X-ray Lithographic Alignment and Overlay Applied to Double-Gate MOSFET Fabrication

by

Mitchell W. Meinhold

B.S. Electrical Engineering
Lehigh University, 1994

B.S. Engineering Physics
Lehigh University, 1994

M.S. Electrical Engineering
MIT, 1996

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2003

Author.....
Department of Electrical Engineering and Computer Science
May 1, 2003

Certified by .....
.....
Henry I. Smith
Keithley Professor of Electrical Engineering
Thesis Supervisor

Accepted by.....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

X-ray Lithographic Alignment and Overlay Applied to Double-Gate MOSFET Fabrication

by

Mitchell W. Meinhold

Submitted to the department of electrical engineering
on May 1, 2003 in Partial Fulfillment of the
Requirements For the Degree of Doctor of Philosophy

Abstract

Double-gate MOSFETs represent a significant solution to transistor scaling problems and promise a dramatic improvement in both performance and power consumption. In this work, a planar lithographic process is presented that is capable of producing double-gate MOSFET (DGFET) gate structures with 50 nm physical gate length and <5 nm alignment between upper and lower gates. Because a self-aligned approach is not taken, the central challenge in fabrication is to define each gate in separate lithographic steps with precision alignment of upper to lower-gate masks. In order to obtain optimum device performance, the position of the lower-gate should be aligned to the upper-gate to better than 10% of the gate length. The gates are defined using X-ray lithography (a close-proximity shadow printing scheme). The associated alignment scheme, Interferometric Broad Band Imaging (IBBI), has been proven to yield nanometer level sensitivity. While the IBBI alignment system offers superior alignment detectivity, it must be complemented by comparably successful mask pattern placement in order to yield structure details within the desired 5 nm tolerances. This work addresses the details of a novel mask design and fabrication scheme as well as its incorporation into the process flow of the DGFET. Additionally, the

parasitic effects of strain that result from wafer bonding and thermal effects have been measured and analyzed.

Thesis Supervisor: Henry I. Smith
Title: Keithley Professor of Electrical Engineering

Acknowledgments

I would like to sincerely thank Hank Smith and Dimitri Antoniadis for providing me with the opportunity to work with them, for sharing their resources, guidance and support and for their patience as I worked through the difficult challenges that this project presented. I thank Jongwan Jung for the great deal of time and effort he put into wafer processing and for sharing his expertise with me. I thank Euclid Moon for his diligent work on the IBBI alignment system as it was upgraded over the years to meet higher levels of precision. Thanks to Mark Mondol for assistance with e-beam writing and for keeping the machine up and running. Thanks to Jim Daley and Jimmy Carter for many evaporations, filling lab supply requests, repairs and countless other forms of assistance. Thanks to Anthony Lochtefeld for his work on wafer processing and discussion that expanded my understanding of the devices. And I am of course very grateful to all the students and staff in NSL and MTL, past and present, who have assisted me with this work – I will not forget your kindness.

On a personal note, I express deep gratitude towards my mother, Barbara Stevenson, my father, Walter Meinhold, and my brother, Derrick Meinhold, all of whom fostered tremendous support over the past 8 years during my career at MIT. And finally, I thank Kristy for her loving support as I worked through the final and challenging phases of my project.

Contents

Table of Figures

1 Introduction

The uniqueness of this project arises not only from its implementation of X-ray lithography in double-gate MOSFET (DGFET) fabrication, but also in the fact that a specialized alignment system is used when patterning both the wafers *and* the X-ray masks themselves. The central goal is to present a manufacturable process for the fabrication of DGFETs with patterned gate lengths of 50 nm and alignment to 10% of the gate length – 5 nm. It is shown here that such tolerances are achievable with the exception that wafer strain must be eliminated in order to maintain overlay between the lithography tool and the wafers during processing. Strain disrupts the overlay during lithographic patterning and must be mitigated to attain device yield over larger die areas.

1.1 Overview of MOSFET Scaling and Advanced Devices

Demand for high performance and high density integrated circuits has driven MOSFET scaling to nanometer regimes. As this scaling occurs, circuit requirements demand long channel behavior while parasitic short channel effects are minimized. In addition, high current drive requires thinner gate oxides in short channel devices. This section provides a summary of the effects of scaling, the limits imposed by scaling and how the DGFET, a novel approach, speaks to these demands.

With the drive to make devices that occupy less space, consume less power, and operate at faster speeds, the past thirty years has seen a shrinkage of feature sizes by about 2 orders of magnitude. Reducing power consumption is key not only in reducing costs but in enabling minimal battery power consumption.

Shrinkage of MOSFET dimensions results in short channels. By one definition, short channels exist when the gate length is on the order of the length of the depletion region of the source and drain junctions. Alternatively, it can be defined as when the effective channel length is the same length as the source and drain implant depth [**1**]. In this regime, several detrimental effects come into play.

1.1.1 Velocity Saturation

As gate length decreases, fields in the channel perpendicular to the gate (parallel to the channel) increase for a given source-drain potential. When the field reaches 10^5 V/cm, the electron velocity saturates at 10^7 cm/s. This governs the maximum electron drift velocity which in turn saturates the drain-source current. Thus, for shorter gate lengths, the maximum source-drain current becomes a function of gate length rather than gate-source voltage [**1**].

1.1.2 Threshold Voltage Reduction (V_t Roll-off)

Threshold voltage (V_t) roll-off is one of the most serious consequences of device scaling [**2**]. As the effective channel length is decreased below approximately 0.2 μm, the threshold voltage begins to tend towards zero [**3**]. With smaller threshold voltages, the sub-threshold current (the off-state current) increases. Thus, there is a tradeoff between higher performance and lower power consumption.

Simple models that predict the dependency of V_t on gate length relate the two based on the skewed geometrical profile of charge beneath the gate. Drain Induced Barrier Lowering (DIBL) introduced in 1979 by Troutman [**4**], further refines this model. With increasing potential between source and drain, the depletion region of the drain reduces the electron energy barrier of the source making it easier for electrons to be injected into the channel. This degrades the ability of the gate to control the channel current.

1.1.3 Mobility

Mobility is the ratio of the speed of a carrier to the electric field. In the channel, there are two fields present, one parallel and one perpendicular to the gate oxide interface. As the perpendicular (or vertical) field increases, there is greater scattering of electrons near the Si/SiO_2 interface which in turn degrades the mobility. Since short channels induce greater vertical fields, mobility is degraded [**4**]. Mobility is also degraded when doping concentrations are high.

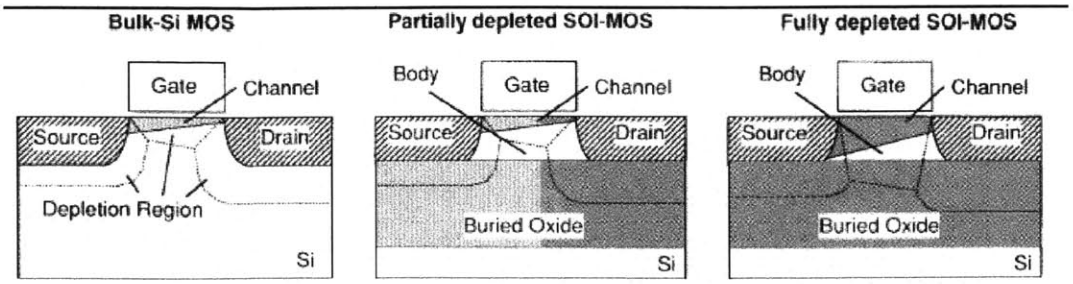
1.1.4 Silicon on Insulator (SOI)

SOI provides excellent performance at very short channel lengths. SOI wafers consist of a single-crystal film of Si with an oxide layer beneath it. SIMOX

(Separation by IMplanted Oxygen) involves implanting oxygen to a certain depth beneath the wafer surface and then annealing to form a film of SiO_2. BESOI (Bonded and Etchedback SOI) is another type of SOI in which a Si film is bonded to an oxidized wafer. The Si film can be isolated from its starting wafer with either an etch stop or a layer of implanted hydrogen. In both cases, the starting wafer must be removed to expose the Si film after bonding. If an etch stop is used, the original wafer is etched to the "stop" layer, and the Si film is exposed. If the hydrogen implantation method is used, the bonded wafer pair is heated to activate the hydrogen thus forming a pocket of gas which breaks the Si away from the rest of the original wafer.

The SOI doping profile and thickness can be adjusted to form either fully depleted (FD) or partially depleted (PD) channels. For FD devices, the entire channel thickness is depleted of majority carriers. For PD devices, there is a region of Si between the source and the drain that is not depleted. This area can be contacted or left floating depending on the desired device characteristics. If left floating, problems known as floating body effects (FBE) result in degraded device performance. Table 1, taken from the work of Eimori, et al., summarizes some of the pros and cons between bulk, PD and FD devices.

TABLE III
FEATURES OF SOI MOSFET's



Items		Bulk-Si MOS	Partially depleted SOI-MOS		Fully depleted SOI-MOS	
			Body-floating	Body-tied	Body-floating	Body-tied
(1) Vth controllability		Good	Good (but not small variation)	Good	Not Good (Gate material work function)	
(2) Back -bias effect	(i)Si-sub bias	Large	Small		Medium	
	(ii)SOI film bias (=Body bias)	——	Large		Small	
(3)Break down voltage		Good	Small	Good	Small	
(4) Current drivability		Standard	Almost same as the Bulk-Si MOS		Larger then the Bulk-Si MOS	
(5) S-factor		Standard	Almost same as the Bulk-Si MOS		Smaller then the Bulk-Si MOS	

Table 1 Comparison between bulk-Si, PD-SOI and FD-SOI devices. [5]

Reports of 25% and higher speed improvements of SOI relative to bulk-Si have been reported [**6**]. One of the main sources of improvement comes from the decrease in parasitic source and drain junction capacitance for SOI devices. This is because the lower surfaces of the depletion regions terminate at the buried oxide layer. In bulk devices, since high substrate doping ($5x10^{18}$) is required when effective channel lengths are reduced to 50 nm and below, leakage tunneling current occurs from drain to substrate. The lower doping concentrations available in SOI devices eliminate this problem.

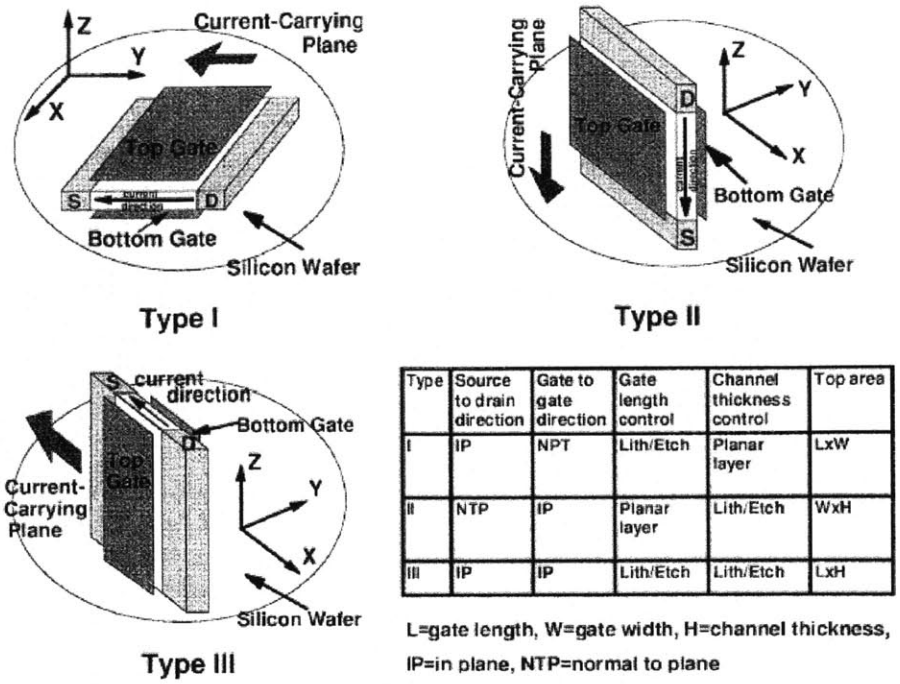
The existence of the buried oxide also prevents the presence of electric field lines that would otherwise begin on the gate and terminate in the body. This reduction of vertical field improves the problem of mobility reduction in the channel due to scattering at the Si/SiO_2 interface.

1.1.5 The Double-gate

The double-gate transistor was demonstrated as early as 1987 [**7**]. By adding a lower-gate to the SOI transistor a second channel is introduced at the

lower surface of the SOI. Figure 1 shows three possible orientations for the DGFET channel, each requiring a unique fabrication approach [**8**]. In type I, the channel thickness is defined by the SOI film thickness, and since it does not depend on lithography, it has the main advantage of best thickness uniformity. Type II and III have an advantage of higher device packing density, but they pay a penalty in performance and suffer from a dependence on uniformity of lithography and etching.

In addition to electric field screening of the second gate which reduces short channel effects, simulations predict an effect called *volume inversion* [**7**]. In the case in which the SOI is thinner than the depletion widths induced by the two gate voltages, a greater portion of the cross-section area of the SOI is inverted. Since more participating carriers are further from the Si/SiO_2 interface, less scattering takes place and effective mobility is increased leading to greater current drive. In the nominal case, current drive increases by a factor of 2. Table 2 provides a summary of strengths and weaknesses of the various families of SOI-based devices [**8**]. Alignment of the gates is critical. Wong showed that a 25% misalignment led to a 33% increase in delay [**9**].



Type	Source to drain direction	Gate to gate direction	Gate length control	Channel thickness control	Top area
I	IP	NPT	Lith/Etch	Planar layer	LxW
II	NTP	IP	Planar layer	Lith/Etch	WxH
III	IP	IP	Lith/Etch	Lith/Etch	LxH

L=gate length, W=gate width, H=channel thickness,
IP=in plane, NTP=normal to plane

Figure 1 Taken from Wong [8]. Process topologies for the DGFET.

Variant	Strengths	Weaknesses
Partially depleted SOI	1. Channel design bulk-like 2. VT insensitive to BOX interface	1. Very susceptible to floating body effects (but solutions are available) 2. Same scaling constraints as bulk
Fully depleted SOI	1. Elimination of floating body effects 2. Elimination of punch-through currents 3. Elimination of drain-body tunneling	1. VT sensitive to SOI thickness and back interface 2. Back-channel potential may be influenced by drain voltage 3. Difficulty of contacting thin SOI
Ground Plane (GP)	1. Same as FD SOI 2. GP shields channel from drain 3. GP permits electrical control of VT 4. GP may be used as second gate	1. VT sensitive to SOI thickness 2. Difficulty of contacting thin SOI 3. Degradation of subthreshold slope by close GP
Double Gate (DG)	1. Maximum electrostatic control of channel and best scaling potential 2. Best current drive and performance 3. OR logic function within single device	1. Difficult to fabricate 2. Mis-aligned top and bottom gates result in extra capacitance and loss of current drive 3. VT control difficult by conventional means
Stacked SOI (ST)	1. High functional density 2. Shorter wires therefore higher performance and lower power	1. Fabrication complexity 2. Difficult to cool

Table 2 Taken from Wong [8]. Evolution of SOI devices; strengths and weaknesses.

1.2 DGFET Process Overview

A typical device cross-section is shown in Figure 2. Essentially, two X-ray masks are used, in turn, to print lower and upper gate patterns onto a Si channel layer (see Figure 4). The key to visualizing this approach is to recognize that the gate lithography is performed on both surfaces of a thin Si film from which the device channel is built (20-40 nm of single crystalline Si). Obviously, these steps occur at two separate points in the device process between which the channel material is flipped with respect to the substrate. Since the gates are to be constructed successively, precise alignment and overlay is necessary when patterning the upper gates with respect to the lower gates. In addition to possessing gate patterns, both masks must contain alignment marks. The relative placement of the alignment marks and gate features on the X-ray masks is critical. However, in the case of our system, alignment is determined by a phase reading that changes monotonically with displacement. Therefore, it is not the absolute position of the alignment marks that is critical, but rather the position at which two superimposed marks read zero phase. An easy way to describe these overlay conditions is to consider the following thought experiment. Begin by bringing the two X-ray masks into face-to-face contact. Next, "slide" the masks with respect to one another until the alignment marks read zero phase. With the masks aligned in this fashion, check to see that for every pattern on the first mask there is an identical pattern on the second that is in exactly the same position. (This concept will be elucidated in the next few sections.) The alignment system itself is a rather sophisticated scheme called *interferometric broad band imaging* (IBBI). It is capable of sub-1 nm detectivity of position in X and Y.

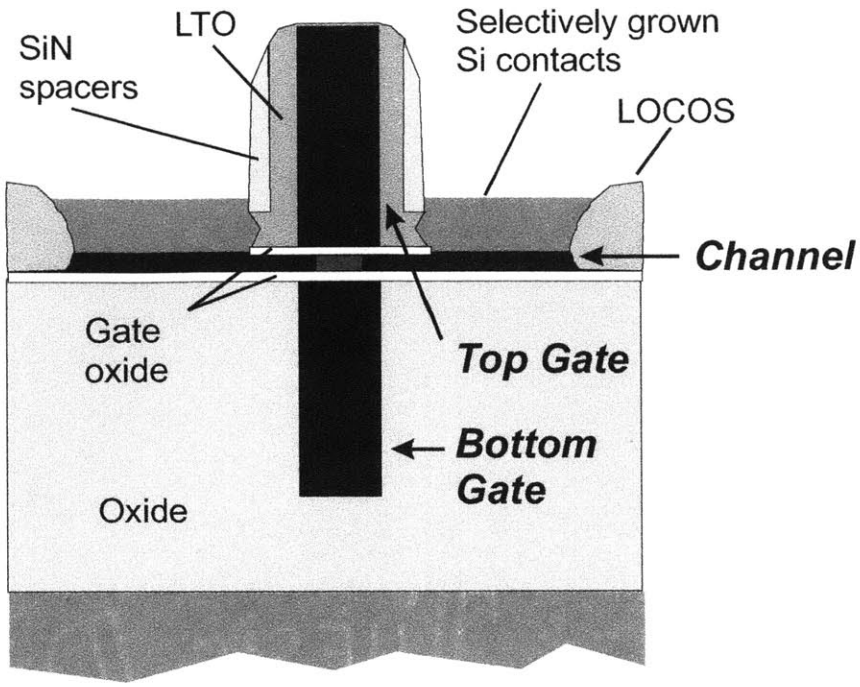


Figure 2 DGFET device cross-section.

Since the X-ray mask patterns match when the masks are placed face-to-face, two distinct X-ray masks are required (rather than attempting to use one mask to define both the upper and lower gates[1]). *Overlay* is the condition that partnering gate features (one from each mask) are located at the same coordinates of each mask surface. There are six degrees of freedom for the relative spatial position of two planar surfaces: X, Y, angle about the Z axis ϕ, the average gap (Z) between the surfaces and the two angles, α and β that bring the surfaces out of parallel. This is shown in Figure 3. In these experiments, leveling (α and β) and gap (Z) are set prior to exposure while X, Y and ϕ are continuously monitored and corrected throughout X-ray exposure using IBBI alignment feedback.

[1] Specifically, the patterns on one mask are "right-handed" while those on the opposing mask are "left-handed". Bringing the two masks together face to face is akin to bringing two hands together, palm to palm.

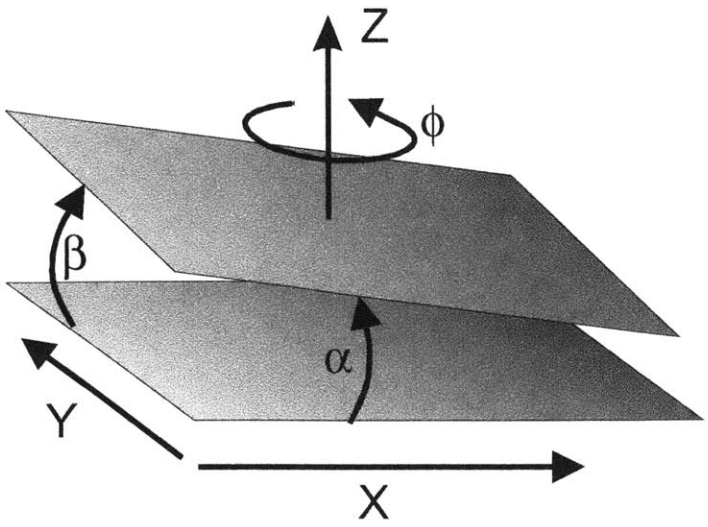


Figure 3 Depiction of mask-sample surfaces (or mask-mask surfaces) that shows degrees of positional freedom.

1.3 Mask Requirements

The purpose of this section is to summarize the lithographic requirements in a double-gate fabrication scheme where direct alignment is used to ensure accurate placement of upper gates with respect to lower gates. Mask overlay, alignment precision and in-plane distortion effects are the three conditions necessary to achieve pattern placement of a given tolerance over a particular area.

1.3.1 Mask Overlay

As mentioned, in order to ensure precise pattern placement, partnering gate patterns on upper and lower mask must match in position when masks are in face-to-face contact. This can be seen in Figure 4 where a cross-section of the Si channel is shown as free-standing material that is being patterned consecutively by lower and upper gate masks. (This example is simplified to one dimension.) Given a spatial relationship L between gate patterns and the effective center of the alignment marks on the lower gate mask, it is clear that the same spatial relationship (given a tolerance of say 5 nm) must exist on the upper gate mask.[2] As shown in the figure, the lower gate alignment marks are built from the same material as the lower gates – typically 150 nm polysilicon.

[2] This analysis neglects magnification effects that will be addressed later.

In this project, mask pattern overlay is obtained by replicating the lower gate X-ray mask from the upper gate mask. Essentially, the upper gate patterns are "contact printed" onto the lower gate mask. This replication scheme will be detailed in the chapter on mask processes.

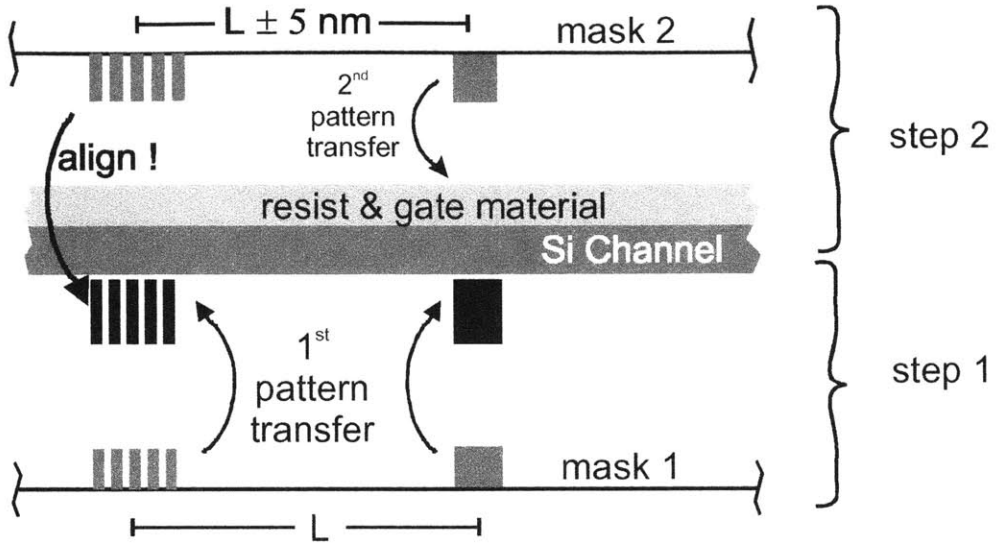


Figure 4 Depiction of lower (step 1) and upper (step 2) gate patterning steps where the Si channel is conceptualized as free standing.

1.3.2 Magnification and Distortion Errors

Magnification error is a form of pattern distortion that arises from the presence of a gap between the X-ray mask membrane and the sample surface. The X-ray source used in these experiments is a point source that emits radiation radially within a given solid angle.

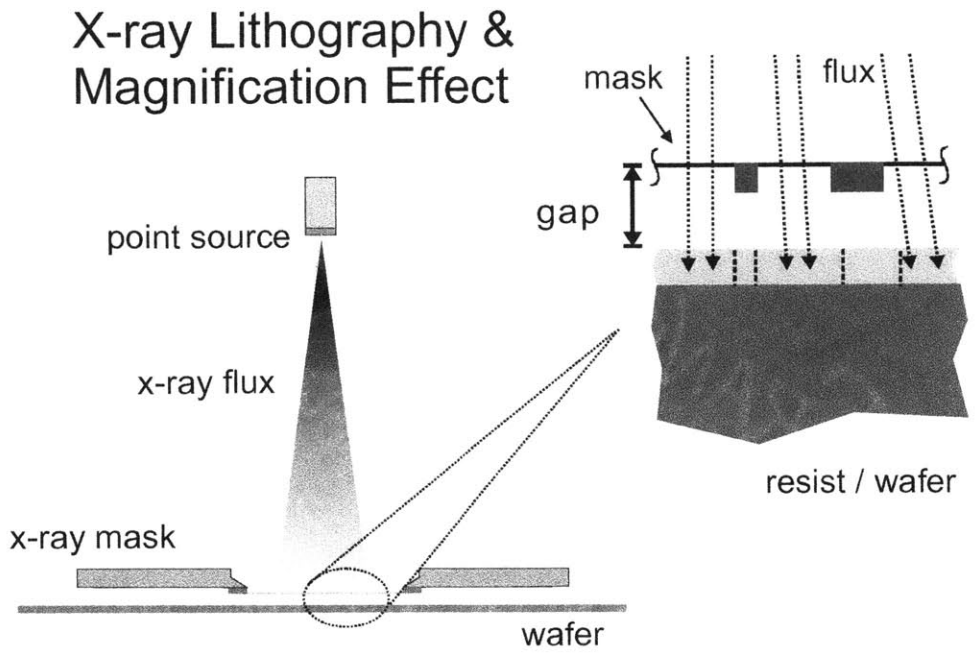


Figure 5 X-ray lithography scheme and depiction of the magnification effect from point source of X-ray emission. The shadow image cast through a flat membrane from such a source results in a pure magnification (enlargement) of the mask patterns.

Pattern placement errors can arise from in-plane distortion of the X-ray masks and strain of films on the Si wafer. The addition or removal of various films on the mask membranes or device wafer can alter the stress of the material causing expansion or contraction of the patterns. The high-temperature deposition and annealing processes that the wafer undergoes are other factors that must be addressed.

Chapter 2 explains the alignment system along with the X-ray source alignment scheme. Chapter 3 describes the device process flow with emphasis on the gate lithography steps. Chapter 4 details the X-ray mask fabrication process. Chapter 5 presents the results of the X-ray lithography as it is applied to the double-gate process. And Chapter 6 addresses the presence of wafer strain and introduces a moiré technique, similar to IBBI, that was used to quantify wafer strain resulting from the DGFET process.

2 The Alignment System

2.1 IBBI

The purpose of this section is to provide a basic understanding of the alignment system. For a more detailed description of the system and technical specifications not addressed here, please refer to reference [**10**].

Once again, IBBI detects the interference between gratings on the mask and the sample. Through spatially coherent illumination, interference of these alignment marks produces a set of dark and light fringes whose spatial position correlates to the relative position of the marks. The marks are designed such that small lateral displacements of the gratings lead to large displacements of the fringes, thereby providing highly sensitive detectivity of relative position.

Figure 6 shows the basic IBBI setup. A broad-band laser source (usually containing about 5 peak wavelengths) is fed into the optics system (a beam-splitter and lenses) via a fiber. The beam is steered toward the mask/wafer alignment marks by a mirror mounted just outside the anterior lens. The beam is collimated such that its cross-sectional area matches that of the alignment marks (approximately 400 μm square)[1]. Periodicity of the alignment mark on the wafer causes a 1st order diffracted signal that matches the angle of the input illumination. This back-diffracted signal is encoded with the interference between the marks. It is imaged by lenses in the optics tube onto the CCD camera. The fringe images are analyzed by computer which can then correct relative position and/or rotation by moving either the mask or the wafer.

A typical IBBI mark consists of two adjacent regions each of which produces a set of fringes (see Figure 7). Displacements between the mask and wafer cause the fringes to shift. For a given displacement of the mask, the fringes on one side of the mark move in opposition to those on the other side of the mark. If p_1 is the grating pitch on the mask and p_2 is the pitch on the sample, then fringes will move in opposition to the motion of the mask if p_1 is larger than

p_2. Conversely, if p_1 is less than p_2, fringes will move in the direction of the mask. Detection of displacement is done by comparing the relative position (or relative spatial phase) of the fringes.

One potential problem occurs when the mask is moved far enough: the fringes will return to their zero-phase position. This is referred to as *phase-ambiguity*. Therefore, it is necessary to perform coarse alignment to eliminate this ambiguity. Coarse alignment is accomplished by aligning a set of "bar-marks" on the mask and wafer. These marks can be seen in Figure 9 on page 27.

[1] The alignment marks can be designed significantly smaller (100 μm square) to comply with industry manufacturing requirements. The use of larger sized marks facilitates coarse alignment which is done manually in these experiments.

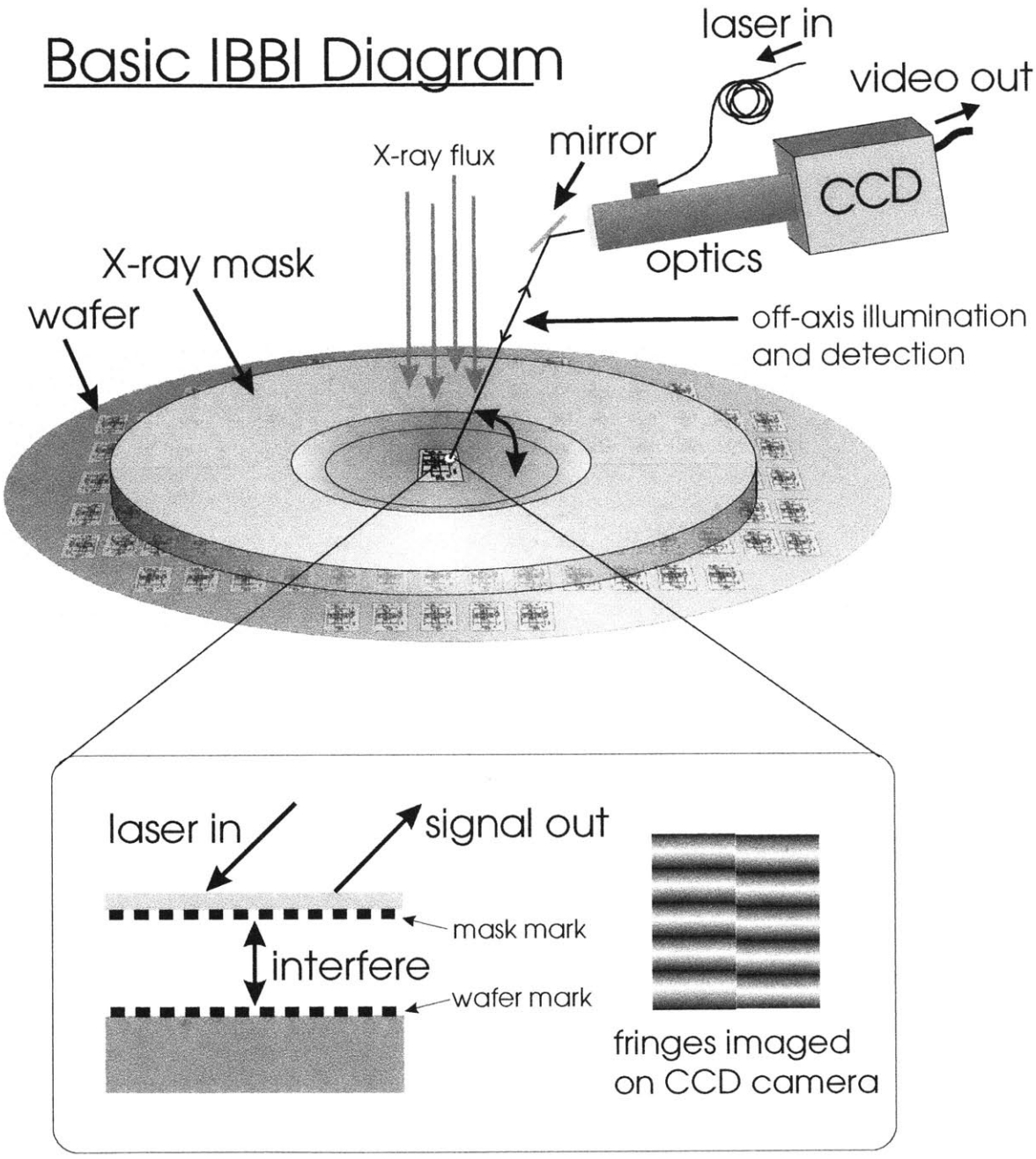


Figure 6 Schematic of IBBI setup. Since illumination and detection of IBBI signal is at an angle to the mask normal (15°), alignment does not interfere with X-ray exposure. In this project, a single 1 cm die is implemented on the mask. f/10 optics are used with a working distance of 11 cm. A 400 μm square area on the sample fills the CCD detector area.

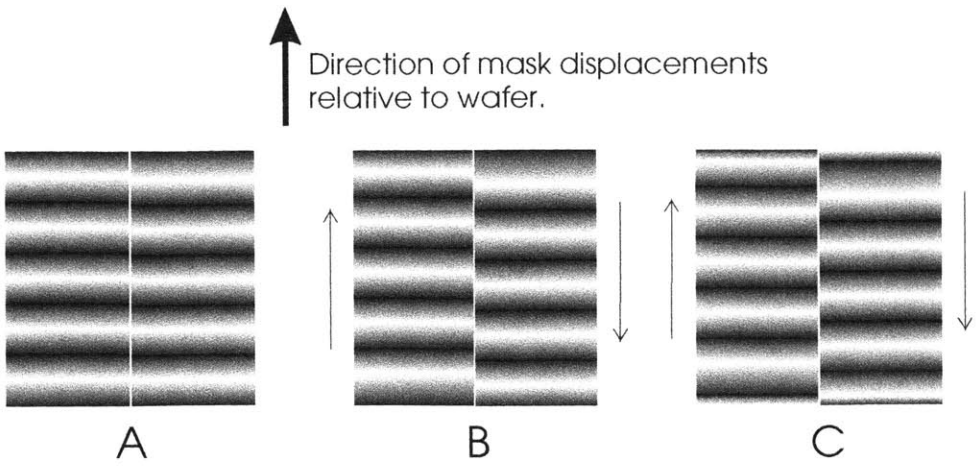


Figure 7 IBBI fringes. In (A) the fringes are matched. If the mask is displaced by a certain amount, the fringes on the left will shift upward while those on the right will shift downward (B). (C) depicts an additional displacement. Clearly, the fringes will "re-align" at a number of evenly-spaced displacements. This is referred to as phase-ambiguity. A coarse alignment routine is used to eliminate this problem.

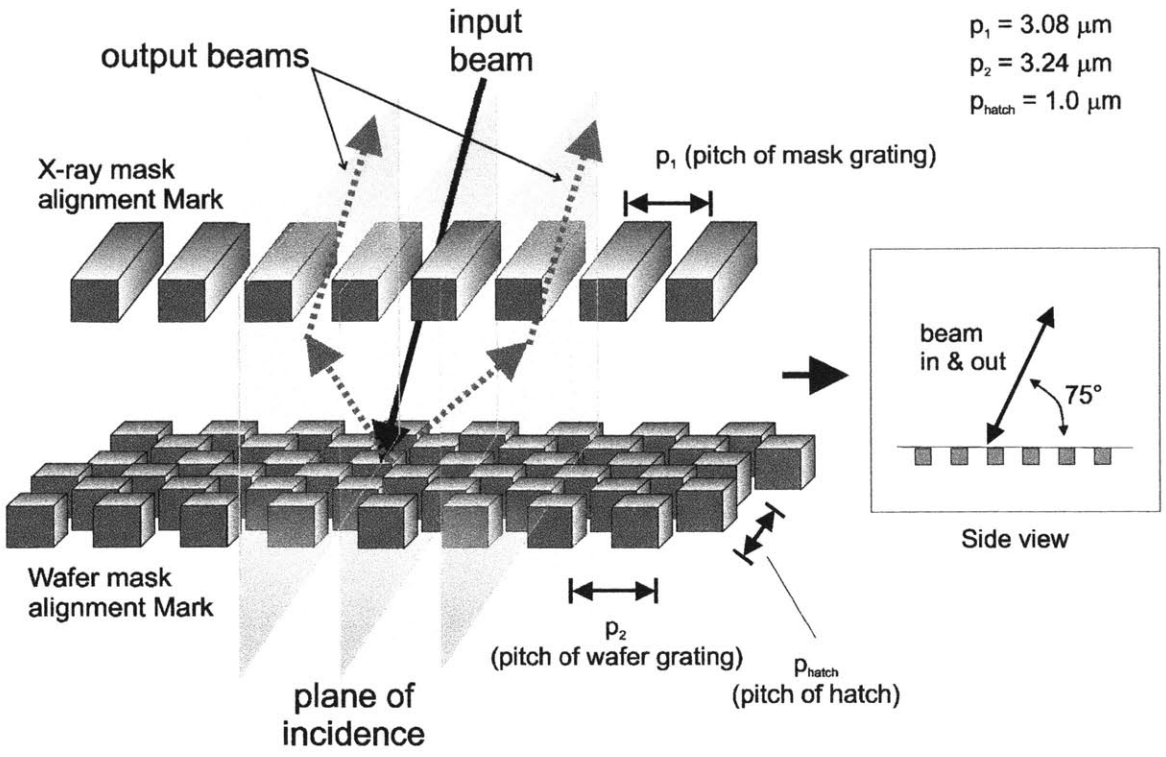


Figure 8 3-D diagram of IBBI alignment marks containing rays that indicate the directions of input and output beam paths. Input illumination passes through the mask grating within the plane of incidence (as depicted) and assumes an angle of 15° with respect to the mask normal. It is encoded with pitch p_1. Four 1st order diffracted beams are generated by the checker-board pattern on the substrate. The two forward beams are lost while the two back-diffracted beams return through the mask grating as depicted by the dashed rays. Since p_1 and p_2 are nearly the same pitch, the return beams are brought back into the plane of incidence. Furthermore, p_{hatch} is chosen so that these back-diffracted rays assume the same angle with respect to the mask normal as the input ray (shown in the "side view"). Fringes occur as the spatial beat frequency between p_1 and p_2.

Figure 8 explains how alignment information is generated by the IBBI alignment marks. The period of the fringes is a function of p_1 and p_2 as given by

$$p_{fringe} = \frac{p_1 p_2}{|p_1 - p_2|}.$$

while the magnification of displacement is

$$M = \frac{1}{2}\frac{p_1 + p_2}{|p_1 - p_2|}.$$

This value expresses the ratio of fringe displacement to mask-sample displacement. This is a major contributor to the alignment sensitivity of the system. Counter-motion of the fringes adds a factor of 2 to this figure, and magnification of the fringes by the optics adds further sensitivity. Figure 9 provides dimension information of the IBBI marks used in this project.

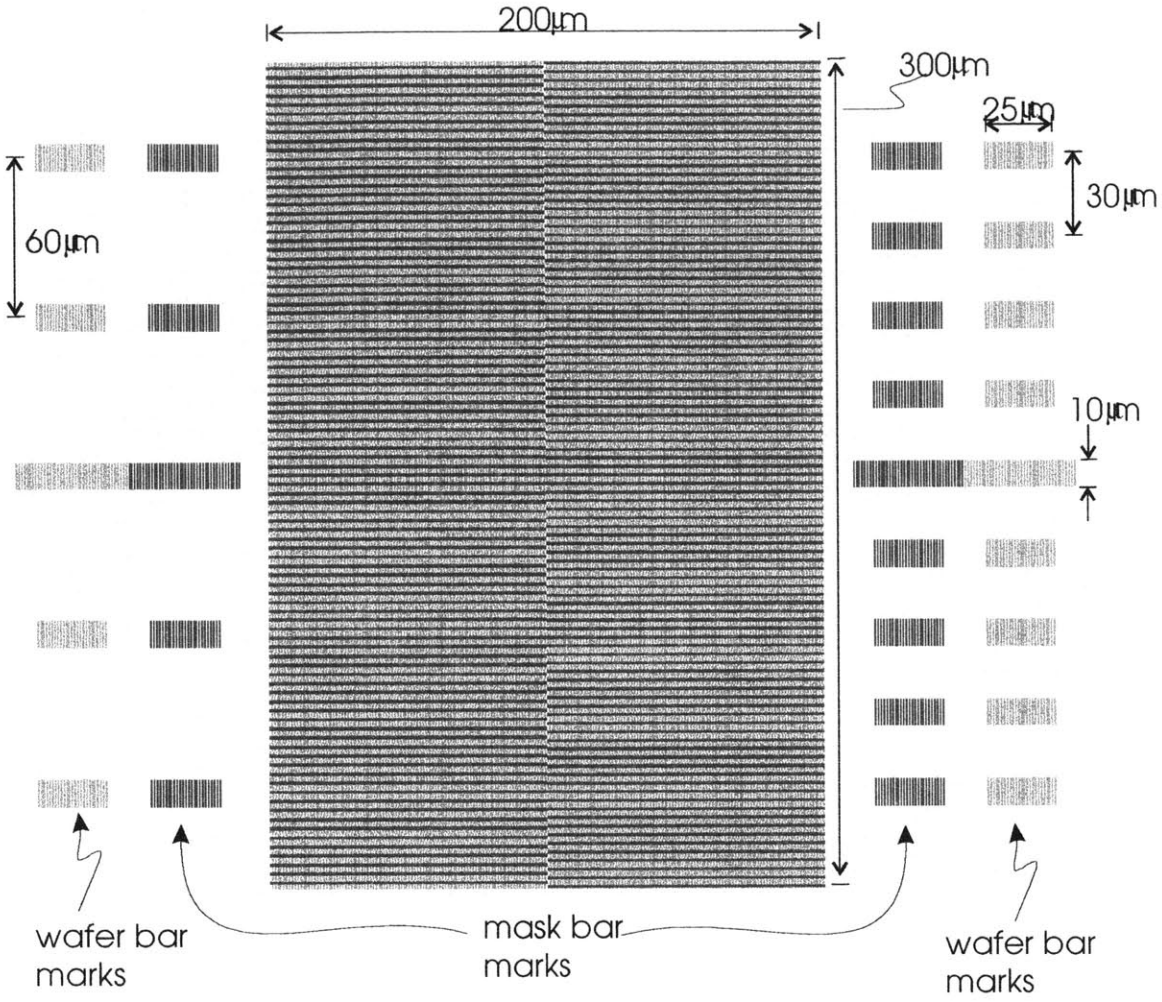


Figure 9 Dimensions of IBBI alignment mark components.

2.2 Gapping

Gapping is accomplished by inspecting green light interference within the mask-to-sample gap and through the use of IBBI-compatible gapping marks. This technique can be used to easily level the mask to within a fraction of the wavelength of green light since each visible fringe represents a half wavelength change in gap.

The gap detection marks consist of a chirped grating through which a laser signal is diffracted. The diffracted beams travel to the substrate and are then reflected back through the grating as depicted in Figure 10. The locations at which the return beams pass through the grating depend on gap such that larger

gaps result in larger transverse deviations from the ray input. The resulting interference signal from this scheme is a set of fringes that have low spatial frequency at small gaps and higher spatial frequency at large gaps. In designing this mark, it is preferable to chirp the grating such that the resulting fringes are evenly spaced and vary in a uniform manor with gap.

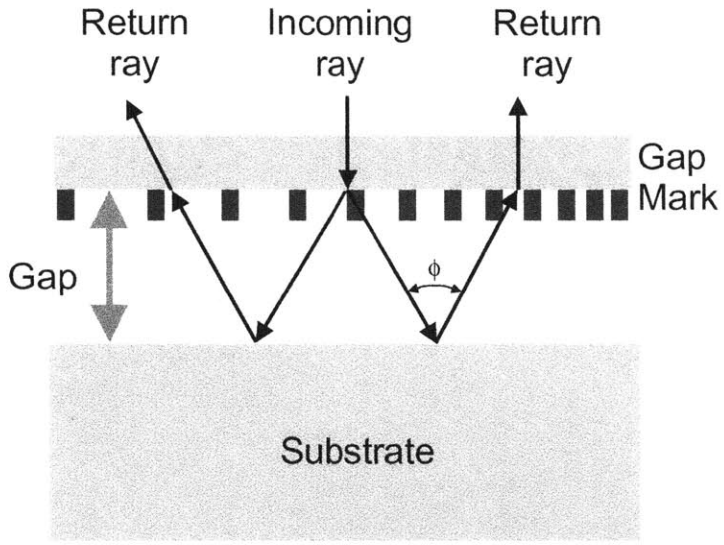


Figure 10 Schematic of gap detection mark. Each incoming ray is diffracted to the substrate and then reflects back up to the grating. The locations at which the return beam pass through the grating depend on gap such that larger gaps result in larger transverse deviations from the ray input.

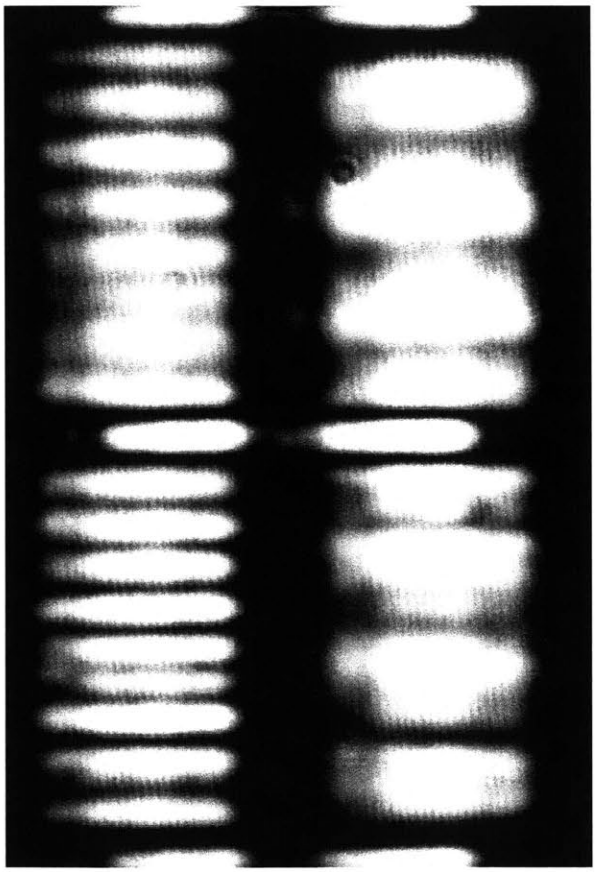


Figure 11 Image of gap mark signals. The signals from the left side of the mark differ from those on the right because the grating has approximately 20% additional chirp. That is, the pitch ranges from 1 - 1.3 μm compared to 1 – 1.1 μm. As gap is decreased, the pitch of both sets of fringes broaden. By using gratings with various chirps, a larger range of gap can be measured with greater sensitivity.

2.3 Grid/Grating Software

The IBBI alignment system, innovated by Euclid Moon [**10**], advanced in many ways over the past decade. In order to facilitate the design and generation

of new alignment mark ideas, I created software to automatically produce grid and grating patterns will a high degree of flexibility. Figure 12 shows the interface for this software. The flexibility of the algorithm allows for gratings or grids in "street hatch" or "checkerboard" format as well as control of pitch or duty cycle. Additional software allows designs to be imported from a spreadsheet file.

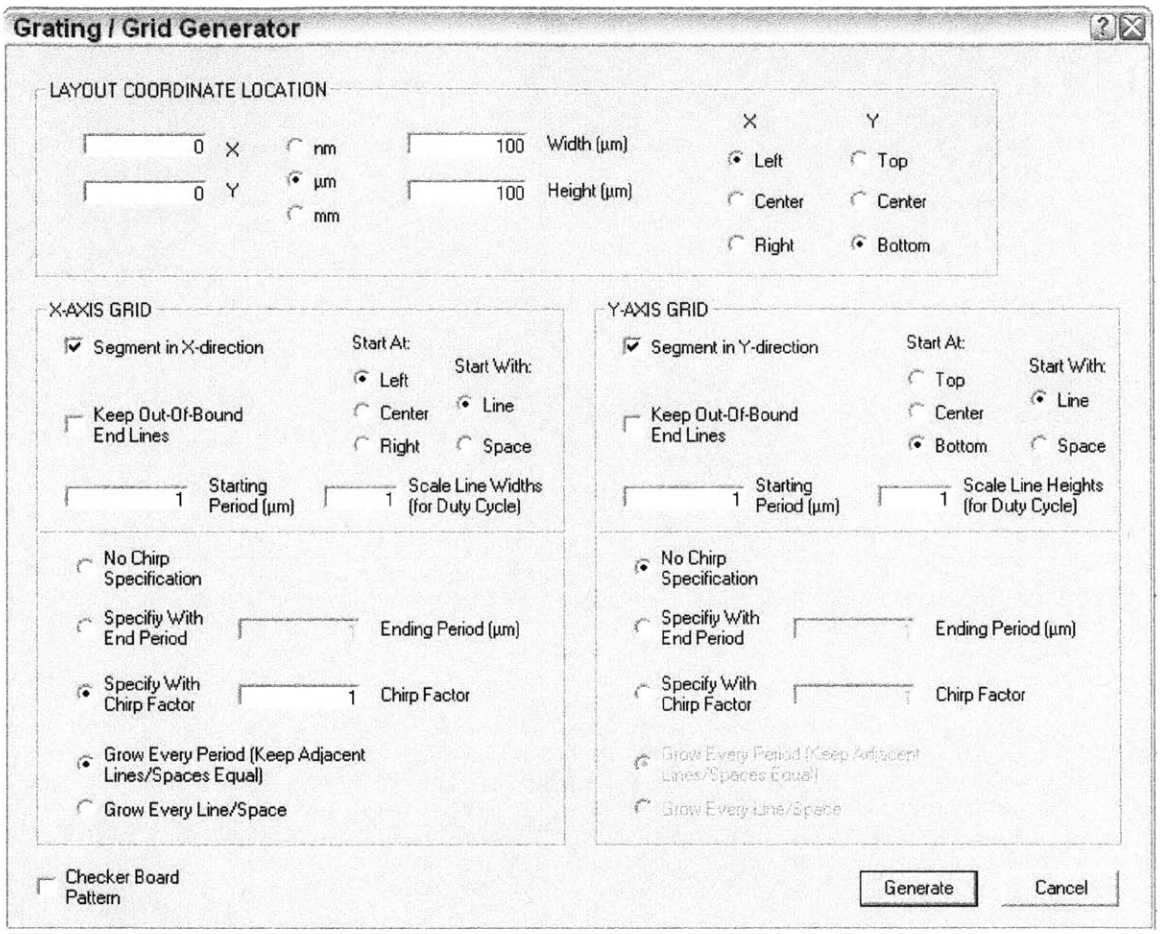


Figure 12 Grating/Grid generator software interface.

2.4 X-ray Source Alignment

During X-ray exposure, a gap is maintained between the mask membrane and the sample. Therefore, if the direction of impinging X-ray flux is not normal to the mask, mask patterns imaged on the wafer will be displaced laterally from their position on the membrane. The propagation direction of X-ray flux, in turn, is dependent on the type of source used. In NSL, X-rays emit from a point on a copper target via electron bombardment. Technically speaking, in such point-

source systems, the direction of X-ray flux is normal to all spheres whose center is the point of X-ray emission. The resulting pattern magnification that takes place during gapped exposures is addressed mathematically in Appendix D. In short, the effects of magnification do not pose a problem in pattern placement as long as gap and the X-ray axis are known. By keeping track of this information, effects of magnification from previous exposure steps can be corrected. The X-ray axis can be defined as a line normal to the mask/sample surface that passes through the point of X-ray emission.

A scheme was developed to align an arbitrary point on the mask with the X-ray axis. The following set of figures and captions details a step-by-step process to accomplish this.

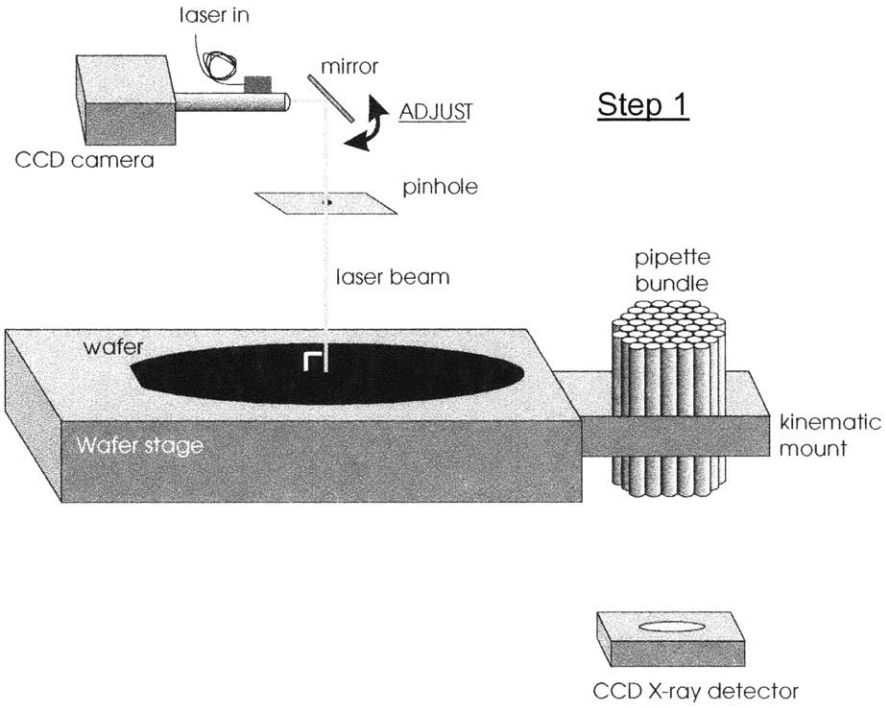


Figure 13 The X-ray source alignment scheme consists of a wafer stage to which is mounted a pipette bundle and CCD X-ray detector. An IBBI scope (imaging scope) emits a collimated laser beam that can be directed by a mirror via a kinematic mount. In the first step of the alignment process, the beam is adjusted so that it is perpendicular to the wafer. This is done by passing the beam through a pinhole and ensuring that it returns through the pinhole.

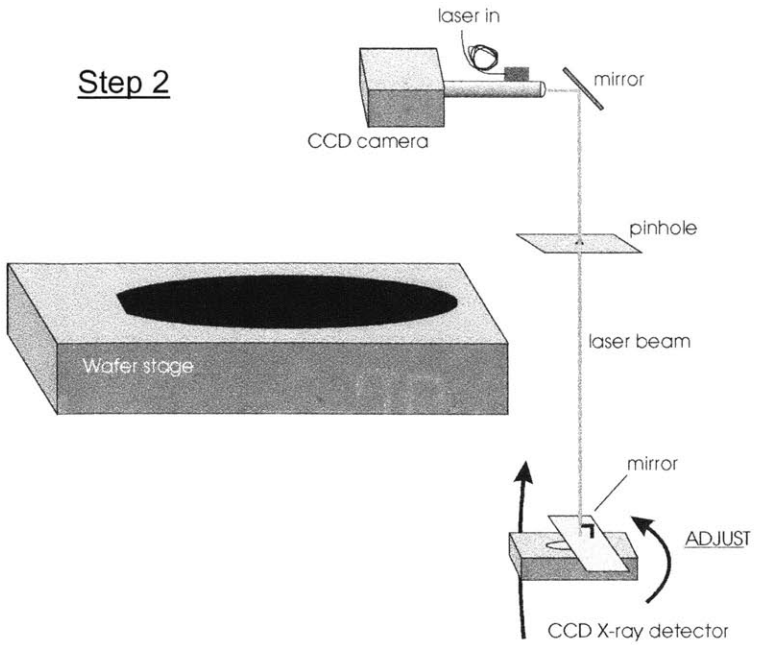


Figure 14 In this step, the stage assembly is shifted so that the X-ray CCD detector is beneath the vertical beam. A mirror is placed on the CCD mount and the angle of the detector is adjusted so that the beam is directed back through the pinhole. This step ensures that the X-ray CCD is parallel to the wafer stage.

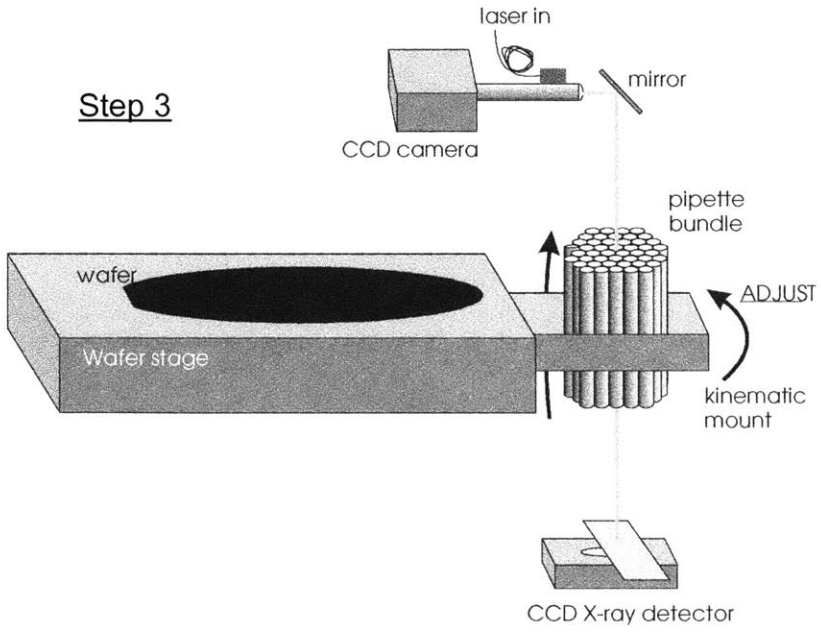


Figure 15 Here, the pipette bundle is mounted to the stage. With the mirror still in place on the X-ray CCD and the beam still directed perpendicular to the X-ray CCD, the angle of the pipettes is adjusted so that the axis of the pipettes is parallel to the beam. The scope images the light modes that occur within one of the pipettes. When the pipettes are perpendicular to the X-ray CCD, a single bright spot can be seen in the pipette. At this point, the pipette axis is perpendicular to the X-ray CCD and the wafer stage.

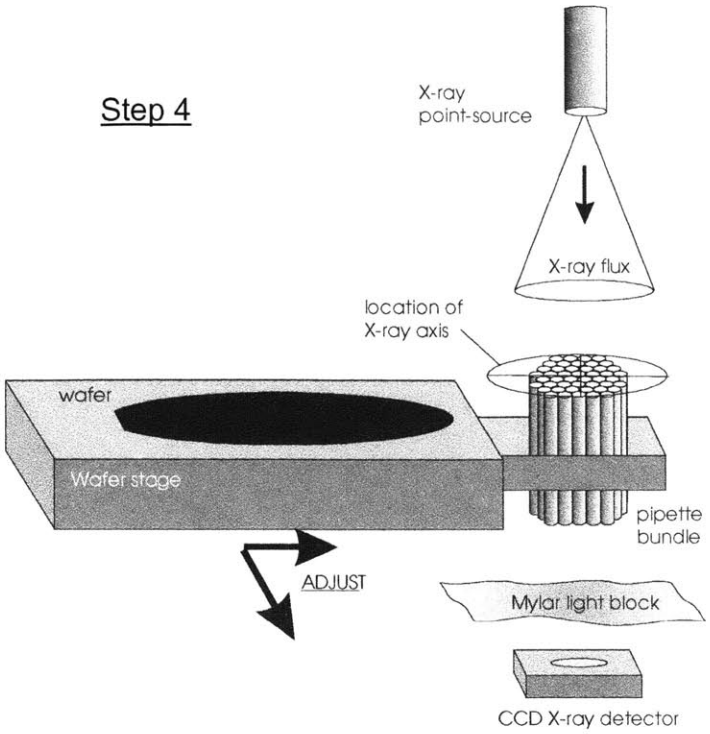


Figure 16 Now, the X-ray CCD is activated and the X-ray source is turned on above the pipettes. A metalized Mylar film is placed over the CCD to filter out ambient light. The position of the stage assembly is adjusted such that the image of the X-ray radiation as seen through the pipette bundle is centered within the pipette bundle. The center axis of the pipette bundle is now collinear with the X-ray axis. The stage position is locked in place.

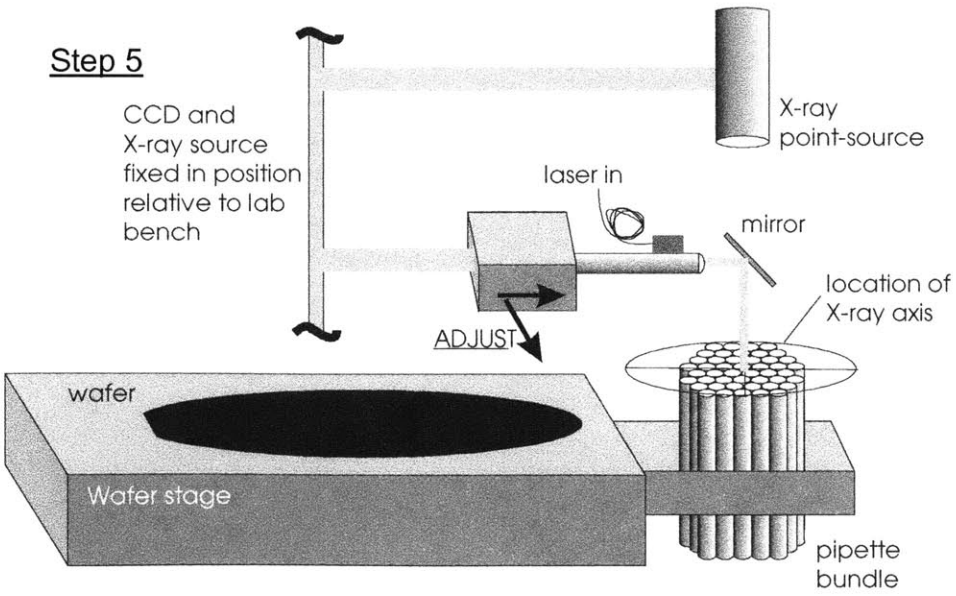


Figure 17 Next, the scope is re-focused onto the top of the pipette bundle and crosshairs (in computer software) are positioned at the center of the bundle. The scope and X-ray source remain stationary with respect to the lab bench. The scope has essentially captured the location of the X-ray axis.

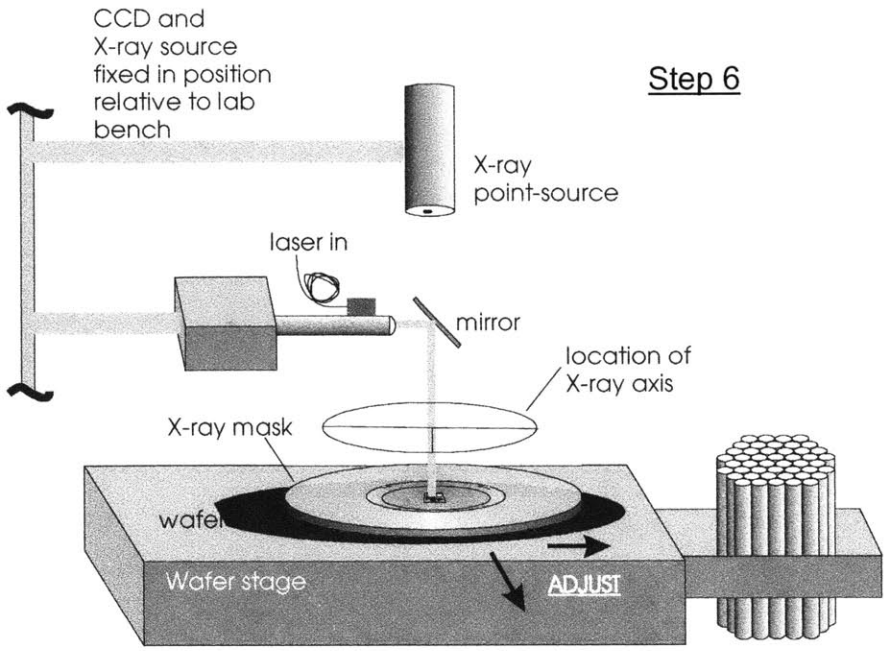


Figure 18 The wafer stage is slid so that the mask and wafer are imaged by the scope. The position of the mask can now be adjusted so that any location within the die can be set onto the X-ray axis. The wafer is then aligned to the mask.

2.4.1 When Is X-ray Source Alignment Necessary?

X-ray source alignment is necessary when performing the X-ray mask replication and when patterning the upper-gates. It is not necessary when patterning the lower-gates. This is because it is permissible for the lower-gate patterns to be shifted several 10's of nanometers. For example, if the source-to-sample distance is 30 cm, the gap is 5 μm, and the X-ray source is displaced by 3 mm, the resulting pattern shift would be 50 nm. This would only affect alignment of the lower-gates with respect to the optically (stepper) patterned layers such as the active layer. Since the stepper tolerance is on the order of a micron or so, the added misalignment due to X-ray pattern shift is only 5%. On the other hand, as discussed earlier, knowledge of gap is always important since magnification during any X-ray step will affect the overlay between upper and lower-gate levels. Furthermore, it should be clear that pattern shift either during mask replication or upper-gate patterning would be unacceptable because despite the fact that the alignment marks may be aligned perfectly, pattern overlay would be lost.

Also to be considered is what happens to the lower-gate alignment marks when they are patterned onto the wafer (please refer to the layout diagram in Figure 31 on page 48.) The X-ray axis is consistently set on the center set of the main X-Y alignment marks. This means that all patterns (including the outer sets of alignment marks) will spread away from the center when lower-gate patterning takes place. The center mark is responsible for X-Y alignment. An additional alignment mark is used for angular alignment (an additional Y-mark). It is chosen such that its alignment direction is perpendicular to its shift due to the magnification effect. In theory, the alignment readings from an IBBI alignment mark are immune to small shifts perpendicular to the direction of alignment.

3 Device Process

3.1 Double-gate MOSFET Process Overview

In order to understand the lithographic requirements for this device process as outlined in the following section, the device process itself is summarized here. The process scheme chosen for this project is based on the use of SOI (Silicon-on-insulator) wafers. Conceptually, the upper and lower surfaces of the Si film are individually exposed to lithographic patterning at distinct points in the process flow. Figure 19 shows a cross-section of a starting SOI substrate and the surfaces to be patterned.

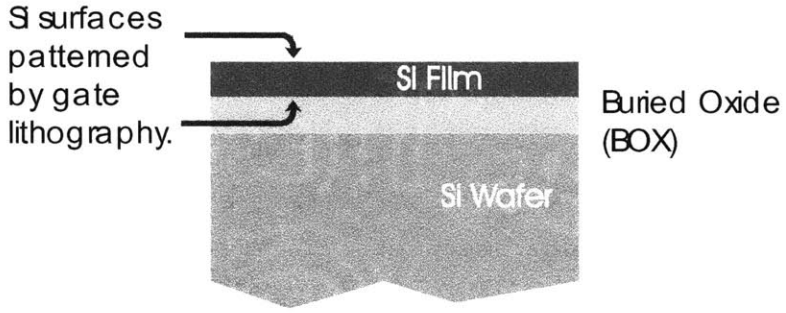


Figure 19 SOI wafer cross-section.

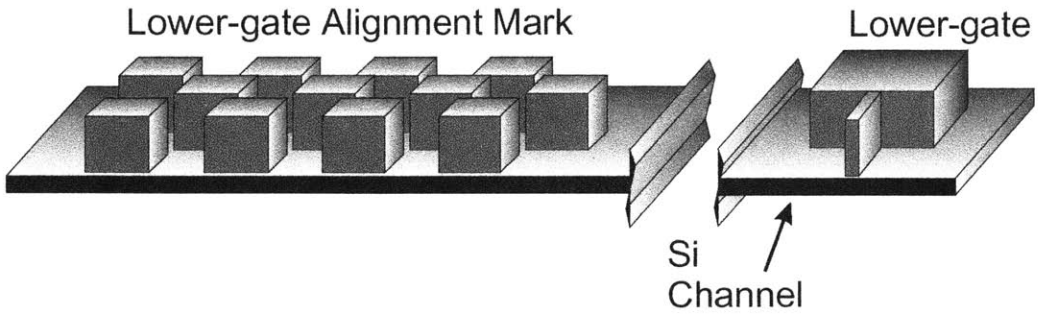


Figure 20 Patterning lower-gate and associated alignment marks.

In the first gate lithography step, the lower gates and associated IBBI alignment marks are patterned on the top surface of the SOI wafer. As shown in Appendix A (steps 1-3), this is done by patterning resist on a layer stack

consisting of Low Temperature Oxide (LTO), gate-poly-Si and thermal oxide.[1] Once the pattern is transferred into the LTO, the LTO behaves as a hardmask for the gate-poly. The gate-poly is etched down to the thin thermal oxide layer (gate oxide). An illustration of this step is shown in Figure 20 where only the Si film and gate materials are depicted for clarity.

Steps 4,5 and 6 in Appendix A show the flip-and-bond process. The poly pattern is buried in a thick LTO layer; the top surface is polished smooth and the wafer is bonded to the oxide of a bare wafer (called the handle wafer). Material from the SOI wafer is removed such that the opposite Si surface is exposed. Again, an LTO / Poly / gate-oxide stack is formed (step 7). Step 8 shows the aligned x-ray exposure that defines the upper-gates with respect to the lower gates. Here, the upper-gate mask is aligned to the lower-gate patterns that are buried beneath the gate stack and Si film of the device wafer. This is also shown in Figure 21.

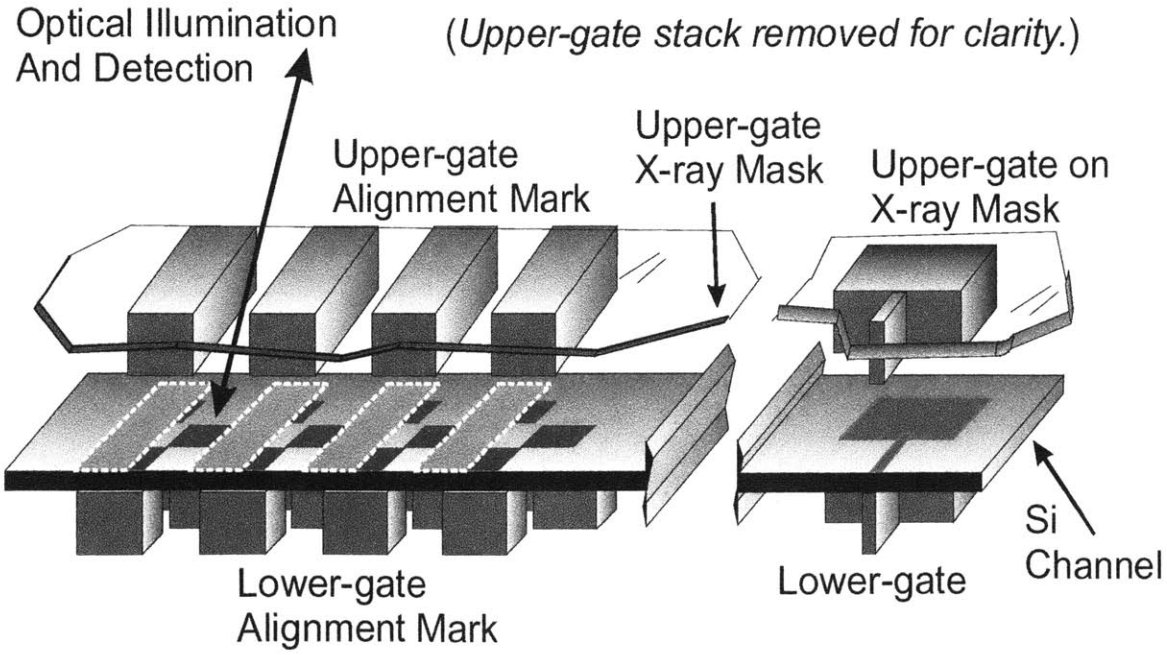


Figure 21 Depiction of aligned exposure of upper-gate mask to the buried lower-gate of the device wafer. (Note: for clarity, the upper-gate stack is not shown.)

[1] A more sophisticated gate stack is presented in section 3.3

Clearly, alignment of the gate patterns in this critical step will not be satisfactory unless the placement of upper-gate device patterns matches the placement of the lower-gate patterns. To ensure pattern overlay, the lower-gate x-ray mask is made by direct replication from the upper-gate mask (Figure 22). In essence, the same aligned exposure that is used in this mask replication step is used when patterning the upper-gates on the device wafer. This is because the same alignment marks are used in both exposures. The flip-bond wafer process lends itself to this face-to-face x-ray mask replication process since the device process requires a mask pair with left-handed/right-handed patterns respectively.

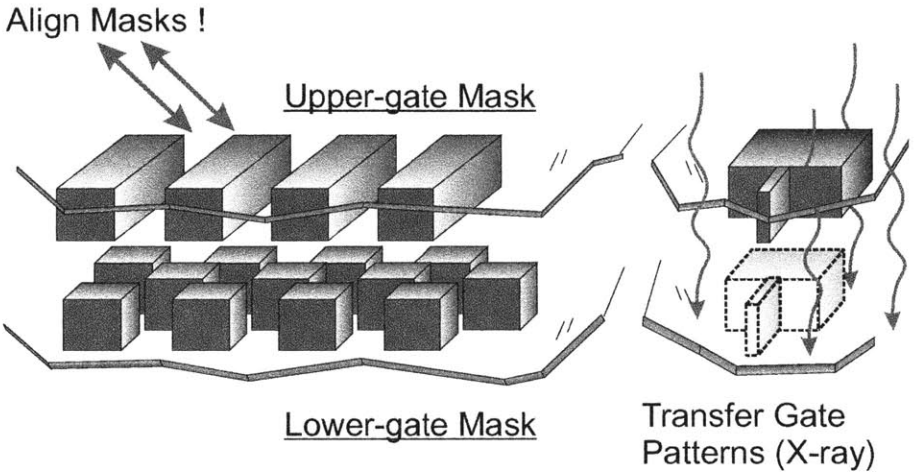


Figure 22 Depiction of the mask replication procedure. Here, device patterns are transferred from the upper-gate mask to the lower-gate mask while the masks are held under alignment.

3.2 Mix and Match Scheme: X-ray & Optical Lithography

Several forms of lithography are employed in this project including X-ray lithography, e-beam writing, optical contact printing and G-line optical projection printing. The device wafers are patterned using only X-ray and G-line. X-ray is used for the upper and lower gates while G-line is used for active area, contact cuts, metalization and preliminary alignment patterning. The X-ray masks are designed with a single 1 cm square die centered on its 30 mm diameter membrane. The FETs are designed with a minimum 2 μm tolerance between the

edges of gate features and certain optically patterned features (such as contact cut edges). Therefore, it is important that when the gate features of the upper-gate mask are patterned using the e-beam tool, their placement matches that of the optical stepper to some degree. That is, the 2-D pattern distortion of the stepper must be matched by the e-beam tool so that X-ray patterns (e.g. gates) match subsequent stepper patterns (e.g. cuts, metal). Refer to the section 4.2 for a detailed explanation of e-beam registration to X-ray masks.

3.3 Wafer Preparation: Backgate X-ray Patterning

When patterning the lower-gates using X-ray, it is important to measure the gap between the membrane and the wafer. This is because we wish to keep track of the pattern magnification in the process. Examining the overall process, there are three X-ray exposures: 1) the aligned X-ray mask replication, 2) the lower-gate exposure and 3) the aligned upper-gate exposure. Magnification of the gates occurs first when the masks are replicated such that the lower-gate mask contains an enlarged image of the upper-gate mask. Then, when the lower-gates are patterned onto the device wafer, this image is enlarged once more. Since these enlargements (or magnifications) are linear with gap, the total enlargement (now realized on the wafer) is the sum of the two magnifications corresponding to the sum of the gaps in these two exposure steps. Therefore, when it is time to pattern the upper-gate onto the wafer, it is clear that the upper-gate mask is being aligned to an image with a cumulative magnification from previous steps. The patterns, however, can be matched simply by setting the gap of this final exposure to the sum of the gaps of the two previous exposures. There is, however, a tradeoff if one is to enlarge the gap of the final exposure since this modified gap may be too large to resolve the finest linewidths (see, for example, section 4.4.2.1). If linewidth broadening occurs due to an increased gap, then corrections to the linewidth can be made using the technique in section 3.3.1.1.

As an example, assume that the minimum gap in the final X-ray exposure is approximately 15 μm and that we apply the diffraction model discussed in section 4.4.2.1. According to the graph in Figure 45 plotting gap versus minimum

resolvable linewidth, (say alpha is 2.0 - a reasonable assumption), one would expect to achieve 80 nm lines in resist. Since the target minimum linewidth is 50 nm, (for 25 nm effective channel length gates), one would have to reduce the physical gate lengths by 15 nm on either side.

Additionally, if a liftoff process is used when replicating the lower gate mask, an unwanted pattern enlargement occurs (as discussed in section 4.4.3). Although it may be acceptable for the lower-gate to be slightly wider than the upper-gate, it may be necessary to correct its size.

An effective approach was developed to make such modifications. The following processes that were employed to reduce lower-gate linewidth also greatly enhanced our poly-Si gate sidewall profile. It is clear that similar thinning processes are common in industry by observing the disparity between gate linewidths in resist versus linewidths after gate etch [**11**]. Whether some of these gate etch schemes are entirely dry processes or a combination of dry and wet processes is most likely proprietary. Nonetheless, it is clear that a 20 nm reduction in linewidth is possible under the uniformity constraints imposed by the industry. In our case, a wet etch scheme was developed to correct lower-gate linewidth by about 30 nm.

3.3.1.1 Modified Gate Etch

In this approach, the gate stack is modified simply by adding a thin layer of poly on top of the LTO (Figure 23). The poly layers are now referred to as top-poly and bottom-poly. With the resist as an etch mask, the top poly is etched, and with the top-poly as an etch mask, the LTO is etched (Figure 24). At this point, one has the option of performing the linewidth reduction by immersing the sample in buffered HF solution (Figure 25). Finally, the lower-poly is etched using the LTO as an etch mask. The top-poly is about one tenth the thickness of the lower-poly. Although there was concern that the overhang of the top-poly would lead to non-uniformity in the vicinity of the line when etching the bottom poly, it was found that the bottom-poly could be cleared without visible residue on the oxide and no evidence of oxide punch-through was present.

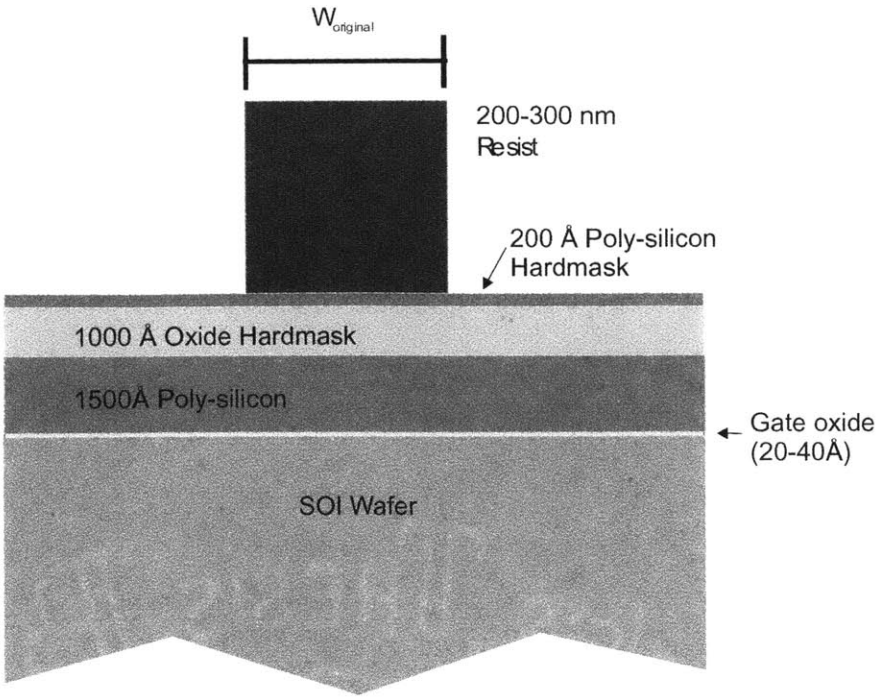


Figure 23 Initial resist pattern on modified gate-stack consisting of a double-hardmask (poly-Si & LTO).

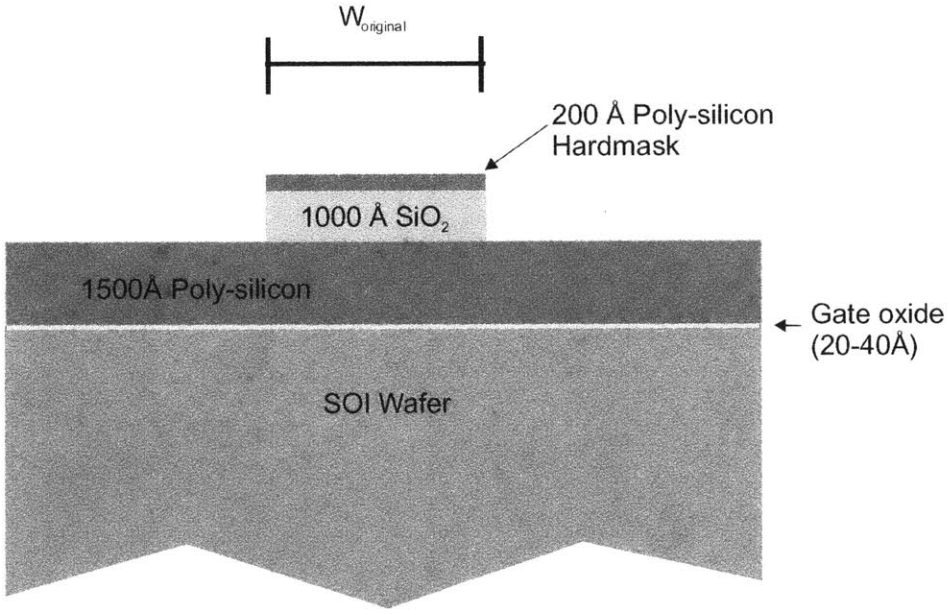


Figure 24 Transfer of resist pattern into the hardmask layers.

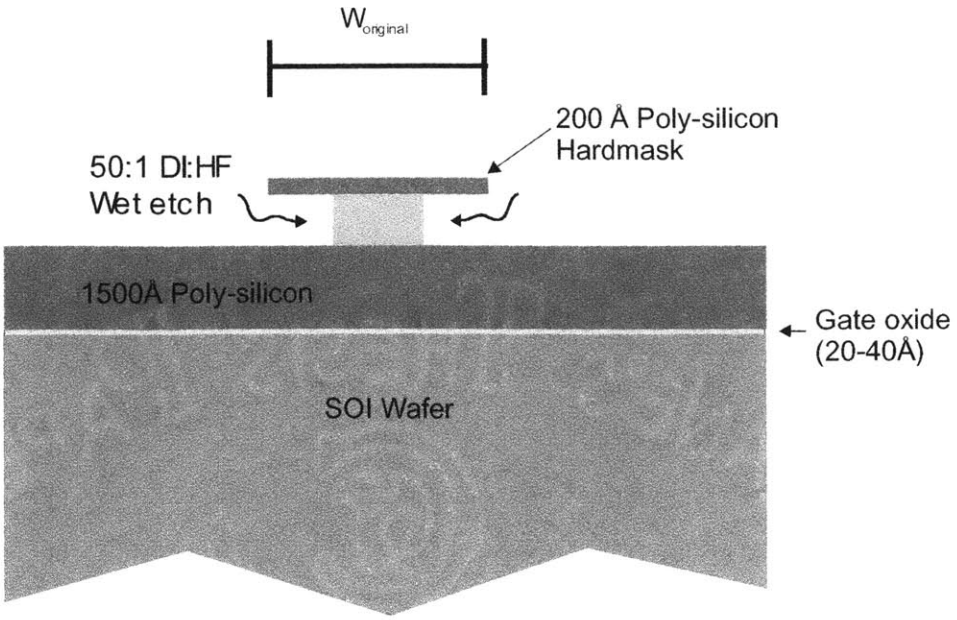


Figure 25 Reduction of linewidth though anisotropic wet etch of LTO hardmask interlayer using 50:1 DI:HF solution.

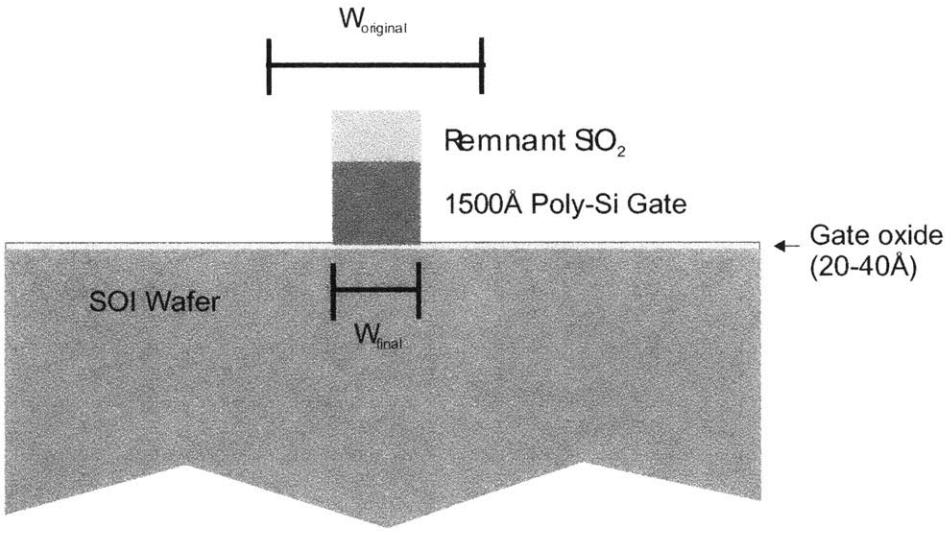


Figure 26 Transfer of the modified hardmask into the Poly-Si Gate layer. In this step, the overhang of the upper-poly hardmask is etched along with the gate poly.

This process was developed using large area 300 nm pitch gratings fabricated using interference lithography. A Loyd's mirror setup, (operating at 325 nm wavelength), was used to expose test wafers with the triple-gate stack and Ohka THMR-iN negative chemically amplified resist. Figure 27 shows etch results using this system. In this example, some inward lateral etching of the poly-Si suggests that a proximity effect caused some isotropic etching.

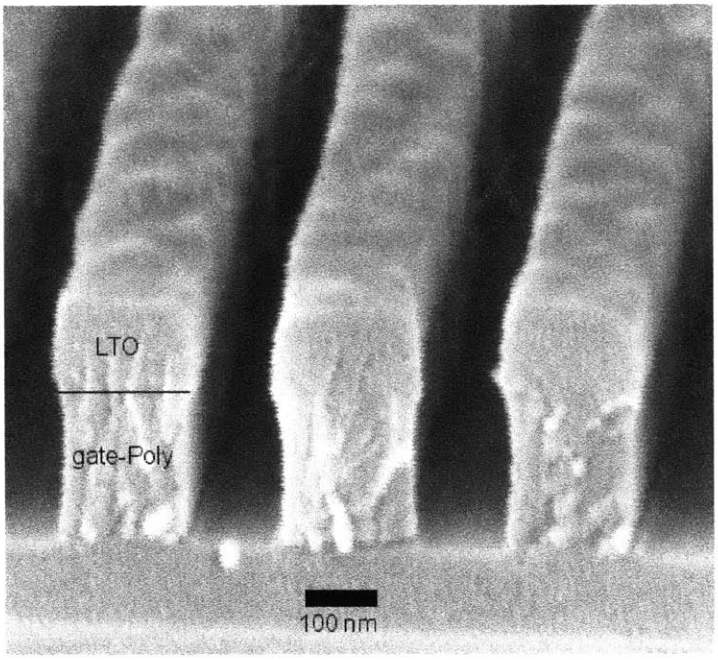


Figure 27 SEM micrograph of poly-Si grating etching using poly/LTO hardmask defined by interference lithgoraphy. Sidewalls show slight inward bow.

Figure 28 shows the grating just after HF wet etch of the LTO layer (prior to etching the lower-poly). Apparently, the wet etch process was quite robust and well controlled. Figure 29 shows etch results with the thinning step employed. Here, the sidewalls were more vertical than in the case of broader lines (Figure 27). Finally, we tested this process for compatibility with X-ray lithography using UV-5 resist to pattern isolated lines down to 60 nm. Figure 30 shows this result for which about 10 nm of LTO thinning was applied.

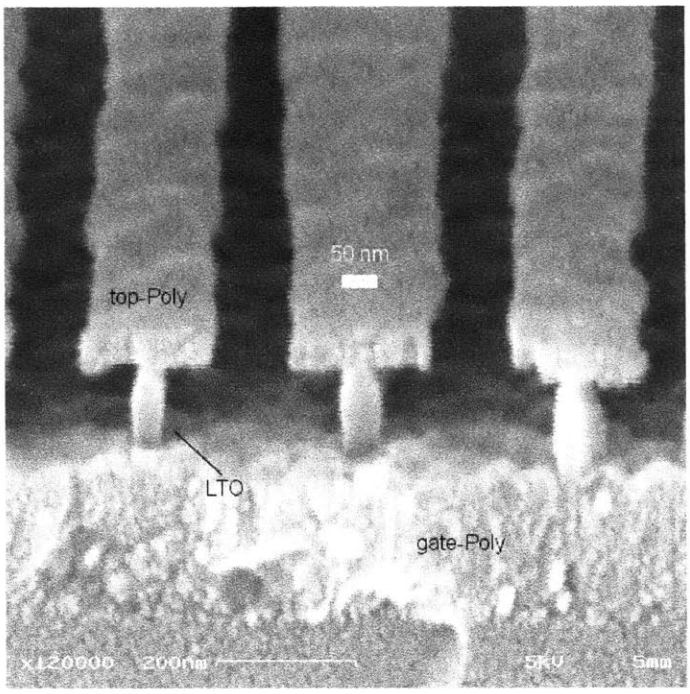


Figure 28 Gate etch process after upper-Poly and LTO are etched, followed by a timed HF wet etch to shrink the LTO layer laterally.

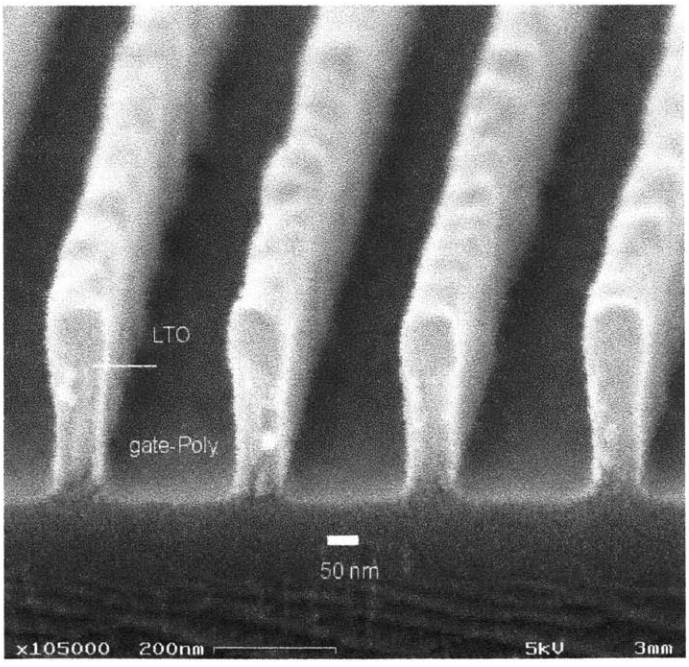


Figure 29 Gate etch process of test grating employing the HF thinning step. At this feature size, sidewalls are highly vertical.

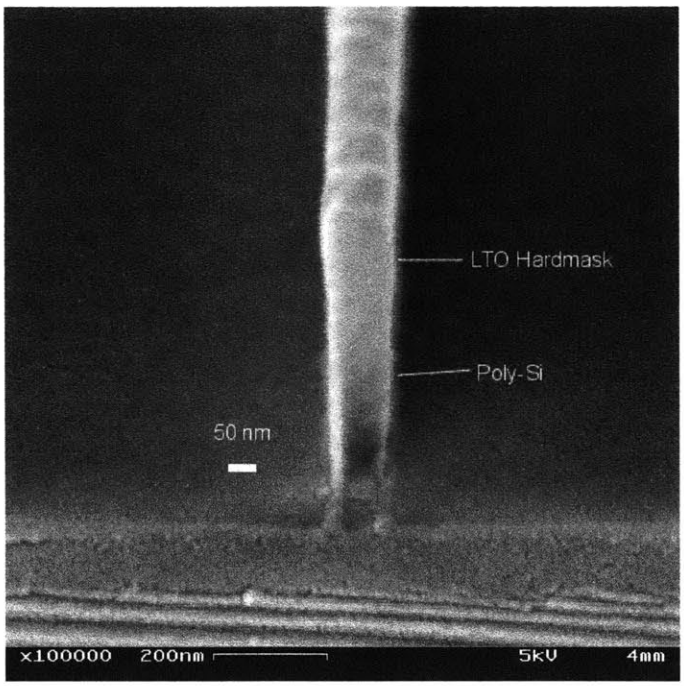


Figure 30 A gate stack etch test of isolated features to test the etch process for compatibility with X-ray patterning.

3.3.2 Planarization & Bonding

The planarization step consists of deposition of LTO followed by chemical-mechanical polishing (CMP). Appendix B contains a table showing deposition and CMP parameters applied to the wafers. Good planarization is critical for successful bonding. Quality bonding, in turn, eliminates pitting and blistering of the films during the etchback of the handle wafer. *Voids* are regions where air traps or delamination occur between wafers. Infrared transmission photographs can be used to detect the their presence. Voids are generally the result of uneven topography of the patterned wafer or the presence of particles. During earlier stages of process development, voids of consistent shape and orientation had systematically appeared adjacent to the IBBI alignment marks. In order to eliminate this problem, the LTO deposition & CMP steps were repeated until the surface roughness topology was below 5 nm on dense features.

The bonding process used in this phase of the project was a manual scheme, consisting of shims and a Teflon® roller. An RCA clean was performed immediately preceding this step. The shims were 0.2 mm and were set at the periphery of either the patterned wafer or the handle wafer. The mating wafer

was then set on the shims and pushed into contact in the center. The shims were then removed and the upper wafer was pushed into contact using the roller.

Following mechanical bonding, an anneal was done at 850C for 1 to 1.5 hours in order to increase the oxide-oxide bond strength.

3.3.3 Grinding, Etchback & Upper-gate Stack

Prior to mechanical grinding of the SOI wafer, SiN_x was deposited onto the wafer pair to protect the handle wafer from being etched during subsequent wet chemical etches.[2] Wafer pairs are sent to Silicon Valley Microelectronics for grinding of the SOI wafer. The thickness of the bonded pairs, initially at about 1 mm, is reduced to 600 μm after grinding leaving about 100 μm of the original SOI wafer. This remaining Si is etched in 25% TMAH at 85°C from 2.5 to 4.5 hours thereby exposing the buried oxide of the SOI wafer. This oxide is then removed in 6:1 DI:buffered HF solution.

[2] This CVD step is done twice. For the second deposition, the wafers are rotated in the quartz boats in order to re-deposit over at support points along the wafer edge where the boat blocks growth.

4 X-Ray Mask Process

4.1 Design of Layout

The layout (Figure 31) is designed with the following goals in mind:

- Ability to analyze alignment results either by 1) cleave of long double-gate test structures leading to a SEM of double-gate cross-sections or 2) access via FIB (focused ion beam milling) of similar yet smaller test structures arrayed across the die.
- Suitable array of devices containing DGFETs with various gate lengths and gate widths.
- Incorporation of IBBI alignment marks and auxiliary marks. Marks must be positioned such that they are immune to magnification error shifts. They must be set far enough apart to allow for sufficient angular alignment detectivity.
- Alignment marks and devices are clustered together as tightly as possible about the axis of the X-ray exposure system to maximize yield of well-aligned devices. The goal is to mitigate the effects of X-ray lithographic magnification effects and other contributions to 2-D distortion.
- Must contain alignment marks for alignment of lower-gate X-ray mask to stepper patterns.
- Coarse feature test structures are placed near the perimeter of the die.

Various methods to determine the effectiveness of the alignment and overlay scheme are supported in this design. The most straightforward is the inclusion of 3 mm-long double-gate structures that mimic the real DGFET in cross-section. These structures, as well as the FIB structures, include patterns that define upper and lower gates, active region and contact cuts. The so-called cleave lines (or SEM lines) are written in sets of 15 lines of varying linewidths and there are 4 sets spaced evenly, both horizontally and vertically, near the center of the die. With this setup, a single cleave yields alignment data at 4

positions along a given axis. If the remaining halves of the die are cleaved once more, alignment can be measured in the orthogonal axis.

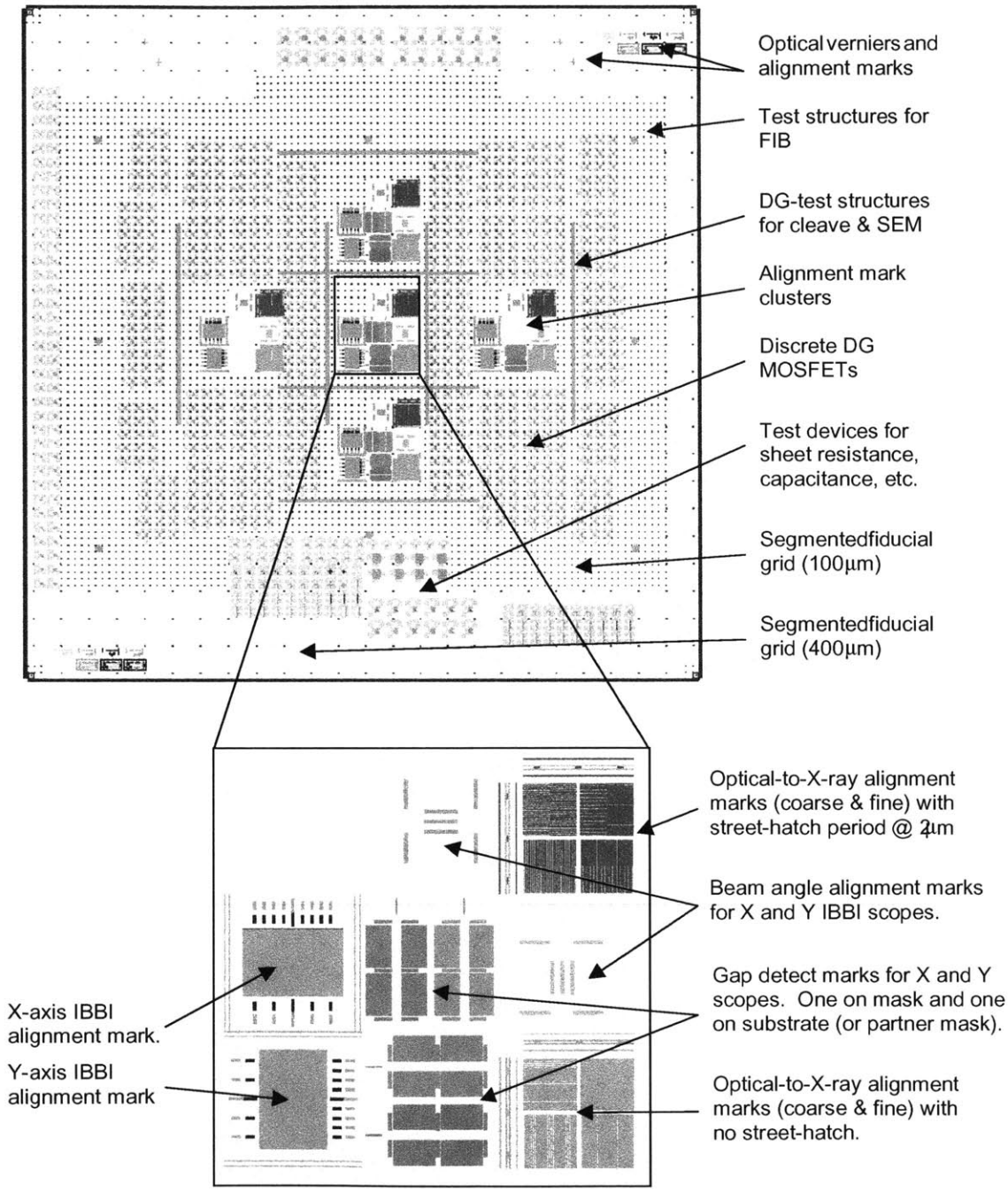


Figure 31 Arrangement of layout. This view shows all layers including: fiducial grid for e-beam registration, upper and lower X-ray mask patterns and nMOS layers.

The design rules for the DGFETs were chosen to maximize the yield of devices given the tolerances that arise from the mix and match scheme (see Figure 32). For example, consider alignment of contact cuts in the active area

with respect to the gates. This is actually a "tertiary alignment" since these cuts are actually aligned to the active area which, in turn, is aligned to the gates. Given a 1 μm alignment tolerance for each of these steps, a 2 μm design rule is the minimum allowable spacing for these features. Chosen gate widths, (defined by the width of the active region), are 2, 10 and 30 μm. Gate lengths (based on linewidths on the upper-gate X-ray mask) range from 60 – 120 nm in 5 to 10 nm increments and coarse gates at 0.2, 1, 2 and 5 μm. Contact cuts are 3 μm square, gate pads are 8 μm square and metal pads are 60 μm square.

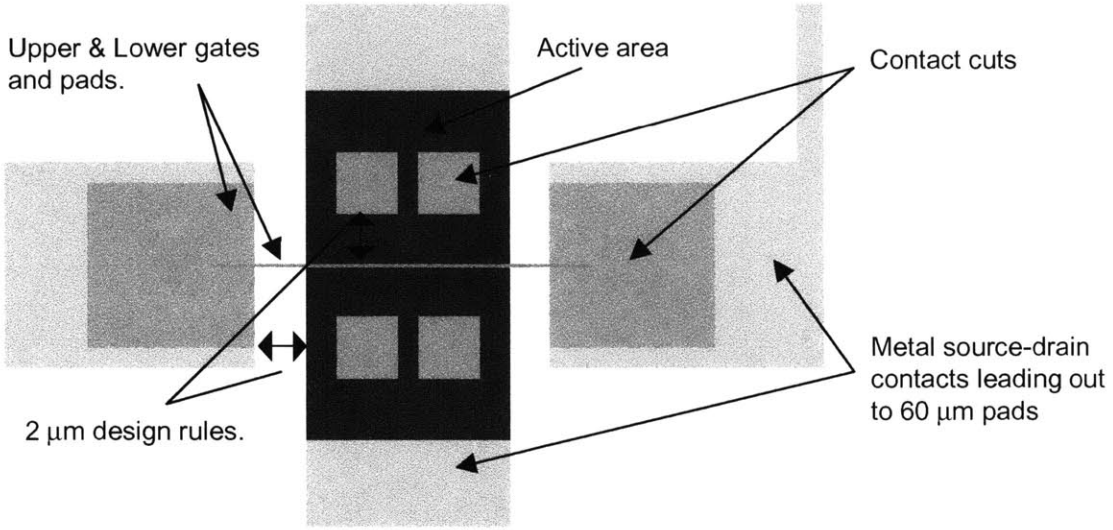


Figure 32 Typical DGMOSFET layout showing design rules of 2 μm between gate pad & active. Also between active contact cuts and gate. Note that metal contact over active region is cut-away to make region visible.

4.2 X-ray Mask Preparation For E-beam Patterning

As mentioned earlier, since gate features of the upper-gate mask are patterned using the e-beam tool, their placement must match that of the optical stepper in order to maintain overlay in post-gate patterning steps. In order to accomplish this, a fiducial grid patterned by the stepper is placed on the X-ray masks. It is then scanned by the e-beam system so that features can be appropriately placed. The device layout contains 100 μm and 400 μm field sizes. Devices and test structures reside in 100 μm fields while alignment marks and

other IBBI related marks reside in 400 μm fields. This pattern of registration marks thus identifies field boundaries. It consists of a 1.5 to 2.5 μm dot placed near the corner of every field as shown in Figure 33.

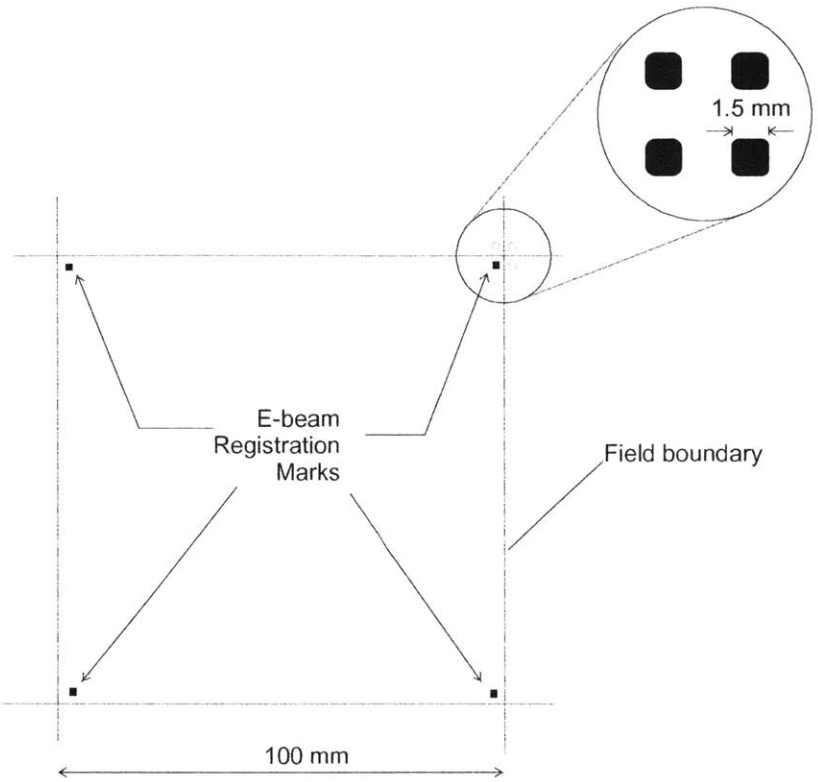


Figure 33 A 100 μm e-beam field and its associated registration marks. These marks, which are initially patterned onto the x-ray mask, are scanned by the e-beam writer. The e-beam system, in turn, places features in a coordinate system corresponding to the registration marks. (Placement accuracy within a field is estimated at roughly 30 nm.)

The registration patterns are printed onto the X-ray masks using 4" chrome-patterned quartz wafers. The quartz wafers, in turn, are prepared by printing a 10x emulsion mask of the registration pattern onto the quartz using a G-line stepper. Thus, the distortion of the stepper is captured by the quartz mask, which is subsequently transferred to the X-ray masks. Figure 34 shows the contact printing scheme used to transfer these patterns to the X-ray mask. For this step, 200 nm of PMMA was spun on top of Au plating base on the X-ray mask. 10 μm glass fibers were placed on the Si mesa of the X-ray mask to help bring the two masks out of contact when the exposure was complete. The quartz

mask was then positioned and weighted down onto the X-ray mask. A small amount of nitrogen pressure was used to bring the membrane into contact with the chrome/quartz surface. A green light was used to monitor the quality of contact as the monochromatic light interferes between the membrane and quartz mask forming dark and light fringes. DUV at 220 nm was used to expose the PMMA.

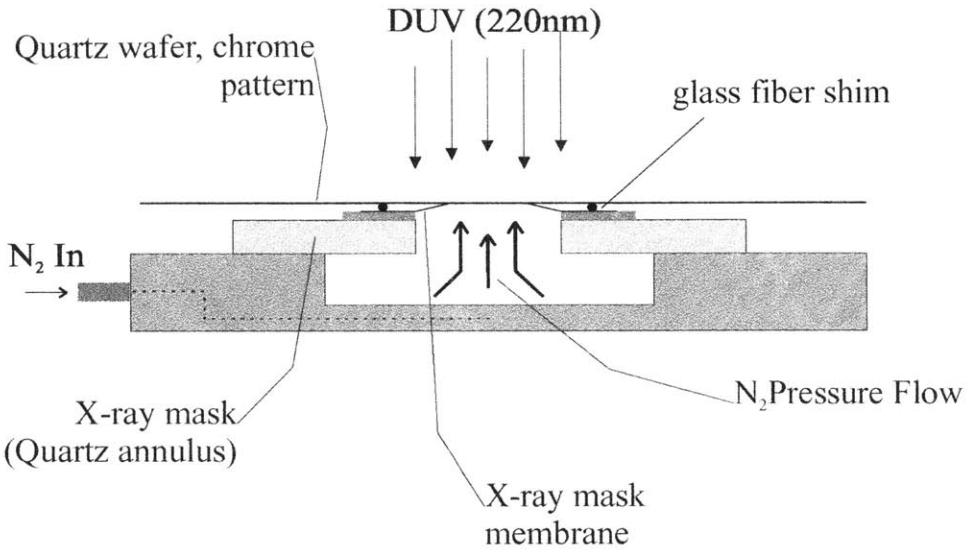


Figure 34 Cross-section of contact printing scheme used to pattern registration pattern on X-ray mask membrane.

This step can cause the membrane to stick to the quartz mask and rupture. Rinsing the quartz mask in a sequence of acetone – methanol – DI seems to improve the ability of the two surfaces to delaminate. After the PMMA was developed, the registration pattern was electroplated in Au. It is important to note that the PMMA must be stripped and re-spun prior to further processing of the X-ray mask. This is because the UV exposure can cause deterioration of the PMMA in regions close to the registration marks.

As a final note, recall that the upper-gate mask contains e-beam written alignment marks and device patterns; the lower-gate mask contains only e-beam written alignment marks. With regard to this, there are three subtle points that highlight important aspects of registration and the mix and match scheme. Because a right-handed and left-handed version of the registration pattern is

needed for each mask, two distinct quartz registration masks are needed for each of the two X-ray masks. A second quartz mask is made by replicating the first onto a quartz blank by means of contact printing. By contact printing the registration pattern in this fashion, the original registration pattern switches from right-handed to left-handed and thus conforms to the requirements of the X-ray masks. The second subtlety is that by using registered e-beam patterning to create alignment marks on the X-ray masks, it is guaranteed that the alignment marks will overlay between the two masks. This overlay suffices to overlay the alignment marks to better than 0.1 μm based on estimated cumulative e-beam registration errors. However, it is not precise enough to guarantee overlay between upper and lower-gate features. It is estimated that the cumulative intra-field distortions of the e-beam system are as large as 30 to 40 nm[1]; it is for this reason that the X-ray mask replication scheme is employed. Finally, when designing the mix and match scheme, it is important to select the appropriate quartz wafer for each of the two X-ray masks. This is because if the wrong quartz mask is selected for the upper-gate X-ray mask, the stepper distortion will be flipped relative to the mask, leading to a loss of distortion matching in the optical printing steps. The logic in selecting the appropriate quartz mask is as follows:

- The master quartz mask captures stepper distortion that will be present in optical wafer patterning after the flip and bond steps.
- The distortion captured by this quartz mask is flipped twice: once by contact print to the X-ray mask and again by printing the X-ray mask onto the device wafer. Hence, printing this X-ray mask will yield distortion-matched patterns on a wafer (correct handedness).
- Therefore, the master quartz mask should be used for the upper-gate mask since the upper-gate mask is used after the flip and bond, when optical patterning steps take place.

[1] This error level assumes conventional registration schemes. Nanometer level intra-field distortion can be achieved but this does not address global pattern placement accuracy.

4.3 E-beam Calibration, Patterning and Electroplating

4.3.1 E-beam Calibration & Patterning

After the e-beam registration pattern is electroplated onto the masks, they are ready for e-beam patterning and electroplating. Recall that only alignment marks are written on the lower-gate mask while alignment marks, gates, pads and other test structures are written on the upper-gate mask. Prior to e-beam patterning, the masks were stripped and cleaned in hot NMP (70° C) followed by acetone, methanol and DI. 3% PMMA was spun to between 230 and 250 nm and baked at 140° C for 90 minutes. The e-beam tool is an IBM VS2 system running at 50 KeV. Colloidal gold was deposited near opposite corners of the die for focusing. It was found that the focus setting was indistinguishable between the two deposits. This implies that the membrane was level to within the effective electron beam waist or focal depth. The minimum beam diameter is estimated to be 50 nm for this system. Unfortunately, beam shape was not consistent and it was difficult to control when patterning sub-100 nm lines. The stigmation may favor either horizontal or vertical linewidth depending on whether the oblong beam cross-section is oriented horizontally or vertically. Therefore, it is advantageous for the layout to contain both horizontal and vertical devices and SEM lines.

Figure 35 shows a plot of vertical and horizontal linewidths from a rather successful write in which poor stigmation was not particularly problematic.

	Beam Current	Pixel Freq.	Repetition
Fine Lines (50-120 nm)	50 pA	240 kHz	4
Coarse Lines and Boxes (>200 nm)	1.14 nA	750 kHz	1

Table 3 E-beam parameters used for X-ray mask patterning.

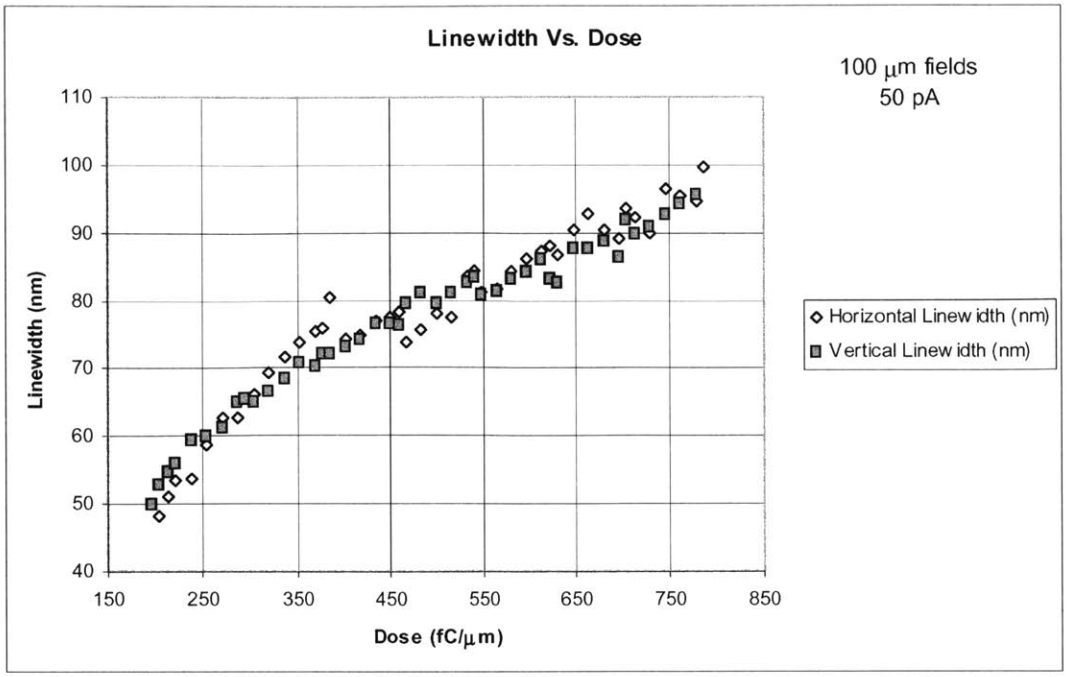


Figure 35 Plot of linewidth versus linear dose. Measured linewidths in electroplated Au within PMMA trenches defined by e-beam.

E-beam parameters used to pattern the masks are listed in Table 3. Fine lines are written as a row of single pixels and are consequently referred to as single-pass lines. The beam, however, can be swept multiple times in order to smooth over beam noise. Figure 36 shows a sample of nested L's after the plating step while Figure 37 shows one line at high magnification revealing some of the gold texture and roughness. Figure 38 is a similar micrograph showing the calibration patterns for coarse features (0.20 μm and up).

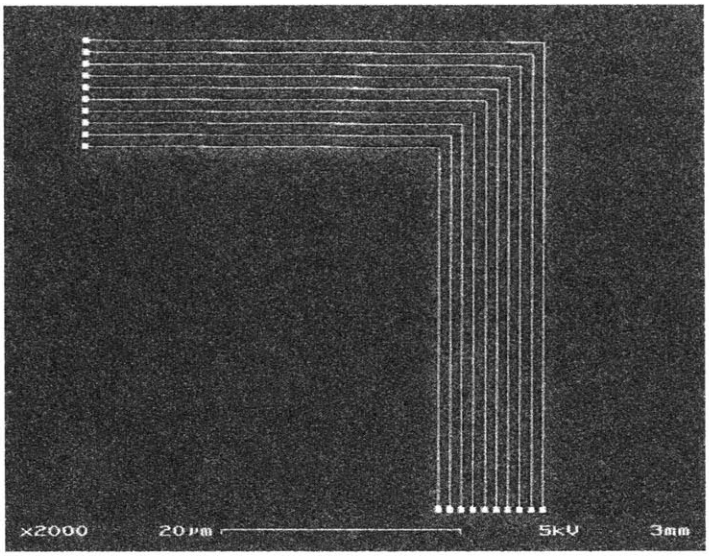


Figure 36 SEM image of gold lines on mask membrane after e-beam patterning, developing and Au electroplating. The lines are written horizontally and vertically with varying dose.

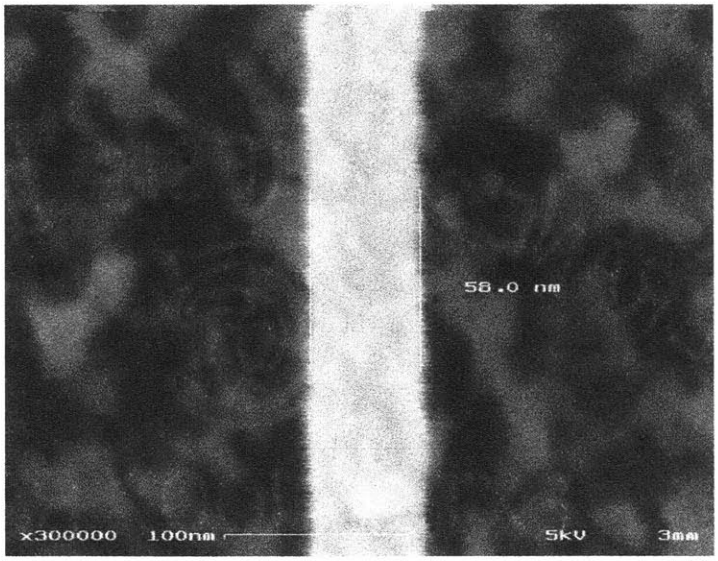


Figure 37 E-beam written line at 60 nm. Image reveals good quality gold grain and low degree of roughness.

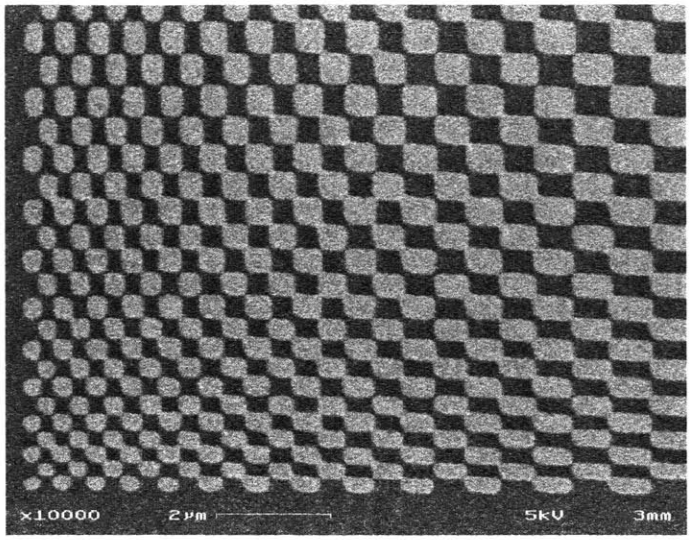


Figure 38 Coarse feature calibration. This 2-D chirped grid is used to determine optimal e-beam parameters for writing patterns at 0.25 μm and larger.

Experiments were also conducted on a Raith e-beam tool in order to investigate the possibility of achieving very fine patterns on X-ray masks. Although a number of technical difficulties inhibited timely incorporation of this system into the process, dose calibration as well as a demonstration of Au electroplated features down to 25 nm were performed. Figure 39 contains a plot of linewidth versus dose from one such experiment. In this system, linewidth was varied by adjusting linear e-beam dose as well as varying the width (in pixels) in a given line. Figure 40 shows a high resolution result from this e-beam system using the same electroplating scheme.

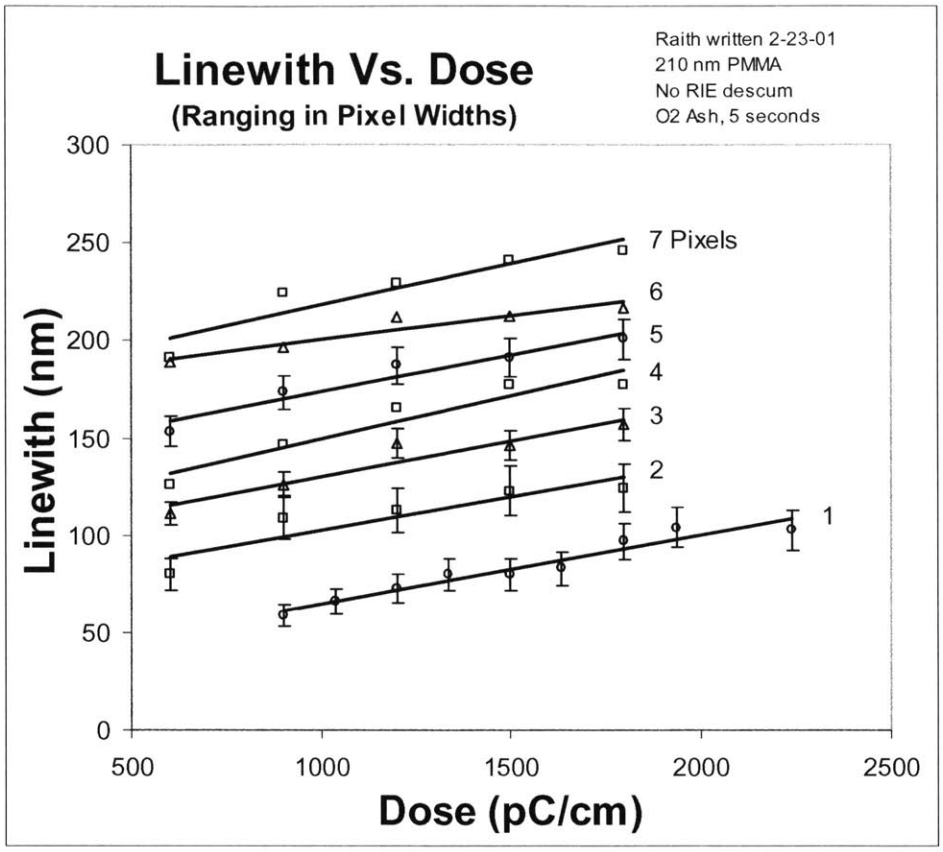


Figure 39 Linewidth versus dose amongst various pixel widths. Patterns written in PMMA and Au electroplated.

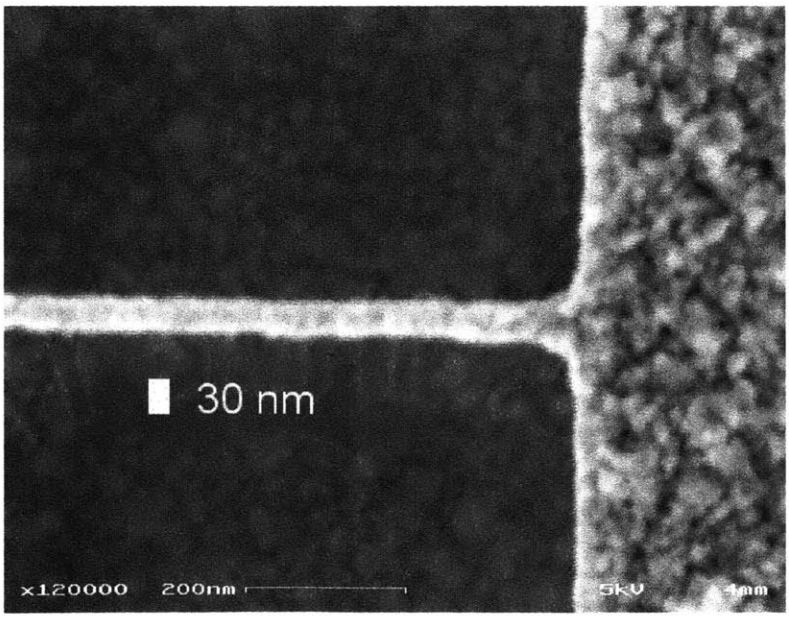


Figure 40 Electroplated Au line and pad showing results at 30 nm.

4.3.2 Electroplating

Good quality mask absorber is necessary to achieve proper profiles of patterns in resist. Since gold density is affected by grain structure, plating rates and solution conditions should be monitored and adjusted to maintain small, tightly packed grains. Furthermore, the thickness of the gold should reach approximately 200 nm in order to achieve a factor of 10 attenuation for Cu_L (X-ray wavelength ~ 1.3 nm). Experiments revealed that electrodeposition rates may be a function of the aspect ratio of the plating mold. For example, under certain bath conditions, fine lines (below 100 nm) plated as much as 50% less than coarse features. This effect is presumed to be largely a function of the age of the plating solution. More precisely, the various constituents of the solution should be maintained at appropriate concentrations.

4.4 Aligned X-ray Mask Replication

In order to replicate X-ray masks without pattern reversal[2] using X-ray lithography, one must choose from several replication methods. The technology available for producing absorber on X-ray masks in NSL is gold electroplating. Therefore, the mask replication should result in trenches of some non-conducting medium with a layer of plating base beneath. The trenches can then be used as a plating mold in which a conducting layer exists at the base providing an cathodic surface for electrodeposition.

The most straightforward approach would be to use a negative resist which would directly yield trenches on the lower-gate mask. A number of resists were studied for this application including NEB-22, SAL-601 and HSQ, but a process window could not be found that produced the desired 50 nm trenches with vertical sidewalls. This is unfortunate because the liftoff scheme is not only more complex, but it imposes more change in stresses on the membrane because of the necessary removal of the resist film during liftoff. Nonetheless, a reversal scheme using liftoff was employed, as described in the next section.

---

[2] Here, pattern reversal means that a feature on the original mask would result in the negative of that feature on the copied mask. For example, Au absorber lines would become trenches on the replicated mask. Thus, a non-reversal process would produce lines from lines and trenches from trenches.

4.4.1 Preparation

Prior to the following process steps, the lower-gate mask is prepared according to the mask process outlined in the previous section. Thus, it contains electroplated Au patterns consisting of registration marks and IBBI marks. Once the mask is stripped and cleaned, a film stack containing SiO_2, ARC and PMMA is put onto the existing plating base as shown in Figure 41. Absent from this cross-section are the existing alignment mark patterns that lie beneath the stack in various locations on the membrane. These so-called *alignment mark regions* require special processing as described later.

The plating mold is constructed from the ARC layer by etching trenches using an oxygen reactive ion etch. Since the O_2 RIE etch easily sputters gold, a thin SiO_2 etch stop is placed between the ARC and the plating base. PMMA is the resist of choice because of its high resolution in X-ray lithography. The liftoff step described below is designed to dissolve the PMMA rapidly without harming the ARC layer. The upper SiO_2 interlayer serves two purposes. It protects the ARC layer from the solvent used during liftoff in addition to serving as an etch mask for the ARC.

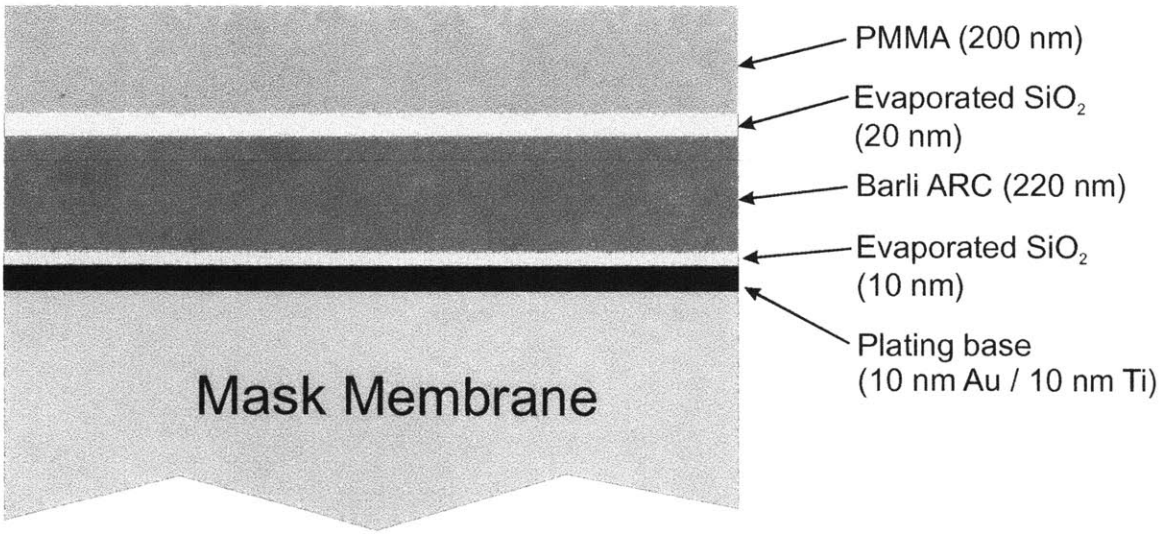


Figure 41 Mask membrane cross-section showing initial stack on the original plating base. (Not shown are existing electroplated Au patterns elsewhere on the membrane.)

4.4.2 X-ray Exposure For Replication

At this point, the mask is ready for the aligned replication exposure. Laser illumination from the IBBI scopes passes through the ARC-oxide-PMMA layers on the lower-gate mask. Figure 42 depicts this aligned exposure and shows how lines on the upper-gate mask transfer as lines in positive resist (as opposed to trenches). Figure 43 is an SEM micrograph showing an X-ray exposed PMMA line on a similar oxide-ARC stack. From here, it is necessary to reverse the pattern (into trenches) so that a plating mold can be made.

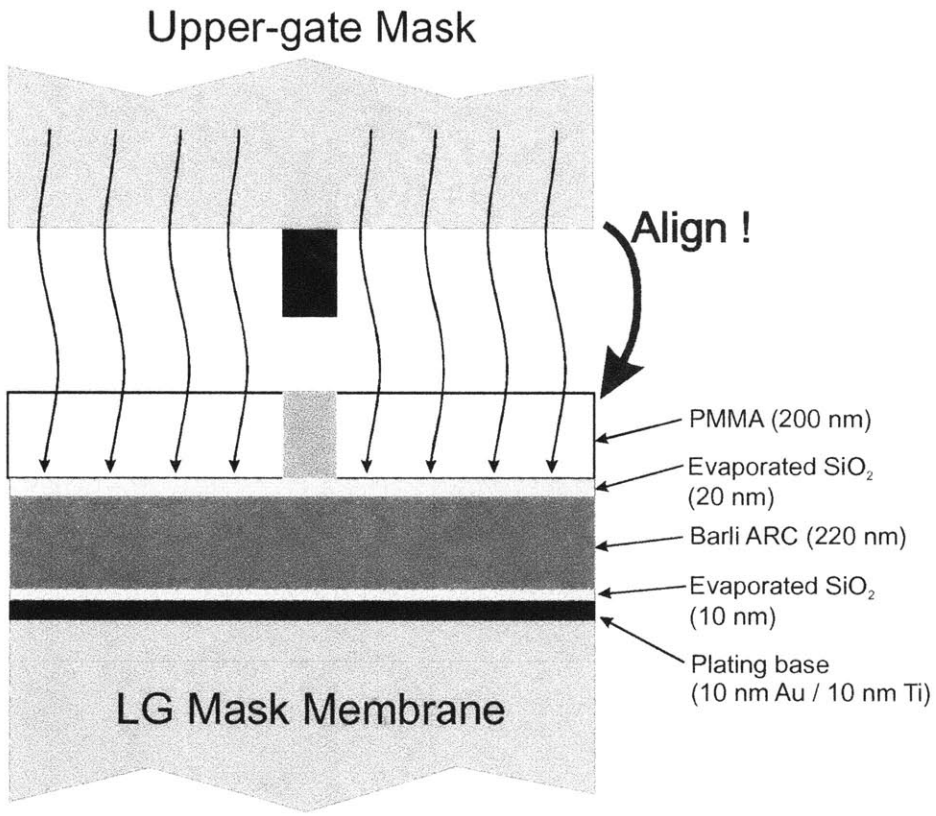


Figure 42 Align upper-gate mask to lower gate mask. Expose and develop PMMA.

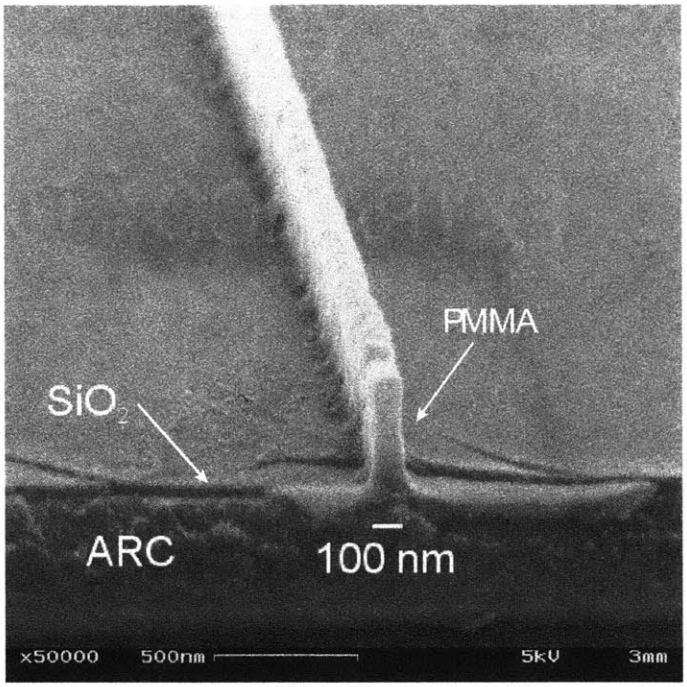


Figure 43 SEM micrograph of PMMA resist on SiO_2 stack. The base of the PMMA contains some unwanted roughness (webbing).

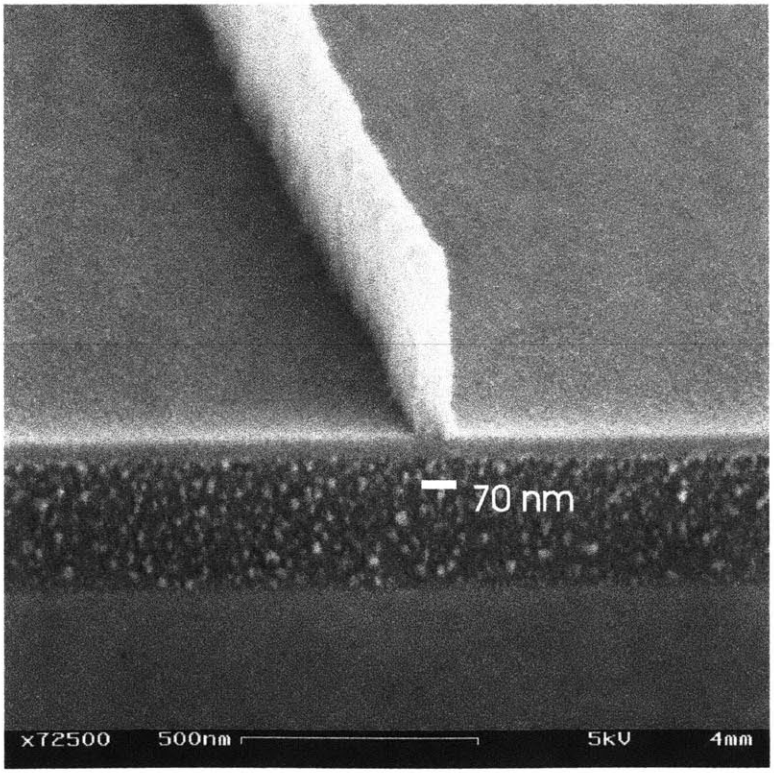


Figure 44 Use of UV5 resist was also explored as a candidate for transferring liftoff patterns to stack. Although good quality profiles could be attained, the resist behaved poorly during liftoff compared to PMMA.

The sidewall angle of the resist is a critical process parameter to control and maintain. Non-vertical profiles can arise from X-ray diffraction and penumbra.

4.4.2.1 X-ray Diffraction

The minimum printable feature size, W_{min}, can be related to the gap, G, the X-ray wavelength, λ and a scaling parameter α as modeled in the following equation

$$G = \frac{\alpha (W_{\min})^2}{\lambda}$$

One must theoretically and experimentally determine the appropriate value for alpha. This factor varies based on the electromagnetic models used to predict the intensity of radiation at the wafer. X-ray radiation can be emitted from various types of source and pass through intermediate media such as vacuum windows and low-Z gases. The X-ray mask will introduce a spatially dependent attenuation and phase shift to incoming photons. The sources used today can vary in brightness uniformity over solid-angle, spectrum profile and coherence. The models used to calculate diffraction effects can therefore become quite complex making it difficult to accurately predict the dose profile in the resist. However, it is likely that values of alpha greater than 1.5 reflect the true constraints on gap. Diffraction effects become convolved with penumbral blur, (as explained in the next section), which has a beneficial effect on resolution. Figure 45 plots the required gap for minimum linewidths up to ¼ micron. Assuming an alpha of 1.5, features between 50 and 100 nm correspond to gap requirements of 3 to 10 μm respectively. Typically achievable gaps range from 3 to 10 μm when using a stud-controlled gapping approach or when implementing dynamic alignment. Gap is most affected by wafer topography and particulate contamination. Since gaps between 3 to 5 μm were achieved in the mask replication and wafer patterning exposures, these exposures are probably operating close to diffraction limits for linewidths in the 50 nm range.

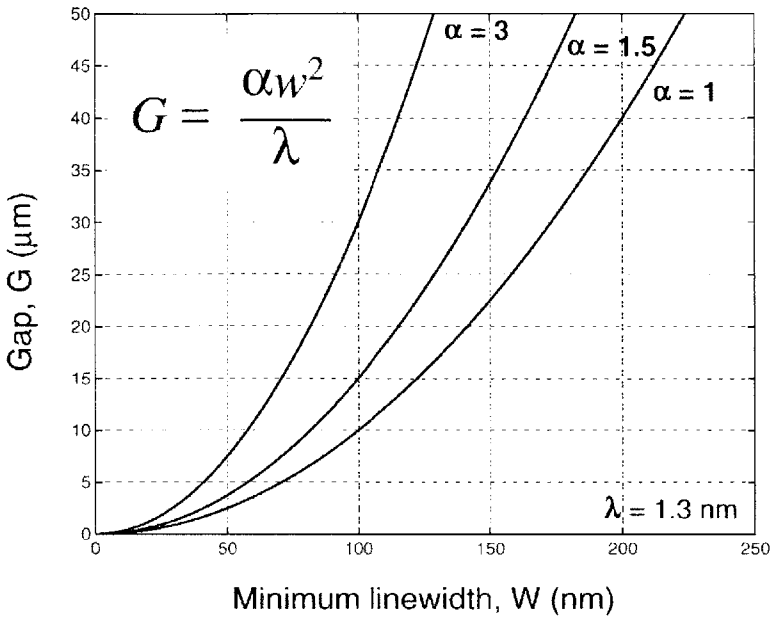


Figure 45 Plot of maximum possible mask-wafer gap as a function of minimum printable linewidth for an X-ray wavelength of 1.3 nm and α of 1, 1.5 and 3.

4.4.2.2 Penumbra

Penumbra is a measure of the non-sharpness of shadow edges cast by the X-ray mask absorber onto the substrate. This edge blur arises from the fact that the X-ray source used in these experiments is not a true point source but rather a small spot.

In these experiments, X-ray emission comes from a spot on the anode with a typical diameter of about 1 to 1.5 mm. In practice, the spot is usually elongated in a particular axis. For example, on one particular anode, the spot was imaged with a blue filtered CCD camera[3]. It was observed to be roughly 1 mm wide by 1.5 mm long. Essentially, multiple rays of illumination can be traced from the source towards a particular point on the mask. The result is a transition region at the edge of all patterns where the X-ray intensity climbs from its minimum value well beneath the absorber to its maximum value well clear of the absorber. The distance over which this transition takes place, in addition to the response of the resist to this varying X-ray dose, results in a mechanism for sloped sidewalls to occur. The size of this blur can easily be calculated based on the geometry of

the exposure system. Based on mask-sample gap, G, distance of mask-to-source, D, and the source spot-size, d, the penumbral blur length[4]

$$penumbra = G\frac{d}{D}$$

Given typical values of gap (3 μm), source size (2 mm) and source-to-sample distance (28 cm), the penumbra is estimated at 20 nm. If the resist height is 200 nm, the approximate sidewall angle in the dose profile 0.1 rad. This would be the angle of the resist sidewall if the resist was perfectly linear and one disregards the fact that upper portions of resist experience more develop time than lower portions. If the sidewall angle is 0.1 rad and if the metal thickness in the liftoff step is 7 nm, than the sidewall coating is 0.7 nm. This thickness of Ni should not interfere with one's ability to remove the patterned metal. If the gap is doubled to 6 μm, the predicted sidewall coating doubles to 1.6 nm. Again, because of resist roughness, this is not expected to cause a problem with liftoff. More accurate sidewall profiles are found experimentally or through simulations that incorporate the non-linear effects of the resist.

4.4.3 Metal Evaporation and Liftoff

Figure 46 and Figure 49 display the metal evaporation and liftoff step. For this step to be successful, it is necessary to ensure that the Ni does not coat the resist sidewalls. It is also necessary to remove the PMMA and Ni without collapse or re-deposition of the metal. This must be accomplished without harming the underlying ARC which can be attacked by NMP, especially at higher temperatures.

---

[3] Here we assume that most of the X-ray emission comes from the blue glow on the anode surface.

[4] Note that penumbra should be calculated in both the x and y axis as it can differ based on the orientation of an oblong source emission spot.

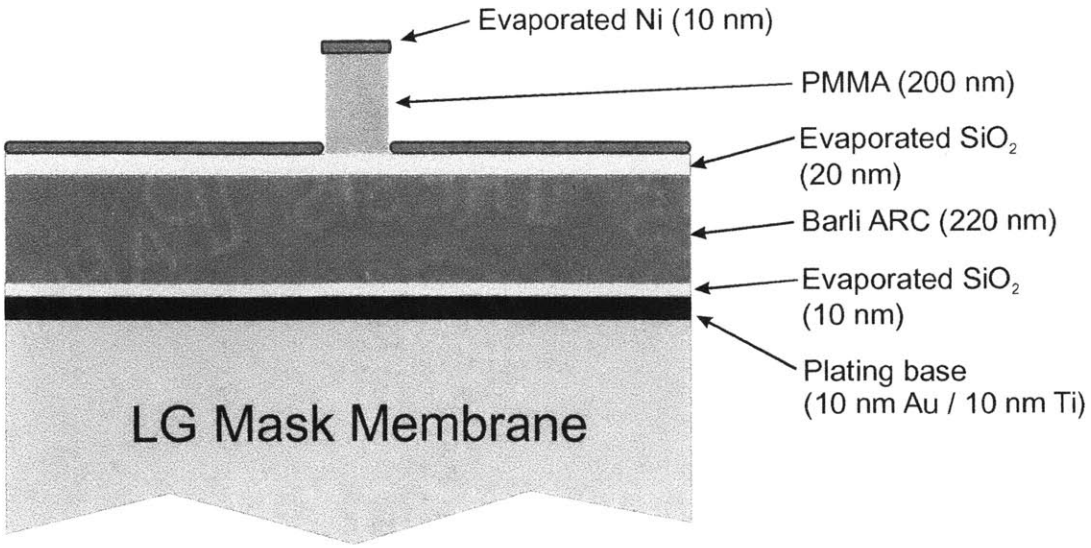


Figure 46 10 nm Nickel is evaporated onto the mask.

Figure 48 is an SEM micrograph taken after a Ni evaporation onto a PMMA line. It is evident in this view that there is a clear discontinuity between the Ni deposit on the substrate versus the deposit on the PMMA. Figure 47 shows an example of a failed liftoff attempt in which case the Ni re-deposited adjacent to the line. It is possible that a sidewall coating contributed to this failure.

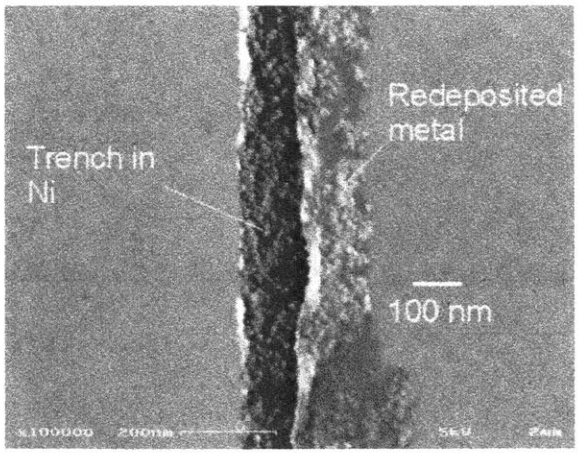


Figure 47 Re-deposited metal from a failed liftoff attempt.

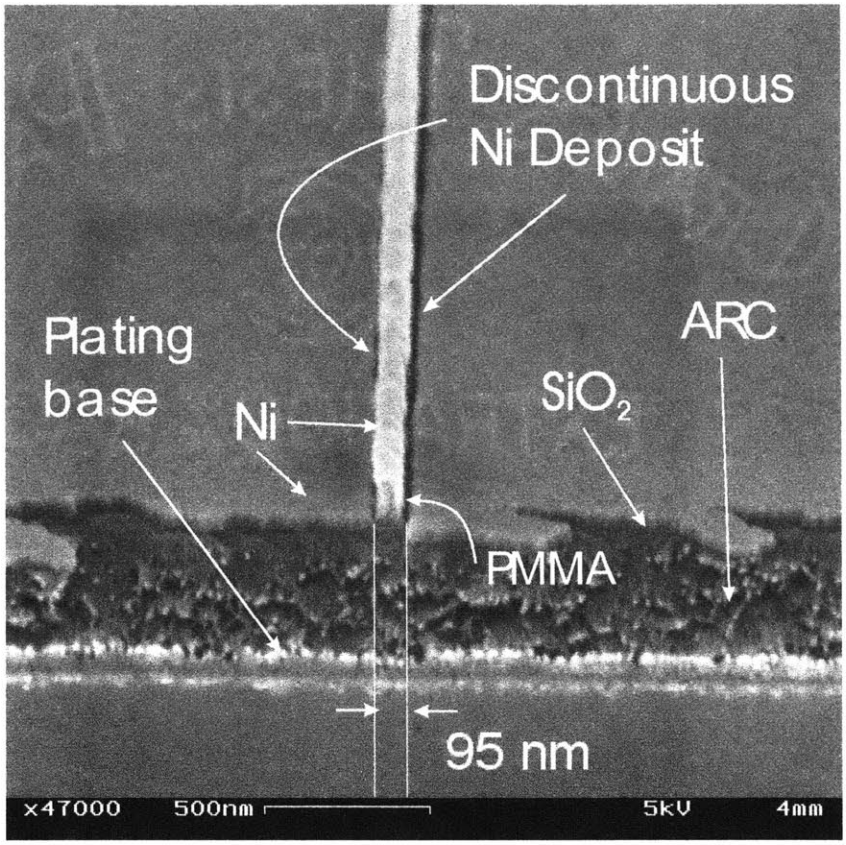


Figure 48 SEM micrograph of PMMA line with evaporated Ni for liftoff. It is fairly clear in this image that there is no sidewall coating of metal on the resist line.

One of the disadvantages of introducing the liftoff step is the expansion of linewidths. Based on measurements made of X-ray exposed resist lines and the resulting trenches in liftoff metal, the typical enlargement is approximately 15-20 nm. One reason this occurs is because a lateral metal accumulation occurs on the upper edges of the resist. During metal deposition, this overhang shields the edge of the metal trench from forming. This can be seen in Figure 48. Furthermore, the subsequent etch steps can only aggravate the problem since some degree of isotropic etching can occur. It has been observed that a 70 nm line on the upper gate mask can expand to a 100 nm line on the lower gate mask. If this enlargement is unacceptable, a corrective method can be applied in the wafer process as discussed in section 3.3.1.1.

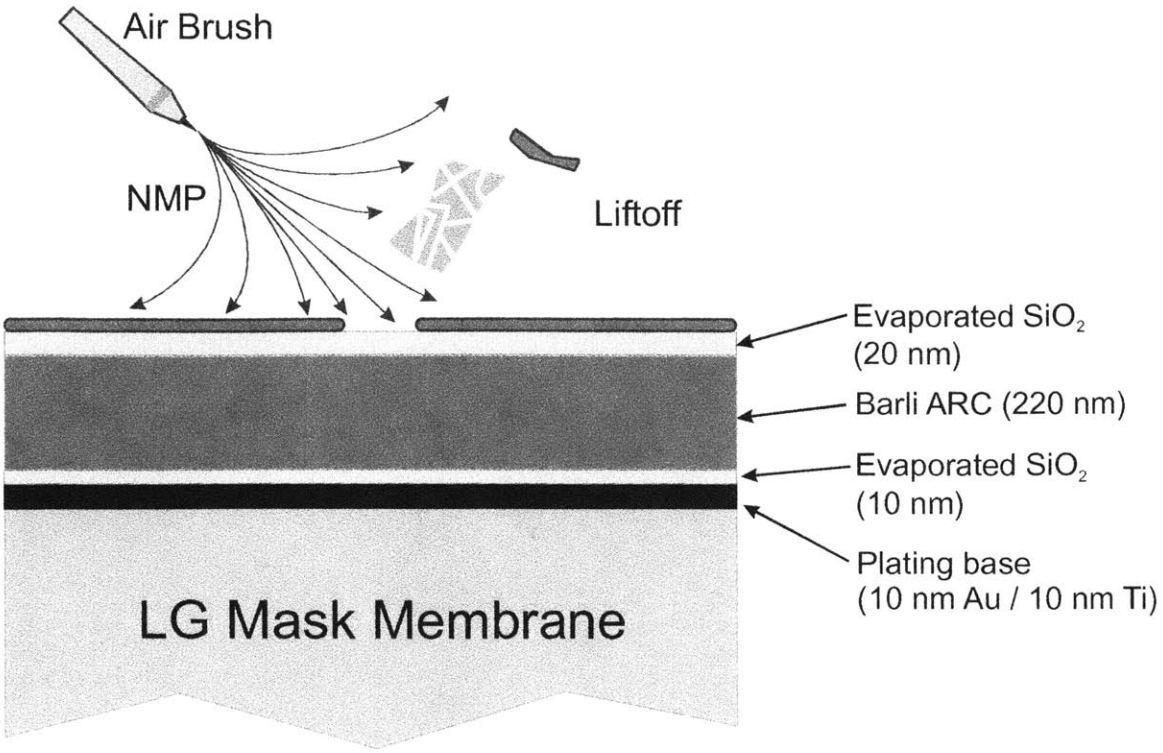


Figure 49 Liftoff: spray technique. NMP at room temperature is sprayed at PMMA features from a distance of about 1 cm. This very effective liftoff technique should be used in conjunction with immersion in NMP at 60° C which helps remove large area patterns. In this step, it is necessary to remove the PMMA (and Ni) without attacking the underlying ARC.

In Figure 51, the oxide interlayer etch is shown along with RIE parameters and Figure 52 shows the cross-section after Ni removal using a wet etch. At this point, there is usually a noticeable residue that appears on the surface. However, it does not interfere with the subsequent ARC etch which actually eliminates this contamination. Figure 50 shows a top-down SEM micrograph of lower-gate X-ray mask after Ni liftoff showing successful transfer of 80 nm feature.

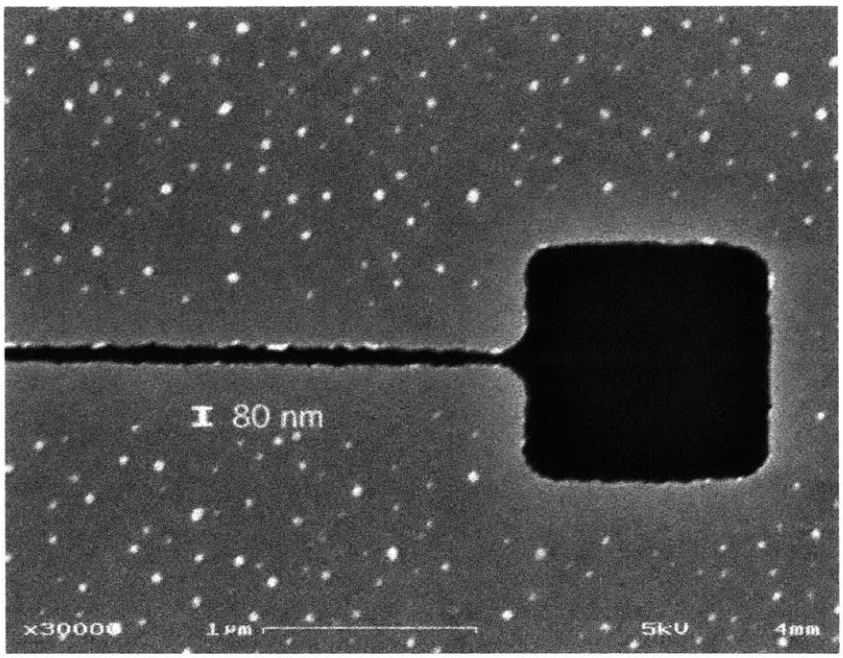


Figure 50 Top-down SEM micrograph of lower-gate X-ray mask after Ni liftoff showing successful transfer of 80 nm feature.

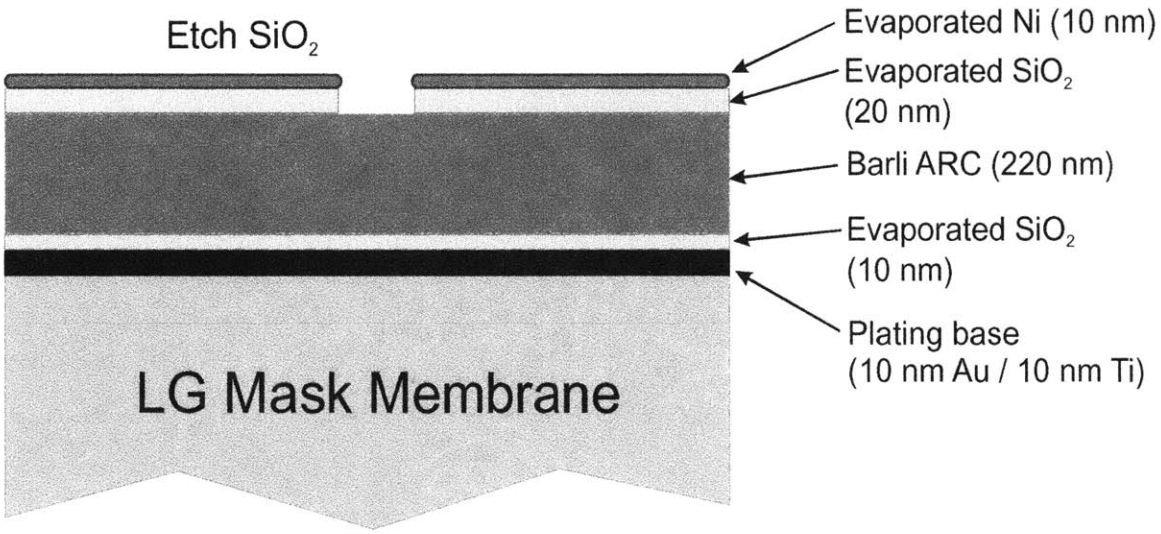


Figure 51 SiO_2 is etched using Ni mask. Recipe: CHF_3 @ 10 mT & 15 scm. DC bias 150 V, power 190W. Time 30-40 sec.

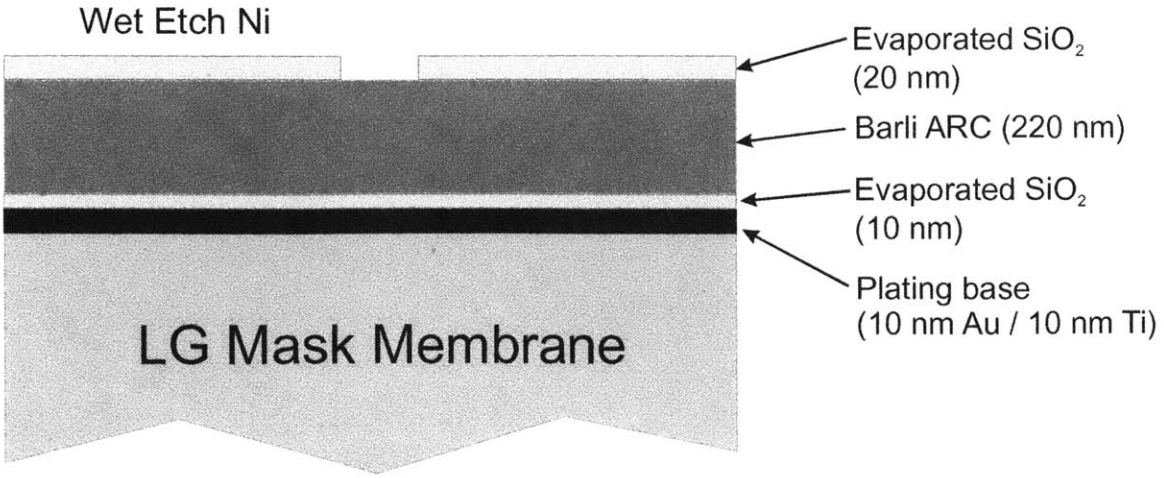


Figure 52 Wet etch Ni in commercially available solution at room temperature with mile agitation. 10 nm Ni typically requires 15-25 sec to completely remove.

In the next step, the ARC is etched in O_2. Figure 53 shows the resulting cross-section along with etch parameters.

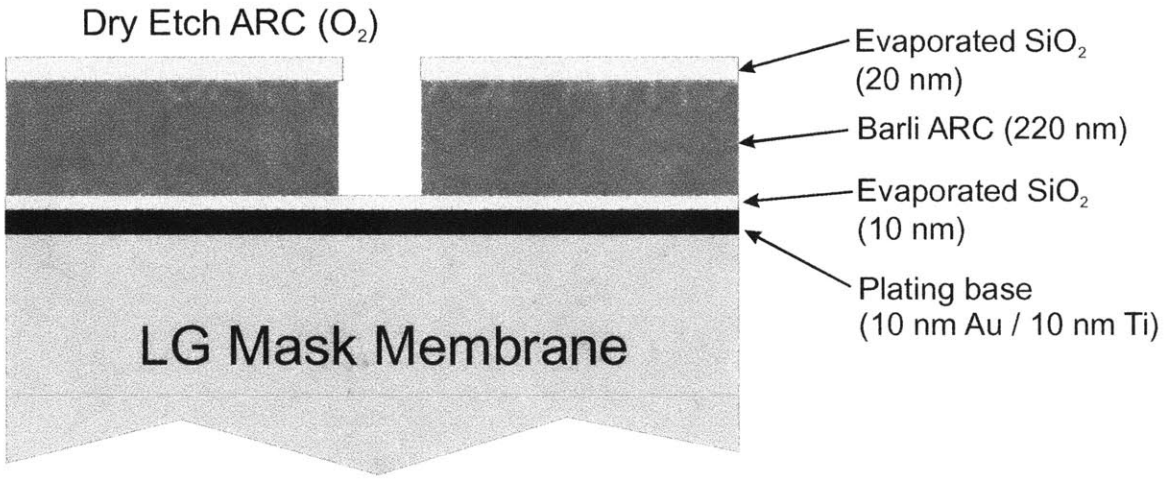


Figure 53 Cross-section after ARC etch using O_2. O_2 @ 5 scm & He @ 10 scm. Pressure 7 mT. DC bias 280 V. Power 280 W. Etch time approximately 4 min.

By producing gratings on the same SiO_2 / ARC stack that is used on the mask, the etching of oxide and ARC was studied on wafers in order to inspect the quality of the ARC sidewalls, the preservation of linewidth and the ability to avoid degradation of the plating base at the end of the ARC etch. Figure 54 displays a comparison of linewidths before and after etching and reveals that etching only contributes approximately 5 nm to the expansion of pattern size.

Figure 55 shows the detrimental effect that plating-base sputtering has on the gold electroplating process.

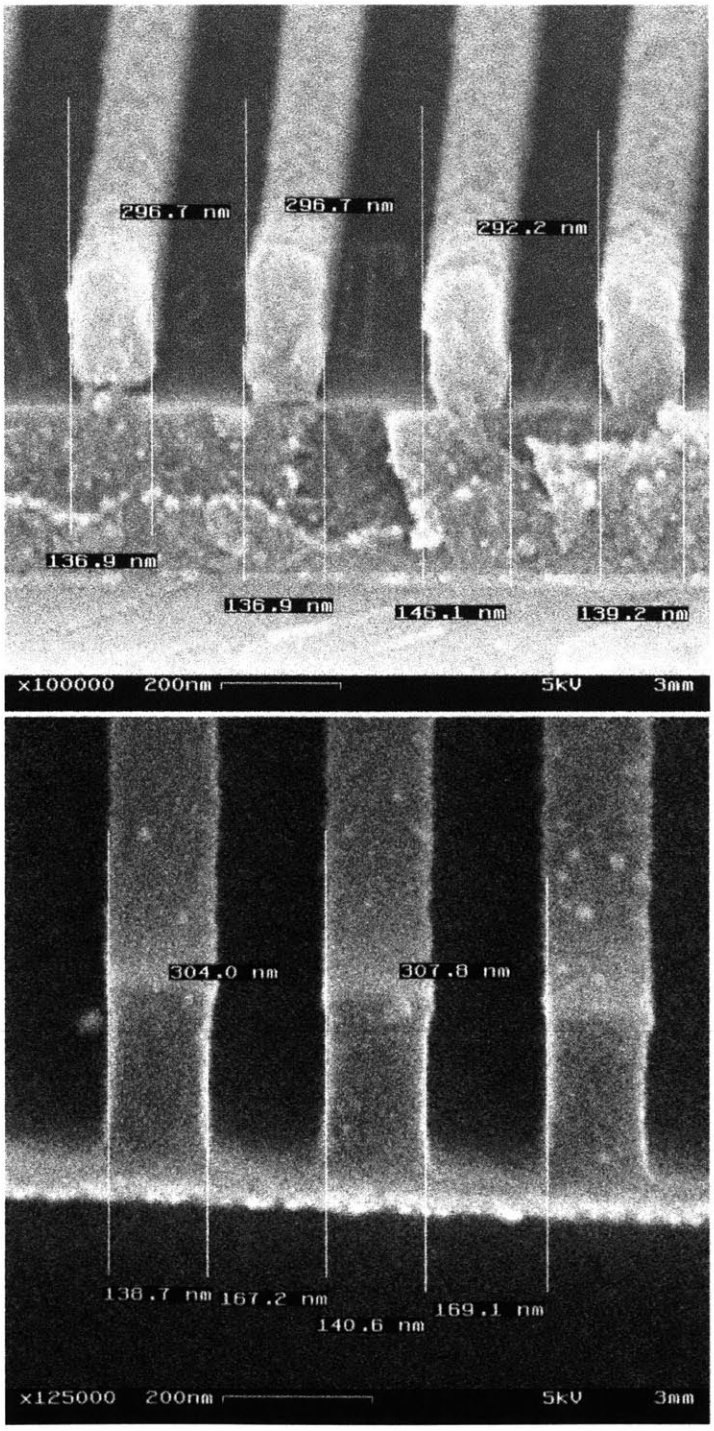


Figure 54 Inspection and measurement of lines etched into oxide / ARC stack. Gratings in resist are produced by interferometric lithography and Ohka THMR-iN negative chemically amplified resist at 325 nm UV. Using uniform gratings with 300 nm pitch, linewidths can be measured before and after the etch to ensure the etch does not expand patterns beyond acceptable limits. Here, it is evident that there is only about 10 nm expansion of spaces between resist lines. Thus, 5 nm of lateral etching is taking place on each pattern edge.

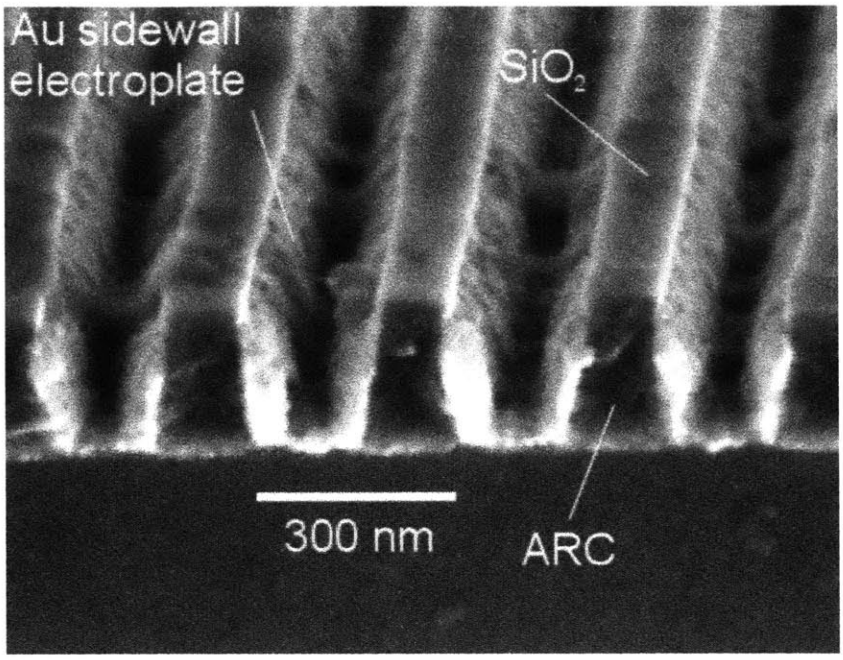


Figure 55 Result of plating-base sputtering after ARC etching using O_2 etch gas. Au electrodeposit accumulates on sidewalls because plating base is present on sidewalls prior to plating. This problem is alleviated by introducing a protective oxide etch stop just below the ARC layer.

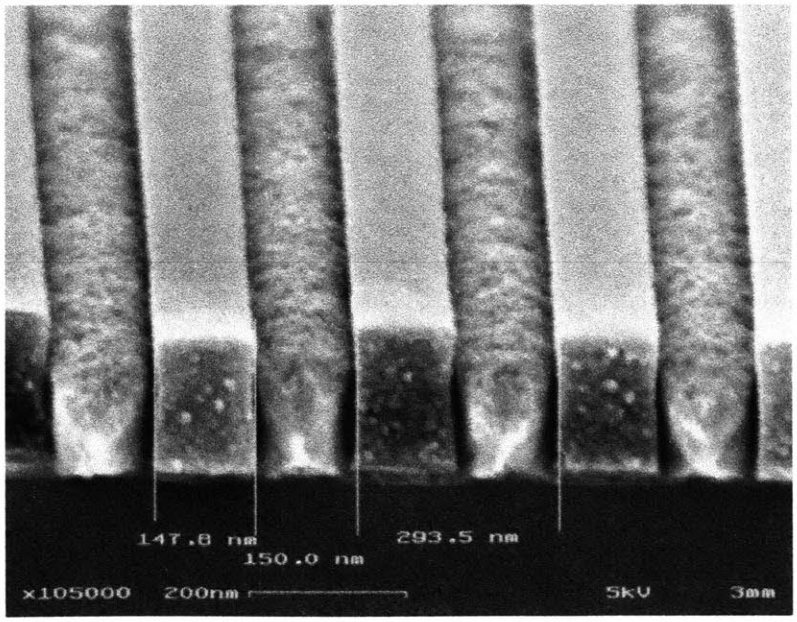


Figure 56 Electroplating results in trenches formed by IL grating. This demonstrates the efficacy of introducing a thin oxide etch stop between the ARC and gold plating base.

In this sample, the sputtered gold caused the plating base to be transferred to the sidewalls such that electro-deposition occurred on the ARC walls instead of growing from the base of the trenches. Figure 56 is a micrograph of the same process step except that the oxide etch stop layer is used to prevent sputtering.

Up to this point, the process flow drawings (Figure 51-53) have shown the mask cross-section through a device area. We have not yet considered how process steps have affect the alignment mark regions. The alignment marks, which exist on the lower-gate mask prior to replication, must not be altered. Figure 57 displays how the process has affected the alignment mark areas. The trenches in the ARC over the alignment marks are a result of the image produced by the top-gate alignment marks during the aligned exposure. Clearly, material must be added to these areas to protect the existing patterns from acquiring additional gold electrodeposit which would render the lower-gate alignment marks useless.

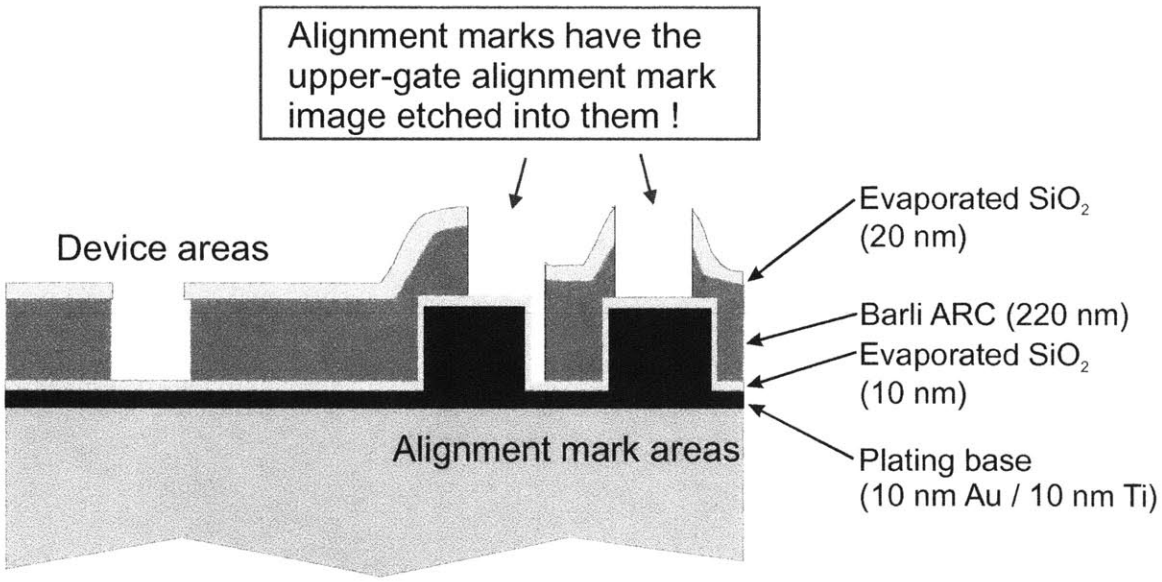


Figure 57 Cross-section that includes the alignment mark areas of the mask. At this point, it is necessary to protect the alignment marks from further processing so that no Au deposit it added to them.

The best solution is to pattern a PMMA protective film over the alignment marks. Photoresist has proven to be a poor choice since TMAH-based[5]

[5] TMAH tetra-methyl-ammonium-hydroxide

photoresist developers attack the lower-oxide interlayer. In doing so, Au will electroplate into a void etched into the foot of the trenches thereby expanding the effective linewidths of the device patterns.

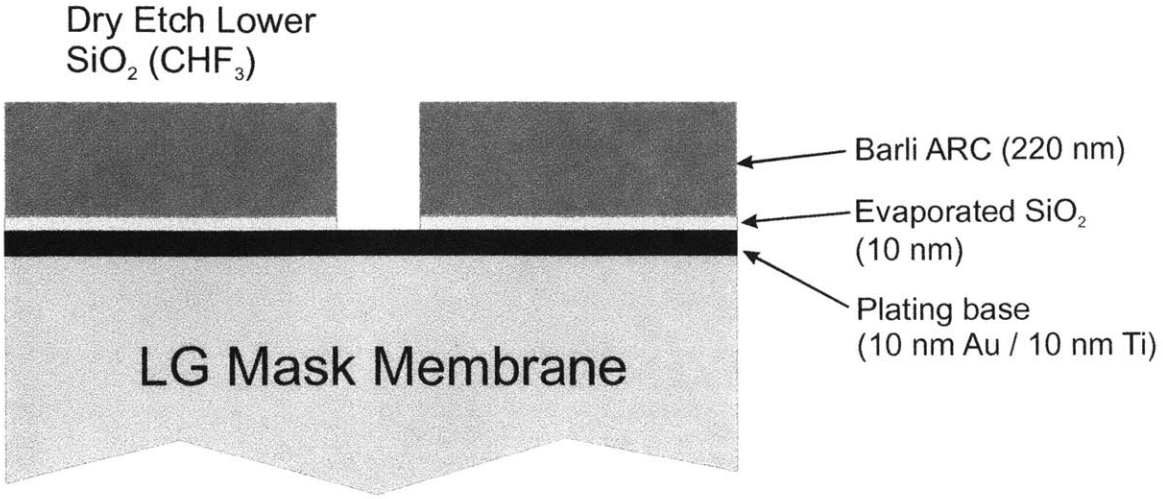


Figure 58 Profile after removal of lower oxide interlayer. Etch conditions are the same as for the upper oxide interlayer with the exception of a reduced duration: 20-30 sec.

Electroplating Au is described in section 4.3.2. Figure 59 shows the final cross-section. It should be noted that electroplating can only be carried out if the Pyrex®-Au mask surface, (the annulus supporting the membrane), is free of deposited films because the Pyrex® surface serves as a consistent surface area for electroplating. This was accomplished through the followings steps. First, ARC was spun onto the mask (and inevitably coated the Pyrex®). Next, the membrane was warmed using a mesa-style hotplate fixture at 90°C. In doing so, most of the solvents in the ARC on the membrane were driven off preventing the film thickness from becoming non-uniform. Meanwhile, the ARC on the Pyrex® remains at lower temperatures due to the relatively slow heat transfer through the Pyrex®. In the next step, ARC on the Pyrex® can easily be swabbed off using NMP, Acetone and Methanol. Once the Pyrex® ring is sufficiently clean, the mask can be baked at 140°C for 1 hour in order to cure the ARC on the membrane.[6]

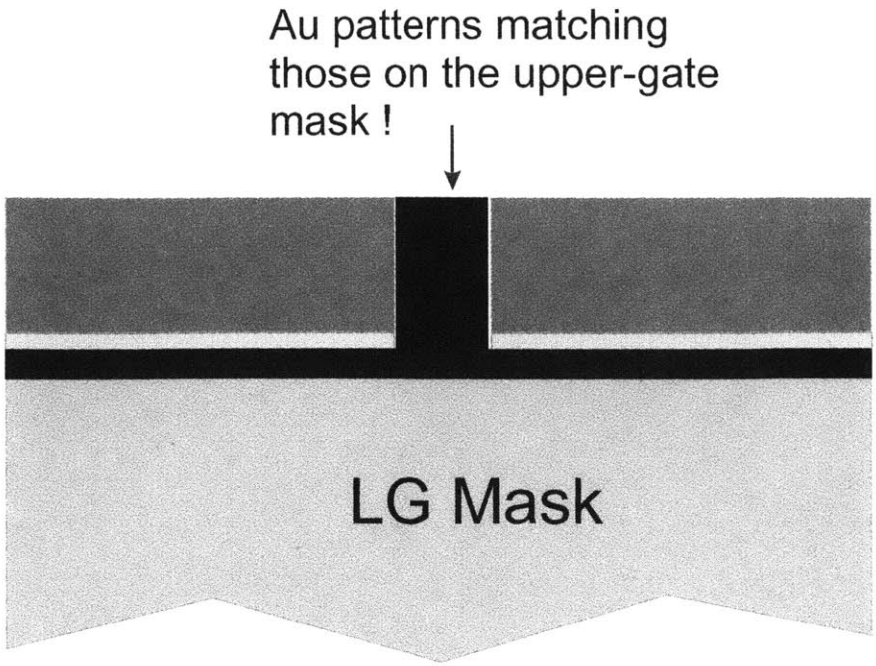


Figure 59 Final cross-section after Au plating. The ARC can be kept on the membrane without adversely affecting X-ray exposures. In doing so, mechanical stress induced by the ARC will remain unchanged reducing the potential for in-plane distortion which would lead to pattern shift and misalignment.

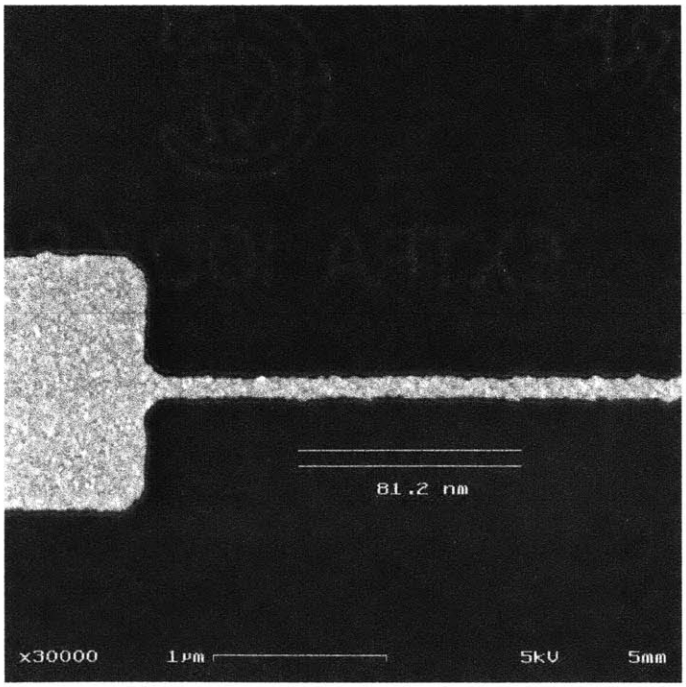


Figure 60 Top-down SEM micrograph of lower-gate mask showing successfully transferred 80 nm gate feature realized in electroplated Au absorber.

---

[6] When ARC is baked on the 10 nm oxide interlayer, the oven temperature must be ramped from 90°C to 140°C (and must not exceed 150°C). This prevents the oxide from cracking.

4.4.4 Other Mask Preparation Steps

After electroplating the lower-gate mask, the ARC may be left in place without hindering the performance of the mask. In doing so, any stress induced by the ARC will remain unchanged. This is desirable since one wishes to preserve the overlay that has been captured in the replication process.

However, in order to comply with CMOS processing guidelines, the possibility of gold contamination from mask to wafer must be eliminated since gold atoms behave as a strong electron trap in Si; therefore, a protective layer of PVA (~150 nm) was spun onto the masks to solve this problem. PVA is a simple solution because it requires no baking and can easily be stripped with water. It also serves to absorb photoelectrons that are a by-product of incident radiation on gold. If allowed to reach the resist, these electrons may affect the quality of exposure. This issue is more prominent when using sensitive negative resists.

5 X-Ray Alignment Results

This section covers the experiments that were carried out in order to obtain the following measurements:

- lower-gate X-ray mask membrane distortion
- repeatability and precision of IBBI alignment
- sensitivity of overlay to changes in gap
- ability to induce distortion corrections thermally

The results from two X-ray mask sets and three wafer lots are discussed below. UG and LG will be used to refer to upper-gate and lower-gate respectively. The mask/wafer descriptions are summarized in the following table:

Mask Set		Description	Wafer Lot	Description
Set 1	Upper-gate	Min. linewidth = 90 nm	Lot 1	Voids present after bonding. Lower-gate patterning gap set high (> 5 μm). LG poly thickness low (100 nm), thicker poly desired for IBBI signal.
	Lower-gate	Min. linewidth = 120 nm (low yield)		
Set 2	Upper-gate	Min. linewidth = 50 nm	Lot 2	No voids present. Lower-gate patterning gap set at approx. (5 μm). LG poly thickness increased (150 nm) and LG linewidths thinned to 50 nm.
	Lower-gate	Min. linewidth = 90 nm	Lot 3	No voids present. Lower-gate patterning gap set to zero. LG linewidths thinned to 80 nm.

Table 4 Mask sets and wafer lots; labels and description.

Initially, alignment experiments between the UG in mask set 1 and LG patterned wafer lot 1 were done at minimal gap. The patterns of interest were located within 300 μm of the central X and Y alignment marks.

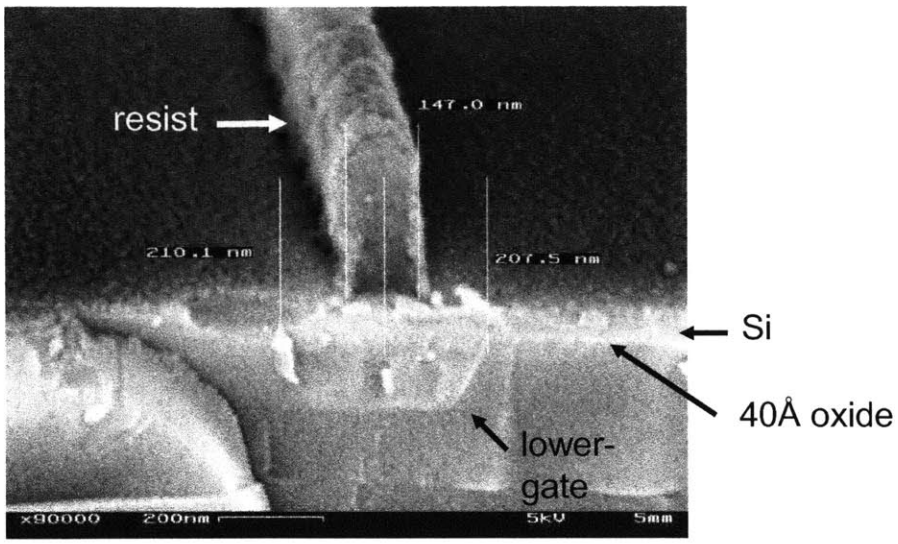


Figure 61 Alignment results after initial IBBI exposure tests. Upper-gate structure is UV-5 resist and lower-gate is poly-Si embedded in LTO. Sub-5 nm alignment demonstrated.

This location of the die (i.e., in close proximity to the central alignment marks), was maintained within the X-ray source axis during exposure. The results show that alignment in this region is good to about 5 nm based on SEM measurements of the structure cross-section. Outside of this region, however, alignment measurements reveal an alignment "run-off". This means that misalignment values vary linearly when measured at equal intervals along a given cut across the die. Experiments were done to pinpoint the source of this apparent distortion.

5.1 Pattern Distortion Measurements

The UG mask of set 2 was aligned to LG patterned lot 2 and alignment measurements were taken across the die. Top-down SEM inspection provided pattern alignment measurements between the UG and LG SEM lines. Figure 62 contains a plot of misalignment in the Y-axis direction of the die. The slope of the line reveals a pattern run-off of about 120 nm/mm. In contrast, pattern mismatch induced by magnification error is theoretically about 6 nm/mm for every 1 μm error in gap[1]. Certainly, this run-off cannot be entirely due to gap error. Some combination of mask membrane and wafer distortion exists. In order to differentiate the two effects, the X-ray mask replication experiment was

"repeated" in order to measure overlay amongst the masks after LG mask processing.

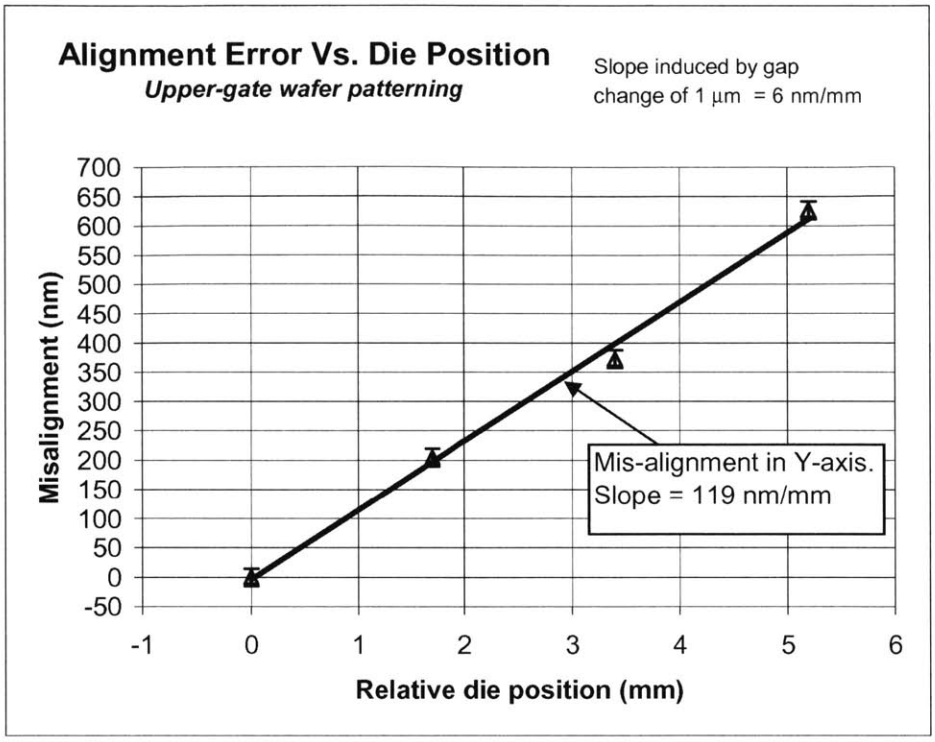


Figure 62 Alignment error measured along the Y-axis direction for IBBI aligned X-ray exposure of DGFET wafer. The slope of the line indicated a pattern run-off of 120 nm/mm.

In this so-called "mask re-replication" experiment, UV-5 resist was spun onto the LG mask of set 2. An adhesion promoter compatible with noble-metal substrates was applied to the existing Au patterns and plating base. IBBI alignment was carried out in the same fashion as the original replication experiment. The gap was set to within a micron of the original gap. The plot in Figure 63 reveals that the pattern run-off on the LG mask is about 26 nm/mm in X and 32 nm/mm in Y. Thus, the LG mask is stretched relative to the UG mask which is likely due to the change in materials present on the membrane. If this expansion is subtracted from that of the apparent wafer expansion, it can be inferred that the wafer of lot 2 had expanded by about 90 nm/mm.

[1] Assuming 28 cm source-to-sample distance. This change in gap can be the cumulative gap error amongst mask replication, LG patterning and UG patterning.

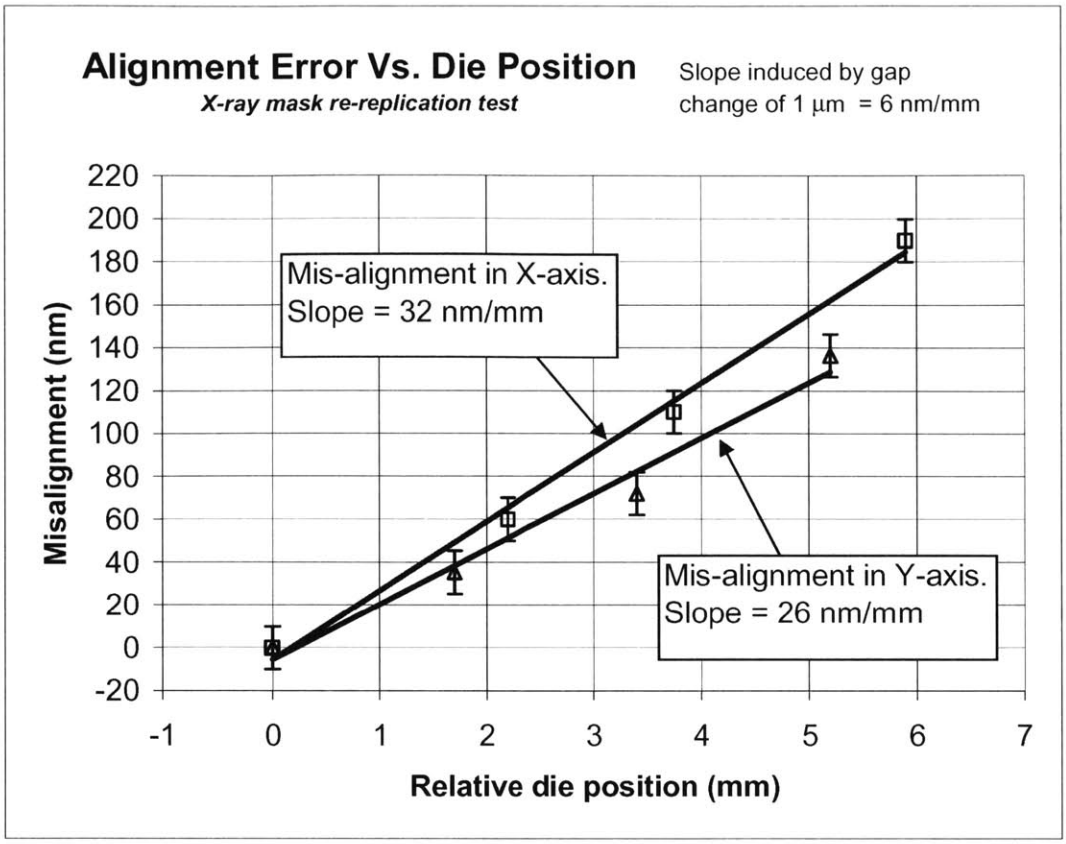


Figure 63 Alignment error measured along the Y-axis direction for IBBI aligned X-ray exposure of LG mask (re-replication test).

In order to pinpoint the cause of the strain of the membrane, we can consider the films that were removed and added after the mask replication exposure. PMMA, nickel, and evaporated oxide were removed. Additionally, the ARC was removed in order to inspect the gold deposit at the base of the lines. In contrast, a region of gold deposit surrounding the die was added during the replication process and is depicted in Figure 64. This region of the membrane had served as an area to monitor etching and electrodeposition quality.

To study the strain contribution of gold deposit, the lower-gate mask was patterned with photoresist such that only this area of excess gold was exposed. A wet etch was performed to completely remove this film and the mask re-replication experiment was repeated. Figure 65 shows a plot of runoff data resulting from the mask re-replication experiment following Au film removal. The

Y-axis slope increased by about 10 nm/mm leading to the conclusion that the Au film was under compressive stress.

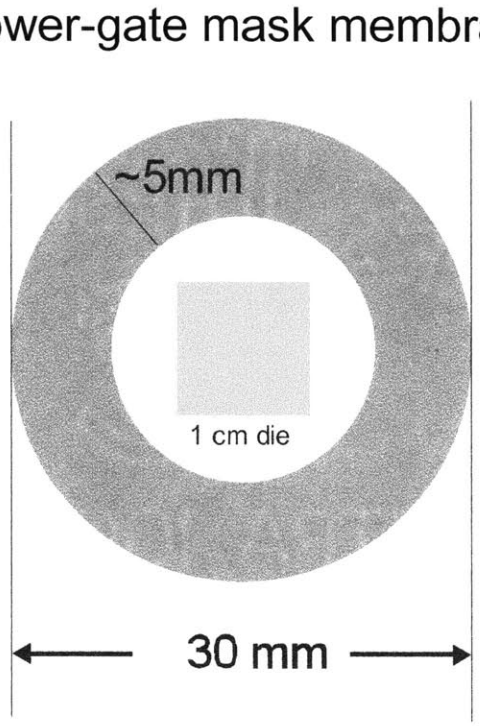


Figure 64 Depiction of LG-mask membrane. An annulus of 5 mm radial extent at the perimeter of the 30 mm membrane was removed and the resulting change in strain was measured.

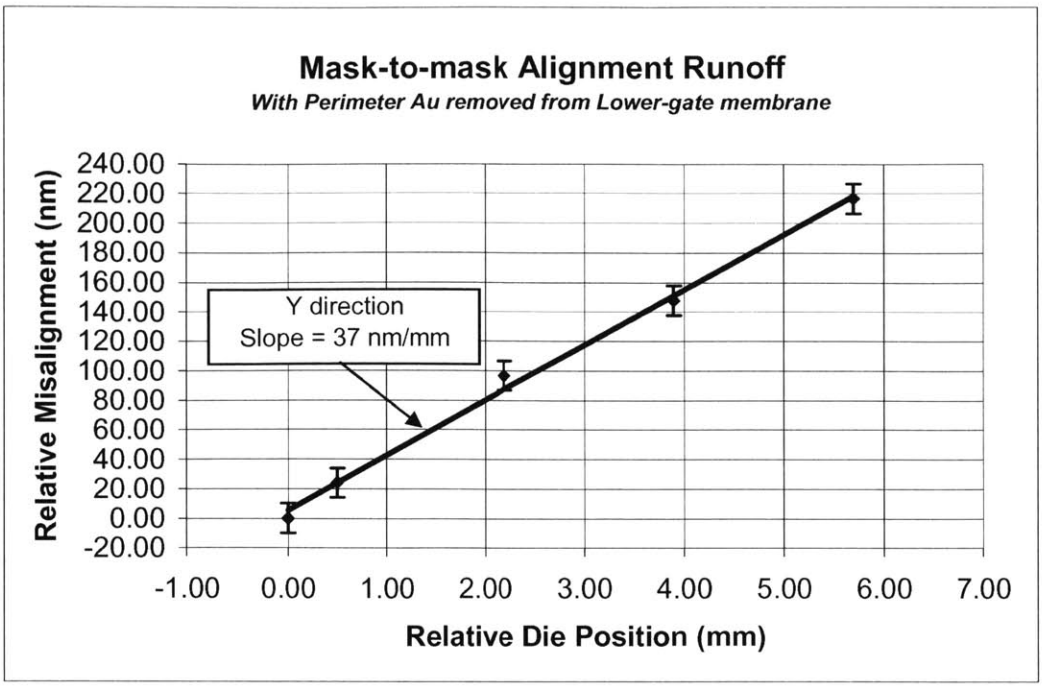


Figure 65 Pattern runoff in the Y-direction resulting from a mask re-replication experiment with Au film removed from the lower-gate mask. Comparing Y-axis strain to Figure 63, the change in strain is about 10 nm/mm.

5.2 Improving the Replication Process: Low Stress Scheme

Based on observations of the strain of the LG-mask, it is evident that the ARC, oxide and/or PMMA invoked tensile stress on the membrane and that by removing these materials, the membrane had relaxed. In contrast, the gold had produced compressive stress. It would be useful to consider the result of the modified replication scheme in which a negative resist is employed to transfer the gate patterns. In this case, the majority of changes in stress are assumed to result from the electroplated gold and to a lesser degree from the dissolution of the resist during developing. In order to estimate the contribution of strain from gold in such a process, the stress of the gold must be calculated. This was done based on the measurements described in the previous section.

In order to estimate the strain in the gold, a MATLAB simulation program written by Ken Murooka in NSL was used. Shown in Figure 66 is the Y-axis displacement over a 30 mm circular membrane. The program calculates the displacements induced by the presence of the gold annuls deposited on an otherwise pristine mask membrane. The gold stress magnitude, (a parameter of

the calculation) was adjusted until 10 nm/mm appeared over the center square centimeter of the membrane corresponding to the measurement in the re-replication experiment. The value of gold stress that yielded a match is 0.5 GP.

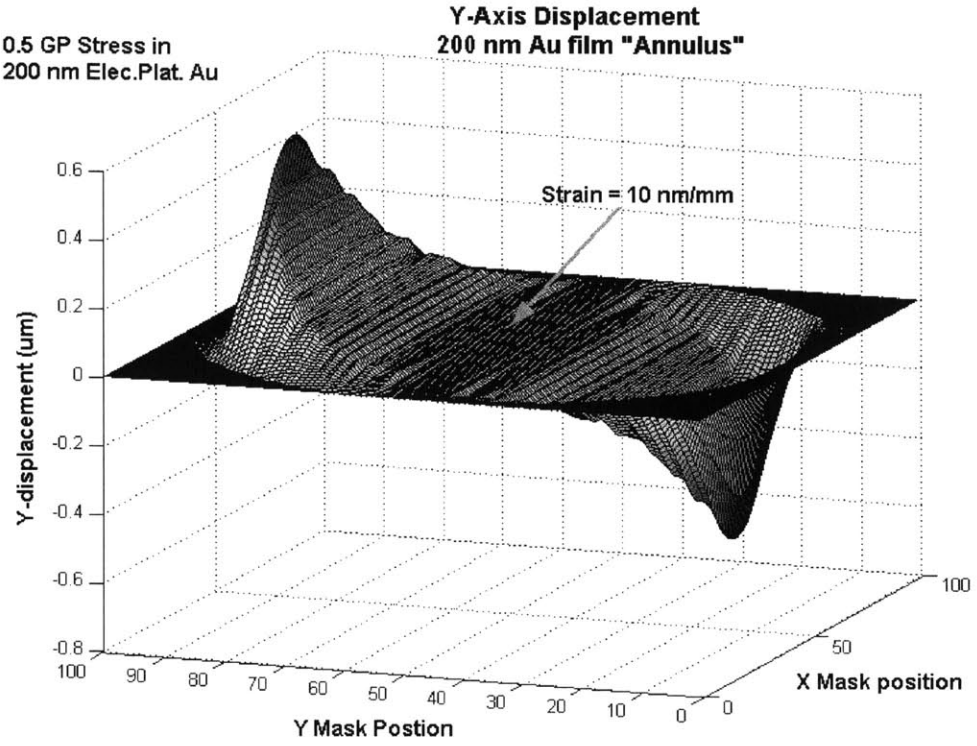


Figure 66 Surface plot of Y-axis displacement over 30 mm circular membrane. The gold stress magnitude, (a parameter of the calculation) was adjusted until 10 nm/mm appeared over the center of the membrane corresponding to the measurement in the re-replication experiment. (Axes show percent distance over membrane.)

This value was then used in to model a 1 centimeter square die for which it is assumed that 10% gold pattern coverage exists (this would be primarily due to the gate pad density). The result of this calculation is shown in Figure 67. The result indicates that the magnitude of strain in such a case is 0.01 nm/mm. The result was obtained using two approaches: 1) Assume 100% gate pattern coverage of the square die, calculate the strain and then scale by 10%, 2) scale the stress of Au by 10%, assume 100% coverage of Au over the square die and calculate strain. In both cases, the strain was uniform over the die and peaked in a rather non-linear fashion just outside of the die. Thus, modification of the

replication process to a negative resist scheme should reduce the impact of mask strain to a negligible level.

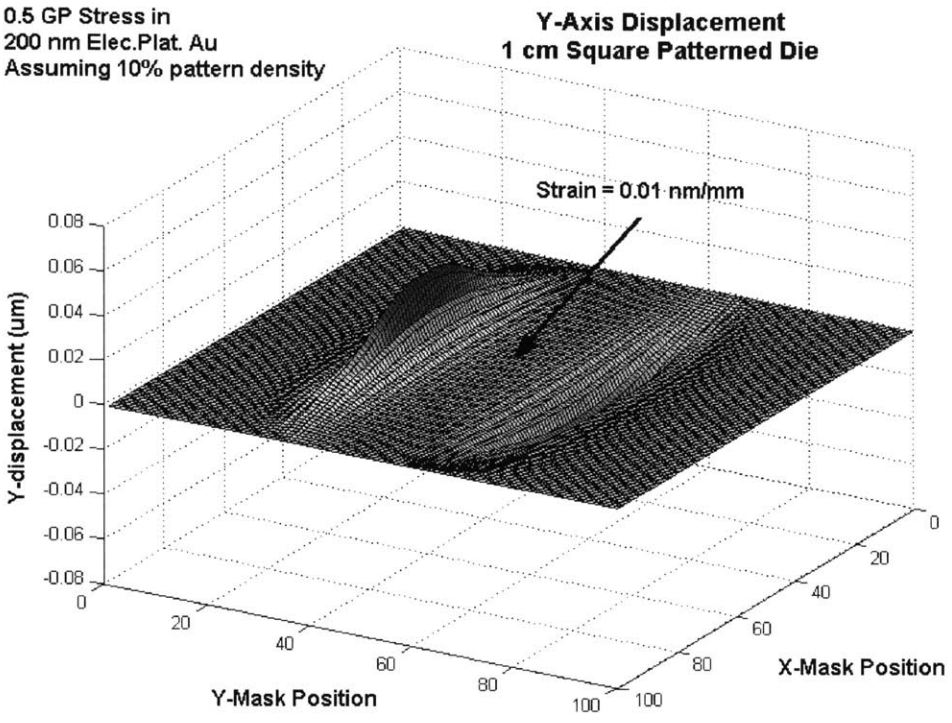


Figure 67 Simulation of Y-axis displacement on a 30 mm circular SiN_x membrane on which is patterned gold with 10% pattern density and 0.5 GP stress.

5.3 Alignment Results

Although the existing wafers and mask set contain a degree of stress-induced distortion, an experiment was done to establish the efficacy of alignment locally within the die. Specifically, mask set 2 was used with a remaining wafer of lot 2 to fabricate the double-gate structures located in close proximity to the X-ray axis of the exposures. These lines were placed within 100 μm of the adjacent alignment mark. Theoretically, the strain-induced misalignment would have been 10nm assuming 100 nm/mm net strain. Figure 68 shows this result.

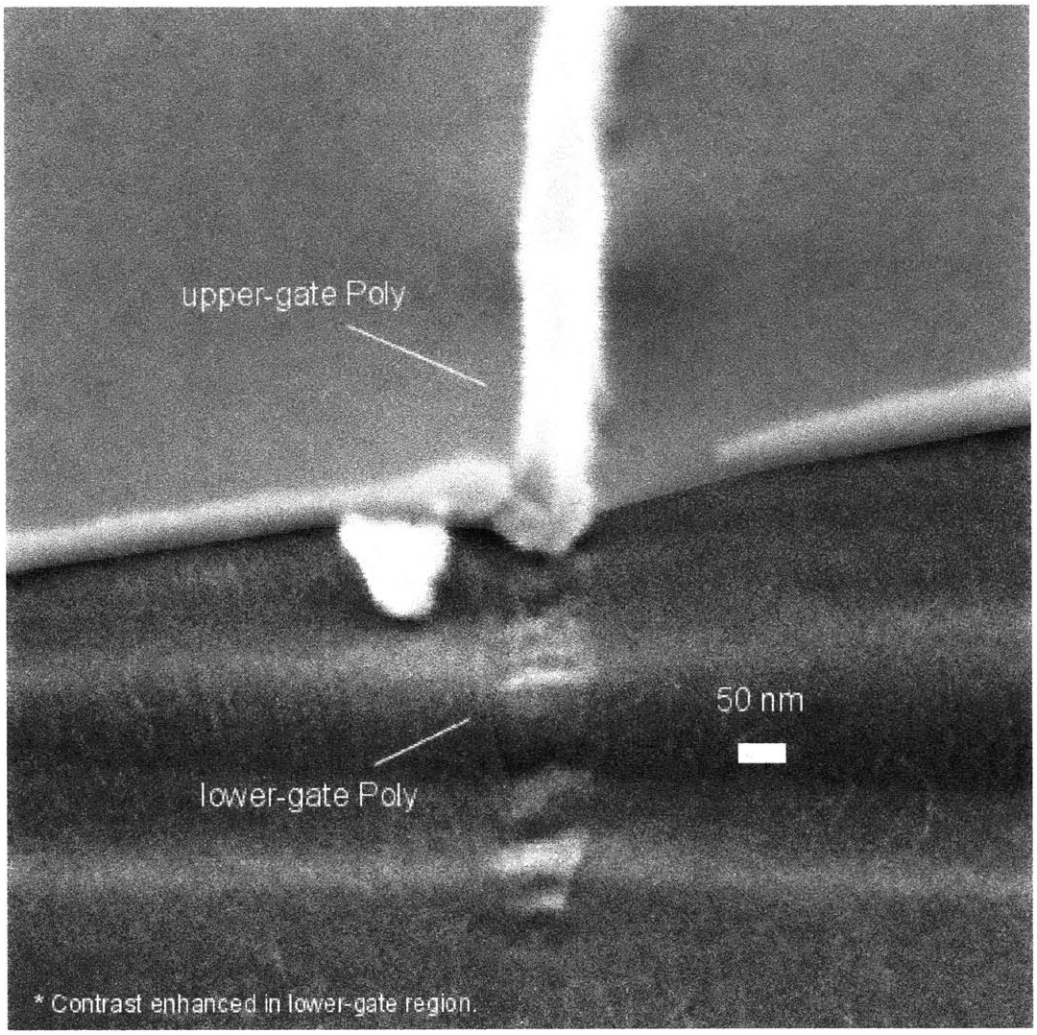


Figure 68 SEM micrograph of 50 nm gate structures fabricated using IBBI-aligned X-ray lithography. Misalignment is approximately 5 nm. Due to process induced distortion of wafers and masks, good pattern alignment was achieved within a sub-section of the die. This "selected area" was held at the X-ray axis of the exposures, adjacent to the central alignment marks (about 100 μm).

6 Wafer Strain Measurements

This section explores the experiment and analysis of film strain due to bonding and thermal cycling. The strain measurement methods presented here are well suited for obtaining the strain of SOI films that are patterned, flipped and then patterned again with alignment to the initial features. The moiré mark design used here can be flipped about its y-axis and superimposed with its original, thus producing interference fringes that provide sensitive alignment information. Because of this, the lithography tool need only print one mask, (with constant, well controlled magnification), for this scheme to work well. While the current method only yields strain in the y-axis, it could, through calibration, be extended to measurements in the x-axis. In a similar fashion, it could also be extended to measurement of films that remain un-flipped.

Certain methods for measuring film strain require the presence of a crystalline film. They include x-ray diffraction (XRD) [4] and high-resolution Raman spectroscopy [5]. Another approach is to quantify strain indirectly through measurement of wafer bow [6, 7]. The most prominent feature of the moiré technique is that it will serve those who wish to directly characterize lithographic capabilities of alignment and overlay for specialized processes that incorporate bonding and/or susceptibility to film strain.

6.1 Experiment

6.1.1 Wafer Preparation

Starting wafers were 100 mm SOI with 400 nm Si and 400 nm buried oxide. A sacrificial oxidation step was used to thin the Si film to 50 nm. Thermal oxide of 4-10 nm was grown, followed by chemical vapor deposition (CVD) of 150 nm poly-Si and a 100 nm low temperature oxide (LTO) hardmask layer. A GCA 4800 G-line stepper was used to define the patterns containing ~2 μm pitch gratings in 1 μm thick photo-resist. After HMDS surface treatment, 1 μm thick photoresist was spun onto the wafers, softbaked at 90°C for 60 sec, exposed and then hard-baked at 130°C for 60sec. The LTO was then etched in an AME5000 reactive ion etcher (RIE) using CHF_3. The resist was then ashed and the poly

layer was etched using a Cl_2 and HBr gas mixture. The etch was stopped on the thermal gate oxide. 1 μm of LTO was deposited and annealed at 800°C for 1.5 hours. The wafers were planarized using CMP. This step was repeated up to three times in order to maximize surface smoothness and procure higher bond strengths. 500 nm of LTO was removed with each CMP cycle. We ensured that step heights are no greater than 5 nm over 2 μm pitch gratings as measured by AFM.

The SOI wafers were bonded to "handle wafers" which are Si wafers with a thermal oxide of 100nm. Prior to both bonding processes, handle wafers and patterned SOI wafers are cleaned using the standard RCA method.

6.1.2 Bonding Method I

In this method, wafer bonding was done manually. Either the patterned wafer or the handle wafer was placed face-up on a Teflon® block. Four 0.2 mm-thick Teflon® shims were placed at the periphery of the wafer and the mating wafer was lowered onto the shims. Once the wafer flats were aligned, a Teflon® roller was used to apply pressure to the upper wafer. Contact was initiated at the center of the wafer pair, the shims were removed, and force was repeatedly applied from center to edge. It is estimated that the Teflon® support of the lower wafer deflects by a significant fraction of a millimeter which leads to bow of the wafer pair during bonding. The repercussions of this effect are discussed below.

6.1.3 Bonding Method II

This method was performed in an EV-501 automatic bonding machine. Either the patterned wafer or the handle wafer was placed face-up on a steel chuck. Three shims of 1mil (25 μm) thickness were set into place at the periphery of the lower wafer and the mating (upper) wafer was set into place on the shims with the wafer flats aligned. The upper wafer was in contact with a Teflon® pad (which is the lowest component of the machine's piston). A vacuum of 3 mTorr is applied to the bond chamber. A pin applied force to the center of the upper wafer initiating the bond. The shims were removed and the piston delivered 1500N of force to the wafer pair for 60sec.

6.1.4 Post Bonding Process

After bonding, the wafers were annealed for 2 hours at 775°C. The bond quality was inspected using IR imaging. A 300 nm SiN_x protection layer was deposited on the wafer pairs using LPCVD. This was done in two deposition steps between which the wafers were rotated in the quartz boat in order to achieve complete coverage (at the boat supports). The wafers were sent to an outside vendor for grinding service in order to begin removal of the SOI wafer. The target thickness of the wafer pair after grinding is 600 μm assuming a ~500 μm thickness for the handle wafer and a 100 μm thickness for the remaining SOI wafer. The wafer pairs were dipped in 7:1 BOE (HF) solution for 10sec to remove native oxide and then were etched in a 25% TMAH bath at 80°C. The time required to etch the Si varied from 2.5 to 4 hours.

The buried oxide film was wet etched in a BOE bath. Removal of this film not only enhances contrast of the interference fringes but plays a role in the strain of the wafer. Lithography identical to the previous patterning step was performed with coarse alignment to the embedded poly-Si patterns. This alignment ensures that the gratings overlay sufficiently for the fringe measurements while fine alignment of the gratings becomes a function of strain. Measurements were made by analyzing the interference of the photoresist gratings with respect to the embedded poly-Si gratings.

6.1.5 Measurements

Figure 69 shows a view of the grating layout for a 10 mm die. Highlighted are the Y-axis gratings aligned along the center vertical axis. This scheme will not detect X-axis strain. In the exploded of Figure 69, the grating on the left has pitch p_1=2.000 μm and on the right has pitch p_2=2.008 μm. Superposition of the gratings results in a side-by-side pair of moiré fringes with period

$$p = \frac{p_1 p_2}{|p_1 - p_2|} = \frac{p_1 p_2}{\Delta p} \cong 500\mu m, \tag{6.1}$$

which is equal to the vertical size of the grating pair. Consequently, on each side of the mark, one fringe appears after resist patterning. Figure 70 shows an optical micrograph of the resist/polysilicon moiré measurement. The position of

the fringes in these marks is sensitive to Y-direction shift of the resist relative to the poly-Si grating beneath it. For a positive vertical displacement of the resist grating, left-side fringes move downward and right-side fringes move upward. The ratio of fringe displacement to actual displacement is

$$2\frac{p_{avg}}{\Delta p} \cong 0.50\,\mu m/nm \tag{6.2}$$

where p_{avg} is the average of the two periods. The factor of two arises because of the counter-motion of fringes that doubles sensitivity. Using an optical microscope, a user can measure the position of a fringe to better than 10 µm, using a 20x objective and the built-in scale of the eyepiece. Thus, the sensitivity of strain measurement can be as fine as 10 nm per data point using this technique. By implementing a design with a greater number of fringes and introducing computer image analysis of the marks, nanometer-level sensitivity can be achieved.

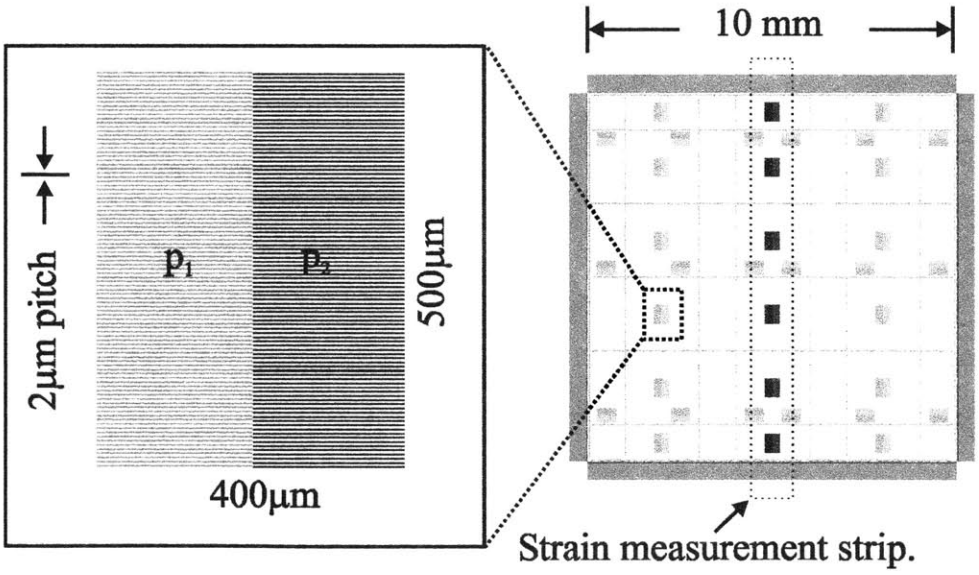


Figure 69 Schematic of layout.

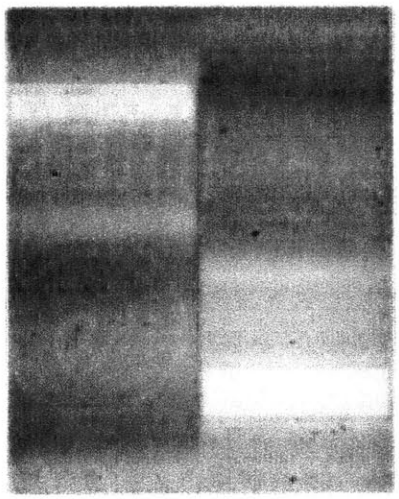


Figure 70 Optical micrograph of moiré mark showing fringes resulting from interference between resist grating and underlying poly-Si grating.

The lithography system that is used to pattern these gratings need not perform precise alignment because measurements of strain are relative. In our analysis, the fringe offset of the lowest mark is set as a reference, and the variation of fringe offsets of the other marks are taken relative to it. The direction in which the fringes drift as one measures across the die determines whether the film stack is under tensile or compressive strain. This drift versus position is termed "runoff". The marks are also immune to angular error because rotation of the marks leads to identical fringe motion on both sides of the mark.

While precise alignment is not a requirement, repeatability of magnification of the lithography tool is essential. As shown in Figure 69, strain measurements are made only in the center vertical strip of the die . In doing so, the same portion of the lens system is used to define these central gratings before and after the wafer is flipped. This is because this center strip is an axis of symmetry of the die pattern.

6.2 Results

6.2.1 Data

Figure 71 presents typical data from two wafers processed using bonding method I (manual bonding). The wafer pair corresponding to Figure 71A had

force applied to the handle-wafer while those of Figure 71B had force applied to the patterned SOI wafer. Shim thicknesses of 0.2mm and 0.4mm were used in these experiments, but a strong statistical relationship between strain magnitude and shim thickness has not yet been established. The average strain was found by calculating the average of the slopes of straight line fits to each trace. Each trace is taken from a 10mm die on the wafer, and the die are arranged in an 8x8 array. For the case of force applied to the handle wafer, the calculated average strain is -23nm/mm. For the case of force applied to the SOI wafer, the calculated average strain is 46nm/mm. These numbers are representative of the wafers processed using manual bonding.

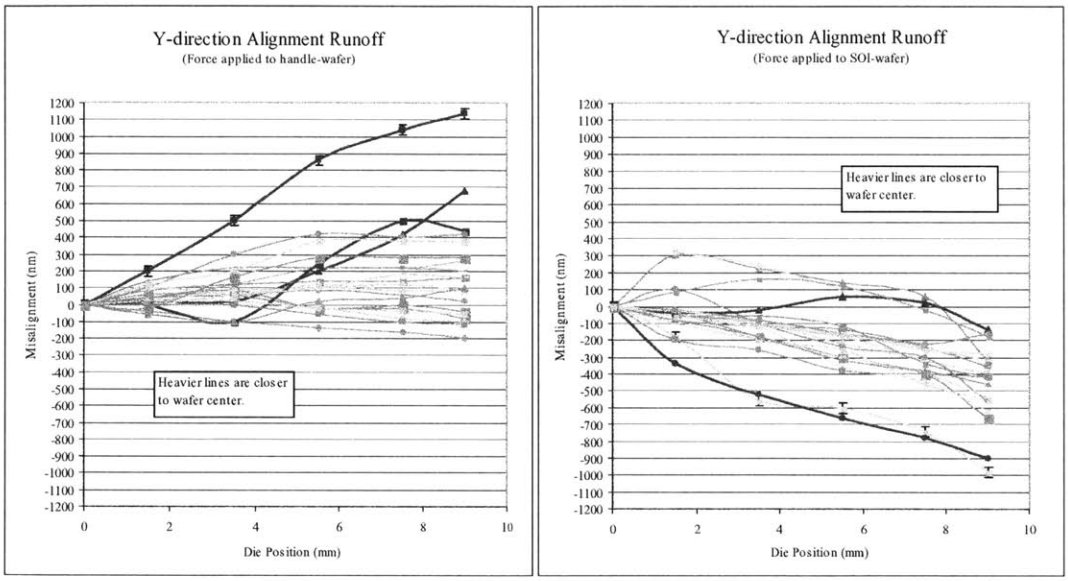


Figure 71 Strain results for two wafers bonded manually. Each trace is taken from a single die of an 8x8 array as printed on the wafer. The left plot (A) represents data from bonding with force applied to the handle-wafer while plot (B) contains data from a wafer set with force applied to the SOI (patterned) wafer. A positive slope corresponds to tensile strain while a negative slope corresponds to compressive strain of the SOI film. The 20nm error bars that are displayed on one trace in each plot represent the measurement error in determining the fringe locations.

Figure 72 presents data from wafers processed using bonding method II (bonding machine). The results of 3 wafers are presented here. *Wafer Pair A* was bonded such that the handle wafer rested on the chuck and force was applied to the patterned SOI wafer. *Wafer Pair B & C* was bonded with SOI wafer on the chuck and force applied to the handle wafer. This was done in an

attempt to investigate if a relationship exists between strain type (tensile/compressive) and direction of applied force in machine-type bonding. Figure 72 contains surface plots of strain of the three wafers. The surface plots reveal some tendency for greater strain magnitude towards the center of the wafers. They also suggest the possibility of a systematic positive offset between 10 and 20nm/mm.

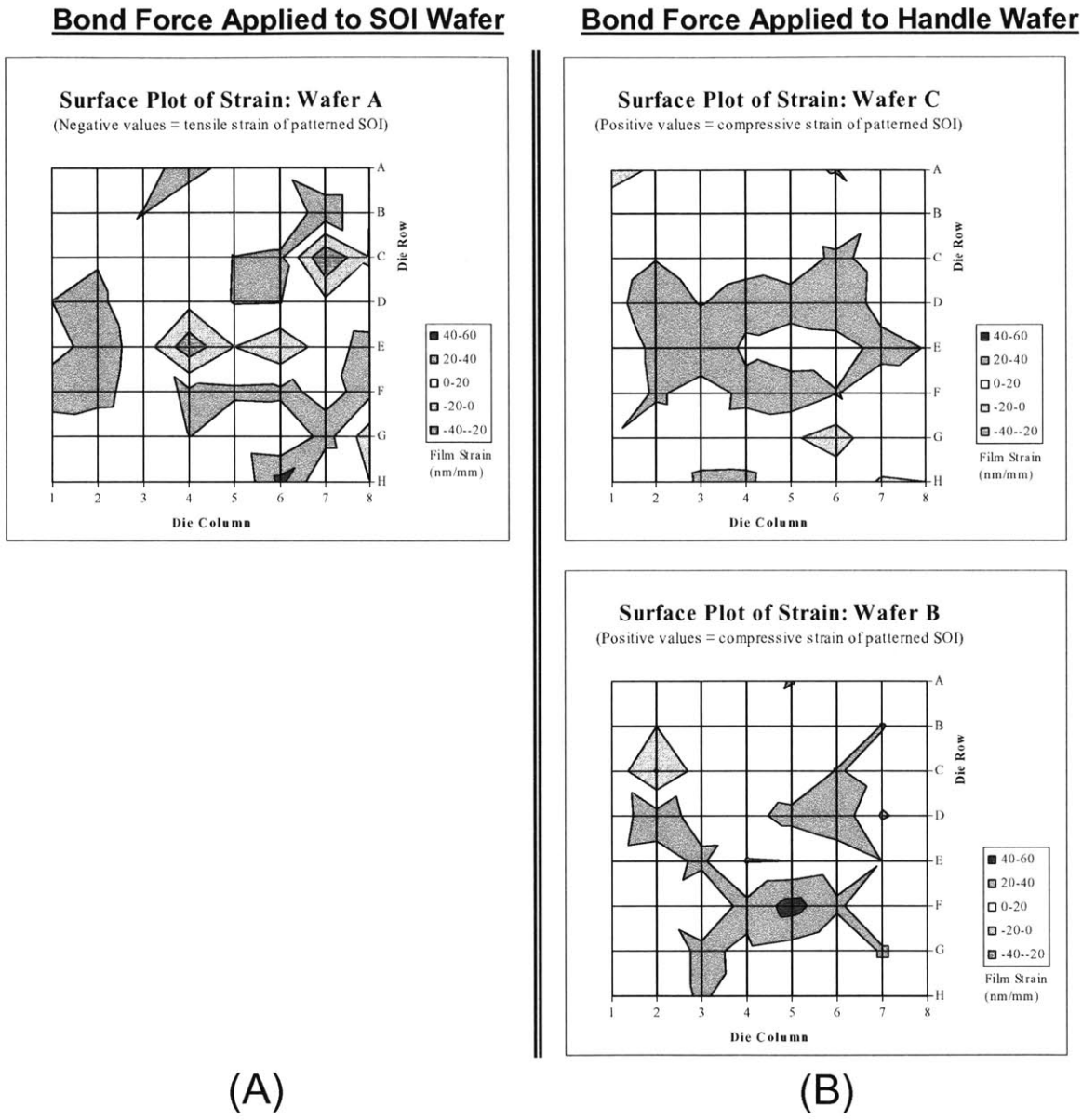


Figure 72 Strain results for three wafers bonded using bonding machine. Plot (A) represents data from bonding with force applied to the SOI (patterned) wafer while plot (B) contains data from wafers with force applied to the handle wafer.

To quantify the repeatability of stepper magnification as well as the repeatability of fringe measurements, The measurements were repeated for wafer A and C by stripping the resist, re-spinning resist and re-exposing the gratings. Figure 73 shows these consecutive measurements side-by-side. Additionally, Figure 74 contains histograms that categorize the number of die that changed by varying degrees between these two measurements. This experiment incorporated an additional test, namely: magnification versus wafer thickness non-uniformity. This is important because wafer height variation will lead to some degree of projected image magnification. In order to test whether this was the case, the depth of focus limits were established through a focus array. For this G-line system with NA of 0.28 and depth of focus ~6 μm. A ~2 μm shift of the lens was used in two sequential tests of exposure and measurement. These two focus settings were used, sequentially, to obtain the data in Figure 74A. It was found that magnification change due to this change is focus settings did not result in a discernable shift in strain measurement. Given that typical wafer thickness variation is about 0.5 μm, this height variation has a negligible impact on magnification at the current sensitivity level. With regard to sensitivity, the data in Figure 74 indicates that this method is repeatable to within ±10 nm/mm. The time interval between these measurements was approximately two weeks while the interval between pre-bond and post-bond lithography was about 6 weeks.

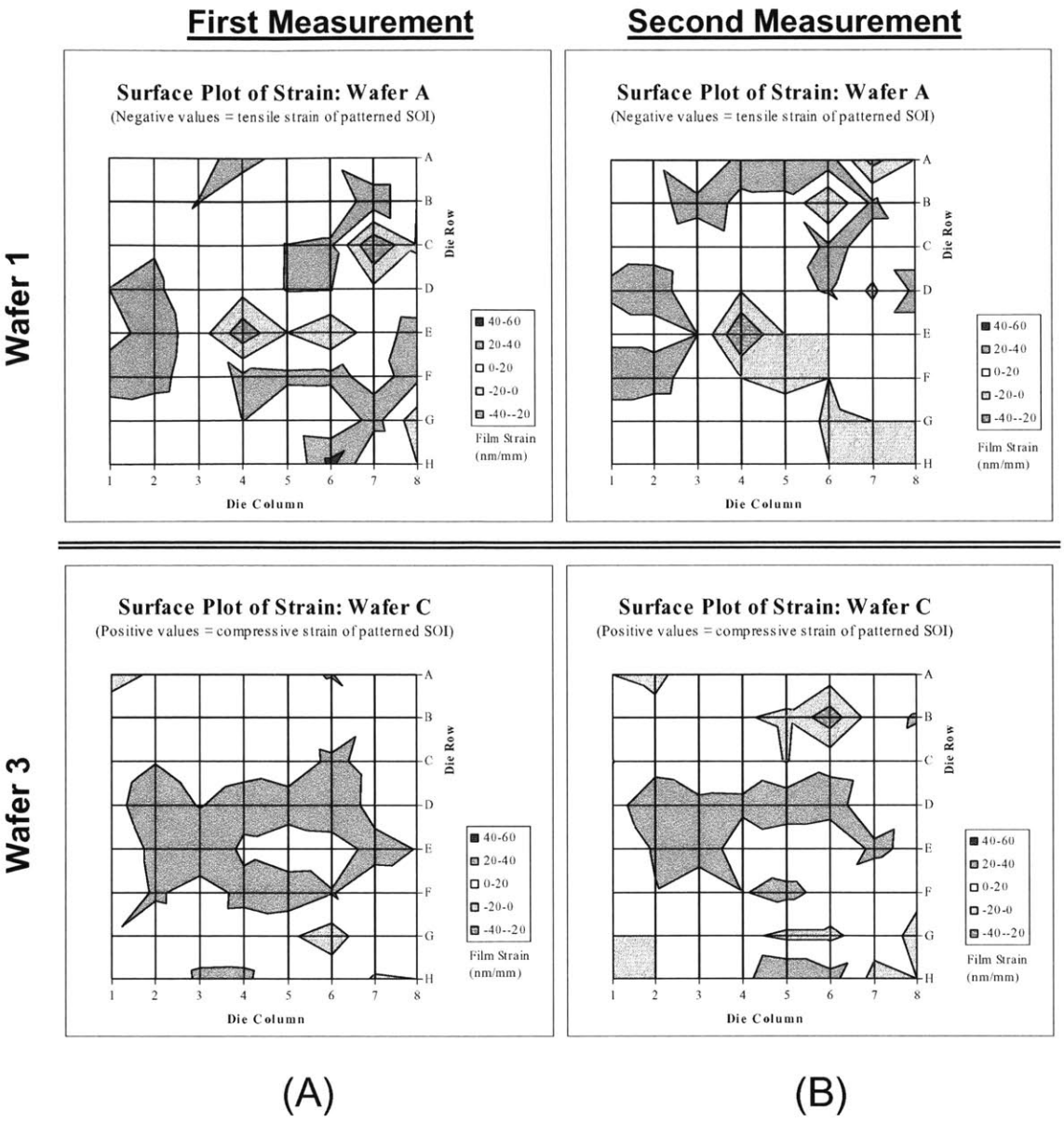


Figure 73 Comparison between the repeated measurements of strain in wafer A and wafer C. After the first measurement is taken, the resist is stripped, re-spun and then re-exposed in the stepper. This experiment is done to quantify the repeatability of the magnification of the lithography tool as well as the repeatability of fringe measurements.

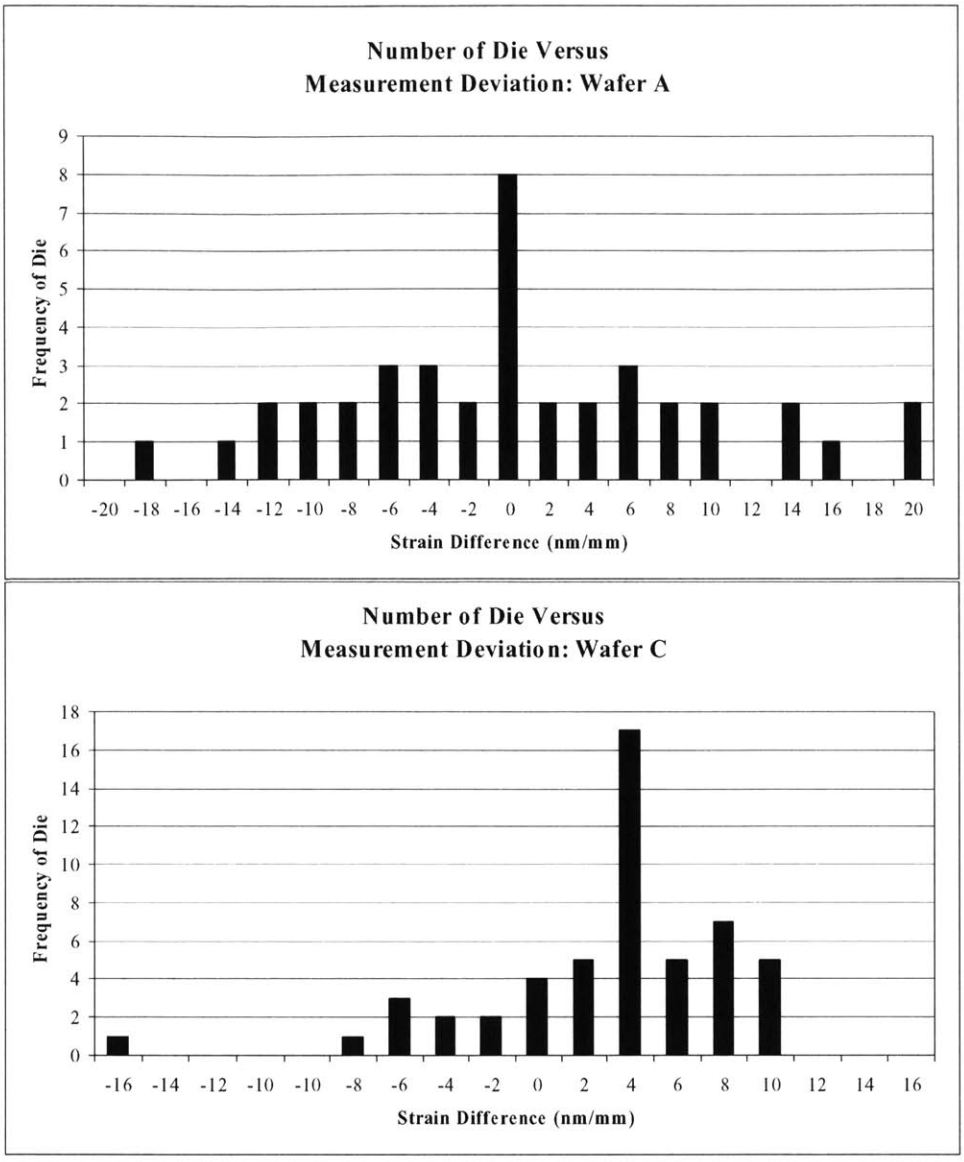


Figure 74 By repeating the final pattering and measurement of a particular wafer, the strain exhibited by each die may change, yielding a measure of repeatability of magnification of the lithography and fringe offset measurement. This change in value is designated here as the *deviation*. The frequency of die is plotted against 2nm/mm ranges of deviation.

6.2.2 Analysis

The data indicates that the bonding method can have an impact on film strain. It is clear that the manual bonding method yields grossly strained SOI films and that furthermore, the direction of strain correlates well with the direction of the force applied to the wafer pair.

In the case of machine bonding, the strain maps in Figure 72 reveal two subtle features. In all cases, there is a positive level of strain over the entire wafer, and it is evident that there is a higher degree of strain towards the center of wafers A and C and below the center of wafer B. If one dismisses the possibility that the overall positive offset is due to a stepper magnification drift by virtue of the repeatability measurements, then the strain must arise from either the stress of deposited/annealed films or from the presence of wafer bow during bonding. Evidence supporting the latter mechanism is presented in Figure 72 where some correlation exists between the polarity of strain and the direction of bond force (as we saw in the case of manual bonding.) In Figure 72B shows some positive maxima which corresponds to compressive strain of the SOI film. Given that the bonding force was applied to the handle wafer in this case, compressive strain may be a result of the relaxation of convexity of the handle wafer bond surface. The presence of a positive strain in the remaining areas of the wafers would be better explained by film stress. The following sections present models that help to characterize these two sources of strain.

6.2.2.1 Bonding-induced distortion

For further insight into bond-induced distortion, a relationship between wafer non-flatness and strain is presented here. In this case, the strain is induced by bowed wafers at the moment of bonding. A diagram of a bent section of wafer is shown in Figure 75. A measure of non-flatness is indicated by *h* over the distance L. The neutral line is the plane having zero strain and is assumed to bisect the wafer thickness. The radius of curvature of the bend is R and the wafer thickness is *t*. δ/L is the lower-surface strain of the wafer. A relationship between h, L and δ will relate surface strain to wafer non-flatness (bow).

Considering the right-half of the wafer segment, the length of the lower surface is

$$L+\delta=\theta\left(R+\frac{t}{2}\right)=\frac{L}{R}\left(R+\frac{t}{2}\right) \tag{6.3}$$

and the lift is

$$h = \left(R + \frac{t}{2}\right)\left[1 - \cos(\theta)\right] \cong R\left[1 - \cos(\theta)\right] \quad for \quad R >> h\,. \tag{6.4}$$

For small deflections, R>>L and θ=L/R so the cosine term in (6.2) can be approximated by the first two terms of the Taylor expansion

$$h \cong R\left[1 - \left(1 + \frac{L^2}{2R^2}\right)\right] = \frac{L^2}{2R} \tag{6.5}$$

Thus $R=L^2/2h$, and this is substituted into (6.1) yielding

$$L + \delta = L + \frac{th}{L} \qquad \therefore \ \delta = \frac{th}{L} \tag{6.6}$$

Since strain of the surface is δ/L,

$$Strain = \frac{\delta}{L} = \frac{t\,h}{L^2} \tag{6.7}$$

Figure 76 plots strain versus non-flatness, (*lift*), for various lengths of wafer based on equation 6.7. For example, to induce a strain of 20nm/mm, about 100μm lift needs to be present over a 50 mm length. Once the bond takes place, the wafers will equilibrate to balance the stress at the interface. Thus, the strain of the bowed wafer surface will be reduced somewhat due to the compliance of the mating wafer. However, once one wafer is removed from the pair leaving behind the transferred film, the mating wafer will relax back to its original state. Assuming that the transferred film is relatively compliant, the magnitude of the strain will be the same as that of the bowed wafer surface during bonding.

Thus, in the case of machine bonding, assuming wafer non-flatness contributions of 25 μm from the shims, and about 20 μm from cumulative wafer non-flatness, according to Figure 76, we would expect 8 nm/mm strain if this lift occurred over 50 mm and 50 nm/mm if it occurred over 20 mm. This is at least of the correct order of magnitude as the experimental results.

Another consideration is whether local CMP-induced thickness variation could cause a distortion contribution. Given a ~10:1 difference between patterned and un-pattered surface area in the moiré grating layout, and with a poly-Si thickness of 100 nm, the variation in LTO thickness is expected to be

10% of this thickness; i.e. 10 nm occurring over a 3 mm range. According to this model, strain induced by this contribution of wafer bow would of the order of 1 nm/mm. This is clearly not a detectable level in these experiments.

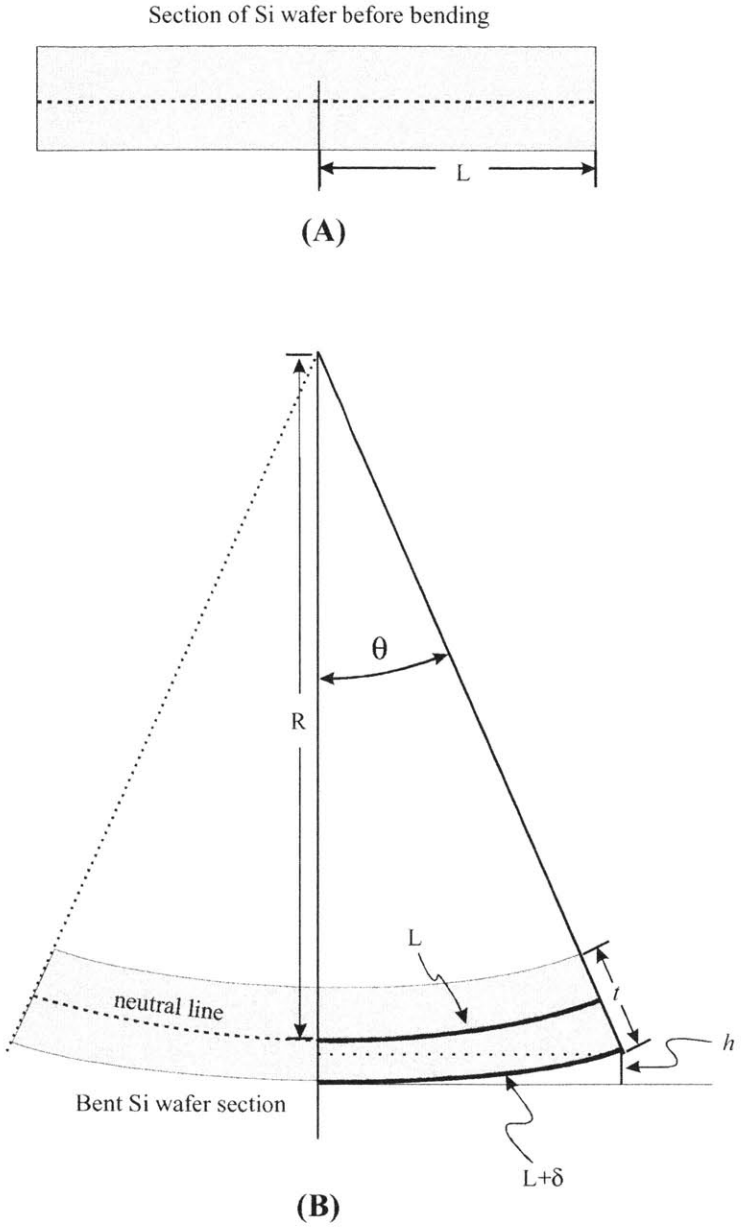


Figure 75 Cross-section of a segment of an Si wafer before and after bending. A measure of non-flatness is indicated by *h* over the half-distance L. The neutral line is the plane having zero strain and bisects the wafer thickness. The radius of curvature of the bend is R and the wafer thickness is *t*. Δ/L is the lower-surface strain of the wafer-section.

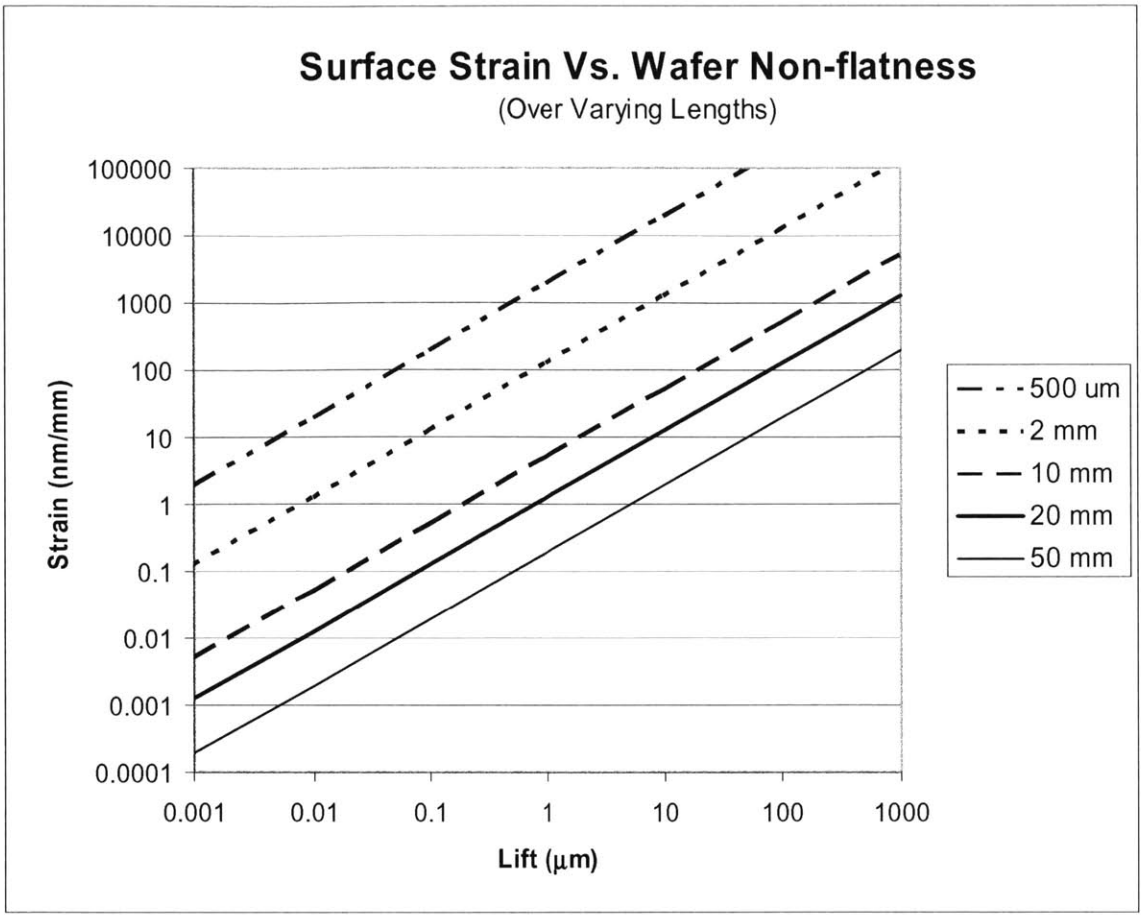


Figure 76 Surface strain versus wafer non-flatness over various lengths of a Si wafer of 500 µm thickness.

6.2.2.2 Stress in annealed and deposited films

Considering the films present on the wafer and the end of processing, 100nm of SiO_2 and 300nm of SiN_x is present on the back-side while about 600nm of LTO and 50nm Si is present on the front-side (with 150nm embedded poly-Si patterns). It is conceivable, for this particular process, that these films contribute to the observed positive stress. Given that the photoresist is sparsely patterned on the wafer, its stress contribution is not considered. A relationship derived by Townsend [**19**] expresses the strain throughout a Si substrate that is pulled to a flat surface as a function of its thickness and the unrestrained wafer curvature

$$\varepsilon_s^{flat} = -\frac{t_s K}{6} \tag{6.8}$$

where t_s is the substrate thickness and K is the curvature. Based on measurements of the curvature of the wafers using an interferometer, the radii of curvature are about -20m (convex). Using equation (6.8) to approximate the strain for a 500 μm thick wafer, the strain is 4 nm/mm and is compressive. This result helps to support the assumption that film stress is contributing to the constant positive strain seen in all three wafers in Figure 72 as it lies within the correct range, i.e. 0 to 20 nm/mm.

7 Conclusions

This project has addressed and solved a number of formidable fabrication challenges associated with the planar double-gate design approach. Two distinct problems areas exist: 1) strain of the X-ray mask membrane and 2) strain of the wafers.

In the case of mask membrane strain, it was shown that a liftoff process had led to detrimental distortion of the mask membrane. The viability of the alternative process idea, (use of a negative resist), was studied based on measurement of Au stress. It was found that Au stress should not pose a problem when pattering gate features with 10% feature density; thus the mask replication scheme is a potentially excellent method for achieving overlay.

Regarding negative resists, it may be that the major hurdle in achieving the required trench-widths is the maximum achievable absorber contrast of the upper-gate mask. It may be possible to obtain thicker gold absorber on this mask at the required resolution. Alternatively, one may tradeoff resolution for greater contrast and employ the trimming process to correct for the larger linewidths. If this does not prove to be possible, then alternative resists with suitable dose response may become available.

One of the difficulties of the mask replication process is the protection of the alignment marks during the final electroplating step. A simpler alternative would be to electrically isolate the alignment marks by eliminating the plating base along the perimeter of the alignment mark clusters.

Although statistical data of alignment could not be obtained, the effectiveness of the IBBI alignment system was demonstrated several times yielding structures aligned to about 5 nm.

The strain that a wafer incurs through the flip-and-bond process is perhaps a more difficult obstacle to overcome. However, it should not be difficult to extend the moiré experiments into a more sensitive regime in order to test and improve the bond process and the lithography tools used. Based on IBBI results achieved in NSL, a mark design similar to the IBBI-based mark could be used in

the G-line process. It would produce 6 or more finer fringes that would be fed into a computer-based analysis tool to obtain nanometer level measurent sensitivity of the mark. A further extension to the moiré setup would be to use much longer marks and acquire the fringe images perhaps by forming a "panorama" with an automated optical microscope stage. This would yield data that has extent over the wafer surface as opposed to discrete measurement points. In such an automated scheme, one could also partially overlap the die prints so that moiré mark clusters gain more coverage over the wafer.

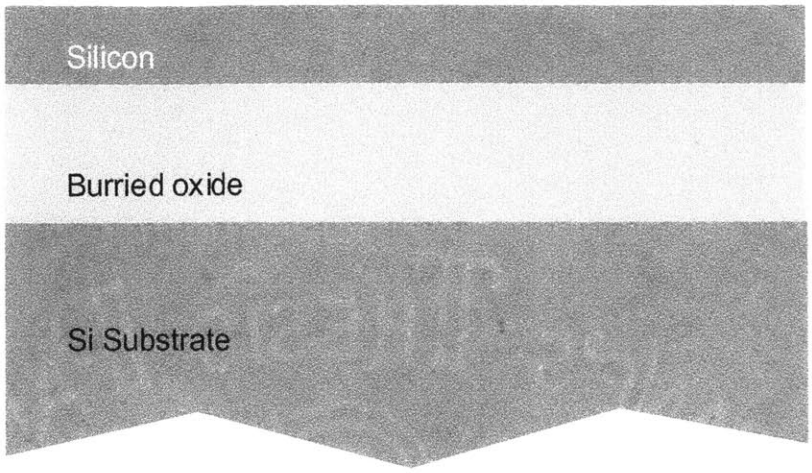
Clearly, wafer flatness must be maximized during bonding. With the current tools, it is difficult to determine the exact flatness that exists during the bond. Based on the literature, it is not clear as to whether or not wafer contact must be initiated at a single point if the wafers had a high degree of flatness to begin with. In order to get the best results, the wafers should be first mounted to chucks that can provide flatness correction. Since bonding in a vacuum would most likely be preferable, a method of obtaining adhesion to these chucks must be found since vacuum could not be used. The handle wafer should be flat to start with, or an alternative optically flat substrate can be used as long as it is compatible with subsequent process steps.

It is not completely clear to what degree film stress will cause a problem in a complete CMOS process in view of the extreme tolerances of the 10 nm double-gate. In our experiments, the major film stress contribution came from a 300 nm SiN_x film on the backside of wafers. Un-needed films such as these should obliviously be removed in order to optimize the process. Another possibility may be to pattern films that offer opposing stress to the backside of the process wafer or to used thicker substrates.

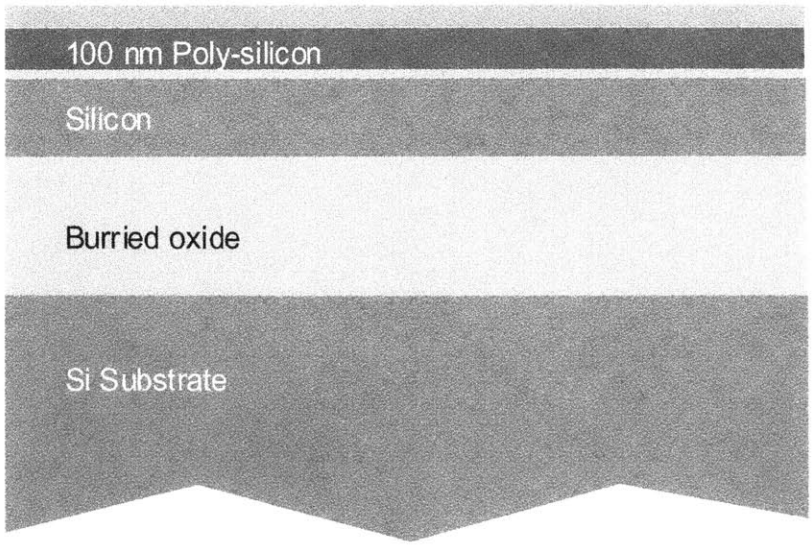
The author hopes to see the development of this work continue. I firmly believe that this process can now be used to fabricate DGFETs at a research level given some of the aforementioned improvements. Given the extreme tolerances of this *ultimate transistor*, the scheme presented in this work may prove to be the best (or only) method of achieving these goals.

Appendix A: Device Flow Diagram

(1) Initial SOI wafer, typical values: 400 nm Si (before thinning) and 400 nm buried oxide.



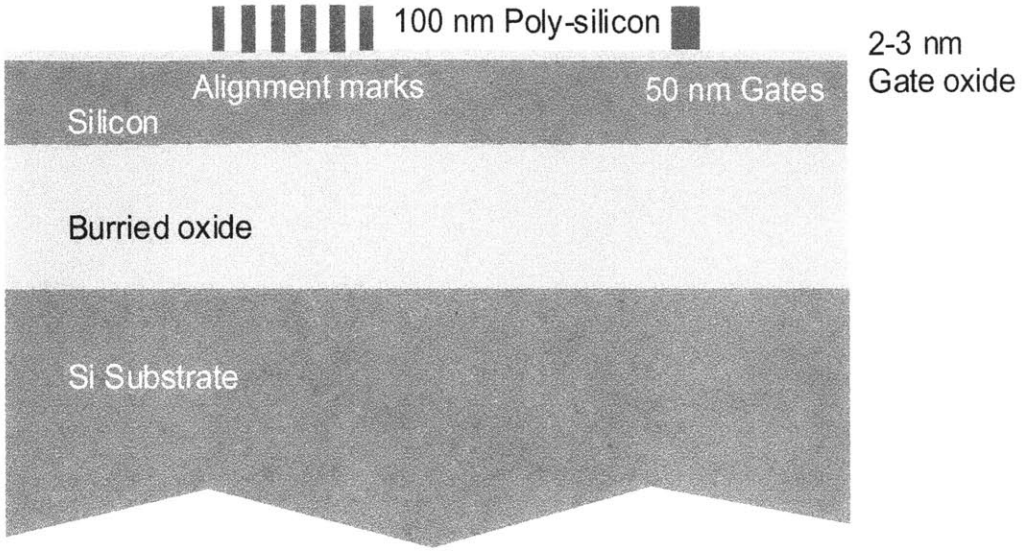
(2) Deposit gate stack consisting of 2-3 nm thermal oxide, 100 nm poly-Si gate material, dielectric



Nitride or Oxide Hardmask

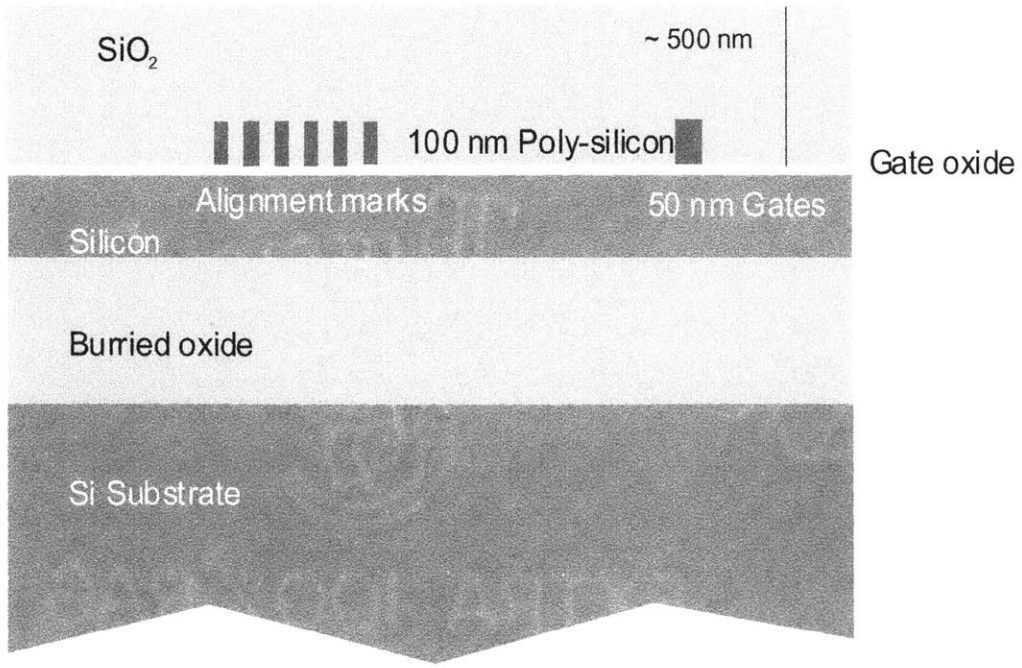
2-3 nm Gate oxide

(3) Define hardmask using X-ray lithography and perform gate etch which includes the gates/pads and the IBBI alignment marks.

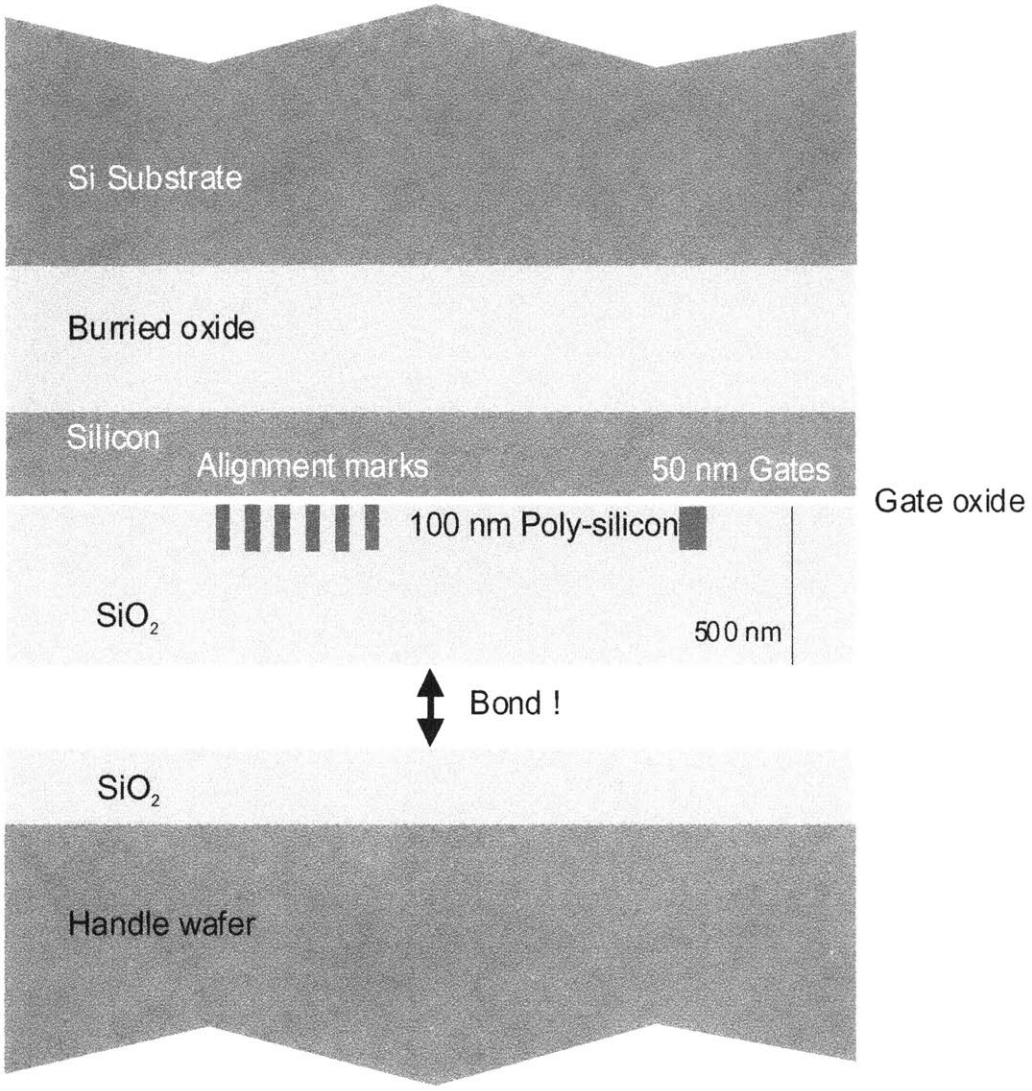


(4)

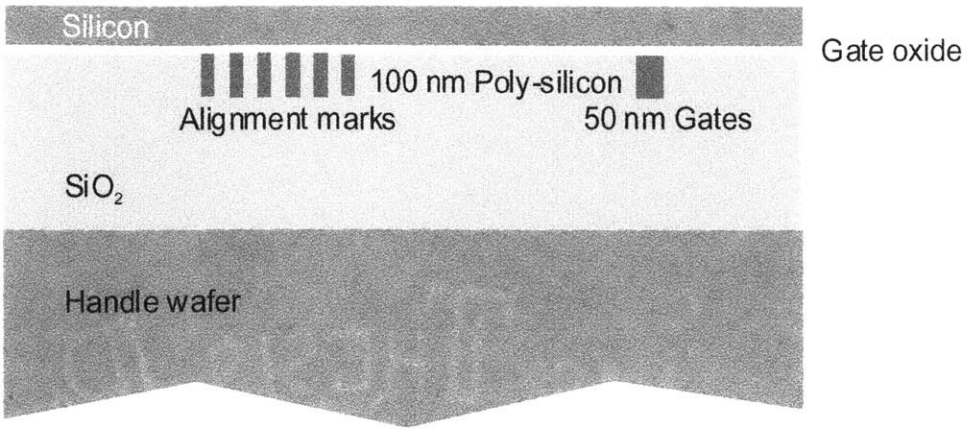
Deposit SiO_2 and CMP prior to bonding.



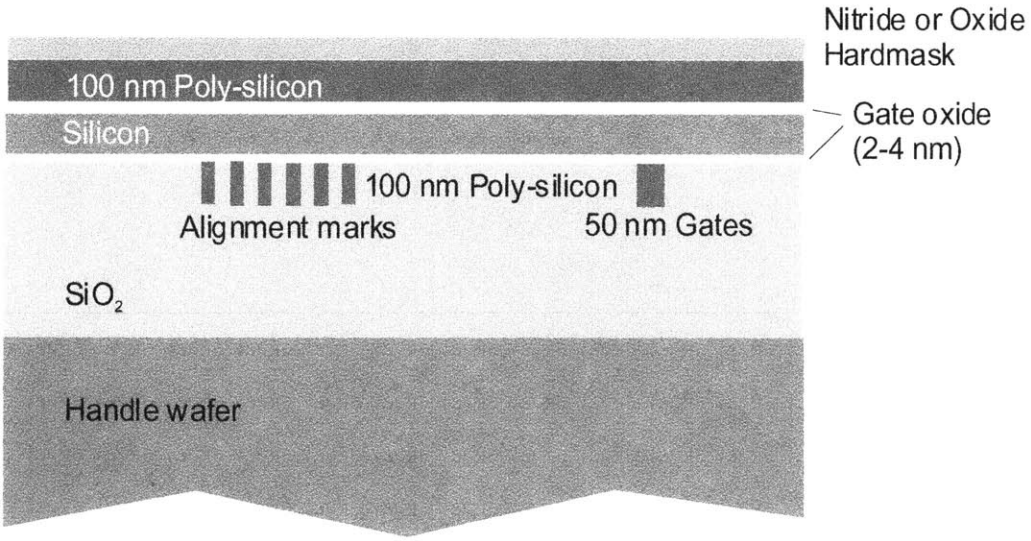
(5) Flip SOI wafer and bond polished surface of oxide to handle wafer.



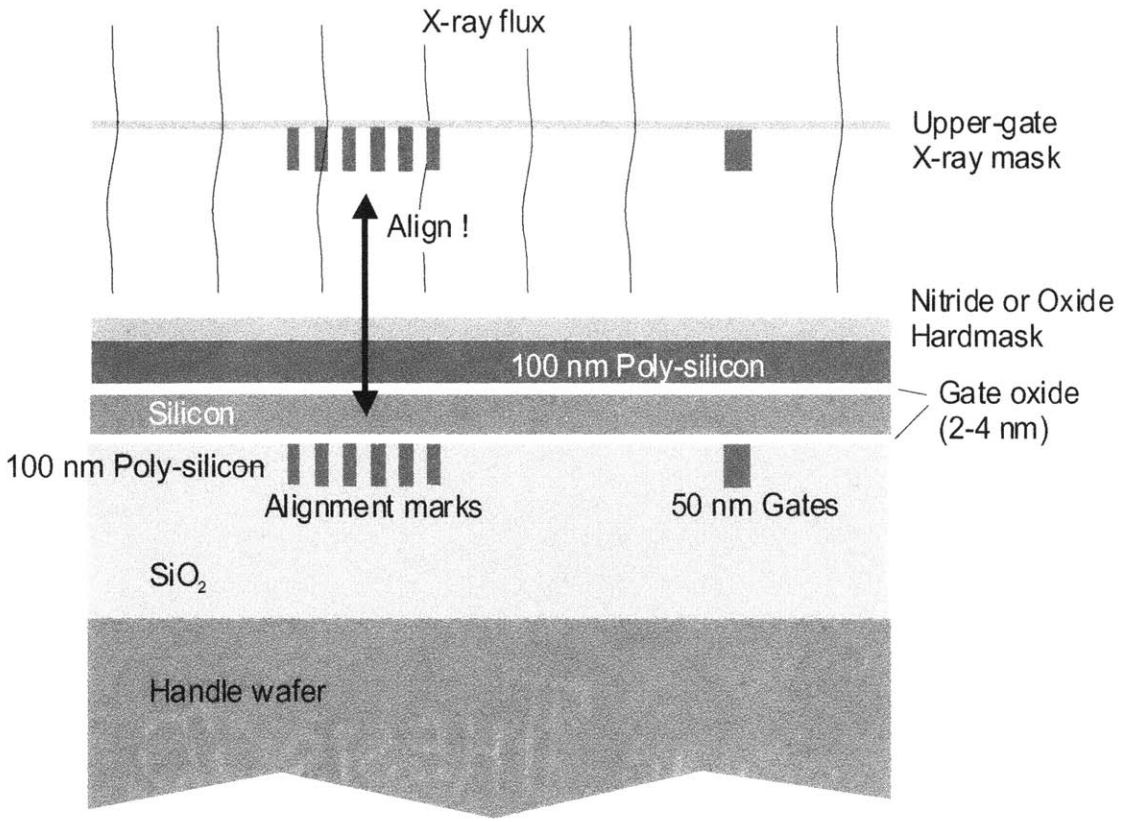
(6) After etch-back of SOI wafer.



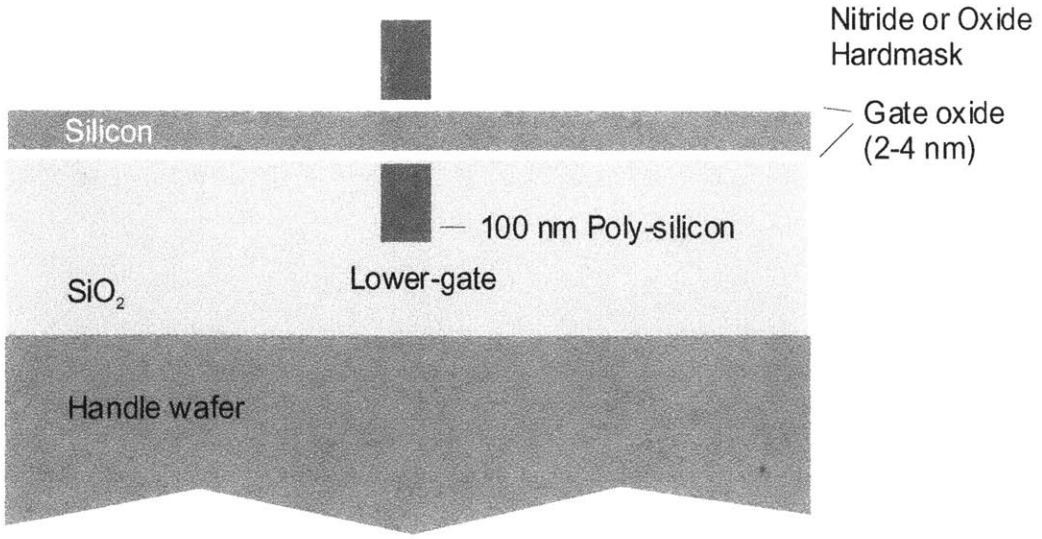
(7) Deposit top-gate stack consisting of 2-3 nm thermal oxide, 100 nm poly-Si gate material, dielectric



(8) Align upper-gate X-ray mask while exposing upper-gate patterns.



(9) Final, simplified device cross-section.



Appendix B: Standard NSL X-ray Mask Process

The following X-ray mask process is common to both the upper-gate mask and the lower-gate mask. This process begins with the selection of a suitable X-ray mask blank. The blank consists of a optically flat, 1.0 μm thick membrane of silicon nitride supported on an annulus of silicon. The silicon, in turn, is bound to an optically flat ring of Pyrex and is termed the mesa. A cross-section is shown in Figure 77.

Figure 77 X-ray mask cross-section.

The membrane is inspected for flatness and particle contamination. A fiber light is used to determine which side of the membrane specific particles reside and microscope-mounted vernier is used to measure their size. Particles away from the mesa and smaller than 5 μm in height seem to be benign in subsequent lithography steps. Figure 78 shows an interferometer measurement of a somewhat distorted mask blank. Some fringes can be seen on the mesa, and in this case, non-flatness of the mesa can cause un-level gapping when this mask is brought into proximity with a sample. Each fringe corresponds to a height change of roughly 250 nm.

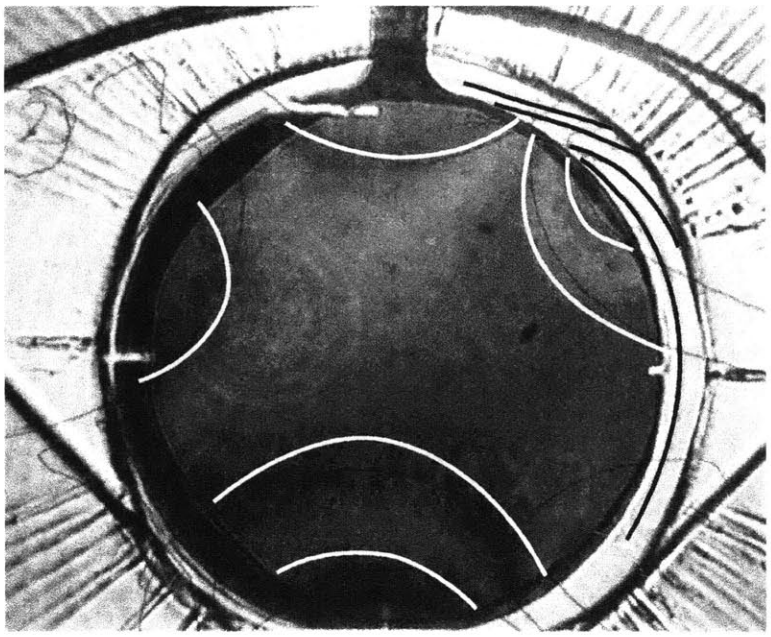


Figure 78 Interferometer image of X-ray mask with PMMA. The white lines highlight fringes appearing on the membrane while black lines highlight fringes appearing on the mesa.

Mask blanks are cleaned using a piranha bath (3:1 sulfuric acid :H_20) for 10 minutes followed by an RCA SC1 clean (5:1:1 H20 : peroxide : ammonium hydroxide) at 75° C. Following the clean, 10 nm titanium followed by 10 nm gold is evaporated onto the membrane. The Ti behaves as an adhesion layer. Extra metal is evaporated onto the Pyrex to enhance conduction to the membrane plating base. The cross section appears in Figure 79.

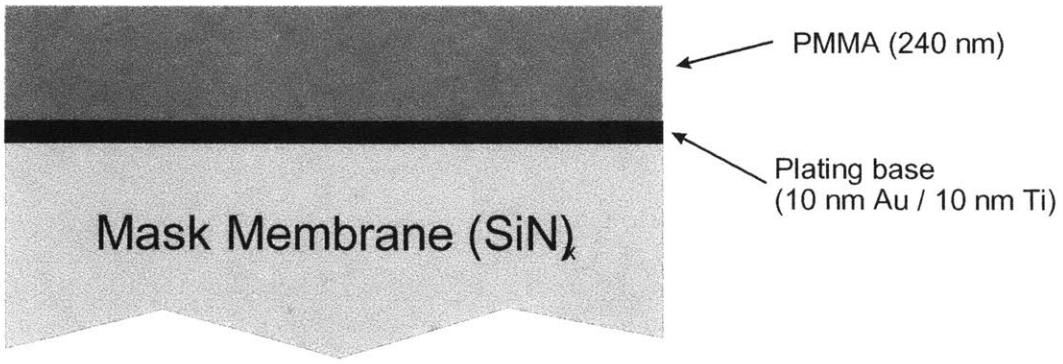


Figure 79 Initial mask cross section.

The PMMA is patterned accordingly (Figure 80): for device patterning, an e-beam tool is used (see section 4.3) while for registration marks, DUV contact printing is used (see section 4.2).

Figure 80 PMMA after patterning – forms a mold for Au electroplating.

After exposure, (but before developing), the PMMA on the Pyrex is flood exposed with 220nm UV so that most of it can be removed during the develop step. E-beam patterned PMMA is developed in 2:1 IPA:MIBK for 90 sec 25.0 °C, immersed with slight agitation. DUV exposed PMMA is developed in 3:2 IPA : MIBK, sprayed at room temperature for 60 sec.

After developing, the mask is rinsed in IPA followed by DI. DI should not be used on dense, high aspect ratio patterns, (for example, a 150 nm pitch grating) as the surface tension of H_20 can cause these patterns to collapse towards one another. Before further processing, it is critical to ascertain that the Pyrex surface is free of residual PMMA. Remnant resist can cause the electroplating surface area to be drastically reduced and result in improper plating rates and possible over-plating. Flood exposure of the PMMA in this region results in an electromagnetic node (zero intensity) at the PMMA-Au interface resulting in a thin layer of unexposed PMMA on the Au surface. An effective way to clean the Pyrex is to manually polish it using a swab saturated with acetone, chlorobenzene or NMP.

Regardless of the lithography tool used to pattern the PMMA, it was found that consistently good plating results were achieved when a brief oxygen RIE (reactive ion etch) step was used prior to electroplating. It was difficult to establish the necessity of this step because good plating results were sometime obtained without treating the sample. PMMA can erode in the etch, so the etch power/bias should be kept relatively low (below 100 V @ 10-20 mT) and the etch duration should not exceed 6 sec (4 sec works well).

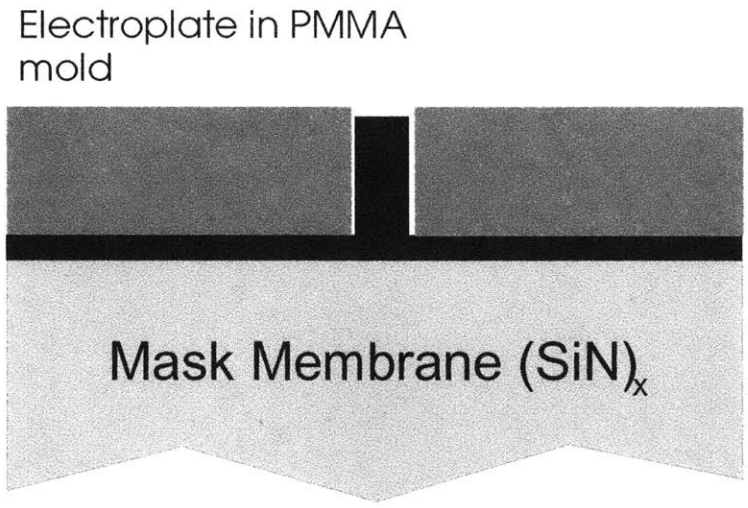


Figure 81 Cross-section after electroplating.

The plating bath used is BHT-510 manufactured by Ethone Inc. Masks are plated at a current between 12 and 16 mA depending on the exposed surface area. The total charge necessary to plate up to 200 nm is generally about 120 mA-min. The solution is held between 31 to 34 °C with continuous flow and filtration. The conductivity should fall between 69-76 mS-cm. Safeguards for reliable results include plating a test sample to check the quality of the deposit. In addition, a reliable procedure for fully plating the sample (i.e. bringing the gold deposit up to the height of the PMMA mold and no higher) is to incrementally deposit the metal. In this method, gold is deposited based on a calculation that leads to 50% deposit thickness. AFM is used directly on the membrane to determine the depth of the patterns. An additional deposit is then added,

followed by an AFM measurement which then leads to a reliable deposition rate. One or two more deposits usually results in completely filled trenches. No additional surface preparation is needed between electroplating steps. It is usually found that finer features plate at a lesser rate than coarse features. This transition occurs at about 100 nm feature size. The difference in rates is usually at least 10% and can be as large as 30%. This difference is most likely dependent on the condition (freshness) of the bath but can also occur if the plating rate is incorrect. An experiment done to demonstrate plating uniformity versus bath age is shown in Figure 82. The optimal rate for this system is approximately 25 nm per minute.

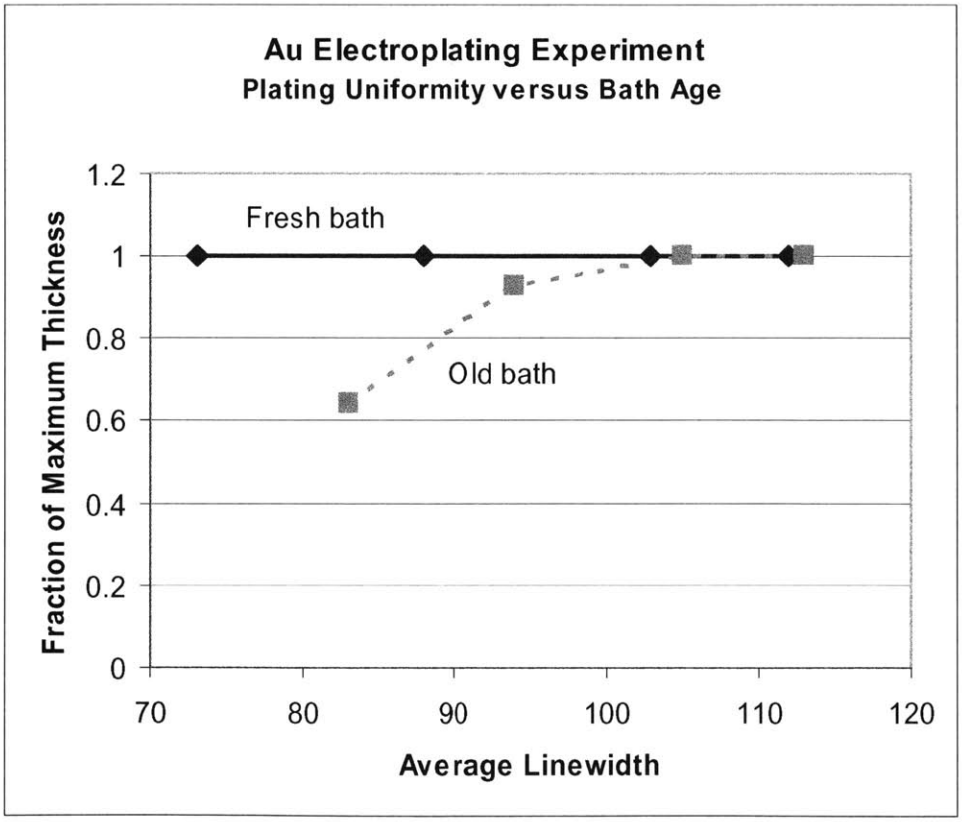


Figure 82 Au electroplating uniformity (over 80 to 110 nm linewidth range) for old versus fresh plating bath.

After electroplating, the PMMA can be stripped using NMP (@ 70° C) or chlorobenzene. It is usually necessary to remove the plating base in the vicinity of the alignment marks to improve the IBBI signal obtained from those marks. The plating base can be removed using a commercially available wet etch diluted

10:1 in DI. To do this selectively, windows in photoresist are created over the alignment mark regions using an optical microscope to expose the resist. Drops of the diluted etchant are applied for 10-20 sec at room temperature until sufficient plating base is removed (without harming the patterns). Finally, it must be taken into account that wafers intended for NMOS device processing are sensitive to Au contamination. To diminish the possibility of Au contamination on wafers, PVA (poly vinyl alcohol) can be spun onto the membrane. This material is water-soluble and will spin from 100 to 200 nm depending on concentration. A final mask cross-section is shown in Figure 83.

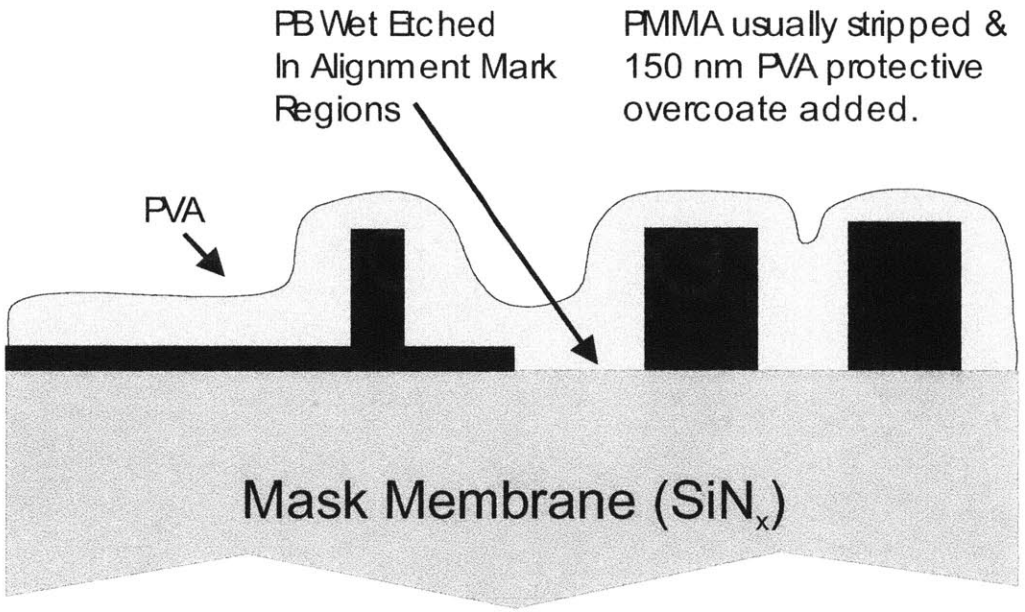


Figure 83 Final mask cross-section. PVA top-coat is used to prevent Au-contamination on NMOS device wafers. Removal of plating-base in alignment mark regions enhances IBBI signal strength.

Appendix C: Wafer Process Parameters

1	Recipe	Recipe	Machine	Step	Details	time	
	RCA ICL	RCA	RCA/ICL	SC1	(DI+NH4OH+H2O2)=5:1:1	10min	
				Dump rinse			
				50:1HF		15sec	
				Dump rinse			
				SC2	DI:HCl:H2O2=5:1:1	15min	
				Dump rinse			
				Dry			
2	Sac. Oxide	122	TubeA3/ICL		1000C Wet O2	2h10min	Oxide thickness= 6574A
3	Gate oxide	144	TubeA1/ICL		800C, DryO2	30min	Oxide thickness=45A
4	Gate poly	461	TubeA6/ICL		650C	29min	Poly thickness=1700~1800A
5	LTO	462	TubeA7/ICL		400C LTO	8min	LTO thickness=1000A
6	Denisify LTO	181	TubeA2/ICL		850C N2	1hour	
7	Top gate poly	461	TubeA6/ICL		650C	3m20s	Poly Thickness~150A
	Etch BG/TG stack						
8	Etch BG stack(Etch poly)	DUHEONCP	AME5000/ICL	main etch	350W, 90mtm, HBR:Cl2=20:30scc	7~10sec	Etch rate =80~100A depending on chamber condition
	Etch BG stack(Etch LTO)	SONGLTO			26" (500W, CHF3 30scc)	26~30"	Etch rate ~40A
	Remove PR	Piranha	Pre-metal/ICL		H2SO4:H2O2=3:1	10min	
	Undercut LTO	50:1HF	Pre-metal/ICL		50:1HF		Undercut rate= ~0.5nm/sec (30nm/min pre both side). Time depends on lindwidth
	Etch BG poly	DUHEONCP	AME5000/ICL	main etch	350W, 90mtm, HBR:Cl2=20:30scc	~15sec	Etch rate poly using LTO hardmask is ~1.4 times higher than that using PR mask!!.
		DUHEONCP	AME5000/ICL	over etch	150W, 90mt, HBR=30scc	~20sec	
9	Deposit LTO for CMP(twice)	462	TubeA7/ICL		400C LTO	80min	Thickness 9000~10000A
10	CMP (twice)	8	CMP/ICL		Table/Quil speed=25/16rpm Down force/Back pressure=4.0/6.0psi Slurry =150ml/min Pad down force=4psi	first CMP/second CMP total CMP time=16~17m 8min for each	After second CMP: Step height over 50~100um wide stripe: unnoticeavle Step height over 1mm wide pattern ~60A
11	Densify LTO before bonding	181	tubeA2/ICL		850C 1hour	1hour	This step was done before CMP each step after CMP depostion in JM07 In JM08, the densification was done after CMP just one time!
12	Wafer bonding	Manual bonding	gadget/ICL				Before bonding full rca was perfomed for SOI/handle wafer. Just before bonding, just SC1(10m)+50:1HF(15sec) was applied.
13	Post anneal	181	tubeA2/ICL		850C, N2	1~1hour 30min	Post anneal after CMP
14	SiNx depostion	460	tubA5/ICL			total 110min	first deposit SiNx with flat zone up for 50m after that, tilt wafers after boatout and deposit again for another 50min
15	Grinding wafers	Silicon valley microelectronics					Total thickness of bonded pair=1000~1100um The total thickness after grinding=610um. (the thickness of SOI wafer by grinding reduced to 60~110um)
16	etch SI substrate		TMAH/KOHhood/ ICL		TMAH 25% solution @ 85C	4h~4h30min	After ~4hour, the patterns can be seen.
	Etch TG stack on window						
17	Etch TG stack(Etch poly)	DUHEONCP	AME5000/ICL		350W, 90mtm, HBR:Cl2=20:30scc	7~10sec	Etch rate =80~100A depending on chamber condition
	Etch BG stack(Etch LTO)	SONGLTO	AME5000/ICL		26" (500W, CHF3 30scc)	26~30"	Etch rate ~40A
	Etch BG poly	DUHEONCP	AME5000/ICL	main etch	350W, 90mtm, HBR:Cl2=20:30scc	~15sec	Etch rate poly using LTO hardmask is ~1.4 times higher than that using PR mask!!.
		DUHEONCP	AME5000/ICL	over etch	150W, 90mt, HBR=30scc	~20sec	
	Etch gate oxide	SONGLTO	AME5000/ICL		350W, 90mtm, HBR:Cl2=20:30scc	10sec	
	Etch channel	DUHEONCP	AME5000/ICL		40" (150W, CHF3 30sec)	~40sec	In this step, the oxide thickness can be measured by Elipsometer.
	Etch LTO around Algn mark/SEM ine	SONGLTO	AME5000/ICL		40" (500W, CHF3 30scc)	~40sec	The step height over alignmark can be measured by P10

Appendix D: Geometry of Magnification Distortion In Point-source X-ray Systems

The purpose of this section is to show mathematically that the projection of patterns from a flat mask membrane onto a parallel sample surface using point source illumination results in perfectly linear displacement and magnification of the features. This may not be obvious because a point source produces spherical wavefronts that pass through the membrane onto the sample.

Consider the geometry in Figure 84 where radiation emits from a source on the Z-axis, passes through an arbitrary point on the mask, P_{mask}, and is absorbed on the sample at P_{sample}. This distance between source and mask is D and the distance between the mask and sample is the gap g. $\mathbf{r}_m$ is the position vector of a point on the mask while $\mathbf{r}_s$ is the position of its image on the sample. We wish to describe $\delta_r = \mathbf{r}_m - \mathbf{r}_s$, the difference in lateral position amongst mask and sample as a function of: gap, source-to-sample distance and $\mathbf{r}_m$. This is easy to do. If the path between P_{mask} and P_{sample} is projected onto the X-Z plane, we find the X component of $\delta\mathbf{r}$ by similar triangles

$$\delta_x = \frac{g}{D} X_m$$

and the X components of position vectors $\mathbf{r}_m$ and $\mathbf{r}_s$ can be related as follows

$$X_m + \delta_x = X_s .$$

Substituting the previous equation

$$X_s = X_m + \frac{g}{D} X_m = \left(1 + \frac{g}{D}\right) X_m = k\, X_m$$

where we introduce a scaling constant k = $1 + \frac{g}{D}$. Similarly, we find that

$$Y_s = k\, Y_m .$$

This simply means that components of $\mathbf{r}_m$ are scaled by a factor k to produce $\mathbf{r}_s$ and k only depends on gap and source-to-sample distance.

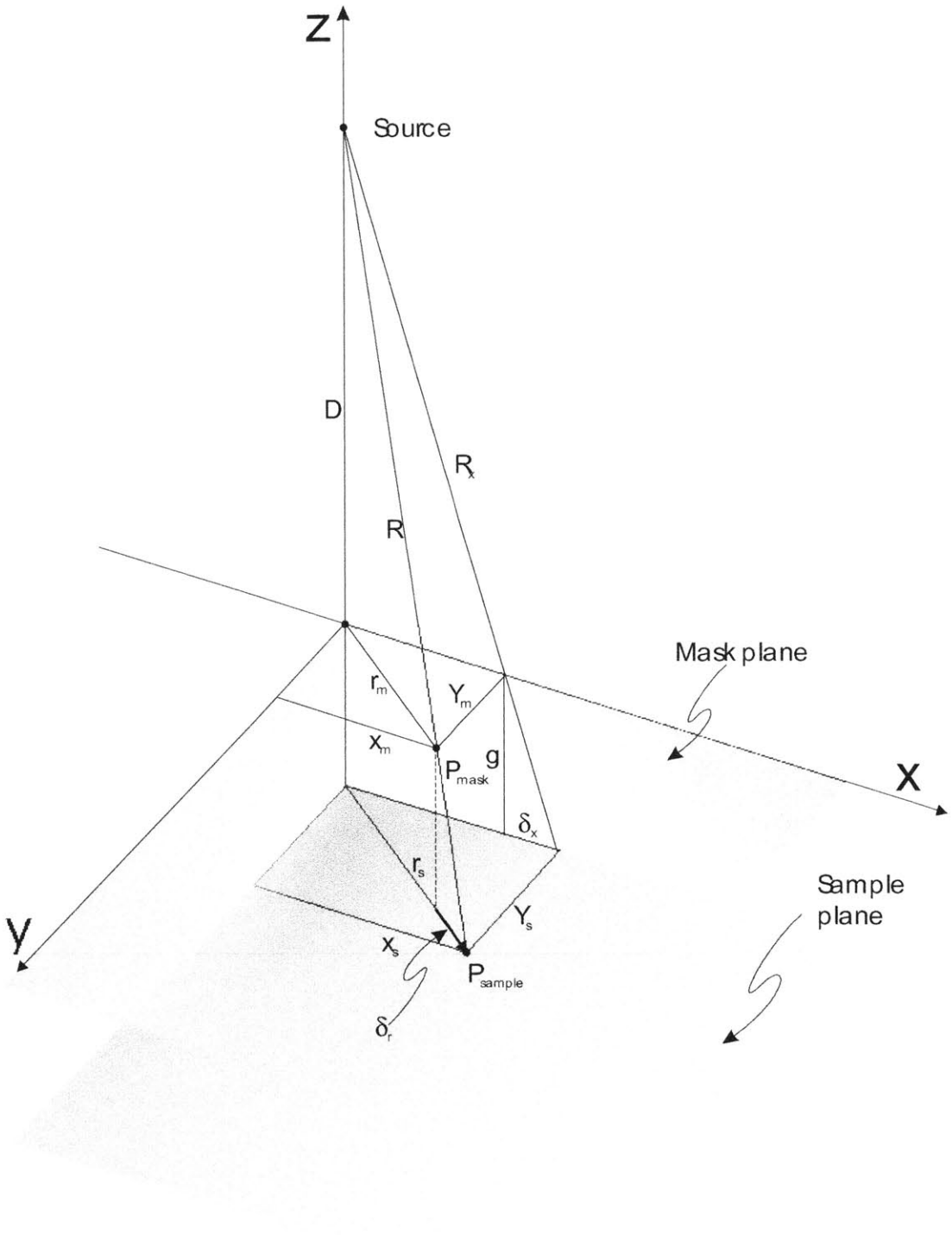


Figure 84 Geometry of point-source with mask plane parallel to sample plane.

In order to demonstrate magnification, consider an arbitrary vector on the mask and call it $\mathbf{r}_{\text{mask feature}}$ as shown in Figure 85. The length of $\mathbf{r}_{\text{mask feature}}$ is

$$\left|r_{mask\ feature}\right| = \left|r_2 - r_1\right| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}\ .$$

Now we know from the previous argument that in order to determine the length of this vector after it is lithographically projected onto the sample, we can simply scale all the components by k. This gives

$$\left|r_{sample\ feature}\right| = \sqrt{(kx_2 - kx_1)^2 + (ky_2 - ky_1)^2} = k\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

and therefore

$$\left|r_{sample\ feature}\right| = k\left|r_{mask\ feature}\right|$$

which is linear magnification. Clearly, the placement of $\mathbf{r}_{feature}$ will also be shifted by a factor of k. That is

$$\boldsymbol{r}_{sample\ position} = k\,\boldsymbol{r}_{mask\ position}$$

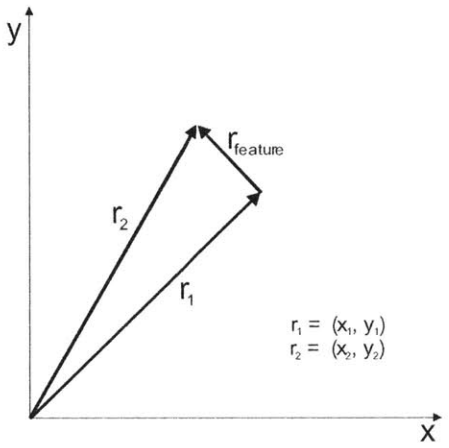


Figure 85 Arbitrary vector that represents a feature on the mask.

If the source-to-sample distance is 25 cm and the mask-to-sample gap is 3 μm, the magnification is 12 parts per million or k = 1.000012 Now consider a feature that is 2 mm from the alignment marks. Assuming zero alignment error, let's say two successive exposures are done for which the gap is changed from 3 to 4 μm between exposures. The difference in position of this feature on the sample would be 2 mm multiplied by k, or a shift of 24 nm!

References

[1] S.M. Kang, Y.Leblebici, "CMOS Digitial Integrated Circuits Analysis and Design", The McGraw-Hill Co., Inc., New York, 1996.

[2] B.Yu, Et al., Short Channel Effect Improved by Lateral Channel-Engineering in Deep-Submicrometer MOSFET's, IEEE Transactions on Electron Devices, Vol. ED-44, 1997, pp. 627-634.

[3] B.Davari, R.H.Dennard, G.G.Shahidi, CMOS Scaling for High Performance and Low Power – The Next Ten Years, Proceedings of the IEEE, Vol. 83, 1995, pp. 595-606.

[4] S.Wolf, "Silicon Processing for the VLSI Era, Volume III – The Submicron MOSFET", Lattice Press, California, 1995.

[5] T. Eimori, et al., Approaches to extra low voltage DRAM operation by SOI-DRAM, IEEE Transactions on Electron Devices, Vol. ED-45, 1998, pp. 1000-1009.

[6] J.W. Park, et al., Performance Characteristics of SOI DRAM for Low-Power Application, IEEE Journal of Solid State Circuits, Vol. 34, pp. 1446-1453, 1999.

[7] F.Balestra, Double Gate Silicon-On-Insulator Transistor with Volume Inversion: A New Device with Greatly Enhanced Performance, IEEE Electron Device Letters, Vol. EDL-9, pp. 410-412, 1987.

[8] H.S. Wong, et al., Nanoscale CMOS, Proceeding of the IEEE, Vol. 87, pp. 537-565, 1999.

[9] H.-S.Wong, D.Frank, Y.Taur, and J.Stork, "Design and performance considerations for sub-0.1μm double-gate SOI MOSFET's," in Proc. Int. Electron Devices Meeting, 1994, p747.

[10] E. Moon, "Dynamic nanometer alignment for nanofabrication and metrology", Thesis E.E. S.M., MIT, 1998.

[11] The International Technology Roadmap for Semiconductors, 2002 [Online]. Available: http://public.itrs.net/Files/2002Update/2002Update.pdf

[12] J.Z.Y. Guo and F. Cerrina, "Verification of Partially Coherent Light diffraction Models in X-ray Lithography", *J. Vac. Sci. Technology,* B 9, 3207 (1991)

[13] Q.-Y. Tong, U. Gosele, "Semiconductor wafer bonding: recent developments," *Materials Chemistry and Physics*, vol. 37, pp. 101-127, 1994.

[**14**] W.P.Maszra, "Wafer bonding: SOI, generalized bonding, and new structures," *Microelectronic Engineering*, vol. 22, pp. 299-306, 1993.

[**15**] Tsutomu Iida, et al., "Residual lattice strain in thin silicon-on-insulator bonded wafer: Thermal behavior and formation mechanisms," *Journal of Applied Physics*, vol. 87, pp. 675-681, 2000.

[**16**] H.D.Fuchs, et al., "High-resolution Raman spectroscopy of Ge-rich c-$Ge_{1-x}Si_x$ alloys: Features of the Ge-Ge vibrational modes," *Physical-Review-B-Condensed-Matter.* vol 44, no. 23, pp. 13120-13123, Dec. 1991.

[**17**] J.A. Floro, E.Chason, S.R.Lee, "Real time measurement of epilayer strain using a simplified wafer curvature technique," *Mat. Res. Soc. Symp. Proc.*, vol. 405, pp. 381-386, 1996.

[**18**] E.Chason, J.A. Floro, "Measurements of stress evolution during thin film deposition," *Mat. Res. Soc. Symp. Proc.*, vol. 428, pp. 499-501, 1996.

[**19**] P.H. Townsend, D.M.Barnett, T.A.Brunner, "Elastic relationships in layered composite media with approximation for the case of thin films on a thick substrate," *J. Appl. Phys.*, vol. 62, no. 11, 1987.