of 9.4 μm was obtained. Because of the prior targeted milling step, a large range of RST values is observed initially at 22.6 μm . With polishing, this range is observed to stay approximately constant, indicating that the bow and thus profile introduced by the targeted milling is largely maintained. Looking into the local RST within the AOI window, a slightly lower amount of material removal at 17.8 μm was obtained. While this can be attributed to the variation in polishing rates throughout the die, it can also be the result of the irregular profile milled. Looking at the contour map for milling and the resultant RST map, the small dip at the AOI likely resulted in lesser material removed as the surface was smoothened through polishing. However, the smoothening is also observed to have a beneficial effect on the local RST range as it is observed to drop from an initial 1.5 μm down to a smaller 0.9 μm . Having achieved a final 1.6 μm RST with a variation of 0.9 μm locally, it should meet the RST specifications necessary for VLP.

Table 4.1: Table of RST values through sequential polishing steps from left to right and difference between global and local RST.

	After 3D	15 μ m polish	9 μ m polish	3 μ m polish	1 μ m polish	0.5 μ m polish	0.5 μ m polish	Overall polish
Global RST (μ m)								
Average	30.6	27.6	24.7	21.7	12.6	12.2	9.4	21.2
Max	41.3	35.6	32.0	30.2	27.1	26.6	26.1	15.2
Min	18.7	16.3	14.6	12.6	4.6	4.2	1.2	17.4
Range	22.6	19.3	17.5	17.6	22.6	22.4	24.8	-2.2
Local RST (μ m)								
Average	19.4	17.2	15.2	13.2	5.2	4.6	1.6	17.8
Max	20.1	18.4	16.3	13.9	5.9	5.1	2.2	18.0
Min	18.7	16.3	14.6	12.6	4.6	4.2	1.2	17.4
Range	1.5	2.1	1.7	1.3	1.3	1.0	0.9	0.5
Difference								
Difference	11.2	10.4	9.4	8.5	7.4	7.6	7.8	-
RST removed	-	3.0	3.0	3.0	9.1	0.4	2.8	21.2

On the other hand, it is also necessary to maintain thicker RST distribution on areas outside the AOI for thermal stability. In Table 4.1, the minimum RST values between those tabulated globally and locally are the same throughout polishing. This indicates that the thinnest sections of the device remain within the AOI as intended even after polishing. Thus, targeted milling is suitable both for localising ultra-thin silicon and preventing exposed circuitry in other regions while targeting a low RST at the AOI.

Taking the difference between the global and local averages, it is then observed that the non-uniformity induced by the targeted milling drops progressively from 11.2 μ m down to 7.8 μ m. This implies that with extended polishing, the profile is smoothened out, undoing the milled profile, and introduced non-uniformity. It is thus desirable to reduce polishing to preserve the milled profile and RST non-uniformity for thermal stability. However, inadequate polishing may leave behind surface damage that degrades resolution during SIL imaging. Thus, careful tool dressing and parameter selection during milling is needed to reduce milling induced damage, the polishing needed and thus polishing non-uniformity.

For the 22.3 mm x 14.3 mm sample with a 1.9 mm x 2.4 mm AOI window, a final RST map was then taken with a visible light spectrometer on the measurement tool as shown in

figure 4.14 a). A global RST of average $9.4\ \mu\text{m}$ with a range of $24.8\ \mu\text{m}$ and local RST of average $1.6\ \mu\text{m}$ with a range of $0.9\ \mu\text{m}$ at the AOI was measured. This meant that VLP SIL requirements of under $1\ \mu\text{m}$ variation at the AOI has been met while maintaining thicker silicon elsewhere for heat dissipation. Under optical imaging as shown in Figure 4.14 b), it can also be observed that only the AOI thinned to $1.6\ \mu\text{m}$ for light transmission has its circuitry visible at where it is targeted. Upon electrical testing, multiple samples prepared this way were then found to survive better than samples thinned globally. Thus, it has been demonstrated that increasing the non-uniformity would improve thermal stability through testing.

*all values in x, y, z (RST) given in μm

Figure 4.14: a) Final RST map and b) optical image of ultra-thinned device with AOI visible.

A large area die image was then taken with an IR camera at 1x magnification as shown in Figure 4.15. When imaged with IR light as opposed to visible light as before, a clearer floorplan of the circuitry layout can be observed due to the higher transmission of IR light in silicon. Furthermore, Newton rings are also observed as concentric rings with alternating bright and dark fringes radiating from the AOI. This is formed by interference between signals reflected off the top surface and the bottom surface of silicon subject to local conditions relating to the path difference. Because the signal passes through the silicon thickness and is reflected at the bottom back up with a 180° phase shift, the path difference is equivalent to double the silicon thickness. Constructive interference thus results when double the silicon thickness is at an integer multiple (m) of the signal wavelength in silicon, giving rise to bright fringes. On the other hand, at silicon thicknesses at $(m + \frac{1}{2})$ multiples of wavelength, destructive interferences results, giving rise to dark fringes. Because the refractive index of silicon at infrared wavelengths is approximately 3.5, the wavelength of the imaging light is reduced by a factor of 3.5 within the silicon. The difference in silicon thickness between each successive bright fringe can then be correlated to half the wavelength of the light in silicon.

Figure 4.15 Large area die image taken with IR camera at 1x magnification.

When imaged with a 1340 nm laser using a laser scanning microscope, similar fringe patterns (Figure 4.16) were observed as in the IR camera image because of the same phenomenon. With a higher magnification, a large area spanning two square blocks or tiles can now be observed within the approximately 1.8 mm central ring. Compared to the device prepared by 2D milling alone, this represents a 50% increase in area uniformity. It can then be deduced that the RST variation within the area is under a quarter that of the wavelength of the laser. Scaling the wavelength of the 1340 nm laser by the 3.5 refractive index of silicon, the uniformity is thus calculated at under $(0.25 * 1340 \text{ nm} / 3.5) = 383 \text{ nm}$ within the field. Similarly, local variation in RST variation is also observed mirroring the underlying C4 bump topology.

Figure 4.16: Laser scanning microscope image taken with 1340 nm laser at a) 5x, b) 10x, c) 20x, and d) 50x magnification.

4.3 Polishing Non-Uniformity After Targeted Milling

Having the ability to target ultra-thinning on the device, a difference of 7.8 μm RST between the local and global thicknesses was obtained in the device prepared. This thickness differential offers adequate thermal stability for electrical testing. However, this may not be true for devices that run at higher temperatures or are subjected to extended testing lasting days. It is thus useful to increase the thickness differential to improve the thermal stability of the thinned device. On the previous samples prepared, a cursory glance at the difference in material removed globally and locally points to a polishing differential. Thus, it is instructional to look at the polishing non-uniformity to retain the non-uniformity introduced by targeted contour milling.

4.3.1 Experimental Setup and Process

A 22.3 mm x 14.3 mm (319 mm²) sample was similarly taken through the targeted contour milling workflow described in the previous section. First, an AOI was specified in a window near the upper right quadrant of the device and the sample was then levelled across it. Subsequently, it was grinded down and a targeted contour map was generated for milling. After milling to 20 μm at the thinnest, the sample was then polished with progressively finer grits of diamond paste as did in the previous section. Here however, the full RST maps were collected after every polish step and studied to understand the polishing behaviour.

4.3.2 Results and Discussion

As with the previous device, a concave RST profile was observed centering at the specified AOI. In this case, the AOI was specified over a smaller 1.9 mm x 1.8 mm window at the upper right quadrant of the device. The device thickness measurements were taken and tabulated in Table 4.2 throughout the process. A global minimum RST of 18.6 μm was located within the AOI window with a maximum of 38.9 μm outside. With an intentional non-uniformity milled into the device, an initial global thickness range of up to 20.3 μm

was observed while those within the AOI remained at 2.2 μm . Overall, the global RST range increased slightly through polishing from 20.3 μm to 24.4 μm ., implying an increase in non-uniformity. However, the difference between average RST measured globally and locally decreased from an initial 11.8 μm to 8.0 μm implying the reverse.

Table 4.2: Table of RST values through sequential polishing steps from left to right and difference between global and local RST.

	After 3D	15 μm polish	6 μm polish	1 μm polish	0.5 μm polish	Overall polish
Global RST (μm)						
Average	31.1	24.2	20.6	15.3	13.0	18.1
Max	38.9	33.2	31.3	29.2	29.0	9.9
Min	18.6	14.1	11.2	7.2	4.5	14.0
Range	20.3	19.1	20.2	22.1	24.4	-4.1
Local RST (μm)						
Average	19.3	14.9	12.5	7.8	5.0	14.2
Max	20.8	16.1	13.6	8.8	5.5	15.3
Min	18.6	14.5	11.9	7.5	4.5	14.0
Range	2.2	1.6	1.7	1.4	1.0	1.2
Difference	11.8	9.3	8.1	7.5	8.0	-
RST removed	-	6.9	3.6	5.3	2.3	18.1

To make sense of the previous statistics, it is useful to take a closer look at the changes in the RST map with every polish step (15 μm , 6 μm , 1 μm) in Figure 4.17. First looking at the difference map between the initial map and final map after polishing, it is readily seen that material has been polished off unevenly. Initially thicker areas after contour milling were polished preferentially over the AOI, with the difference map mirroring the initial map. This means that while the as-milled device has regions with much thicker RST, the final RST profile shares the same profile but with a smaller thickness differential. Thus, the drop-in difference between the local and global averages is accounted by the uneven, preferential polishing at the thicker parts. On the other hand, lower polishing rates are also observed around the edges, accentuating the higher thickness of the edge after milling. This then results in a larger RST range post polishing while the overall RST profile becomes more uniform.

Figure 4.17: RST maps after a) 3D milling, b) 6 μm polish, c) final 1 μm polish, and d) overall polished RST.

The difference in polishing rate can be attributed to the tool being run in floating force mode of feedback during polishing. As the rotating tool is rastered across the device, a constant force is exerted to maintain uniform removal. However, when run across contours such as those milled into the device, the tool contacts and polishes the tilted face preferentially. With extended polishing, the resulting surface would then be smoothed into a flat-top similar to that obtained after 2D milling. This means the profile and RST non-uniformity introduced by contour milling would be undone. The implications are two-fold, that a larger non-uniformity than that obtained by 2D milling cannot be obtained

without milling a concave profile and that lesser polishing is necessary to preserve the introduced non-uniformity. In spite of the polishing non-uniformity, the resulting RST profile still maintains the AOI at the thinnest as it was before. This is attributed to the device being levelled about the AOI initially, with polishing restores a flat-top profile centered around the AOI being the thinnest. This implies that the RST non-uniformity cannot be increased beyond that with a flat-top profile and fine milling is necessary to reduce surface damage and thus polishing non-uniformity after contour milling.

Thus far, new techniques have been presented to target ultra-thinning on a device through CNC contour milling. By doing so, it is possible to eliminate two key challenges to ultra-thin sample preparation and VLP by introducing RST non-uniformity. The first lies in the improved thermal stability while the second lies in the maintenance of the AOI as the thinnest section of the device. Because the AOI is maintained at the thinnest throughout the thinning process, there is no risk of over polishing other areas while getting the AOI to low target thicknesses. Thus, the lowest thickness attainable is no longer limited by the global thickness variation as with normal contour milling workflows.

4.4 Conclusion

In this chapter, different targeted milling techniques were studied to localise ultra-thin silicon for visible light probing while maintaining thermal stability. First, targeted 2D milling was achieved through sample tilt and subsequent 2D milling. The sample processed yielded a large thickness differential globally and poor local thickness uniformity. Subsequently, 3D methods of targeted contour milling were also developed and tested on actual devices. The samples ultra-thinned this way yielded greater local uniformity at the expense of a poorer thickness differential globally. With either technique, targeted milling to localise ultra-thin of silicon at AOI was shown to meet the needs for visible light probing while maintaining thermal stability. In anticipation of the need for further ultra-thinned samples, localised FIB milling techniques are then explored in the next chapter

Chapter 5

Localised FIB milling

Mechanical thinning techniques were presented in the previous chapters to ultra-thin onto 5 μm remaining silicon thicknesses for visible light probing. Looking towards the horizon however, developments in the use of shorter wavelength visible or even electron beams for higher resolution probing can be anticipated. This demands yet thinner remaining silicon thicknesses or even thinning onto STI levels. To this end, mechanical methods of thinning will still fall short due to the non-deterministic nature of polishing. This thus limits the use of polishing to target final amounts of material for removal. To address this, precise material removal through FIB milling with Ga and Xe FIBs will be explored in this chapter. The milling rates across different individual systems will be studied to gain a baseline so as to guide precise material removal where necessary. Subsequently, to minimise the introduction of artifacts and possibility of altered circuit performance, the surface of samples milled with different parameters are characterised through TEM. Finally, the results will guide a methodology where ultra-thinning can be localised with precise material removal while minimising induced sample damage.

5.1 Ga FIB Milling Rate Characterisation

In this section, various trenches are milled into silicon substrates with varying beam conditions on a Ga FIB system and subsequently cross sectioned to measure the etch depth. By correlating the dosage of ions administered and the etch depth, the etch rate for the Ga FIB system can be characterised. Subsequently, the gas assisted etching characteristics are also examined with fluorine introduced during milling. Limitations relating to the system and the process are then discussed in relation to its future use case.

5.1.1 Experimental Setup and Process

For subsequent experiments in this section, the Thermo Fisher Scientific Helios G4 FX DualBeam FIB system is used for milling. The FIB system features a gallium source with ion beam energies from 500 V - 30 kV and current ranges between 0.1 pA - 65 nA. The system is also fitted with a multi gas injection system delivering XeF_2 for gas assisted silicon etching. For the silicon substrates, bare dies with full thickness silicon ($\sim 770 \mu\text{m}$) were retrieved from varied diced wafers. The wafer surfaces were all similarly polished to a mirror finish prior and used as is through all experiments. The die was then secured silicon side up to an aluminium sample stub with silver paste applied around and under the die for electrical grounding before heating at 80°C to dry.

To first study the Ga FIB etch rate, a die was loaded into the FIB and a pristine section of the die was located under SEM imaging for milling. The stage was tilted at 52° for normal incidence to the ion beam and brought to eucentric height for the coincidence of electron and ion beam field of views. A square pattern was then defined along with specified parameters of window size, beam current, and time for milling. Raster parameters such as scan pattern, dwell time and line spacing were kept at default during all tests for consistency.

After milling, the resulting window was then inspected and cross sectioned *in-situ*. The window was similarly located under SEM imaging and navigated to one of the edges for trenching. Subsequently, the stage was tilted at normal incidence to the ion beam and

brought to eucentric height. An adequately deep cross-sectional trench of between 20 to 60 μm long was then milled across the edge as illustrated in Figure 5.1. After milling, a SEM image was then taken while tilted at normal incidence to the ion beam. Having corrected for the angled incidence of the electron beam to the visualised face, the height difference between the surface within and outside the window was then measured as the mill depth. The area-normalised dosage delivered was then calculated as follows and further divided by the etch depth to identify the dosage necessary to etch a micron of silicon.

$$\text{Dosage} \left(\frac{\text{nC}}{\mu\text{m}^2} \right) = \frac{\text{beam current (nA)} * \text{time (s)}}{\text{length (um)} * \text{width(um)}}$$

Figure 5.1: Illustration of FIB interface with *in-situ* ion beam milling and electron beam imaging.

For the next experiment, the above was repeated but with XeF_2 introduced during milling to characterise rate enhancements and the milling finish. To perform the above, the sample was tilted to 52° at normal incidence to the ion beam and the gas nozzle was brought in at preset positions of approximately 150 μm over the AOI. Subsequently during milling, the valve to the XeF_2 reservoir was opened to dispense the gas under default conditions. The

gas flow rate used throughout was then kept at default so as not to compromise the system vacuum and risk damage.

5.1.2 Results and Discussion

The etch rate of a gallium FIB was first studied by milling a square window onto a silicon substrate. A $100\text{ }\mu\text{m} \times 100\text{ }\mu\text{m}$ window typical for circuit edit purposes was specified and milled with the maximum beam current of 65 nA over 10 minutes to characterise its maximum etch rate.

As shown in Figure 5.2, striations are found across the initial surface due to the back grinding of the wafer before dicing. After milling, a well-defined square trench was then observed with a relatively smooth milled surface. A rectangular trench was then milled across the edge of the window and the etch depth was measured as the difference between the height of the original surface and milled surface at $0.88\text{ }\mu\text{m}$. With a total ion charge of $39\text{ }\mu\text{C}$ delivered at 65 nA in 10 minutes over $10000\text{ }\mu\text{m}^2$, the dose delivered thus works out to be $3.90\text{ nC}/\mu\text{m}^2$. Furthermore, with the known etch depth, it can be surmised that the etch depth per unit dose is $0.23\text{ }\mu\text{m}^3/\text{nC}$. This value approximates literature reported gallium sputter rates of $0.27\text{ }\mu\text{m}^3/\text{nC}$ gathered through experimentation and simulation. Minor deviations can be attributed to variations in individual systems, including beam current fluctuations during milling as well as beam performance that will be discussed later.

Figure 5.2: a) Initial surface SE image, b) milled window surface image, and c) trench cross section.

Across the edge of the window, a smooth slope is also observed transitioning between the milled surface and the original surface over approximately $2\ \mu\text{m}$. The surface of the slope region is darkened presumably by beam exposure and can be accounted for by two possible reasons. The first lies in the redeposition of sputtered material from the milled trench onto the edge. While the low depth and thus aspect ratio of the trench ensures little obstruction in the path of the sputtered material, there may still be some redeposit material by the edge. The second lies in the interaction of the ion beam tail with regions outside of the defined milling rectangle. Due to the inherent radial distribution of the beam density, a larger spread

of ions and thus beam tail can be assumed at higher beam currents. Physically, this is manifested in a larger spot size and beam tail surrounding the milled regions.

To illustrate, the ion beam was used to mill holes on silicon substrate across different beam currents as seen in Figure 5.3. While it is immediately obvious that higher beam currents form larger beam spots, the associated beam tails are also larger along the periphery as well. Another point to note is the need for beam optimisation, so that the resulting spot is round such as those below 9.3 nA rather than oblong as the 21 nA one. Wherever possible, it would then be prudent for the beam to be optimized to ensure precise milling and consistent sputter rates.

Figure 5.3: Beam spot sizes with associated beam tails at varying beam currents.

For electron beam probing, while $100 \times 100 \mu\text{m}^2$ windows as milled here may suffice for localised failures, larger windows might also be necessary for poorly isolated failures. Moreover, for visible light probing, a larger window would be necessary to accommodate

the SIL tip for landing. Depending on the specifics of the SIL design, this could then put the minimum window size at $400 \times 400 \mu\text{m}^2$. Taking the previously calculated sputter rate of $0.23 \mu\text{m}^3/\text{nC}$, milling a mere $1 \mu\text{m}$ depth would then require $4.43 \text{ nC}/\mu\text{m}^2$ dose per unit area or a $70.9 \mu\text{C}$ total charge over a $400 \mu\text{m}$ square window. Milling with the maximum current of 65 nA , it would then take approximately 3 hours to trench $1 \mu\text{m}$ deep. For a proposed workflow where the sample is mechanical thinned to $5 \mu\text{m}$ before localised milling to under $1 \mu\text{m}$ for shorter wavelength VLP, this means the FIB operation can take up to 15 hours. Because the FIB is a costly tool primarily intended for TEM sample preparation, it might not be practical to spend excessive tool time on silicon trenching and other means to speed up milling rates should be explored.

In order to accelerate silicon trenching, XeF_2 was then introduced for gas assisted etching. For comparison, similar beam conditions to that before were used for milling so a net total charge of $39 \mu\text{C}$ or dose of $3.90 \text{ nC}/\mu\text{m}^2$ was administered. With the trench imaged at tilt as illustrated in the Figure 5.4, the differences in etch profiles across the left and right edge can be observed. Upon cross sectioning, it was then confirmed that while the left edge was etched to $7.92 \mu\text{m}$, the right edge was etched deeper to $10.8 \mu\text{m}$. Taking the average etch depth of $9.36 \mu\text{m}$, a sputter rate of $2.77 \mu\text{m}^3/\text{nC}$ can then be calculated. This represents an enhancement in milling rates at up to 10 times than without gas-assisted etching. Comparing to literature reported enhancement rates of up to 100 times, those found here fall short but can be attributed to limitations in instrumentation leading only to modest enhancements.

Figure 5.4: a) Etched trench SE image at tilt, b) left edge cross section, and c) right edge cross section.

In terms of milling profile, a similar sloped transition was observed across both the left and right edge due to beam tail interactions. However, it was also noted that while the right slope is steep and relatively smooth, the left slope as observed in Figure 5.4 b) is gentler and riddled with pits that results in curtaining effects during the cross section. The same pit-like structure was also observed on the surface surrounding the trench in contrast to a smoother etched surface as in Figure 5.5. The phenomenon can be attributed to the spontaneous reactivity of the introduced XeF_2 with silicon over an extended etch time. Because the GIS needle was inserted to the right of the window, a round shadow of surface pitting was correspondingly imprinted slightly offset to the right.

Figure 5.5: SE images of a) trench with GIS in, b) remnant shadow, c) etched surface, and d) surrounding surface.

With the gas dispensed from a single nozzle from the right during milling, the gas flow expectedly declines away from the orifice. With higher gas pressures to the right of the window, etching rates are also expectedly higher. Thus, the etch depth on the right edge of the window was found to be higher. At the same time, there is also neither redeposited material nor pitting along the slope observed as chemical etching dominates sputtering and the volatile products are effectively pumped out of the system. On the contrary, lower local gas pressures on the left edge further from the nozzle may result in the surface gas molecules being depleted prior to milling, leading to a lower etch rate and depth as measured. Consequently, sputtering dominates chemical etching and a higher rate of redeposition on the left edge can be observed.

In the absence of specialized coaxial nozzles for gas flow uniformity, periodic rotation of the sample during etching could improve the etch uniformity just by using existing

hardware. For example, the trench could be etched for a minute in a fixed orientation before rotating the sample by 90° for another minute. The same can be repeated two more times for the other two edges of the trench such that each side is etched at equal interval. Smaller increments of rotations used can also be expected to improve the uniformity of the trenched surface but at the expense of time. However, an anticipated issue would then be that a resulting convex profile would be formed in the trench rather than a flat surface. With such a profile, intimate contact between the SIL and the silicon surface may not be possible and imaging resolution may deteriorate with an air gap present. Thus, further optimisation on etch uniformity was not explored further.

In summary, it can be gathered that the gas flow and distribution by this GIS configuration is non-ideal and compromises etch rates, uniformity, and surface finish. Should purpose-built coaxial GIS nozzles be available, higher uniform gas flows can be delivered to the trench for higher and uniform etch rates. In addition, the gallium FIB is current limited for large area silicon trenching within practical timeframes. While the use of XeF_2 allows for an approximate 10 times enhancement in etch rates, the resulting window is non-uniformly etched even over a modest $100\text{ }\mu\text{m}$ square window. To overcome limits in etch rates without dedicated GIS hardware for silicon trenching, the Xe PFIB becomes necessary with its higher beam current range, sputter yield and thus etch rate.

5.2 Xe PFIB Milling Rate Characterisation

Due to limitations in the maximum beam current available to a gallium FIB and the lack of hardware for uniform distribution of gas for etching while milling, its applications for large area silicon trenching is limited. In this section, the use of a Xe PFIB system for its potentially higher beam current and sputter rate is explored by having various trenches milled into silicon substrates under varying beam conditions. The trenches are similarly cross sectioned to measure the mill depth and by correlating it with the dosage of ions administered, the mill rate for the Xe PFIB system is derived. Subsequently, its gas assisted etch characteristics is also examined with fluorine introduced during milling. Limitations relating to the system and the process are then discussed in relation to its use case.

5.2.1 Experimental Setup and Process

For subsequent experiments, the Tescan Xeia3 PFIB system was used for silicon trenching while cross sectioning was performed on a Thermo Fisher Scientific Helios G4 FX DualBeam FIB system. The PFIB system features a xenon plasma source with ion beam energies between 3 kV - 30 kV and current ranges between 1 pA - 2 μ A. The system is fitted with a five-nozzle gas injection system, of which one dispenses XeF₂ for gas assisted silicon etching. For the silicon substrates, similar bare dies with full thickness silicon (770 μ m) were retrieved from varied diced wafers. The wafer surfaces were all similarly polished to a mirror finish prior and used as is through all experiments. The die was then secured silicon side up to an aluminium sample stub with silver paste applied around and under the die for electrical grounding before heating at 80°C to dry.

To first study the Xe PFIB mill rate, a die was loaded into the PFIB and a pristine section of the die was located under SEM imaging for milling. The stage was then tilted to normal incidence to the ion beam and brought to eucentric height for the coincidence of electron and ion beam field of views. A square pattern was then defined along with parameters of window size, beam current and time for milling. The beam current was then measured by deflecting the beam into an off-axis faraday cup and recorded before milling. Raster parameters such as scan pattern, dwell time and line spacing were kept at default during all tests for consistency. For gas assisted etching experiments, the gas nozzles were brought in at preset positions approximately 150 μ m over the AOI to dispense XeF₂ gas during etching.

After milling, the resulting window was inspected and loaded into a gallium FIB for cross sectioning. The window was similarly located under SEM imaging and navigated to one of the edges for trenching. The stage was then tilted at normal incidence to the ion beam and brought to eucentric height. An adequately deep cross-sectional trench of between 20 to 60 μ m long was then milled across the edge as did before. After milling, a SEM image was then taken while tilted at normal incidence to the ion beam. Having corrected for the angled incidence of the electron beam to the visualised face, the height difference between the surface within and outside the window was then measured as the mill depth. The dosage

delivered was then calculated as before and further divided by the mill depth to identify the area-normalised dosage necessary to etch a micron of silicon.

For comparison, a 100 μm square window similar to that before was first milled using comparable beam currents. Subsequently, preset optimised beams with higher beam current and energies were used for milling to quantify the sputter rate. Larger windows up to the upper limits available were then milled to check for the milling profile, rate, and finish. Finally, for gas assisted etching, XeF_2 was introduced during milling to characterise rate enhancements and to study the milling finish at larger scales.

5.2.2 Results and Discussion

First, a frequently used beam preset (15 kV/121 nA) with an intermediate beam energy was first loaded onto the PFIB and tuned so that the beam current was decreased to match the maximum on the gallium FIB at 66 nA. While milling at 30 kV would be a more equitable comparison, its use is limited to higher beam currents and the sputter rates would have to be quantified and compared separately. Subsequently, three similar 100 μm x 100 μm windows were then milled under default conditions over 1, 5 and 10 minutes to characterise the mill rate.

For the silicon substrate used for this section, the backside was polished to a mirror finish, so that the surface was largely pristine. Correspondingly, the surfaces also remained largely smooth even after milling. During sample transfer between the FIBs however, some contamination was introduced as seen as the white spots on the surface. Upon SEM imaging as shown in Figure 5.6, the large transition region around the periphery of the milled windows is immediately obvious. Measuring up to 22.6 μm , this is approximately 10 times that on the gallium FIB and is likely the result of a broader xenon beam used for milling. However, when milling with a gallium FIB at 65 nA previously, it occurred under a moderate mill rate regime where a volume encompassing 100 μm x 100 μm x 0.88 μm of silicon was milled in 10 minutes for a mill rate of 14.7 $\mu\text{m}^3/\text{s}$. According to Figure 2.10, the beam spot size of the gallium FIB should be larger than a that of xenon PFIB within

such a regime and the larger Xe beam tail experienced here is anomalous. Further, an abnormally low etch depth of $0.65\ \mu\text{m}$ and sputter rate of $0.16\ \mu\text{m}^3/\text{nC}$ was also measured from the cross section, even accounting for the lower beam energy used. Taking a closer look at the other windows milled, a similarly large beam extension and broad range of low sputter rate between 0.10 and $0.29\ \mu\text{m}^3/\text{nC}$ was also calculated. To confirm the milling characteristics, the beam preset (15 kV/121 nA) was then loaded and used as is to mill another series of three windows.

Figure 5.6: SE images of a) milled windows, b) 10 min window milled trench, and c) trench cross section.

On this series of windows milled with the preset beam and measured currents of 115 nA, a smooth milled surface was similarly observed (Figure 5.7). However, it is also apparent that the transition region along the periphery is now smaller at about $6.7\ \mu\text{m}$. Upon cross-sectioning, milling depths of $0.24\ \mu\text{m}$, $1.20\ \mu\text{m}$ and $2.44\ \mu\text{m}$ were measured for the 1 min, 5 min and 10 min windows, respectively. Calculating the sputter rate, all three $100\ \mu\text{m}$ square windows then yielded $0.35\ \mu\text{m}^3/\text{nC}$.

Figure 5.7: SE images of a) milled windows, b) milled trench across 10 min window, and c) trench cross section.

Upon comparison to the fluctuating results obtained previously, it then suggests that the beam could be unoptimized for operation at 66 nA. Indeed, during installation and regular preventive maintenance, the system undergoes mechanical and beam alignments for each

of the frequently used beam presets. Should the beam be used outside of these presets, an unoptimised beam characterised by a larger spot size and unstable beam current would result. To study its effect on beam spot size, a frequently used preset was loaded and the optimised beam was used to mill a spot in silicon. Subsequently, its condenser voltage was changed, and the beam was used to mill another spot. As seen in Figure 5.8, the optimised beam milled a spot with smaller beam extension compared to the one milled by an unoptimised beam on the right.

Figure 5.8: Associated spot sizes and beam tails of optimised (left) and unoptimised (right) beams.

To further visualise the effects of milling with an unoptimised beam, the condenser voltage was then tuned again so that the beam current was at 300 nA as compared to the preset 115 nA. Milling under these conditions then yielded rounded windows with larger beam extensions relative to that milled with an optimised beam as illustrated in Figure 5.9. Due to the presence of a thin layer of native oxide on the silicon, differences in thickness due to milling by the stray beam is observable by the colour gradient within and around the windows as observed in Figure 5.9. With this in consideration, further experiments should

be limited to existing optimised beam presets for optimal milling performance. To add on, due to instrument limitations, different beam energies are limited to different field of views and current ranges. For example, 15 kV beams are limited to about 285 nA and 784 μm field of views. Meanwhile, 30 kV beams are able to go up to 2 μA beam currents but are limited to 392 μm field of views.

Figure 5.9: Optical images of various windows milled across different currents and time.

Having milled and measured sputter rate with small 100 μm x 100 μm windows, larger windows up to that possible at 15 kV beam energies were then attempted. For a 15 kV beam, the largest window corresponding to its maximum field of view of 700 μm and a medium size window of 200 μm were then milled. As shown in Figure 5.10, milling with optimised beams yielded regular and moderate beam extensions around the window edges. To mill a 200 μm window to substantive depths as achieved in the 100 μm window, the dosage needs to scale up with the increase in area. Thus, a milling time of up to 40 mins was chosen to maintain the same dosage per unit area at the same beam current. A similar

longer mill time of up to 12 hours was also used for yet larger windows of 700 μm to achieve substantive mill depths for comparison and calculation.

Figure 5.10: SE images of a) 200 μm and b) 700 μm windows and its cross sections.

For the 200 μm windows, two were milled for 20 mins and 40 mins with a preset beam current of 112 nA to a depth of 1.27 μm and 2.64 μm , respectively. With the milling depth of the 40 min window at about double that of the 20 min window, an approximate sputter rate of $0.38 \mu\text{m}^3/\text{nC}$ was then calculated for these 200 μm windows. The 700 μm windows gave a similar trend, with the 12 hour milling window about 4 times deeper than a 3 hour milling window at 4.01 μm versus 1.08 μm . Sputter rate calculations for the 700 μm window milled with a 15 kV beam then yielded an approximate $0.36 \mu\text{m}^3/\text{nC}$.

Accounting for the sputter rates thus far, an average of $0.35 \mu\text{m}^3/\text{nC}$ was calculated which matches literature reported values for the Xe ion beam, though at 30 kV. While the calculated sputter rates remain relatively consistent within the same window size, a small spread is observed amongst different sized windows. The minor variation between the different sized windows can possibly be accounted by the time scales each was milled to standardise the dose per unit area. With smaller windows, milling was done through a shorter period of up to an hour so the beam would be stable through this period and perform

consistently. When moving up to 40 min mill times with 200 μm windows and 12 hours with 700 μm windows, the period through which the beam was operated lengthens significantly and thus beam fluctuations during then might be more significant. Despite the minor variations however, the milling rate was found to remain largely proportionate to the dosage delivered. The dosage delivered thus far was then varied as a function of the milling time and area with the beam current kept constant. Similar scaling is expected with beam current though and to exemplify it, another preset with 225 nA beam current was used to mill another 700 μm window. When compared to the 700 μm window milled at 133 nA for 3 hours, milling with approximately double the current gave about double the mill depth of 2.05 μm over 1.08 μm . Indicative of an optimised ion beam, a well-defined square window with a moderate beam extension of 10 μm was also observed as shown in Figure 5.11.

Figure 5.11: SE images of a) window milled at 225 nA and b) its cross section.

The mill rate has been established thus far to be dependent on the dosage delivered per unit area. The dosage in turn, is then dependent on the product of beam current and time. With larger windows, the gallium FIB is crippled by long mill times due to low maximum beam currents limiting the dosage deliverable. On the contrary, the xeon PFIB is capable of higher maximum beam currents of 225 nA at 15 kV presented before and up to 2 μA at 30 kV. However, it should be noted that the maximum window size is capped at 392 μm at 30 kV due to optimisations of the scan coils for high beam currents.

To examine the increase in milling rate associated with a magnitude of increase in beam current, two 200 μm windows were milled with 1 and 2 μA presets at 30 kV. Similar durations as that milled with 121 nA at 20 mins before was then chosen to for comparison. According to Figure 5.12, regular square windows are observed with significant beam extension diffused around the window for the 2 μA window. Measured at between 54 μm and 92 μm , the larger extension is attributed to the larger beam spot size and beam tail at such high beam currents. On the 1 μA window, a square dark imprint was overlaid over the trench, masking the beam extension as the ion beam was rastered over the surface before the image was taken. Further, white streaks are observed over the surface as the result of surface contamination during sample transfer between the PFIB and FIB for cross section.

Figure 5.12: SE images of a windows milled at a) 1 μA with c) its cross section, and at b) 2 μA with d) its cross section.

Looking at the cross sections, milling depths of 13.8 μm and 26.1 μm were measured for the 1 μA and 2 μA windows respectively. Before milling, a beam current of 1144 nA and 2206 nA was also measured for the 1 μA and 2 μA presets respectively. Accordingly, the milling depth increased proportionately by about 10 and 20 times that milled at 112 nA. Considering a 20 min mill time, a dose of 34.3 nC/μm² and 66.2 nC/μm² was thus calculated respectively, accounting for a similar sputter rate of 0.40 μm³/ nC each.

Compared to that calculated at 15 keV for xenon beam, sputter rates attained at 30 keV here are expectedly higher due to the greater energy of the impinging ions at greater accelerating voltages. Similarly, when the same 30 keV accelerating voltage is used, the heavier xenon ion yields a higher sputter rate ($0.40 \mu\text{m}^3/\text{nC}$) than the lighter gallium ions ($0.23 \mu\text{m}^3/\text{nC}$) by virtue of its mass and thus incident energy. With the higher milling rate and window depth, more redeposited material is observed on the side wall of the window as a result. During cross sectioning, this is also reflected in the curtaining effects observed under the slope as surface debris induce differential milling rates along the cutting face. In relation the window's purpose for visible light or electron probing however, it is not anticipated to be much of an issue.

Subsequently, the window size was then taken to its limit of $392 \mu\text{m}$ and 3 windows were milled up to 60 mins. Consequently, milling for 10 mins, 30 mins and 60 mins yielded proportionate etch depths of $3.97 \mu\text{m}$, $10.0 \mu\text{m}$ and $21.45 \mu\text{m}$ respectively to give an average sputter rate of $0.42 \mu\text{m}^3/\text{nC}$. As seen in the cross-sectional image in Figure 5.13, redeposit material by the slope wall increases at above $10 \mu\text{m}$ mill depths but should otherwise be insignificant at the mill depths intended. From the trials undertaken by milling at high currents, the PFIB has thus proven capable of consistently trenching large windows to accommodate moderate SIL lenses under reasonable time frames.

Figure 5.13: Cross-sectional image of windows milled at $2 \mu\text{A}$ for a) 10 mins, b) 30 mins, and c) 60 mins.

Finally, experiments were carried out to study gas assisted etching characteristics for windows etched in the PFIB. First, a 200 μm window was milled with the 15 kV/121 nA beam preset at 146 nA while XeF_2 was dispensed over the surface from a nozzle. As seen in Figure 5.14, concentric rings of shadows can be observed around the window. The innermost ring is likely the result of the beam extension as before due to the relatively smooth surface finish. The outer rings are then characterised by numerous pitting and can be attributed to spontaneous fluorine etching as encountered in the gallium FIB. Looking at the window, a slightly roughened surface is observed with mild pitting along the edges. With an appreciable etch depth of 3.91 μm , gas flow along the edges of the window might be lower and non-uniform, resulting in the observed pits manifested during etching. Overall, etching is noted to be uniform across the entire 200 μm window and indicative of even gas distribution across the window. This is well within expectation as per the instrument's specifications for uniform gas flow at under 200 μm .

Figure 5.14: SE images of a) 200 μm window etched over 10 mins at 146 nA, b) surface pitting, and c) its cross section.

When cross sectioned, an etch depth of $3.91\ \mu\text{m}$ was then measured for a dose of $2.24\ \text{nC}/\mu\text{m}^2$ delivered. Normalising the etch depth to the dose delivered, a sputter rate of $1.79\ \mu\text{m}^3/\text{nC}$ was then calculated. In comparison to sputter rates of $0.35\ \mu\text{m}^3/\text{nC}$ for the ion beam alone, etching with XeF_2 thus enabled an approximate 4.8 times increase in milling rate. While not as high as with the gallium FIB obtained before, it could be attributed to instrumental differences with reference to the proximity of the gas nozzle to the surface and gas pressure used.

A larger $300\ \mu\text{m}$ window was then milled for 23 mins to maintain the same dosage delivered. As seen in Figure 5.15, it yielded a similar uniform roughened milling finish with pitted peripheries. Here however, a lower milling depth of $1.77\ \mu\text{m}$ was obtained, giving a sputter rate of $0.79\ \mu\text{m}^3/\text{nC}$. While it is still about 2.1 times than without gas etching, enhancements here are under half of what was obtained with the $200\ \mu\text{m}$ window.

Figure 5.15: SE images of a) $300\ \mu\text{m}$ window etched over 23 mins at 146 nA, b) surface pitting, and c) its cross section.

Following, a 400 μm window was then milled for 40 min under similar conditions to maintain the same dosage as before. As Figure 5.16 shows, severe surface pitting can now be observed within the window. Taking a cross section, an etch depth of 5.87 μm was also measured from the left edge of the window. A high sputter rate of 2.68 $\mu\text{m}^3/\text{nC}$ was then calculated, representing an increase in rate of about 7.2 times over that without gas assisted etching. Comparing between the previous windows milled, a wide range of enhancement rates between 2.1 and 7.2 times thus suggests high variability for etch depth and pitting depending on the etch conditions.

Figure 5.16: SE images of a) 400 μm window etched over 40 mins at 146 nA, b) surface pitting, and c) its cross section.

Notably, the pitting here was visibly severe near the top left corner and tapers off towards the bottom right corner of the window. As seen in Figure 5.17, the orientation of the graded profile corresponds to the offset direction from which XeF_2 was dispensed towards the

window. Furthermore, the sidewall here is higher than before where only mild pitting was encountered around the window edges. Thus, it can be deduced that there is poor coverage by the top left quadrant resulting in shadowing effects and pitting in those areas. Consequently, gas assisted etching is not recommended for large area processing due to the lack of dedicated gas delivery systems to ensure uniform gas distribution.

Figure 5.17: SE images of a) GIS nozzle configuration and b) etch window with inset of higher magnification images at various sections over 400 μm etched window.

To conclude, Table 5.1 summarises all the windows milled and the associated parameters for both the gallium FIB and xenon PFIB. Through the experiments conducted, the product of the current (nA) and exposure time (s) gives the net charge of ions (nC) delivered for milling. When normalised over the window size (μm) and thus milling area (μm^2), a dosage

value ($\text{nC}/\mu\text{m}^2$) was then derived for correlation with its actual mill depth. The dose normalised mill depth then gave the sputter rate for each system at a particular beam energy.

Table 5.1: Table of milling parameters and results.

Preset	Beam Energy (kV)	Current (nA)	Window Size (μm)	Area (μm^2)	Exposure time (s)	Dose ($\text{nC}/\mu\text{m}^2$)	Mill depth (μm)	Sputter rate ($\mu\text{m}^3/\text{nC}$)
Gallium FIB								
30kV/65nA	30	65	100	10000	600	3.90	0.88	0.23
30kV/65nA (XeF ₂ GAE)	30	65	100	10000	600	3.90	9.36	2.77
Xenon PFIB								
15kV/121nA	15	66	100	10000	60	0.40	0.11	0.29
					300	1.98	0.21	0.10
					600	3.96	0.65	0.16
15kV/121nA	15	115	100	10000	60	0.69	0.24	0.35
					300	3.45	1.20	0.35
					600	6.90	2.44	0.35
15kV/121nA	15	112	200	40000	1200	3.36	1.27	0.38
					2400	6.72	2.64	0.39
15kV/121nA	15	133	700	490000	10800	2.93	1.08	0.37
		128			43200	11.28	4.01	0.36
15kV/285nA	15	225	700	490000	10800	4.96	2.05	0.41
30kV/1 μA	30	1144	200	40000	1200	34.32	13.83	0.40
30kV/2 μA	30	2206	200	40000	1200	66.18	26.14	0.40
30kV/2 μA	30	2118	392	153664	600	8.27	3.97	0.48
		2267	392	153664	1800	26.56	10.00	0.38
		2256	392	153664	3600	52.85	21.45	0.41
15kV/121nA (XeF ₂ GAE)	15	146	200	40000	600	2.19	3.91	1.79
			300	90000	1380	2.24	1.77	0.79
			400	160000	2400	2.19	5.87	2.68

Plotting the mill depth against the dose as seen in Figure 5.18 then showed a linear relationship with a factor of sputter rate. With a sputter rate of $0.35 \mu\text{m}^3/\text{nC}$ at 15 kV and higher $0.40 \mu\text{m}^3/\text{nC}$ at 30 kV, it is deduced that a higher mill rate can be achieved using higher beam energy. At the same beam energy (30 kV), it is also noted that heavier xenon ions yield higher sputter and thus mill rates as opposed to lighter gallium ions at $0.23 \mu\text{m}^3/\text{nC}$.

Figure 5.18: Plot of mill depth versus dose at different beam energy.

As the objective of using the FIB is to locally trench ultra-thinned silicon up to 5 μm deep for visible light or electron beam probing, windows up to 400 μm and beyond could be necessary. To this end, the gallium FIB is limited to slow mill rates for large windows by its moderate sputter rates ($0.23 \mu\text{m}^3/\text{nC}$) and low maximum beam current (65 nA). On the other hand, the xenon PFIB offers higher sputter rates ($0.35 - 0.40 \mu\text{m}^3/\text{nC}$) and a magnitude increase in beam currents (2 μA). The combination of higher sputter rates and beam currents then allows for large windows to be trenched in reasonable time just using the ion beam alone. With the ion beam mill rate characterised in relation to its sputter rate, controllable amounts of silicon can then be trenched on prior thinned samples. With XeF_2 gas delivery systems for gas assisted etching, both FIBs exhibit significant, albeit variable enhancements on the milling rate. However, due to the adequate ion beam mill rate at high currents, trenching is still better performed with the beam alone for controllability and uniformity over large scales.

5.3 Xe PFIB Milling Finish Characterisation

In localising ultra-thin silicon for electron beam probing, the xenon PFIB is proposed as a means to remove material controllably and within reasonable time frames. After CNC milling to approximately 5 μm RST at the AOI, the PFIB is used to trench silicon in a window down to 1 μm or to the STI level with about 100 nm of silicon left until the active circuitry. While previous results show a largely smooth surface after milling, FIB induced damage remains an inherent reality of FIB milling. The induced damage can include

amorphisation of the otherwise mono-crystalline silicon, implantation of the incident ions and surface roughening. At target sub-micron thicknesses, these then threatens to introduce artifacts and device alterations delirious to further testing. In this section, the damage induced to the silicon is thus studied as a function of the milling parameters to preserve the sample at ultra-thin thicknesses.

5.3.1 Experimental Setup and Process

For subsequent experiments, the Tescan Xeia3 PFIB system was first used for silicon trenching while TEM sample preparation was subsequently performed on a Thermo Fisher Scientific Helios G4 FX DualBeam FIB system. For the silicon substrate, a bare die with full thickness silicon (770 μm) was retrieved from a polished and diced wafer. The die is then secured silicon side up to an aluminium sample stub with silver paste applied around and under the die for electrical grounding before heating at 80°C to dry.

On a pristine silicon substrate, a window was milled using the PFIB with procedures described in the last section. Beam conditions including beam energy and current as well as the use of gas assisted etching were varied to simulate different milling finishes obtained through PFIB milling. To start with, a 200 x 200 μm^2 window was milled with a moderate 15 kV beam at 124 nA for 10 mins as a baseline for which other parameters could be evaluated against. Subsequently, a lower beam current of 11.2 nA was used to mill at the same 15 kV for 10 mins to examine if different dosages administered impact the surface finish differently. After which, windows were milled with 30, 10, 5, and 3 kV beams to study beam induced damage in silicon as a function of beam energy. Finally, the impact of gas assisted etching was also studied by milling a 200 x 200 μm window with a 15 kV beam at 128 nA for 10 mins under XeF_2 . After milling all the windows on the same substrate over different pristine sections of silicon, the sample was transferred over to the gallium FIB for TEM sample preparation. A thin lamella was then prepared off the center of each window and thinned to under 100 nm thickness using *in-situ* lift out techniques detailed below for optimal imaging.

As widely reported in literature, the *in-situ* lift out method was used to prepare a site-specific lamella from within the milled window to under 100 nm thickness. To maintain the mechanical integrity of the sample, it was prepared as an inverted wedge with the AOI as the thinnest section. Meanwhile, the bulk silicon section was maintained at a higher thickness for mechanical stability to resist warping as the lamella was thinned.

Towards final thinning, progressively lower ion beam energies of 5 kV and finally 2 kV was used to polish the lamella so as to reduce gallium ion beam induced damage on the sample. Upon TEM inspection, any damage observed or xenon implantation in the lamella should thus be induced only by initial PFIB milling rather than through sample preparation on the gallium FIB. After the lamella was attached and thinned on a copper grid, it was transferred to a Thermo Fisher Scientific Titan Themis G2 300 TEM for imaging. TEM and EDS were then performed with a 300 kV beam at varying magnifications to examine the sample crystallinity and elemental composition.

5.3.2 Results and Discussion

To summarize the milling conditions examined in this section, the list of parameters for each window milled is summarized in Table 5.2. With the beam presets optimised, the actual beam current was measured manually measured before use and recorded. Similar 200 x 200 μm^2 windows were then milled over 10 mins. As the objective in this set of experiments is to study the surface finish, the etch depth was not measured as before.

Table 5.2: Table of milling parameters used to mill 200 μm square windows over 10 mins.

	Preset	Beam energy (kV)	Current (nA)	Dose (nC/ μm^2)
1	15kV/121nA	15	124	1.86
2	15kV/12nA	15	11.2	0.17
3	30kV/11nA	30	14.5	0.22
4	10kV/21nA	10	16.8	0.25
5	5kV/2nA	5	3.3	0.05
6	3kV/2nA	3	1.4	0.02
7	15kV/121nA (F)	15	128	1.92

For sample 1, a 200 μm window was milled with a 15 kV beam at 124 nA for 10 mins. A lamella was then prepared off the window seen in Figure 5.19 and final thinning was done with a 2 kV ion beam glancing angle polish for subsequent TEM analysis.

Figure 5.19: a) Sample 1 SE image of window at 52° tilt and b) deposited marker.

Upon imaging, a 38 nm thick amorphous layer is seen (Figure 5.20) sandwiched between the deposited protection layers and the monocrystalline bulk silicon. Running uniformly along the surface, a smooth interface with a roughness of under 2 nm was measured. At higher magnification, two distinct layers can be further resolved with a 21 nm patchy upper layer and a 17 nm homogenous lower layer within the amorphous layer.

Figure 5.20: TEM image of sample 1 at various magnification.

Looking to the EDS maps in Figure 5.21, the upper layer was xenon rich while the remainder of the amorphised layer was not. Taken together, a total 38 nm of silicon was amorphised with 21 nm of it being implanted by the incident xenon ions. Compared to literature reported amorphisation depths of 14 nm at 5 kV and 54 nm at 30 kV, the values here strike an intermediate as expected of a 15 kV beam. To explain the observations, the amorphous layer in silicon is a loss in lattice periodicity due to the impingement of energetic xenon ions during milling. As the incident xenon ion impinges onto the crystalline sample, it collides and loses energy to the lattice in a collision cascade before coming to rest within the sample at its projected range. This then forms the xenon rich amorphous layer as observed. On the other hand, near total momentum transfer during face to face collisions between the incident ions and the lattice atoms can also recoil silicon atoms. This then sets additional recoiled silicon atoms off into the sample, resulting in the lower amorphous layer void of xenon atoms. The net damage induced by the incident xenon ions and recoiled atoms then totals 38 nm while xenon is only implanted in the upper 21 nm. Noting the spontaneity of silicon oxidation in atmosphere, a thin layer of native oxide is also observed on the surface.

Figure 5.21: STEM images and EDS maps of sample 1.

To further examine variations in damage with different beam currents, the above was repeated in sample 2 using a much lower current with the resulting window as seen in Figure 5.22. Compared to sample 1, the milled window is visibly shallower but otherwise similar.

Figure 5.22: a) Sample 2 SE image of window at 52° tilt and b) deposited marker.

From Figure 5.23, sample 2 shows a similar 35 nm thick amorphous layer on the surface along a 2 nm smooth interface. At higher magnification, a similar two-layer structure was observed with an implantation depth of 24 nm as well. Compared to sample 1, similar beam damage profile was observed even at one-tenth the dosage. This is within expectations as beam damage is induced as a function of the incident beam energy rather than current. Subsequent experiments then examine the variation of the induced damage with beam energy rather than current.

Figure 5.23: TEM image of sample 2 at various magnification.

Taking an EDS map at the interface as shown in Figure 5.24, a similar profile is observed with xenon implanted in the upper amorphous layer. On pristine silicon, a thin native oxide layer is normally formed on the surface due to energetic dangling bonds reacting with atmospheric oxygen. After milling, a similar oxide layer is also observed on the surface.

Figure 5.24: STEM images and EDS maps of sample 2.

After establishing the extent of damage at 15 kV beam energy and its invariability with beam current, 30 kV milling was then examined in sample 3. A 200 μm window was milled at 14.5 nA for 10 minutes. Milling at normal incidence with a 30 kV beam, an approximately 54 nm thick amorphisation layer can then be expected according to literature reported values in Table 2.2. As illustrated in Figure 5.25, this is confirmed with a larger implantation depth of 36 nm and total amorphisation layer of 64 nm measured. Correspondingly, the damage induced at 30 kV roughly doubles that induced at 15 kV by the higher energy of the incident ions.

Figure 5.25: TEM image of sample 3 at various magnification.

Taking an EDS map of the interface as shown in Figure 5.26, a similar map was observed as before. The upper 37 nm was confirmed to be implanted with xenon ions while the lower 27 nm was found without. As this is the highest beam energy available to the PFIB system, an upper limit in damage induced can thus be placed at about 64 nm. Furthermore, because the highest currents available to the system can only be accessed at 30 kV, a lower limit of about 60 nm remaining silicon thickness is necessary for the largest windows so as not to amorphise active devices. Practically however, the difficulty and precision of end point detection of the milling will limit the thinnest silicon possible with the PFIB to about 300 nm from the active fin base. Regardless, should a 30 kV beam be used to thin down to STI level, a large proportion of the remaining silicon would be amorphised with possible device degradation. It would thus be prudent to perform final thinning with lower beam energies as studied in the following samples.

Figure 5.26: STEM images and EDS maps of sample 3.

When milling onto thinner silicon, lower beam energies would be necessary to reduce beam damage that could deliriously affect device performance. Here, sample 4 was prepared with a 10 kV, 16.8 nA beam using similar parameters for milling. Shown in Figure 5.27, the window milled is characterised by softer rounded edges and a visibly shallow mill depth. When measured, the size of the window also significantly exceeds the 200 μm defined for milling at 280 μm and the discrepancy can be due to the indistinguishable beam extension from the window itself. Between the large beam extension and diffused edges, both can be attributed to an inherently broader ion beam at lower beam acceleration.

Figure 5.27: a) Sample 4 SE image of window at 52° tilt and b) deposited marker.

Taking a look at the TEM image for sample 4 in Figure 5.28, another similar damage profile was observed with an implantation depth of approximately 17 nm and amorphisation layer of 24 nm. With an amorphisation layer of 24 nm at 10 kV, 35 nm at 15 kV and 64 nm at 30 kV, the empirical data suggests that the bottom wall damage induced approximately doubles with the beam energy.

Figure 5.28: TEM image of sample 4 at various magnification.

Looking to the STEM and EDS map in Figure 5.29, a similar profile is described with a 17 nm xenon-rich layer and distribution of oxygen within the amorphised silicon.

Figure 5.29: STEM images and EDS maps of sample 4.

Subsequently, yet lower beam energies were studied with sample 5 milled at 5 kV and 3 nA over 10 mins as shown in Figure 5.30. As before, the milled window is visibly shallower and the surface is relatively smooth.

Figure 5.30: a) Sample 5 SE image of window at 52° tilt and b) deposited marker.

Upon imaging at low magnification in Figure 5.31, a 67 nm thick layer is observed above the monocrystalline bulk silicon. At higher magnification, the layer is further resolved into two distinct layers with a lower layer with observable patterns and an upper layer without. Collecting the diffraction patterns from each layer, the upper 14 nm layer yields a diffuse ring pattern indicative of amorphous silicon. Meanwhile, the lower 53 nm layer yields a similar pattern with distinct spots indicative of a mix of amorphous and polycrystalline silicon.

Figure 5.31: TEM image of sample 5 at various magnification, with insets showing the diffraction patterns.

Performing STEM and EDS as seen in Figure 5.32, the 14 nm amorphous layer is confirmed to be implanted with xenon while the underlying 53 nm polycrystalline layer was found without. Compared to previous samples, the xenon-free semi-crystalline layer takes the place of an amorphous layer. Considering the low beam energy and implantation depth however, the same mechanism causing the amorphisation layer in the other samples is unlikely to explain the 53 nm thick semi-crystalline layer here. Furthermore, when milling with 5 kV here, the actual mill depth could be very low, and the expected interaction volume should not induce damage to the whole 53 nm layer.

Figure 5.32: STEM images and EDS maps of sample 5.

As the semi-crystalline layer is thick relative to the amorphised layer, it is unlikely to be beam induced. To better understand its structure, a control sample was then prepared from the original silicon surface without any ion beam exposure and imaged in Figure 5.33.

Figure 5.33: Control sample's a) TEM image, b) STEM image and c) associated EDS map.

Looking at the TEM image, a similar 71 nm thick layer of polycrystalline silicon is observed on the surface above the monocrystalline silicon. Depending on the wafer processing history, different surface treatment on silicon backside can sometimes leave either gettering or polysilicon layers such as these to entrap contaminant ions. A STEM image was taken with the high angle annular dark field detector along with its associated EDS map which also shows a lack of implanted xenon in the absence of PFIB milling.

Comparing between sample 5 and the control, the observed polycrystalline layer can thus be understood as the remnants of a pre-existing layer. Assuming a uniform layer thickness across the die, the net 67 nm of amorphous and polycrystalline silicon thus implies that a mere 4 nm of material was sputtered off the sample surface with originally 71 nm of polycrystalline layer. The lower mill depth here thus imply a lower sputter rate using a 5 kV beam. Understandably, ions with low energies would result in smaller momentum transfer to displace surface atoms and thus lower sputter yield. As evident in the low amorphisation and implant depth of 14 nm, damage induced by a lower 5 kV beam here is firmly under that of a higher 10 kV beam. However, it should be noted here that due to the polycrystalline layer here, exact quantification of the amorphised layer could be obscured.

Subsequently, another window was then milled as sample 6 at a yet lower 3 kV beam. As seen in Figure 5.34, the actual mill depth appears superficial as does the windows.

Figure 5.34: a) Sample 6 SE image of window at 52° tilt and b) deposited marker.

Taking a TEM image across the interface, a similar composition of amorphous and polycrystalline silicon atop bulk silicon is observed in Figure 5.35. From the surface, the amorphous layer measures at 11 nm thick while the polycrystalline layer is 62 nm thick. With a combined thickness of 73 nm, it suggests that negligible material has been milled off the surface. Instead, only 11 nm of material at the surface has been fully amorphised due to the ion beam.

Figure 5.35: TEM image of sample 6 at various magnification.

Upon EDS mapping in Figure 5.36, the 11 nm amorphous layer is found to be implanted with xenon. Mirroring the control sample, oxygen was also found within this layer presumably as SiO_2 . As with sample 5, while the upper amorphous layer is induced by the ion beam milling, the lower polycrystalline layer is likely the remnants of the pre-existing layer. Noting the low mill depths stopping within the polycrystalline layer, amorphisation measurements might not be exactly accurate. However, as with the previous 5 kV beam milling, the observed 11 nm of amorphisation here with a 3 kV beam shows qualitatively that a lower beam energy induces lesser damage in silicon.

Figure 5.36: STEM images and EDS maps of sample 6.

Thus far, it has been shown that use of a lower beam energy can reduce the induced damage in silicon. This means that in theory, the milling process can be ended with a low kV polish for an optimal finish. In practice however, the lower beam energy presets are usually optimised for lower beam currents and to undo the prior induced damage, an amount corresponding to the amorphised layer thickness needs to be milled off. Considering the highest amorphisation thickness of 64 nm at 30 kV and the low milling rates of the low energy beams, the necessary polishing step can then take a prohibitively long time. Consequently, the time savings of milling with a high current at a high beam energy could be outstripped by the need for a time-consuming low kV polish. This thus suggests that for ion beam milling, the use of an intermediate beam energy of about 15 kV would be ideal to strike a balance between moderate milling rates and low amorphisation. Asides from

low kV milling however, many studies have also shown that gas assisted etching have benefits in minimizing beam induced damage on top of milling rate enhancement.

Finally, sample 7 was then etched under similar conditions with a 15 kV beam milling at 128 nA for 10 mins. The GIS was then inserted with XeF_2 dispensed over the surface during milling. As observed in Figure 5.37, a uniformly etched 200 μm window was observed as before and the necessary markers were deposited over a largely smooth surface.

Figure 5.37: a) Sample 7 SE image of window at 52° tilt and b) deposited marker.

Inspecting under the TEM in Figure 5.38, an amorphisation layer was similarly observed above the bulk silicon. However, two features differentiate it from a similar window milled without gas. First, the amorphous layer is much thinner at 15 nm for a 15 kV beam as opposed to 35 nm when milled without. Secondly, the layer formed is uniform and presumably without implantation.

Figure 5.38: TEM image of sample 7 at various magnification.

The absence of xenon implantation was then confirmed under EDS mapping as observed in Figure 5.39 with a noisy xenon elemental map. To account for the absence of implanted xenon and smaller amorphisation layer, discussion around the mechanism with which material is etched would be necessary. During ion beam milling, the primary mode of material removal lies in mechanical sputtering where momentum is transferred from the incident ions to the surface atoms in a collision cascade. Recoiled atoms can then either be sputtered off or induce further damage within the material. The ion will then eventually come to rest within the sample and remain as an implanted layer. During gas assisted etching however, the above-mentioned sputtering runs synergistically with chemical etching processes to remove material. As the gas is introduced over the surface, it chemisorbs to the silicon to form a non-volatile SiF_2 intermediate. The incident ion can then either sputter off this layer or induce recombination to form volatile SiF_4 that is pumped out of the system. As material is removed, fresh silicon surface is then exposed leading to further chemisorption and removal.

Figure 5.39: STEM images and EDS maps of sample 7.

Under such a mechanism, a larger volume of material is thus removed than with the ion beam milling or etching alone and lower dosages of ions can be used. As material is now removed by sputtering of surface SiF_2 or SiF_4 volatilisation, any implanted xenon ions are also removed simultaneously. As such, a 0 nm implantation depth and a low 15 nm amorphisation layer is observed here with chemical etching outstripping mechanical sputtering processes. With a thinner amorphous layer introduced than without XeF_2 , the use of GAE is thus a viable method of reducing ion beam damage where the window size permits uniform etching. To summarise, the measured lamella characteristics from all the samples prepared is shown in Table 5.3.

Table 5.3: Table of lamella characteristics.

	Beam Energy (kV)	Current (nA)	Surface roughness (nm)	Amorphous Si (nm)	Implantation depth (nm)	Polycrystalline Si (nm)
1	15	124	1	38	21	-
2	15	11.2	2	35	24	-
3	30	14.5	6	64	37	-
4	10	16.8	2	24	17	-
5	5	3.3	0	14	14	53
6	3	1.4	0	11	11	73
7	15	128	3	15	0	-

In examining the relationship between beam energy with its induced amorphised silicon and implantation depth, Figure 5.40 was then plotted. From the trend lines, higher beam energies used are shown to induce higher amorphisation layer thickness and implantation depths. In particular, the amorphisation layer thickness approximately doubles with the beam energy in nanometers as reported by other sources [25-28].

The implantation depth, however, appears to converge towards the amorphisation layer thickness and equalize at low beam energies. According to the model of incident ions recoiling silicon atoms into further amorphisation, the gap between the amorphisation thickness and the implantation depth is accounted for by the incident beam energy. This means that at low beam energies, the incident ions do not possess adequate energy to recoil silicon atoms into further displacement within the crystal lattice. Instead, the ions are merely implanted within the silicon at a high enough dose to disrupt the crystal lattice and amorphise locally.

Figure 5.40: Plot of amorphisation layer thickness and implantation depth versus beam energy.

To conclude, the use of the ion beam to trench silicon inherently leaves behind a damage layer due to sample-beam interactions. The damage is then manifested both as implanted xenon and as an amorphisation layer with a thickness roughly double that of the beam energy used in nanometers. When targeting low RSTs, caution thus needs to be paid so that active devices are not damaged by these interactions during milling. This then necessitates the use of low kV polishing steps to thin onto target or the use of gas assisted etching wherever the milling conditions facilitates uniform etching. Noting the low currents and sputter rates available to lower energy beams, a stepped process involving the use of progressively lower beam energies for milling is thus recommended to reduce induced damage under reasonable milling times.

5.4 Localised FIB Milling on Ultra-thinned Samples

The objective of this section is to study the use of FIB milling to localise further ultra-thinning for electron beam and shorter wavelength visible light probing. In the previous sections, the process was characterised for its milling rate as a function of dosage and for its induced damage relative to the beam energy used. In practice, a device is to be ultra-thinned mechanically prior to FIB milling. Subsequent FIB then involves the bulk removal of silicon at moderate beam energies and current based on calculated doses. If necessary, a final low beam energy milling or gas assisted etching step can also be used to polish the silicon onto target thicknesses while removing damage induced prior. In this section, the procedure is then put through the steps, and the challenges as well as limitations along the way are discussed.

5.4.1 Experimental Setup and Process

Depending on the intended probing technique, the silicon may need to be thinned down to STI levels or left at thicker sub-micron thicknesses. As such, two small packaged 14 nm devices with 12.3 mm x 8.8 mm dies were prepared for different intended thicknesses. The first is targeted to approximately 500 nm over a large 700 μm window for possible visible light probing applications. On the other hand, the second is thinned down to STI levels over a smaller 50 μm window for possible electron beam probing applications.

The first sample was taken through an overall workflow as illustrated in Figure 5.41 to ultra-thin onto 500 nm. It is first mechanically thinned to 5.0 μm at an arbitrary AOI by targeted contour milling techniques described in section 4.3. Subsequently, an RST map was collected over the die and a few markers were inked onto the surface to demarcate the AOI at 5.0 μm . The sample was then secured onto an aluminium sample stub with silver paste applied around and under the package before heating at 80°C to dry. It was then loaded into the Tescan Xeia3 PFIB system for silicon trenching. Subsequently, the sample was navigated to within the demarcated AOI under electron beam imaging. Due to the relatively high RST target exceeding the projected induced damage at 15 kV, the 15 kV

287 nA beam was used to thin all the way to target thickness. A large 700 μm x 700 μm window possible only at 15 kV and under was then specified for milling over the AOI. To thin onto a target 500 nm RST, approximately 4.5 μm depth of silicon thus needs to be sputtered off. Assuming the previously derived sputter rate of 0.38 $\mu\text{m}^3/\text{nC}$ at 15 kV, the current is then measured and an appropriate length of time is specified for milling. A RST map is then collected over the resulting window to confirm the final RST profile. The sample is also inspected under the LSM to check the same.

Figure 5.41: Flowchart of process to locally ultra-thin silicon to 500 nm by mechanical and FIB milling.

The second sample intended for thinning onto STI was similarly milled but to around 1.0 μm instead via global contour milling techniques described in section 3.4 and Figure 5.42.

A RST map was then collected over the die with no particular AOI specified across the die due to relatively uniform thicknesses. The sample was then secured onto an aluminium sample stub and copper taped across the package for electrical grounding. It was then loaded into the Thermo Fisher Scientific Helios G4 FX DualBeam FIB system for silicon trenching due to its superior electron beam imaging capabilities and the smaller windows necessary. With no particular AOI intended, the sample was then navigated to pristine sections under electron beam imaging for subsequent milling.

Figure 5.42: Flowchart of process to locally ultra-thin silicon to STI by mechanical and FIB milling.

First, the highest current 30 kV, 65 nA beam preset was loaded and a small 50 μm x 50 μm window was then specified for milling. The process was also monitored simultaneously

with the surface imaged intermittently every 10 s with a 10 kV electron beam to end point the milling to the STI. When nearing the target layer, a slight polish was also performed with a 16 kV beam at 10 nA with XeF₂ introduced.

To illustrate etching non-uniformity challenges encountered before, another window was etched with XeF₂ introduced throughout the entire process. The resulting surface was then inspected under the SEM and the challenges associated with ultra-thinning is discussed. To also highlight the effects of ion implantation, two 200 μm windows were milled under similar conditions but with one undergoing GAE with XeF₂ introduced. Subsequently, the two windows were then studied under the LSM for differences.

5.4.2 Results and Discussion

For the first sample, a 12.3 mm x 8.8 mm die was ultra-thinned to 5.0 μm at the bottom right quadrant with the procedure described in section 4.3. As shown in Figure 5.43 a), the sample was initially levelled about the AOI as shown in a) so that the AOI is at highest elevation. Subsequently, the silicon was grinded down with a diamond grinding disc to 70 μm before it was contour milled to about 50 μm with the initially collected profile. The resulting RST map collected was then edited to increase the thickness at the AOI as seen in Figure 5.43 b) to generate a map for milling. As seen in Figure 5.43 c), the generated map is similar to the initial profile but with a dimple over the specified AOI to target ultra-thinning there. Subsequent contour milling and polishing then localised the thinnest 5.0 μm of silicon to the AOI shown in Figure 5.43 d).

Figure 5.43: Maps of a) initial contour profile, b) edited RST profile after milling, c) generated contour profile, and d) final RST profile.

After dismantling the sample from the glass slide and cleaning off the wax, the sample was then affixed to an aluminum sample stub with copper tape and silver paste. With the help of the visible light spectrometer, the AOI with $5.0\ \mu\text{m}$ was then confirmed again, marked with a felt tipped pen and loaded into the PFIB for milling. Here, it is important to point out that the sample was chosen for its relatively minute dimensions to fit within the general-purpose FIB systems. As the sample was tilted to normal incidence versus the ion beam column for milling, larger samples might crash into other components such as the electron column pole piece in close proximity. With dedicated FIB systems, this should not present much of an issue with a singular ion column mounted normal to the sample as in a SEM. The AOI was then located based on the marker under electron beam imaging and a $700\ \mu\text{m} \times 700\ \mu\text{m}$ window was defined for milling. Based on the initial RST of 5.0

μm , thinning down to 500 nm would require a 4.5 μm mill depth, corresponding to approximately $2.2 \times 10^6 \mu\text{m}^3$ of silicon to be sputtered off. Working with a 15 kV beam and a measured current of 230 nA at previously calculated sputter rate of $0.35 \mu\text{m}^3/\text{s}$, a mill time of 25910 s or 7.2 hours was then specified to mill overnight.

After milling, a distinct 700 μm square window was then observed beside the marker under SEM imaging as shown in Figure 5.44. At intermediate magnification, the optical image shows a slight tint from the active circuitry below along with minor scratches within the window. As the product of incomplete polishing from prior CNC thinning, the scratches were subsequently propagated down with the silicon during milling. While there was still sufficient silicon to buffer possible damage from the scratch in this case, future applications should ensure a pristine finish before FIB milling as surface defects like this will propagate down as is. At higher magnification and brighter illumination, active circuitry and features such as discrete cache blocks can then be observed owing to the low sub-micron RST.

Figure 5.44: a) SE image of milled window beside marker, b) low mag image of window, and c) high mag image under brighter illumination.

To confirm the profile across the window after FIB milling, a RST map was then collected across the window and its peripheries as seen in Figure 5.45. While the edges display a couple of anomalous readings due to the imperfect alignment of the sample during measurement, thicknesses are found to generally range between 0.3 μm and 0.8 μm within the window. On the other hand, the thicknesses immediately outside the edge of the

window range between 4.9 μm and 5.4 μm . Looking at the RST variation across the area both within and outside the window shows a trend of decreasing thickness values from the bottom left to top right corner, with a distinct drop of about 4.6 μm within the window. This thus confirms that while the milling removed approximately 100 nm more silicon than intended, the removal is largely uniform across the window with the resulting RST variation of about 300 nm within approximating the 400 nm difference immediately outside the window. Also, while the original RST map shows the AOI uniform at 5.0 μm within a 700 μm area, the map here shows that the 5.0 μm area is offset to the top right. This thus suggests that the placement of the window is slightly off centered from the intended 5.0 μm window, instead being placed a little lower and to the left where the silicon is slightly thicker at 5.4 μm .

Figure 5.45: a) RST map across window, b) measurement at center, and c) physical representation.

The sample was then imaged under a laser scanning microscope with a 1340 nm laser as seen in Figure 5.46. At low magnification, concentric newton rings are observed emanating from a central spot due to targeted ultra-thinning at the AOI and thus RST variation. With reference to the central spot, the milled window is slightly offset to the bottom left rather

than over it. This then explains the 400 nm variation in RST within the window because the window placement was slightly off. Furthermore, the similar fringe patterns within the window to that outside suggests that RST variation remains constant after milling and that uniform amounts of material was sputtered. At higher magnification, the spotted patterns were also observed here due to the underlying bump topology. Finally, it is also notable that even within the window implanted with xenon, the collected signal remains the same as without, showing that xenon beam milling is likely compatible with optical probing.

Figure 5.46: LSM image taken at a) 5x, b) 10x, c) 20x, and d) 50x magnification.

In this experiment, PFIB milling has been demonstrated to controllably trench large windows of sub-micron silicon necessary for shorter wavelength visible light probing within reasonable time frames. Should smaller windows suffice, the higher beam energy and current presets accessible can be used to further reduce the time necessary for milling. However, due to beam induced damage and ion implantation, caution needs to be taken so as not to alter circuit functionality. As observed here, a pristine surface finish also needs to be ensured prior to milling so it does not get propagated down with milling,

Practically, while the exact placement of the window need not be as precise as for circuit edit applications owing to its size, *in-situ* infrared imaging capabilities would still be necessary for navigation. The circuit features can then be discerned for accurate placement of the milling window over the actual AOI. To accurately monitor the RST within the window, an in-chamber spectrometer for *in-situ* measurements would also prove invaluable. Further work outside the scope of this thesis would also be necessary to characterise prepared samples electrically to check for changes in behavior. As seen in the poor thermal stability of globally ultra-thinned samples previously, the viability of these samples for subsequent testing at much lower thickness also remains to be seen.

Moving onto sub-100 nm of thinned silicon for electron beam probing applications, smaller windows are now possible without the need to physically accommodate for SIL contact. With the low target thickness at the STI level, it then becomes necessary to precisely detect the endpoint of the milling process by imaging the milling window. With its superior imaging capabilities and the lower beam currents necessary for smaller window, the gallium FIB was then used to mill windows off sample 2 that had been globally ultra-thinned with methods described in section 3.4. As shown in Figure 5.47, the die was ultra-thinned globally to between 0.5 μm and 6.0 μm thick with the circuitry visible under illumination.

Figure 5.47: a) Final RST map, b) RST visualization, and c) optical image of ultra-thinned die.

The sample was then cleaned, affixed onto a sample stub, and loaded into the FIB for trenching. Next, the sample was navigated to the rough vicinity of the thinned part with 0.6 μm RST and a 50 μm x 50 μm window was specified. A 16 kV, 10 nA beam was then loaded for milling to reduce beam induced damage and to maintain controllability in end-pointing the process. With a target mill depth of 0.6 μm and an approximate sputter mill rate of 0.23 $\mu\text{m}^3/\text{s}$, a mill time of 652 s or 10.8 mins was used. Towards the end of the process after 9 mins, a 10 kV electron beam was then used to intermittently monitor the window in 10s intervals. The process was then continued until the silicon was thin enough so that circuitry was observable under electron beam imaging as shown in Figure 5.48 a). It is also worth noting here that ion beam secondary electron imaging is ideally used in place of high energy electron beam imaging for end pointing. Due to its lower interaction volume of the ion beam over the electron beam, the underlying active devices is less likely to be affected from the target STI layer.

Depending on the specific requirements for probing, this could be an acceptable end point for ultra-thinned silicon. However, GAE was then used to further polish the window down to STI by etching with a 16 kV, 4.4 nA beam under XeF_2 while monitoring the SE image in 1 s intervals. Upon exposure of the STI as shown in Figure 5.48 b) and c), the process was then stopped. At higher magnification, aside from the brighter contrast, the STI can also be identified from the individual blocks of active silicon observed. As noted in prior etching trials however, slight etching non-uniformity can also be observed here as well with some remnant silicon observed on the top left quadrant of the window.

Figure 5.48: SE images of a) thin silicon approaching STI, b) STI exposed as observed at 52°, c) STI exposed as observed at 0°, and d) STI imaged at higher magnification.

To further illustrate the differential etching rates, another 50 μm x 50 μm window in the vicinity was then thinned entirely through GAE. After initiating the process, the progress was monitored until the STI is observed to be fully exposed throughout the window as seen in Figure 5.49 a). Initially at silicon, a roughened window is observed in Figure 5.49 a)

with slight pitting along the peripheries due to spontaneous etch by the introduced XeF_2 gas. Through extended etching until the STI is fully exposed as in Figure 5.49 b), the pitting becomes more pronounced due to lengthened exposure. Also notably, the etching took about 2 minutes till the exposure of the STI, representing an approximate 5 times increase in milling rate over ion beam milling. Subsequently, a gradient between the bottom right corner with STI exposed to the top left corner with remnant silicon is observed. At higher magnification, while the STI is just barely exposed by the top left corner in Figure 5.49 c), there is over-etch observed by the bottom right corner in in Figure 5.49 d). This is observed with the parallel fins being etched off into trenches within the individual islands. While this is an undesirable outcome in maintaining device functionality, it highlights the rate enhancement and selectivity of the GAE process in etching the silicon fin over the oxide isolation around it.

Figure 5.48: SE images of a) window with silicon remaining, b) STI partially exposed, c) STI intact at top left corner, and d) STI over-etched in bottom right corner of window.

Between these windows, the previous one where GAE was used for slight polish and the windows milled for rate characterisation, a faster etch rate is observed at the bottom right corner of the window over the opposite edge. This is attributed to the positioning of the gas nozzle as illustrated previously in Figure 5.5, delivering a higher gas pressure closer to the nozzle than away. As with prior data on milling rate, this thus suggests that GAE is not ideal for silicon trenching through thick silicon and over large areas as the GIS configuration is unoptimised.

However, as previous data on ion beam induced damage suggests, ion beam milling alone is not suitable for thinning onto STI due to silicon amorphisation and ion implantation. This is especially so for gallium milling here as opposed to xenon, the implanted metallic gallium ions are not only likely to alter device operations but will also affect signal transmission through silicon. This is exemplified with two large $200\text{ }\mu\text{m} \times 200\text{ }\mu\text{m}$ windows milled at 16 kV and 10 nA over 5 mins on the same die. While one of the windows was left to mill with the ion beam alone, the second was etched with XeF_2 under the same conditions. LSM images were then taken over both windows as shown in Figure 5.50.

Figure 5.50: a) SE image of both windows, and LSM images of b) both windows at 20x, c) milled window at 50x, and d) etched window at 50x.

From the SE image shown in Figure 5.50 a), the left window was milled with the ion beam alone while the right window was etched with XeF_2 . While both were milled for the same length of time, the gas etched window on the right is visibly deeper with its bottom right corner almost exposed. Imaging the two windows under the LSM, the newtons rings are remarkably absent from within the milled window but present within the etched window. At higher magnification, it is also obvious that the circuit features are obscured within the milled window relative to that outside and is indicative that the laser signal is somewhat absorbed or reflected by the implanted gallium ions. On the other hand, the contrast within the etched window remains similar to that outside. Unlike the window milled with the ion

beam alone, this is likely attributed to a lack of implanted ions when using GAE. However, as the vertical fringes within the window show, GAE thins the silicon non-uniformly between the left and right edges as prior data suggests.

In the experiments with the gallium FIB, milling in conjunction with GAE for final polish has been demonstrated to trench small windows down to STI levels for potential electron beam probing applications. While GAE has been shown to enhance milling rates and to avoid the effects of ion implantation, its use is limited to small windows by non-uniform etching due to an unoptimised GIS configuration.

5.4.3 Comparison Between Different Milling Processes

In this section, the various techniques presented thus far were used to demonstrate sample ultra-thinning to sub-micron thicknesses for high resolution backside probing. The techniques presented can broadly be categorized into mechanical milling and FIB milling with its associated advantages and disadvantages to be discussed.

For mechanical CNC milling, the greatest advantage lies in the ability to remove large amounts of material in a relatively short time. Thinning a typical 300 mm^2 device globally from full thicknesses of $700\text{ }\mu\text{m}$ down to $5\text{ }\mu\text{m}$, equivalent to 208.5 mm^3 , takes about 5 hours of process time. Accounting for bulk silicon grinding, contour milling and polishing to the mirror finish required, this translates to a material removal rate of approximately $0.70\text{ mm}^3/\text{min}$. Another benefit is the relatively lower cost of ownership and consumables for a CNC miller. Combined with its higher throughput, the cost of processing a single sample is drastically lower than that of using a multi-million FIB system for the sole purpose of sample thinning. Most importantly, the mature CNC milling techniques are also time tested and come with lower risks of sample damage compared to other chemical etching techniques not discussed. Because the milling processes are isolated to the die, it does not damage any other components on the device like the chip capacitor and solder balls.

However, while a high removal rate is advantageous, its removal and end-pointing precision is limited when thinning onto ultra-thin samples. In the final steps of mechanical thinning, polishing is necessary to obtain a mirror finish for SIL contact and optimal signal collection. Although the amount of material removed during polishing can largely be controlled through the parameters used, slight variability in the realm of microns is still inevitable. Thus, it cannot be used for ultra-thinning onto sub-micron thicknesses for the risk of over polishing and sample destruction due to process variability during polishing.

To address challenges in ultra-thinning due to polishing variation, FIB milling was then presented as a complementary technique for final thinning onto sub-micron thicknesses. After characterizing the mill rates for the different systems, FIB milling was found to be consistent in the amount of material removed based on the sputter rate of the individual system. With an estimated sputter rate of $0.23 \mu\text{m}^3/\text{nC}$ on the gallium FIB and $0.40 \mu\text{m}^3/\text{nC}$ on the xenon PFIB at 30 kV, precise volumes of material can be removed locally over the area of interest by specifying appropriate parameters of area, current and time. For applications that require thinning onto STI levels, the milling process can also be safely end-pointed through *in-situ* SEM imaging.

As a complementary technique to mechanical milling, FIB milling is limited to much lower ranges of mill rates. FIB mill rates range from between a maximum of about $0.90 \times 10^{-6} \text{ mm}^3/\text{min}$ for a 30 kV gallium beam at 65 nA to a maximum of about $0.48 \times 10^{-4} \text{ mm}^3/\text{min}$ for a 30 kV xenon beam at 2 μA . On top of the limited milling area in a FIB system, the milling rates also lag the $0.70 \text{ mm}^3/\text{min}$ of mechanical techniques by between 4 to 6 orders of magnitude. Accounting for the differences and complementary range of material removal in Table 5.4, mechanical milling can be used to globally pre-thin the sample before final thinning to target in the FIB.

Table 5.4: Comparison between mechanical and FIB milling

	Mill rate (mm ³ /min)	Mill area (μm)	Throughput	Cost	Precision range
Mechanical	0.70	unlimited	High	Low	μm
FIB	(Ga) 0.90 x 10 ⁻⁶	400 μm @ 30 kV	Low	High	nm
	(Xe) 0.48 x 10 ⁻⁴	400 μm @ 30 kV, 700 μm @ 15 kV	Mid		

For practical applications, it is also worth noting that the choice of techniques used in the milling processes is contingent upon several considerations. Primarily, this depends on the sample requirements that are in turn dependent on the device type and fault isolation technique necessary. For devices fabricated using older process technologies, resolution requirements are more lenient and may only require IR beam probes with 50 μm samples thinned mechanically. On the other hand, devices fabricated on cutting edge process technologies will necessitate higher resolution through visible light probing. With longer wavelength, visible light lasers require between 5 - 10 μm thinned samples, mechanical thinning techniques can still be pushed to its limits. Should future resolution needs push beam probes onto shorter wavelength or even electron beams, the thinner samples needed would then require FIB milling to avoid overmilling and sample destruction. Furthermore, mechanical pre-thinning for thinner samples will also need to be targeted so as to maintain the thermal stability of the device during testing.

Other considerations such as the size of the region of interest, device type, device size and warpage also inform the choice of techniques to be used. For mechanical milling, a larger region of interest would require greater silicon uniformity and limit the use of targeted contour milling techniques. For FIB trenching, larger regions of interest and windows needed may necessitate the use of the PFIB over the FIB. On the device type, a higher power draw and thus heat generation may require that targeted ultra-thinning be performed for improved thermal stability. On the contrary, global ultra-thinning could be viable for low power devices without thermal runaway risks. Finally, larger devices with larger warpage usually come with poorer uniformity and thus higher over polishing risk during global contour milling. This then limits it to targeted contour milling techniques.

5.5 Conclusion

In this chapter, FIB milling was studied as a controllable and precise method to further ultra-thin silicon locally for yet higher resolution probing techniques. First, the milling rate was characterised across both gallium FIB and xenon PFIB systems for precise material removal. Sputter rates at known beam energies were then identified for precise material removal during trenching. Subsequently, the effects of xenon ion beam milling on the sample was then studied through from samples of the milled surfaces. TEM analysis then revealed higher damage and ion implantation induced by the ion beam at higher beam energies. With both mill rate and finishing structure characterised, a workflow was then proposed and validated on ultra-thin samples prepared as per earlier chapters. Challenges and limitations with the process and instrumentation were then discussed. The advantages and disadvantages of the different milling processes were then discussed in relation to the sample preparation requirements.

Chapter 6

Conclusion and Recommendations

The objective of this work is to develop a controllable and repeatable process with which silicon can be ultra-thinned for high resolution backside probing. In this chapter, all the work undertaken is concluded in relation to the beforementioned objective. Subsequently, recommendations are presented for future work that would further efforts for high resolution backside probing.

6.1 Conclusion

The motivations behind this work is first presented in relation to the aggressive scaling of devices and the accompanying need for high resolution fault isolation techniques. In chapter 2, these visible light probing techniques were discussed along with its requirements in relation to the final remaining silicon thicknesses. Methods of silicon thinning including CNC milling techniques were reviewed alongside anticipated sample warpage behavior. FIB milling techniques were also examined across different ion sources for their impact on the sample.

In chapter 3, global contour milling techniques were studied and sample relaxation during milling was identified as a key challenge. By the characterisation of sample relaxation behavior and subsequent modifications to the workflow, the technique was then successfully adapted for global ultra-thinning. However, subsequent electrical testing with the prepared samples highlighted challenges associated with the thermal stability of globally ultra-thinned samples.

Subsequently in chapter 4, different methods of targeted ultra-thinning to 5.0 μm for visible light probing were then studied to improve the prepared devices' thermal stability. This

was achieved in a 2D fashion through sample tilt with samples yielding a larger thickness differential for heat dissipation but poorer local thickness uniformity. On the other hand, 3D methods of targeted contour milling were also developed and tested on actual devices. Here, greater local uniformity was achieved at the expense of a poorer thickness differential instead. In either case, targeted ultra-thinning to localise 5.0 μm of silicon at AOI was shown to meet the needs for visible light probing while maintaining thermal stability.

In anticipation of higher resolution probing techniques that need yet thinner remaining silicon thicknesses, FIB milling was then studied as a controllable and precise method in chapter 5. First, the milling rate was characterised across different systems for precise material removal. With the low target thicknesses, the effects of the beam on the sample was then studied through TEM inspection of the milled surfaces. It was found that the milling process induces an amorphisation layer in the otherwise monocrystalline silicon, with higher beam energies inducing thicker layers. As a rule of thumb, the amorphisation layer is in nanometers, roughly double the used beam energy in kilovolts. With both the mill rate and finishing structure characterised, a workflow was then proposed and validated on ultra-thin samples prepared as per prior chapters. Challenges and limitations with the process and instrumentation were then discussed. Chiefly, this includes the etching non-uniformity when GIS is used for large area trenching, with the area closest to the needle being etched quicker than the areas further away.

6.2 Recommendations for Future Work

In this work, CNC contour milling was chiefly studied as a means to rapidly ultrathin silicon due to its prevalence and suitability in the industry today. However, several other techniques including adaptive milling and polishing with specialized equipment could be studied to better localise ultra-thinning. As encountered in targeted 3D contour milling experiments, introduced silicon non-uniformity were often undone though polishing. Through the ability to modulate polishing at different position with adaptive thinning, ultra-thin silicon can be better localised for greater thermal stability.

Currently, the localisation of ultra-thin is also maintained arbitrarily to improve thermal stability. Should further detailed work be conducted to characterise heat dissipation through the sample, the necessary difference in remaining silicon thickness can be derived to better guide the thinning process. Furthermore, appropriate heat map simulations of the device during testing can also set a lower limit on the thinnest silicon possible so as to avoid destruction of the actual device before it happens.

In later sections, methods were presented for further localised ultra-thinning of samples to under 5.0 μm for shorter wavelength visible light probing and beyond. While the samples thinned to 5.0 μm were subsequently validated through electrical testing and probing with currently available hardware, testing on those thinned even further were limited by unavailability of the necessary equipment. This included SILs intended for shorter wavelength visible light probing and electron beam probing. Should the necessary hardware become available, further work on testing the samples prepared here would be invaluable in enabling yet greater resolution in defect localisation. In the same vein, further work in characterising device alterations due to artifacts induced by the FIB milling would also be essential to further fine tune the process for practical applications.

Finally, it is believed that should the necessary specialized hardware be available, further work on FIB milling can better localise ultrathin silicon. The ideal FIB system for such an application would resemble those for circuit edit with a singular FIB column mounted normal to the sample for greater ion beam resolution and to accommodate large devices, The availability of infra-red cameras along with in-chamber visible light spectrometers would then enable accurate navigation of the sample to the AOI as well as *in-situ* RST measurements after milling. Additionally, the configuration of a coaxial GIS nozzle can then allow for uniform gas flow during GAE, allowing quick large window etching, surface polishing and circumvent challenges encounter in this work with GAE.