# PHOTOVOLTAIC RETINAL PROSTHESIS

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

LELE WANG

JUNE 2012



This dissertation is online at: http://purl.stanford.edu/pj955yy5239

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**James Harris, Primary Adviser**

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**Daniel Palanker**

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**Theodore Kamins**

Approved for the Stanford University Committee on Graduate Studies.

**Patricia J. Gumport, Vice Provost Graduate Education**

*This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.*

# ABSTRACT

Degenerative retinal diseases are among the leading causes of vision loss in the developed world. In these diseases, the "image capturing" photoreceptor layer slowly dies while the "image processing" inner retinal layers are preserved to large extent. One approach to restore vision of patients suffering from the selective photoreceptor loss is to deliver information to the visual system by patterned electrical stimulation of the remaining retinal circuitry. Several retinal prosthetic systems have achieved promising results in clinical studies, with the most successful prostheses allowing blind people to read large letters. However, the current retinal prosthetic designs require a trans-scleral cable to deliver power or data to the retinal stimulating array.

We designed a photovoltaic prosthetic system with a silicon photodiode array placed behind the retina, in which each pixel of the array photovoltaically converts patterned pulsed near-infrared light projected from video goggles into pulses of bi-phasic electric current to stimulate the nearby retinal neurons. The light-powered 30-μm thick photodiode array does not require any wired power connections, which greatly simplifies the surgical procedure and reduces chances of infection. Several arrays could be juxtaposed to tile a larger area on the retina and thereby expand the visual field. The local return electrodes in each pixel help to reduce the spatial spread of electric current, thereby reducing the cross-talk between pixels and improving the spatial resolution of the implant. Each pixel contains three photodiodes connected in series to increase charge injection levels. Deep reactive-ion etched trenches are used to isolate the three photodiodes in each pixel. Trenches between the pixels are open to allow perfusion of nutrient, which is essential for in-vitro electrophysiological experiments, and may also help with the in-vivo integration of the implant with the retina. We fabricated 1 and 2 mm wide arrays of photodiode for implantation in small and large animals. Three sizes of pixels (280 μm, 140 μm and 70 μm) were fabricated to explore the resolution that could be achieved by the implant.

The development of the fabrication process for the silicon photodiode array is presented, which involves eight mask layers, including deep isolation trench etch, $n^+$ region predeposition, $p^+$ region ion implantation, first via etching, first metal deposition

and liftoff, second via etching, polysilicon etching in selected trenches, second metal deposition and liftoff. A detailed optoelectronic characterization of the fabricated photodiode array is also presented. The initial electrophysiological results of retinal stimulation in-vitro and in-vivo show that the fabricated devices are able to reliably elicit retinal responses at safe near-infrared light irradiances, with good acceptance of the photodiode array in the subretinal space.

# ACKNOWLEDGEMENT

This dissertation would not have been possible without the help of many people who contributed their expertise and assistance that enabled the completion of this study.

First and foremost, I would like to thank my advisor, Professor James Harris, whose guidance and support is essential to my PhD life and will be a valuable asset for my future development. It has been a great time being a "Harris" group member, helping each other and being helped. I thank him for connecting me to this interesting and meaningful project where I can explore the area I am most interested in.

Professor Daniel Palanker, my co-advisor, who leads the retinal prosthesis project, effectively organizes the group to achieve exciting results and move the project forward. His insights on the future directions of the project guide us to achieve one milestone after another.

I thank the guidance and assistance of Dr. Ted Kamins and Dr. Keith Mathieson, whose expertise in silicon and semiconductor processing guides the fabrication in the right direction. Their support is of much importance for the smooth progress of the device fabrication.

Ludwig Galambos helped me to get familiar with the clean room in a short time, and I thank him for his generous help debugging the tools during tool operation.

I appreciate the help of Jim Loudin, Georges Goetz, Philip Huie and Professor Alexander Sher with the characterization of the fabricated devices. I thank them for sharing with me their knowledge of the biological systems.

I thank other graduate students in Harris group and Palanker group from whom I learned over the years, especially Angie Lin, who shared her knowledge with me about the MBE systems and patterned nonlinear optical waveguides.

I thank the SNF staff, for their help and support with various processing problems, and the SNL staff, for their instructions in using the SEM and FIB equipment.

I appreciate the help from Gail Chun-Creech, Heidi Wong and the administrative staff of the department of Electrical Engineering for the help with the administrative processes.

Last but not least, I thank my parents for being there for me in every situation and my friends for their supports.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# 1. INTRODUCTION

Degenerative retinal diseases are among the leading causes of vision loss in the developed world. The leading cause of blindness in the Western world, Age-Related Macular Degeneration (AMD), affects 0.2% of the population aged 55 to 64 years, and rises to 13% of the population older than 85 years [1]. The leading cause of inherited blindness, Retinitis Pigmentosa (RP), affects 1: 3943 persons according to a nation-wide study in Denmark [2]. In these diseases, the photoreceptor layer slowly dies while the inner retinal neurons are preserved to large extent [3, 4, 5]. Electronic retinal prostheses aim to restore vision of patients suffering from degenerative retinal diseases by electrically stimulating the remaining circuitry of the retina. This chapter starts with a brief introduction of the structure of the retina. Then various retina stimulation methods are introduced, with a comparison of the advantages and disadvantages of each method. Finally, an overview section discusses the overall organization of this thesis.

## 1.1 Retina

The human visual system detects light and converts it into neural signals, which allow visual perception of the surroundings. The eye consists of an optical imaging system, which including the cornea and crystalline lens, and a thin layer of tissue in the back of the eye, called the retina. A schematic diagram of the eye is shown in Figure 1.1. Light passes through the cornea and lens and is focused on the retina in the back of the eye. The retina transforms light signals into neural signals and relays the signal via the optic nerve to the brain.

The retina consists of several layers as shown in Figure 1.2. The signal output layer, the ganglion cell layer, is closest to the incoming light. While the photoreceptor layer, which is the signal input layer, lies in the outermost part of the retina. Light travels through the transparent retina to activate the photoreceptor layer and generate neural signals. The dark bottom layer, retinal pigment epithelium (RPE), absorbs the light penetrating beyond photoreceptors to prevent scattering inside the eye and metabolically supports photoreceptors. The outer segments of photoreceptors contain rhodopsin – a

light sensitive protein, which initiates conversion of photons into neural signals. There are two types of photoreceptors in the human retina: rods, which are mainly used in night vision; and cones, which operate at ambient daylight levels and enable color vision [6]. The neural signals from photoreceptors are processed by the inner neural layers and generate action potentials in the retinal output layer, the ganglion cell layer, which transmits the signals via optic nerve to the brain. Signals from about 100 million photoreceptors are delivered to the brain by about 1 million ganglion cells in the human retina [ 7 ]. Although very significant progress has been recently achieved in understanding the retinal signal processing, not all the details of it are fully understood.

Retinal signals are carried and processed by the neural cells. Each cell has a cell membrane that separates the interior of the cell from the extracellular surrounding electrolyte. Permeability of the ion channels in the membrane depends on the presence or absence of neurotransmitter molecules and on the trans-membrane potential drop. Trans-membrane proteins, called ion pumps, actively maintain the gradient of ion concentration across the cell membrane [ 8 ]. In an equilibrium state, the pumps maintain the intracellular potential at a polarized state of about -50 to -70 mV [9]. When retinal ganglion cell is polarized in external electric field, and trans-membrane potential is decreased below threshold value, sodium ion channels open initiating a rapid (1 ms) rise of intracellular potential, followed by a similarly rapid drop due to outflow of the potassium ions. The resulting pulse is called an action potential. The cell then returns to equilibrium with a brief hyperpolarizing potential overshoot after the spike. The threshold voltage step to elicit an action potential in the cell depends on pulse duration, and for 1 ms pulse it is about -5 mV, compared to the resting potential [8]. Stimulation of cells can be achieved intracellularly, by piercing a cell membrane, or extracellularly, by sending electric current through the bulk tissue. Extracellular stimulation creates a potential gradient outside the cell, causing parts of the cell to be depolarized and parts of the cell to be hyperpolarized. The locations of the polarized parts or hyperpolarized parts depend on the geometry of both the cell and the stimulation electrodes [10]. Stimulation of the non-spiking neurons, such as bipolar cells in the retina results in the change of their cellular potential, affecting the rate of release of neurotransmitters at synaptic connections to other neurons. This way stimulation of the non-spiking inner retinal neurons can lead to

generation of the action potentials in the post-synaptic neurons – the retinal ganglion cells, an effect called the network-mediated stimulation.

In diseases such as Retinitis Pigmentosa (RP) and Age-Related Macular Degeneration (AMD), the photoreceptor layer is significantly degraded while the inner retinal neurons are relatively well preserved [3, 4, 5]. Retinal prostheses aim at restoring vision by bypassing the degenerated photoreceptors and stimulating the remaining retinal circuitry to deliver visual information.

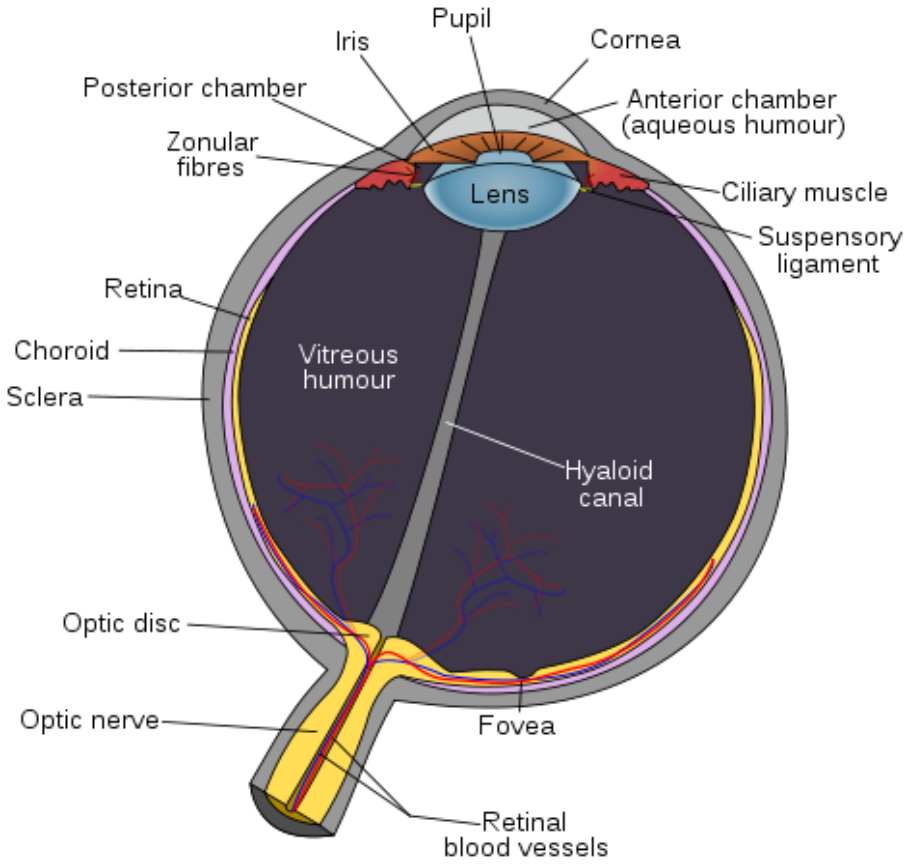



Figure 1.1. Schematic diagram of the structure of the eye (Image adapted from Wikipedia and is a part of the public domain). Light passes through cornea and lens and is focused onto the retina. The retina then converts the optical signals into neural signals and transmits them to the brain via the optic nerve.

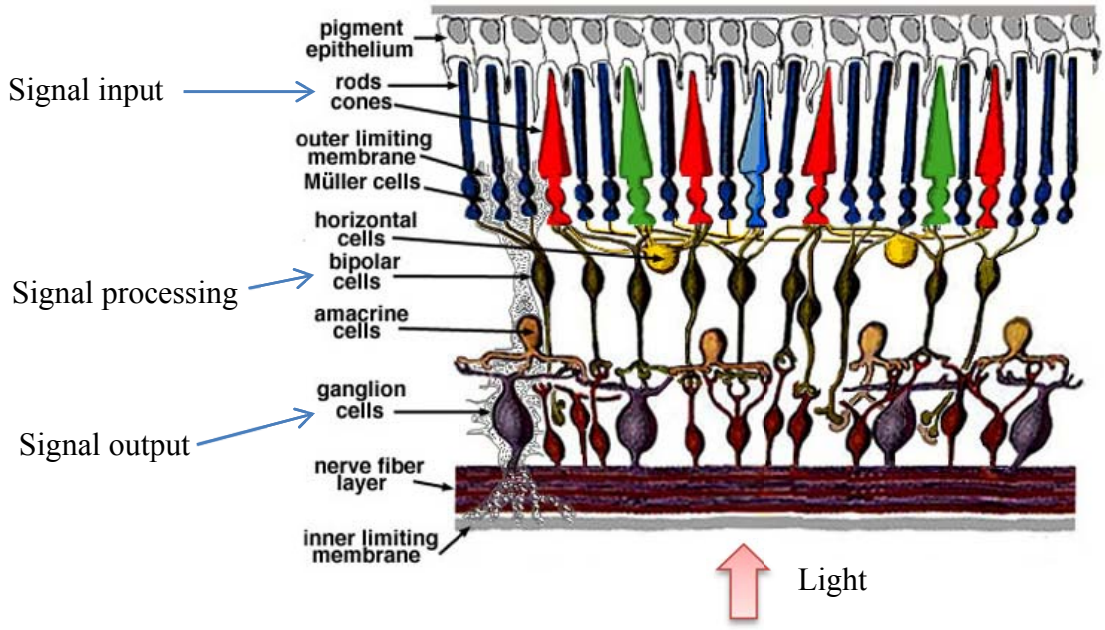



Figure 1.2. Schematic diagram of the layers of the retina (Image adapted from Webvision, http://webvision.med.utah.edu/). The retinal layers are organized into signal input, signal processing and signal output layers.

## 1.2 Prosthetic devices

The research of prosthetic devices that generate perceptions of light by electrical stimulation of the visual cortex was first carried out by Brindley et al. as early as the 1960's using an array with 80 platinum electrodes [11]. Dobelle et al. found that electrical stimulation of the visual cortex by multiple electrodes produces visual responses, which reinforces the possibility of a functional visual prosthesis [12]. However, cortical visual prosthesis needs to account for the very complex signal processing, and positioning the electrodes precisely in the primary visual cortex is a challenge [6]. In addition, power cables used in those systems introduce serious surgical challenges [13].

Retinal implants address visual system earlier and are better suited for proper retinotopic mapping of the visual scene. Depending on the location of the stimulating array, retinal implants are classified into two fundamental types: subretinal, where the implant is located between the retina and the RPE; and epiretinal, where the implant interfaces with the ganglion cells. Epiretinal stimulation of the retina in animal experiments was shown to produce localized retinal responses at low current levels [14,

15, 16]. Humayun et al. demonstrated that local electrical stimulation of the retinal surface results in focal light perception in patients blinded from end-stage RP [17]. By increasing the number of the stimulating electrodes from 3 to 25, patterned electrical stimulation of the retina of human subjects results in perception of simple forms [18]. Various aspects of retinal prostheses, including the encoder needed to replace the signal processing of the retina [19], the specific parameters needed to stimulate the retina [20], and major considerations to develop a retinal prosthesis [21] are being investigated. By implantation of a 16 electrode epiretinal implant, Humayun et al. reported perception of light in human subjects, with the location of the percept corresponding to the stimulated electrode and a stimulus threshold as low as 28 μA with 1 ms pulses[22]. In addition, the subjects could use the camera-driven stimulus to recognize simple shapes [23] and detect motion [24]. Recent studies with patients implanted with the epiretinal ARGUS II prosthesis system having 60 electrodes (Second Sight® Medical Products, Inc., Sylmar, CA) show improvements of visual function in certain tasks; some patients were able to identify letters, with equivalent visual acuity of 20/1200 [25, 26, 27]. However, the positioning of the epiretinal array close to the retinal neurons remains a challenge, which leads to higher stimulation thresholds. Since visual information is delivered from the camera to the retinal electrode array via RF transmission, the eye orientation does not affect the stimulation pattern, and therefore normal eye scanning is excluded from the visual perception. Patients are required to move their heads for scanning of the object [28]. Due to wired connection to the electrode array, the number of simultaneously addressed electrodes is difficult to scale up.

Subretinal prostheses involve implantation of a microphotodiode array between the bipolar cell layer and the RPE, with the goal of stimulating the neurons in the inner nuclear layer [29]. The advantages of the subretinal approach include proximity of the prosthesis to the remaining inner retinal neurons, and partial preservation of the retinal signal processing due to natural conversion of the stimulation of the bipolar cells into bursting of the ganglion cells [6, 30]. Chow et al. built a subretinal prosthesis (Artificial Silicon Retina (ASR, Optobionics Corp.)) consisting of about 5000 microphotodiodes equipped with micro-electrodes, with the belief that a subretinal implant could be powered solely by ambient light [31, 32, 33]. Although the subjects reported

improvements in visual perception [34], the prosthesis failed to stimulate adjacent neurons due to insufficient energy generated by the small light sensors with ambient light [35]. In addition, continuous ambient illumination cannot provide sustained stimulation current without irreversible electrochemical reactions at the electrode-electrolyte interface [8]. Zrenner and colleagues have been developing a microphotodiode array since 1996 [36, 37]. By implantation of their prosthesis in animals, they explored various parameters for successful electrical stimulation, long-term biocompatibility and stability of the subretinal implants [38, 39, 40]. The subretinal microchip they developed contains an array of 1500 photosensitive pixels, each having its own photodiode, stimulation electrode and amplifier, and an array of 16 wire-connected electrodes, which allows light-independent direct stimulation [41, 42]. Their clinical pilot study demonstrated that the subretinal implant provides meaningful visual information to blind patients [41]. The implant is powered via a cable connected to the extraocular power source, which greatly complicates the surgical procedure [41]. In addition, the lack of local return electrodes in each pixel is likely to increase the cross-talk between the pixels that have a common global return, resulting in limited resolution and contrast sensitivity.

## 1.3 The Stanford retinal prosthesis

In our optoelectronic prosthetic system, each pixel in the subretinal array photovoltaically converts patterned pulsed near-infrared (NIR) light projected from video goggles into pulses of bi-phasic electric current to stimulate the nearby retinal neurons [43, 44, 45]. Each pixel in the array of the subretinal implant converts the pulsed incident light into pulsed electric current. This current flowing through the retinal tissue between the central active electrode and circumferential return electrode in each pixel stimulates the nearby inner retinal neurons. Advantages of this system include: preservation of the natural link between eye movements and visual information; parallel transmission of visual information to each pixel, which allows scaling up the number of pixels; adjustable stimulation parameters (pulse intensity, duration and repetition rate) and image processing algorithms; and the capability to superimpose NIR images on any remaining natural vision [43, 44]. Operating the photodiode array photovoltaically avoids complex wired power connections and data decoding implants, which greatly simplifies the

surgical procedure, and reduces chances of infection. The local return electrode in each pixel helps to reduce the spatial spread of the electric current, which is likely to reduce the cross-talk between pixels and improve the spatial resolution of the implant. In the photodiode array, each pixel contains three photodiodes connected in series, with trench-isolation between them, to increase the charge injection levels. We also opened trench channels between the pixels to allow perfusion of nutrient through the implant. We fabricated three different sizes of pixels to explore the maximum resolution of the subretinal stimulation, and two sizes of the photodiode arrays for use in small and large animals. Expanding of the visual field is possible by juxtaposing several arrays to tile a larger area.

## 1.4 Thesis overview

This chapter is a brief review of the efforts in developing a functional retinal prosthesis aimed at restoration of sight to the blind. Chapter 2 discusses the various parameters taken into account in the design of the photodiode array. The detailed fabrication process of the photodiode array is presented in Chapter 3. The optoelectronic characterization results of the fabricated photodiode array are presented in Chapter 4. Initial electrophysiological results of retinal stimulation in-vitro and in-vivo are presented in Chapter 5; these results are the cooperative work of a large group of people. To improve the device performance, a redesign and fabrication of the next generation photodiode array is ongoing and will be presented in Chapter 6.

# 2. DEVICE DESIGN

This chapter will focus on the design parameters of the photodiode array. Topics include the material properties of the photodiode array, the structure of each pixel and the anti-reflection coating of the photodiode array.

## 2.1 Light absorption in silicon

Silicon only absorbs light in a certain wavelength range. The upper wavelength limit is determined by the energy band gap of silicon. If the energy of the photon is larger than the band gap energy, electron-hole pairs are photo-generated. If the energy of the photon is smaller than the band gap energy of silicon, silicon is transparent to the incoming light [46]. The optical wavelength (λ) can be converted to the corresponding photon energy by the following equation:

$$\lambda = \frac{1.24}{E_{ph}} \qquad \text{(Eq. 2. 1)}$$

where $\lambda$ is expressed in μm and $E_{ph}$ is in eV. For silicon with a band gap energy of 1.12 eV at 300 K, the expected response of silicon is very small for wavelengths above 1.1 μm. The absorption coefficient of silicon for wavelengths from ultraviolet to near infrared is provided by Green et al. [47]. At the wavelength of 880 nm used in much of this work, the absorption coefficient is 383 $cm^{-1}$, corresponding to an absorption depth of about 26 μm [47].

When light strikes a semiconductor material, part of it will be reflected and part of it will be transmitted. Here we assume that the light is monochromatic with a wavelength, λ, and the material is silicon with a thickness, L, as shown in Figure 2.1. If the corresponding energy of the photon is larger than the band gap energy of silicon, light will be absorbed as it passes through the material and generate electron-hole pairs. The intensity of light will decrease exponentially as described by the following equation:

$$I(x) = I_{inc} \exp(-\alpha x) \qquad \text{(Eq. 2. 2)}$$

where $\alpha$ is the absorption coefficient of silicon at the wavelength $\lambda$; $x$ is the depth; $I_{inc}$ is the initial light intensity when light just enters the silicon substrate.

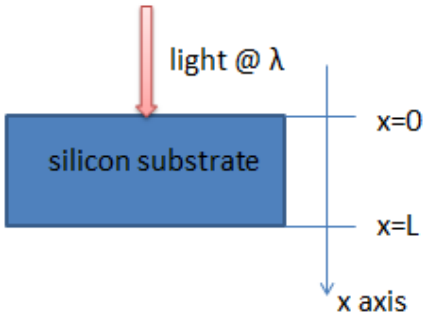



Figure 2.1. Light with a wavelength of λ penetrates into a silicon substrate with a thickness of L. The intensity of light decreases exponentially as described by (Eq. 2.2).

A plot showing the exponential decrease of the light intensity as a function of penetration distance is shown in Figure 2.2. We can see that at the penetration distance of 30 µm, the light intensity drops to about 30% of its initial value at 880 nm wavelength. Therefore, we chose to use silicon on insulator wafers with a device layer thickness of 30 µm to absorb a large portion of the incident light, while being still thin enough to be inserted beneath the retina.

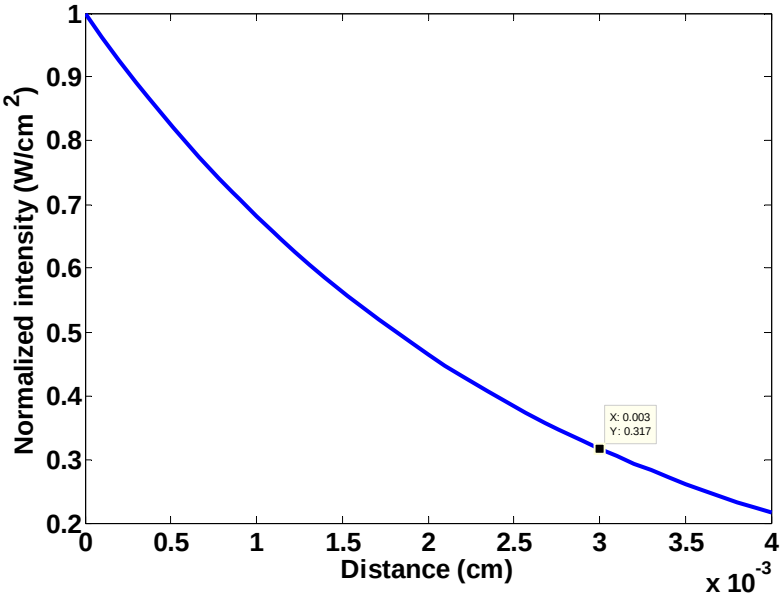



Figure 2.2. Normalized light intensity as a function of penetration distance. At the penetration distance of 30 µm, the light intensity drops to about 30% of its initial value for λ=880 nm.

## 2.2 Photodiode

Photodiodes are optoelectronic devices that convert light into electric current. A p-n junction photodiode can be made of any semiconductor in which a p-n junction can be formed. The light-to-current conversion process includes: absorption of light and generation of electron-hole pairs; diffusion of the generated carriers to the depletion region, with or without gain; collection of carriers by drifting across the depletion region to form photocurrent [48]. The external quantum efficiency η of a photodiode is defined as the ratio of the number of collected carriers to the number of incident photons by the following equation:

$$\eta = \frac{I_{ph}/q}{P_{inc}/h\nu} = \frac{I_{ph}}{q} \cdot \frac{h\nu}{P_{inc}} \qquad \text{(Eq. 2. 3)}$$

where $I_{ph}$ is the photocurrent; $P_{inc}$ is the incident light power; $v$ is the frequency of the incoming light; $h$ is the Planck's constant and $q$ is the charge of an electron. Another important factor in photodiodes is the photo-responsivity, defined as the ratio of the collected current to the incident light power, as described in the following equation:

$$R = \frac{I_{ph}}{P_{inc}} = \frac{\eta q}{h\nu} = \frac{\eta\lambda(\mu m)}{1.24}(A/W) \qquad \text{(Eq. 2. 4)}$$

where $\eta$ is the quantum efficiency defined in (Eq. 2. 3), and $\lambda$ is the wavelength of light in μm. For longer wavelengths with the corresponding photon energies smaller than the band gap energy of the material, the absorption coefficient is very small, so the expected photo-response is negligible.

When an external reverse bias voltage is applied to the p-n junction, the photo-generated carriers are accelerated in opposite directions by the reverse bias and give rise to a photocurrent. This kind of operation is usually called the photodetector mode. While in the solar cell mode, which is also called the photovoltaic mode, there is no external applied voltage. In the dark, there is a small reverse saturation current due to the drift of thermally generated minority carriers across the junction. If the junction is illuminated by light with photon energy larger than the band gap energy of the material, electron-hole

pairs are generated, and the resulting current-voltage characteristic is described by the following equation:

$$I = I_S(e^{qV_A/nkT} - 1) - I_{ph} \qquad \text{(Eq. 2. 5)}$$

where $I_s$ is the reverse saturation current; $V_A$ is the applied voltage; $I_{ph}$ is the photocurrent, and $n$ is the ideality factor. The important factors for the solar cell mode are the open circuit voltage $V_{oc}$, and short circuit current $I_{sc}$, which are defined as follows:

$$V_{OC} = \frac{nk_BT}{q}\ln[\frac{I_{ph}}{I_S} + 1] \qquad \text{(Eq. 2. 6)}$$

and

$$I_{SC} = I_{ph} \qquad \text{(Eq. 2. 7)}$$

Since the forward bias voltage in the fourth quadrant of the current-voltage characteristic under illumination is created by the photocurrent flowing in the opposite direction, this device can deliver power under illumination.

## 2.3 Metal-semiconductor contacts

Metal-semiconductor contacts that are Ohmic with low impedances regardless of biasing polarity are required for most electronic devices. In practice, Ohmic contacts are realized by heavily doping the semiconductor surface beneath the contact regions. When the surface of the semiconductor is heavily doped, any depletion region is very thin so that carriers can tunnel through. Especially, when the doping concentration exceeds about $10^{19}$ $cm^{-3}$, the barrier at the semiconductor and metal interface is so narrow that even low-energy majority carriers can easily tunnel through the potential barrier at the interface and transfer between the semiconductor and metal [46]. The tunneling model was first studied by Kröger et al. [49] and later pursued by others [50, 51, 52]. Contact resistances of Al and Pt on n-type silicon for concentrations ranging from $10^{18}$ to $2\times10^{20}$ $cm^{-3}$ measured at room temperature and liquid nitrogen temperature are in good agreement with the electron tunneling model [53]. In our devices, the $n^+$ and $p^+$ regions are usually heavily doped to a surface concentration larger than $10^{19}$ $cm^{-3}$ to reduce contact resistances.

## 2.4 Device structure

This section discusses the various parameters that are taken into consideration during design of the photodiode array. In each pixel of the photodiode array, we use three photodiodes connected in series to optimize the charge injection into the retina. Iridium oxide electrodes are used to provide a large charge capacitance. We have fabricated the cathodal version of the device where the current flows into the central electrode; the anodal version where the current flows out of the central electrode is under fabrication. The reasons for the above mentioned designs are discussed as follows.

### 2.4.1 Three series-connected photodiodes

The "water window", which is the range of potentials through which an electrode can be polarized without causing the electrolysis of water, is a frequently used safety criterion for physiological stimulation. The water window ranges over 1.4 V (-0.6 V to 0.8 V) [8]. At physiologically safe light intensities, the photo-voltage of one single silicon photodiode is limited to 0.5 volts [44]. To make full use of the water window, three photodiodes can be connected in series to increase the generated photo-voltage, enhancing the charge injection levels in retinal stimulation [45]. Assuming each diode is the same and is illuminated equally, with the series resistance taken into account, the I-V characteristic for m series connected photodiodes can be expressed in the following equation [8]:

$$I = I_S(e^{(\frac{V_A}{m} - I_D R_S)/(nV_T)} - 1) - I_p \qquad \text{(Eq. 2. 8)}$$

where $V_T=kT/q$ and is about 26.7 mV at body temperature (310 K) [54]; $R_s$ is the series resistance of each photodiode circuit. Although three photodiode connected in series can triple the voltage of one photodiode, it also requires three times as much light power for illumination to generate the same photocurrent [8].

### 2.4.2 Iridium oxide electrodes

Stimulation current is driven from an active electrode to a return electrode. Having a local return electrode will greatly decrease the interference between pixels, although it also reduces the field's penetration depth into the tissue [44]. Ionic current across an

electrode/electrolyte interface can be produced by capacitive coupling, as well as electron transfer due to electrochemical reactions at the electrode surface (faradaic processes) [43]. The maximum amount of charge that can be driven into the electrode without causing irreversible Faradaic reactions is called the charge injection limit, which depends on the electrode material and how the electrodes are driven [8]. Capacitive coupling between the electrode and the liquid medium is preferred since it does not involve any electrochemical reactions that could affect the tissue [43]. Therefore, materials with large surface area, such as TiN, can increase the capacitance at the electrode/electrolyte interface [55, 56]. Platinum electrodes have also been used in cell stimulation [18, 57]. Most of the studies have focused on the charge-injection and electrochemical properties of iridium oxide electrodes to maximize the electrochemical charge capacity within the water window [58-63]. The iridium oxide injects charges by a fast and reversible faradaic reaction, which involves reduction and oxidation between the $Ir^{3+}$ and $Ir^{4+}$ states of the oxide [43, 64]. The reactions involved are confined to the oxide film without any generation of soluble species required for charge transformation. The maximum amount of charge that can be delivered by an iridium oxide electrode without corrosion is 4 $mC \cdot cm^{-2}$ [43]. While for platinum, the corresponding value is only 0.15 $mC \cdot cm^{-2}$ to 0.4 $mC \cdot cm^{-2}$ [65, 66].

### 2.4.3 Cathodal vs. Anodal Configuration

When an electrode is driven more positively than the return electrode, it is said to be "anodally stimulated". If the electrode is driven negatively relative to the return electrode, it is said to be "cathodally stimulated" [67]. It has been reported that the subretinal stimulation threshold is several times higher for cathodal stimulation than for anodal stimulation [68-71]. In our currently fabricated devices, the $n^+$ regions are connected to the central active electrodes, which will result in the cathodal stimulation mode. The anodal version of the devices is under fabrication, where the $p^+$ regions are connected to the central active electrodes, aiming to reduce the threshold of retinal stimulation.

## 2.5 Anti-reflection coating

To improve the efficiency of light to current conversion, we designed an anti-reflection coating consisting of 60 nm of thermally grown silicon dioxide and 70 nm of silicon nitride, to minimize light reflection from the device surface in water. According to the multiple reflection method for electromagnetic waves propagating in layered dielectric structures [72], the reflectivity at the water/silicon nitride interface at 880 nm is plotted as a function of the silicon oxide and silicon nitride thicknesses as shown in Figure 2.3. We chose to have a combination of 60 nm of silicon oxide and 70 nm of silicon nitride, which will reduce the reflectivity to below 10%, while still providing a good passivation for the photodiode array.

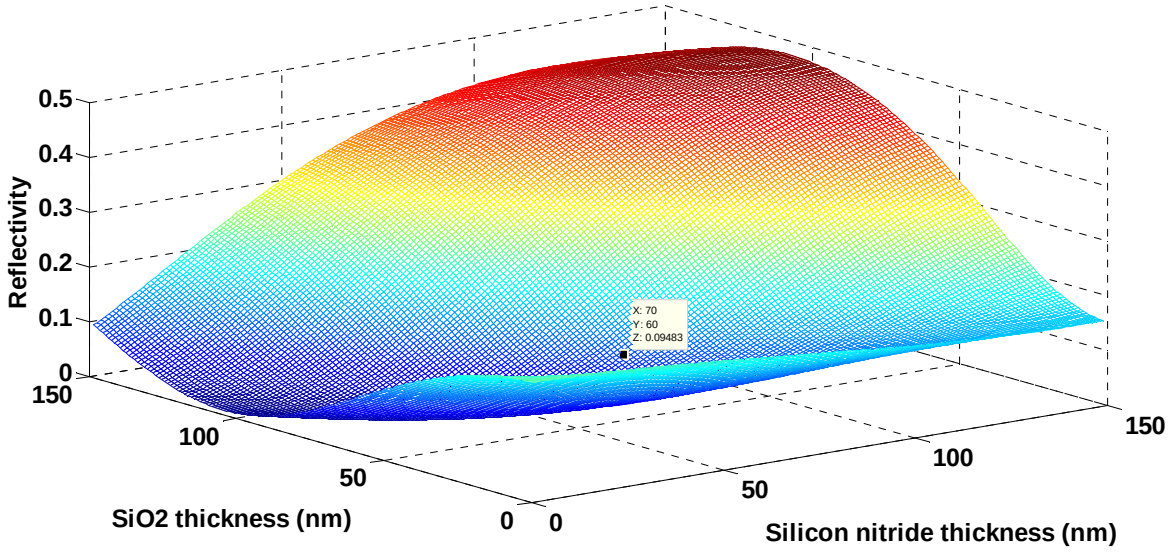



Figure 2.3. Reflectivity at the water/ silicon nitride interface as a function of the silicon nitride and silicon oxide layer thicknesses at the wavelength of 880 nm (Reprinted from L.Wang, K. Mathieson, T. I. Kamins, J. Loudin, L. Galambos, J. S. Harris and D. Palanker, Proc. SPIE 8248, 824805 (2012); http://dx.doi.org/10.1117/12.909104 ©SPIE).

# 3. DEVICE FABRICATION

After describing the background and motivation of this project, this chapter will focus on the detailed fabrication process of the silicon photodiode array. We start with a brief introduction of the process tools used to build the device. Then the detailed process conditions for each step are presented. The effects of various process parameters on the device performance will also be discussed and analyzed.

## 3.1 Introduction to the fabrication process

The photodiode array fabrication has been performed at the class 100 cleanroom of the Stanford Nanofabrication Facility (SNF). The fabrication process flow of the silicon photodiode array has incorporated many of the standard CMOS fabrication processes, such as standard cleaning procedures, photolithography, oxidation, dopant pre-deposition and ion implantation, plasma dry etching, chemical wet etching, low-pressure chemical vapor deposition, plasma-enhanced chemical vapor deposition, metal sputter coating and liftoff, chemical-mechanical polishing, etc., although many had to be adapted to the specific requirements of this array as a retinal prosthesis. Several fabrication processes that are popular to the Micro-Electro-Mechanical Systems (MEMS) fabrication, such as deep reactive-ion etching (DRIE) and photoresist spray coating, are also incorporated into the fabrication process to optimize the device performance for our specific bio-medical application. The mechanisms of the main process steps will be presented in this section.

### 3.1.1 Substrate cleaning

Since semiconductor device fabrication steps consist of introducing dopants with concentrations of parts per million, and depositing thin films with thickness control in the nm range, etc., unwanted impurities must be kept below the parts per million range, and stray particles should be eliminated. Front-end processes often involve high temperature processes, such as oxidation, annealing and thin-film deposition, where contaminants diffuse rapidly into the thin films or into silicon itself. Therefore, removing the contaminants from wafer surfaces by cleaning procedures is essential. For back-end processes, cleaning is also important but not as critical, since those processes are

normally at low temperature. Wafers are usually cleaned in a combination of chemical baths to remove the organics, metal traces and particles from the wafer surfaces [73].

In our fabrication process, the cleaning is done using the standard SNF clean room and wafer cleaning procedures. SNF is a class 100 clean room, where class 100 means in each cubic foot of air in the factory, there are less than 100 total particles greater than 0.5 μm in size. Wafer cleaning procedures include the initial cleaning of substrates, cleaning steps after photolithography and other procedures. For processing steps after most photolithography steps, the cleaning procedure usually begins with removal of photoresist. Photoresists are organic compounds which can be decomposed into $CO_2$ and $H_2O$ by an acid ($H_2SO_4$) and a strong oxidant ($H_2O_2$). Oxygen plasma cleaning is also used to convert photoresist to gaseous byproducts.

For photoresist stripping and cleaning, the wafers are immersed in 90% sulfuric acid ($H_2SO_4$)/10% hydrogen peroxide ($H_2O_2$) ("Piranha") at 120˚C, for stripping resist and removing wafer scribe dust. In pre-diffusion cleaning, the wafers first undergo a hot $H_2SO_4/H_2O_2$ bath at 90 ˚C for 10 minutes to further remove organic residues; then a short hydrogen fluoride (HF) (water: HF=50:1) dip at room temperature for 30 seconds to remove any native oxide on the wafer surface; then a hot hydrogen chloride (HCl)/$H_2O_2$ bath at 70˚C to remove inorganics, such as metals. Pre-diffusion clean is an essential step before any high-temperature furnace step to alleviate any problems of contaminant diffusion during high-temperature steps and to help prevent furnace contamination. Wafers that have metal on top will have a different cleaning procedure since the acid solutions will attack the metal. In our fabrication steps, wafers with metal on top are usually cleaned at the solvent wet bench by immersing the wafers in solvent solutions (acetone, methanol, and isopropanol) at room or elevated temperatures. Ultrasonic cleaning is also used to clean photoresist after the metal liftoff process, which will be discussed later.

### 3.1.2 Photolithography

Before the photolithography step, a mask set is designed using one of the commercial software tools--Tanner L-edit. The photolithography process uses conventional optical

lithography techniques, including; spin coating of photoresist, 5:1 reduction projection system for exposure, and standard photoresist development. The concept of the process is: a thin layer of photoresist that is sensitive to light is first spun onto the wafer surface; a light source then selectively exposes certain areas of the photoresist by shining light through a predesigned mask which contains the pattern information for each layer to be fabricated. The photoresist is then developed, resulting in pattern transfer from the mask to the wafer. The patterned photoresist is then used as a mask to etch an underlying thin film or to block the dopants from ion implantation, etc. For all of the lithography steps, we use positive photoresists because they generally have better resolution than do negative photoresists. Process steps, including exposure, baking and development should be carefully controlled and optimized to achieve diffraction limited resolution of the photoresist images.

In the fabrication process, the wafers are first treated in the prime oven, which bakes the wafer and dispenses the adhesion promoter, Hexamethyldisilazane (HMDS). Photoresists, in the form of liquid are then placed on the wafer using an automated track system for dispensing photoresist: SVG (Silicon Valley Group) coater. By spinning the photoresist at several thousand revolutions per minute (RPM), a uniform layer of photoresist is formed on the wafer surface; the thickness is determined by the viscosity and spin speed of the photoresist. The solvent composition of the photoresist determines its viscosity. Once the photoresist is spun onto the wafer, a prebake step is usually performed to drive off the remaining solvent from the photoresist. The baking process also improves the adhesion of the photoresist to the substrate. Stresses in the photoresist that were generated in the spinning process are relieved by thermal relaxation. For our process, specifically, the wafers are first baked and vapor primed in a 150 ˚C Yield Enhancement Systems (YES) oven. Positive photoresist of Shipley 3612 (Shipley Company, 455 Forest St. Marlborough, Massachusetts, 01752) is used to define photoresist patterns on the wafer. 3612 is a positive photoresist with a standard film thickness of 1 μm or 1.6 μm. After spinning the photoresist, a prebake at 90 ˚C for 1 minute is usually performed to drive off the remaining solvent.

Next, the exposure system transfers the mask image to a latent photoresist image. For positive photoresist, the latent photoresist image is removed in the developer later. The exposure tool we used during the fabrication process is a 5:1 reducing stepper with maximum die width of 18 mm and maximum die height of 22.6 mm. It is an i-line (365 nm) exposure system with alignment accuracy of 60 nm and typical resolution of 0.45 μm.

Then a post-exposure bake helps to minimize standing wave effects in the photoresist. Photoresist is then developed in standard developer (MEGAPOSIT MF-26A Developer, Rohm and Haas Electronic Materials LLC, 455 Forest Street, Marlborough, MA, 01752) on SVG develop tracks. A final post-bake step is then done on a 110 ˚C hot plate for 60 seconds for the 1.0 μm thick 3612 photoresist. The post-bake process completes the cross-linking of the photoresist and makes it strong enough to withstand subsequent processing steps, such as wet or dry etching processes.

### 3.1.3 Oxidation and characterization

#### *3.1.3.1 Silicon oxide*

The silicon surface is usually passivated by silicon oxide, which can be easily grown thermally on silicon or deposited. Silicon oxide can be used to block dopants and many other unwanted impurities, and to act as a hard mask during plasma dry etching. It is also a very good insulator and has very stable bulk properties. Because of the very low mechanical and electrical defect density, the silicon/silicon-oxide interface is stable over time. The oxidation process involves a volume expansion since oxygen atoms insert between silicon atoms to form Si-O bonds. There is a 2.2 times volume expansion of the oxide compared to the volume of the silicon that undergoes oxidization [73].

#### *3.1.3.2 Strain formation*

Oxide layers grown on silicon are amorphous since the lattice constant of the crystalline form of $SiO_2$, quartz, does not closely match that of the silicon substrate. At the silicon/silicon di-oxide interface, since the growing oxide can only expand upward but not laterally, the grown oxide layers are normally in compressive stress. For high oxidation temperatures above 1000˚C, the oxide can release part of the stress by viscous flow. While at low oxidation temperatures, there is not much stress relief because of the

high oxide viscosity. In addition, the large difference in the thermal expansion coefficients of silicon and $SiO_2$ result in an additional compressive stress in the oxide. The backside of the silicon substrate is in tension, which is not as serious since the silicon substrate is many times thicker than the oxide layer and the tension scales down in proportion to the thickness ratio. One possible issue during fabrication is that the oxidation process will grow oxide on both the front side and back side of the wafer and the stresses from the two oxide layers are balanced. But once the oxide is chemically removed from one side of wafer surface, for example, the backside of the wafer, a curvature will develop on the wafer, which may affect the alignment process during wafer exposure.

#### *3.1.3.3 Oxide defects*

The electrical properties of the silicon/$SiO_2$ interface have a significant effect on some of the device characteristics such as reverse-bias leakage current in diodes. Such effects have been extensively studied to determine the optimized process conditions and how to control them. Most of the defects at the interface are dangling or untied bonds of silicon due to incomplete oxidation of the silicon atoms. For today's device processing technology, only 1 out of $10^5$ atoms have such defects [73]. Electrical defects at the silicon/$SiO_2$ interface and in $SiO_2$ layers were extensively studied by Deal in the 1970s and 1980s [74]. There are several types of defects or charges present at the silicon/$SiO_2$ interface: fixed oxide charge, interface trapped charge, mobile oxide charge and oxide trapped charge [73]. Some of the charges come from broken Si-O bonds from ionizing radiation or other process steps. Therefore, after the exposure of the oxide to energetic ions, electrons and other neutral species during processes like plasma dry etching, a high-temperature anneal is required to repair the broken Si-O bonds. If the oxide is not well annealed after exposure to ionizing radiation, the traps in the oxide can capture the injected electrons and holes during device operation, which causes a shift in the device characteristics. A high-temperature anneal in nitrogen or argon before metal deposition processes and a lower temperature anneal at about 400 ˚C in forming gas (nitrogen/hydrogen) at the end of the process will minimize the various charges in the oxide. In our process, a 1000 ˚C anneal in nitrogen before the first via opening and 425

˚C annealing in forming gas (nitrogen/hydrogen) after the first metal deposition are performed to minimize the oxide charges.

#### *3.1.3.4 Oxidation systems*

Oxidation systems are horizontally or vertically oriented with oxidation species coming from the back of the tube. Quartz boats which support the wafers while they are in the tube can carry 50 or more wafers. For dry oxidation, oxygen is injected directly into the furnace. For wet oxidation, $O_2$ and $H_2$ are burned in the back of the furnace to produce steam. The temperature is maintained by resistive heating elements and monitored by thermocouples. In order to avoid crystallographic defects caused by large thermal gradients across the wafers, wafers are loaded at around 800 ˚C and then the furnace temperature is ramped up to the oxidation temperature after wafers are positioned in the central uniform-temperature zone. A control circuit will ramp the furnace temperature up and down to minimize thermal stress in the wafer [73].

#### *3.1.3.5 Process characterization*

Oxidation process characterization includes physical measurements, optical measurements and electrical measurements. For physical measurements, the grown oxide is etched away in selected regions, and a small needle stylus is used to record the height of the step at the edge of the etched film by mechanical force. During the fabrication process, cross-section scanning electron microscopy images are also used to calibrate the grown oxide thicknesses.

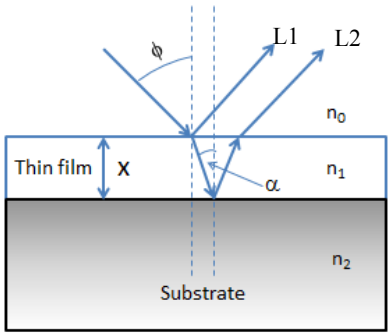



Figure 3.1. Light refraction and reflection at layered structure interface. Thin film with a thickness of X is grown on a thick substrate. The incident light angle is $\phi$ and the refraction angle is $\alpha$. Usually, $n_0$ is the refractive index of air; $n_1$ is the refractive index of the thin film whose thickness is to be measured; $n_2$ is the refractive index of the substrate material, which is usually silicon.

Optical characterization is shown in Figure 3.1. Monochromatic light with wavelength $\lambda$ incident from the air/thin-film interface is partially reflected (L1) and partially refracted. The refracted fraction of the light will be reflected again at the bottom thin-film/substrate interface; this reflected light then propagates back up to the thin-film/air interface and be refracted to become L2. If L1 and L2 are in phase with each other, they will interference constructively and create a maximum in the reflected light intensity, and vice versa to produce a minimum when they are out of phase. The maximum and minimum happen at wavelengths that satisfy the following equations:

$$\lambda_{\min,\max} = \frac{2n_1 X \cos\alpha}{m} \qquad \text{(Eq. 3. 1)}$$

where

$$\alpha = \sin^{-1}[\frac{n_0 \sin\phi}{n_1}] \qquad \text{(Eq. 3. 2)}$$

In the above equations, $m = n$ for maxima and $m = (2n-1)/2$ for minima, for $n$=1, 2, 3….

By sweeping the wavelength of the light with a fixed incident angle, the thickness of the film can be measured by relating the corresponding wavelengths for maxima and minima to the thin film thickness by (Eq. 3. 1) with a known refractive index $n_1$ of the thin film. This technique works very well for film thicknesses greater than a few tens of nanometers.

Color charts, which were originally described by Pliskin and Conrad [75], can be used to provide an estimate of the film thickness by observing the color of the thin film under white light illumination. The mechanism is that when white light is incident on a transparent film on a reflecting substrate, destructive and constructive interferences occur for specific wavelengths, resulting in a characteristic color of the reflected light.

#### *3.1.3.6 Modeling and simulation*

The first model to simulate the oxidation process is a linear-parabolic model developed by Deal and Grove in 1965 [76]. Oxide grows by diffusion of oxidizing species to the oxide/silicon interface. The chemical reaction is described by the following equation:

$$Si + O_2 \rightarrow SiO_2 \qquad \text{(Eq. 3. 3)}$$

or

$$Si + 2H_2O \rightarrow SiO_2 + 2H_2 \qquad \text{(Eq. 3. 4)}$$

In the Deal-Grove model, three sequential steps happen during the oxidation process, including: the transport of the oxidant in the gas phase to the oxide surface, diffusion of the oxidant through the oxide to the $Si/SiO_2$ interface and the reaction at the $Si/SiO_2$ interface. The oxide thickness can be calculated by the following equations:

$$x_o = \frac{A}{2}\left\{\sqrt{1+\frac{t+\tau}{A^2/4B}}-1\right\} \qquad \text{(Eq. 3. 5)}$$

where

$$\tau = \frac{x_i^2 + Ax_i}{B} \qquad \text{(Eq. 3. 6)}$$

In the above equations, $x_i$ or $\tau$ account for the initial oxide thickness before oxidation. *B* and *B/A* are the parabolic and linear rate constants respectively, which are extracted from experimental data.

The linear-parabolic model is widely used to calculate oxide thicknesses. However, for growth of thin oxide (less than 20 nm) in dry $O_2$, the Deal-Grove model cannot accurately predict the growth rate of the oxide. Other models proposed are from Reisman et. al. [77], Han and Helms [78], Ghez and van der Meulen [79]. Those models aim to improve the linear-parabolic model. However, none of them is as widely accepted as the linear parabolic model. Massoud and colleagues [80] did an extensive experimental study and reported that by adding a term to the Deal-Grove model, the oxide growth in the thin regime could be fitted as described by the following equations:

$$\frac{dx_o}{dt} = \frac{B}{2x_o + A} + C\exp(-\frac{x_o}{L}) \qquad \text{(Eq. 3. 7)}$$

where

$$C = C^0\exp(-\frac{E_A}{kT}) \qquad \text{(Eq. 3. 8)}$$

Process simulation software such as SUPREM IV has integrated the above models to better match the experimental data over the whole thickness range.

The oxidation rate is also dependent on pressure because of the linear and parabolic coefficients, B and B/A are pressure dependent. The crystal orientation also affects the oxidation rate because of the different surface density on different crystal planes [73].

Shaped silicon structures have a different oxidation rate than flat surfaces. Kao et al. [81, 82] found that retardation of oxidation rate is very significant at sharp corners, a factor of two in oxide thickness for nominally 500 nm oxide. The retardation is more pronounced for low-temperature oxidations than that for high temperatures, with virtually no corner effect for oxidations done at 1200 ˚C. And the interior (concave) corners show a more pronounced effect than exterior (convex) corners, with both of them significantly retarded compared to flat surfaces.

It has been shown many times that the oxidation rate of highly doped substrates is much higher than that of lightly doped substrates, especially at lower temperatures and for thinner oxides. This effect is also more pronounced for regions that are heavily doped n-type compared to p-type. One possible explanation is that for highly doped regions, the total number of vacancies is much higher than for lightly doped material and the vacancies are able to provide sites for oxidation reactions. For $p^+$ regions, since the total vacancy number is less than for $n^+$ regions and dopants tend to deplete into silicon oxide, dopant effects on the oxidation rate are not as significant as those for the $n^+$ regions.

Most of the process steps and complex geometries can be simulated by a computer program SUPREM IV to better understand various process steps.

### 3.1.4 Dopant diffusion and characterization

For electronic devices, the control of active doping regions is critical to obtain the desired electrical characteristics. For a uniformly doped region, the current density $J$ is:

$$J = nqv = nq\mu\varepsilon \qquad \text{(Eq. 3. 9)}$$

where $n$ is the doping, $q$ is the electron charge, and $v$ is the velocity, which is the product of the mobility $\mu$ and the electrical field $\varepsilon$. The conductivity $\sigma$ and resistivity $\rho$ of a doped region are defined as:

$$\sigma = nq\mu \qquad \text{(Eq. 3. 10)}$$

$$\rho = \frac{1}{nq\mu} \qquad \text{(Eq. 3. 11)}$$

The resistivity is the resistance of a cube with any dimensions. In device design, it is useful to specify the sheet resistance of any doped region without knowing many specific dimension parameters. Sheet resistance is defined as the resistance measured between two sides of a square with a junction depth of $x_j$, as described in Figure 3.2.

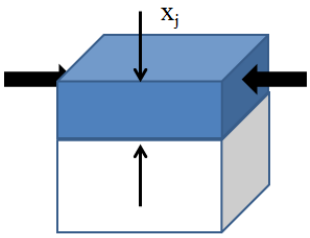



Figure 3.2. The sheet resistance is the resistance between the two sides of a square of any dimension with a junction depth of $x_j$.

Sheet resistance $\rho_s$ is described by the following equation:

$$R = \frac{\rho}{x_j}\Omega / sq \equiv \rho_s \qquad \text{(Eq. 3. 12)}$$

Two methods are widely used to introduce dopants. Predeposition has been used since the 1960s to controllably introduce a certain amount of dopant into silicon, followed by solid-phase diffusion from glass layers deposited on wafer surfaces or high-temperature gas-phase depositions in a furnace. However, predepositions are limited to introducing dopants at solid solubility levels. An alternative to the predeposition method is ion implantation, which has far better control of the dose and is not limited by solid solubility. Implantation became popular from the mid-1970s as it is highly flexible and manufacturable. However, it causes damage to the silicon crystal. Therefore, a high-temperature annealing process is required to anneal the crystal structure of silicon, which will cause dopant diffusion and redistribution, which will be discussed later.

#### *3.1.4.1 Diffusion models*

Solid solubility is defined as the maximum concentration of a dopant that can be dissolved in silicon under equilibrium conditions, without forming a separate phase [83]. Electrical solubility is the corresponding electrically active dopant concentration that may be considerably lower than solid solubility. Dopants above the electrical solubility limit

form electrically-neutral inactive complexes which do not contribute to the free carrier concentration.

The diffusion process is described by Fick's law, which relates the diffusion of dopant atoms to the concentration gradient. Fick's second law of diffusion which relates the change of concentration with time to the concentration gradient can be described by the following equation:

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \qquad \text{(Eq. 3. 13)}$$

where $D$ is the diffusivity.

Two of the analytic solutions of Fick's law that are interesting for silicon processing are the Gaussian solution and error-function solution. For the Gaussian solution in an infinite medium, a solution of Fick's second law that satisfies the boundary condition is:

$$C(x,t) = \frac{Q}{2\sqrt{\pi Dt}}\exp(-\frac{x^2}{4Dt}) = C(0,t)\exp(-\frac{x^2}{4Dt}) \qquad \text{(Eq. 3. 14)}$$

where $Q$ is the introduced dose and $C$ is the concentration gradient. When a fixed dopant dose $Q$ is introduced at a surface such as during ion implantation, the diffusion equation for the dopant to diffuse near the surface can be described as:

$$C(x,t) = \frac{Q}{\sqrt{\pi Dt}}\exp(-\frac{x^2}{4Dt}) = C(0,t)\exp(-\frac{x^2}{4Dt}) \qquad \text{(Eq. 3. 15)}$$

Another solution is the error-function; in an infinite medium, the solution of the diffusion equation is:

$$C(x,t) = \frac{C}{2}\left[1 - erf\frac{x}{2\sqrt{Dt}}\right] \qquad \text{(Eq. 3. 16)}$$

with $C$=0 at $t$=0 for $x$>0 and $C$=C at $t$=0 for $x$<0.

The error function solution is valid when the surface concentration is fixed, such as the diffusion from a gas ambient with concentrations higher than the solid solubility of the dopant in the solid. The corresponding diffusion profile can be described as:

$$C(x,t) = C_s\left[1 - erf\left(\frac{x}{2\sqrt{Dt}}\right)\right] \qquad \text{(Eq. 3. 17)}$$

where $C_s$ is the constant source concentration.

The error function solution is used when there is an infinite supply of dopant, which means the dose that is introduced into the substrate is increasing to maintain a constant surface concentration. For a Gaussian solution, the dose is a constant, so the surface concentration will drop as the dopants diffuse into the bulk.

#### *3.1.4.2 Dopant diffusion coefficients*

The diffusion coefficients of most of the dopants in silicon are exponentially dependent on temperature, as follows:

$$D = D^0 \exp\left(\frac{-E_A}{kT}\right) \qquad \text{(Eq. 3. 18)}$$

where $k$ is the Boltzmann constant and $T$ is the temperature in Kelvin. $E_A$ is the activation energy with units of eV; $E_A$ is usually 3.5 eV for boron and 3.68 eV for phosphorus. $D^0$ is 1.0 $cm^2$ $sec^{-1}$ for boron and 4.70 $cm^2$ $sec^{-1}$ for phosphorus.

For multiple diffusion steps, if each of the diffusion steps is at a constant temperature with the same diffusivity, the effective $Dt$ product is as follows:

$$(Dt)_{eff} = D_1t_1 + D_2t_2 + \ldots \qquad \text{(Eq. 3. 19)}$$

In order to describe the dopant profiles of modern semiconductor devices more accurately, numerical solutions are needed which take into account the electric field effects, concentration-dependent diffusion, dopant segregation at the interfaces between different materials, and interfacial dopant pileup for shallow junctions. These effects are discussed in [73] and implemented in process simulation programs like SUPREM IV.

#### *3.1.4.3 Characterization*

Sheet resistance of a doped region is usually measured by the 'four point probe' method, as shown in Figure 3.3. Typically, the surface is contacted by four probes, with the outer two probes forcing a current $I$ and the inner two probes measuring a voltage $V$

to obtain the resistivity. The resistivity of a doped region can be described by the following equation:

$$\rho = \frac{1}{q\mu_n n + q\mu_p p}\Omega \cdot cm \qquad \text{(Eq. 3. 20)}$$

where $n$ and $p$ are the carrier concentrations, $\mu_n$ and $\mu_p$ are the mobility of electrons and holes.

The sheet resistance calculation with various correction factors is discussed in [84] and can be expressed by the following equation:

$$\rho_s = \frac{\pi}{\ln 2}\frac{V}{I} = 4.532\frac{V}{I}\Omega/square \qquad \text{(Eq. 3. 21)}$$

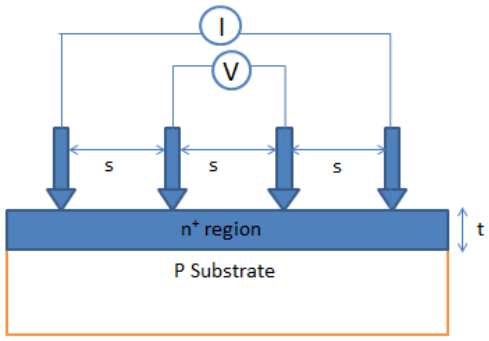



Figure 3.3. Four probe measurement of sheet resistance of a heavily doped n region on p-type substrate. The outer two probes force a current to flow through the sample; the inner two probes measure the voltage drop.

### 3.1.5 Ion implantation

#### *3.1.5.1 Models*

Ion implantation is the dominant doping method for the past 30 years because of its ability to control precisely the distribution and dose of various dopants in silicon. The distribution of implanted ions can be described to the first order by a symmetric Gaussian distribution by:

$$C(x) = C_p \exp(-\frac{(x-R_p)^2}{2\Delta R_p^2}) \qquad \text{(Eq. 3. 22)}$$

where $R_p$ is the average projected range normal to the surface, $\Delta R_p$ is the standard deviation and $C_p$ is the peak concentration where the Gaussian distribution is centered. The dose, which is defined as the total number of implanted dopant ions, can be calculated by integrating the above equation which gives:

$$Q = \sqrt{2\pi}\Delta R_p C_p \qquad \text{(Eq. 3. 23)}$$

Usually, a relatively thick photoresist or silicon dioxide layer is used as a mask layer to block the dopant implantation. The range and standard deviation are very similar in silicon and silicon dioxide [73]. The criterion for efficient masking can be described as:

$$C^*(x_m) = C_p^*(\exp - \frac{(x_m - R_p^*)^2}{2\Delta R_P^{*2}}) \le C_B \qquad \text{(Eq. 3. 24)}$$

where $C^*(x_m)$ is the concentration inside the mask at the mask/substrate interface, $x_m$ is the thickness of the mask and $C_B$ is the background concentration in the substrate. Setting $C^*(x_m) = C_B$, the required mask thickness $x_m$ can be expressed as:

$$x_m = R_P^* + \Delta R_p^* \sqrt{2\ln(\frac{C_P^*}{C_B})} = R_P^* + m\Delta R_p^* \qquad \text{(Eq. 3. 25)}$$

where $m$ is determined by the levels of masking efficiency.

Because of the symmetry of the crystalline lattice of silicon, planar and axial channels can have a significant effect on the implant profile. Once an ion enters a channel, it will experience small-angle scatterings from the atoms that line the walls of the channel and travel a long distance along the channel before coming to rest [73]. The channeling effect causes a tail in the implant dopant profile which continues an unexpected distance into the substrate. Higher implanted doses are less prone to channeling effects as the crystal structure is damaged by the implanted ions. A screen oxide, which is amorphous, is often used to minimize channeling by randomization of the incident beam.

#### *3.1.5.2 Mechanism*

The reason for the energy loss of ions was proposed by Bohr to be due to positive atomic cores of the target and the electronic energy loss to the free electrons in the target. Due to the damage caused by the ion implantation process, most dopants are not

electrically active after the ion implantation process. A high temperature anneal process will activate most of the dopants. The anneal process will remove the damage caused by the implant and restore the silicon lattice to its perfect crystalline state. The amorphous substrate can regrow layer by layer by solid-phase epitaxy. However, for higher-dose implants, the End-Of-Range (EOR) defects, which are stable dislocation loops at the interface of amorphous and crystalline silicon after anneal, are difficult to remove and can cause large leakage current if they are located in the depletion region of a pn junction [73].

### 3.1.6 Etching

When a thin film or a silicon substrate needs to be selectively etched to create certain patterns on the wafer surface, a "wet" or "dry" etching process is used. Dry etching involves using gas-phase etchants in a plasma; dry etching usually occurs by a combination of physical and chemical processes. Photoresist, silicon oxide or silicon nitride can be used as an etch mask. Oxide and nitride stand up better in plasma etching conditions and are usually called hard masks. Selectivity and directionality are the two main parameters considered during etching. Selectivity is the ratio of the etch rate of the film material to be removed to the etch rate of the mask material, with a reasonable etch rate of the film. Directionality is a measure of the etch rate in different directions. If the etch rate laterally is the same to the etch rate vertically, the etch process is called isotropic etching. Anisotropic etching has less lateral etching. For complete anisotropic etching, there is no lateral etching. The etch directionality and selectivity are related to the physical and chemical components of the etch process: the more the physical component, the more the directionality of the process, with less selectivity; while the more the chemical component, the more the selectivity of the process, with less directionality. Other requirements during an etch process are the ease of delivery of the etch gases and liquids to the wafer and easy removal of the etch byproducts. This means that the byproducts of the wet etching process need to be soluble, and the byproducts of the dry etching process need to be gaseous or volatile [73].

#### *3.1.6.1 Chemical wet etching*

During wet etching, wafers are immersed in the etchant solution and the exposed material is removed by chemical processes. The common chemical reaction for silicon di-oxide can be expressed as follows:

$$SiO_2 + 6HF \rightarrow H_2SiF_6 + 2H_2O \qquad \text{(Eq. 3. 26)}$$

Ammonium fluoride ($NH_4F$) is often added to HF to keep the etch rate constant over time; this etchant is known as "Buffered Oxide Etch".

Most wet etching processes are isotropic, with the exception that some wet etching rates are dependent on crystallographic orientation. For example, for KOH etching of silicon, the (110) plane is the fastest etching surface since it has a more corrugated atomic structure than the (100) and (111) planes. Since the (111) plane is very closely packed, the etch rate is extremely slow [ 85 , 86 ]. The orientation dependence of tetramethylammonium hydroxide (TMAH) is similar to KOH [85, 87]. The temperature and composition dependence of the KOH and TMAH etch rates are also discussed and summarized in [85].

#### *3.1.6.2 Plasma dry etching*

Plasma dry etching is preferred in some of the fabrication processes because of its capability of directional etching. The system uses a low pressure (1 mtorr – 1 torr) gas and two electrodes with high voltage applied between them. First, the high electrical field across the two electrodes causes ionization of the gas atoms, generating a plasma which is partially ionized gas consisting of equal number of positively charged ions and negatively charged electrons, a number of unionized neutral molecules, ionized fragments of broken-up molecules, and free radicals. The reactive neutral species diffuse to the surface of the wafer to have chemical reactions with the material to be removed, forming volatile byproducts. The ions provide the physical part of the etching which performs the physical etching or sputtering. When both reactive neutral species and ions act together, the etching process is called ion-enhanced etching.

#### 3.1.6.2.1 Chemical etching component

Chemical etching of material usually arises from free radicals, which are neutral species, such as the fluorine free radical F, which comes from $CF_4$ and electrons as the following reaction:

$$e^- + CF_4 \rightarrow CF_3 + F + e^- \quad \text{(Eq. 3. 27)}$$

The incomplete bonding structure of free radicals makes them highly reactive. F has seven electrons in its outer shell, resulting in the fast reaction with other species to pair with the electrons for a more energetically favorable state. For the etching of silicon, the reaction of fluorine free radicals with silicon can be described by:

$$4F + Si \rightarrow SiF_4 \quad \text{(Eq. 3. 28)}$$

The byproduct $SiF_4$ can be removed from the surface as a gas. Certain amount of oxygen is often added to the $CF_4$ plasma to reduce the recombination of the dissociated $CF_4$ species ($CF_3$, $CF_2$) with F, which enhances the chemical etching.

Chemical etching is nearly isotropic since the free radicals are neutral and have an angular distribution, as well as low sticking coefficients (defined as the ratio of the reacted or deposited flux of atoms attaching to the surface to the flux of incident atoms). Since free radicals or reactive neutral species react by chemical reactions, the process is selective.

#### 3.1.6.2.2 Physical etching component

Because of the high electrical field developed between the plasma and the electrodes, ions will be accelerated toward the wafer surface, strike the atoms at the surface of the material and dislodge them, resulting in a highly directional and anisotropic etch. The etch process is more physical and less selective.

#### 3.1.6.2.3 Ion-enhanced etching

In ion-enhanced etching, both the chemical components and physical components take part in the etching process, which achieves good selectivity as well as anisotropic etch profiles. The possible explanation is that ion bombardment enhances the surface adsorption or etching reaction or helps with the removal of the byproduct of the chemical etch process. Some chemically inert residue or species from the plasma may deposit on

the surface, which inhibits chemical etching by blocking chemical-etching species or reacting with them. Ion bombardment can help to remove or prevent the formation of the inhibitors, resulting in anisotropic etching due to the directionality of the ions. If the ion flux is not perfectly normal to the wafer surface or the inhibitor formation or deposition rate is higher than the etch rate of the inhibitor, bowed or sloped sidewalls can form. The sidewall slope can be tailored by changing the deposition/etch rates.

#### 3.1.6.2.4 Reactive ion etching systems

For systems with two parallel plate electrodes, the electrode on which the wafers sits can be made smaller than the other electrode, resulting in a large voltage drop from the plasma to the wafers and more energetic ion bombardment of the wafer. In this Reactive Ion Etching (RIE) mode, both the physical ion component and the chemical reaction part are important. The directionality of the etching can be improved by lowering the gas pressure (10-100 mtorr), which reduces ion collisions during transit.

#### 3.1.6.2.5 Focused-Ion-Beam technique

In the Focused-Ion-Beam (FIB) technique, ion beams can be focused to a small diameter and scanned to sputter etch very small and specifically defined areas. The commonly used ions are $Ga^+$, $Ar^+$, $Xe^+$, $He^+$ and $H^+$. In our process, the FIB technique is used to selectively etch metal lines connecting the common ground electrode of the pixels.

#### 3.1.6.2.6 Models

The simplest model of etching assumes that the chemical component and physical component of the etching process are independent of each other and can be combined linearly. In this model, the etch rate of each of the components is assumed to be linearly proportional to the flux of the corresponding etching component, which results in:

$$\text{Etch rate} = \frac{(S_C K_f F_C + K_i F_i)}{N} \qquad \text{(Eq. 3. 29)}$$

where $S_C$ is the sticking coefficient defined as the ratio of the reacted or deposited flux of atoms on the surface to the flux of incident atoms. $N$ ($cm^{-3}$) is the density of the film being etched. $F_C$ is the flux of the reactive neutral species or free radicals, and $F_i$ is the ion flux. $K_f$ and $K_i$ are relative rate constants, which are usually used as fitting parameters.

The linear model works in cases where the etch process is purely chemical or physical, or the chemical and physical etching components act independently of each other.

When the chemical and physical components act together, such as in ion-enhanced etching, the saturation/adsorption model is better suited to calculate the combined etch rate. In ion-enhanced etching, the etch rate is not simply the linear addition of the chemical and physical components, but is usually much higher. Usually, both components are important in the etching process. Without ion bombardment to remove the inhibitor layer, the chemical etchant may not attach to the surface to etch the material; without the chemical etching component, the physical etching by ions may be negligible. From the saturation/adsorption etch model, the etch rate can be expressed as:

$$\text{Etch Rate} = \frac{1}{N}\left(\frac{1}{K_i F_i} + \frac{1}{S_C F_C}\right)^{-1} \qquad \text{(Eq. 3. 30)}$$

where the definition of the parameters is the same as in that of (Eq. 3. 29). As can be seen from the above equation, when there is no ion flux or chemical flux, the etch rate is zero. When the ion flux is increased with a certain chemical flux, the etch rate will increase linearly initially but saturates later, and vice versa. The slower process, either chemical or physical, will limit the combined etch process.

### 3.1.7 Thin film deposition

Various materials, including silicon oxide, silicon nitride, polysilicon, and various metals may be deposited during semiconductor device fabrication processes. For thin film deposition, processes need to be well controlled to achieve the desired composition, mechanical strength and electrical properties, while keeping contamination levels and defect density low enough. Strong adhesion of the deposited film to the underlying layers is also important. With nonplanar wafer topography, the film usually becomes thinner across the steps, which may cause poor step coverage. For example, metal lines deposited across steps may become very thin or discontinuous, resulting in a high resistance and possible mechanical cracking and failure. Conformal coverage provides the same film thickness on the top of the steps and on the sidewalls of the steps. Another issue is the filling of trenches in topographical structures. If the filling is not complete, voids will

form inside the trenches, which result in a mechanically weak structure and cause stability issues by trapping processing chemicals or moisture. The aspect ratio, which is defined as the height of a feature to its width, is an important parameter that affects the filling and coverage of a feature.

#### *3.1.7.1 Chemical vapor deposition*

In chemical vapor deposition (CVD) systems, the components of the film are introduced as gases into the deposition chamber where they react and form the film. In a CVD process, first, the reactants are transported to the deposition region; then, reactants will diffuse from the main gas stream through the boundary layer to the wafer surface; the reactants are typically pyrolyzed and absorbed where surface chemical reactions occur; finally, byproducts are desorbed and transported away from the deposition region. The deposition velocity can be expressed as:

$$v = \frac{k_S h_G}{k_S + h_G} \frac{C_T}{N} Y \qquad \text{(Eq. 3. 31)}$$

where $N$ is the number of atoms per unit volume in the film; $h_G$ is the gas-phase mass transfer coefficient (in cm $sec^{-1}$); $k_S$ is the surface chemical reaction rate (in cm $sec^{-1}$); $C_T$ is the concentration of all molecules in the gas phase and $Y$ is the mole fraction of the incorporating species in the gas phase. From (Eq. 3. 31), we can see that the deposition velocity is determined by the smaller of $k_S$ or $h_G$.

$$\text{If} \quad k_S << h_G, \; v \cong \frac{C_T}{N} k_S Y \;, \qquad \text{(Eq. 3. 32)}$$

which is the surface reaction limited case, and

$$\text{If} \quad h_G << k_S, \; v \cong \frac{C_T}{N} h_G Y \;, \qquad \text{(Eq. 3. 33)}$$

which is the mass transfer limited case.

At high temperature, the reaction rate is very high and the reactant species will react at the wafer surface as soon as they arrive. Therefore, mass transfer from the gas phase controls the process, and the configuration of the equipment and placement of the wafers are important to ensure uniform deposition from wafer to wafer. At lower

temperatures where the surface reaction is the limiting factor, the deposition rate is sensitive to temperature. In this regime, the configuration of the system and placement of the wafers do not have much effect on the deposition uniformity from wafer to wafer, and wafers can be stacked.

#### 3.1.7.1.1 Low pressure chemical vapor deposition (LPCVD)

For uniform deposition, the surface reaction limited regime is preferred since the deposition rate is not determined by mass transport, which will cause non-uniformity because of the variability in the boundary layer composition and thickness [73]. Decreasing the total gas pressure will extend the surface reaction limited regime to higher temperatures because the gas-phase diffusivity of the reacting species will increase. In an LPCVD system, wafers are vertically stacked, and the pressure is controlled in the 0.25-2.0 torr range.

The deposition of polysilicon in a LPCVD system is very conformal, with a sticking coefficient as low as 0.001, resulting in good step coverage or trench filling. In the polysilicon deposition process, silane is often used, and the following reaction occurs:

$$SiH_4 \rightarrow Si + 2H_2 \qquad \text{(Eq. 3. 34)}$$

Although silicon chlorides can also be used in the deposition of polysilicon, the silane process has better nucleation on amorphous materials such as silicon oxide, and silane decomposes at lower temperatures than that of the chloride reactions. Nitrogen may be used as a carrier gas instead of hydrogen since hydrogen increases the reverse reaction (etching) in (Eq. 3. 34). The deposition temperature of polysilicon is around 620 ˚C. If the deposition temperature is below 570-600 ˚C, the deposited film will probably be amorphous.

#### 3.1.7.1.2 Plasma-enhanced chemical vapor deposition (PECVD)

When depositing thin films on wafers with exposed metals, lower temperature deposition is required since the melting point of metals is relatively low. LPCVD systems can use lower temperatures, but both deposition rate and film quality are not satisfactory. PECVD systems can provide reasonable deposition rates, good film quality, and a certain degree of conformality at low deposition temperatures. In PECVD systems, a plasma

source is used to provide the energy needed during the chemical reaction process, so a lower temperature, in the range of 200-350 ˚C, can be used during the deposition process. Many PECVD systems have two parallel plate electrodes, with wafers placed on the lower plate electrode. A plasma is formed and sustained between the two electrodes by a high electric field, usually at 13.56 MHz. The gas pressure is usually between 50 mtorr and 5 torr. High energy electrons decompose the reactant gases into a variety of species, including ionized and excited molecules, neutral species and free radicals. Free radicals such as SiO and $SiH_3$ have incomplete bonding and they are very reactive. Ion and electron bombardment of the surface deliver more energy to the reactive species and enhance the surface reactions.

The non-equilibrium nature of the processes in a PECVD system leads to nonstoichiometric composition of the deposited films and incorporation of byproducts. Incorporation of byproducts such as hydrogen, oxygen and nitrogen will result in issues of film quality, such as peeling or cracking of the film during further processing.

Silicon nitride is usually used as a mask layer or for passivation. Because of the poor interface properties, such as interface charges and stress, silicon nitride is usually not used in direct contact with silicon. As a passivation layer near the end of the fabrication flow, silicon nitride should be deposited at temperatures lower than 450 ˚C in a PECVD system. Silane and ammonia are used to form the film following the equation below:

$$3SiH_4 + 4\ NH_3 \rightarrow Si_3N_4 + 12H_2 \qquad \text{(Eq. 3. 35)}$$

The typical deposition temperature is around 300 ˚C. The resulting film is not stoichiometric $Si_3N_4$ but $Si_XN_Y$, or $Si_XN_YH_Z$ with the incorporation of hydrogen.

Near the end of the fabrication flow, a forming gas anneal is often performed to reduce the interface charges at the silicon/silicon-oxide interface. The anneal process in hydrogen should be performed before the nitride film deposition since a nitride film is a barrier to hydrogen diffusion.

### 3.1.8 Metal sputter coating and liftoff

#### *3.1.8.1 Sputter deposition*

Metal sputtering systems have two parallel plate electrodes, similar to the PECVD system. A plasma is formed between the two electrodes. A target which is a plate of metal is placed on the electrode that is negatively biased, and the wafer is placed on the other electrode which is usually grounded. Accelerated positive ions in the plasma strike the target and dislodge the target atoms, which then travel to the surface of the wafer and deposit a metal film. The step coverage can be improved by using large-area targets because of the wide distribution of the arrival angles of the atoms.

The target needs to be conductive if the sputter deposition is operated in DC mode. Deposition of insulating material usually uses AC method to capacitively couple an RF voltage to sustain the plasma [73].

#### *3.1.8.2 Liftoff*

Metal patterning in the micrometer range is sometimes done by first depositing a metal film and then lifting-off the metal film in selected regions. The lift-off technique is used for patterning of deposited films which are not easily dry etched. During this process, first, a photoresist pattern is developed by photolithography. Metal is then blanket deposited over the wafer surface, which covers both the photoresist region and the regions where photoresist has been removed. Next, during the lift-off process, solvent will remove the photoresist together with the film on top of it, leaving the metal films which directly contact the substrate and forming the desired metal pattern. The metal pattern formed in the lift-off process is the opposite of the metal pattern formed by the metal dry etching process.

Deposited films that satisfy the following requirements may be considered for the lift-off process: the deposition temperature does not reach a point where photoresist will be burned; the film quality is not critical since photoresist will outgas slightly in vacuum systems; adhesion of the film to the substrate needs to be good; the film should be easily wetted by the solvent and be thin enough for solvent to seep underneath; the film is not elastic and should be thin enough to tear along adhesion lines [88].

The lift-off process is illustrated in Figure 3.4. First, a layer of photoresist is applied on the wafer surface. Standard exposure with a pre-designed mask pattern then exposes the photoresist in selected regions as shown in Figure 3.4(a). For positive photoresist, development of the wafer in standard developer will dissolve the exposed regions of the photoresist, finishing the pattern transfer of the desired metal pattern from the mask to the wafer as shown in Figure 3.4(b). A thin layer of metal is blanket deposited on the wafer surface covering both the regions with photoresist and regions that are cleared of photoresist as shown in Figure 3.4(c). A relatively long time soak in acetone with a direct stream rinse of the wafer surface with acetone from a squeeze bottle will remove photoresist together with the metal films on top, leaving the desired metal pattern on the wafer as shown in Figure 3.4(d). The substrate should be immersed in acetone until no traces of particles are left on the wafer, since it is quite difficult to remove the particles once they dry.

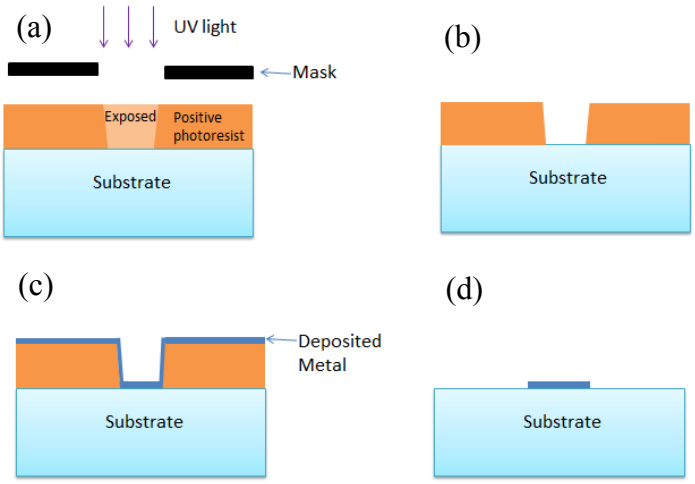



Figure 3.4. Lift-off process with standard photoresist processing: (a) Exposure of the photoresist with mask. (b) Photoresist development. (c) Deposition of metal film. (d) Photoresist liftoff by solvent.

A better way to do the lift-off process is by applying a dual layer photoresist process, with a polymer called LOL2000 (SHIPLEY MICROPOSIT LOL™ 2000). LOL2000 is not sensitive to UV light and can be etched with most of the standard developers. The dual layer process with LOL2000 can achieve submicron lift-off capability. The dual photoresist layer process is illustrated in Figure 3.5. In this process, first, the LOL2000 is spun on the substrate to a thickness of about 1.3 times the thickness

of the metal film to be deposited. Then the LOL2000 goes through a bake process (Figure 3.5(a)). Then a standard photoresist (for example, positive photoresist Shipley 3612 as mentioned before) is spun and baked (Figure 3.5(b)). Next, the photoresist is exposed by the standard exposure process (Figure 3.5(c)). Then standard developer clears the exposed photoresist regions for positive photoresist and also etches away the LOL2000 (Figure 3.5(d)). With precise control of the development time, the etch undercut in LOL2000 is the right size that the subsequent film deposition of metal is not continuous on the sidewall (Figure 3.5(e)). The photoresist is then lifted-off by acetone and a resist removal solvent such as MICROPOSIT 1165 helps remove any photoresist residue left on the wafer surface (Figure 3.5(f)). The undercut size needs to be well controlled so that narrow features are not completely undercut by controlling the development time or the bake procedure of LOL2000. Less development time will cause a smaller undercut provided that the development time is long enough to resolve the standard photoresist on top. Higher bake temperature (around 200 ˚C) will also result in less undercutting.

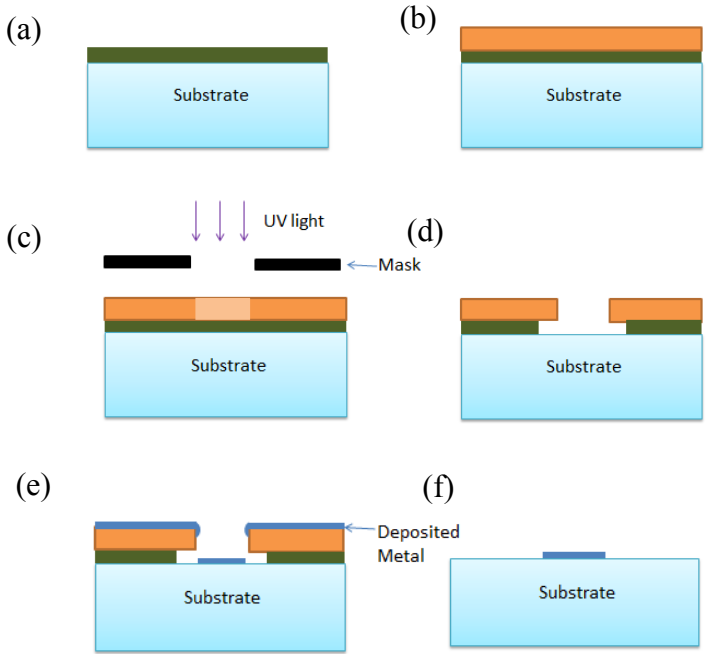



Figure 3.5. Lift-off with the dual photoresist layer process: (a) Spinning of LOL2000 on substrate. (b) Spinning of a standard photoresist. (c) Exposure of the photoresist with mask. (d) Development of the photoresist, resulting in an undercut in the LOL2000 layer. (e) Metal deposition. (f) Photoresist liftoff by solvent.

### 3.1.9 Chemical mechanical polishing

Wafer planarization is needed during processing of substrates with surface topography. Common methods of planarization of a wafer surface are conformal deposition of the dielectric, deposition of the dielectric and etchback, application of Spin-On-Glass (SOG), and Chemical-Mechanical Polishing (CMP) [73].

In the CMP process as illustrated in Figure 3.6, the wafer is held by a vacuum chuck facing down against a spinning, semi-rigid polishing pad. A colloidal silica slurry is sprayed onto the polishing pad. The composition and pH (Potential Hydrogen) of the slurry are adjusted so that the chemical reactions at the wafer surface accelerate the abrasion of the surface material by the silica slurry. The semi-rigid polishing pad adjusts the polishing rate of the nonuniform surface so that elevated features are mostly removed while the lower areas are barely etched, resulting in a planarized surface.

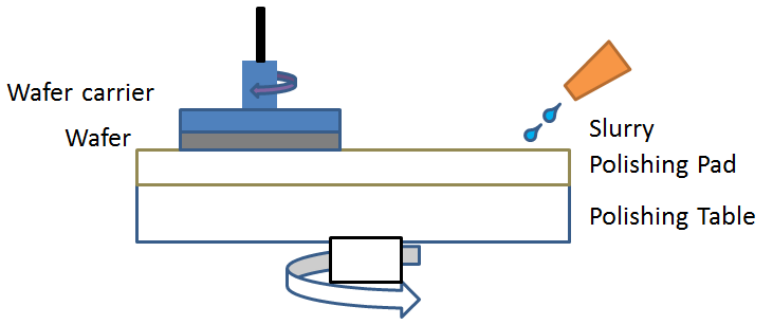



Figure 3.6. Illustration of the CMP process.

### 3.1.10 Spray coating

For relatively planar substrates, photolithography is usually carried out by spin coating of photoresist on the wafer surface to form a thin, uniform layer, which is then baked, exposed and developed to form the desired pattern. The disadvantage of spin coating is that the utilization of photoresist is not efficient and it can only uniformly coat relatively flat surfaces. When spin coating a wafer surface with a degree of topography, such as trenches or mesas, photoresist may not spin on evenly because it tends to draw back and tear from edges and pool at the bottom of trenches. The spray coating technique can be used to conformally coat deep trenches and wells with high aspect ratios [89].

The spray coating system deposits photoresist as an aerosol onto the surface of a slowly rotating wafer from an ultrasonic spray nozzle. The nozzle oscillates to produce microscopic resists with a typical diameter of 20 μm [90]. The chemicals used in the spray coating system need to be diluted with solvents that are compatible with the chemicals' base solvents to a viscosity below 20 centistokes [90]. The droplets produced by the ultrasonic nozzle are propelled toward the surface of the wafer using a stream of nitrogen. The substrate is spun at a low speed-around 50-100 rpm. Most of the parameters, such as the nozzle pressure, the spin speed of the substrate, substrate temperature, and velocity profile of the nozzle as it sweeps across the wafer, can be adjusted by software. The velocity profile of the nozzle needs to be divided into multiple indices to accommodate different coverage area as the nozzle moves over the wafer. This is because near the center of the wafer, the required chemical coverage shrinks and the sweeping speed of the nozzle should be higher. The velocity profile of the nozzle can be optimized by calculating the relative area of each index. The parameters that affect the thickness, uniformity and roughness of the photoresist include: solid content of the resist, angle of the spray nozzle, resist dispense rate, scanning speed of the nozzle, spinning speed and temperature of the substrate.

### 3.1.11 Scanning electron microscopy

Scanning electron microscopy (SEM) is frequently used to characterize the structure and material properties during the fabrication process by scanning a sample surface with a beam of electrons. First, an electron beam is thermionically emitted from an electron gun with an energy of 0.2 keV to 40 keV. The electron beam is then focused with electromagnetic lenses onto the sample, ejecting secondary electrons and X-rays. In the most common imaging mode, the detector collects secondary electrons which are ejected from the atoms of the sample and shows the three-dimensional appearance of the sample on an analogue or digital display.

In our fabrication process, the SEM is usually used to image the cross section of the wafer and characterize the deep trench etching and filling processes.

## 3.2 Process simulation

Most of the models discussed in section 3.1 are incorporated into the simulation program TSUPREM4. TSUPREM4 is a version of SUPREM IV, which is a computer program to simulate various processing steps in the fabrication of silicon integrated circuits, such as inert-ambient dopant drive-in, thermal oxidation of silicon, ion implantation and etching of various materials. The change of semiconductor structures by various manufacturing processing steps, such as the distribution of impurities and various layer thicknesses, can be simulated.

The simulation of our process flow uses TSUPREM4 in the one-dimensional mode with the structural grid in the depth direction of the device. The structure is first initialized on 5 μm thick <100> silicon; the grid size is set to be 0.01 μm; the substrate boron concentration is set to be $1\times10^{15}$ $cm^{-3}$. Simulation in TSUPREM4 provides information such as the various silicon oxide thickness and dopant distribution profiles, which are important factors in device performance. By simulating various process conditions, parameters such as oxidation time and temperature, annealing time and sequence, and the dose of dopant ion implantation can be optimized. Three cross-sections in the final structure are simulated as shown in Figure 3.7.

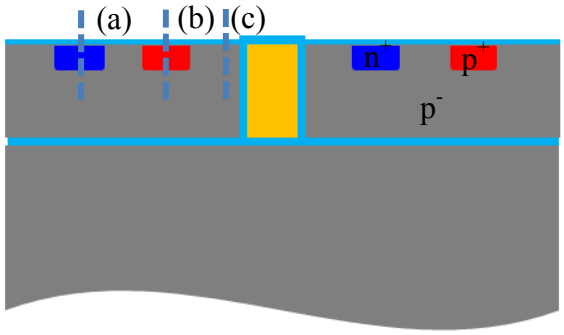



Figure 3.7. Cross-section of the structure being simulated: (a) cross-section of the $n^+$ region. (b) cross-section of the $p^+$ region. (c) cross-section of the lightly-doped region.

As in Figure 3.7, the process flow starts with a lightly boron doped substrate marked as $p^-$. Region (a) is the heavily doped $n^+$ region formed by a phosphorus predeposition process. Region (b) is the heavily doped $p^+$ region formed by a boron ion implantation process. After forming the $n^+$ and $p^+$ regions, the oxide on the surface of the wafer is etched. A thermal oxidation and annealing process grows a thin layer of thermal

oxide on top of the wafer surface and drives in the dopant to the final depth. In the following subsections, the doping profiles of the $n^+$, $p^+$ and lightly doped regions are simulated. The $n^+$ region is the most important region in the simulation since the pn junction thus formed is where light-generated carriers are collected. In addition, the sheet resistances of the $n^+$ and $p^+$ regions need to be high enough to make an Ohmic contact to the following metal layer. The oxide thicknesses on various regions are also simulated for the design of the antireflection coating.

### 3.2.1 Simulation of the $n^+$ region

The simulation of the heavily doped n regions starts with a lightly boron-doped substrate with a substrate doping of $1\times10^{15}$ $cm^{-3}$. In the simulated process flow, first, a 30 minute, 1000 ˚C phosphorus predeposition is performed; then a 20 minute wet oxidation is performed to block the boron implantation; next, a dry oxidation at 950 ˚C for 15 minutes is performed on the $n^+$ regions during the growth of the screen oxide on the $p^+$ regions; finally, all of the oxide on top of the $n^+$ regions is etched, and an 83 minute dry oxidation at 1000 ˚C followed by a 75 minute anneal is performed. The simulated doping profile of the heavily doped n region is shown in Figure 3.8.

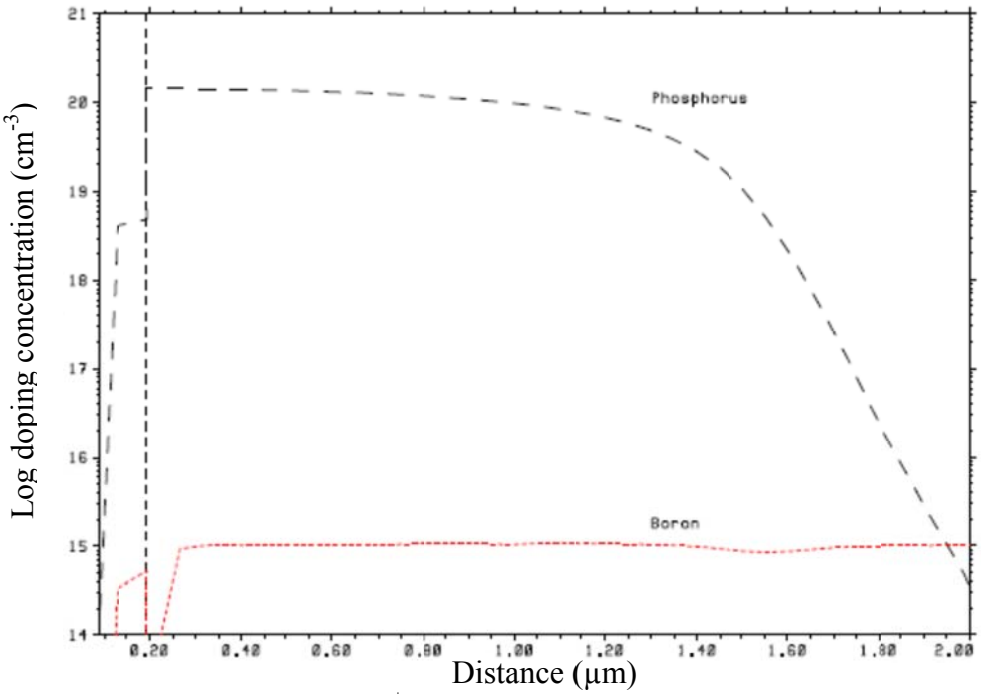



Figure 3.8. Doping profile of the $n^+$ region (Reprinted from L.Wang, K. Mathieson, T. I. Kamins, J. Loudin, L. Galambos, J. S. Harris and D. Palanker, Proc. SPIE 8248, 824805 (2012); http://dx.doi.org/10.1117/12.909104 ©SPIE).

From the simulation results, the oxide thickness on top of the $n^+$ region is 104 nm; the sheet resistance of the $n^+$ region is 7.2 Ω/sq; the junction depth is 1.76 μm. From the simulation profile, we can see that the doping concentration of phosphorus at the surface is above $1\times10^{20}$ $cm^{-3}$, which is high enough to make a low-resistance Ohmic contact to the following metal layer. The junction depth of 1.76 μm is not critical since for our application at a wavelength near 900 nm, the absorption coefficient in silicon is relatively small corresponding to an absorption length of about 30 μm.

### 3.2.2 Simulation of the $p^+$ region

The simulation of the heavily doped p regions starts with a lightly boron doped substrate with a substrate doping of $1\times10^{15}$ $cm^{-3}$. First, a dry oxidation at 950 ˚C for 15 minutes is performed during the growth of the screen oxide on the $p^+$ regions; next, a boron ion implantation with a dose of $1.3\times10^{15}$ $cm^{-2}$ is performed; then, the oxide on top is etched; finally, an 83 minute dry oxidation at 1000 ˚C is performed followed by a 75 minute anneal. The simulated doping profile of the heavily doped p region is shown in Figure 3.9.

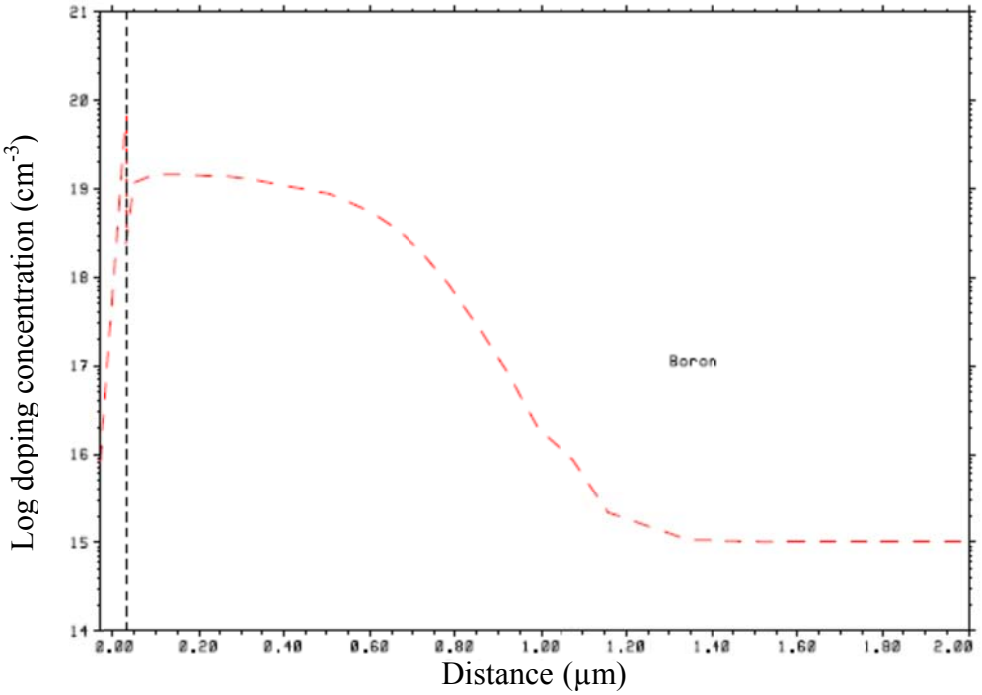



Figure 3.9. Doping profile of the $p^+$ region (Reprinted from L.Wang, K. Mathieson, T. I. Kamins, J. Loudin, L. Galambos, J. S. Harris and D. Palanker, Proc. SPIE 8248, 824805 (2012); http://dx.doi.org/10.1117/12.909104 ©SPIE).

From the simulation results, the oxide thickness on top of the $p^+$ region is 60.3 nm, and the sheet resistance of the $p^+$ region is 115 Ω/sq. From Figure 3.9, we can see that the surface doping concentration of boron is about $1\times10^{19}$ $cm^{-3}$, which will make a low resistance Ohmic contact to the following metal layer. The $p^+$ region is used to make a good contact between the lightly doped silicon and the metal layer; there is no pn junction formed in this region.

### 3.2.3 Simulation of the lightly doped region

The simulation of the lightly doped p regions starts with a lightly boron doped substrate with a substrate doping of $1\times10^{15}$ $cm^{-3}$. First, a wet oxidation process at 1000 ˚C for 100 minutes is performed to grow 500 nm thermal oxide; then a wet oxidation process at 1000 ˚C for 20 minutes is performed during the oxidation process for $n^+$ regions; next, a dry oxidation at 950 ˚C for 15 minutes is performed on the lightly doped regions during the growth of the screen oxide on the $p^+$ regions; finally, the oxide on top is etched, and an 83 minute dry oxidation at 1000 ˚C is performed followed by a 75 minute anneal. The simulated doping profile of the lightly doped n region is show in Figure 3.10.

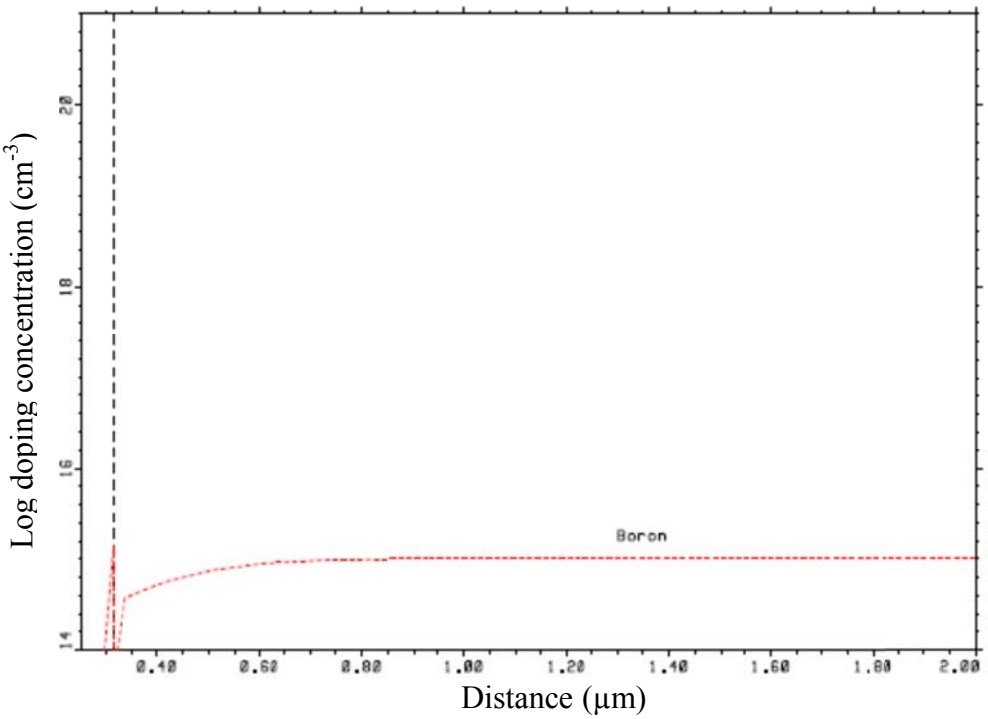



Figure 3.10. Doping profile of the lightly doped region.

From the simulation results, the oxide thickness on top of lightly doped region is 61 nm; the sheet resistance of the lightly doped region is 636 Ω/sq.

| Parameters | $n^+$ | $p^+$ | lightly doped region |
|---|---|---|---|
| Sheet resistance | 7.2 Ω/sq | 115 Ω/sq | 636 Ω/sq |
| Oxide thickness | 104 nm | 60.3 nm | 61 nm |

Table 3.1. Simulated parameters including sheet resistances and oxide thicknesses for the $n^+$, $p^+$ and lightly doped regions.

From the simulated results for the $n^+$, $p^+$ and lightly boron doped regions, which are summarized in Table 3.1, we can see that the oxide thickness is much larger for the $n^+$ regions (104 nm) than for the $p^+$ and lightly doped regions (61 nm). This is because for highly doped regions, the total number of vacancies is much higher than that for lightly doped material. The vacancies are able to provide sites for the oxidation reaction. For $p^+$ regions, since the total vacancy number is less than that for $n^+$ regions and dopants tend to deplete into silicon oxide, dopant effects on the oxidation rate are not as significant as those for the $n^+$ regions, as discussed in section 3.1.3. The issue with different oxide thicknesses in different regions is that as we design the anti-reflection coating on wafer surface, the optimal point cannot be achieved for the whole surface. Besides the anti-reflection coating, oxide is also used to passivate the surface and provide bio-compatibility in biological systems. Therefore, the oxide thickness needs to be above a certain value, so the final cycle oxidation time is chosen to be 83 minutes. However, the thicker oxide on the $n^+$ region degrades the antireflection coating and increases the reflection of light on the $n^+$ surface due to the deviation of the grown oxide layer thickness (104 nm) from the target value (60 nm).

## 3.3 Fabrication flow

The fabrication process uses eight mask layers, involving most of the tools discussed in section 3.1. This process has integrated most of the standard CMOS fabrication processes, including the standard cleaning procedure, photolithography, oxidation, dopant pre-deposition and ion implantation, plasma dry etching, chemical wet etching, low pressure chemical vapor deposition, plasma-enhanced chemical vapor

deposition, metal sputter coating and liftoff, and chemical mechanical polishing, etc. Some fabrication processes that are usually involved in the fabrication of Micro-Electro-Mechanical Systems (MEMS), such as deep reactive-ion etching (DRIE) and spray coating of photoresist are also incorporated into the fabrication process to optimize the device performance for bio-medical applications. For a detailed description of the fabrication flow, please refer to the appendix.

### 3.3.1 Brief discussion of the fabrication batches

We initiated four fabrication batches during the progress of this research. The first batch was fabricated on p-type SOI wafers with 30 μm thick device layer and a buried oxide layer of 100 nm, using a seven layer mask process consisting of $n^+$ region, $p^+$ region, isolation trench etch, first via etch, first metal deposition and liftoff, second via etch and second metal deposition and liftoff. The three photodiodes in each pixel are connected by 1 μm silicon bridges. The issue with this batch was that the silicon bridges were not thin enough to be fully depleted and to isolate the three photodiodes in each pixel. Therefore, the current-voltage characteristics for the three in-series connected photodiodes are similar to those of a single photodiode, due to additional current paths introduced by the silicon bridges. Another issue is that the reverse bias dark current was too high (on the order of sub-μA) and the reverse breakdown voltage was very low (about 1.5 V). The reason for the higher dark current is probably from the design of the mask: the $n^+$ regions and $p^+$ regions extend to the edges of the isolation trench etch, resulting in higher densities of recombination and generation centers in the depletion regions. Due to the inadequate isolation between the three photodiodes in each pixel and the high dark current, we developed a new mask design to keep the $n^+$ and $p^+$ regions far away from the edge of the isolation trench etch (at least 2 μm), and used polysilicon to fill in the trenches to avoid using the bridges as connections; this design was used for the second batch of devices. The second batch was fabricated on p-type SOI wafers with 30 μm thick device layer and a buried oxide layer of 100 nm. The fabrication process of the second batch of devices consisted of eight layers, including the isolation trench etch, $n^+$ regions, $p^+$ regions, first via, first metal deposition and liftoff, second via, polysilicon etch in selected regions, and second metal deposition and liftoff. The resulting devices from the

second batch have good current-voltage characteristics, which will be discussed in Chapter 4. To reduce the simulation threshold on the retina, a new generation of devices with reversed $n^+$ and $p^+$ polarity was designed for the third batch of devices. The third batch of wafers was fabricated on n-type SOI wafers with 30 μm thick device layer and a buried oxide layer of 1 μm. However, due to the out-of-range deposition thickness during the LPCVD process and the thinner handling wafer thickness (400 μm instead of 500 μm), most of the wafers were broken in the LPCVD furnace so that a fourth batch is under fabrication. The arrays in the fourth batch are round, instead of rectangular as in the previous batches. The round shape has advantages from the surgical point of view. The fourth batch of wafers is being fabricated on n-type SOI wafers with 30 μm thick device layer and a buried oxide layer of 500 nm. Different buried oxide thicknesses were chosen to facilitate the fabrication process and provide good passivation to the arrays. The fabrication process of the second batch will be discussed in detail in the following sections.

### 3.3.2 Mask design

The mask pattern was designed using Tanner L-edit. In our layout, each pixel has a central active electrode and a local return electrode. The local return electrode in each pixel helps to reduce the spatial spreading of the electrical stimulation, which is likely to reduce the cross-talk between pixels and improve the spatial resolution of the implant. Iridium oxide electrodes are used to provide a large charge capacitance. In the photodiode array, each pixel contains three photodiodes connected in series, with trench-isolation between them, to increase the charge injection levels. Ti/Pt metal lines are used to connect the three photodiodes in series in each pixel. Between pixels, trench channels are opened to allow perfusion of nutrient during in-vitro electrophysiological experiments. Dopants are introduced to form the photosensitive regions and make Ohmic electrical contacts. We fabricated three different sizes of pixels with edge to edge distance (excluding trenches) of 280 μm, 140 μm and 70 μm to explore the maximum resolution of the subretinal stimulation, and two sizes of the photodiode arrays (1.9 mm × 1.9 mm and 1.2 mm× 0.8 mm) to provide retina coverage in both large animals and small animals. The overall mask pattern is shown in Figure 3.11.

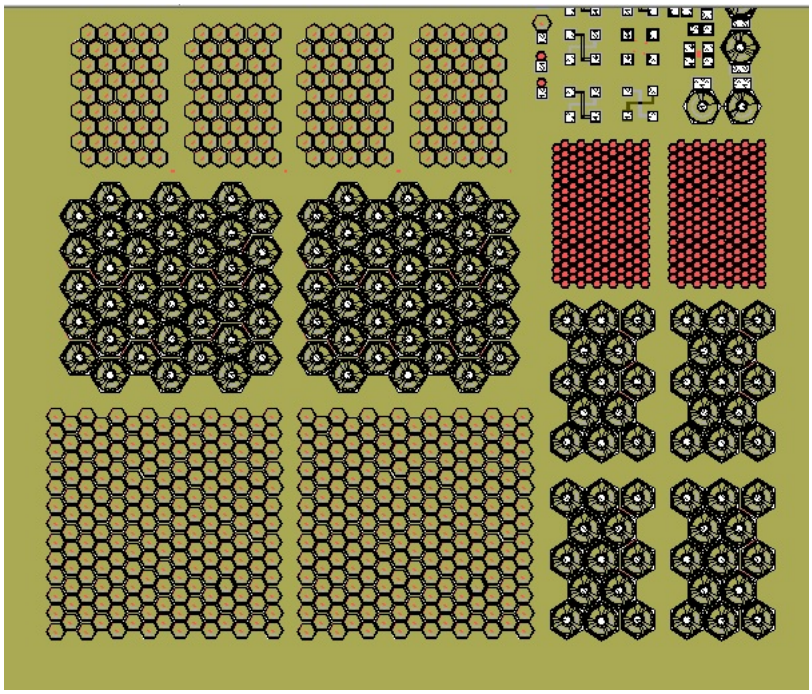


Figure 3.11. Overall mask pattern including three pixel sizes (280 μm, 140 μm and 70 μm) and two array sizes (1.9 mm × 1.9 mm, 1.2 mm× 0.8 mm) in L-edit.

Three pixel sizes 280 μm, 140 μm and 70 μm were designed and the generated current scales with the exposed area of silicon. Large-size pixels will generate more current, but the pixel density (about 13 $mm^{-2}$) is very low, resulting in poor resolution to detect a light pattern. Medium-size pixels will generate moderate current with a pixel density of 45 $mm^{-2}$, which will provide better resolution. Small-size pixels will generate the least current, but the pixel density is very high (174 $mm^{-2}$), resulting the maximum possible resolution of the three sizes of pixels. A test structure region at the top right of the mask area was designed to measure sheet resistances, contact resistances and single pixel characteristics, etc.

As an illustration of the pixel design process, the layer-by-layer design process for the medium-size pixel is illustrated in Figure 3.12. First, an active region is drawn which defines the silicon region after the DRIE trench etch. The white areas in Figure 3.12(a) are the exposed regions to be etched by the DRIE process, which isolates the three photodiodes in each pixel and opens regions between pixels for nutrient flow in biological systems. Pixels are connected by 16 μm silicon bridges in the large-size and medium-size pixels and 10 μm silicon bridges in the small-size pixels to mechanically

and electrically connect all of the return electrodes together. Figure 3.12(b) shows the design of the $n^+$ regions on the three regions inside each pixel. The $n^+$ regions occupy a relatively large area of the active regions since they will form pn junctions on the lightly boron doped substrate, where light generated carriers are collected under light illumination. When designing this region, the edge of the heavily doped n region needs to be 2 μm away from the edge of the trench, to make sure the depletion region is far away from the etched sidewall of silicon, which may have recombination and generation centers. The areas of the three $n^+$ regions in each pixel of silicon should be approximately equal so the generated photocurrent will be close to equal in each diode.

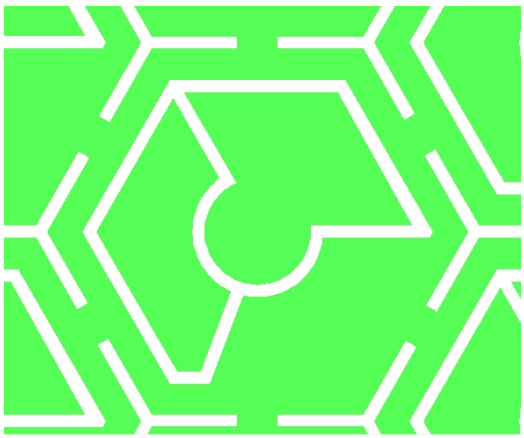


Figure 3.12.(a) active region.

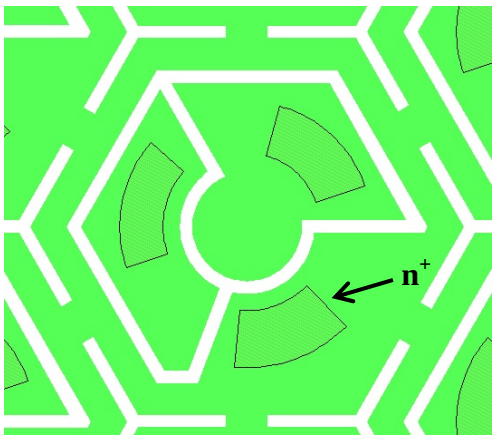



Figure 3.12.(b) $n^+$ region formed on (a) .

Figure 3.12(c) shows the added $p^+$ regions on the pixel design. It is also better to make sure the edge of the $p^+$ regions is far away from the edge of the silicon trench, although the $p^+$ regions are mainly to make low resistance contacts to the following metal layer. The area of the three $p^+$ regions should also be similar so that the electrical characteristics of the three photodiodes are similar. The distance between a $n^+$ region and a $p^+$ region should be 3-4 μm so that lateral diffusion of dopants after long-time high-temperature processes does not cause the two regions overlap. Figure 3.12(d) shows the first via etch. These via regions are opened for the following metal layer to contact the highly doped silicon layer so that current can be conducted between photodiodes and out to the stimulating metal electrodes. The sizes of the via regions should be larger than 3 $\mu m^2$ so that contact resistance will not become a problem. The via regions need to be well inside the heavily doped regions, and the edge of the via etch should be $\geq$ 1 μm inside the

edge of the heavily doped regions so that if there is accidental misalignment, the following metal layer will not accidently touch any other regions, causing shorts. In higher current regions, such as the $n^+$ regions on the current design, two via regions are designed in the medium-size and large-size pixels so that current can be directed out efficiently without causing too much series resistance in the current pathway. The smaller $p^+$ regions have one via region in the center.

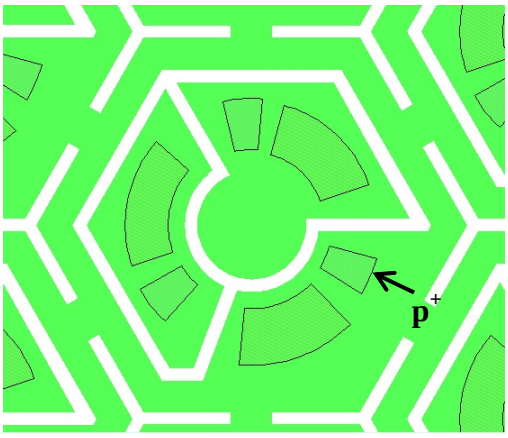



Figure 3.12. (c) $p^+$ region formed on (b).

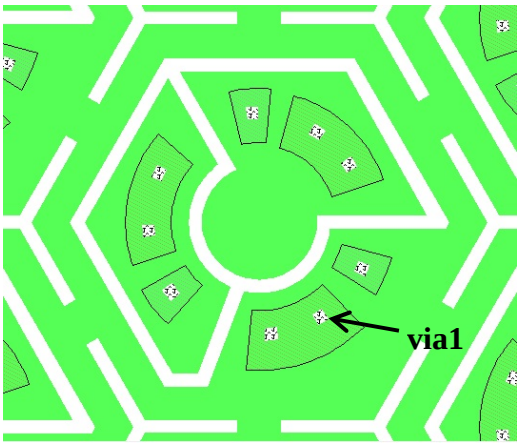
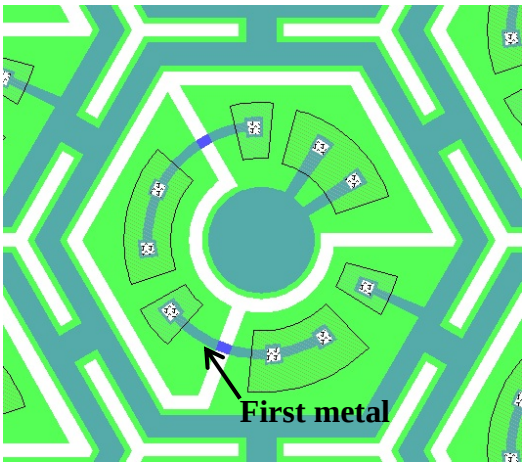



Figure 3.12. (d) First via formed on (c).

Figure 3.12(e) shows the design of the first metal layer, which connects the three photodiodes in each pixel in series with the central active electrode and local surrounding return electrode. The area of each metal pad needs to be large enough to cover the corresponding via region. This is because the metal liftoff process for this layer tends to tear the metal at the pattern edges. With the edge of the metal about 1 μm outside the edge of the via region, the via region will be well sealed during the liftoff process. In the current design, the edge of the first metal layer is 0.5 μm inside the edge of silicon to avoid any possible shorts during the metal liftoff process. Figure 3.12(f) shows the design of the second via regions. The second via regions open contacts in the center active and surrounding return electrode regions to connect the first metal (Ti/Pt) layer to the second metal (iridium oxide) layer. The edges of the second via layers are designed to be inside the edges of the metal layers by 0.5-1 μm to avoid any sealing problem during the metal liftoff process. Figure 3.12(g) is the design of the polysilicon etch mask which removes the polysilicon in selected regions between pixels. The close-up of the polysilicon etch

mask is shown in Figure 3.12(h), where the edge of the polysilicon etch mask extends 0.5 μm from the edge of silicon into the trench. This is needed because, in the polysilicon etch process, the tops and the sidewalls of the trenches are all protected by silicon oxide and the top oxide is opened by the polysilicon etch mask. This mask expands into the trench so that the oxide etch process does not expose any silicon regions that can be etched later by the polysilicon etch process.



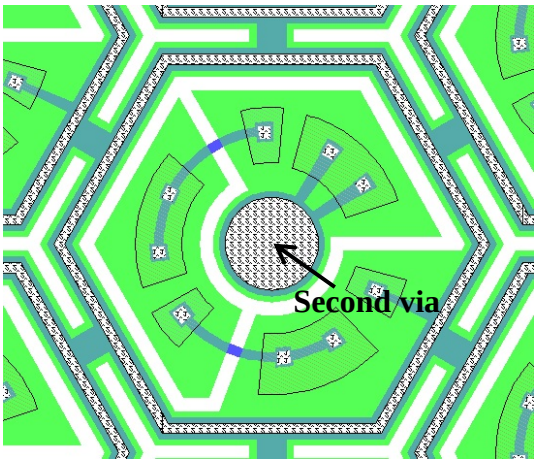
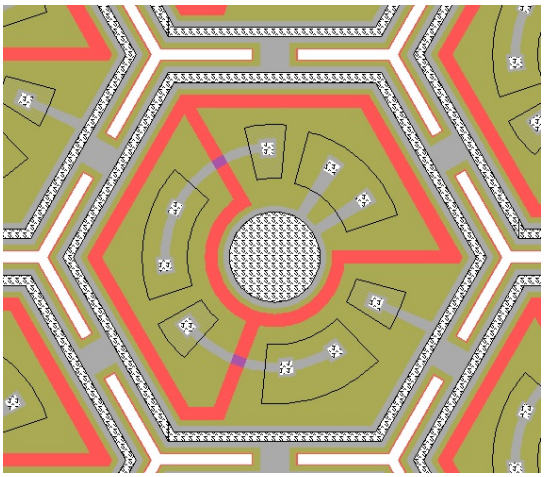




Figure 3.12. (e) First metal formed on (d). Figure 3.12. (f) Second via formed on (e).



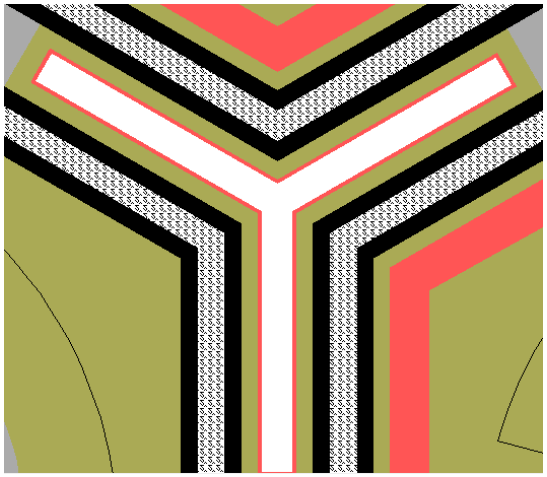


Figure 3.12. (g) Polysilicon removal (red regions). Figure 3.12. (h) Close-up of (g).

Figure 3.12(i) shows the second metal layer added to the pixel. The second metal layer overlaps the first metal layer in the central active and surrounding return electrode regions. Charges generated by the photodiode can then flow from the first metal through the path in the second via regions to the second metal regions. The second metal which is

iridium oxide has a large charge capacity and provides a stable neural interface in biological systems. By capacitive coupling, charges can then be conducted from the second metal layer to the surrounding tissues. Figure 3.12(j) shows an array of medium-size pixels; the array has an area of 1.2mm × 0.8mm with rounded corners to facilitate surgery.

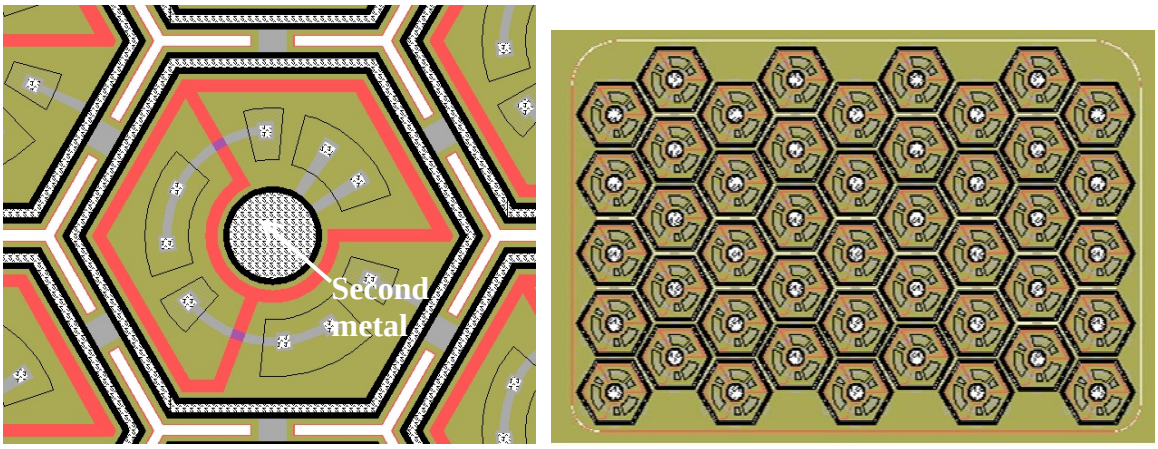
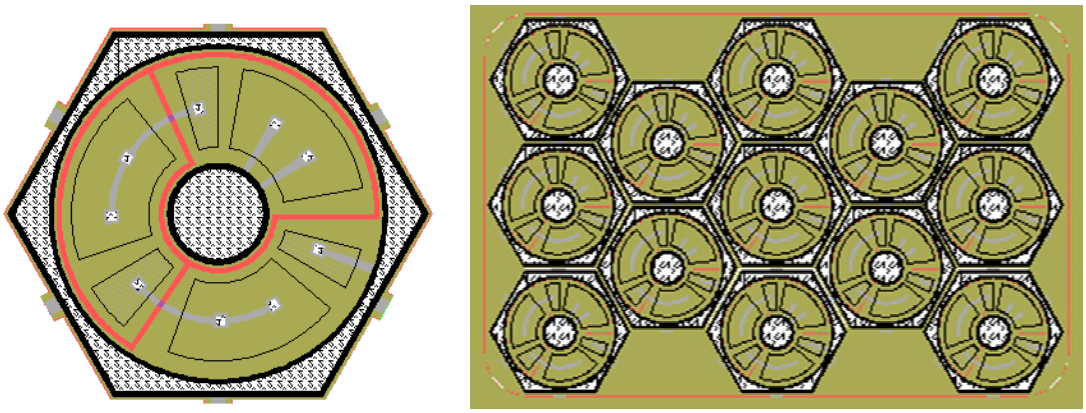



Figure 3.12. (i) Second metal (iridium oxide). Figure 3.12. (j) Mask of medium-size pixel array.

Figure 3.12. Layer by layer design process of the medium-size pixel.

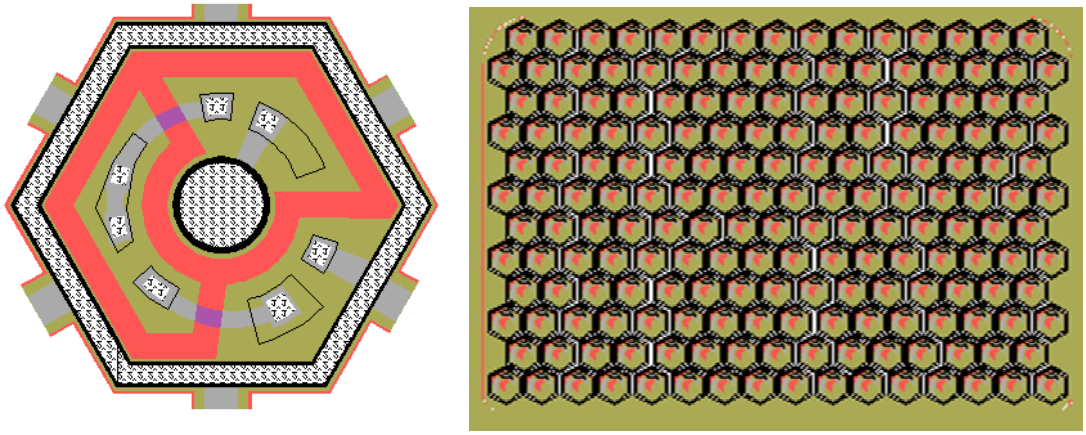
The designs of the large-size pixel and small-size pixel are shown in Figure 3.13 and Figure 3.14. The general device structure is similar for all of the three sizes of pixels. The central electrode diameter is 80 μm for large-size pixels, 40 μm for medium-size pixels and 20 μm for small-size pixels. The trench width does not scale with the three sizes of pixels, and remains 5 μm in width for all three sizes of pixels. Having the same trench width across the wafer simplifies some of the processing steps, such as the DRIE etch process, the polysilicon refill process and the photoresist spray coating process.

Masks in our fabrication process are made by COMPUGRAPHICS USA INC. The final mask set consists of two quartz plates with four images on each plate. Each image represents the information of one layer and corresponds to a die size of 6.5 mm × 6.5 mm on the wafer. Patterns on the first plate are $n^+$, $p^+$, first via, and first metal. Patterns on the second plate are active region, polysilicon etch, second via and second metal.



(a) (b)

Figure 3.13. Mask pattern of large-size-pixel array: (a) close-up of a large-size pixel. (b)1.2 mm×0.8 mm array with large-size pixels.



(a) (b)

Figure 3.14. Mask pattern of small-size-pixel array: (a) close-up of a small-size pixel. (b)1.2 mm×0.8 mm array with small-size pixels.

### 3.3.3 Fabrication process

The fabrication process starts with cleaning a batch of silicon on insulator (SOI) wafers and etching the alignment marks. The main fabrication modules for the second batch of devices include: deep trench etching and isolation; pn junction formation; metal deposition and liftoff; and device release. This section is focused mainly on the fabrication of the second batch of devices; issues during fabrication of the other batches are also discussed.

#### *3.3.3.1 Wafer cleaning and oxidation*

Firstly, wafers are marked using a diamond pen with wafer names reflecting the batch number and the name of individual wafers to record the fabrication processes for each wafer. Several bulk test wafers are also included.

Then the wafers are cleaned with a standard procedure at 'Wet-bench Nonmetal'. First, wafers are cleaned by a solution consisting of 90% sulfuric acid and 10% hydrogen peroxide ("Piranha") at 120 ˚C for 20 minutes. Then the wafers are cleaned by a dump rinse in DI wafer for 5 minutes. Finally, the wafers are dried by a spin/rinse dryer (280 seconds rinse; 120 seconds spin dry).

Then the wafers are cleaned in the pre-diffusion wet bench before the oxidation process. First, a solution of 4:1 $H_2SO_4$:$H_2O_2$ at 90 ˚C for 10 minutes removes the organic contaminants from the wafers. Next, a room temperature HF (50:1) dip for 30 seconds removes the chemical oxide put on during the first step. Next, a 5:1:1 $H_2O$: HCl: $H_2O_2$ solution at 70 ˚C for 10 minutes removes any metal contaminants on the wafer surface. This is most effective after the HF dip step which removes the chemical oxide on top of the wafer. An extra HF dip can be done to remove the chemical oxide which is grown on the wafer surface during the previous cleaning step. A dump rinse in DI wafer for 5 minutes should be inserted between each chemical cleaning step. The wafers are spin dried before loading into the furnaces. Oxidation is performed by one of the Tylan tools in SNF. An oxide layer with a thickness of about 500 nm is grown by a wet oxidation process at 1000 ˚C. This layer of oxide will be used as a hard mask for the trench etch process.

#### *3.3.3.2 Job file and alignment mark*

A job file for each fabrication process defines the image information, layer information and exposure information for the photolithography. The exposure tool we used is the ASML PAS 5500/60 model (ASM Lithography), which is a 5:1 reducing stepper system with a maximum die size limit of 18 mm × 22.6 mm. The parameters that need to be defined are discussed as follows. First, we need to specify the die size, which is the step and repeat size on our wafers. In our fabrication process, the die size is the same as the size of each layer, which is 6.5 mm × 6.5 mm. About 100 die can be placed

on each 4 inch wafer. Next, the alignment strategy needs to be defined, including the position and type of the alignment mark. We use 'XPA' marks distributed near the edge of the wafer. Then, images on the reticles are defined by inputting the position and size of each of the eight images. Next, nine layers are defined including the zero alignment layer, isolation trench etch layer, $n^+$ layer, $p^+$ layer, first via layer, first metal layer, second via layer, polysilicon etch layer, second metal layer and the mark clear-out layer. Finally, layers need to be linked with images by specifying which images should be exposed under a certain layer name.

Alignment is important if there is more than one layer of photolithography in the fabrication process. In the fabrication process, after the initial wafer cleaning process, alignment marks are placed on the wafer, which is defined as the zero layer. All of the following layers such as the deep trench etch, $n^+$ opening, $p^+$ opening, first via opening, first metal deposition and liftoff, second via opening, polysilicon etch in selected regions, second metal deposition and liftoff layers will then align to the zero alignment mark layer.

The ASML aligner uses an i-line light source (365 nm) with an alignment accuracy of 60 nm and a typical resolution to 0.45 μm. The reticle, which is a quartz plate with a single layer of chip layout at 5 times the actual size, has a size of 5 inch, which is usually made of quartz with 0.090 inch thickness. The ASML alignment marks are diffraction gratings in both x and y directions. A HeNe laser at a wavelength of 632.8 nm illuminates the alignment mark and a diffraction pattern of bright and dark lines is reflected to a sensor for alignment of the wafer [91]. The wafer marks are etched into the starting wafer as lines and spaces. The etch depth should be $\lambda/4n$, which is determined by the maximum contrast of the diffracted pattern, where $\lambda$ is the wavelength of the laser light (632.8 nm) and $n$ is the refractive index of the material above the marks (usually photoresist or oxide). Therefore, the etch depth can be calculated as follows:

$$\text{Etch depth} = \frac{\lambda}{4n} = \frac{632.8nm}{4 \times 1.45} = 110nm \qquad \text{(Eq. 3. 36)}$$

Alignment marks are first patterned by photolithography which is pre-defined by the job file. A standard photolithography process is performed to pattern the alignment mark layer. First, wafers are primed in the YES oven in SNF, which applies HMDS on

the wafer surfaces. Next, a positive photoresist of 1μm thick 3612 is spin coated on the wafer surface and baked at 90˚C for 60 seconds on the SVG coater. Next, the photoresist exposure is done at the ASML aligner with the standard alignment mark reticle (ID: 45023981A009) at a dose of 80 mJ/cm$^2$. Then, post exposure bake at 110˚C for 60 seconds and development at the SVG developer equipment using MICROPOSIT MF-26A with a standard develop program are performed.

After the lithography process, a plasma DESCUM process with oxygen plasma is used to remove photoresist residue on the exposed areas. Because our wafers have a $SiO_2$ layer of 500 nm on top of the silicon substrate, the etching of the alignment marks needs to first etch through the $SiO_2$ layer, and then etch into silicon. Plasma dry etching processes in the AMT etcher (Applied Materials Technologies 8100 Hexode plasma etcher) first etches through silicon oxide in the exposed regions using a standard silicon oxide etch program, and then etches into the silicon substrate using a standard via-etch program to a depth of about 110-120 nm. The etch conditions are summarized in Table 3.2.

| AMT etcher | Std Oxide Etch program | Via Etch program |
|---|---|---|
| Gas1-$O_2$ (sccm) | 6 | 30 |
| Gas2-$CHF_3$(sccm) | 85 | 50 |
| Pressure (mTorr) | 40 | 40 |
| DC bias (-V) | -530 | -530 |
| Max RF (W) | 1600 | 1600 |

Table 3.2. Etching recipe for standard oxide etch and silicon via etch in AMT etcher (Adapted from SNF website: https://snf.stanford.edu/SNF/).

After etching the alignment marks, the photoresist is removed with an oxygen plasma which burns the photoresist off the wafer. Then wafers are further cleaned by the "Piranha".

#### *3.3.3.3 Deep trench isolation*

After etching the alignment marks, the wafers are ready for the first main module of the fabrication process: deep trench etching and isolation. First, standard lithography

(as discussed in section 3.3.3.2) is used to define the photoresist pattern of the trench etching process. Then the 500 nm oxide is etched from the top of the isolation regions by the standard oxide etching process as discussed in section 3.3.3.2. A schematic diagram of the cross section of the wafer is shown in Figure 3.15.


Figure 3.15. Schematic diagram of the cross section of the trench region after opening $SiO_2$ windows.

A deep reactive-ion etch (DRIE) process is used to etch the 5 μm wide, 30 μm deep trenches in our fabrication process. The technology we used is widely known today as the BOSCH process, invented by Lärmer and Schilp [92]. A more refined process, called the Advanced Silicon Etch (ASE) process, is based on the BOSCH process, which alternates between the passivating $C_4F_8$ plasma and the silicon etching $SF_6$ plasma. The $SF_6$ plasma supplies fluorine radicals that etch the exposed silicon spontaneously. The $C_4F_8$ deposits a $(C_xF_y)_n$ polymeric layer to protect exposed surfaces [93]. The directional energetic ions remove the passivation from the bottom of the trench and expose the silicon for spontaneous etching. The etching components and passivation components need to be balanced so that etching of trenches with vertical and smooth sidewalls can be achieved. Figure 3.16 shows the schematic diagram of the cross section of the trench region after the DRIE process.


Figure 3.16. Schematic diagram of the cross section of the trench region after the DRIE process.

Figure 3.17 shows the cross-section SEM image of the trench profile after the DRIE process. The etch rate of the DRIE process is related to the specific feature size, so we designed all of the trenches to be 5 µm wide to have better control of the etch process. The etch rate is 2-3 µm/min in the Surface Technology Systems (STS) multiplex inductively coupled plasma system in SNF. The etch stops on the buried oxide layer.

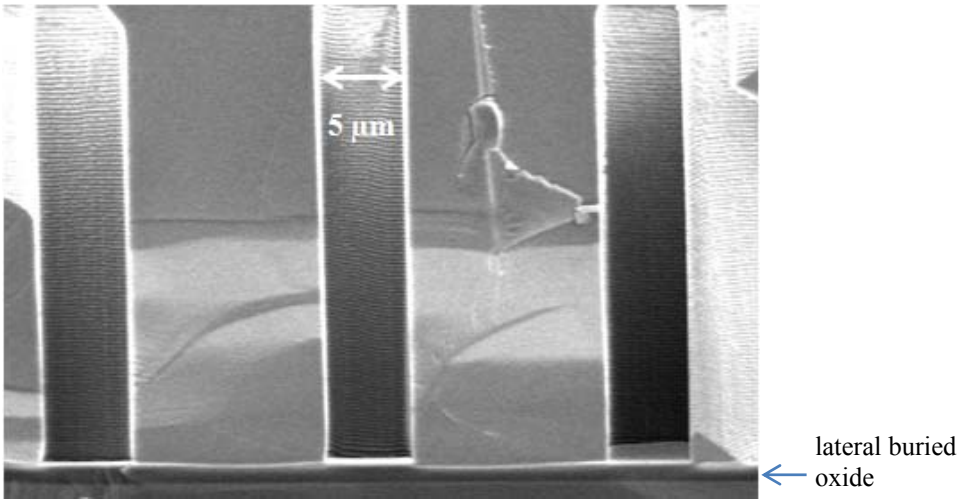



Figure 3.17. Cross section SEM image of the trench profile after the DRIE process.

Figure 3.18 shows a close-up of the cross-section SEM image of the trench profile after the DRIE process. We can see that the etching stops at the bottom buried oxide layer. The etching also creates scallops with a period of hundreds of nanometers corresponding to one etch/coat cycle on the sidewall of the trench. The sidewall smoothness is not a critical parameter for our application, so this degree of sidewall smoothness is acceptable.

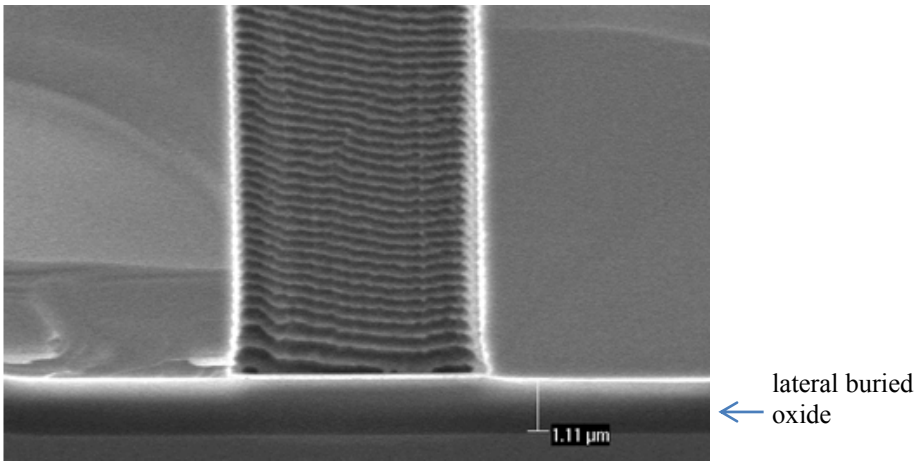



Figure 3.18. Close-up of the cross-section SEM image of the trench profile after the DRIE process on an SOI wafer with about 1 µm thick buried oxide.

After the DRIE process, the photoresist on the wafers is first removed by photoresist burning in a plasma and then a "Piranha" cleaning step. Then the wafers go through a further pre-diffusion cleaning process as discussed in section 3.3.3.1. A wet oxidation process grows 500 nm of thermal oxide on the wafer surface to passivate the sidewalls and provide isolation between different electrically active regions, as will be discussed later. After the oxidation process, the wafers are directly loaded into the LPCVD furnace, and a 4 μm polysilicon layer is deposited by the LPCVD process, which is discussed in detail in section 3.1.7.1.1. Deposition of 4 μm polysilicon seems to be adequate to fill in the 5 μm wide trenches. The LPCVD process is very conformal, which deposits polysilicon that fills the trenches and planarizes the wafer surfaces, as shown in the schematic diagram in Figure 3.19.

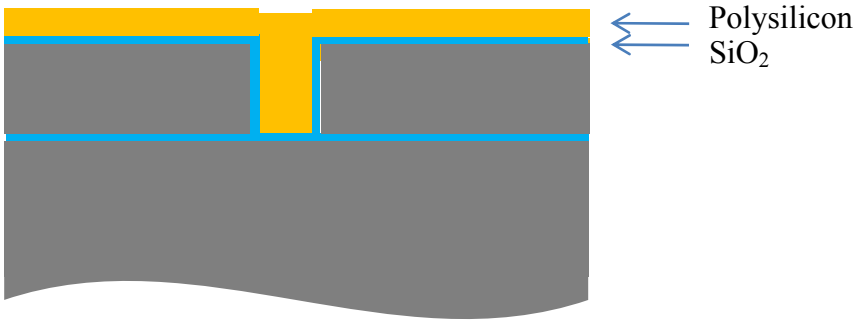



Figure 3.19. Schematic diagram of the cross section of the trench region after the LPCVD process.

The cross-section SEM image in Figure 3.20 shows the conformal deposition of the LPCVD polysilicon; the trenches are filled.

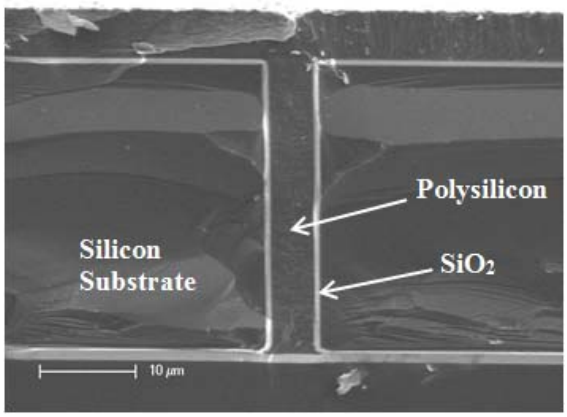



Figure 3.20. Cross-section SEM image of the trench profile after the LPCVD process. The conformal deposition of polysilicon fills the trenches.

After the LPCVD process, a CMP process is performed to polish off the polysilicon on top of the wafer surface and planarize the wafer surface. The CMP process is discussed in detail in section 3.1.9. Figure 3.21 is the schematic diagram of the cross section of the trench region after the CMP process. The top layer of the polysilicon is polished off, and the polish stops on the thermal oxide layer. Figure 3.22 is a cross-section SEM image of the trench region after the CMP process. We can see from the figure that the top layer of polysilicon is polished off and the etch stops on the thermal oxide layer. There are some charging effects due to the oxide on top of the wafer surface.

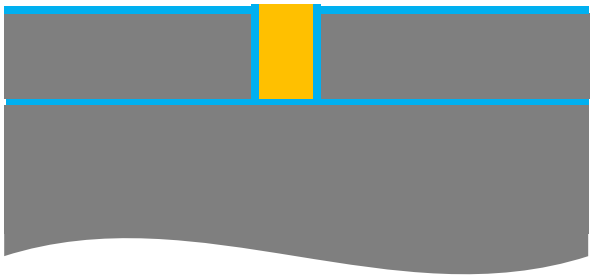


Figure 3.21. Schematic diagram of the cross section of the trench region after the CMP process.

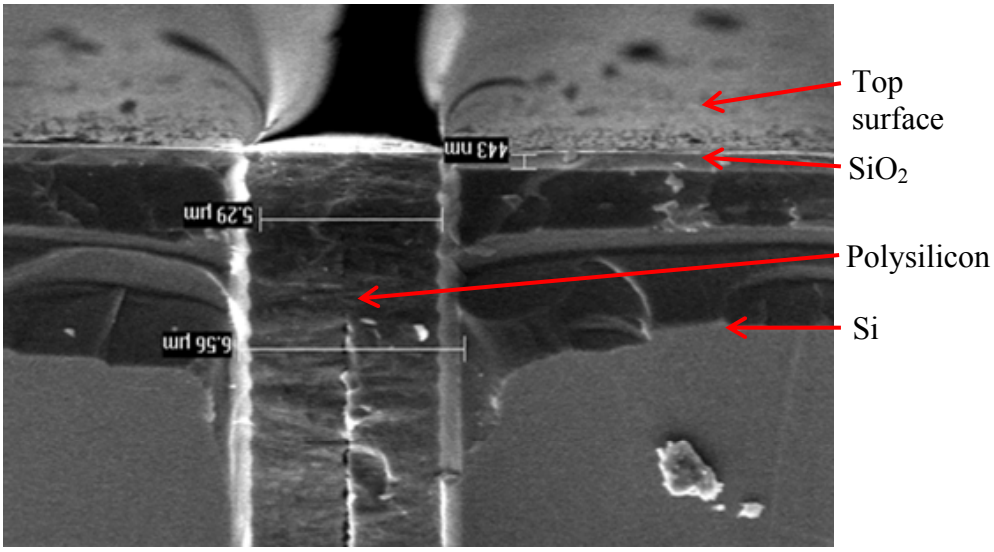



Figure 3.22. Cross-section SEM image of the trench region after the CMP process. The top layer of polysilicon is polished off and the etch stops on the thermal oxide layer.

Figure 3.23 shows an optical microscope image of the wafer surface after the CMP process. In the pictures, the green or pink regions are the thermal oxide; the white regions are polysilicon filled trenches. The CMP process polishes off the 4 μm polysilicon on the

wafer surface and stops after removing about 40 nm of the underlying thermal oxide, leaving a planarized wafer surface. This planarization process isolates the three photodiodes in each pixel and helps with the following photolithography processes, especially the photoresist spinning process. This is because photoresist will form islands on the wafers with unfilled trenches as shown in Figure 3.24.

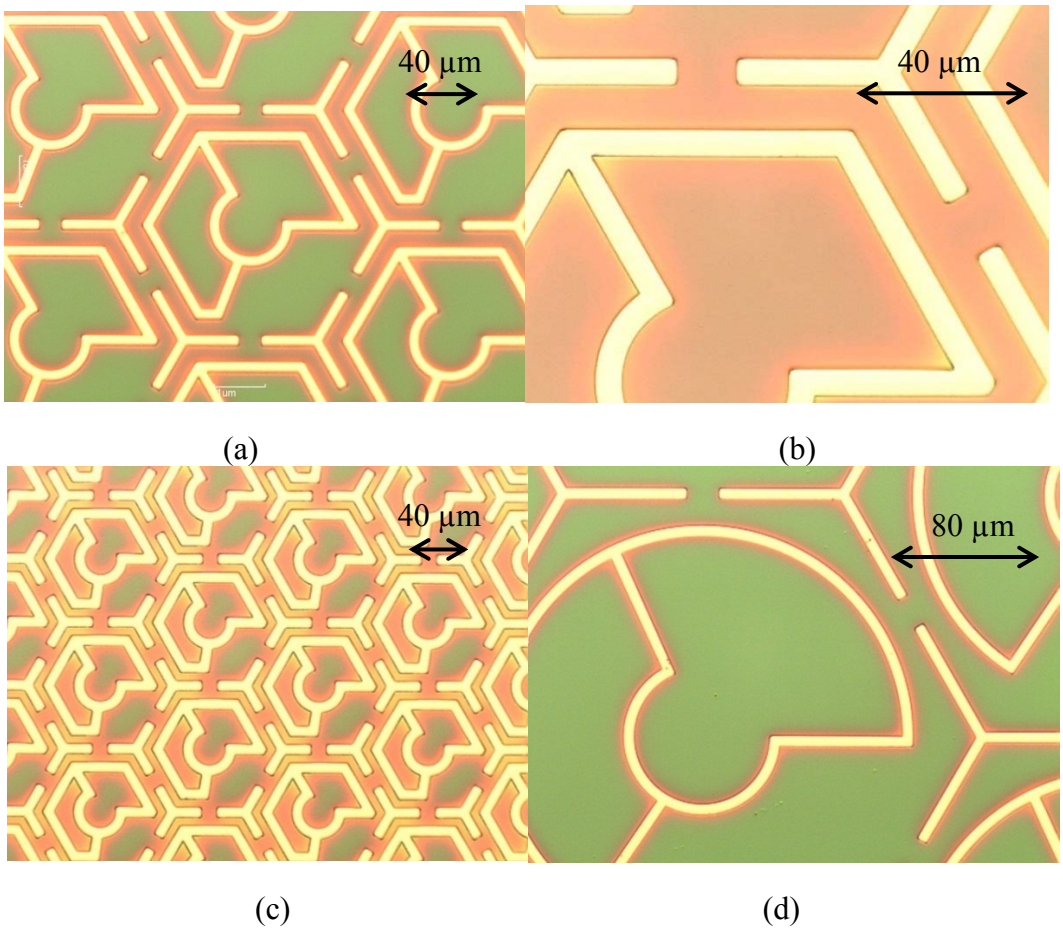
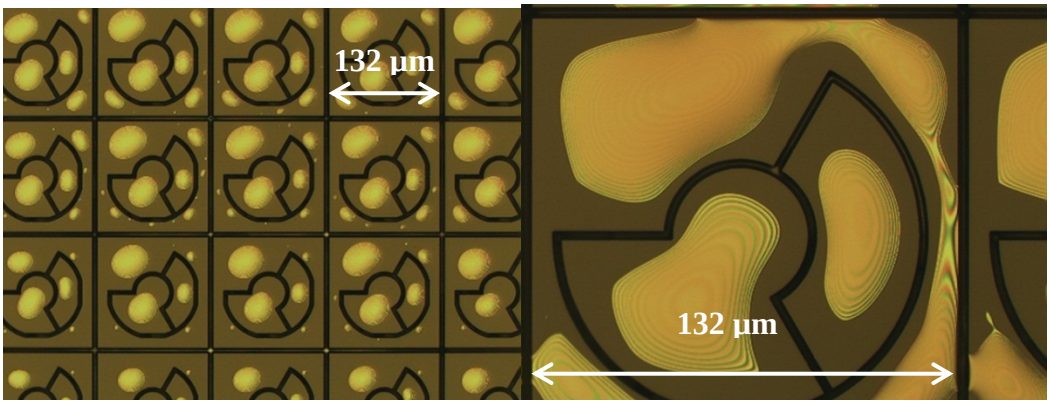



Figure 3.23. Optical microscope images of the wafer surface after the CMP process: (a) Medium size pixel. (b) Close-up of a medium-size pixel. (c) Small-size pixel. (d) Large-size pixel.

After the CMP process, any possible contamination of potassium, which is contained in the CMP slurry, should be removed. The decontamination process is necessary since potassium ($K^+$) is a lifetime killer and an extremely fast-diffusing alkali metal ion. In this cleaning procedure, wafers are cleaned in a self-heating 1:1:1 $H_2O$: $H_2O_2$: HCl solution for 20 minutes in quartz or Teflon labware. Pyrex beakers should be avoided since they contain 5% sodium. Polypropylene is also not acceptable since it has a low melting temperature. The cleaning solution is formed by pouring $H_2O$ first, then 30%

$H_2O_2$, then slowly adding the concentrated HCl. This mixture self-heats and boils and effectively decontaminates the wafers. KOH decontamination can also be performed in a solution of 5:1:1 $H_2O$: $H_2O_2$: HCl at 70 ˚C for 20 minutes. Then the wafers should go through dump rinse and spin/rinse dryer. After KOH decontamination, the wafers are considered clean and ready for further processing. Then a brief oxide etching process is performed using 6:1 buffered oxide etch (BOE) for 3 minutes. The etch rate in 6:1 BOE is about 90 nm/min. The BOE etch process will remove about 270 nm of thermal oxide on top of the wafer. This step is to further reduce any possible potassium contamination on the wafer surface.




(a) (b)

Figure 3.24. Spin coating problem on wafers with surface topology: photoresist forms islands on wafers with unfilled trenches. (a) Optical microscope image of the medium-size pixel array of an earlier version of the devices. (b) Close up of a medium-size pixel with photoresist islands on the surface.

#### *3.3.3.4 pn junction formation*

After the wafer planarization process as discussed in section 3.3.3.3, wafers are oxidized to protect the top of the polysilicon from the subsequent doping processes. Then wafers are ready for forming the $n^+$ and $p^+$ regions. A standard lithography step is performed to pattern the wafers, using the mask for $n^+$ regions. The oxide on top of the $n^+$ regions is again etched by the standard dry oxide etch program as discussed in section 3.3.3.2. After opening the $n^+$ regions, photoresist is removed by oxygen plasma resist burning and “Piranha”. Then wafers are further cleaned using the prediffusion cleaning process as discussed in section 3.3.3.1. A gas-phase phosphorus predeposition process then heavily dopes the opened $n^+$ regions. During the phosphorus predepositon process, a

highly resistive layer of phosphosilicate glass will form, which can be easily removed by a wet oxide etching process. We use the 6:1 BOE solution to etch the wafers for 1 minute, which removes all of the phosphosilicate glass. The sheet resistivity is measured using the four-point probe method, and the sheet resistance obtained for the $n^+$ region is about 4 Ω/sq. The formation of $n^+$ regions is illustrated in the schematic diagram of Figure 3.25. The phosphorus predeposition process heavily dopes the $n^+$ regions, which will form a pn junction in the lightly boron doped substrate. The heavily doped n regions will also form Ohmic contacts to the following metal layers.

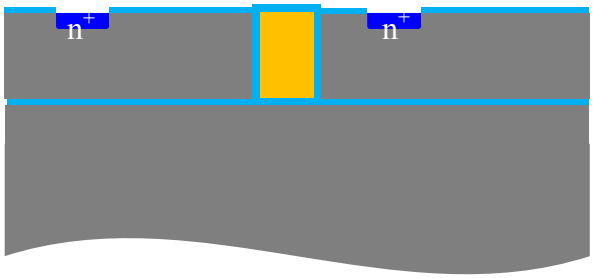



Figure 3.25. Schematic diagram of the cross section showing the formation of $n^+$ regions.

After the phosphorus predeposition process, wafers are cleaned at the prediffusion wet bench. An oxide mask is grown on the $n^+$ regions to block the subsequent boron ion implantation. The boron ion implantation process uses an ion energy of 20 keV. The corresponding range $R_p$ is 0.0826 μm and the standard deviation $\Delta R_p$ is 0.0384 μm [73]. An oxide mask with a thickness greater than $R_p + 3\times \Delta R_p \approx 198$ nm is used to block the boron dopant. Therefore, a wet oxidation process is performed at 1000 ˚C for 20 minutes, which grows about 260 nm of thermal oxide on top of the $n^+$ regions to protect them from the following boron ion implantation process.

Next, the standard photolithography process is used to define the $p^+$ regions. Oxide on top of the $p^+$ regions is etched by the standard dry oxide etch program as discussed in section 3.3.3.2. A thin layer of screen oxide with a thickness of 13 nm is grown by a dry oxidation process at 950 ˚C for 15 minutes. The screen oxide, which is amorphous, can minimize channeling by randomization of the incident beam, as discussed in section 3.1.5.1. The boron ion implantation process is done at INNOViON Corporation. $B^{11}$ singly-charged boron with a dose of $1.3\times10^{15}$ $cm^{-2}$ is implanted with an

energy of 20 keV at 7˚ tilt angle, which forms the $p^+$ region. Figure 3.26 is a schematic diagram showing the cross section of the wafer after the boron ion implantation process. The sheet resistivity measured on the $p^+$ regions after annealing is about 200 Ω/sq. The $p^+$ regions are used to make Ohmic contacts to the following metal layers.

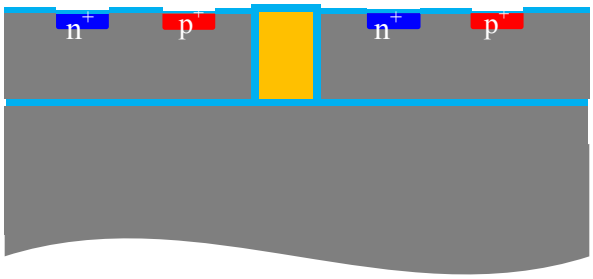



Figure 3.26. Schematic diagram of the cross section showing the formation of $p^+$ regions.

Next, a BOE etch removes all of the oxide on the wafer surface, and wafers are cleaned at the prediffusion wet bench. A dry oxidation process at 1000 ˚C for 83 minutes grows a 60 nm layer of $SiO_2$, which passivates the surface of silicon and forms the first layer of the anti-reflection coating. A subsequent 1000 ˚C anneal for 75 minutes activates the dopants and drives them to the final depth. Figure 3.27 shows the cross section of the wafer after the formation of the 60 nm thermal oxide and the final dopant drive-in and anneal process.

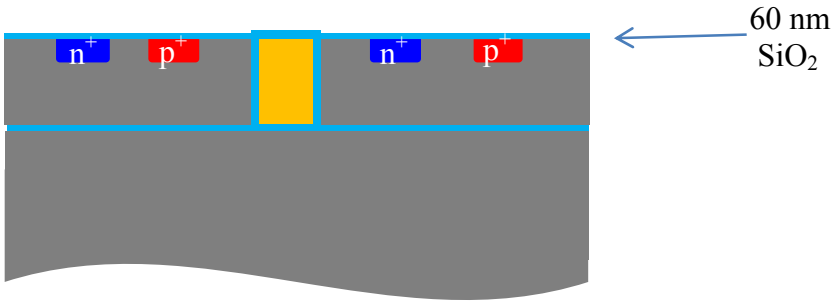



Figure 3.27. Schematic diagram of the cross section after forming the 60 nm anti-reflection oxide coating and 1000 ˚C anneal.

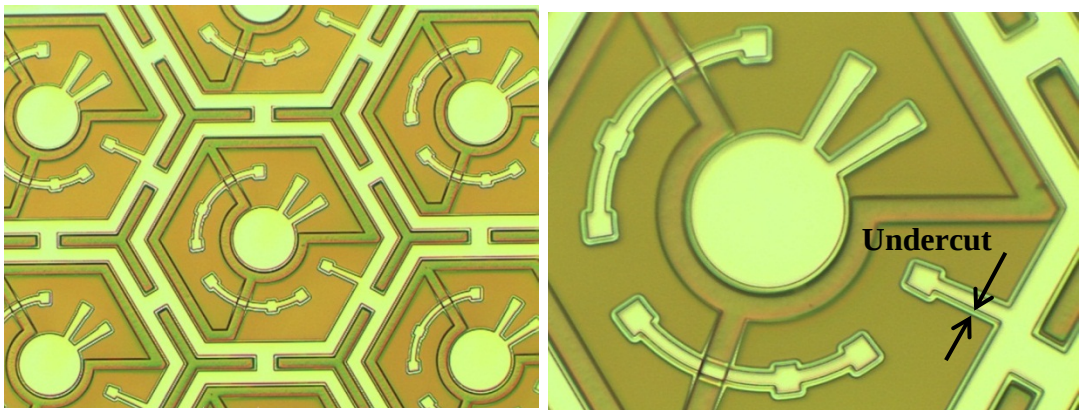
The doping profiles of the $n^+$ and $p^+$ regions were simulated in section 3.2. The simulated sheet resistance is 7.2 Ω/sq for the $n^+$ regions and 115 Ω/sq for the $p^+$ regions. The measured results are generally close to the simulated results: the measured results for the $n^+$ regions are about 3 Ω/sq, and the measured results for the $p^+$ regions are about 107

Ω/sq. The simulated results need to be corrected by a correction factor to reflect the experimental values.

#### *3.3.3.5 Metal deposition and liftoff*

After the pn junction formation in section 3.3.3.4, the standard photolithography is used to define the first via layer, and the standard oxide etch program is used to etch the oxide in the via regions. The vias are opened for making contacts to the following metal (Ti/Pt) layer, which then connect the three photodiodes in each pixel and connect the three series-connected photodiodes to the central active electrode and surrounding local return electrode.

Before the metal deposition, a dual-layer photoresist process is used to define the first metal layer. The dual-layer photoresist process is discussed in section 3.1.8.2. First, a 350 nm of LOL2000 is spun on the wafer surface at a spin speed of 1000 rpm. Then the wafers are oven baked at 210 ˚C for 30 minutes, which is critical in the control of the lateral undercut of the patterns after development. Then a 1 μm thick layer of 3612 is spun on the top of the LOL2000. Standard exposure and development define the pattern of the first metal layer, forming a small undercut in the underlying LOL2000 layer, which helps with the liftoff process for the metal layer. The photoresist pattern after development is shown in Figure 3.28, which shows the undercut of the underlying LOL2000.




(a) (b)

Figure 3.28. Optical micrograph of photoresist pattern after development showing the undercut in the LOL2000 layer: (a) Photoresist pattern on the medium-size pixels. (b) Close-up of (a).

Figure 3.29 shows the schematic diagram of the wafer cross section after the metal deposition and liftoff. The metal layer consisting of 20 nm of Ti and 250 nm of Pt is sputter deposited as discussed in section 3.1.8.1. The 20 nm layer of Ti acts as an adhesion layer. We chose to deposit 250 nm of Pt to have good step coverage since the wafer surface has some polysilicon steps as illustrated in the following graphs. Figure 3.30 shows the formation of the polysilicon steps. First, the LPCVD process deposits about 4 μm polysilicon on 800 nm thermal oxide (Figure 3.30(b)). After the CMP process, the wafer surface is planarized (Figure 3.30(c)). After the dopant drive-in and anneal process, the 60 nm oxide layer replaces the previous 800 nm oxide layer, and the polysilicon steps with a step height of about 400 nm form on the surface of the wafers (Figure 3.30(d)).

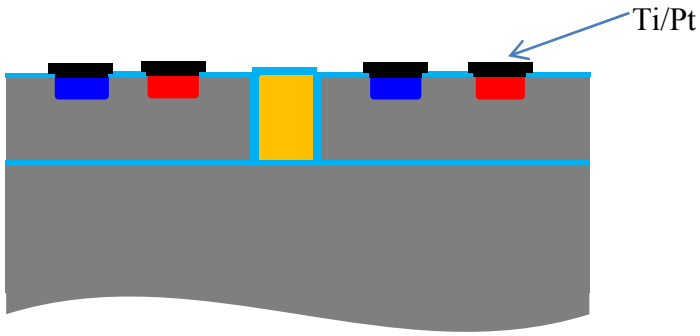



Figure 3.29. Schematic diagram of the wafer cross section after Ti/Pt deposition and liftoff.

Figure 3.31 shows a cross-section SEM image of the wafer surface with the visible polysilicon steps. With these polysilicon steps, one concern is whether the metal deposition has good step coverage across the polysilicon steps. Poor coverage on the sidewalls of the polysilicon steps will cause an open circuit since the metal lines that connect the three photodiodes in series need to go across the polysilicon steps. The metal sputter deposition system can provide good step coverage since the target metal atoms have a wide range of arriving angle as discussed in section 3.1.8. However, the target area we used is small compared to the wafer so we deposited a thick layer (250 nm) of Pt to achieve a better step coverage. The thickness of 250 nm of Pt is chosen since it can provide a better step coverage while still being thin enough to be lifted-off by the standard lift-off method as discussed in section 3.1.8. To test the conductance of the metal lines across the polysilicon steps, a current-voltage measurement is performed to

compare the metal resistance across the polysilicon steps and the metal resistance elsewhere for similar length of the metal lines.

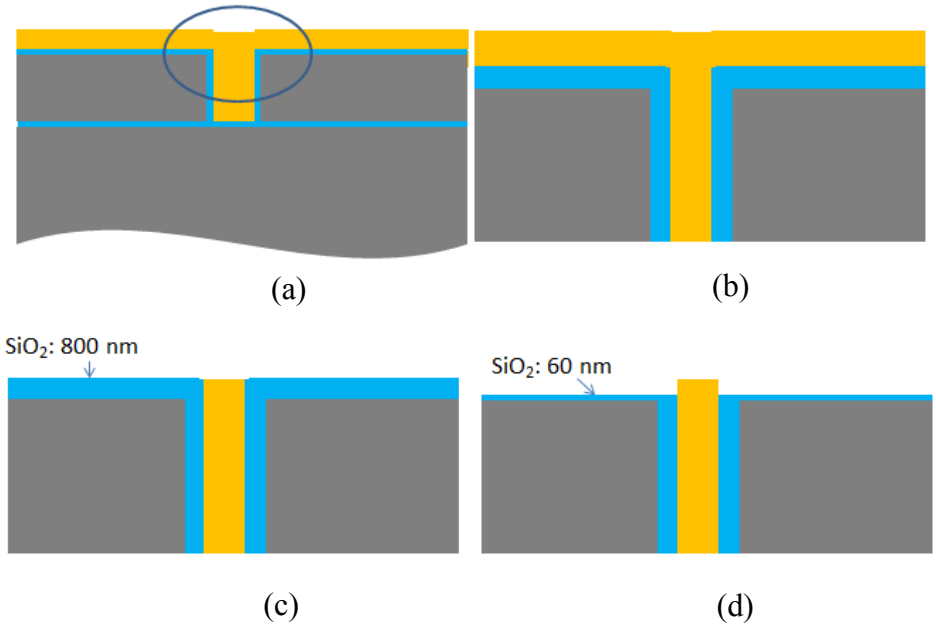



Figure 3.30. Schematic illustration of the formation of polysilicon steps: (a) Cross section of wafer surface after LPCVD polysilicon deposition. (b) Close-up of the circled area in (a). (c) After the CMP process, there is about 800 nm of thermal oxide on top of wafer surface. (d) After finishing dopant drive-in, 60 nm of oxide growth and 1000 ˚C annealing, the oxide thickness on top of the wafer surface is 60 nm, resulting a polysilicon step height of about 400 nm.

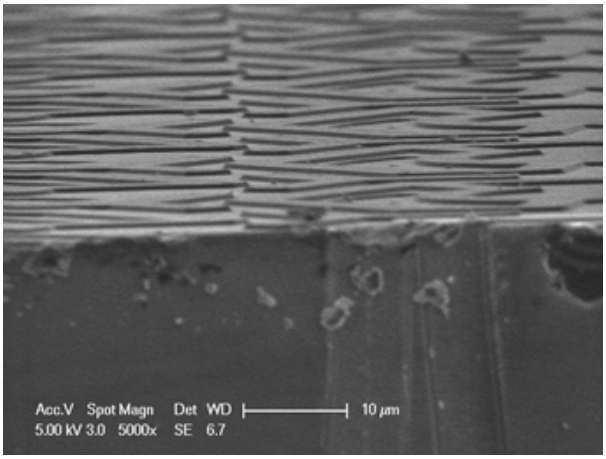



Figure 3.31. Cross-section SEM image of the polysilicon steps on the top of the wafer surface.

From Figure 3.32, the measured resistance from the inverse of the slope of the curves is 40.3 Ω for the metal line with one polysilicon step along the current pathway and 12.3 Ω for the metal line without any polysilicon step along the current pathway. This means the metal deposition is continuous across the sidewall of the polysilicon steps and the conductance is about 3 times poorer with polysilicon steps along the current pathway. Although this is the case, the added metal resistance is not a big problem since it is still within the 100 Ω range and should not cause too much series resistance with an estimated current of 10 μA in our application.

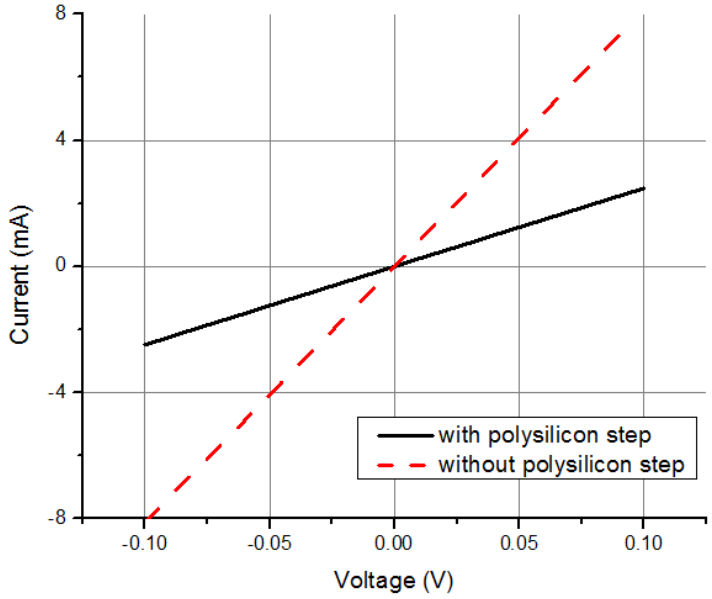



Figure 3.32. Current-voltage characteristic of the metal line measured across one polysilicon step (black line) and without any polysilicon step in the current pathway (red line).

After the metal deposition and liftoff, forming gas consisting of 4% $H_2$ in $N_2$ anneals the structure at 425 ˚C for 30 minutes, which improves the contact resistivity and reduces the $Si/SiO_2$ interface states. A 70 nm silicon nitride layer is then deposited by plasma enhanced chemical vapor deposition to provide electrical isolation and form the second layer of the anti-reflection coating. The PECVD process is discussed in detail in section 3.1.7.1.2. Since thin PECVD nitride layers tend to have pinholes in them, rotation of the wafers half way through the deposition process will help improve the electrical isolation.

A schematic diagram of the wafer cross section after the PECVD process is shown in Figure 3.33.

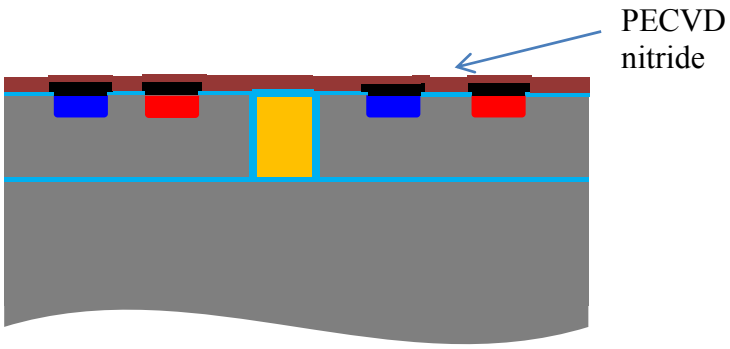



Figure 3.33. Schematic diagram of the wafer cross section after PECVD nitride deposition.

After the PECVD deposition of nitride, the second via layer is defined by standard photolithography. In the photoresist coating step, it is better to switch to the 1.6 μm thick photoresist process since the 1 μm thick photoresist does not protect the edge of the polysilicon steps well during the following nitride etch process. The second via mask layer is used to connect the first metal (Ti/Pt) layer to the second metal (iridium oxide) layer on the central active electrode and surrounding local return electrode. The nitride in the opened via region is etched by plasma dry etching using an $SF_6$ plasma. The etch rate is about 65 nm/min with the recipe summarized in Table 3.3.

| $SF_6$ | F116 | Pressure | Power |
|---|---|---|---|
| 50 sccm | 33 sccm | 200 mTorr | 90 W |

Table 3.3. Etch conditions for PECVD nitride etching (Adapted from SNF website: https://snf.stanford.edu/SNF/).

After the nitride etch process, another forming gas anneal process by 4% $H_2$ in $N_2$ at 425 ˚C for 30 minutes is performed to anneal the structure from the plasma damage during the PECVD nitride deposition and plasma dry etching of nitride. This will reduce the dark current and improve device performance. A schematic diagram of the wafer cross section after the second via etching is shown in Figure 3.34.

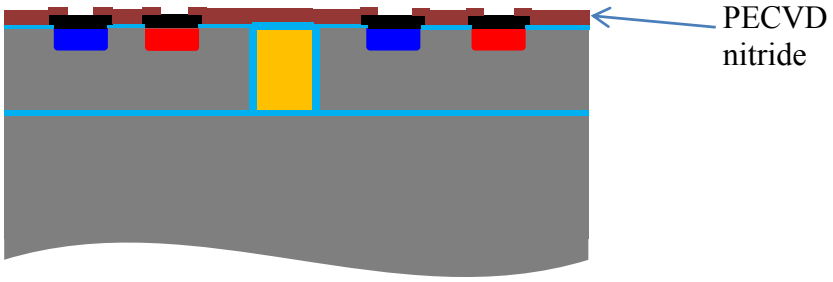



Figure 3.34. Schematic diagram of the wafer cross section after the second via etching.

After defining the second via layer, the standard photolithography is used to define the pattern of the selected polysilicon etch, with a photoresist thickness of 3 μm. The use of a thicker photoresist is to make sure that the photoresist is not etched completely during the series of etching steps in this layer. The series of etching steps include etching of the nitride on top of the selected polysilicon regions, etching of the thermal oxide underneath the nitride in the exposed regions, etching of polysilicon by $XeF_2$ etching and the directional oxide etch at the bottom of the trenches for some of the wafers. The nitride and oxide is etched by the standard programs discussed before. The polysilicon is removed in selected regions by $XeF_2$ (xenon difluoride) etching. The etching process can be described as follows:

$$2XeF_2 + Si \rightarrow 2Xe + SiF_4 \qquad \text{(Eq. 3. 37)}$$

The etch process is a vapor phase chemical etch, which is isotropic and highly selective to silicon dioxide and silicon nitride. This etch is dry and vacuum based without any plasma activation, which minimizes the damage to other materials on the wafer. There are several issues during processing. First, $XeF_2$ etching involves a chemical diffusion process, so the etch rate and etch profiles are highly dependent on loading effects. In our fabrication process, the etch rate of polysilicon on wafer pieces is about 1 μm/min. While the etch rate of polysilicon on 4 inch wafers is non-uniform, with a small etch rate in the center of the wafer and high etch rate near the edge of the wafer. The etch time needed to remove polysilicon from selected trenches across the whole wafer is significantly larger than that for etching wafer pieces (25 cycles to etch 30 μm on wafer pieces compared to 125 cycles to etch 30 μm on a 4 inch wafer). Another issue is to avoid the exposed silicon at the edge of the wafer, which leads to more etch non-uniformity. Edge bead removal of the photoresist should be avoided so that the edge of the wafer will

be protected by photoresist. The third issue is that before $XeF_2$ etching, the native oxide on the surface of the exposed area should be removed. In our process, the wafers are directly loaded into the $XeF_2$ etch chamber after the oxide etch. The fourth issue is that since the reaction between $XeF_2$ and silicon is exothermic, a delay step between a certain number of cycles should be used to cool the wafer to minimize any observed thermal issues. In our process, a five minute delay is added between every 25 cycles of etching to cool the wafer, which will minimize the heating of the photoresist. A schematic diagram of the wafer cross section after $XeF_2$ etching of polysilicon is shown in Figure 3.35.

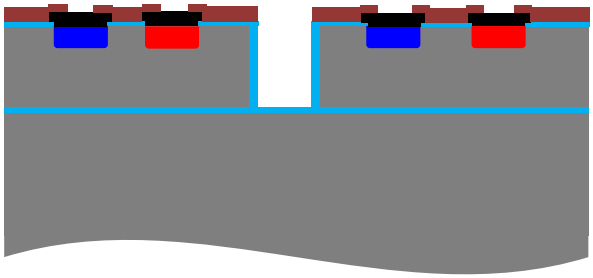


Figure 3.35. Schematic diagram of the wafer cross section after $XeF_2$ etching of polysilicon in selected regions.

To aid the array release process which will be discussed later, the exposed buried oxide at the bottom of the trenches is also etched on some of the wafers. The directional oxide etch is done in the P5000 etcher (Applied Materials Precision 5000 Etcher), which is a Magnetically-Enhanced Reactive Ion Etch (MERIE) system. In the etch chamber, wafers are placed on an RF (13.56 MHz) powered electrode; the electrode that the wafers are sitting on is smaller than the grounded counter electrode, which develops a large dc bias on the wafer electrode and results in high energy (100 to 600 eV) ion bombardment of the wafer surface; a magnetic field (30 to 100 gauss) is applied parallel to the wafer surface, which increases the plasma density and rate of ion bombardment to increase the etch rate and directionality. The etch rate of thermal oxide is about 380 nm/min for the oxide on the wafer surface using the standard oxide etch program. The schematic diagram of the wafer cross section after the directional oxide etch is shown in Figure 3.36.

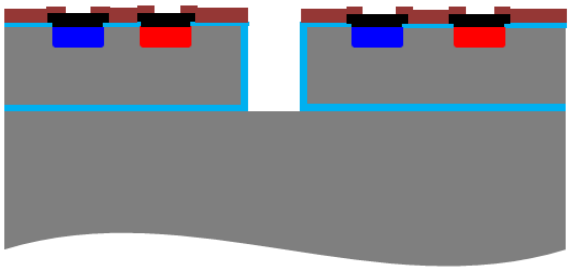


Figure 3.36. Schematic diagram of the wafer cross section after the directional oxide etch at the bottom of the exposed trenches.

Figure 3.37 is a cross-section SEM of the bottom of the trench after a directional oxide etch in the P5000 oxide etch chamber for 2 minutes. Before the directional oxide etch, the oxide thickness is similar for the sidewall and the bottom of the trench. After the 2 minutes directional oxide etch, most of the thermal oxide at the bottom of the trench is removed while the oxide on the sidewall of the trench is barely etched. We can see from the figure that about 500 nm thermal oxide at the bottom of the trench is etched, corresponding to an etch rate of 250 nm/ min, which is smaller than the etch rate of the oxide on the surface of the wafer (380 nm/min).

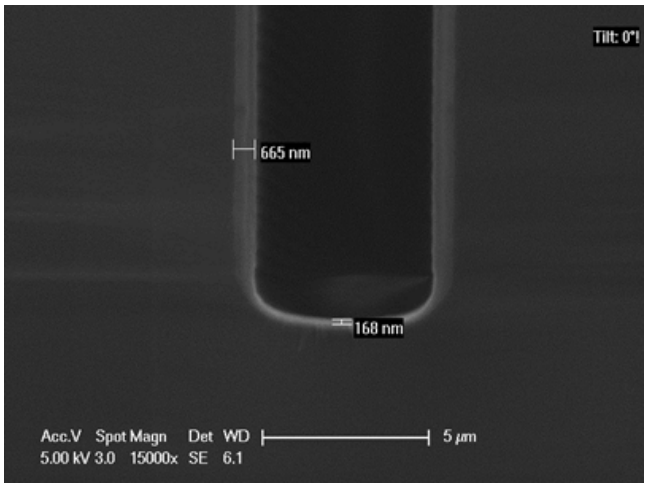


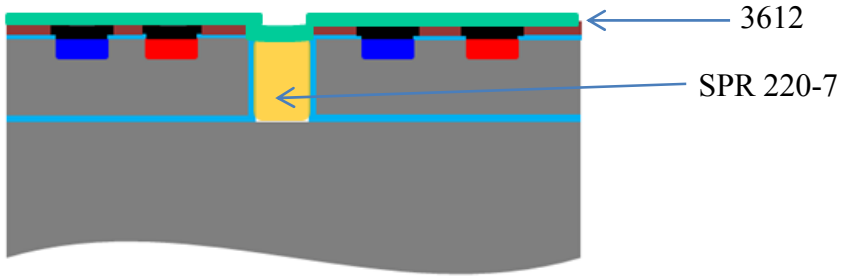

Figure 3.37. Cross-section SEM of the bottom of the trench showing that the directional oxide etch removes most of the thermal oxide at the bottom of the trench while the oxide at the sidewall is barely etched.

After the polysilicon etch in selected regions, the wafers are annealed in forming gas at 425 ˚C for 30 minutes to anneal the damage caused by the various plasma processes. Wafers are then ready for depositing the second metal layer. However, the

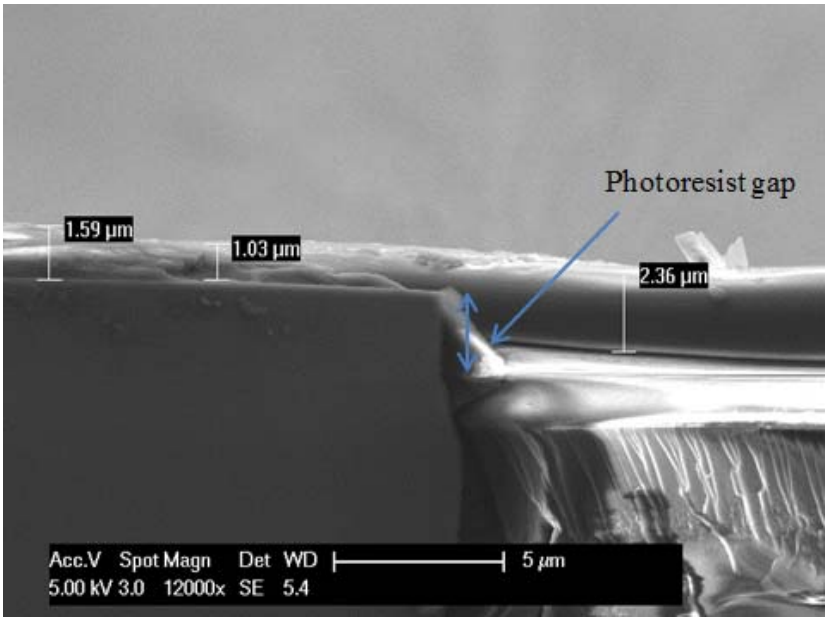
wafer surface has a rough topology which makes it impossible to spin a thin layer of photoresist as described in Figure 3.24. We developed a two layer photoresist photolithography process that defines the pattern of the second metal layer. First, we use the spray coater as discussed in section 3.1.10 to coat the wafer with 28 μm photoresist. The photoresist we used to coat the first layer of photoresist is 1:3 SPR 220-7: MEK (Methyl Ethyl Ketone). The spray coater coats the wafer with four runs, resulting in a photoresist thickness of about 28 μm after a 10 minutes bake on a 90 ˚C hotplate. We chose to spray coat about 30 μm thick photoresist layer since the trenches will be filled with photoresist and the wafer surface will be planarized. The coating of photoresist is not uniform, with photoresist thickness variations of about 1 μm from the edge to the center. A blanket exposure by a contact exposure system (KarlSuss MA-6 Aligner) is performed to expose the top layer of photoresist. The first exposure is in contact mode using 5 exposure cycles with 3 seconds for each exposure. Then wafers are manually developed in beakers by the standard developer MF-26A. The development time should be long enough so that the exposed photoresist will be fully developed. A second exposure consisting 5 exposure cycles of 2 seconds each exposes the remaining photoresist on the wafer surface. A manual development after the exposure with long enough time will remove the remaining layer of photoresist on the top of the wafer surface, leaving the exposed patterned substrate. Since the thickness of the photoresist after spray coating will vary, an extra 1 second exposure may be needed to be added to expose any remaining photoresist, which will be developed away in the manual development process. After the multiple exposure and development processes, the top surface of the wafer is clear of photoresist, while all of the trenches are filled with photoresist. The wafer surface is planar again with trenches filled with photoresist. A spin coating process then coats the wafers with a thin, uniform layer of photoresist to define the patterns of the next layer. The schematic diagram and cross-section SEM image in Figure 3.38 show the two-layer photoresist process. From this figure, it is clearly seen that the first photoresist from the spray coating fills in the trenches, and then the second photoresist is spun on the top surface, which is thin and uniform in most of the regions. However, due to the small gap of photoresist filling on the top of the trenches, the second photoresist is not uniform near the trench areas. In Figure 3.38, we can see that the spin

coated second photoresist varies from about 1.6 µm to 1.03 µm near the edge of the trench; and the spin coated second photoresist layer fills in the gap at the top of the trench area with a thickness of about 2.4 µm. In some regions, the corners of the trenches are only covered by a thinner layer of photoresist as shown in Figure 3.39, where the photoresist thickness at the corner of the trench is only about 400 nm, while regions away from the trench are covered by a layer of photoresist with a thickness of 1.4 µm. The much thinner photoresist coverage at the corners of the trenches means that if any subsequent processes involves dry etching, the etch time will be limited since the corners of the trenches will lose protection from photoresist after a certain length of time. For the case of the second batch of devices, this is not a problem since the polysilicon is removed after the second via etch, so the two layer photolithography process is only needed for the last mask layer: the second metal (iridium oxide) formation. Due to the relatively large features of the second metal layer, we use a 4 µm thick SPR 220-3 photoresist layer as the second layer of photoresist.

With the first layer of photoresist filling the trenches, the second layer of photoresist can then be spin coated to form a thin, uniform layer of photoresist on the wafer surface. Figure 3.40(a) shows the photoresist coating on wafers with surface topology with only spin-coated photoresist, which forms islands and cannot be used for defining photolithography patterns. Figure 3.40(b) shows that with spray coated photoresist filling the trenches, the spin coated photoresist forms a thin, uniform layer of photoresist, which can then be used to define the next photolithography pattern.




(a)




(b)

Figure 3.38. Schematic diagram of the cross-section structure (a) and cross-section SEM image (b) of the trench area showing the two layer photoresist: trench filled with SPR 220-7 and the photoresist on the top surface is replaced by 1 µm 3612.

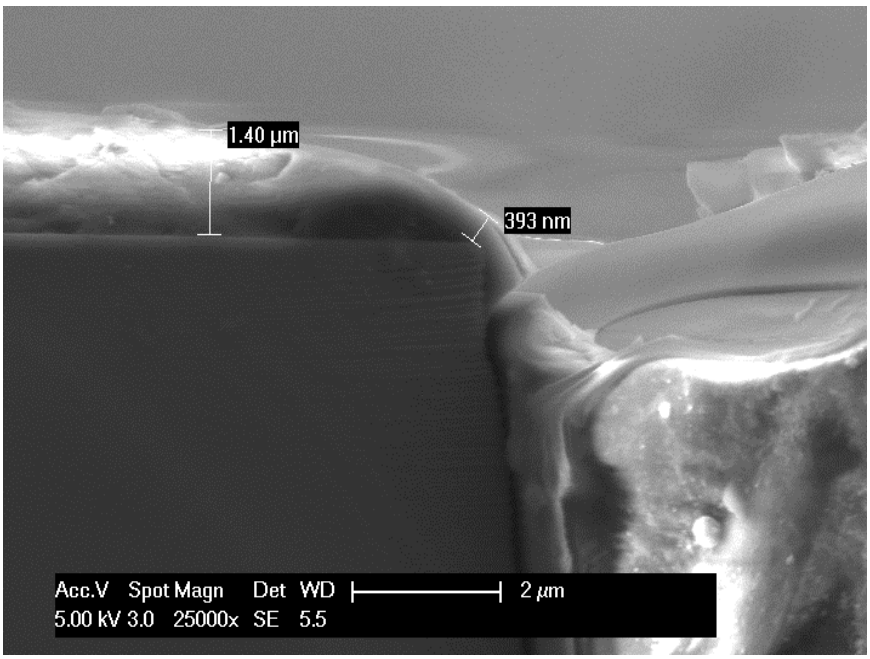
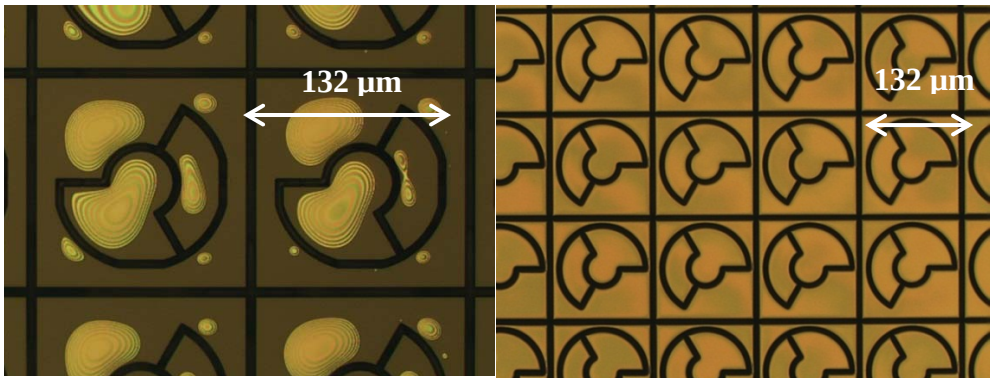



Figure 3.39. Cross-section SEM image of the trench area showing the two layer photoresist: trench filled with SPR 220-7 and the photoresist on top surface is replaced by 1 µm 3612. The corner of the trench is covered only by a thinner layer of photoresist.




(a) (b)

Figure 3.40. Photoresist coating on wafers with surface topology: (a) spin coated photoresist forms islands; (b) with spray coated photoresist filling in the trenches, spin coated photoresist forms a thin, uniform layer of photoresist.

By using the two-layer photoresist method mentioned above, the pattern of the second metal layer is defined by the standard exposure and development process. A short $O_2$ plasma descum process will remove the photoresist residue on the exposed areas of the wafers. Then the wafers are sent to EIC Laboratories, Inc. for iridium oxide coating. After that, the iridium oxide is lifted off by soaking the wafers in acetone for about 2 hours, and a Q-tip is usually used to gently sweep across the surface of the wafer to help with the liftoff process. After acetone soaking, the wafers are taken out of the beaker with a thorough acetone stream rinse. Photoresist residue is further cleaned using methanol and isopropanol for 10 minutes each. A schematic diagram and an optical microscope image show the wafer surface after the formation of the iridium oxide electrodes in Figure 3.41.

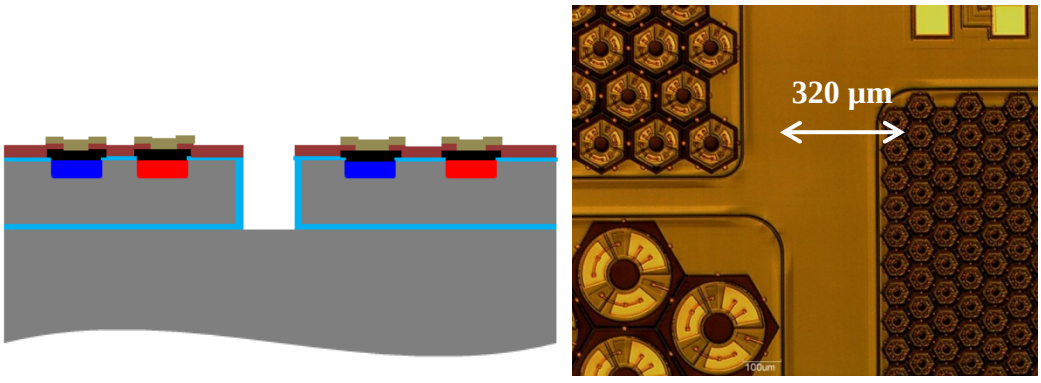



Figure 3.41. Schematic diagram (a) and optical microscope image of the wafer surface (b) after the formation of the iridium oxide electrodes.

### *3.3.3.6 Device release*

After finishing the second metal deposition and liftoff, we released the individual arrays from the SOI handle wafer so that they are thin enough to be implanted into the back of the retina. The release process first involves protecting the front side of the wafers with ProTEK (Brewer Science, Inc.), which is a spin-applied polymeric coating that protects wafers from wet etching chemicals such as alkaline or acid etchants. In our wafer release process, we protect the front side of the wafer with ProTEK B3 thin film and use tetramethylammonium hydroxide (TMAH) to etch the back side of the wafers. Before spinning ProTEK, a ProTEK primer is spun manually by the Headway coater in SNF at 1500 rpm and baked on a hot plate for 60 seconds at 205 ˚C. The ProTEK film is then spun on the wafer surface at 1500 rpm, which results in a film thickness of about 7 μm. Then the ProTEK is first soft baked on a hotplate for 120 seconds at a temperature of 130 ˚C. Then the coated ProTEK film is hard baked on a hot plate at 205 ˚C for 60 seconds. Figure 3.42 is a cross-section SEM of the wafer surface of the trench area showing that the 7 μm ProTEK layer protects the top surface of the wafer, and also fills the trenches, which will protect the sidewalls of the trenches from the TMAH etch.

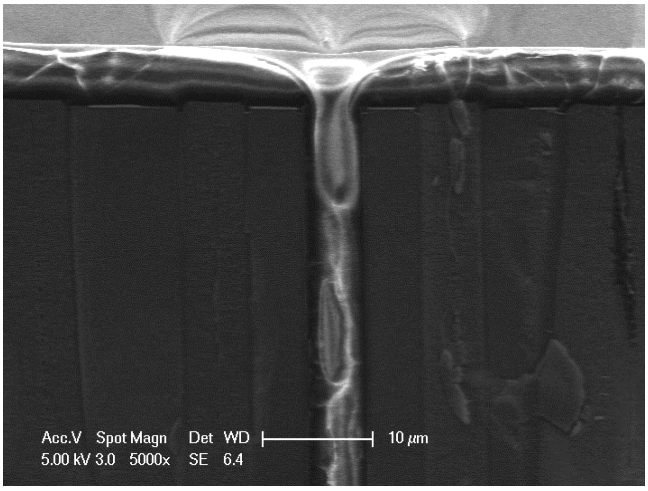



Figure 3.42. Cross-section SEM image of the trench area after applying the 7 μm thick film of ProTEK.

After coating the wafers with ProTEK, the wafers are sent to Silicon Quest International, Inc. for backside grinding. Before grinding, a UV release tape is applied on the top surface of the wafers for protection during backside grinding. The backside grinding process thins the wafers from a total thickness of 537 μm (500 μm handle wafer

+ 30 μm device layer + 7 μm ProTEK layer) to a total thickness of about 80 μm. When the desired thickness is achieved by the backside grinding, the UV release tape is UV exposed so that it is ready to be peeled off. When the wafers are shipped back, a TMAH etching process is performed with UV release tape left on top of the wafers. A mechanical clamp with O-ring sealing is used to further protect the wafer surface from the TMAH etching, with only the backside of the wafer exposed to the TMAH solution. Before the TMAH etching, a brief 50:1 HF dip is used to remove any native oxide on the backside of the wafers since TMAH etches native oxide in a much lower rate. After the HF dip, wafers are rinsed in a DI water bath but not dried before the TMAH etch. The TMAH etch was discussed previously in section 3.1.6.1. We use a solution of 10 % TMAH in DI water at 90 ˚C to etch the remaining silicon (about 40 μm) of the handle wafer, which has an etch rate of about 1-1.2 μm/min. As we can see from [85], for a certain range, the etch rates of TMAH at 90 ˚C decreases with increasing TMAH concentration as summarized in Table 3.4. Therefore, the etch rate is expected to decrease during the etching process since a small portion of the water will be evaporated during the etching process.

| Concentration | Etch rate at 90 ˚C, (100) direction |
|---|---|
| 5% TMAH 95% $H_2O$ | 1.4 μm/min |
| 10% TMAH 90% $H_2O$ | 1.2 μm/min |
| 22% TMAH 78% $H_2O$ | 0.9 μm/min |

Table 3.4. Etch rate dependence on TMAH concentration at 90 ˚C for (100) direction: for the cited concentration range, the etch rate decreases with increasing concentration [85].

The total etch time to remove about 40 μm of silicon is longer than expected, about 1 hour. TMAH etch will stop on the buried oxide layer. There are two ways to tell where we should stop the etching process: first, the bubbling phenomenon is gone when the etching process reaches the buried oxide layer; second, the patterns on the device layer appear as shown in Figure 3.43.

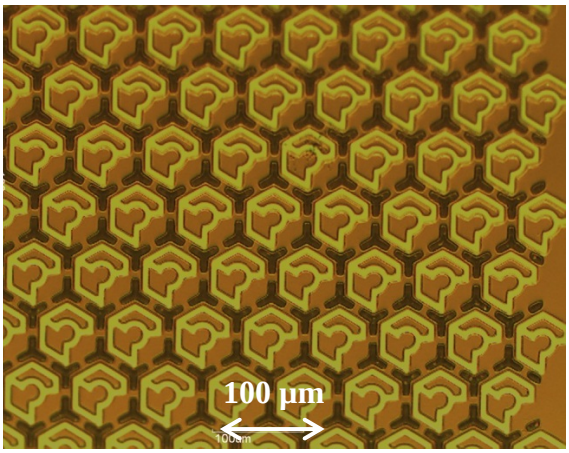



Figure 3.43. Backside of the wafer after TMAH etch, exposing the buried oxide layer.

One issue during the TMAH etching is that the etch rate is not quite uniform across the wafer. Therefore, the etching of the remaining silicon on the backside of the wafer completes at different times. As shown in Figure 3.44, most of the areas on the backside of the array are cleared of silicon with the buried oxide exposed, but the top left corner still has remaining silicon due to the etch rate difference in different areas of the array. Therefore, there is a compromise: the longer the etch time, the more cleared the backside is of remaining silicon, but the longer etch time also means increasing risk of damaging the front side of the wafers. Figure 3.45 shows an example of the wafer surface with incomplete protection by the ProTEK, which means that the etch time for this area was too long that the wafer lost protection from the ProTEK layer on top, so TMAH may attack the arrays from the front.

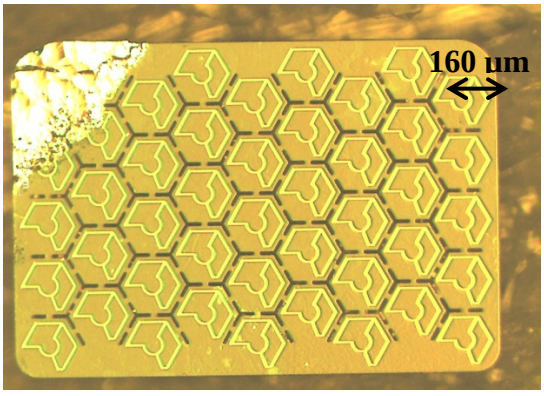



Figure 3.44. Backside of an array after TMAH etch, with the buried oxide exposed on most of the regions, but the top left corner of the array still has remaining silicon, due to the etch rate difference in different areas of the array during etching.

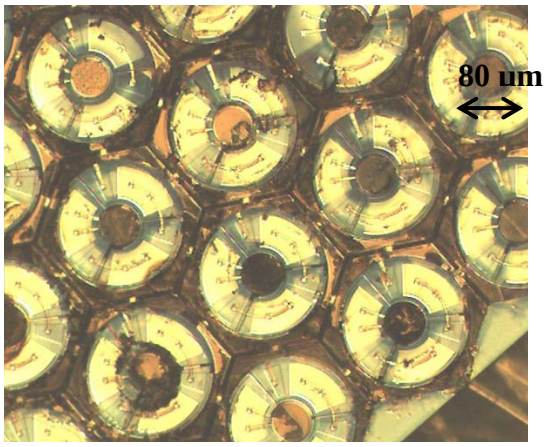



Figure 3.45. One piece of the wafer surface showing the incomplete etch protection by the ProTEK.

After TMAH etching, the wafers are water rinsed for a long time and then soaked in acetone. Acetone will remove the UV release tape on the front of the wafer and remove the ProTEK protection layer. Wafers are then separated into pieces by rinsing with acetone, so that most of the photodiode arrays are released. A schematic diagram in Figure 3.46 shows the cross section of the wafer after backside grinding, TMAH etching and acetone cleaning. For the wafers with a directional buried oxide etch, the arrays are separated or connected only by a very thin layer of suspended buried oxide.

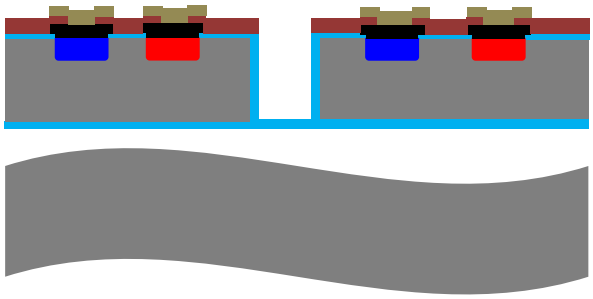


Figure 3.46. Schematic diagram of the wafer cross section after the release process.

Most of the arrays are not released since they are connected by the suspended buried oxide at the bottom, which can be broken by a short ultrasonic process. Figure 3.47 schematically shows that after 1-5 minutes ultrasonic agitation, the suspended buried oxide is broken so that the arrays are disconnected from each other. The length of the ultrasonic process depends on the previous directional oxide etch conditions and the releasing process on each individual wafer.

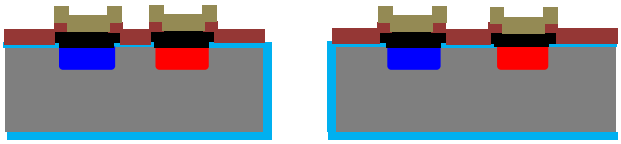
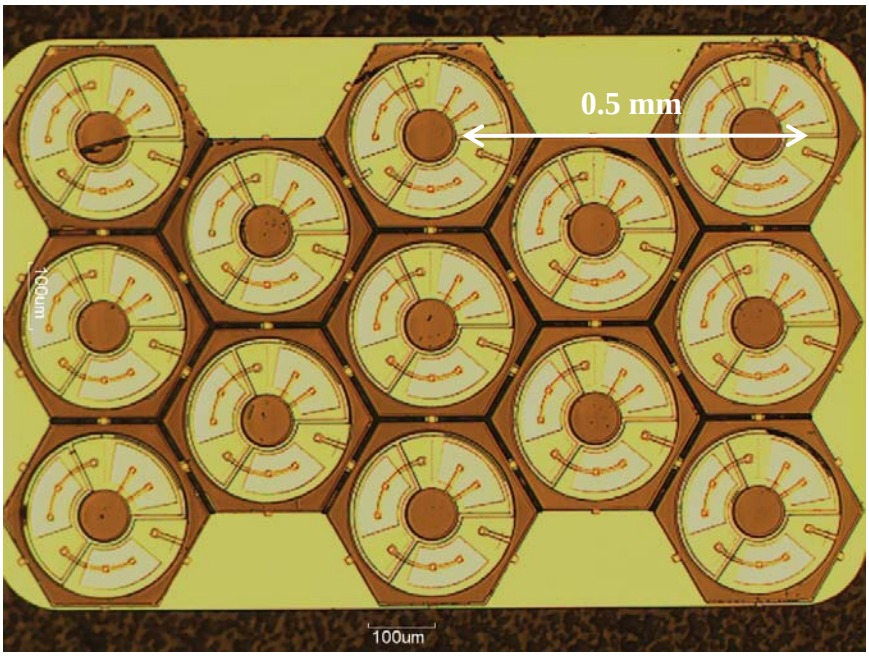
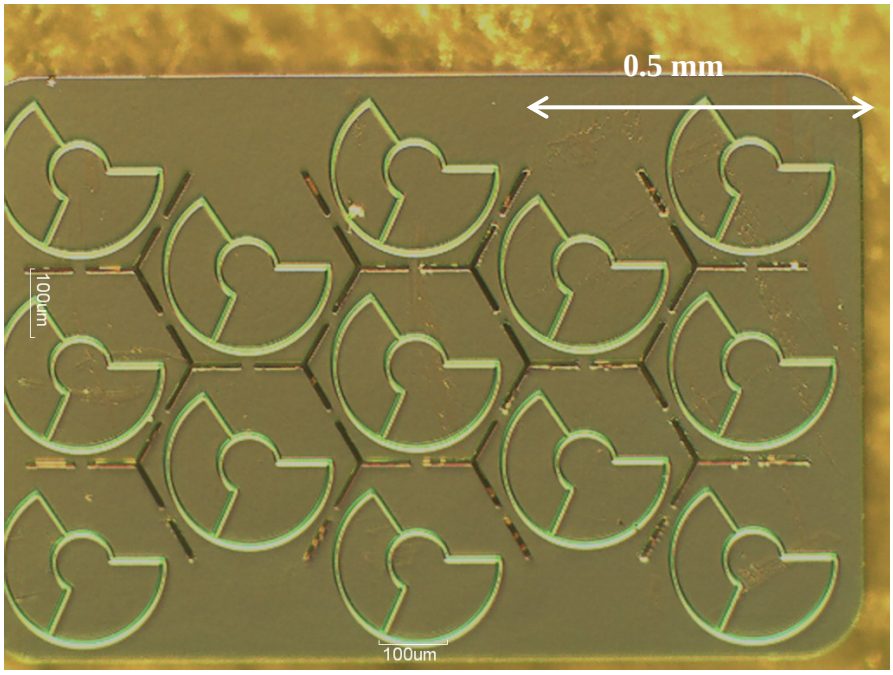


Figure 3.47. Schematic diagram of the wafer cross section after a short ultrasonic process.

Optical microscope images of the released arrays are shown in the following pictures. Figure 3.48 shows the front-side and back-side optical microscope images of an array with large-size pixels. We can see that the trenches between the pixels are etched through, which will provide a channel for nutrient flow in biological systems. Figure 3.49 and Figure 3.50 show optical microscope images of the front-side and back-side of the arrays with medium- and small-size pixels, respectively. Each of the arrays is 1.2 mm × 0.8 mm in size, and 30 µm thick, which can be easily implanted into the subretinal space. The TMAH etching and ultrasonic process will cause some damage on some of the arrays, such as etching into the surface of the arrays or damaging the structure of the arrays; therefore, each device needs to be carefully examined to select good arrays. A good array will have uniform metal coverage on electrodes, no ProTEK residue on the top surface, and cleared trenches.

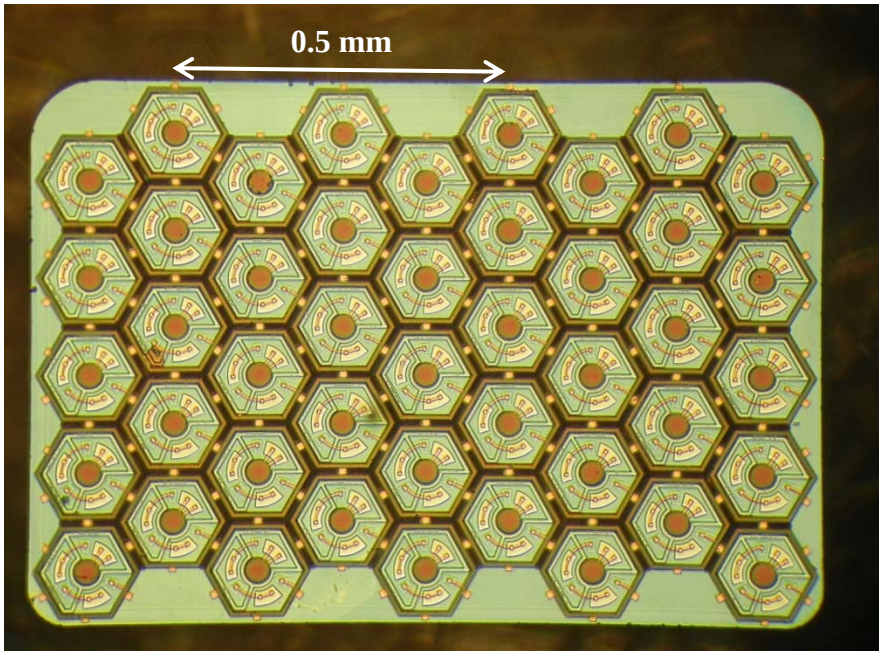



(a)

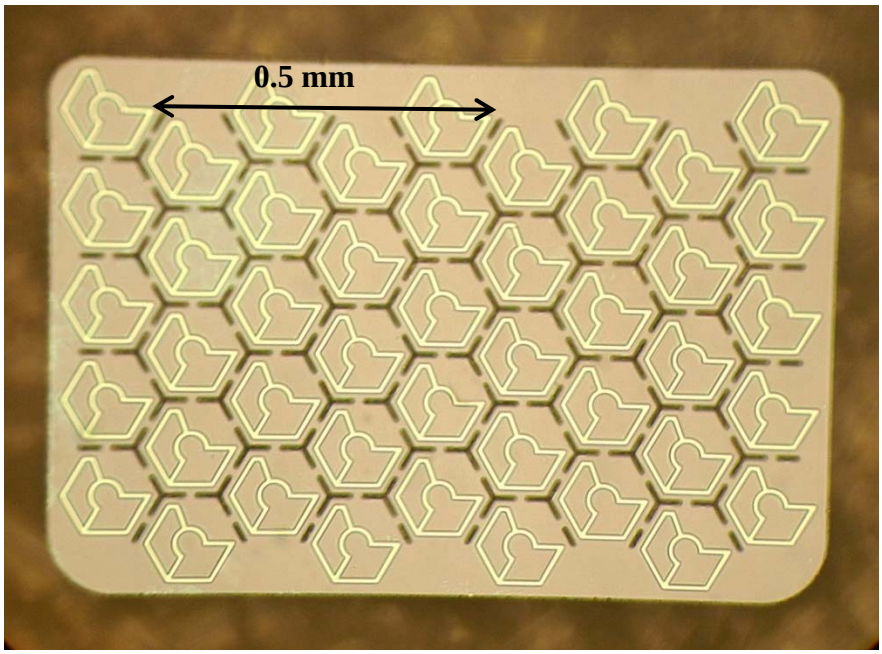



(b)

Figure 3.48. Optical microscope images of an array with large-size pixels: (a) front side of the array; (b) back side of the array.

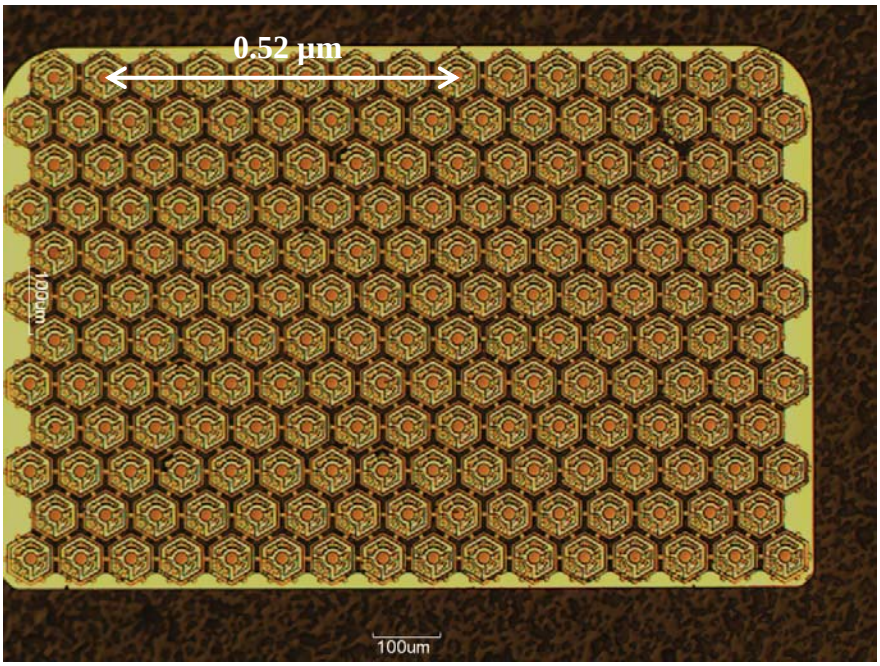



(a)

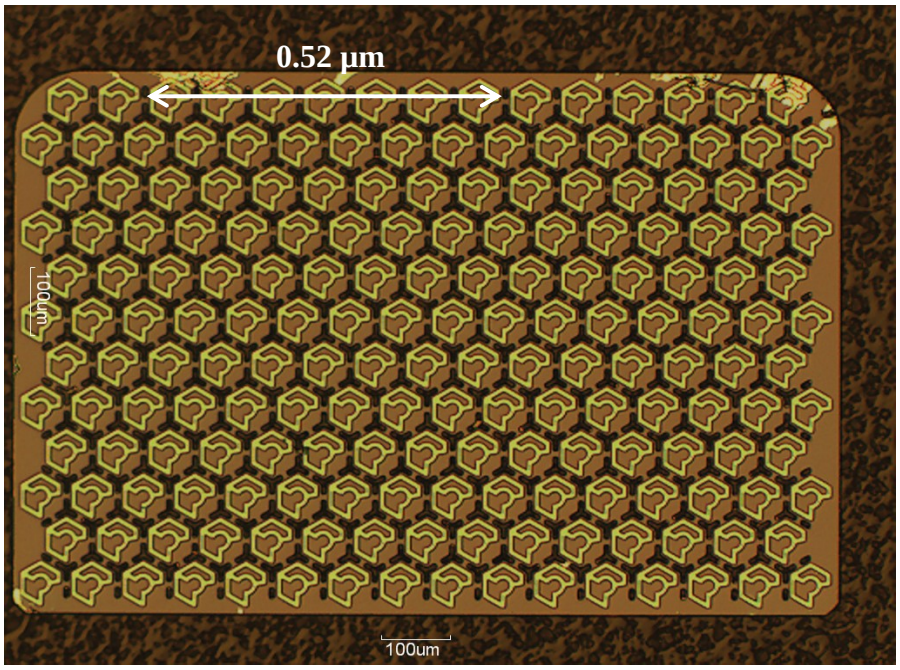



(b)

Figure 3.49. Optical microscope images of an array with medium-size pixels: (a) front side of the array; (b) back side of the array.




(a)




(b)

Figure 3.50. Optical microscope images of an array with small-size pixels: (a) front side of the array; (b) back side of the array.

Released arrays are stored in 4 inch wafer holders shown in Figure 3.51(a). The membrane box shown in Figure 3.51(b) can be used to transfer the released arrays during shipment. The vacuum pen shown in Figure 3.51(c) is used to handle the released arrays. All these tools come from Ted Pella, Inc.

(a)

(b)

(c)

Figure 3.51. Tools used to handle, store and transfer the released arrays: (a) Storage box for the released arrays. (b) Transfer box for the released arrays. (c) Handling tool for the released arrays. Tool pictures come from Ted Pella, Inc.

#### *3.3.3.7 Fabrication issues*

Besides the fabrication issues discussed in the previous section, there are several other issues that affect the formation of the structures, such as the lateral over etching of the SOI wafers at the silicon/buried-oxide interface during the DRIE process, the undercut at the mask/silicon interface during the DRIE process causing shrinkage of features such as bridges, the removal of ProTEK residue from the released arrays, and the incomplete filling of trenches by an inadequate LPCVD polysilicon layer thickness. These issues will be discussed in this section.

##### **3.3.3.7.1 Notching**

During the deep trench etching process of the SOI wafers, one potential problem is a lateral undercut when the silicon trench etching reaches the bottom buried oxide layer. This phenomenon is called "notching" and is usually caused by charging of patterned structures due to the different directionality between ions and electrons. Increasing sidewall protection with increased polymer deposition on the sidewall can reduce notching, but the etched profiles vary with different pattern spacing. The notching can be reduced by lowering the rf bias frequency from 13.56 MHz to 400 kHz under most conditions [94]. By employing low frequency, the charged ions on the insulator layer are allowed to escape after the etching cycle, which prevents charge accumulation [95]. The pulsed mode, when used in conjunction with the low frequency, can further reduce the charge build up and thus notching. In the pulsed mode, the platen power is turned off for a desired time interval, and the charged ions are allowed to escape during the off step [95]. Other methods, such as a conformal PECVD oxide coating and anisotropic oxide etch between several segment of the DRIE etching to protect the sidewall are also useful in solving the notching problem regardless of the feature sizes [96].

In our DRIE etching process, the notching was serious for the third batch of wafers, where the buried oxide thickness is 1 μm thick. This thick buried oxide layer probably causes serious charging at the bottom of the trench, which deflects the direction of the impinging ions to the sidewall at the bottom of the trench and causes the notching. For the second batch of devices, notching was not a serious problem since the buried oxide layer is 100 nm thick and the charging is not as serious. Figure 3.52 shows the

serious notching problem at the bottom of the trench where the 30 μm thick silicon device layer interfaces with the 1 μm thick buried oxide layer. The over-etch time for this sample is 20% using only the high frequency rf power supply in STS1 at SNF. The extent of the lateral etching is as large as about 4 μm due to the serious charging of ions on the buried oxide layer at the bottom of the trench. Figure 3.53 shows that a part of the small-size pixel is completely undercut and released from the wafer. Due to the serious notching problem in the third batch of devices, almost all of the smaller features were released from the wafer, and the large notching for the relatively larger features may cause reliability issues for the following processing steps.

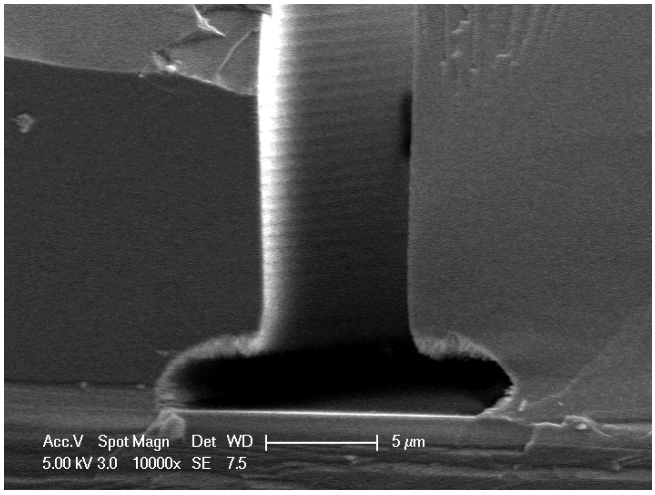



Figure 3.52. Cross-section SEM image showing the serious notching problem at the bottom of the trench: notching happens where the 30 μm thick silicon device layer interfaces with the 1 μm thick buried oxide layer.

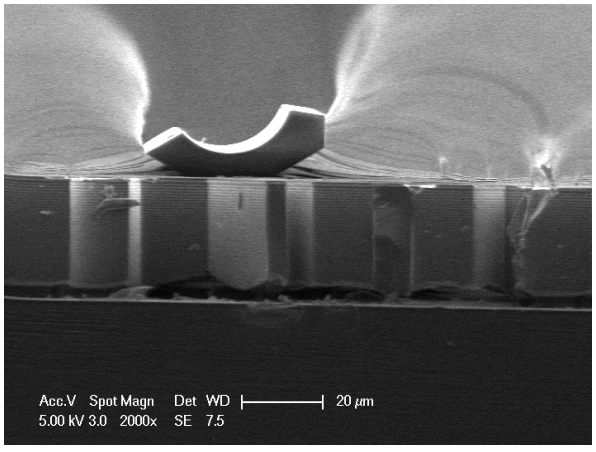



Figure 3.53. Cross-section SEM image showing that a part of the small-size pixel is completely undercut and released from the wafer.

The solution is to use a low bias frequency to dissipate the charges, which reduces the notching on the SOI wafers from the STS2 system in SNF. A vertical and smooth sidewall is achieved with no visible notching at the bottom of the trenches, as shown in Figure 3.54 and the close-up image in Figure 3.55.

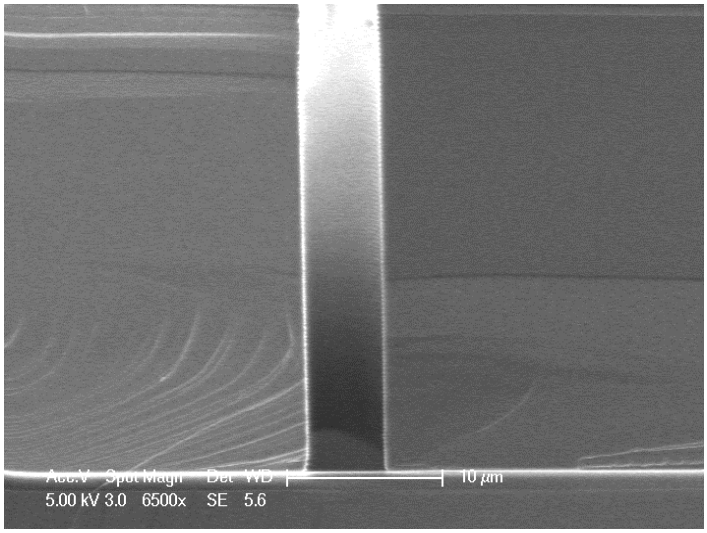



Figure 3.54. Cross-section SEM image of the trench profile after the DRIE etch, showing vertical and smooth sidewalls with no visible notching at the silicon/buried oxide interface.

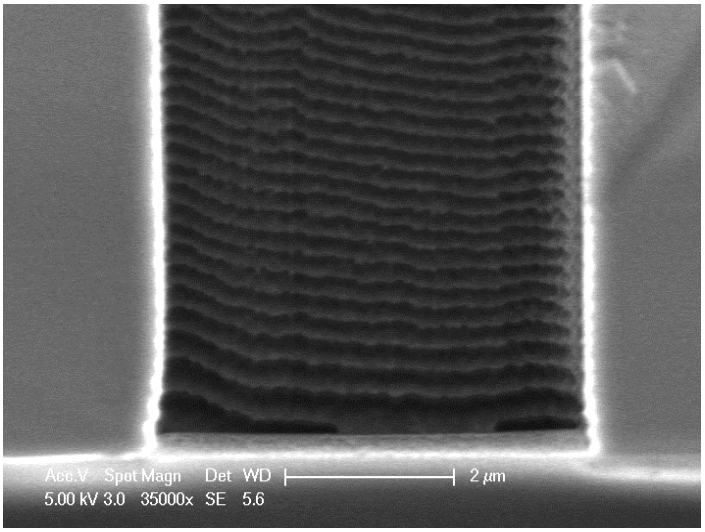
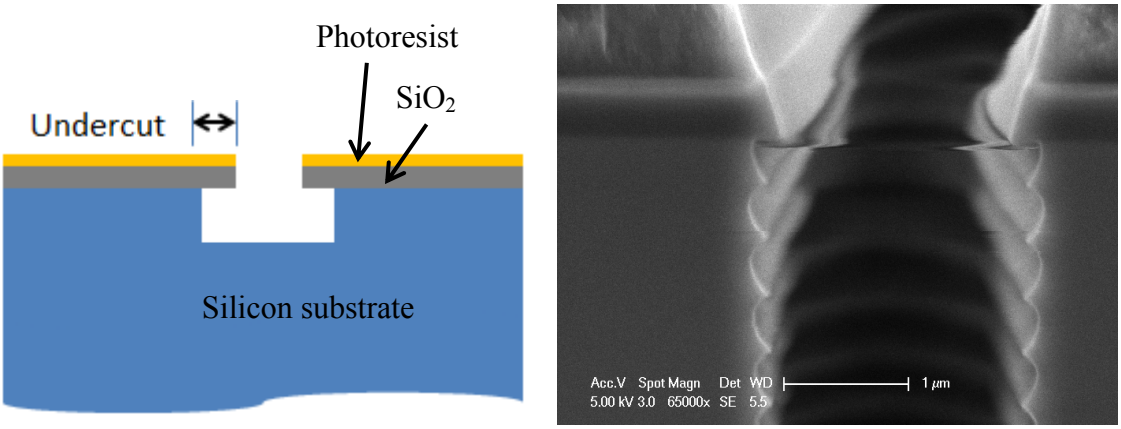



Figure 3.55. Close-up of the cross-section SEM image of the trench profile after the DRIE etch, showing vertical and smooth sidewalls with no visible notching at the silicon/buried oxide interface.

#### 3.3.3.7.2 Undercut

The undercut problem also comes from the DRIE etch process. In the advanced silicon etch process discussed in section 3.3.3.3, the deposition of the passivation layer and etching alternate. Lateral etching will occur if the polymeric layer on the sidewall is removed before each etching cycle finishes. A constant lateral etching was observed as the etching time goes in the advanced silicon etching process [97]. The formation of the undercut at the mask/silicon interface is probably caused by two mechanisms: the diffusion of active species on the mask surface and the isotropic chemical etching of silicon by fluorine radicals mainly at the beginning of the process [98]. Figure 3.56 shows a schematic diagram and SEM image of the undercut at the mask/silicon interface. The size of the undercut is smaller for a shorter etch-phase time. The size of undercut can be minimized by adjusting various etch parameters such as the ratio of the etch-phase time to the passivation-phase time, the etch gas flow and deposition gas flow, the platen temperature, the pressure, etc.




(a) (b)

Figure 3.56. Schematic diagram of the undercut at the mask/silicon interface (a) and cross-section SEM of the trench region showing the undercut at the mask/silicon interface with an etch-phase time of 2.5 seconds (b).

The undercut issue has a large effect if the feature is very narrow. During the etching of the narrow bridges in the first batch of devices with a designed bridge width of 1 μm, the undercut from each side of the bridge reduced the size of the bridge to as thin as 0.1 μm after the DRIE trench etch. Therefore, when designing the mask, the bridge widths on the mask need to be slightly larger to compensate for the undercut during the

DRIE process. Another thing to notice is that the 5 µm wide trenches in the second batch of devices will expand after the DRIE isolation trench etch, to a width of about 6 µm due to the undercut, which will affect the subsequent process parameters, such as the deposition thickness of the polysilicon.

#### 3.3.3.7.3 ProTEK cleaning

We used a 7 µm-thick ProTEK layer to protect the front surface of the wafer during the TMAH etching and used acetone to remove the ProTEK after finishing the etching process. However, the ProTEK layer cannot be completely removed by a prolonged time in acetone on some of the devices in the second batch, which results in a low yield. Figure 3.57 shows the ProTEK residue on an array with large-size pixels after soaking in acetone for 48 hours.

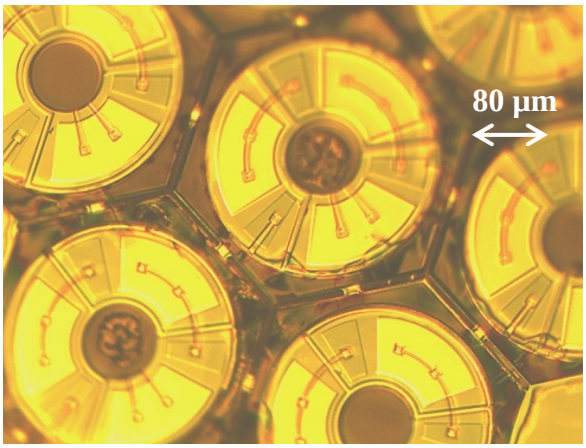
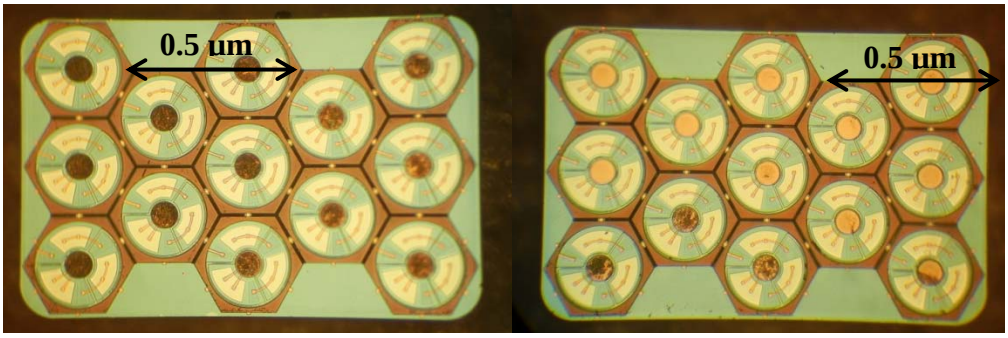



Figure 3.57. Optical microscope image of an array with large-size pixels, showing ProTEK residue on the wafer surface after soaking in acetone for 48 hours.

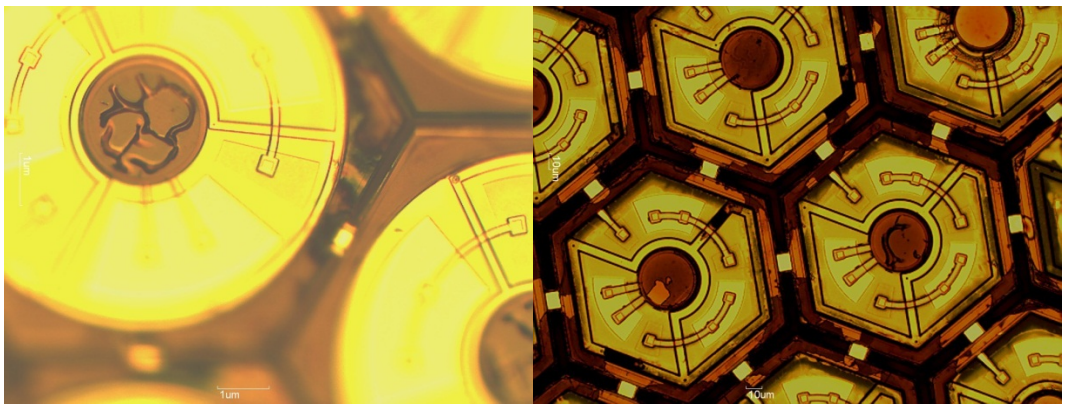
The above mentioned issue is more serious on the large-size pixels. The possible reason for this issue is that the ProTEK adheres strongly on the pixels with large surface areas so that it is hard to be removed from the surface. A recommended two-bath removal process of ProTEK from Brewerscience Inc. is to use Air Products' ACT XT-1100 at room temperature for 30 minutes and then Air Products' ACT 412 at 80 ˚C for 30 minutes, followed by isopropanol cleaning and water rinse. However, this cleaning solution is strong, and the iridium oxide layer might be damaged by the cleaning process. Figure 3.58 shows that after cleaning with the recommended ACT 412 cleaning solution, the central electrode of the large pixel array changes color, which suggests that the

iridium oxide on the central active electrode is probably damaged. A better cleaning solution is needed to improve the yield of the fabrication process. Another possible reason is that the iridium oxide does not adhere well enough to the wafer surface after the various process steps, as shown in Figure 3.59.




(a) (b)

Figure 3.58. A large-pixel array is cleaned by the recommended cleaning bath from Air Products' ACT 412: (a) before cleaning, showing ProTEK residue on the array surface; (b) after cleaning, showing that the central active electrode changes color.



(a) (b)

Figure 3.59. Iridium oxide peeling off of the central and return electrodes: (a) a large-size pixel showing that the iridium oxide peels off at the central electrode; (b) medium-size pixels, showing that the iridium oxide peels off in the central active and return electrode regions.

Figure 3.59 shows that for some of the second batch of devices, the iridium oxide does not adhere well and peels off at the central and return electrodes. The poor adhesion of iridium oxide may come from the iridium oxide deposition process or the subsequent ProTEK high temperature (205 ˚C) bake processes, which may damage the

iridium oxide. Therefore, any improvement in the iridium oxide deposition process or process optimization for a more compatible fabrication process will improve the yield.

#### 3.3.3.7.4 Polysilicon filling problem

For the second batch of devices, during the LPCVD polysilicon deposition process discussed in section 3.3.3.3, we chose to deposit 4 µm thick polysilicon to fill the trenches. However, during the deposition process, the deposition rate is not quite uniform across the wafers and also varies from wafer to wafer. In addition, at the corners of the pixels and edges of the arrays, there are wider trench regions so that the filling of 4 µm polysilicon is not enough to completely fill the trenches, leaving a small pit on the surface as shown in Figure 3.60. The incomplete filling of the wider area of the trenches may cause non-uniform photoresist coverage near the area of the unfilled regions. In our fabrication process, the unfilled regions on some of the wafers in the second batch cause poor photoresist coverage so that during the $XeF_2$ etching of polysilicon in selected regions (trench regions between pixels), some of the polysilicon in the regions inside each pixel is also etched, causing instability in the structure. This is because without the polysilicon connecting the isolated regions to the whole array, these regions may be detached from the whole array. Another issue is broken metal tracks after the $XeF_2$ etching of the polysilicon in trenches inside each pixel due to the poor photoresist coverage in the trenches if not completely filled by polysilicon, as shown in Figure 3.61. To solve this problem, a thicker photoresist layer during photolithography was used in Batch 2 and a thicker layer of the polysilicon deposition is needed in future fabrications.

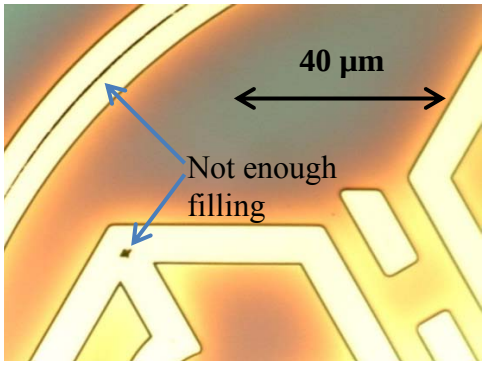



Figure 3.60. Wafer surface after the CMP process, showing pits on some of the wider trenches, indicating incomplete filling of the trenches by the polysilicon deposition process.

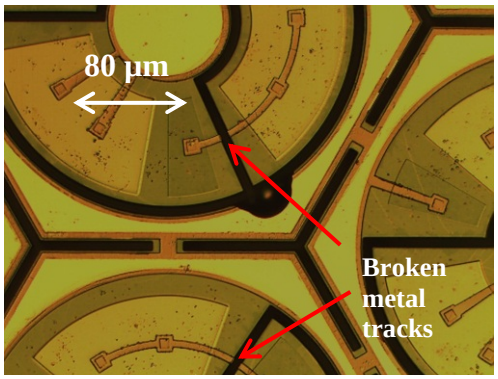
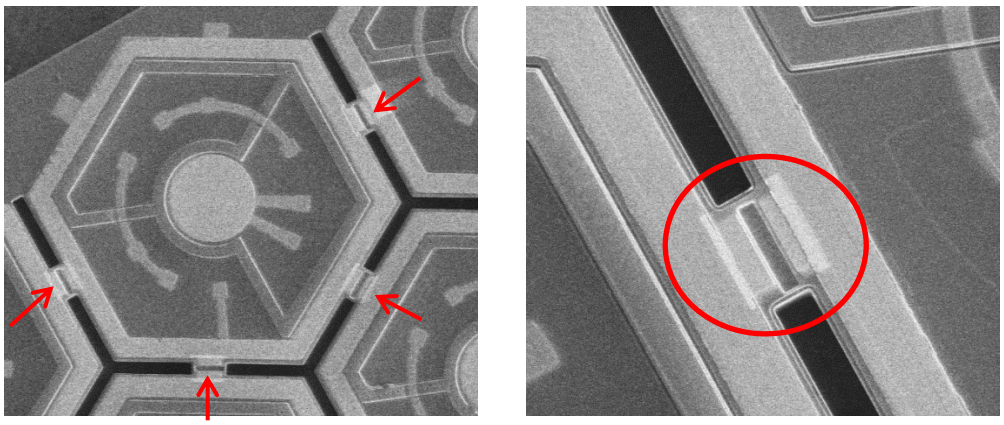



Figure 3.61. Optical microscope image of the large-size pixel showing that the polysilicon inside the pixel is also removed during the $XeF_2$ etching, causing broken metal tracks.

## 3.4 Focused ion beam etching

The local return electrodes in the fabricated photodiode arrays are connected by metal lines (10 µm width for large and medium size pixels, 6 µm width for small size pixels) to form a common ground. We observed current spreading which can be seen in Figure 4.23 of Chapter 4. Ion-beam etching was performed to study the properties of the pixels with the surrounding metal bridges disconnected from the common ground. The main ion used for the ion beam etching is $Ga^+$. Several pixels are disconnected from the common grounds by etching the metal bridges between the pixels in the medium-size pixel array. Figure 3.62 shows the SEM image of a medium-size pixel with the surrounding metal bridges cut by the ion beam.



(a) (b)

Figure 3.62. SEM image of medium-size pixels after ion-beam etching: (a) a medium-size pixel with the 4 surrounding metal bridges cut. (b) Close-up of the cut region.

# 4. DEVICE CHARACTERIZATION

*Portions of Chapter 4 reprinted, with permission, from Lele Wang, K. Mathieson, T. I. Kamins, J. D. Loudin, L. Galambos, G. Goetz, A. Sher, Y. Mandel, P. Huie, D. Lavinsky, J. S. Harris and D. V. Palanker, "Photovoltaic retinal prosthesis: implant fabrication and performance", accepted by Journal of Neural Engineering, ©2012 IOP.*

After finishing the fabrication and releasing the devices, the structural, electrical and optical device properties are characterized, which will be discussed in detail in this chapter. For structural characterization, we examine whether the structures of the pixels are well formed, such as if the metal lines are connected as designed; the metal electrode looks uniform and is not peeling; the antireflection coating is uniform and has no visible damage; the backside oxide is uniform and is not damaged by the TMAH etching, etc. For electrical characterization, we are focusing on dark current in the reverse bias region, which should be low enough to allow detecting a weak light signal. We also focus on the current-voltage characteristic for the three photodiodes connected in series, which indicates the efficiency of isolation between photodiodes. For optical characterization, photo-responsivity is the key factor that measures the efficiency of light to current conversion. Measurements of the photodiode arrays in conductive solution show that the electrical current generated by the photodiode array is converted to an ionic current in the conductive solution.

## 4.1 Structural and material characterization

We characterize the structural properties of the fabricated devices by optical microscopy and scanning electron microscopy. Figure 4.1 shows optical microscope images of large-, medium- and small-size pixels before the final release step. In the central and return electrode regions, the iridium oxide shows as a dark color. The metal tracks that connect the three photodiodes in series are Pt, which shows as a light color. The trenches inside each pixel are filled with polysilicon, while the trenches between pixels are cleared of polysilicon. Silicon bridges connect the pixels together to form an array. Edge-to-edge distance is 280 μm for large pixels, 140 μm for medium pixels and 70 μm for small pixels.

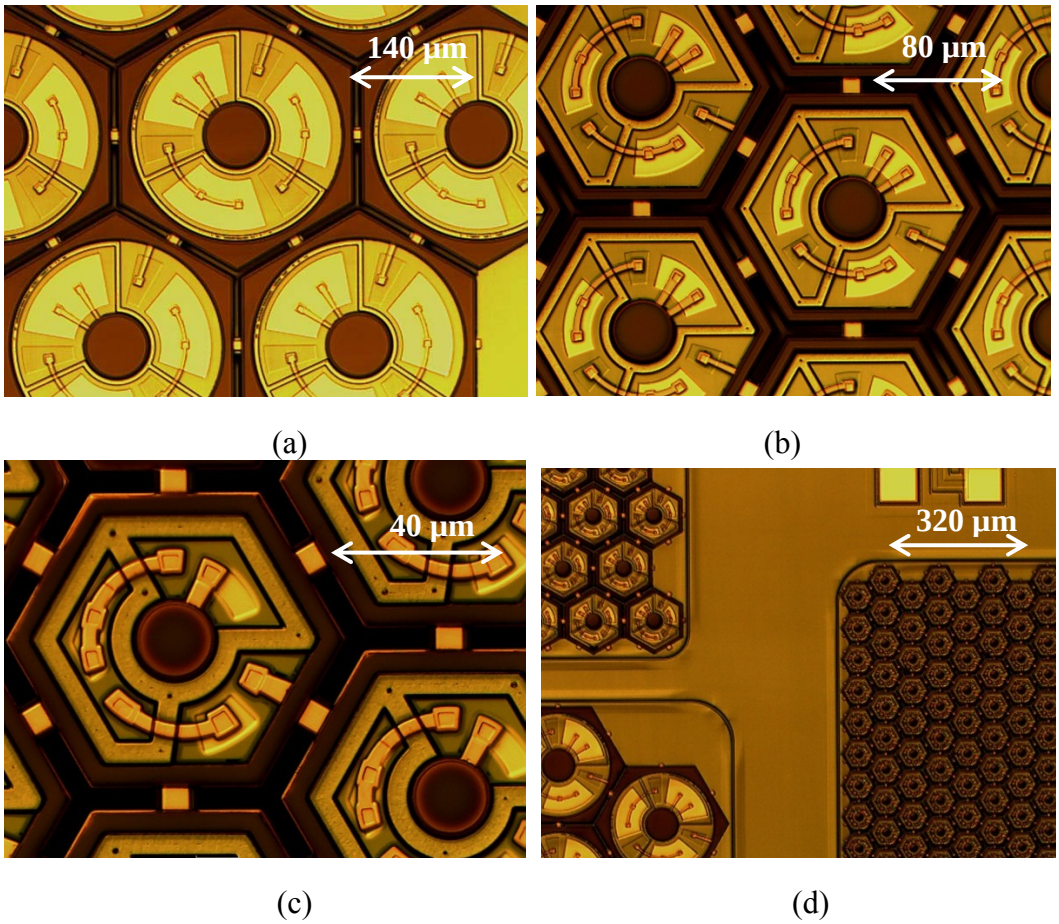
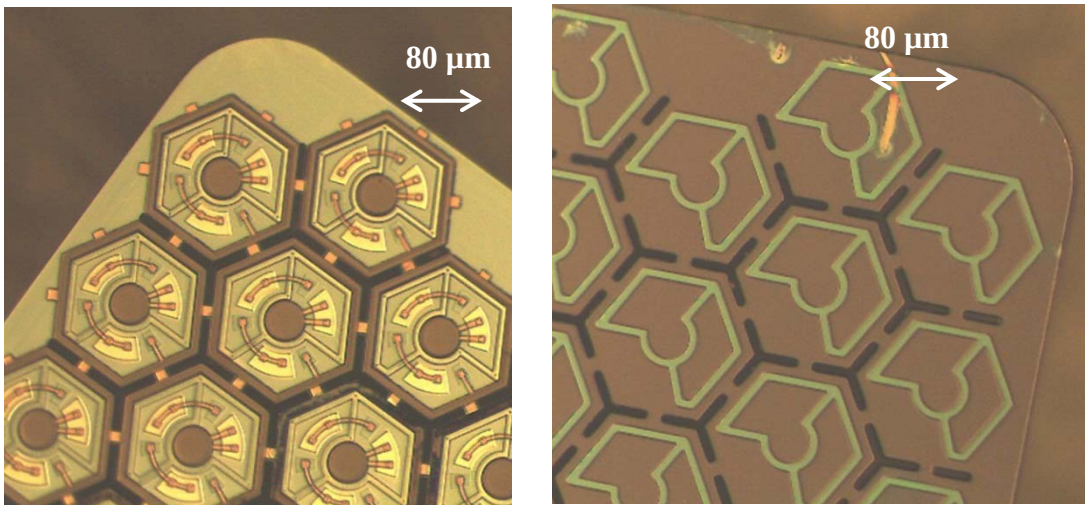



Figure 4.1. Optical microscope image of the arrays before release: (a) array with large-size pixels; (b) array with medium-size pixels; (c) array with small-size pixels; (d) comparison of all three size pixels on wafer.

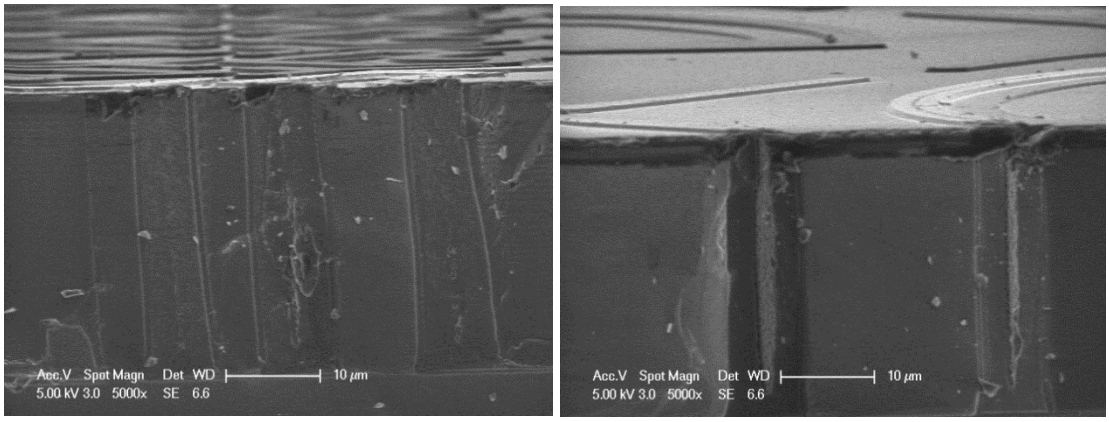
Figure 4.2 shows a medium-size pixel array after being released. From Figure 4.2(a), we can see that for a good array, the color of the field regions is uniform. The color of the central and return iridium oxide electrode should be similar to that of the arrays before release, as shown in Figure 4.1. The trenches inside each pixel should be filled with polysilicon. Partially etched trenches inside pixels may lead to a defective array, affecting the performance of the array. This is because the polysilicon inside each pixel provides mechanical support of the three photodiodes as well as support for the metal tracks that connect the three photodiodes in series. If the polysilicon inside the pixels is etched during the $XeF_2$ etching step due to insufficient photoresist coverage as discussed in section 3.3.3.7.4, some regions of the pixel may be detached and the metal tracks may be broken, affecting the normal function of the pixel. From the backside image of the array in Figure 4.2(b), it is clear that the trenches between pixels are etched through and the

trenches inside each pixel are filled, which can be seen by the color difference between them. Optical measurement as discussed in section 3.1.3.5 is used to measure the backside oxide thickness after release. The backside of the array has 100 nm of thermal oxide and appears as a color of purple.




(a) (b)

Figure 4.2. Optical microscope image of the medium-size pixel array after release: (a) front side; (b) back side.




(a) (b)

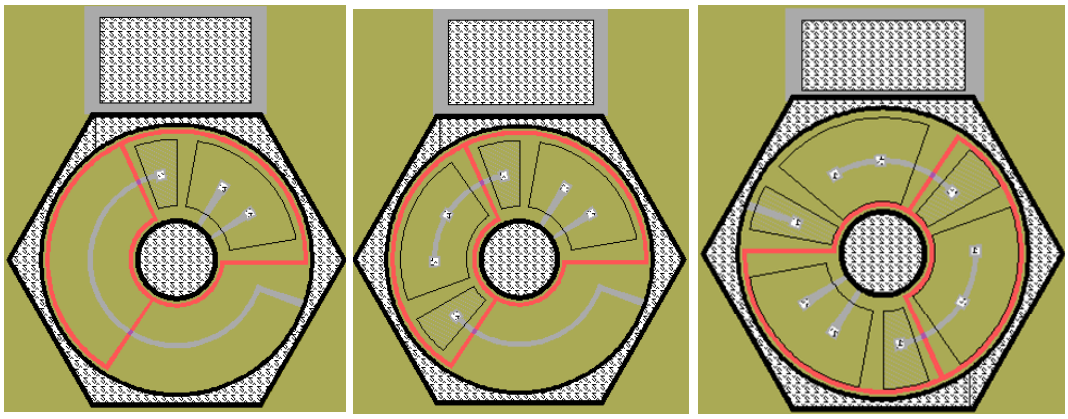
Figure 4.3. Cross-section SEM of the trenches: (a) trench filled by polysilicon; (b) a visible seam in the middle of the filled trenches.

In addition to optical microscope images, SEM images also show us important structural information. Figure 4.3 is a cross-sectional SEM showing the trench area. We can see that the trenches are 5 µm wide, 30 µm deep with a 500 nm oxide layer on the trench sidewalls. Polysilicon fills the trenches and planarizes the wafer, with polysilicon

steps on the surface of the wafer (Figure 4.3(a)). Figure 4.3(b) shows that for some trenches, there is a visible seam in the middle of the trench after polysilicon refilling, which is probably due to the marginal polysilicon deposition thickness.

## 4.2 Electrical characterization

The electrical characteristics of all three size pixels (280 μm, 140 μm, and 70 μm) were tested by a semiconductor parameter analyser (HP 4155A, Hewlett Packard Co., Palo Alto, CA). The reverse-bias dark current was in the range of 1-100 pA. The reverse breakdown voltage was larger than 20 V. We measured the one-diode, two-series-connected-diode and three-series-connected-diode test structures as shown in Figure 4.4. Figure 4.5 shows the typical current-voltage characteristic of the one-diode, two-series-connected-diode and three-series-connected-diode structures of large-size pixels with the turn-on voltages of 0.6 V, 1.2 V and 1.8 V, respectively. This means the three photodiodes are well isolated from each other and are connected in series in each pixel.



(a) (b) (c)

Figure 4.4. One-diode, two-series-connected-diode and three-series-connected-diode test structures for electrical measurement: (a) one-diode test structure; (b) two-series-connected-diode test structure; (c) three-series-connected-diode test structure.

Figure 4.6 shows the semi-log scale I-V characteristic of a large-size pixel with 3 diodes connected in series. The measured dark current was in the range of 1-10 pA, which is adequate for our application. Figure 4.7 and Figure 4.8 show the linear scale I-V characteristics of the one-diode, two-series-connected-diode and three-series-connected-diode structures of medium-size pixels and the semi-log scale I-V characteristic of a

medium-size pixel with 3 photodiodes connected in series. Generally, the current-voltage characteristics are similar for different pixel sizes.

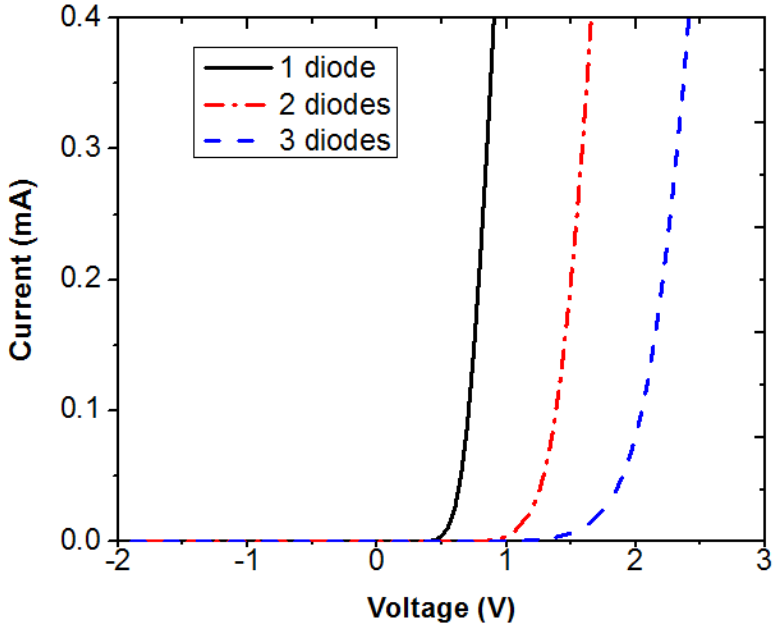



Figure 4.5. Linear scale I-V characteristic showing the turn-on voltages of the one-diode, two-series-connected-diode and three-series-connected-diode structures (large-size pixels) are 0.6 V, 1.2 V and 1.8 V, respectively.

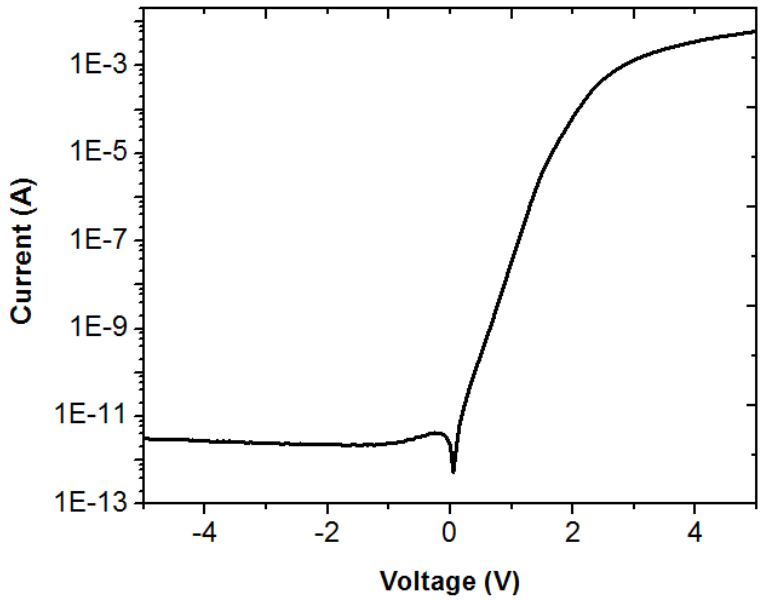



Figure 4.6. Semi-log scale I-V characteristic of a large-size pixel with 3 diodes connected in series, showing the measured dark current in the range of 1-10 pA.

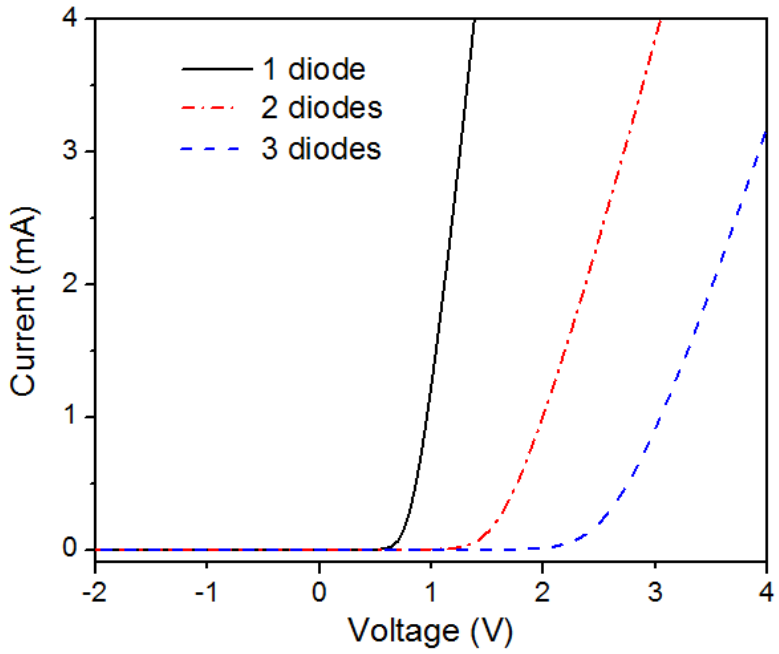



Figure 4.7. Linear scale I-V characteristic of the one-diode, two-series-connected-diode and three-series-connected-diode structures (medium-size pixels).

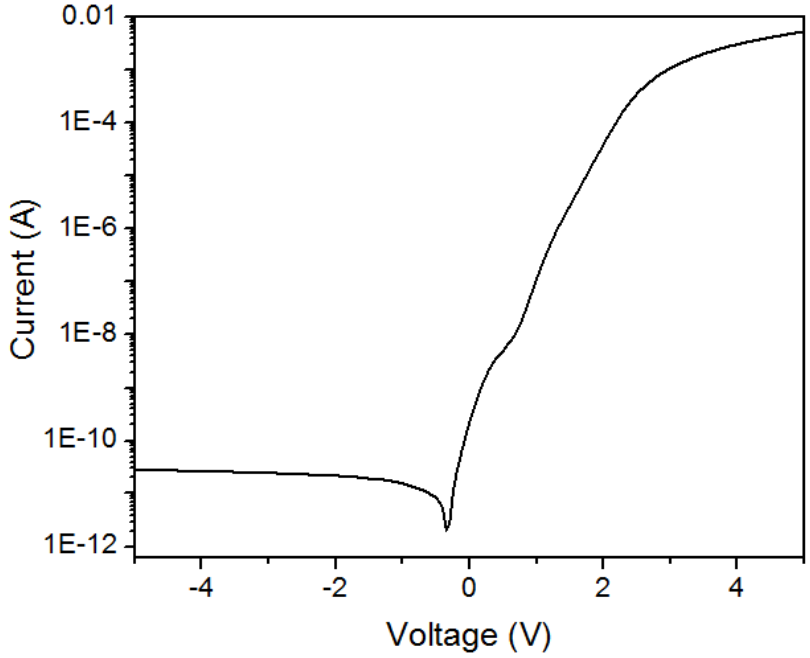



Figure 4.8. Semi-log scale I-V characteristic of a medium-size pixel with 3 diodes connected in series.

In the forward bias region, the ideal diode equation can be simplified to be:

$$I = I_S e^{\frac{qV_A}{nkT}} \qquad \text{(Eq. 4. 1)}$$

For three diodes connected in series, if we assume the total applied voltage is equally split into the three diodes (based on the similar area of the three diodes), the bias voltage for each diode is $V_A/3$ and the current I is the same for each of the diodes. Therefore, the expected q/nkT factor will decrease accordingly by a factor of 2 or 3 for the two- and three-series-connected diode configurations. Figure 4.9 shows the fitted current-voltage curves for the one-diode, two-series-connected-diode and three-series-connected-diode structures of large-size pixels. By fitting the current-voltage characteristic in the forward bias region, the reverse bias saturation current $I_s$ and the ideality factor n can be determined. The fitted results are summarized in Table 4.1, showing that n is 1.48, 2.87 and 5.09 for the one-diode, two-series-connected-diode and three-series-connected-diode structures, respectively, which scales approximately with the number of diodes.

| Number of diodes | $I_s$ (A) | q/nkT ($V^{-1}$) | n |
|---|---|---|---|
| 1 diode | 1.17E-11 | 26.1 | 1.48 |
| 2 diodes | 7.89E-12 | 13.5 | 2.87 |
| 3 diodes | 1.5E-10 | 7.6 | 5.09 |

Table 4.1. Summary of the fitted results for the forward current-voltage characteristic of the one-diode, two-series-connected-diode and three-series-connected-diode structures of large-size pixels.

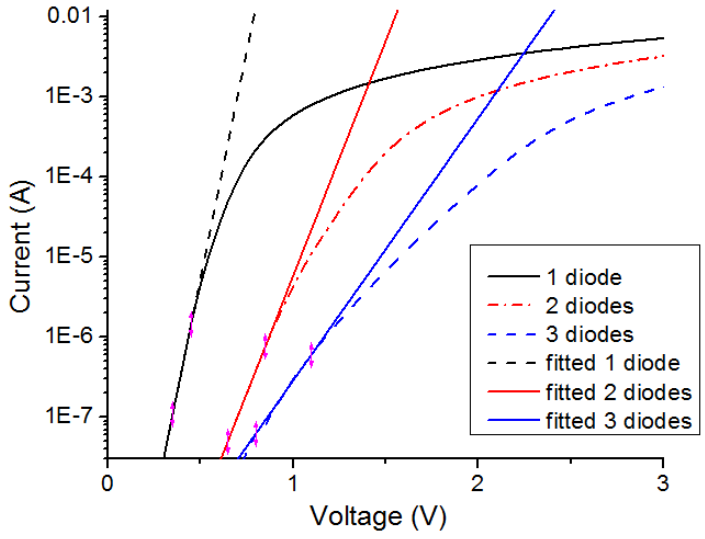



Figure 4.9. Fitted curves of the forward I-V characteristics of the one-diode, two-series-connected-diode and three-series-connected-diode structures of large-size pixels.

We also measured the test structures to test the sheet resistance and contact resistance of the $n^+$ regions and $p^+$ regions. As shown in Figure 4.10, by probing V1 and V3, sheet resistance can be determined (16.5 squares in the $n^+$ or $p^+$ regions). By probing V2 and V3, with the information of the sheet resistance measured in the previous step, contact resistance can be estimated. For contact resistance, a test structure with a via size of 2.3 μm × 2.3 μm is measured. The measurement results are summarized in Table 4.2 and plotted in Figure 4.11.

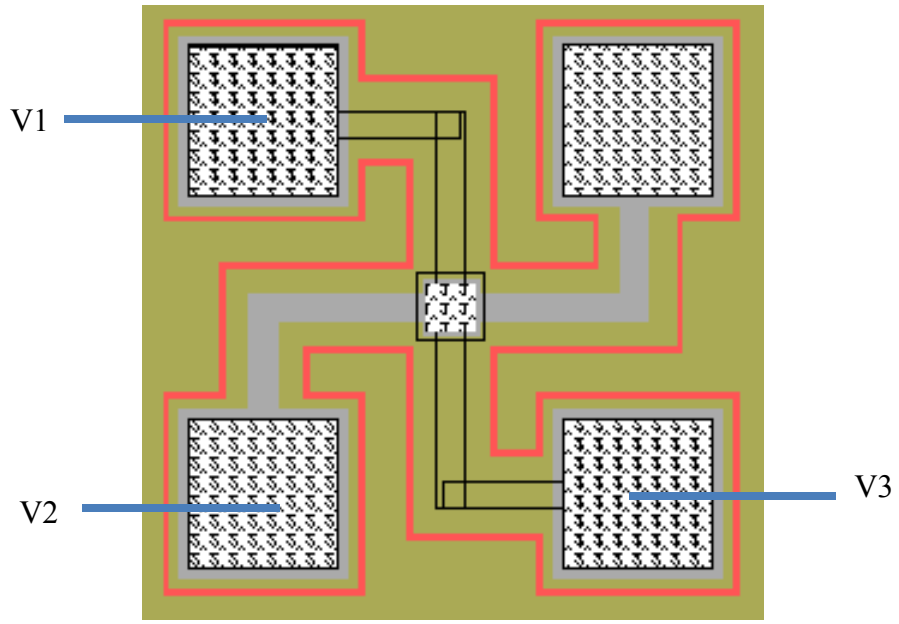



Figure 4.10. Test structures for sheet resistance measurement and contact resistance measurement.

We can see from Table 4.2 that the sheet resistances and contact resistances are close to the simulation results as shown in section 3.2. The sheet resistance and contact resistance for the $p^+$ region are relatively high, which can be improved by using a larger dose in the boron ion implantation process.

| Regions | Sheet resistance (Ω/sq) | Contact resistance (Ω) |
|---|---|---|
| $n^+$ | 2.96 | 15.86 |
| $p^+$ | 106.5 | 67.43 |

Table 4.2. Sheet resistance and contact resistance summary of the $n^+$ and $p^+$ regions.

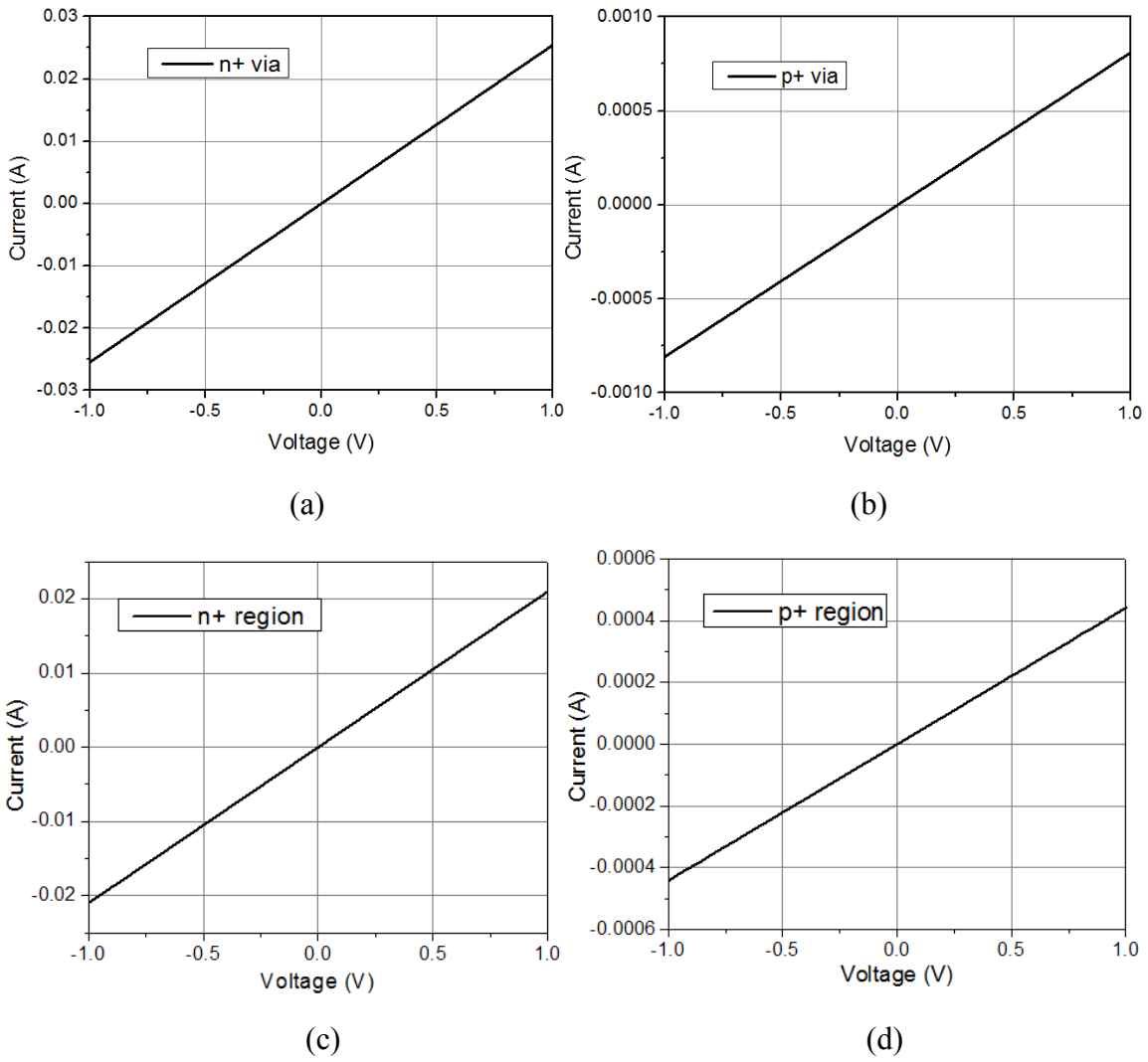



Figure 4.11. Measurement results for the test structures of the $n^+$ and $p^+$ regions: (a) V2-V3 measurement of $n^+$ region. (b) V2-V3 measurement of $p^+$ region. (c) V1-V3 measurement of $n^+$ region. (d) V1-V3 measurement of $p^+$ region. (V1, V2, V3 are as defined in Figure 4.10.)

## 4.3 Optical characterization

The schematic diagram of the optical measurement setup is shown in Figure 4.12. Laser light at 880 nm is collimated by an optical system, which is then partly directed to a light detector and partly directed into the microscope to illuminate the photodiode pixel. The light intensity was uniform across the illuminated area. A 1 kΩ resistor is connected in series with the device under test. The external voltage is swept from -4.5 volts to 1 volt across the device and the resistor, as shown in the figure. The sweeping voltage ($V_1$), the voltage across the 1 kΩ resistor ($V_2$) and the light power measured by the light detector

are recorded. The incident light power at the plane of the device is also measured to calibrate the value recorded by the system during the measurements. The recorded data are imported into MATLAB and analyzed. Figure 4.13 to Figure 4.18 are the measurement results of the photocurrent and photo-responsivity for the large-, medium- and small-size pixels. The light irradiance was varied from 0 mW/mm$^2$ to 35.4 mW/mm$^2$ on the one-diode test structure in our measurements. From the photo-responsivity plots of the different sizes of pixels, we can see that the photocurrent scales linearly with light intensity. The measured photocurrent was 23 μA, 5 μA, and 0.7 μA for large-, medium- and small-size pixels at 6 mW/mm$^2$ light intensity, scaling approximately with the exposed silicon area. The measured photo-responsivities at 880 nm wavelength for the measured large-, medium- and small-size pixels were about 0.33 A/W, 0.40 A/W and 0.40 A/W per diode at zero voltage bias.

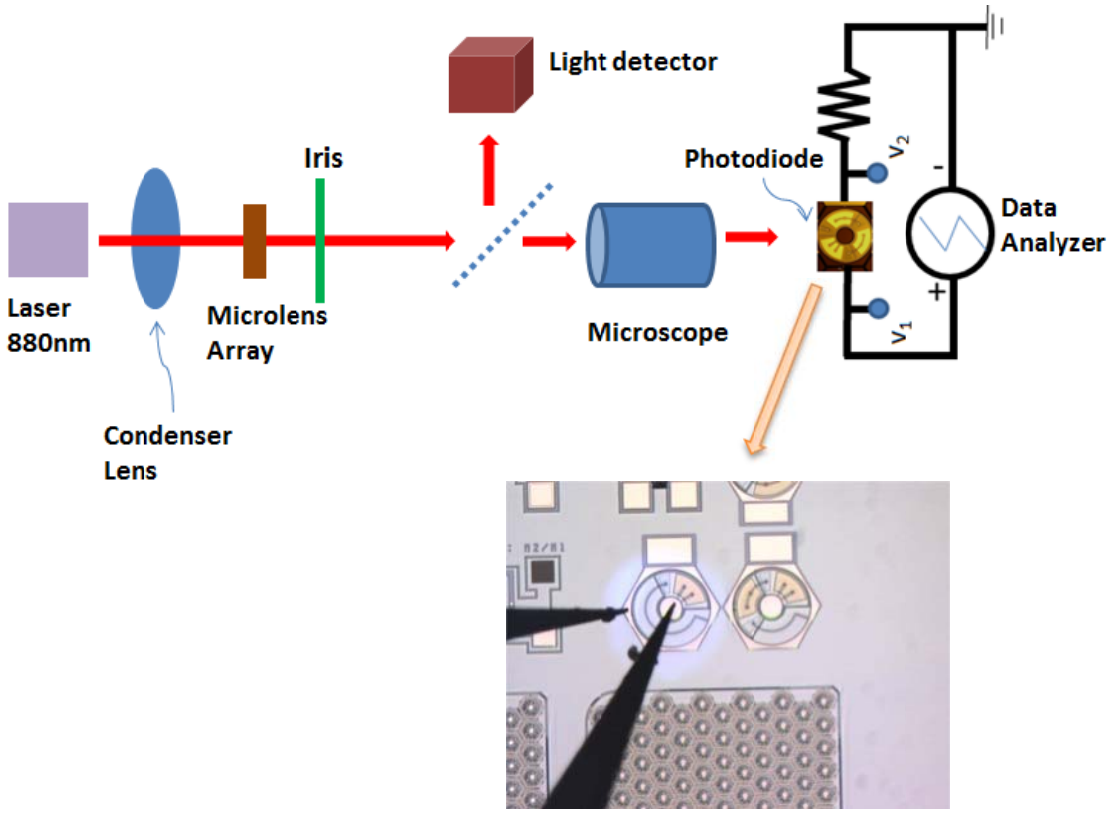



Figure 4.12. Schematic diagram of the system used for optical measurements.

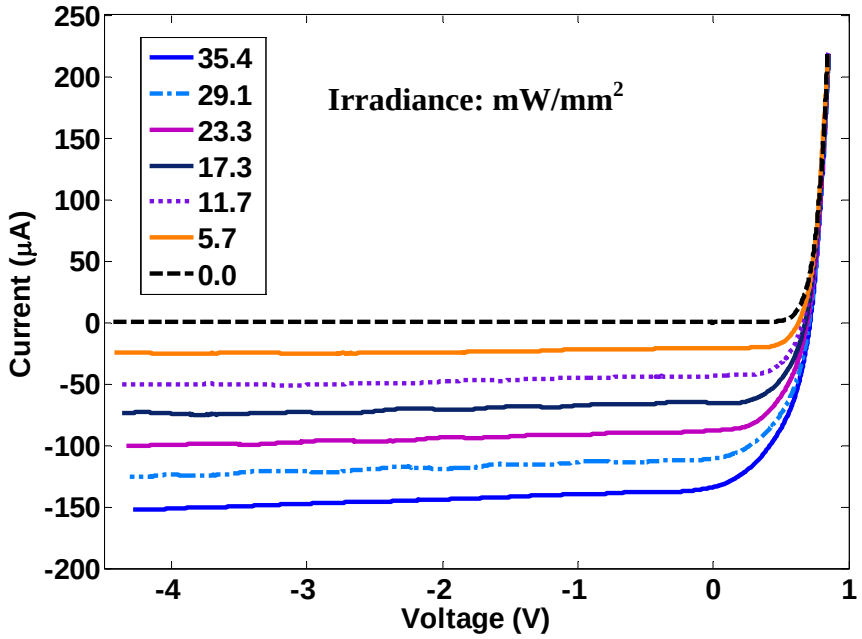



Figure 4.13. Photocurrent as a function of bias voltage at various light intensities for a large-size single-diode test structure.

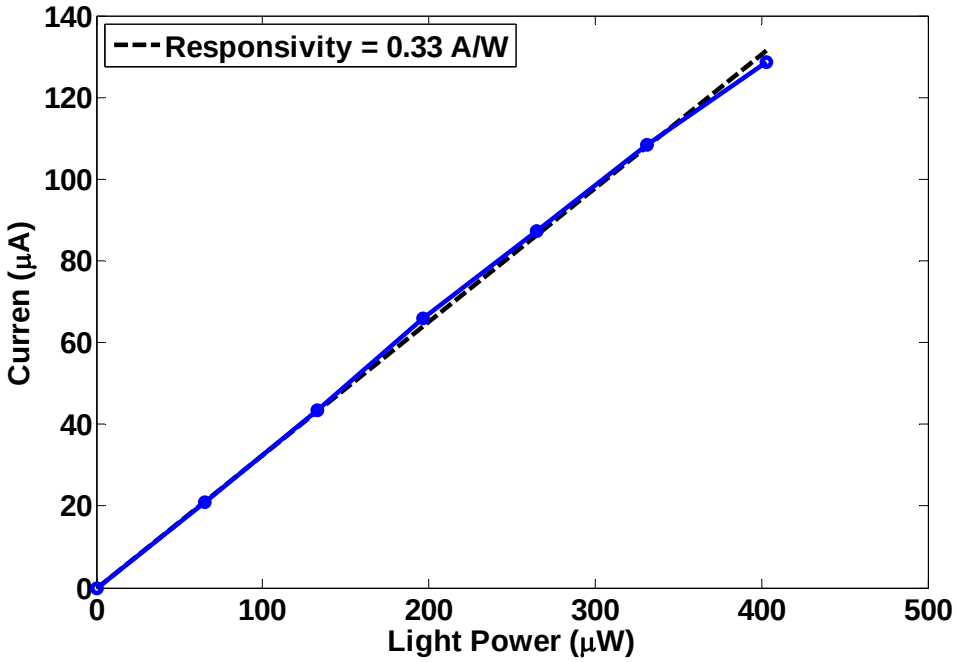



Figure 4.14. Photo-responsivity at zero voltage bias for a large-size single-diode test structure.

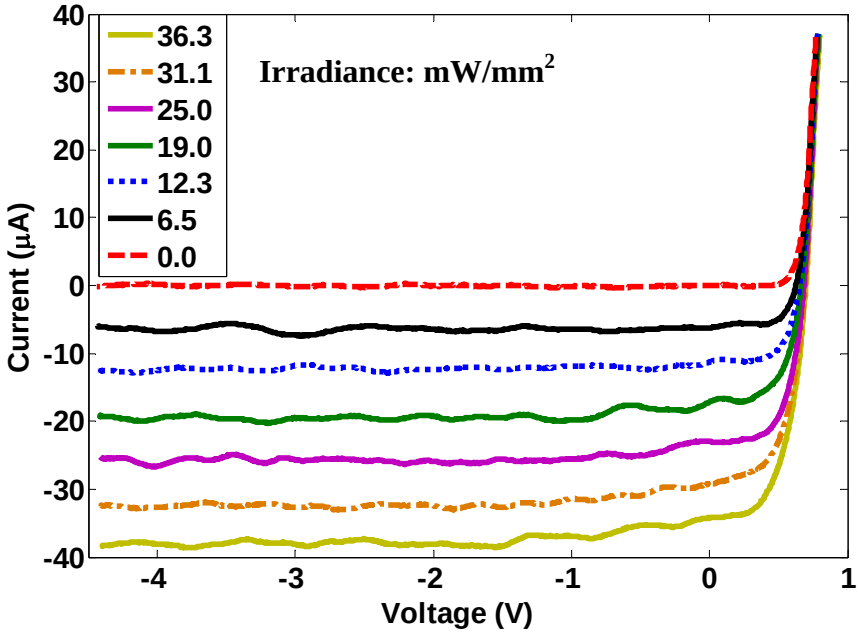



Figure 4.15. Photocurrent as a function of bias voltage at various light intensities for a medium-size single-diode test structure.

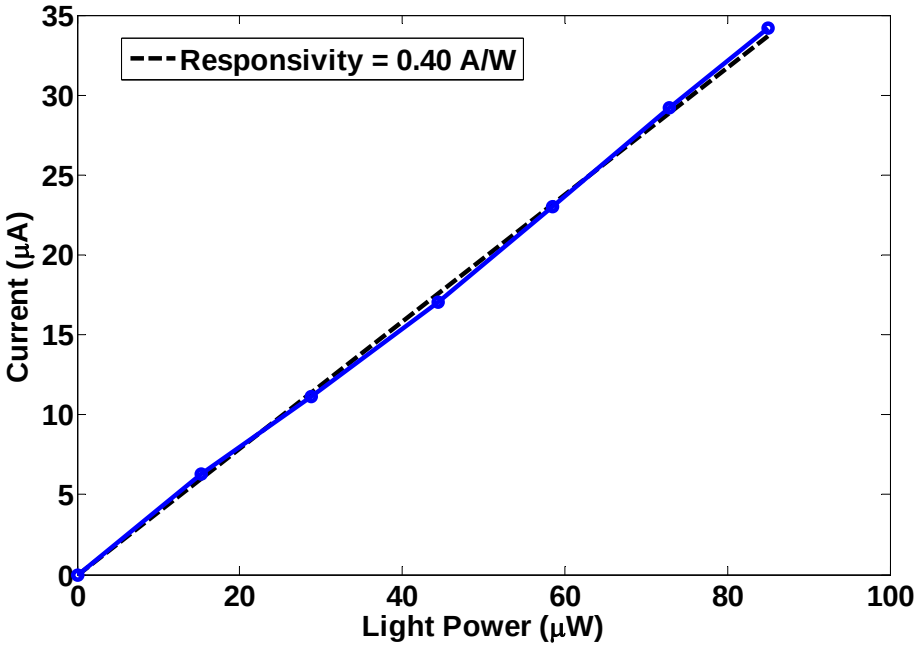



Figure 4.16. Photo-responsivity at zero voltage bias for a medium-size single-diode test structure.

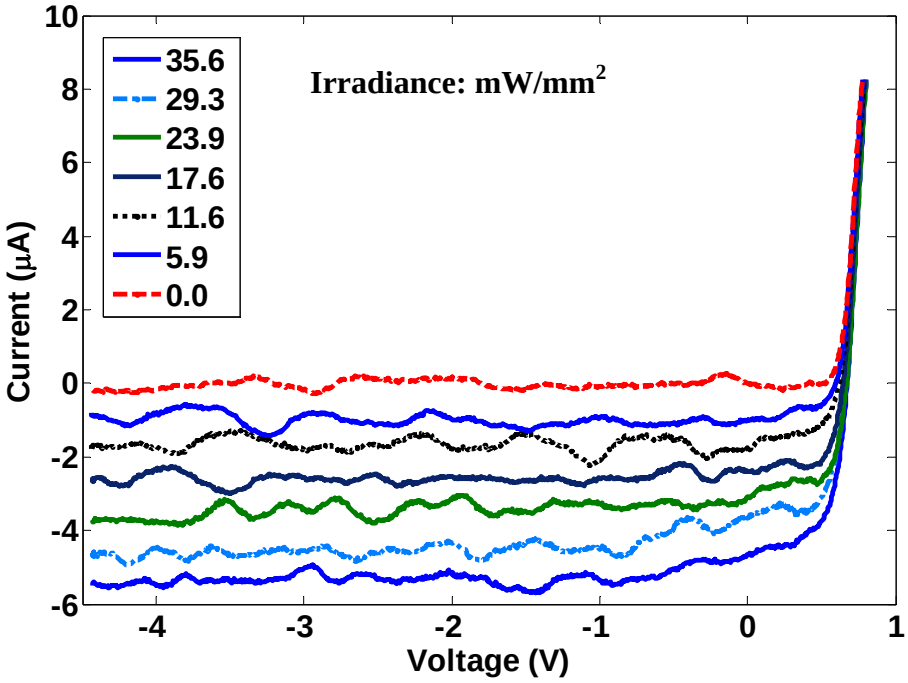



Figure 4.17. Photocurrent as a function of bias voltage at various light intensities for a small-size single-diode test structure.

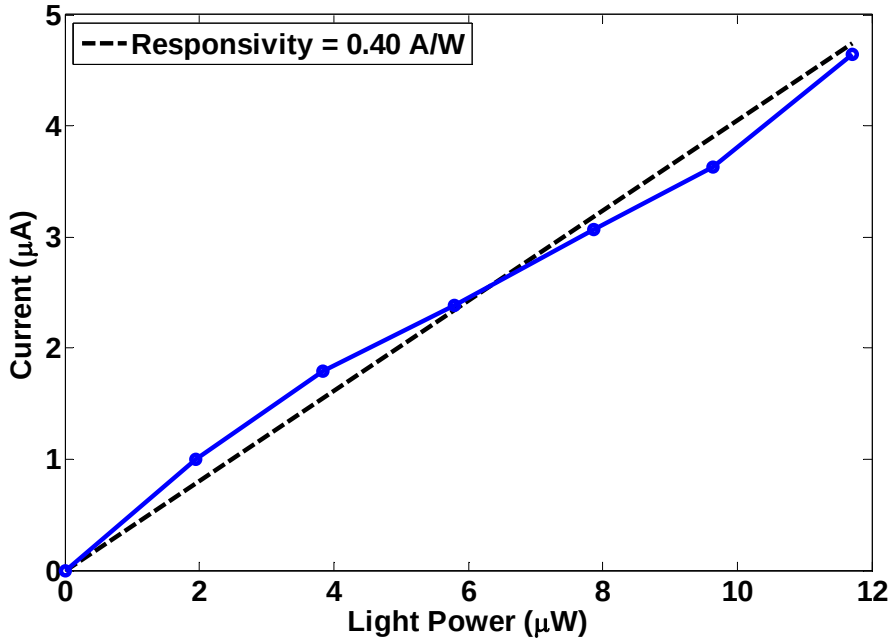



Figure 4.18. Photo-responsivity at zero voltage bias for a small-size single-diode test structure.

The measured quantum efficiency was approximately 56% for medium-size pixels. Due to sub-optimal performance of the anti-reflection coating (deviation of the grown

$SiO_2$ thickness from the target thickness), there was more reflection of the NIR light from the wafer surface than expected. The small-size pixel measurement results in Figure 4.17 show a relatively large noise from the data analyzer. This can be improved by increasing the light irradiance, which results in a larger photocurrent as shown in Figure 4.19. The photo-responsivity is shown in Figure 4.20, with a similar result of about 0.4 A/W as that of Figure 4.18.

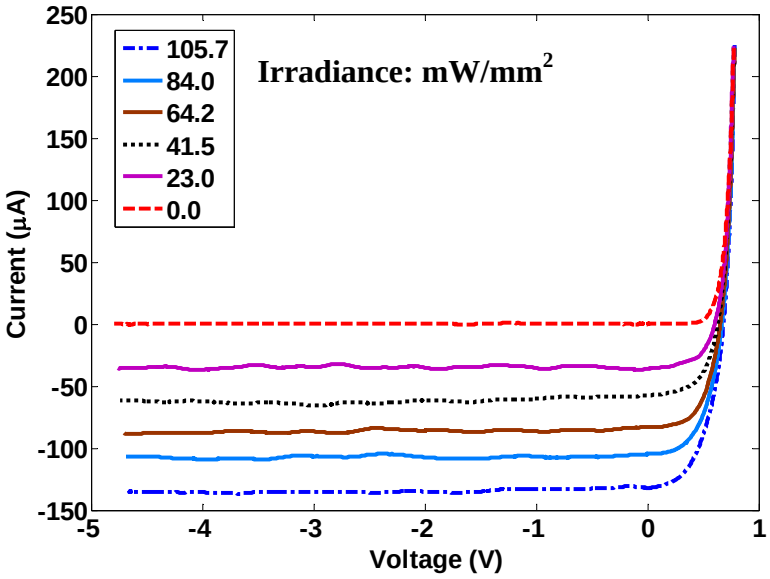



Figure 4.19. Photocurrent as a function of bias voltage at various light intensities for a small-size single-diode test structure.

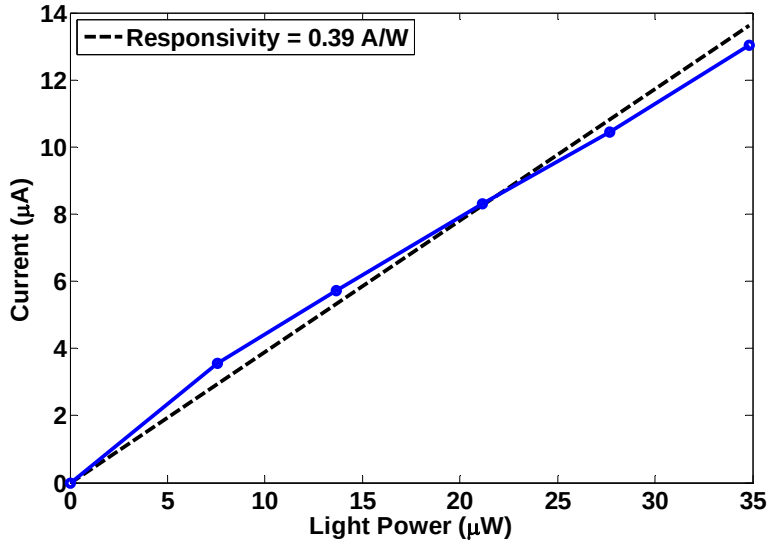



Figure 4.20. Photo-responsivity at zero voltage bias for a small-size single-diode test structure.

## 4.4 Characterization in conductive solution

Following the electrical and optical measurements, we carried out measurements of the device performance in a conductive solution. A schematic diagram of the system used for optical measurements in conductive solution is shown in Figure 4.21. First, a photodiode array is supported by a 160 µm thick filter paper and placed on top of a multi-electrode recording array [99]. The multi-electrode recording array consist 512 platinum electrodes, which are 5 µm in diameter with 60 µm spacing between them. The electrodes were fabricated on a glass substrate with an active area of 0.9 × 1.8 $mm^2$ [99, 100]. Signals were transported by indium tin oxide tracks on the glass substrate and digitized at a 20 kHz sampling rate. The sandwich of the photodiode array, filter paper and the multi-electrode recording array is placed in a chamber full of conductive solution. Infrared light at 880 nm passes through an LCD display and optical system to illuminate the photodiode array through the transparent tracks of the multi-electrode array. The photodiode array will generate current when illuminated, which converts into ion movement in the conductive solution by capacitive coupling and reversible faradaic reaction. The multi-electrode recording array records the resulting signals and channeling the data to a customized software "VISION" to make in-situ plots [99]. Data can also be streamed offline for further processing. All three sizes of pixel arrays are tested both by full field illumination and by spot illumination.

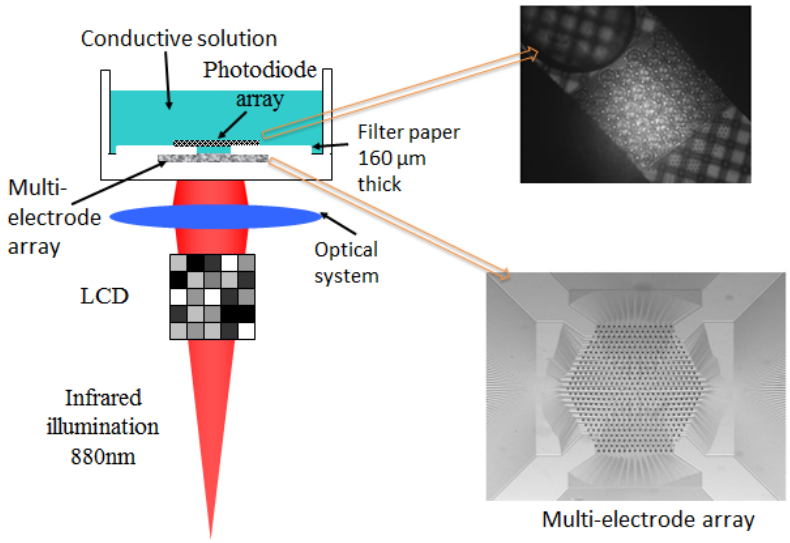



Figure 4.21. Schematic diagram of the system used for optical measurements in conductive solution.

An optical microscope image of spot illumination during measurement of the response of the medium-size pixel is shown in Figure 4.22. The spot size was chosen to be 140 µm in diameter, which corresponds to the edge-to-edge distance of a medium-size pixel. The amplitudes of the stimulation signals as a function of the distance from the illuminated medium-size pixel for different irradiances are plotted in Figure 4.23. Maps of the signal amplitude on the multi-electrode recording array with different irradiances are shown in Figure 4.24. We can see from Figure 4.23 and Figure 4.24 that as irradiance increases, the signal recorded from the multi-electrode recording array also increases. The peak signal scales approximately linearly with the irradiance. However, at large irradiance (6.4 mW/mm$^2$), there is some signal saturation due to the saturation of the amplifiers in the recording circuit, causing the non-linear scaling of the recorded signal.

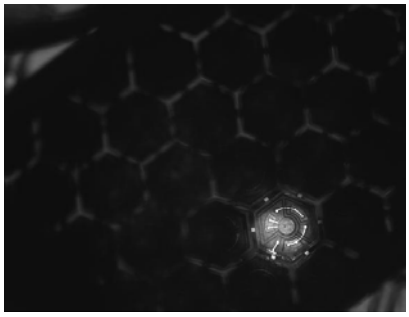


Figure 4.22. Optical microscope image of spot illumination during measurement of the response of the medium-size pixel.

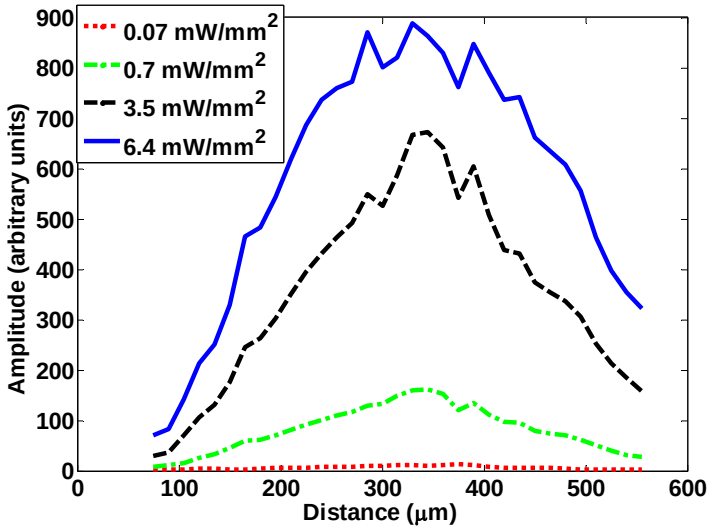



Figure 4.23. Amplitude vs. distance of the response of the medium-size pixel for different irradiances, along the black line drawn in Figure 4.24.

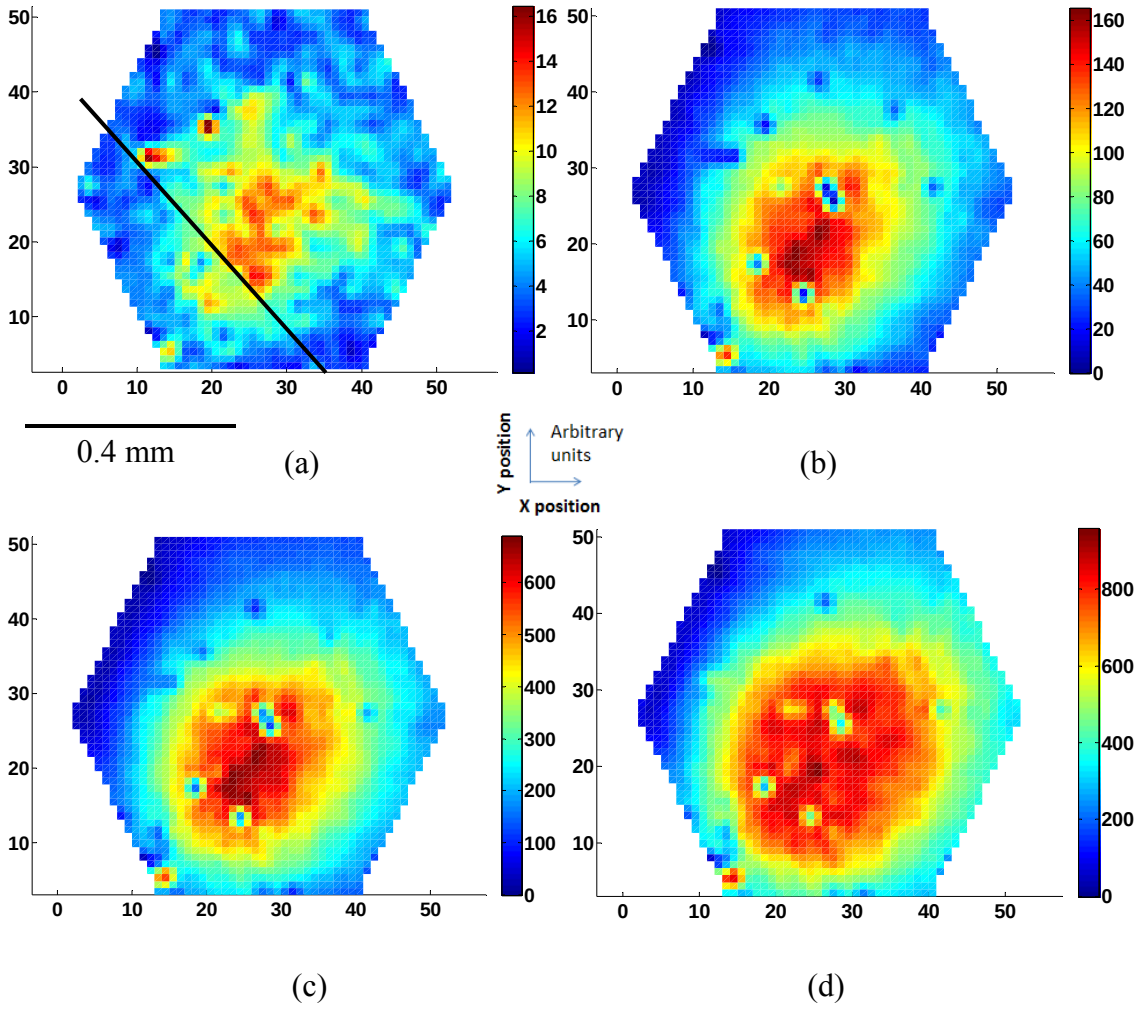



Figure 4.24. Maps of the signal amplitude on the 512-multi-electrode recording array, at the irradiances of (a) 0.07 mW/mm$^2$ (b) 0.7 mW/mm$^2$ (c) 3.5 mW/mm$^2$ (d) 6.4mW/mm$^2$.

# 5. APPLICATIONS

*Portions of Chapter 5 reprinted, with permission, from Lele Wang, K. Mathieson, T. I. Kamins, J. D. Loudin, L. Galambos, G. Goetz, A. Sher, Y. Mandel, P. Huie, D. Lavinsky, J. S. Harris and D. V. Palanker, "Photovoltaic retinal prosthesis: implant fabrication and performance", accepted by Journal of Neural Engineering, ©2012 IOP.*

After characterizing the optoelectronic performance of the fabricated photodiode arrays, we tested the device performance in-vitro (on the isolated functioning retina from a rat's body) and in-vivo (on the retina inside a rat's body). All three size pixels were shown to reliably elicit retinal responses at safe near-infrared light irradiances. The initial in-vitro and in-vivo testing will be discussed in detail in this chapter. These results are the cooperative work of a large group of people, and the dissertation author's main contributions in this part are the sorting and preparation of the implant devices for the retinal experiments after fabricating and characterizing these devices, and participation in the in-vitro study.

## 5.1 Background of in-vitro testing

### 5.1.1 Safety limit

EU safety regulations for implantable devices have a "less than 1 ˚C" safety criterion [101]. The rise of temperature around a disk-shaped heater can be calculated as a function of the dissipated power, the disk diameter and the heat conductivity of the medium [102]. To keep the temperature rise under 1˚C for a 3 mm disk in water, the average retinal irradiance should not exceed 1 mW/mm$^2$ [8]. Thermal effects of near-IR irradiation of the rabbit retina were studied previously [103], and demonstrated that the ocular blood flow in the eye helps to cool the tissue, with the cooling effect depending on temperature. The safety standards of the maximum permissible irradiance that can be delivered to the retina for prolonged exposures are established by the American National Standards Institute (ANSI) [104, 105]. Accordingly, for illumination in the 775-1000 nm wavelength range, the maximum permissible average exposure is 2.8-8 mW/mm$^2$ for chronic retinal prosthesis illumination [8, 45]. For single-pulse exposure, the limit of the peak-irradiance for 0.5 ms pulse duration at 905 nm wavelength is 350 mW/mm$^2$ [8, 106].

## 5.2 In-vitro retinal stimulation with the photodiode array

The ability of the photodiode array to stimulate retina was evaluated in-vitro using the high-density microelectrode array (MEA) system, which was described in section 4.4. Photovoltaic stimulation of the inner retinal neurons leads to the generation of action potentials (spiking) by the retinal ganglion cells. The recording electrodes in the MEA can sense the action potentials as small spikes (~10-100 μV) due to the ohmic potential changes in the extracellular medium which result from the flow of ionic currents. A customized software package "VISION" was used to extract these spikes from the recorded signals, as well as to create electrophysiological images of the spiking neurons [99]. The system is capable of recording extracellular action potentials from hundreds of neurons simultaneously. Data is streamed off-line for further analysis, where spike sorting routines are used to attribute individual action potentials to specific neurons. The electrophysiological experiments began with enucleation of the eye from an anesthetized (35 mg/kg ketamine, 5 mg/kg xylazine, 0.01 mg/kg glycopyrrolate) long-evans rat (age 60 days) that was subsequently euthanized. A small piece of retina (~ 3 mm × 3 mm) was isolated and placed retinal ganglion cells (RGC) side down on the recording array with the prosthesis chip placed on the photoreceptor side. This is shown schematically in Figure 5.1(a) and its image through the optical microscope objective is shown in Figure 5.1(b). The retina was perfused with Ames' solution bubbled with 95% oxygen and 5% carbon dioxide and kept between 25°C and 30 °C.

A NIR laser source illuminates the prosthetic device through the optically transparent tracks of the MEA. A pulse of NIR light is transduced by the photodiode array with each illuminated pixel creating a charge-balanced pulse of electric current in response (Figure 5.1(c)). This results in a large stimulation-induced electrical artifact on the recording electrode, followed by a delay of 10 to 50 ms, whereupon the retina responds by initiating RGC spiking that is time-locked to the NIR pulse. The resultant spiking activity can be quantified by creating a peri-stimulus time histogram (PSTH) plot (Figure 5.1(c)) which illustrates the strength, latency and distribution of the NIR-induced electrical stimulation of the retina. The repetition rate of the NIR pulses was 2 Hz, to ensure that the RGC firing rate had returned to the spontaneous level before the next stimulation pulse.

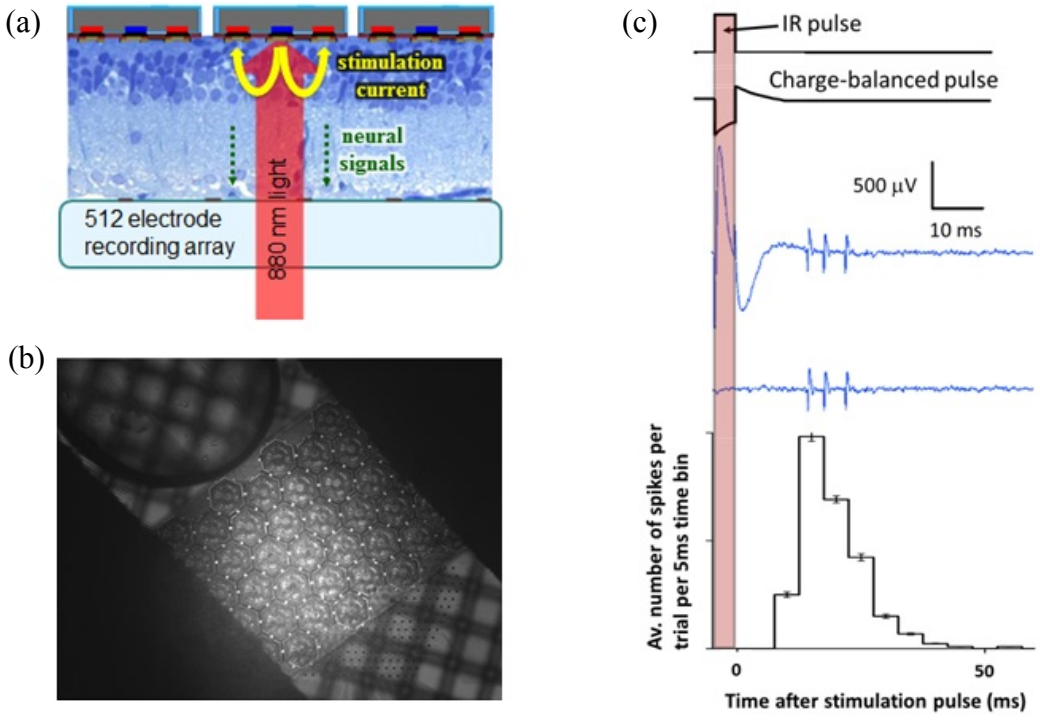



Figure 5.1. a) A schematic of the experimental setup. The retina is placed between the recording array (ganglion cells side) and the photodiode array (photoreceptors side). b) Optical microscope image of the photodiode array on top of the MEA with the retina in between. Picture was taken by a camera placed at the bottom-side of the transparent MEA. Transmission illumination by visible light allows imaging the MEA, while 880 nm illumination reflected off the prosthesis surface, passes through a 50% mirror onto the CCD camera. c) An IR pulse, with variable pulse width and intensity, creates a charge-balanced current waveform in each pixel of the photodiode array. The microelectrode array records the resultant stimulus artifact and retinal responses from each of the 512 electrodes. Stimulation is repeated at least 400 times for each setting. The artifact is then subtracted and the recorded action potentials undergo principal component analysis and automated clustering to attribute spiking waveforms to over 100 RGCs per experiment. A typical peristimulus time histogram representing the average number of spikes of a single neuron per 5 ms time bin per trial is shown at the bottom of (c).

All experimental procedures were conducted in accordance with institutional guidelines and conformed to the guidelines of the Association for Research in Vision and Ophthalmology (ARVO) Statement for the Use of Animals in Ophthalmic and Vision Research.

The experiment outlined above allowed the examination of the RGC activity in wild-type (WT) rat retina in response to NIR-induced electrical stimulation from the subretinal implant (medium-size pixel array, as shown in Figure 3.49). The implant

underwent full field illumination (NIR light, λ=880 nm) with RGCs under the prosthesis responding well to stimulation. Figure 5.2(a) shows a typical response from one of the RGCs in the form of a PSTH. Spiking activity occurs between 10 and 50 ms after the NIR light pulse with a peak latency ~25 ms. Threshold responses (defined as the 50% probability a spike was elicited) occurred at an irradiance level of 0.9 mW/mm$^2$ with pulse duration of 1 ms and 0.3 mW/mm$^2$ with pulse duration of 4 ms. Spike trains of 2-3 spikes were elicited at higher irradiance levels. Figure 5.2(b) shows how reliable trains of spikes (3 spikes per train) can be produced at an irradiance of 6.9 mW/mm$^2$ and a pulse width of 4 ms. An example of the raw data can be seen in Figure 5.1(c), which shows a 3-spike response. Spikes are separated by 3-4 ms, corresponding to an increase in spike rate from a spontaneous background rate of 1-3 Hz to a peak rate of ~250 Hz occurring 25 ms after the stimulation pulse.

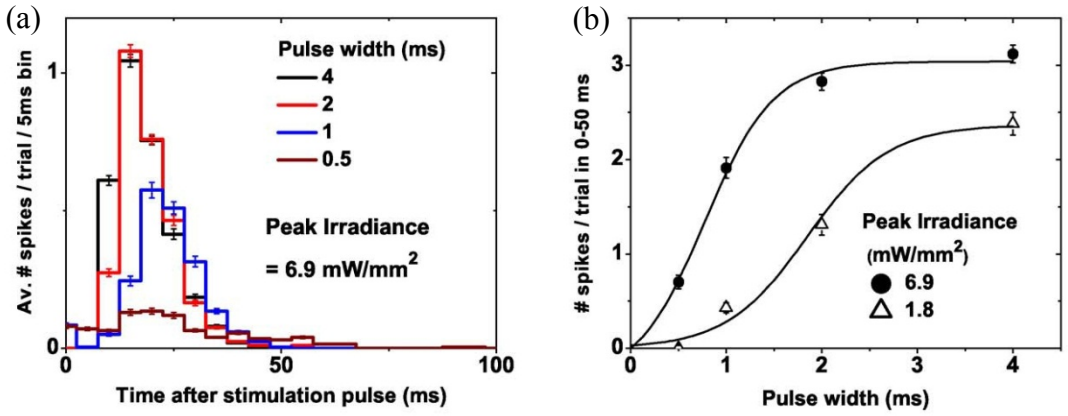



Figure 5.2. (a) PSTH response from a typical RGC under NIR-induced electrical stimulation by the prosthesis chip. Varying the width of the NIR pulse modulates the strength of the RGC response. (b) Integration under the peak response (with the background spontaneous rate subtracted) gives the total number of spikes per trial, which can be varied from 1 to 3 by modulating the pulse width or NIR irradiance.

## 5.3 In-vivo testing of the photodiode array

The fabricated implant is thin, small in size and powered through projected NIR light. This makes the surgical procedure much simpler than that of other retinal prosthetic approaches, which are connected by a cable to a power/data supply. The implantation procedure in the rat involves a small sclerotomy done 1.5 mm posterior to the limbus,

followed by a subretinal injection to create a retinal bleb. The implant is then inserted into the subretinal space using a custom insertion tool [44]. Optical Coherence Tomography (OCT, 870 nm) (Spectralis, Heidelberg Engineering, Germany) was performed during the 90-day follow-up period to examine the implant location and anatomy of the retinal tissue around it. Crucially, the imaging demonstrated how close the stimulation electrodes are to the inner retinal layer. The color fundus photo shown in Figure 5.3(a) demonstrates a medium-size pixel implant located under the retina in an 18-week old Long-Evans rat. Figure 5.3(b) demonstrates an IR image of the same implant, showing retinal blood vessels crossing the device. The OCT image in Figure 5.3(c) demonstrates a cross section of the retina with the subretinal implant (along the line marked in Figure 5.3(b)). OCT measures the optical path length, therefore due to the much larger refractive index of silicon at 870 nm wavelength (n=3.6) [47] than that of water (n=1.3) [107], the implant appears in OCT almost 3 times as thick as it actually is. The OCT image in Figure 5.3(c) demonstrates very close and uniform proximity between the implant surface and the inner nuclear layer 90 days after implantation. Visual evoked potential (VEP) was measured with a 915 nm wavelength beam projected into 1 mm-diameter spot on the subretinal implant in WT and Royal College of Surgeons (RCS) rats [108] using 2 Hz repetition rate of stimulation. Threshold peak irradiance with large- and medium-size pixels using 4 ms pulses was 0.5 $mW/mm^2$ in both species. With small-size pixels the stimulation threshold increased to 1 $mW/mm^2$. When the spot was directed on the retina away from the implant, no response could be detected.

To illustrate positioning of the implant above a retinal pigment epithelium (RPE), and to provide realistic size comparison, an array with small-size pixels was placed on an RPE in a porcine eye, and imaged with a scanning electron microscope. Figure 5.4 shows the hexagonal RPE mosaic with cell sizes of about 15 μm in width, and a 30 μm-thick photovoltaic array with 70 μm pixels.

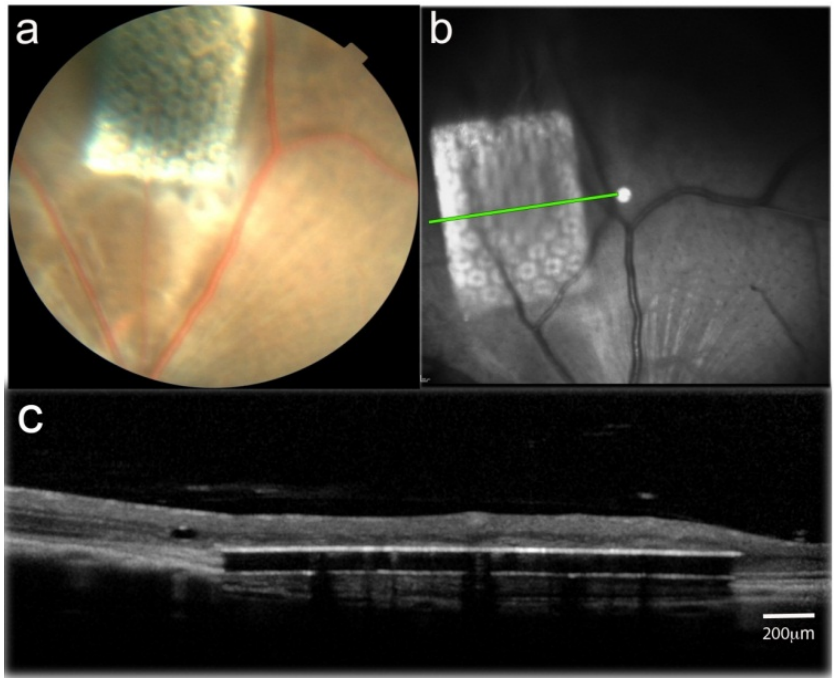



Figure 5.3. Imaging of the subretinal implant (1.2 mm × 0.8 mm) with medium-size pixels in a rat eye 90 days after implantation in-vivo: (a) Colour fundus photo of the subretinal implant sitting on the retina. (b) OCT image of the array, showing retinal blood vessels crossing the subretinal implant. (c) OCT image showing the cross section of the retina with the subretinal implant (along the line marked in (b)).

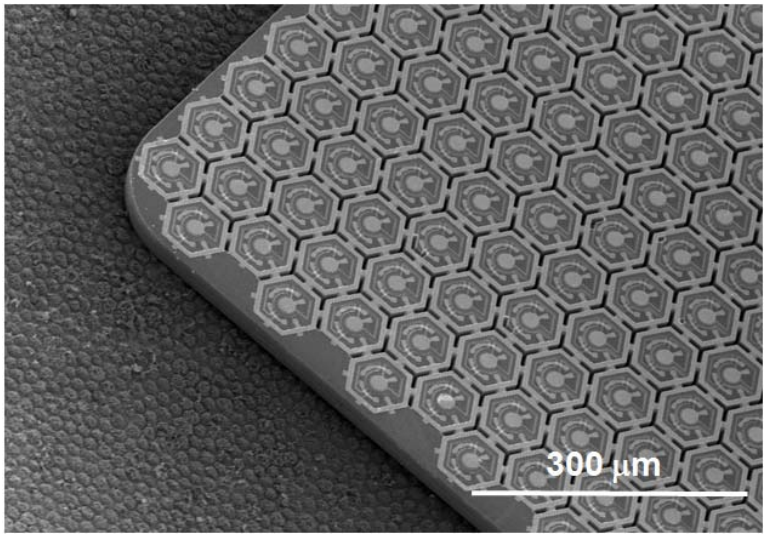



Figure 5.4. SEM image of an array with small-size pixels on retinal pigment epithelium (RPE) in a porcine eye.

# 6. REDESIGN AND FUTURE WORK

As mentioned in section 2.4.3, since the anodal mode has a lower threshold in retinal stimulation, a photodiode array with $p^+$ regions connected to the central active electrode has been designed and is being fabricated. The new mask design and process flow simulation will be discussed in this chapter.

## 6.1 Redesign of the mask

To design the anodal version devices where the $p^+$ regions are connected to the central active electrode, we just need to switch the sequence of the $n^+$ layer and $p^+$ layer, and switch the substrate from lightly boron doped to lightly phosphorus doped. To avoid problems with sharp corners of the rectangular implant during surgery and to make it better match the spherical shape of the eye ball, we re-designed the shape of the arrays from rectangular to round. Figure 6.1 shows the overall layout of the new mask, including large-size 3-series-connected-diode pixel array, medium-size 1-diode, 2-series-connected-diode and 3-series-connected-diode pixel array and small-size 1-diode, 2-series-connected-diode, and 3-series-connected-diode pixel array.

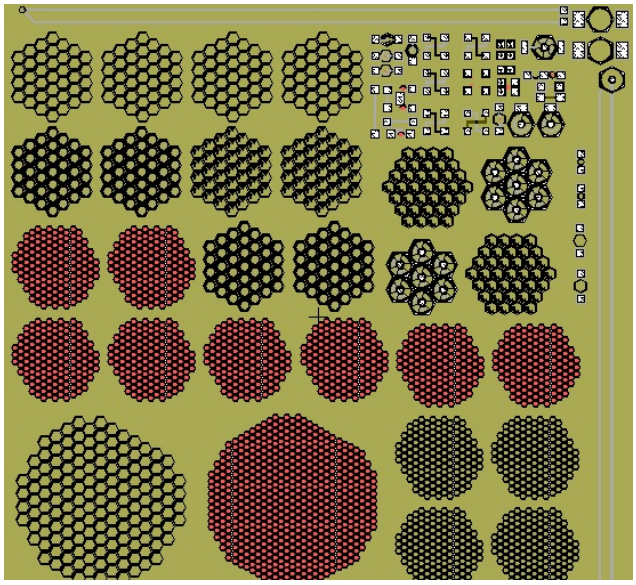
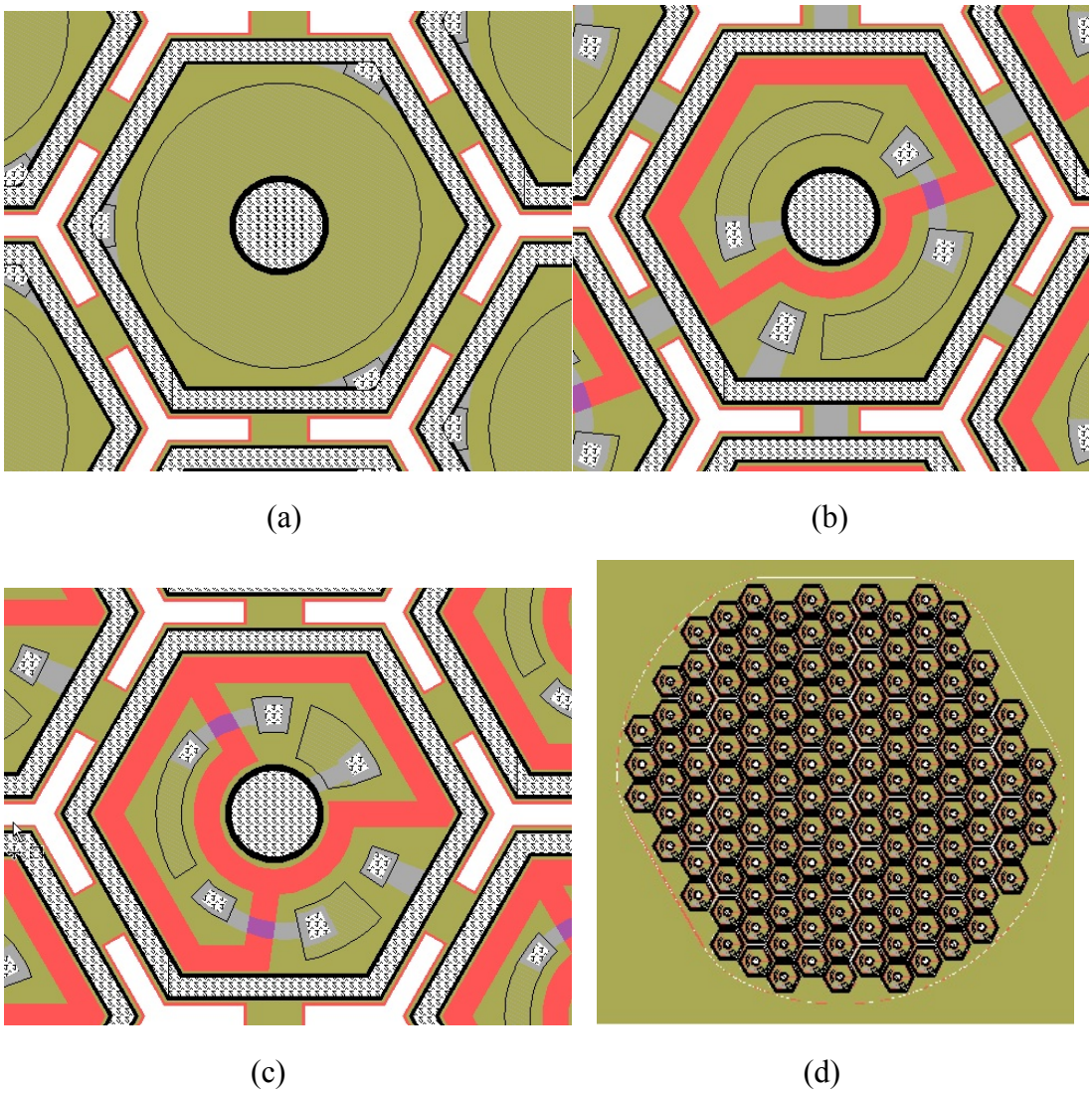


Figure 6.1. Overall mask for the new design, including large-size 3-series-connecteed-diode pixel array, medium-size 1-diode, 2-series-connected-diode and 3-series-connected-diode pixel array and small-size 1-diode, 2-series-connected-diode, and 3-series-connected-diode pixel array.

The design details for a small-size pixel array are shown in Figure 6.2, including a one-diode small-size pixel, two-series-connected-diode small-size pixel, three-series-connected-diode small-size pixel and the overall array of small-size pixels. From Figure 6.2(d), we can see that the array is not symmetrical to better differentiate the front side and back side of the array during the surgery.



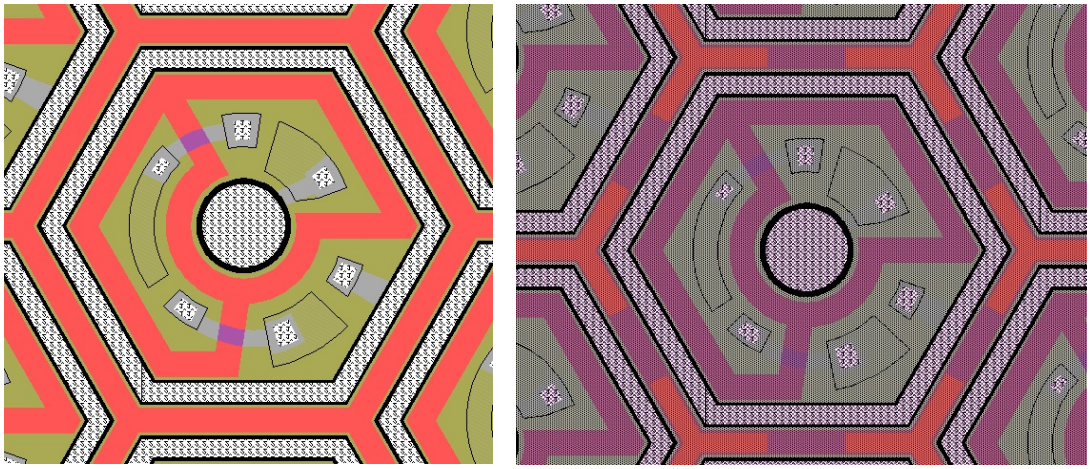
(a) (b)

(c) (d)

Figure 6.2. Mask design of a small-size pixel: (a) one-diode pixel; (b) two-series-connected-diode pixel; (c) three-series-connected-diode pixel; (d) overall small-size pixel array.

To improve the confinement of current, fully isolated pixels and pixels connected by polysilicon bridges are also designed as shown in Figure 6.3. By fully etching the surrounding silicon of each pixel, the common ground which is formed by

connecting the pixels by silicon bridges is disconnected, which may help with the localization of stimulation. Polysilicon bridges can be used to allow opening the trenches between pixels in order to provide channels for nutrient flow.



(a) (b)

Figure 6.3. (a) Fully-isolated pixel design. (b) Pixels connected by polysilicon bridges.

## 6.2 Simulation of the new process flow

In the new fabrication flow, several main parameters are changed. First, since the measured oxide thickness for the anti-reflection coating on the $n^+$ regions of the previously fabricated devices was much larger than expected, the final thermal oxidation time will be reduced from 83 minutes to 60 minutes in the new process flow. Second, since the measured sheet resistance of the $p^+$ regions was larger than expected, the dose for the boron ion implantation will be increased from $1.3\times10^{15}$ $cm^{-2}$ to $4\times10^{15}$ $cm^{-2}$ in the new process flow. In addition, the top layer of oxide after the DRIE trench etch will be removed by BOE in the new process flow to reduce the polysilicon step height across the wafer, which is discussed in section 3.3.3.5.

The simulation of $n^+$ regions starts with a lightly phosphorus doped substrate with a doping concentration of $1\times10^{15}$ $cm^{-3}$. In the simulated process flow, first, a 30 minute, 1000 ˚C phosphorus predeposition is performed; then a 20 minute wet oxidation is performed to block the boron implantation; next, a dry oxidation at 950 ˚C for 15 minutes is performed on the $n^+$ regions during the growth of the screen oxide in the $p^+$ regions;

finally, all of the oxide on top of the $n^+$ regions is etched, and a 60 minute dry oxidation at 1000 ˚C followed by a 75 minute anneal is performed. The simulated doping profile is shown in Figure 6.4.

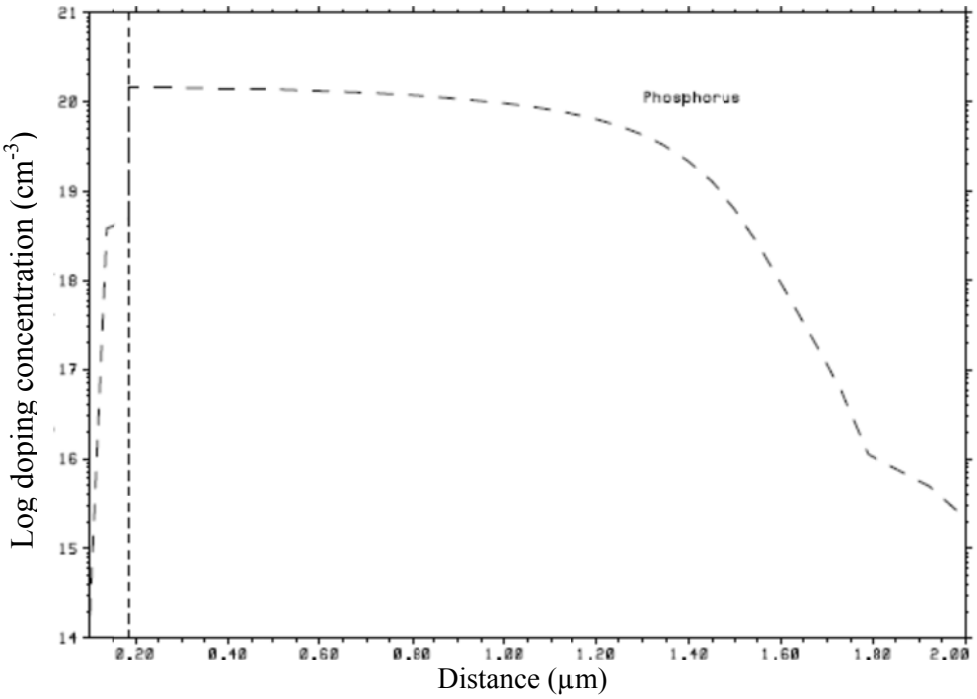



Figure 6.4. Doping profile of the $n^+$ region.

From the simulation results, oxide thickness on top of the $n^+$ region is 86.1 nm; the sheet resistance of the $n^+$ region is 7.1 Ω/sq.

The simulation of the $p^+$ regions starts with a lightly phosphorus doped substrate with a doping concentration of $1\times10^{15}$ $cm^{-3}$. In the simulated process flow, First, a dry oxidation at 950 ˚C for 15 minutes is performed during the growth of the screen oxide in the $p^+$ regions; next, a boron ion implantation with a dose of $4\times10^{15}$ $cm^{-2}$ is performed; then, the oxide on top is etched; finally, a 60 minute dry oxidation at 1000 ˚C followed by a 75 minute anneal is performed. The simulated doping profile of the heavily doped p region is shown in Figure 6.5.

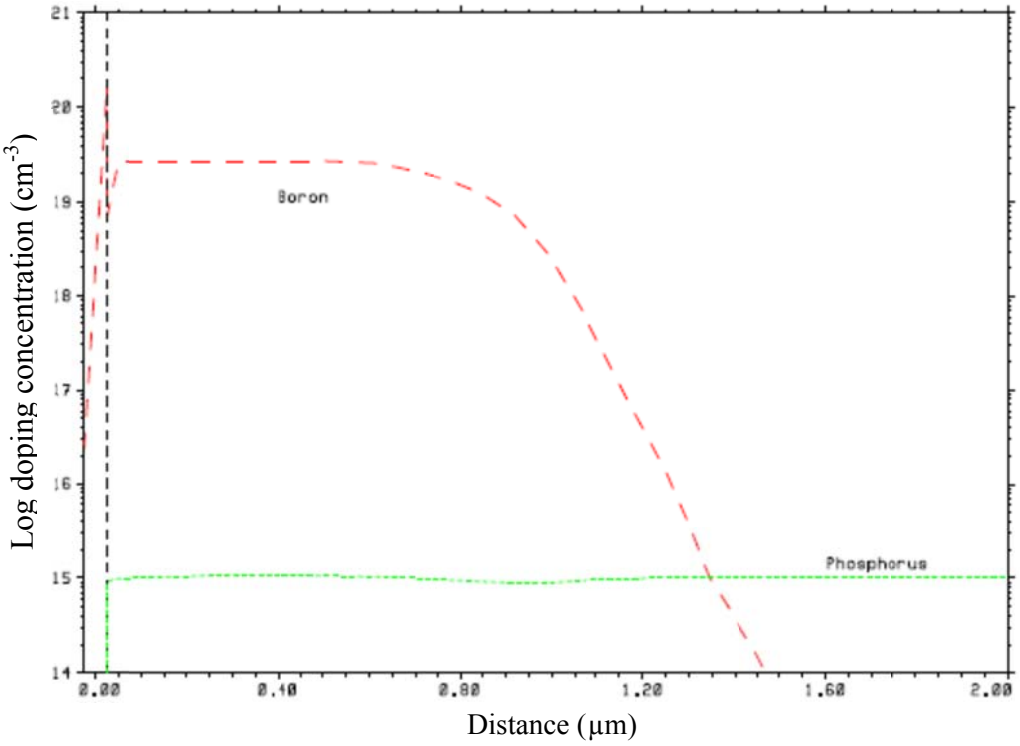



Figure 6.5. Doping profile of the $p^+$ region.

From the simulation results, oxide thickness on top of the $p^+$ region is 50.2 nm; the sheet resistance of the $p^+$ region is 52 Ω/sq; the junction depth is 1.32 μm.

The simulation of the lightly doped n regions starts with a lightly phosphorus doped substrate with a doping concentration of $1\times10^{15}$ $cm^{-3}$. First, a wet oxidation process at 1000 ˚C for 100 minutes is performed to grow 500 nm thermal oxide; then a wet oxidation process at 1000 ˚C for 20 minutes is performed during the oxidation process for $n^+$ regions; next, a dry oxidation at 950 ˚C for 15 minutes is performed on the lightly doped regions during the growth of the screen oxide in the $p^+$ regions; finally, the oxide on top is etched, and a 60 minute dry oxidation at 1000 ˚C followed by a 75 minute anneal is performed. The simulated doping profile of the lightly doped n region is shown in Figure 6.6. From the simulation results, the oxide thickness on top of the lightly doped region is 60.6 nm; the sheet resistance of the lightly doped region is 240 Ω/sq.

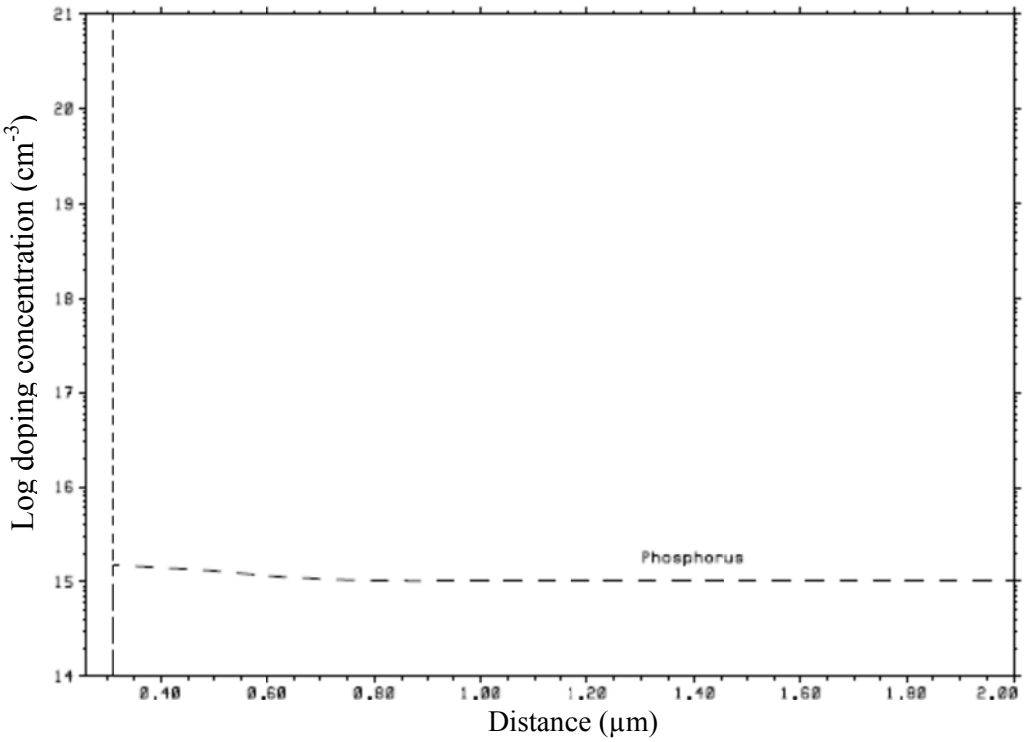



Figure 6.6. Doping profile of the lightly doped region.

The simulated sheet resistances and oxide thicknesses of the $n^+$, $p^+$ and lightly doped regions are summarized in Table 6.1.

| Parameters | $n^+$ | $p^+$ | lightly doped regions |
|---|---|---|---|
| Sheet resistance | 7.1 Ω/sq | 52 Ω/sq | 240 Ω/sq |
| Oxide thickness | 86.1 nm | 50.2 nm | 48.8 nm |

Table 6.1. Simulated parameters including sheet resistances and oxide thicknesses of the $n^+$, $p^+$ and lightly doped regions.

## 6.3 Conclusions and future work

The fabricated 30 μm thick photodiode arrays have low reverse-bias dark current levels (<100 pA) and high breakdown voltages (>20 V), which are capable of reliable subretinal stimulation. Operating the implants photovoltaically avoids power cables and greatly simplifies the surgical procedure. The local return electrodes in each pixel reduce cross-talk between pixels by confining current flow, and the channels opened between the pixels allow perfusion of nutrient to the retina. Retinal stimulation has been achieved

with pixel sizes down to 70 μm. Reliable spike trains were elicited in-vitro at irradiance levels at least one order of magnitude lower than the retinal damage limits at NIR wavelengths [45], as discussed in section 5.1.1. OCT images of the implanted 30 μm-thick device with perfusion channels between adjacent pixels demonstrated good acceptance of the array in the subretinal space and close proximity to the Inner Nuclear Layer.

Future work will involve fabrication and characterization of the anodal version of the devices, exploring the optimum number of diodes per pixel, optimization of the ratio of the photosensitive area to electrode size, adding a shunt resistor in the device design to speed-up the discharge phase of the stimulation waveform, perfecting the anti-reflection coating, adding a biocompatible Parylene coating to better passivate the devices, integrating pillar electrodes that provide much closer proximity to the target neurons to reduce stimulation threshold, and optimizing the process flow to improve array yield.

# 7. REFERENCES

## APPENDIX

Process flow

1. Scribe wafers
   Wafer ID (B=batch, D=device, T=test)
2. Measure wafer resistivity:
   Tool: Prometrix
3. Clean wafers
   Tool: Wetbench Nonmetal
   Piranha Temperature: 120 ˚C Time: 20 minutes
   Spin rinse/dry
4. Pre-oxidation clean
   Tool: Wetbench Diffusion
   $H_2SO_4$: $H_2O_2$ 4:1 Temperature: 90 ˚C Time: 10 minutes
   Water rinse 5 minutes
   HF 50:1 30 seconds
   Water rinse 5 minutes
   $H_2O$: HCL: $H_2O_2$ 5:1:1 Temperature: 70 ˚C Time: 10 minutes
   Water rinse 5 minutes
   Spin rinse/dry
5. Oxidation
   Tool: Tylan2 Temperature: 1000 ˚C Time: 100 minutes
   Recipe: Wet1000 Target thickness: 500 nm Actual thickness: 490 nm
6. Photolithography
   Vapor prime
   Tool: YES oven Temperature: 150 ˚C Time: 35 minutes
   Spin coating
   Tool: SVG coater Photoresist: 1um 3612 Recipe: 7 Bake: 90 ˚C Time: 60 seconds
   Exposure
   Tool: ASML Job file: photodiodepoly Mask layer: Alignment Dose: 80 mj
   Develop
   Tool: SVGDEV Recipe: Post-exposure: 1 Development: 3 Bake: 1
   DESCUM
   Tool: DRYTEK2 Recipe: DESCUM Time: 30 seconds
7. Etch alignment mark
   Tool: AMT etcher Recipe: Prog3 Time: 20 minutes @ 30nm/min
   Prog4 Time: 5.5minutes @20 nm/min
8. Photoresist cleaning
   Tool: gasonics Recipe: 013
   Tool: Wetbench nonmetal
   Piranha: $H_2SO_4$: $H_2O$ 9:1 Temperature: 120 ˚C Time: 20 minutes
   Water rinse Time: 5 minutes
   Spin rinse/dry
9. Photolithography
   Vapor prime
   Tool: YES oven Temperature: 150 ˚C Time: 35 minutes
   Spin coating
   Tool: SVG coater Photoresist: 1um 3612 Recipe: 7 Bake: 90 ˚C Time: 60 seconds
   Exposure
   Tool: ASML Job file: photodiodepoly Mask layer: Isopoly Dose: 80 mj
   Develop
   Tool: SVGDEV Recipe: Post-exposure: 1 Development: 3 Bake: 1

DESCUM
Tool: DRYTEK2 Recipe: DESCUM Time: 30 seconds

10. Oxide etch and trench etch
    Tool: AMT etcher Recipe: Prog3 Time: 20min @ 30 nm/min
    Tool: STS1 Recipe: DEEP Time: 18 minutes
11. Photoresist cleaning
    Tool: gasonics Recipe: 013
    Tool: Wetbench nonmetal
        Piranha: $H_2SO_4$: $H_2O$ 9:1 Temperature: 120 ˚C Time: 20 minutes
        Water rinse Time: 5 minutes
        Spin rinse/dry
12. Pre-oxidation clean
    Tool: Wetbench Diffusion

| | | |
|---|---|---|
| $H_2SO_4$: $H_2O_2$ 4:1 | Temperature: 90 ˚C | Time: 10 minutes |
| Water rinse | | 5 minutes |
| HF 50:1 | | 30 seconds |
| Water rinse | | 5 minutes |
| $H_2O$: HCL: $H_2O_2$ 5:1:1 | Temperature: 70 ˚C | Time: 10 minutes |
| Water rinse | | 5 minutes |

        Spin rinse/dry
13. Oxidation
    Tool: Tylan2 Temperature: 1000 ˚C Time: 100 minutes
    Recipe: Wet1000 Target thickness: 500 nm
14. LPCVD polysilicon
    Tool: Tylan9 Recipe: POLYEMIT Temperature: 620 ˚C
    Time: 5 hours Target thickness: 4 μm Actual thickness: 3.2-3.5 μm
15. Chemical mechanical polishing
    Silicon Quest International, Inc.: polish off the 4 μm polysilicon on the top
    Tool: Wetbench Nonmetal
        BOE 6:1 Time: 3 minutes
16. Pre-oxidation clean
    Tool: Wetbench Diffusion

| | | |
|---|---|---|
| $H_2SO_4$: $H_2O_2$ 4:1 | Temperature: 90 ˚C | Time: 10 minutes |
| Water rinse | | 5 minutes |
| HF 50:1 | | 30 seconds |
| Water rinse | | 5 minutes |
| $H_2O$: HCL: $H_2O_2$ 5:1:1 | Temperature: 70 ˚C | Time: 10 minutes |
| Water rinse | | 5 minutes |

        Spin rinse/dry
17. Oxidation
    Tool: Tylan2 Temperature: 1000 ˚C Time: 35 minutes
        Recipe: Wet1000 Target thickness: 266 nm Actual thickness: 270 nm

18. Photolithography
    Vapor prime
    Tool: YES oven Temperature: 150 ˚C Time: 35 minutes
    Spin coating
    Tool: SVG coater Photoresist: 1um 3612 Recipe: 7 Bake: 90 ˚C Time: 60 seconds
    Exposure
    Tool: ASML Job file: photodiodepoly Mask layer: Nwell Dose: 80 mj
    Develop
    Tool: SVGDEV Recipe: Post-exposure: 1 Development: 3 Bake: 1
    DESCUM
    Tool: DRYTEK2 Recipe: DESCUM Time: 30 seconds
19. Oxide etch
    Tool: AMT etcher Recipe: Prog3 Time: Measure oxide thickness to determine (650 nm/30 min)

20. Photoresist cleaning
    Tool: gasonics Recipe: 013
    Tool: Wetbench nonmetal
        Piranha: $H_2SO_4$: $H_2O$ 9:1 Temperature: 120 ˚C Time: 20 minutes
        Water rinse Time: 5 minutes
        Spin rinse/dry
21. Pre-diffusion clean
    Tool: Wetbench Diffusion
        $H_2SO_4$: $H_2O_2$ 4:1 Temperature: 90 ˚C Time: 10 minutes
        Water rinse 5 minutes
        $H_2O$: HCL: $H_2O_2$ 5:1:1 Temperature: 70 ˚C Time: 10 minutes
        Water rinse 5 minutes
        HF 50:1 30 seconds
        Water rinse 5 minutes
        Spin rinse/dry
22. Phosphorus predeposition
    Tool: Tylan6 Recipe: POCL1000 Temperature: 1000 ˚C Time: 30 minutes
    Phosphorus glass removal
    Tool: Wetbench nonmetal BOE 6:1 Time: 1 minute Resistivity: 4 ohm/sq
23. Pre-oxidation clean
    Tool: Wetbench Diffusion
        $H_2SO_4$: $H_2O_2$ 4:1 Temperature: 90 ˚C Time: 10 minutes
        Water rinse 5 minutes
        HF 50:1 30 seconds
        Water rinse 5 minutes
        $H_2O$: HCL: $H_2O_2$ 5:1:1 Temperature: 70 ˚C Time: 10 minutes
        Water rinse 5 minutes
        Spin rinse/dry
24. Oxidation
    Tool: Tylan2 Temperature: 1000 ˚C Time: 20 minutes
        Recipe: Wet1000 Target thickness: 170 nm Actual thickness: 263 nm @ Nwell
25. Photolithography
    Vapor prime
    Tool: YES oven Temperature: 150 ˚C Time: 35 minutes
    Spin coating
    Tool: SVG coater Photoresist: 1um 3612 Recipe: 7 Bake: 90 ˚C Time: 60 seconds
    Exposure
    Tool: ASML Job file: photodiodepoly Mask layer: Pwell Dose: 80 mj
    Develop
    Tool: SVGDEV Recipe: Post-exposure: 1 Development: 3 Bake: 1
    DESCUM
    Tool: DRYTEK2 Recipe: DESCUM Time: 30 seconds
26. Oxide etch
    Tool: AMT etcher Recipe: Prog3 Time: Measure oxide thickness to determine (460 nm/30 min)
27. Photoresist cleaning
    Tool: gasonics Recipe: 013
    Tool: Wetbench nonmetal
        Piranha: $H_2SO_4$: $H_2O$ 9:1 Temperature: 120 ˚C Time: 20 minutes
        Water rinse Time: 5 minutes
        Spin rinse/dry
28. Pre-oxidation clean
    Tool: Wetbench Diffusion
        $H_2SO_4$: $H_2O_2$ 4:1 Temperature: 90 ˚C Time: 10 minutes
        Water rinse 5 minutes
        HF 50:1 30 seconds
        Water rinse 5 minutes

$H_2O$: HCL: $H_2O_2$ 5:1:1 Temperature: 70 ˚C Time: 10 minutes
Water rinse 5 minutes
Spin rinse/dry

29. Oxidation
Tool: Tylan2 Recipe: DRY950 Temperature: 950 ˚C Time: 15 minutes Target: 15 nm
Actual: 13 nm
30. Ion implantation
INNOViON Corporation, $B^{11}$ singly charged boron Energy: 20 KeV Dose: $1.3\times10^{15}$ $cm^{-2}$
31. Clean wafers and oxide etch
Tool: Wetbench Nonmetal
Piranha Temperature: 120 ˚C Time: 20 minutes
Water rinse Time: 5 minutes
BOE (6:1) Time: 352 seconds for 440 nm
Water rinse Time: 5 minutes
Spin rinse/dry
32. Pre-oxidation clean
Tool: Wetbench Diffusion
$H_2SO_4$: $H_2O_2$ 4:1 Temperature: 90 ˚C Time: 10 minutes
Water rinse 5 minutes
HF 50:1 30 seconds
Water rinse 5 minutes
$H_2O$: HCL: $H_2O_2$ 5:1:1 Temperature: 70 ˚C Time: 10 minutes
Water rinse 5 minutes
Spin rinse/dry
33. Oxidation
Tool: Tylan2 Recipe: DRY1000A
Oxidation: Temperature: 1000 ˚C Time: 83 minutes Target: 60 nm
Actual: 109 nm @ Nwell 80 nm @ Pwell Annealing: Temperature: 1000 ˚C Time: 75 minutes
34. Photolithography
Vapor prime
Tool: YES oven Temperature: 150 ˚C Time: 35 minutes
Spin coating
Tool: SVG coater Photoresist: 1um 3612 Recipe: 7 Bake: 90 ˚C Time: 60 seconds
Exposure
Tool: ASML Job file: photodiodepoly Mask layer: Via1 Dose: 80 mj
Develop
Tool: SVGDEV Recipe: Post-exposure: 1 Development: 3 Bake: 1
DESCUM
Tool: DRYTEK2 Recipe: DESCUM Time: 30 seconds
35. Oxide etch
Tool: AMT etcher Recipe: Prog3 Time: 4.5 minutes @ 30 nm/min
36. Photoresist cleaning
Tool: gasonics Recipe: 013
Tool: Wetbench nonmetal
Piranha: $H_2SO_4$: $H_2O$ 9:1 Temperature: 120 ˚C Time: 20 minutes
Water rinse Time: 5 minutes
Spin rinse/dry
37. Photolithography
Vapor prime
Tool: YES oven Temperature: 150 ˚C Time: 35 minutes
Spin coating
Tool: Headway LOL2000 Time: 60 seconds Speed: 1000 rpm
Oven bake Temperature: 210 ˚C Time: 40 minutes
Spin coating
Tool: SVG coater Photoresist: 1um 3612 Recipe: 7 Bake: 90 ˚C Time: 60 seconds
Exposure

Tool: ASML Job file: photodiodepoly Mask layer: Metal1 Dose: 90 mj
Develop
Tool: SVGDEV Recipe: Post-exposure: 1 Development: 3 Bake: 1
DESCUM
Tool: DRYTEK2 Recipe: DESCUM Time: 30 seconds
Tool: Wetbench nonmetal
HF 50:1 Time: 90 seconds
Water rinse Time: 5 minutes
Spin rinse/dry

38. Metal1 deposition and liftoff
Tool: Metalica
Ti 20 nm/ Pt 250 nm
Tool: Wetbench Solvent
Acetone Time: 2 hours Methanol: 10 minutes Isopropanol: 10 minutes
39. Forming gas anneal
Tool: FGA2 Temperature: 425 ˚C Time: 30 minutes
40. PECVD nitride
Tool: STS PECVD Recipe: Deb5knit Time: 7 minutes Target: 70 nm Actual: 70 nm
41. Photolithography
Vapor prime
Tool: YES oven Temperature: 150 ˚C Time: 35 minutes
Spin coating
Tool: SVG coater Photoresist: 1um 3612 Recipe: 7 Bake: 90 ˚C Time: 60 seconds
Exposure
Tool: ASML Job file: photodiodepoly Mask layer: Via2 Dose: 90 mj
Develop
Tool: SVGDEV Recipe: Post-exposure: 1 Development: 3 Bake: 1
DESCUM
Tool: DRYTEK1 Recipe: DESCUM Time: 30 seconds
42. Nitride via etch
Tool: DRYTEK1
Recipe: Nitride etch Time: 2 minutes
43. Photoresist cleaning
Tool: Wetbench Solvent
Acetone Time: 30 minutes Methanol: 10 minutes Isopropanol: 10 minutes
44. Photolithography
Vapor prime
Tool: YES oven Temperature: 150 ˚C Time: 35 minutes
Spin coating
Tool: SVG coater Photoresist: SPR 220-3 Recipe: 5 Target: 3 μm Bake: 90 ˚C Time: 200 seconds
Exposure
Tool: ASML Job file: photodiodepoly Mask layer: ISOPOLY Dose: 180 mj
Develop
Tool: SVGDEV Development: 7
DESCUM
Tool: DRYTEK1 Recipe: DESCUM Time: 30 seconds
45. Etch nitride, oxide, polysilicon and buried oxide etch
Tool: DRYTEK1 Recipe: Nitride etch Time: 2 minutes 10 seconds
Tool: DRYTEK4 Recipe: Oxide etch Time: 5 minutes 10 seconds
Tool: Xactix Recipe: 25-3-15 file Time: 125 cycles (2 hours)
Tool: P5000 Recipe: Jimox Time: 4 minutes @ 150 nm-200 nm at bottom of trench
46. Photoresist cleaning
Tool: Wetbench Solvent
Acetone Time: 30 minutes Methanol: 10 minutes Isopropanol: 10 minutes
47. Forming gas anneal

Tool: FGA2 Temperature: 425 ˚C Time: 30 minutes

48. Photolithography
    Tool: Spray coater Recipe: RHC6 Target: 30 µm
    Tool: Karlsuss Time: 3 seconds × 5
    Manual develop Time: overdevelop
    Tool: Karlsuss Time: 2 seconds × 5
    Manual develop Time: overdevelop Comment: Top layer of photoresist cleared
    Spin coating
    Tool: SVG coater Photoresist: SPR 220-3 Target: 4 µm Bake: 90 ˚C Time: 200 seconds
    Exposure
    Tool: ASML Job file: photodiodepoly Mask layer: Metal2 Dose: 220 mj
    Develop
    Tool: SVGDEV Development: 7
    DESCUM
    Tool: DRYTEK1 Recipe: DESCUM Time: 30 seconds
49. Iridium oxide deposition and liftoff
    Send to Boston, EIC Laboratories, Inc. Target: 300 nm Actual: 300 nm
    Tool: Wetbench Solvent
    Acetone Time: 2 hours Methanol: 10 minutes Isopropanol: 10 minutes
50. Handle wafer removal and release
    Tool: Headway
    Protek primer Time: 1 minute Speed: 1500 rpm
    Hotplate back Time: 1 minute Temperature: 205 ˚C
    Protek Time: 1 minute Speed: 1500 rpm
    Hot plate bake (soft) Time: 2 minutes Temperature: 130 ˚C
    Hot plate bake (hard) Time: 1 minute Temperature: 205 ˚C
    Silicon Quest International, Inc.: Thin wafer from 500 µm to 80 µm
    Tool: Wetbench General
    Clamp the wafer: HF 50:1 Time: 1 minute
    Water rinse Time: 5 minutes
    TMAH Concentration: 10% Temperature: 90 ˚C Time: 1 hour Etch rate: 1-1.2 µm/min
    Water rinse Time: 30 minutes
    Tool: Wetbench Solvent
    Acetone Time: 24 hours Ultrasonic: 1 minute 30 seconds to 5 minutes
    Use dropper to transfer devices to holder for testing

## RECIPES

Oxide etch recipe in Drytek4:
Process: 3 Pressure: 100 mtorr CHF3: 100 sccm $O_2$: 30%
Forward/Reflected power: 100/0 watts

Buried oxide etch in P5000:
Process: JIMOX CHF3: 25 sccm CF4: 50 sccm Argon: 100 sccm
Pressure: 250 mtorr DC bias: -1000 to 0 RF power: 500 W Magnetic: 60 Gauss

Deep trench etch recipe in STS2:
Process: HAR NarrowGap Dep: 2 seconds Etch: 3 seconds
Dep Gas: $C_4F_8$ 200 sccm Etch Gas: $SF_6$ 200 sccm $O_2$: 20 sccm
Dep coil generator 13.56 MHz: 1200 W Etch coil generator 13.56 MHz: 2000 W
Dep coil generator 380 kHz: 0 Etch coil generator 380 kHz: 85 W

$XeF_2$ polysilicon etch recipe in Xactix:
Time: 25 cycles $XeF_2$ etching time per cycle: 15 seconds $XeF_2$ pressure: 3 Torr

## Bibliography


[1] W. Smith, et al., "Risk factors for age-related macular degeneration: Pooled findings from three continents", *Ophthalmology.*, vol. 108, pp. 697-704, 2001

[2] M. Haim, "Epidemiology of retinitis pigmentosa in Denmark", *Acta Ophthalmol Scand Suppl.*, pp. 1-34, 2002

[3] S. Y. Kim, S. Sadda, J. Pearlman, M. S. Humayun, E. de Juan Jr., B. M. Melia and W. R. Green, "Morphometric analysis of the macula in eyes with disciform age-related macular degeneration", *Retina.*, vol. 22(4), pp. 471-7, 2002

[4] F. Mazzoni, E. Novelli and E. Strettoi, "Retinal ganglion cells survive and maintain normal dendritic morphology in a mouse model of inherited photoreceptor degeneration", *J. Neurosci.*, vol. 28(52), pp. 14282-92, 2008

[5] J. L. Stone, W. E. Barlow, M. S. Humayun, E. de Juan Jr. and A. H. Milam, "Morphometric analysis of macular photoreceptors and ganglion cells in retinas with retinitis pigmentosa", *Arch. Ophthalmol*., vol.110, pp. 1634-1639, 1992

[6] J. D. Weiland, W. Liu, M. S. Humayun, "Retinal prosthesis", *Annu. Rev. Biomed. Eng.*, vol. 7, pp. 361–401, 2005

[7]A. C. Guyton, *Human physiology and mechanisms of disease*, 6th ed. Philadelphia: W.B. Saunders Company, 1997.

[8] J. D. Loudin, *High Resolution Photovoltaic Retinal Prosthesis*, PhD. Dissertation, Stanford University, 2010

[9] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve", *J Physiol.,* vol. 117, pp. 500-44, 1952

[10] R. J. Greenberg *et al.*, "A computational model of electrical stimulation of the retinal ganglion cell", *IEEE Trans Biomed Eng.,* vol. 46, pp. 505-14, 1999

[11] G. Brindley and W. Lewin, "The sensations produced by electrical stimulation of the visual cortex", *Journal of Physiology,* vol. 196, pp. 479-493, 1968

[12] W. H. Dobelle *et al.*, "Artifical vision for the blind: electrical stimulation of visual cortex offers hope for a functional prosthesis", *Science,* vol. 183, pp. 440-4, 1974.

[13] J. Knutson *et al.*, "Electrode fracture rates and occurences of infection and granuloma associated with percutaneous intramuscular electrodes in upper-limb functional electrical stimulation applications", *J. Rehabil. Res. Develop.,* vol. 39, pp. 671-684, 2002

[14] M. Humayun, Y. Sato, R. Propst, E. de Juan Jr., "Can potentials from the visual cortex be elicited electrically despite severe retinal degeneration and a markedly reduced electroretinogram? " , *Ger J Ophthalmol.* vol. 4(1), pp. 57-64, 1995

[15] R. J. Greenberg, *Analysis of electrical stimulation of the vertebrate retina — work towards a retinal prosthesis,* PhD. Dissertation, Johns Hopkins University, 1998

[16] S. J. Katona, *Excitation thresholds for the electrical stimulation of the normal and degenerate mouse retina,* Master's Thesis, Johns Hopkins University, 1998

[17] M. S. Humayun, E. de Juan Jr., G. Dagnelie, R. J. Greenburg, R. H. Propst, and H. Phillips, "Visual Perception elicited by electrical stimulation of retina in blind humans", *Arch Ophthalmol*., vol. 114(1), pp. 40–46, 1996

[18] M. S. Humayun, E. de Juan Jr., J. D. Weiland, G. Dagnelie, S. Katona, R. Greenberg and S. Suzuki, "Pattern electrical stimulation of the human retina", Vision Research, vol. 39, pp. 2569-2576, 1999
[19] R. Eckmiller, "Learning retina implants with epiretinal contacts", *Ophthalmic Research*, vol. 29(5), pp. 281–289, 1997
[20] M. S. Humayun, "Intraocular retinal prosthesis", *Transactions of the American Ophthalmological Society*, vol. 99, pp. 271–300, 2001
[21] J. Rizzo and J. Wyatt, "Prospects for a visual prosthesis", *Neuroscientist*, vol. 3, pp. 251–262, 1997
[22] M. S. Humayun, J. D. Weiland, G. Y. Fujii, R. Greenberg, R. Williamson, J. Little, B. Mech, V. Cimmarusti, G. V. Boemel, G. Dagnelie and E. de Juan Jr., "Visual perception in a blind subject with a chronic microelectronic retinal prosthesis", Vision Research, vol. 43, pp. 2573-81, 2003
[23] M. S. Humayun, R. J. Greenberg, B. Mech, D Yanai, M. Mahadevappa et al., " Chronically implanted intraocular retinal prosthesis in two blind subjects", *Invest. Ophth. Vis. Sci.*44: E-Abstract 4206, 2003
[24] J. D. Weiland, D. Yanai, M. Mahadevappa, R. Williamson, B. V. Mech et al., "Visual task performance in blind humans with retinal prosthetic implants", *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.Soc., 26th, San Francisco, CA, 2004*
[25] M. S. Humayun et al., "Preliminary 6 month results from the ArgusTM II epiretinal prosthesis feasibility study", *Conf. Proc. IEEE Eng. Med. Biol. Soc.,* pp.4566–8, 2009
[26] L. daCruz, B. Coley, P. Christopher, F. Merlini, V. Wuyyuru, J. A. Sahel, P. Stanga, E. Filley, G. Dagnelie and Argus II Study Group, "Patients blinded by outer retinal dystrophies are able to identify letters using the ArgusTM II retinal prosthesis system", *Invest. Ophthalmol. Vis. Sci.* 51: E-Abstract 2023, 2010
[27] A. K. Ahujia, J. D. Dorn, A. Caspi, M. J. McMahon, G. Dagnelie, L. daCruz, P. Stanga, M. S. Humayun, R. J. Greenberg and Argus II Study Group, "Blind subjects implanted with the Argus II retinal prosthesis are able to improve performance in a spatial-motor task", *Br. J. Ophthalmol.,* vol. 95(4), pp. 539-43, 2011
[28] J. D. Weiland, A. K. Cho and M. S. Humayun, "Retinal prostheses: current clinical results and future needs", *Ophthalmology.*, vol. 118(11), pp. 2227–37, 2011
[29] M. Javaheri, D. S. Hahn, R. R. Lakhanpal, J. D. Weiland, M. S. Humayun, "Retinal prostheses for the blind", Ann. Acad. Med. Singapore, vol. 35, pp. 137-44, 2006
[30] M. Volker, K. Shinoda, H. Sachs, H. Grneiner, T. Schwarz et al., "In vivo assessment of subretinally implanted microphotodiode arrays in cats by optical coherence tomography and fluorscein angiography", *Graefe's Arch. Ophthalmol.,* vol. 242(9), pp. 792–99, 2004
[31] A. Y. Chow and N. S. Peachey, "The subretinal microphotodiode array retinal prosthesis", *Ophthalmic Res.,* vol. 30, pp.195-8, 1998
[32] G. Peyman, A. Y. Chow, C. Liang, V. Y. Chow, J. I. Perlman and N. S. Peachey, "Subretinal semiconductor microphotodiode array", *Ophthalmic Surg Lasers,* vol. 29, pp. 234-41, 1998
[33] A. Y. Chow, M. T. Pardue, J. I. Perlman, S. L. Ball, V. Y. Chow, J. R. Hetling et al., "Subretinal implantation of semiconductor-based photodiodes: durability of novel implant designs", *J Rehabil Res Dev,* vol.39, pp. 313-21, 2002

[34] A. Y. Chow, V. Y. Chow, K. H. Packo, J. S. Pollack, G. A. Peyman and R. Schuchard, "The artificial silicon retina microchip for the treatment of vision loss from retinitis pigmentosa", *Arch. Ophthalmol*., vol. 122,pp. 460-9, 2004
[35] E. Zrenner, "Will retinal implants restore vision?" *Science*, vol.295, pp.1022-5, 2002
[36] E. Zrenner, "The development of subretinal microphotodiodes for replacement of degenerated photoreceptors", *Ophthalmic. Res.,* vol. 29, pp. 269-80, 1997
[37] E. Zrenner, A. Stett, "Can subretinal microphotodiodes successfully replace degenerated photoreceptors?" *Vision Res.,* vol. 39, pp. 2555-67, 1999
[38] H. G. Sachs and V. P. Gabel, "Retinal replacement – the development of microelectronic retinal prostheses – experience with subretinal implants and new aspects", Graefes Arch. Clin. Exp. Ophthalmol., vol. 242, pp. 717-23, 2004
[39] F. Gekeler F, K. Kobuch, H. N. Schwahn, A. Stett, K. Shinoda, E. Zrenner, "Subretinal electrical stimulation of the rabbit retina with acutely implanted electrode arrays", *Graefes. Arch. Clin. Exp. Ophthalmol.*,vol. 242, pp. 587-96, 2004
[40] H. G. Sachs, T. Schanze, M. Wilms, A. Rentzos, U. Brunner and F. Gekeler, "Subretinal implantation and testing of polyimide film electrode in cats", *Graefes. Arch. Clin. Exp. Ophthalmol.,* vol. 243, pp. 464-8, 2005
[41] E. Zrenner et al., "Subretinal electronic chips allow blind patients to read letters and combine them to words", *Proc. R. Soc.* B, vol. 278, pp. 1489-97, 2011
[42] R. Wilke et al., "Spatial resolution and perception of patterns mediated by a subretinal 16-electrode array in patients blinded by hereditary retinal dystrophies", *Invest. Ophthalmol. Vis. Sci.* , vol. 52(8), pp. 5995-6003, 2011
[43] D. V. Palanker, A. Vankov, P. Huie and S. Baccus, "Design of a high-resolution optoelectronic retinal prosthesis", *J. Neural Eng.*, vol. 2, pp. S105-S120, 2005
[44] J. D. Loudin, D. M. Simanovskii, K. Vijayraghavan, C. K. Sramek, A. F. Butterwick, P. Huie, G. Y. McLean and D. V. Palanker, "Optoelectronic retinal prosthesis: system design and performance", *J. Neural Eng.*, vol. 4, pp. S72-S84, 2007
[45] J. D. Loudin, S. F. Cogan, K. Mathieson, A. Sher and D. V. Palanker, "Photodiode Circuits for Retinal Prostheses", *IEEE Trans. Biomed. Circuits Syst.*, vol. 5(5), pp. 468-80, 2011
[46] R. F. Pierret, *Semiconductor Device Fundamentals*, Prentice Hall, 1996
[47] M. A. Green and M. Keevers, "Optical properties of intrinsic silicon at 300 K ", *Prog Photovoltaics,* vol. 3(3), pp. 189-192, 1995
[48] P. Bhattacharya, *Semiconductor Optoelectronic Devices*, Prentice Hall, 1994
[49] F. A. Kröger, G. Diemer and H. A. Klasens, "Nature of an ohmic metal-semiconductor contact", *Phys. Rev.*, vol.103, pp. 279, 1956
[50] C. A. Mead, *Ohmic Contacts to Semiconductors*, Electrochemical Society, New York, 1969
[51] J. Vilms and L. Wandinger, *Ohmic Contacts to Semiconductors*, Electrochemical Society, New York,1969
[52] J. W. Conley, C. B. Duke, G. D. Mahan and J. J. Tiemann, "Electron Tunneling in metal-semiconductor barriers", Phys. Rev., vol. 150 (2), pp. 466-69, 1966
[53] A. Y. C. Yu, "Electron tunneling and contact resistance of metal-silicon contact barriers", *Solid-State Electronics*, vol. 13, pp. 239-47, 1970
[54] K. Hess, *Advanced Theory of Semiconductor Devices*, Wiley-IEEE Press, 2009

[55] J. D. Weiland, D. J. Anderson and M. S. Humayun, "In vitro electrical properties for iridium oxide versus titanium nitride stimulating electrodes", *IEEE Trans. Biomed. Eng.*, vol. 49, pp.1574–9, 2002
[56] H. Hammerle, K. Kobuch, K. Kohler, W. Nisch, H. Sachs and M. Stelzle, "Biostability of micro-photodiode arrays for subretinal implantation", *Biomaterials*, vol. 23, pp. 797–804, 2002
[57] L. Hesse, T. Schanze, M. Wilms and M. Eger, "Implantation of retina stimulation electrodes and recording of electrical stimulation responses in the visual cortex of the cat", *Graef. Arch. Clin. Exp.,*vol. 238, pp. 840–5, 2000
[58] X. Beebe and T. Rose, "Charge injection limits of activated iridium oxide electrodes with 0.2 ms pulses in bicarbonate buffered saline (neurological stimulation application)", *IEEE Transactions on Biomedical Engineering,* vol. 35, pp. 494-495, 1988
[59] S. Cogan et al., "Potential-Biased, Asymmetric Waveforms for Charge-Injection with Activated Iridium Oxide (AIROF) Neural Stimulation Electrodes", *IEEE Trans Biomed Eng,* vol. 53, pp. 327-332, 2006
[60] S. F. Cogan, "In vivo and in vitro differences in the charge-injection and electrochemical properties of iridium oxide electrodes", *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine & Biology Society,* vol. 1, pp. 882-5, 2006
[61] S. F. Cogan*, et al.*, "The influence of electrolyte composition on the in vitro charge-injection limits of activated iridium oxide (AIROF) stimulation electrodes", *Journal of Neural Engineering,* vol. 4, pp. 79-86, 2007
[62] S. Negi et al., "Neural electrode degradation from continuous electrical stimulation: Comparison of sputtered and activated iridium oxide", *Journal of Neuroscience Methods,* vol. 186, pp. 8-17, 2010
[63] E. Slavcheva et al., "Sputtered iridium oxide films as charge injection material for functional electrostimulation", *Journal of the Electrochemical Society,* vol. 151, pp. E226-E237, 2004
[64] J. Mozota and B. E. Conway, "Surface and bulk processes at oxidized iridium electrodes: 1. Monolayer stage and transition to reversible multilayer oxide film behavior", *Electrochim. Acta.,* vol. 28, pp. 1–8,1983
[65] L. S. Robblee, J. Mchardy, J. M. Marston and S. B. Brummer, "Electrical-stimulation with Pt electrodes: 5. The effect of protein on Pt dissolution", *Biomaterials*, vol. 1, pp. 135–9, 1980
[66] T. L. Rose and L. S. Robblee, "Electrical-stimulation with Pt electrodes: 8. Electrochemically safe charge injection limits with 0.2 Ms pulses", *IEEE Trans. Biomed. Eng.*, vol. 37, pp. 1118–20, 1990
[67] J. Loudin, R. Dinyari, P. Huie, A. Butterwick, P. Peumans and D. Palanker, "High resolution optoelectronic retinal prosthesis", *Proc. of SPIE,* vol. 7163, pp. 71631B-4, 2009
[68] R. J. Jensen and J. F. Rizzo III, "Responses of Rabbit Retinal Ganglion Cells to Electrical Stimulation With a Subretinal Electrode", *Invest. Ophthalmol. Vis. Sci.*, 46: E-Abstract 5279, 2005
[69] R. J. Jensen and J. F. Rizzo III, "Thresholds for activation of rabbit retinal ganglion cells with a subretinal electrode", *Exp. Eye. Res.*, vol. 83(2), pp. 367-73, 2006

[70] A. Stett, A. Mai and T. Herrmann, "Retinal charge sensitivity and spatial discrimination obtainable by subretinal implants: key lessons learned from isolated chicken retina", *J. Neural Eng.*, vol. 4, pp. S7-S16, 2007
[71] R. J. Jensen and J. F. Rizzo III, "Activation of ganglion cells in wild-type and rd1 mouse retinas with monophasic and biphasic current pulses", *J. Neural Eng.*, vol. 6, pp.1-7, 2009
[72] G. V. Morozov et al., "Multiple reflection method for electromagnetic waves in layered dielectric structures", Quantum Electronics, vol. 31(9), pp. 767-773, 2001
[73] J. D. Plummer, Michael D. Deal and Peter B. Griffin, *Silicon VLSI Technology Fundamentals, Practice and Modeling*, Prentice Hall, 2000
[74] B. E. Deal, "Standardized Terminology for Oxide Charges Associated with Thermally Oxidized Silicon," *IEEE Trans. Elec. Dev.*, vol. ED-27, pp. 606, 1980
[75] W. A.Pliskin and E. E. Conrad, "Nondestructive Determination of Thickness and Refractive Index of Transparent Films," *IBM J. Res. and Dev.*, vol. 8, pp. 43, 1964
[76] B. E. Deal and A. S. Grove, "General Relationship for the Thermal Oxidation of Silicon," *J. Appl. Phys.*, vol. 36, pp. 3770, 1965
[77] A. Reisman, E. H. Nicollian, K. C. Williams, and C. J. Merz, "The Modeling of Silicon Oxidation from $1 \times 10^{-5}$ to 20 Atmospheres," *J. Elect. Mat.*, vol. 16, pp. 45, 1987
[78] C. J. Han and C. R. Helms, "Parallel oxidation Mechanism for Si Oxidation in Dry $O_2$," *J. Electrochem. Soc.*, vol. 134, pp. 1297, 1987
[79] R. Ghez and Y. J. van der Meulen, "Kinetics of Thermal Growth of Ultra-Thin layers of $SiO_2$ on Silicon, part I: Experiment and Part II: Theory," *J. Electronchem. Soc.*, vol. 119, pp. 530 and 1100, 1972
[80] H. Z. Massoud, J. D. Plummer, and E. A. Irene, "Thermal Oxidation of Silicon in Dry Oxygen: Growth-Rate Enhancement in the Thin Regime I. Experimental Results, II. Physical Mechanisms," *J. Electrochecm. Soc.*, vol. 132, pp. 2685 and 2693, 1985
[81] D. B. Kao, J. P. McVittie, W. D. Nix, and K. C. Saraswat, "Two-Dimensional Thermal Oxidation of Silicon---I. Experiments," *IEEE Trans. Elec. Dev.*, vol. ED-34, pp. 1008, 1987
[82] D. B. Kao, J. P. McVittie, W. D. Nix, and K. C. Saraswat, "Two-Dimensional Thermal Oxidation of Silicon---II. Modeling Stress Effects in Wet Oxides," *IEEE Trans. Elec. Dev.*, vol. ED-35, pp. 25, 1988
[83] F. A. Trumbore, "Solid Solubilities of Impurity Elements in Germanium and Silicon," *Bell System Technical Journal*, vol. 39, pp. 205, 1960
[84] D. K. Schroder, *Semiconductor Material and Device Characterization*, John Wiley & Sons, 1990
[85] Virginia Semiconductor, Inc., "Wet-Chemical Etching and Cleaning of Silicon", www.virginiasemi.com, 2003
[86] K. Sato et al., " Characterization of orientation-dependent etching properties of single-crystal silicon: effects of KOH concentration", *Sensors and Actuators A* vol. 64, pp. 87-93, 1988
[87] M. Shikida, K. Sato, K. Tokoro, D. Uchikawa, Dept. of Micro Systems Engineering, Nagoya University, Japan
[88] Stanford Nanofabrication Facility, "Lift Off Procedures", https://snf.stanford.edu/SNF/

[89] Cooper et al., “Conformal Photoresist Coatings for High Aspect Ratio Features”, SUSS MicroTec, Waterbury Center, 2007
[90] C. Brubaker, M. Wimplinger, P. Lindner, and S. Pargfrieder, “Investigating the use of spray-coating technology in MEMS applications”, *Micro Magazine*, pp. 45-55, 2004
[91] L. Fuller and S. Bolster, “Introduction to ASML PAS 5500 Wafer Alignment and Exposure” , Microelectronic Engineering, *Rochester Institute of Technology*, 2011
[92] F. Lärmer, A. Schilp, Method of Anisotropically Etching Silicon, German Patent DE4241045C1, USA patents 4855017 and 4784720.
[93] A. M. Hynes, H. Ashraf, J. K. Bhardwaj, J. Hopkins, I. Johnston, J. N. Shepherd, “Recent advances in silicon etching for MEMS using the ASE$^{TM}$ process”, *Sensors and Actuators* vol. 74, 13-17, 1999
[94] H. Morioka, D. Matsunaga, and H. Yagi, “Suppression of notching by lowering the bias frequency in electron cyclotron resonance plasma with a divergent magnetic field”, *J. Vac. Sci. Technol.* A vol.16(3), 1998
[95] S. Munro, “SOI Wafer Notch Reduction using the Low Frequency Pulsing Option on the STS ICP-RIE”, NanoFab, University of Alberta, 2009
[96] J. Li, Q. X. Zhang, A. Q. Liu, W. L. Goh, and J. Ahn, “Technique for preventing stiction and notching effect on silicon-on-insulator microstructure”, *J. Vac. Sci. Technol.* B vol. 21(6), 2003
[97] J. Ohara, K. Kano, Y. Takeuchi, N. Ohya, Y. Otsuka and S. Akita, “A new deep reactive ion etching process by dual sidewall protection layer”, Research Laboratories, Denso Corporation, Japan, pp. 277-282, 2000
[98] M. Boufnichel, P. Lefaucheux, S. Aachboun, R. Dussart and P. Ranson, “Origin, control and elimination of undercut in silicon deep plasma etching in the cryogenic process”, *Microelectronic Engineering* vol. 77, pp. 327-336, 2005
[99] A. M. Litke et al., “What Does the Eye Tell the Brain?: Development of a System for the Large-Scale Recording of Retinal Output Activity”, *IEEE T. Nucl. Sci.* vol. 51(4), 1434-40, 2004
[100] K. Mathieson et al., “Large-area microelectrode arrays for recording of neural signals”, *Ieee Transactions on Nuclear Science,* vol. 51, pp. 2027-2031, 2004
[101] “Active implantable medical devices”, in *Directive 90/285/EEC,* ed: 2004
[102] W. Rohsenow et al., *Handbook of Heat Transfer Fundamentals*, New York: McGraw-Hill, 1985
[103] H. Sailer, et al., “Investigation of thermal effects of infrared lasers on the rabbit retina: a study in the course of development of an active subretinal prosthesis”, *Graefes Archive for Clinical & Experimental Ophthalmology,* vol. 245, pp. 1169-78, 2007
[104] ANSI, “American National Standard for safe use of lasers (ANSI 136.1)”, vol. ANSI 136.1-2000, ed: The Laser Institute of America, 2000
[105] F. DeLor, et al., “Maximum permissable exposures for ocular safety (ANSI 2000), with emphasis on ophthalmic devices”, *Journal of the Optical Society of America* A, vol. 24, pp. 1250-1265, 2007
[106] K. Mathieson, J Loudin, G Goetz, P Huie, L Wang, T Kamins, L Galambos, R. Smith, J S Harris, A Sher and D Palanker, “Photovoltaic retinal prosthesis with high pixel density”, *Nature Photonics,* vol. 6, pp. 391–7, 2012

[107] G. M. Hale and M. R. Querry, “Optical Constants of Water in the 200-nm to 200-μm Wavelength Region”, *Appl.Opt.*, vol. 12, pp. 555-63, 1973
[108] M. C. Bourne, D. A. Campbell and K. Tansley, “Hereditary degeneration of the rat retina”, *Br. J. Ophthalmol.,* vol. 22, pp. 613–23, 1938