Germanium on Silicon Heteroepitaxy for High Efficiency Photovoltaic Devices

by

Brian Ross Albert

B.S., Columbia University (2010)

Submitted to the Department of Materials Science and Engineering
on March 4, 2016, in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Materials Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2016

© Massachusetts Institute of Technology 2016. All rights reserved.

Author . . . . . .
Department of Materials Science and Engineering
March 4, 2016

Certified by . . . . . .
Jurgen Michel
Senior Research Scientist
Thesis Supervisor

Certified by . . . . . .
Lionel C. Kimerling
Thomas Lord Professor of Materials Science and Engineering
Thesis Supervisor

Accepted by . . . . . .
Donald R. Sadoway
Chair, Departmental Committee on Graduate Students

Germanium on Silicon Heteroepitaxy for High Efficiency Photovoltaic Devices

by

Brian Ross Albert

Submitted to the Department of Materials Science and Engineering
on March 4, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Materials Science and Engineering

Abstract

Optoelectronic devices based on III-V direct gap semiconductors enable efficient energy conversion for photovoltaic cells, light emission for LEDs, and on-chip communication via various microphotonic components. However, widespread adoption of III-V solar cells is limited by the expensive Germanium and III-V standard substrates required, while monolithic integration of III-V devices with Silicon CMOS circuitry is not yet well established. III-V solar cell cost reduction and direct Si/III-V integration can both be realized by depositing a thin layer (e.g. 1 μm) of high quality Ge on relatively inexpensive Si substrates for which the decreased cost is due to Si's greater material abundance and larger possible wafer diameters. Efficient device performance will be retained if the Ge layer maintains a sufficiently low threading dislocation density (TDD) that does not adversely effect carrier lifetimes in epitaxially deposited III-V layers that inherit the Ge film's TDD. Assuming recombination at dislocations is carrier diffusion limited, an acceptable limit for most applications is below 10^6 cm^{-2} due to typical minority carrier diffusion lengths of ~10 μm in III-V materials. However, direct deposition of Ge on Si will initially generate a TDD as high as 10^{12} cm^{-2} to plastically relax the 4.2% lattice mismatch between the two materials. State of the art approaches can reduce the TDD in large-area films to 10^6 cm^{-2} by including a 10^+ μm thick SiGe compositionally graded buffer, while TDD reduction in thinner films (e.g. 1 μm) is limited to 10^7 cm^{-2} after cyclic annealing which enhances dislocation fusion and annihilation reactions. By introducing Ge film edges spaced approximately 10 μm apart to serve as dislocation sinks during dislocation glide, the TDD has been reported to further decrease to 2.3×10^6 cm^{-2} in 1 μm thick patterned Ge. However, these films are limited to areas too small for photovoltaic cells, and the sinks appear ineffective for thread reduction at the edges of faceted, selectively grown Ge. Thus, no solution has previously existed for a thin Ge-on-Si film grown over large areas that achieves a TDD of 10^6 cm^{-2} or below. This thesis first explores the limitations to dislocation reduction by sinks in selectively-grown Ge and provides structure and fabrication modifications to enable patterned Ge films with a TDD below 10^6 cm^{-2} throughout the patterned region. To use these films for large-area applications, overgrowth and coalescence of patterned Ge films are then evaluated in different pattern designs to determine the structures that optimize coalescence in terms of throughput as well as simultaneously avoid generation of additional defects as a result of coalescence.

TDD reduction in patterned Ge films by glide to film edges requires uniform resolved shear stresses and minimum dislocation pinning during cyclic annealing. Because film facets allow for elastic relaxation of the applied thermal strain, the process of selective growth must be reversed: blanket Ge is to be grown instead to avoid faceting, followed by sidewall etching and filling before the cyclic anneal. Thermal expansion mismatch between Ge and the sidewall causes undesirable shear stress components while repulsive image forces are created if the sidewall surface's shear modulus is greater than that of Ge. Therefore, the ideal sidewall is primarily composed of Ge, separated from the primary Ge film by a thin SiO_2 layer. Monte Carlo simulations of dislocation glide were developed to estimate the limitations of glide due to the pinning effect of orthogonal dislocations. For small mesa widths w (or more generally, the spacing between adjacent dislocation sinks), TDD was found to scale with w^a with $a \sim 4$. The threshold of the small width regime and the value of a both increase for greater applied thermal stresses and thicker Ge films.

Due to the high surface energy of the Ge/SiO_2 interface, lateral overgrowth and film coalescence do not readily occur. The rate was observed to strongly correlate with the Ge film perimeter concavity, delayed at convex mesa corners while relatively promoted at the ends of isolated SiO_2 lines surrounded by a concave Ge film perimeter. Ge mesa arrays were staggered to eliminate regions entirely dependent on overgrowth from mesa corners, decreasing the growth time until complete coalescence by at least 50% as compared to a regular gridded array. The faster overgrowth rates over isolated SiO_2 lines was observed to further increase for lines of reduced widths. Due to the facets that develop, orientation of SiO_2 lines relative to intersections of $\{111\}$ planes with the substrate surface further affected overgrowth rates which maximized for slight offsets below 15°.

Etch pit studies of coalesced, selectively-grown Ge films around SiO_2 sidewalls indicated a maximum TDD above the SiO_2 (6×10^7 cm^{-2} for staggered grids) while decreasing to 10^7 cm^{-2} further away in the film. As predicted by modeling, the dislocation pile-up near SiO_2 walls was due to inverted resolved shear stress and the reduced thickness at the Ge film edge. Significant improvement in TDD reduction is expected by these models if blanket Ge is instead grown, followed by etch and fill of sidewalls with additional Ge separated by a thin layer of SiO_2. While fabrication is more involved compared to the selective growth process, the structure will be successful at threading dislocation removal. With isolated line film edges of minimal width, oriented $\sim 5°$ from $\{111\}$ surface intersection directions, the coalescence rate will be maximized. Coalescence-induced defects resulting from lattice misregistry over the SiO_2-coated Ge lines will be prevented as the Ge film is continuous at the line ends prior to overgrowth initiation. Assuming a pinning probability of 50%, a Ge film 1 μm thick with a maximum distance between dislocation sinks < 6 μm is expected to exhibit a TDD of 10^5 cm^{-2}. At this density level, the performance of III-V devices will be unaffected, enabling both lower cost high efficiency III-V solar cells and LEDs as well as III-V/Si monolithic device integration. The multiple perspectives of analysis examined in this thesis are not limited to Ge-on-Si and can readily be applied to other high lattice-mismatched materials systems to obtain a low TDD surface in large areas while maintaining a buffer layer of minimal thickness.

Thesis Supervisor: Jurgen Michel
Title: Senior Research Scientist

Thesis Supervisor: Lionel C. Kimerling
Title: Thomas Lord Professor of Materials Science and Engineering

Acknowledgements

No endeavor can rely on the effort of one person, and the past several years of work, part of which is presented in this thesis, is no exception. I'd first like to acknowledge my thesis advisors, Prof. Lionel C. Kimerling and Dr. Jurgen Michel, whose guidance directed my learning and research path. During meetings with Kim, he always found ways to challenge my thinking at the time, whether it was a modeling concept I had overlooked, an experiment I did not believe was possible to conduct, or a method of analysis I did not previously consider. Often times, Kim's ability to think outside of the box let to crucial readjustments of the research direction. Jurgen provided more detail-oriented guidance, probing the intricacies of research results, suggesting variations to experimental designs, or pointing out inaccuracies in my presentation. Jurgen's discussions related to solar-related technologies provided me with a broader understanding of the practical issues relevant to solar in the field and allowed us to identify how best to address those challenges. Hearing the critiques of my work from both Jurgen and Kim was very constructive, especially when they would disagree with each other. I also had the pleasure to work under Kim for several years as a teaching assistant for his undergraduate subject, Principles of Engineering Practice. Kim not only provided the freedom to construct a new component of the subject from scratch but offered valuable pedagogical guidance as well. Several of the concepts and laboratories added to the curriculum were reflected in my own research areas, allowing the teaching experience to complement my research component of graduate school.

Many of the people with whom I interacted on a day-to-day basis were the post-docs and fellow graduate students in the Electronic Materials Research Group, better known as EMAT. In the first half of my graduate school career, Dr. Lirong Broderick was the person I would often first consult, whether it was a question related to working in the Microsystems Technology Laboratories (MTL), solar cell theory, or the many projects we directly worked on together. Her enthusiasm and dedication for research and always asking the "what-ifs" made our discussions worthwhile. Dr. Yan Cai, Wei Yu, Brian Pearson, and I worked together as fellow teaching assistants over several years, relying on each other to make sure lab sessions moved along smoothly despite often highly constrained situations. During the latter half of my time in graduate school, I interacted more with Brian P. who would often offer useful critiques of my research ideas or results whenever I approached

him with something new. Maintaining the UHVCVD Ge growth reactor, fixing issues when they arose, would not have been possible without his help. Assistance in the Substrate Engineering Lab (SEL) provided by several graduate students in the Fitzgerald group is also appreciated. Corentin Monmeyran briefly worked with me when he first began at MIT. During the duration of that initial project and even more so after I moved to his office, during our discussions, his questions, coming from a strong theoretical background, often lead me to realize that I did not understand concepts as well as I thought, helping me to fill in those gaps soon after. While I did not extensively work with Dr. Vivek Singh, Neil Patel, Zhaohong Han, or Derek Kita during my time at MIT, I enjoyed speaking with them whether the topic of conversation was related to research, EHS issues, or something more mundane.

Reliance on other people at MIT was not limited to those already mentioned. Dr. Anuradha Agarwal, another EMAT PI, was the first person I met within the research group in 2009. Her enthusiasm for research initially drew me to EMAT for an undergraduate research summer internship under her supervision. While my research during graduate school did not coincide with her research projects, Anu always provided useful feedback on presentation style and sought to maintain a cordial work environment within EMAT, making graduate school all the more bearable. Lisa Sinclair, EMAT's administrator, allows us to breathe a bit more easily by making sure nothing fell through the cracks and always offering the chance for a quick chat or a few goodies to eat. The encouragement and interest in my research by Mark Beals of the Materials Processing Center during my entire time at MIT has also been appreciated over the years. All of the experimental components of this thesis would not have been possible without the continued maintenance of the semiconductor fabrication tool set at MTL by its staff, including but not limited to Bernard Alamariu, Paul Tierney, Bob Bicchieri, Eric Lim, Donal Jamieson, Paudely Zamora, and Kurt Broderick. Kurt, in particular, was always helpful in accommodating Kim's class in the Experimental Materials Laboratory as well entertaining unconventional solutions to materials processing issues related to my research.

Finally, I thank my friends and family, particularly my parents, for providing unconditional support not only during my years in graduate school but for all previous chapters in life. I could not have done it without you.

"Perfection has one grave defect:

it is apt to be dull."

- W. Somerset Maugham

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Solar energy provides the world with an average of 1.2×10^8 GW of power.[1] Society has always depended on this source of energy either directly (e.g. agriculture) or indirectly via conversion and temporary storage (e.g. fossil fuels). Since the industrial revolution, energy consumption has grown exponentially as people become increasingly reliant on energy intensive technologies including manufacturing, rapid transportation, residential appliances, and most recently, telecommunications. In 2014, 86% of the global energy demand of 17,000 GW was met by consumption of oil, natural gas, and coal, the reserves of which were expected to last 52, 54, and 110 years, respectively, based on 2014 consumption rates.[2] Even if additional reserves are found in the near future, the combustion of these fuels globally emits over 34 billion tons of CO_2 per year into the atmosphere[3] and is identified as a leading cause of climate change.[4]

Solar energy can provide an energy source alternative that will last beyond Earth's lifetime and do so without directly emitting greenhouse gases. Even if only 0.015% of incident solar energy were captured, the current global energy demand would be satisfied. In 2014, electricity generated from solar energy either directly via photovoltaic cells or initially stored as thermal energy for delayed generation accounted for approximately 1.0% of total global energy generation.[5] In order for solar energy to continue its relative growth within the energy sector, system conversion efficiencies must be maximized while maintaining competitive manufacturing costs. The intermittent supply of solar power must also be addressed. While crystalline Silicon (Si) solar cells have proven to be scalable and cost effective,[6] crystalline Si cells fabricated today have begun to reach their fundamental efficiency limit of approximately 25%[7] (without light concentration) determined by the detailed balance analysis of Shockley and Quissier[8] when accounting for Si's low external radiation efficiency.[9]

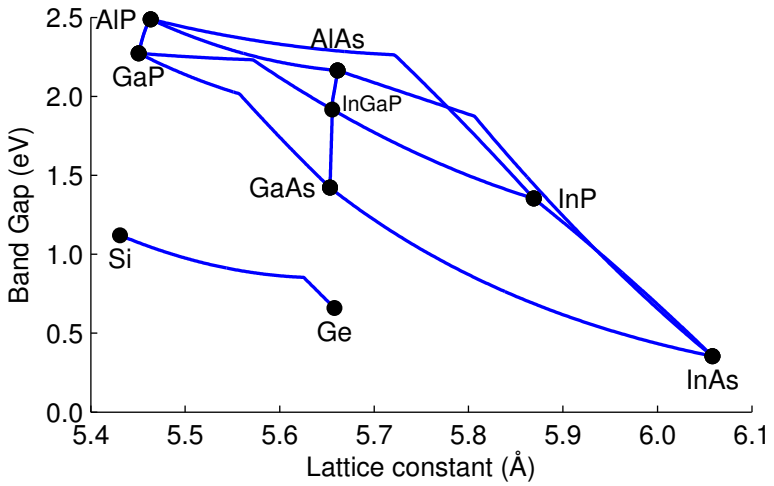


Figure 1.1: Band gap vs. lattice constant for $Si_{1-x}Ge_x$ and III-(As,P) compounds.

Other low-cost photovoltaic materials systems, including microcrystalline and amorphous Si (a-Si), dye-sensitized, chalcogenides (e.g. CIGS, CdTe, CZTS), organics, and perovskites, have not yet reached the efficiency of c-Si.[10] While currently expensive per cell area, GaAs single-junction cells do provide a modest efficiency improvement over c-Si, and III-V based multijunction tandem cells surpass the efficiency limitation of single junction cells by incorporating semiconductors of varying band gaps.[11] However, wide-scale adoption is limited due to both high manufacturing costs resulting primarily from III-V film growth and the substrate as well as the optical concentration and sun tracking systems required to offset the high cost of the cells. The III-V films are deposited as single crystals, having no grain boundaries, in order to maximize cell efficiency. In order to obtain single-crystal films as well as prevent the formation of crystalline defects, the films must be deposited on a single-crystalline substrate with the same crystal lattice spacing as the films to be deposited. For the most common tandem cell architecture, all layer materials nearly share the lattice constant of Germanium (Ge) and Gallium Arsenide (GaAs), appearing directly above the data point for Ge in Figure 1.1 of band gap vs. lattice constant for several materials systems. The typical triple junction, lattice-matched tandem cell consists of junctions composed of Ge, GaAs, and $In_{0.49}Ga_{0.51}P$, for which the specific composition is also indicated in Figure 1.1. Single-crystalline Ge[12] and GaAs[13] substrates account for approximately 50% of fabrication costs for III-V tandem cells.[14]

1.2 Outline of Thesis

This thesis explores efforts to lower the costs of III-V photovoltaic cells, focusing on the semiconductor substrate used to fabricate the solar cell structure. By replacing the Ge substrate with a "virtual substrate" of a thin Ge film deposited on relatively inexpensive Si wafers, high efficiency III-V solar cells will become economically viable for large-scale deployment. Because the lattice constants of Si (a = 5.431 Å) and Ge (a = 5.658 Å) differ by 4.2%, the process of growing single crystal Ge on single crystal Si must carefully minimize the number of crystalline defects that result on the Ge film surface. Requirements for the design, fabrication, and processing of Ge virtual substrates of high material quality are identified through both simulation and experiment. A method to incorporate the low-cost III-V photovoltaic cells utilizing these substrates in a spectrum splitting, hybrid solar energy conversion system is also described as well as a derivative of spectral filter design adapted to the application of structural color.

Chapter 2 reviews both the semiconductor physics relevant to solar cell operation as well as heteroepitaxy, the growth of single-crystalline material on a seed layer of another material. The generation of one-dimensional crystal lattice defects, referred to as dislocations, resulting from lattice-mismatch heteroepitaxy, are described along with previous efforts from literature to reduce the threading dislocation density in epitaxially grown films. The adverse affects of dislocations on solar cell efficiency are also reviewed to identify the target defect density for different devices.

Chapter 3 describes the ultra-high vacuum chemical vapor deposition of Ge-on-Si process, including the pre-epitaxy process flow necessary for selective growth. The method of selective epitaxial growth is used to fabricate Ge mesas array solar cells on Si substrates. Fabrication and device limitations that arise from such a structure are discussed. The growth of a planar Ge virtual substrate is identified as a more promising photovoltaic cell application of Ge-on-Si film growth.

Chapter 4 analyzes finite-element thermal stress simulations to provide insight to which structures surrounding patterned Ge films retain the necessary stress fields during cyclic annealing to enable dislocation glide to the Ge film edges. Simulations of dislocation glide in reduced areas provide additional requirements on the Ge film to avoid dislocation pinning during glide. Fabricated structures that satisfy both sets of requirements laid out should be most successful at producing threading dislocation free material.

In Chapter 5, selective growth is explored in more detail, examining the effects of selective growth window patterns on Ge epitaxial growth and film coalescence, to ultimately create a blanket Ge film with the low defect density of the patterned Ge films. Requirements for structures that enable complete film coalescence are identified. Fabrication of Ge lattice-matched $In_{0.49}Ga_{0.51}P$ single junction solar cells is demonstrated on Ge-on-Si virtual substrates along with evaluation of the effects the dislocation density on InGaP cell efficiency.

Chapter 6 reviews multijunction solar cells and describes a hybrid approach, combining photovoltaics and solar thermal components, with a greater solar to electrical energy conversion efficiency with the added benefit of energy storage. The efficiency of the hybrid system using spectrum splitting is estimated. The design of spectral filters designed to partition the solar spectrum for use in the hybrid system is also discussed. Similar design methodologies are applied to the design of saturated color reflectors, working primarily on thin film interference effects. Design rules to maximize saturation while minimizing angular dependence are identified.

The thesis concludes in Chapter 7 with a summary of the work presented and discussion of further work to be conducted on Ge-on-Si heteroepitaxy to realize the fabrication of low TDD, large area Ge-on-Si films.

Chapter 2

Background

2.1 Solar cell operation

Sunlight provides the earth's surface with a broadband spectrum of photon energies, ranging up to 4.0 eV as shown in the AM1.5 spectra in Figure 2.1. In order to convert this energy to electricity via the photovoltaic effect, semiconductor materials are used to absorb photons, imparting their energy to the creation of excited electron-hole pairs. Charge separation in solar cells is typically achieved by use of a p-n junction. In most solar cell architectures, the junction is composed of a thin, highly doped emitter layer and a relatively lower doped thick base layer. Carriers collected at the metal contacts will induce current flow through an external circuit. The complete process is schematically shown in Figure 2.2.

When current passes through an external load, the voltage drop across the load effectively forward biases the solar cell's p-n junction. The current-voltage characteristics of a solar cell can

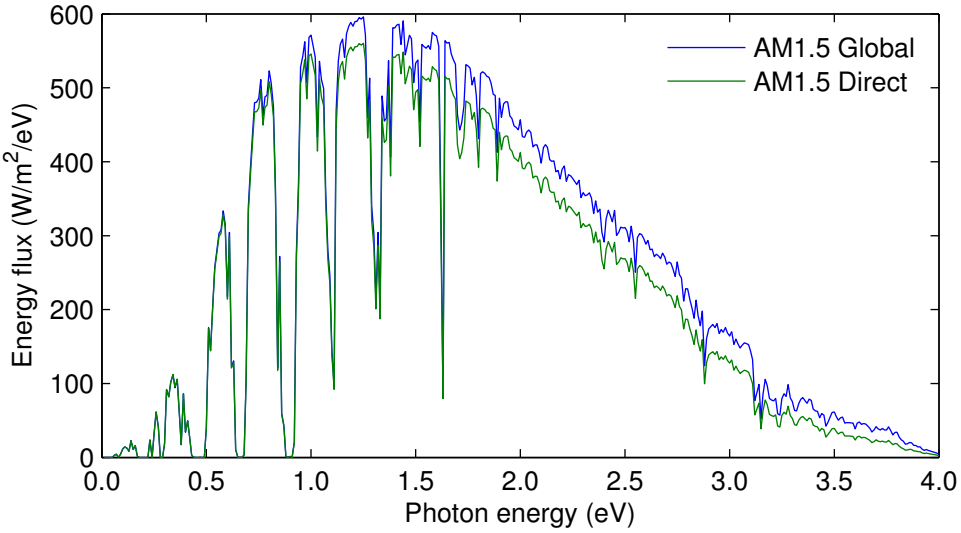


Figure 2.1: Energy flux for AM1.5 Global and Direct solar spectra.[15,16] The global spectrum contains both diffuse and direct sunlight, applicable to non-concentration systems. The direct spectrum only contains the portion of sunlight within a 5° diameter field of view of the sun.

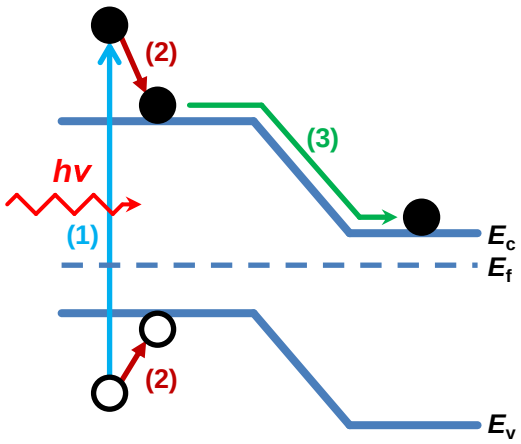


Figure 2.2: Schematic of steps in the photovoltaic conversion of incident light into electricity in a p-n junction: (1) photon absorption, (2) thermalization, (3) charge separation. Lines representing E_c and E_v only indicate the single energy levels of the valley and peak of the respective bands that correspond to E_g. The initial energy states of the electron and hole shown do exist at allowed $E-k$ states which aren't shown for diagram simplicity.

therefore be modeled as a current source in parallel with a diode and its shunt resistance as well as a device resistance in series:

$$J = -J_L + J_0\left[\exp\left(\frac{q(V - Jr_s)}{nk_BT}\right) - 1\right] + \frac{V - Jr_s}{r_{sh}} \tag{2.1}$$

where J_L is the photo-generated current density, J_0 is the diode saturation current density, V is the applied voltage across the solar cell (as determined by the external load or testing apparatus), T is the device temperature, and n is the diode non-ideality factor, typically between 1 and 2, depending on the temperature, semiconductor material, and operating voltage. In an ideal cell, the specific series resistance $r_s = 0$ while the specific shunt resistance $r_{sh} = \infty$.

To operate a solar cell, the voltage is chosen by modifying the effective load to maximize the total power output, $P = IV$. A typical current-voltage (I-V) relationship of a solar cell can be seen in Figure 3.9. The operating voltage of the cell, V_m, is typically 100 mV below the V_{oc} while the current density $J_m \approx J_L$ for ideal devices. The I-V operating point is often compared to the V_{oc} and J_{sc} by definition of the fill factor:

$$FF = V_m J_m / V_{oc} J_{sc} \tag{2.2}$$

The fill factor typically ranges from 0.6 to 0.9, which can be accurately predicted for an ideal device simply by the cell's normalized V_{oc}, $v_{oc} \equiv qV_{oc}/nkT$:[17]

$$FF \approx \frac{v_{oc} - \ln(v_{oc} + 0.72)}{v_{oc} + 1} \tag{2.3}$$

While increasing E_g will increase V_{oc}, doing so will reduce the total photocurrent the cell can produce because the semiconductor will become transparent to a larger portion of the solar spectrum. Because the power extracted by a solar cell is determined by the product of the cell's operating current and voltage, a trade-off appears in cell efficiency vs. E_g, often referred to as the Shockley-Queisser limit.[8] Using the method of detailed balance, the reference AM1.5G spectrum, and diode I-V characteristics, the maximum efficiency for a single junction photovoltaic cell is limited to approximately 33.7%, corresponding to a band gap of 1.34 eV. Actual devices must also consider the effects of imperfect carrier collection or absorption, contact grid shadowing, and reflection. Multijunction and hybrid solar energy conversion approaches that surpass this limitation for single junction solar cells are described in Chapter 6.

2.2 Semiconductor physics and devices

A semiconductor will only generate an electron-hole pair by absorbing a photon when the photon energy, $h\nu$, exceeds the material's band gap energy, E_g, the energy difference between the bottom of the conduction band, E_c, and the top of the valence band, E_v. More specifically, the creation of an electron-hole pair from photon absorption must satisfy both energy and momentum conservation. An electronic band structure indicates the possible electron energy levels, E_n, as a function of wavevector state, $\vec{k}$, which is proportional to the crystal momentum, $\hbar\vec{k}$. Because wavevectors are three-dimensional, electronic band structures are typically visualized along high symmetry directions in reciprocal space. The $E - \vec{k}$ diagrams in Figure 2.3 illustrate the band structures along the [111] and [100] directions for three different semiconductors commonly used for photovoltaic cells, Ge, Si, and GaAs, with room temperature E_g values of 0.67 eV, 1.12 eV, and 1.42 eV, respectively. Since the wavenumber of a typical photon from the sun, $2\pi/\lambda \sim 10^5$ cm^{-1}, is several orders of magnitude smaller than the scale of crystal wavevectors, $2\pi/a \sim 10^8$ cm^{-1},

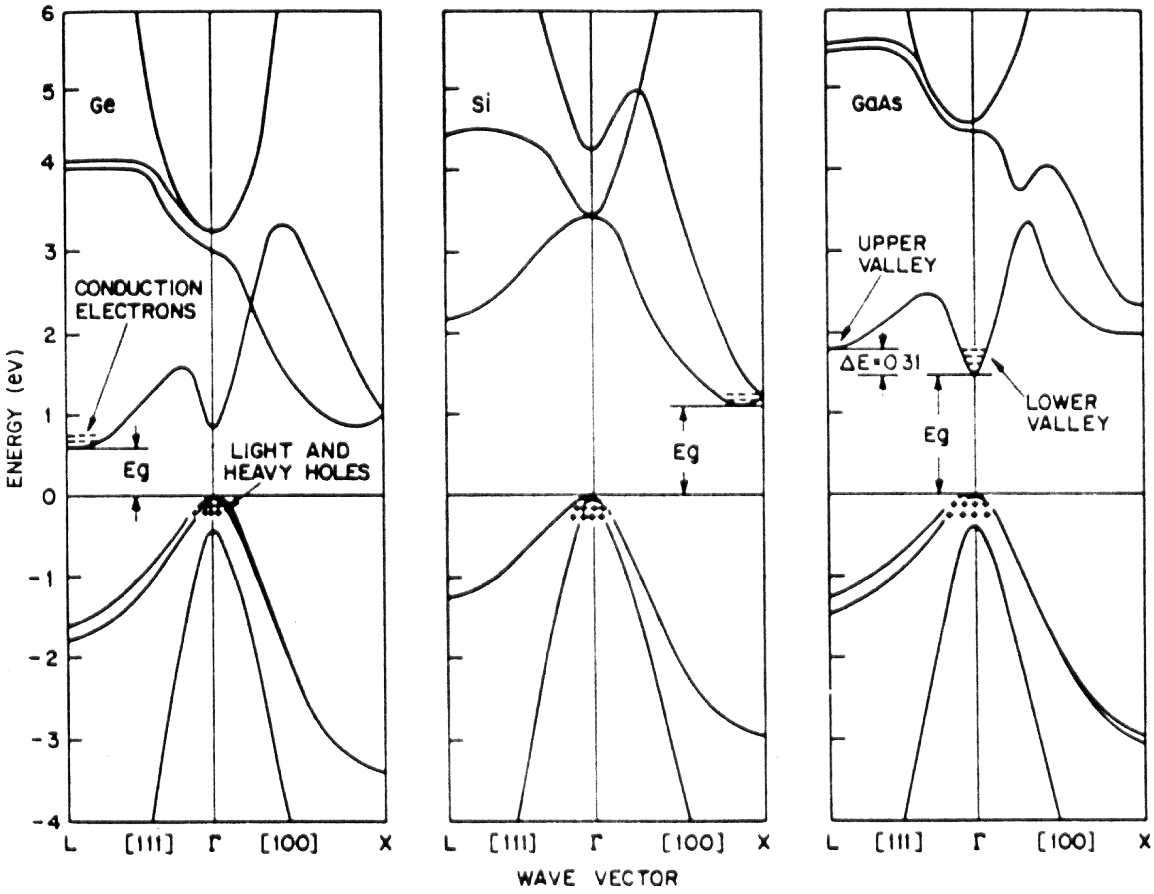


Figure 2.3: Electronic band structure for (a) Ge, (b) Si, (c) GaAs. Adapted from Chelikowsky and Cohen.[18]

where a is the crystal lattice constant, in the absence of phonon interactions, the transition of an electron from a valence state to a conduction state must primarily be vertical as visualized in an $E-k$ diagram. Further, the state in the valence band must be filled while the state directly above in the conduction band must be empty for the transition to occur.

GaAs can absorb photons with energies just above E_g without breaking conservation of momentum because the peak in its valence bands and the lowest valley in its conduction bands coincide at the Γ point, thus making GaAs as "direct" band gap material. In contrast, Si and Ge are indirect semiconductors because their lowest conduction band valley occurs away from the Γ point. In order for Si and Ge to absorb band-gap energy photons, phonons in the material must mediate the process by providing the necessary momentum change during the photon absorption event to satisfy momentum conservation. Because this process requires a three-body interaction, the probability of occurrence is very low when compared to the simpler two-body process for direct

transitions. The absorption length for near band gap photons is considerably longer for Si and Ge (e.g. 1 mm) compared to GaAs (e.g. 1 μm). While c-Si requires relatively thick films on the order of 100 μm to successfully absorb the majority of above-E_g photons in the solar spectrum, the close proximity between Ge's indirect and direct band gap of 0.80 eV allows for film thickness closer to that required for GaAs and other direct gap materials to absorb the majority of above E_g photons.

After photon absorption, excess energy $h\nu - E_g$ is lost by carrier thermalization to the conduction band valley and valence band peak. With the assistance of a built-in electric field from a p-n junction, minority carriers within their diffusion length, $L_{n,p}$, of the junction's depletion region, W, can diffuse and drift to the opposite side of the junction before recombining. The built-in field of the junction is due to the difference in the Fermi energy, E_f, between the two sides of the junction. After charge transport equalizes E_f across the junction, a built-in potential, V_{bi} is created:

$$V_{bi} = \frac{kT}{q} \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{2.4}$$

where N_A and N_D are the doping concentrations on both sides of the junction and n_i is the intrinsic carrier concentration of the semiconductor. The intrinsic carrier concentration is primarily determined by E_g as well as the effective density of states for the conduction and valence bands, N_c and N_v, and temperature:

$$n_i = \sqrt{N_c N_v} \exp\left(-\frac{E_g}{2kT}\right) \tag{2.5}$$

Thus, junctions composed of higher band gap semiconductors will generally have larger built-in potentials. The depletion region width depends on V_{bi}, the applied bias V, as well as the doping densities and permittivity, $\varepsilon_r \varepsilon_0$:

$$W = \sqrt{\frac{2\varepsilon_r \varepsilon_0}{q}\left(\frac{N_A + N_D}{N_A n_D}\right)(V_{bi} - V)} \tag{2.6}$$

$L_{n,p}$ is dependent on the the carrier diffusivity, $D_{n,p}$ and minority carrier lifetime, $\tau_{n,p}$ by the relation $L_{n,p} = \sqrt{D_{n,p}\tau_{n,p}}$. $D_{n,p}$ is primarily a function of the host semiconductor and doping density if above a certain threshold. The carrier lifetime is the characteristic time constant that describes the exponential decay of excess carriers due to recombination. τ is therefore sensitive to all possible recombination mechanisms. Trap-assisted recombination, referred to as Shockley-Read-Hall (SRH)

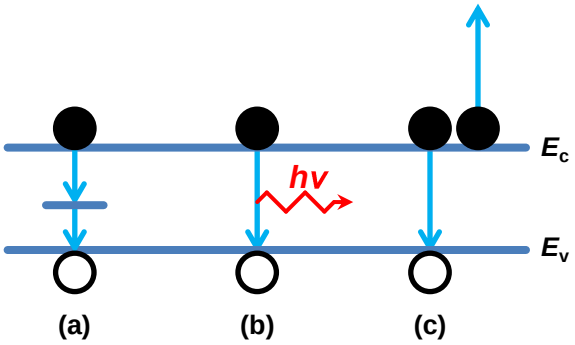


Figure 2.4: Schematic of recombination mechanisms: (a) SRH, (b) Radiative, (c) Auger.

recombination, makes use of intermediate energy levels within the band gap created by defects. Because these energy levels are localized spatially, momentum conservation between the electron and hole alone is not required and thus is the primary recombination mechanism for lightly-doped, indirect band gap semiconductors at low levels of carrier injection. Impurities, crystal defects, and unpassivated interfaces can create deep level traps within the band gap, and thus can greatly reduce τ if at high enough concentrations. In direct band gap semiconductors, in the absence of defects and lifetime killer impurities, recombination occurs mainly by a radiative process in which electrons and holes recombine and transfer their energy by spontaneous emission of a photon. Because absorption of this photon is still possible within the material in a process termed "photon recycling," radiative recombination is the ideal recombination mechanism for most optoelectronic devices.[19] At high carrier injection levels and/or doping levels, recombination can also occur by an Auger process by which an electron and hole recombine and transfer the energy to a third charge carrier (either an electron or a hole) which later relaxes by a thermalization process. These three recombination processes are schematically compared in Figure 2.4.

The overall minority carrier lifetime is determined by combining the recombination rates from each mechanism discussed previously:

$$\frac{1}{\tau} = \frac{1}{\tau_{\text{SRH}}} + \frac{1}{\tau_{\text{rad}}} + \frac{1}{\tau_{\text{Auger}}} \tag{2.7}$$

For well fabricated devices with minimal unintentional impurities and optimized surface passivation, Auger processes dominate the recombination rate in Si solar cells (due to its indirect band gap and

high doping in the emitter) while radiative recombination is the main mechanism in photovoltaic cells composed of direct gap III-V materials. In the presence of impurities or defects with deep level trap states, almost all traps will be filled with the majority carrier due to the sufficient difference between the Fermi and trap energy levels, E_f and E_T, respectively. Once a minority carrier approaches the trap, the recombination mechanism can occur, and is thus solely proportional to the minority carrier concentration. The SRH recombination rate (per minority carrier) can therefore be estimated by the thermal velocity of the minority carrier, v_{th}, the capture cross-section of the defect (σ [=] cm^2), and the defect density (ρ_D [=] cm^{-3}):

$$r_{SRH} = \frac{1}{\tau_{SRH}} = v_{th}\sigma\rho_D \tag{2.8}$$

One-dimensional crystal defects, referred to as dislocations, can also serve as efficient SRH-like recombination-generation centers. Since the trap energy levels of dislocations are typically near the center of a semiconductor's band gap, the recombination rate associated with the dislocation density (ρ_{TD} [=] cm^{-2}) can be approximated as:

$$r_{TD} = \frac{1}{\tau_{TD}} = v_{th}\sigma\rho_T\rho_{TD} \tag{2.9}$$

where ρ_T is the linear density of trap states per dislocation line length. Example values for the $Si_{0.75}Ge_{0.25}$ alloy include $v_{th} = 1.2 \times 10^7$ cm/s at room temperature, $\sigma = 4 \times 10^{-12}$ cm^{-2}, and $\rho_T = 10^6$ cm^{-1}.[20]

Alternatively, the effect of dislocations on minority carrier lifetime can be estimated by the average distance $d = \sqrt{\pi\rho_{TD}}$ between dislocations uniformly distributed with density ρ_{TD} if recombination is carrier diffusion limited. This distance corresponds to the effective diffusion length in the absence of other recombination mechanisms. Through the relation of diffusion length to carrier diffusivity and lifetime, the expected carrier lifetime as a function of ρ_{TD} can be estimated:[21]

$$\frac{1}{\tau} = \frac{1}{\tau_0} + \frac{\pi^3 D\rho_{TD}}{4} \tag{2.10}$$

where τ_0 is the minority carrier lifetime in dislocation-free material. Based on Eq. 2.10, carriers with relatively low carrier diffusivities are more resilient to the effects of dislocations on their

lifetimes. For a given dislocation-free lifetime, n-type and high E_g materials will typically require less stringent requirements on the dislocation density. It should be noted that carrier diffusivities themselves are also affected by dislocations but only at densities greater than 10^8 cm^{-2}.[21] If the effective carrier lifetime is instead determined by the density and cross-section of traps due to dislocations, the second term on the RHS of Eq. 2.10 should be replaced with Eq. 2.9. Section 2.3 discusses the origins of these dislocations relevant to the rest of the thesis.

The open circuit voltage V_{oc} of a solar cell is often used as a measure of the semiconductor junction material quality because of its close connection with junction dark current and therefore, the minority carrier lifetime:

$$V_{oc} = \frac{nk_BT}{q} \ln\left(\frac{J_L}{J_0} + 1\right) \tag{2.11}$$

J_0 represents the current density resulting from minority carriers drifting across the junction without illumination in reverse bias. The sources of minority carriers include those that diffuse to the depletion region edge (and are regenerated in the bulk) as well as carriers thermally generated within the depletion region. The former source, corresponding to the diffusion current of an "ideal" diode ($n = 1$), can be approximated from:

$$\begin{aligned} J_{0,1} &= qn_i^2\left(\frac{D_p}{L_pN_D} + \frac{D_n}{L_nN_A}\right) \\ &= qn_i^2\left(\frac{D_p^{1/2}}{\tau_p^{1/2}N_D} + \frac{D_n^{1/2}}{\tau_n^{1/2}N_A}\right) \end{aligned} \tag{2.12}$$

In the case of an asymmetric junction where $N_D \gg N_A$ or vice versa, $J_{0,1}$ is determined primarily by the minority carriers on the lightly doped side of the junction. In forward bias, the process reverses with the injection of carriers, creating a net forward dark current: $J_{diff} = J_{0,1}\left[\exp\left(\frac{qV}{kT}\right) - 1\right]$.

The second source of diode dark current, $J_{R\text{-}G}$, arises from electron-hole recombination-generation due to SRH trap states within the depletion region of width, W:

$$J_{R\text{-}G} = q\int_{-W_p}^{W_n} \frac{np - n_i^2}{\tau_p(n + n_1) + \tau_n(p + p_1)}\mathrm{d}x \tag{2.13}$$

where $W_{n,p}$ are the portions of the depletion region on each side of the junction, and p_1, n_1 are the carrier concentrations when $E_f = E_T$:

$$n_1 = n_i \exp(E_T - E_i/kT) \qquad p_1 = n_i \exp(E_i - E_T/kT) \tag{2.14}$$

In reverse bias, carriers within the depletion region can be ignored. For mid-gap states ($E_T = E_i$) , the expression in Eq. 2.13 can be simplified:

$$J_{R\text{-}G} = \frac{qn_i W}{2\tau_0} = J_{0,2} \tag{2.15}$$

where $\tau_0 = \frac{1}{2}\left(\tau_p + \tau_n\right)$. In forward bias and/or with photogeneration of carriers, integration in Eq. 2.13 to calculate $J_{R\text{-}G}$ is less straightforward because minority carriers within the depletion region must be considered.[22] τ_0 is replaced with a geometric mean of the carrier lifetimes: $\tau_0 = \sqrt{\tau_{p0}\tau_{n0}}$ and an exponential dependence on bias must be included:[23]

$$J_{R\text{-}G} = J_{0,2}\left[\exp\left(\frac{qV}{2kT}\right) - 1\right] \tag{2.16}$$

For asymmetric junctions, other modifications of the general expression for $J_{R\text{-}G}$ that depend on the energy level of the trap state are required which can include saturation of $J_{R\text{-}G}$ at high enough forward bias levels.[24]

If the minority carrier were defined solely by the local doping type, the total recombination current could be approximated by weighting the recombination rate ($1/\tau$) of the minority carrier for each side of the junction by the side's width in the depletion region:

$$J_{R\text{-}G} \approx \frac{qn_i}{2}\left(\frac{W_p}{\tau_n} + \frac{W_n}{\tau_p}\right)\left[\exp\left(\frac{qV}{2kT}\right) - 1\right] \tag{2.17}$$

Because the fraction of the depletion region on a particular side of the junction is inversely proportional to the relative doping density, the minority carrier lifetime on the lightly doped side of the junction would have a greater effect on $J_{R\text{-}G}$. In a highly asymmetric junction, this assumption suggests that only the carrier lifetime on the lightly doped side determines $J_{R\text{-}G}$.[21] In the analysis of GaAs solar cells with different dislocation densities and n^+/p and p^+/n asymmetric

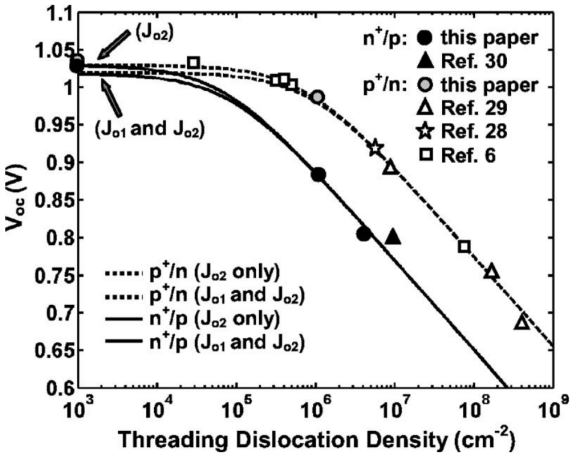


Figure 2.5: V_{oc} of n^+/p and p^+/n GaAs solar cells with different dislocation densities.[25]

junctions, only the carrier lifetime in the lightly-doped base layer was experimentally found to determine the effect of dislocations on the V_{oc} of the solar cell.[25] Because the carrier lifetime of electrons are more sensitive to dislocations due to higher carrier diffusivities,[26] GaAs cells with n-type base layers were found to have a higher dislocation density threshold before device V_{oc} began to decrease as shown in Figure 2.5. Design simulations that consider the effects of dislocations on solar cell efficiency usually assume that the lifetime of the minority carrier type in the base layer is the determining factor of the solar cell dark current.[27]

This reasoning does not fully explain the observed difference between n^+/p and p^+/n cells. Solar cell modeling software PC1D[28] is used to simulate one-dimensional GaAs solar cells composed of 0.1 μm thick emitter layers doped at 2×10^{18} cm^{-3} and 0.9 μm thick base layers doped at 1×10^{17} cm^{-3}. Both n^+/p and p^+/n cells are simulated under AM1.5G illumination at 1× concentration for various combinations of τ_n and τ_p. The lifetime for each carrier is identical in the emitter and base layers for a given simulated device. The V_{oc} values for each cell is summarized in Table 2.1. While the V_{oc} does decrease more when the lifetime of the base layer's minority carrier is reduced by a factor of 10 instead of the lifetime of the emitter's minority carrier, perfect selectivity in the response is not observed. The V_{oc} reductions differ by less than 50% with a reference $\tau_{n,p}$ = 10 ns and less than 10% with a reference $\tau_{n,p}$ = 1 ns.

Under forward bias and illumination, the carrier concentrations change exponentially with distance across the depletion region, causing electrons and holes to change roles as the minority carrier type approximately half-way in the depletion region (see Figure 2.6a). Most recombination

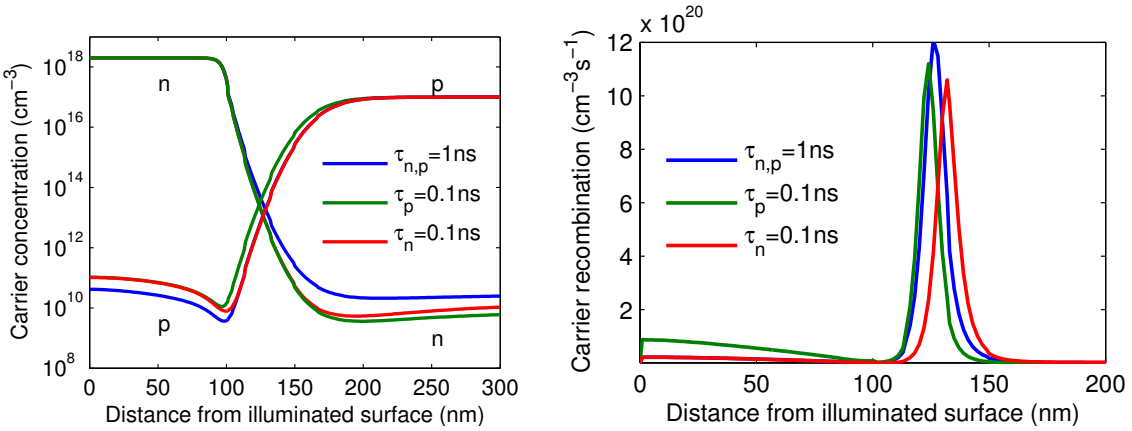


Figure 2.6: PC1D simulation results of (a) carrier concentrations and (b) carrier recombination in n^+/p GaAs solar cells with different value for $\tau_{n,p}$ illuminated and operating at their respective V_{oc} biases.

occurs in the center of the depletion region (see Figure 2.6b) because the np product is maximized there (see Eq. 2.13). While recombination mainly occurs on the lightly doped side ($x > 100$ nm), the lifetimes of both carriers are relevant as both carriers are of the minority in approximately equal regions, at least within an order of magnitude. In order to accurately predict the effect of threading dislocations on the V_{oc} of a solar cell, the recombination current must be simulated in the relevant device structure. Both τ_n and τ_p will effect the simulation result.

An additional potential explanation for the trends observed for GaAs solar cells instead considers the interactions between the dopant atoms and the dislocation core. Dislocations have been observed to getter dopants[29] and other impurities due to the disoriented bonds at the dislocation core.[30] Simulations[31,32] have indicated that certain impurities for different host materials can passivate the

τ_n	τ_p	V_{oc} (n^+/p)	V_{oc} (p^+/n)	ΔV_{oc} (n^+/p)	ΔV_{oc} (p^+/n)
10 ns	10 ns	961 mV	961 mV		
1 ns	10 ns	920 mV	901 mV	41 mV	60 mV
10 ns	1 ns	902 mV	921 mV	59 mV	41 mV
1 ns	1 ns	860 mV	857 mV		
0.1 ns	1 ns	794 mV	801 mV	66 mV	56 mV
1 ns	0.1 ns	797 mV	791 mV	63 mV	66 mV

Table 2.1: PC1D simulation results of GaAs cell V_{oc} as a function of minority carrier lifetimes.

dislocation core, rendering their recombination-generation centers inactive. Passivation of threading dislocation by P in GaAs solar cells has been observed to increase the V_{oc}.[33] Alternatively, dopant atoms can maintain the electrical activity of dislocations. In the absence of impurities, dislocations in GaAs will partially self-passivate through bond reconstruction.[34] If however, an acceptor atom (e.g. Zn) moves into a Ga dislocation or a donor atom (e.g. Te) moves into a As dislocation, bond reconstruction is disrupted due to the difference in atomic bond lengths and valence electrons. While Si is typically used as an n-type dopant, it can substitute for both Ga and As. However, its effect on bond reconstruction on either dislocation type is reduced compared to Zn and Te. If Si and Zn are used to dope n-type and p-type layers, respectively, the dislocation segments in the p-type layers should be expected to be more electrically active. Thus, a solar cell with a n-type base will have less electrically active defect sites at the dislocations, leading to less recombination in the depletion region and a higher V_{oc} as is observed experimentally. The effective distinction between n^+/p and p^+/n GaAs cells is likely not solely due to the difference in minority carrier diffusivities but rather the ability of specific dopant atoms to prevent self-passivation that would otherwise occur through bond reconstruction at the dislocation cores, reducing the overall trap density N_T in Eq. 2.9. If the n-type and p-type dopants used prevent dislocation passivation to different degrees, then the relative portion of the depletion region on each side of the junction is relevant for determining $J_{R\text{-}G}$.

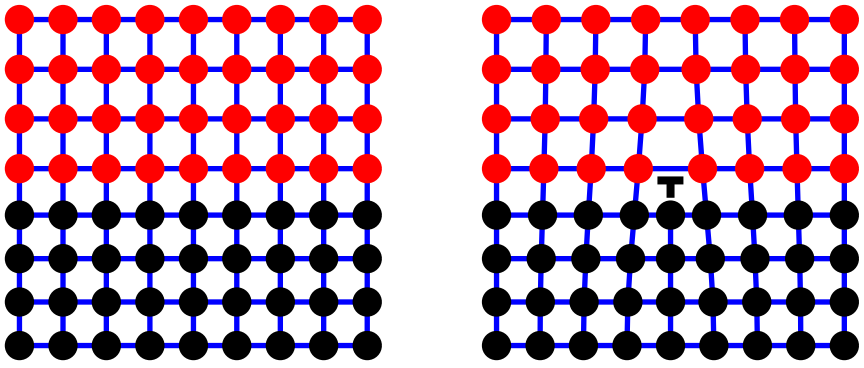
The main differences between $J_{R\text{-}G}$ and J_{diff} are the functional relationships to both n_i and $\tau_{n,p}$. Because $J_{diff} \propto n_i^2$ while $J_{R\text{-}G} \propto n_i$, the diffusion current is dominant for small band gap materials such as Ge while the recombination-generation current is mostly responsible for the I-V characteristics of large band gap materials like GaAs. Si pn junctions are controlled by the former at high forward bias while the latter at low bias due to the different ideality factors in the exponential terms. While $J_{diff} \propto \tau^{-1/2}$, $J_{R\text{-}G} \propto \tau^{-1}$. Since $V_{oc} \propto -\log(J_0)$, a proportional change of the relevant carrier lifetime(s) will produce a linear change in the V_{oc} of a cell dominated by J_{diff} that is half the change in V_{oc} for a cell dominated by $J_{R\text{-}G}$. Thus, smaller band gap solar cells should show a weaker dependence of a change in the magnitude of the dislocation density on the cell V_{oc}. At a given dislocation density however, larger E_g materials are still expected to be more resilient to dislocations due to shorter diffusion lengths. For a material with a larger value of E_g, the intrinsic carrier concentration is reduced, increasing the potential V_{oc} achievable for that material. Because the initial value for V_{oc} will be higher for larger band gap materials before consideration of the

effects of dislocations, the same linear reduction will be a larger proportional decrease for the smaller E_g cells.

2.3 Heteroepitaxy and dislocations

Epitaxy is the growth of a crystalline film on a crystalline substrate. The general growth process finds use extensive use in the fabrication of many devices that require single-crystal material, including but not limited to CMOS strained Si transistors,[35] SiGe microphotonic modulators,[36] as well as III-V optoelectronic devices such as heterostructure FETs and BJTs, solar cells,[37] and lasers.[38] When the crystal structure and lattice constant(s) of the film and substrate are similar, the crystal orientation of the film continues from the substrate's crystal structure, at least initially. One of the primary concerns in epitaxial film growth occurs for cases where lattice matching between the film and substrate is not exact. Figure 2.7a schematically shows the growth of a film with a slightly larger lattice constant than the substrate. While the film can grow with its relaxed lattice constant in the vertical direction, it will remain biaxially strained by the substrate if it continues to adopt the substrates' horizontal lattice without deviation.

One such deviation to reduce the build-up of strain is a dislocation. Dislocations are one-dimensional defects around which atom positions are distorted compared to a perfect crystal. The distortions can be due to an extra half-plane, a sheared half-plane, or aspects of both cases. The associated dislocations are referred to as edge, screw, and mixed, respectively. In the context of



(a) (b)

Figure 2.7: Schematic cross section of a heteroepitaxially grown film (a) before and (b) after introduction of a pure edge misfit dislocation in a simple cubic crystal structure.

epitaxy, dislocation segments at film interfaces can relieve misfit strain and thus are called misfit dislocations. An example of a pure edge, misfit dislocation is shown in Figure 2.7b. Because dislocations cannot terminate within a crystal, unless a misfit dislocation extends across the entire length of the film interface, both ends of a dislocation will extend to the film surface as threading segments, creating a dislocation half-loop.

A dislocation is defined by both its line direction $\vec{\xi}$ and burgers vector $\vec{b}$. The dislocation's burger vector quantifies the magnitude and direction of the distortion associated with the dislocation. Formally, it can be defined by integrating elastic displacements $\vec{u}$ on a large closed loop that encloses the dislocation:

$$\vec{b} = \oint_C \frac{\partial \vec{u}}{\partial l} \mathrm{d}l \tag{2.18}$$

In Figure 2.7b, $\vec{\xi}$ points into the page and $\vec{b}$ extends horizontally by one atomic spacing. For edge dislocations, $\vec{\xi} \perp \vec{b}$ while for screw dislocations, $\vec{\xi} \parallel \vec{b}$. The edge component of a dislocation defines a dislocation's glide plane: $\hat{n} = \vec{\xi} \times \vec{b}_{\text{edge}}$, the plane in which the dislocation line can move without involving net mass transport to or from the dislocation core. $\vec{b}$ defines the direction atoms move as the dislocation glides past. At sufficiently high temperatures, vacancy diffusion enables dislocations to extend or recede their extra half-plane, effectively moving out of their glide plane through a process called climb. Pure screw dislocations do not have single glide plane and thus can glide on multiple planes via cross-slip.[39]

Dislocation burgers vectors will correspond to displacements equal to sums of primitive lattice vectors. Because the elastic energy of a dislocation scales with b^2, dislocations will adopt the smallest primitive lattice vector in order to minimize the elastic energy of the crystal. Dislocations can also split into partial dislocations to further lower the sum: $\sum b_i^2$ but the increase in energy due to the lattice misregistry in the region between the partials, referred to as a stacking fault, forces them to remain relatively close to each other. The glide plane of dislocations is the closest-packed plane because the maximum spacing between planes lowers the stress from atoms on adjacent planes on the dislocation as it glides. For diamond cubic materials, the perfect-dislocation burgers vectors and glide planes are of the types $\frac{a}{2}\langle 110 \rangle$ and $\{111\}$ where a is the crystal lattice constant. These dislocations have been observed to dissociate into partial dislocations of the type $\frac{a}{6}\langle 211 \rangle$.[39]

If the lattice constants between a film and substrate are mismatched, the epitaxial film will initially grow psuedomorphically, adopting the substrate's crystal lattice spacing at the expense of biaxial strain energy in the film:

$$E_{\mathrm{e}} = \epsilon^2 Y h \tag{2.19}$$

where ϵ is the film strain, h is the film thickness, and Y is the Young's modulus under biaxial strain which for the (001) interface in diamond cubic materials, $Y = C_{11} + C_{12} - \frac{2C_{12}^2}{C_{11}}$. The film can reduce its strain energy by the extension of misfit dislocations at the film/substrate interface, allowing the film to relax by approaching its equilibrium lattice constant.[40] However, there is an energy cost associated with increasing dislocation line lengths, which for films with thicknesses greater than the dislocation spacing can be estimated for a misfit dislocation array in both lateral dimensions:

$$E_{\mathrm{d}} = Db^2\left(1 - \nu\cos^2\beta\right)\left[\ln\left(\frac{h}{b}\right) + 1\right]\left(\frac{f - \epsilon}{b_{\mathrm{eff}}}\right) \tag{2.20}$$

where $D = G_{\mathrm{f}}G_{\mathrm{s}}/\left[\pi\left(G_{\mathrm{f}} + G_{\mathrm{s}}\right)(1 - \nu)\right]$ is the geometric average shear modulus at the film/substrate interface, b_{eff} is the burgers vector's in-plane component that contributes to strain relaxation, ν is the poisson ratio, β is the angle between the dislocation line direction and its burgers vector, and f is the lattice misfit. $f - \epsilon$ is the extent of plastic deformation, and $(f - \epsilon)/b_{\mathrm{eff}}$ is the linear density of misfit dislocations at the interface.

At thermodynamic equilibrium, the film will minimize the sum of $E_{\mathrm{e}} + E_{\mathrm{d}}$, generating misfit dislocations only if the total energy of the system decreases. This onset of plastic relaxation occurs at a critical thickness, h_{c}, which can be determined from the following implicit equation, similar to that first suggested by Matthews:[41]

$$h_{\mathrm{c}} = \frac{Db^2\left(1 - \nu\cos^2\beta\right)\left[\ln\left(\frac{h_{\mathrm{c}}}{b}\right) + 1\right]}{2Yfb_{\mathrm{eff}}} \tag{2.21}$$

While Eq. 2.21 predicts the film thickness at which plastic relaxation begins by misfit dislocation extension at the film/substrate interface, the derivation assumes that threading dislocations already exist in the material, presumably from the substrate or film layer below. If however, the substrate is initially dislocation free, half-loops must first nucleate from the film surface and glide down to the substrate before forming misfit segments at the interface. In this case, the appropriate energy

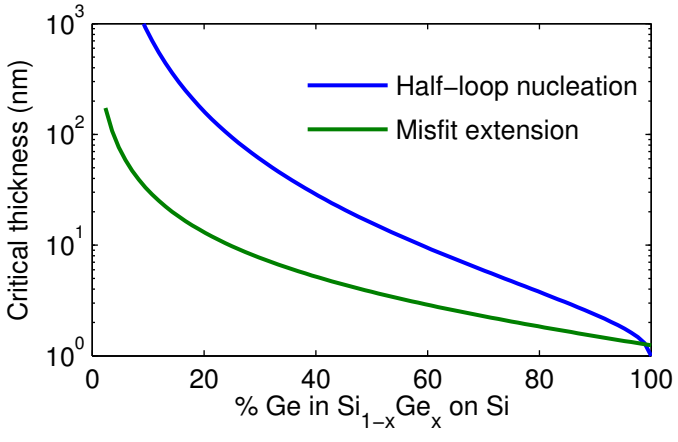


Figure 2.8: Critical thickness calculations based on the models of Matthews et al.[41] and People and Bean.[42]

balance is between the reduction in elastic energy due to misfit relaxation and the self-energy of a screw dislocation that extends through the film:

$$E_{\mathrm{TD}} \approx \frac{Gb^2}{8\sqrt{2}\pi a} \ln\left(\frac{h}{b}\right) \tag{2.22}$$

The critical thickness beyond which a net decrease in total elastic energy results from the creation of a dislocation half-loop that extends to the film interface to relieve misfit strain can be estimated:[42]

$$h_{\mathrm{c}} \approx \left(\frac{1-\nu}{1+\nu}\right)\left(\frac{b^2}{16\sqrt{2}\pi a f^2}\right)\ln\left(\frac{h_{\mathrm{c}}}{b}\right) \tag{2.23}$$

For film/substrate combinations with larger misfit, f, plastic relaxation is expected to occur for thinner film thickness according to the results of either set assumptions made. For growth of Ge on initially dislocation-free Si substrates, Eq. 2.23 is the more relevant equation of the two and has been found to better agree with experimental results over a wide range of $Si_{1-x}Ge_x$ film compositions grown on Si substrates.[42] The two expressions for h_{c} are compared in Figure 2.8 for the $Si_{1-x}Ge_x$ on Si materials system for which the lattice mismatch ranges from 0% to 4.2%. Because the critical thickness for half-loop extension to the film/substrate interface is almost always greater than the critical thickness for misfit extension, the introduction of dislocation half-loops is expected to control the relaxation process for films grown on initially dislocation-free surfaces.

2.4 Ge-on-Si growth techniques

Because Ge and Si have a lattice mismatch of approximately 4.2%, the critical thickness is estimated to be only a few atomic layers, and film relaxation occurs almost immediately during the initiation of film growth. Due to the favorable surface energy of Ge and Si, growth initially proceeds in a two-dimensional layer-by-layer growth mode, referred to as Frank-van der Merwe growth. However, due to the large strain at the Ge/Si interface, film growth will quickly transition to three-dimensional island growth, an alternative mechanism that reduces film strain elastically. This growth mode transition, referred to as Stranski-Krastanov (S-K) growth, is found to be favorable for lattice mismatches greater than 0.8%.[43] S-K growth will occur if the substrate temperature is sufficient for adatom surface diffusion.[44] Typical threading dislocation densities (TDD) in Ge films directly grown on Si in a single-step process are on the order of 10^{10} cm^{-2}.[45] While the formation of misfit dislocations are necessary to plastically relax the lattice misfit between the substrate and the film(s) deposited, misfits are not detrimental to devices if their operation does not rely on the material quality at the lattice-mismatched film/substrate interface. However, because threading dislocations exist in the bulk of the thin film layer(s), there is considerable focus in the literature on the reduction of the final TDD.

Several methods have been established to grow high quality Ge on Si films that avoid islanding and exhibit relatively low TDD values. In the graded buffer approach, instead of directly growing pure Ge on Si, the composition of the film is graded, starting with pure Si and grading at a rate of 10% Ge per μm in steps of 2000 Å for each constant composition.[46] As each layer is deposited, the deposited material relaxes to its equilibrium lattice constant by the extension of misfit segments. After the first lattice-mismatched layer is deposited and relaxed, new dislocations do not need to be introduced for additional layers to relax their strain since the ends of threading dislocations in the previous layer are available to continue as misfits at the newest film interface. The misfit arrays separating $Si_{1-x}Ge_x$ layers of different compositions can be seen in transmission electron microscopy (TEM) cross-section in Figure 2.9. The ability to reuse dislocations in each layer, preventing the formation of additional ones, allows the structure to reach TDD values of 2×10^6 cm^{-2} without post-growth anneal. The main drawbacks to a graded buffer is a thickness in excess of 10 μm, and the chemical mechanical polishing step required after reaching a composition of $Si_{0.5}Ge_{0.5}$ to

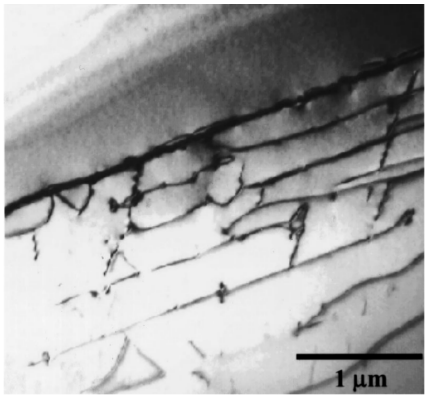


Figure 2.9: Transmission electron microscopy (TEM) cross-section of a graded SiGe buffer with a pure Ge layer at top.[46]

remove cross-hatch. The cross-hatch both exacerbates and results from dislocation pile-up, pinning threading dislocations attempting to glide in orthogonal directions.[47] The effect of dislocation pinning is revisited in Section 4.2.

Because threading dislocation line directions in diamond cubic materials are inclined with respect to the film surface normal, threading dislocations in patterned Ge films with high aspect ratios can terminate at the film sidewalls, leaving the Ge top surface free of dislocations (assuming that the threading dislocations are relatively straight). This dislocation reduction method, often referred to as either aspect ratio trapping (ART) or epitaxial necking, has been investigated by two methods. The first method involves the selective growth of Ge on Si in high aspect ratio SiO_2 trenches.[48] At sufficiently low supersaturation of the Ge source and minimized contamination of the SiO_2 surface, Ge adatoms will deposit on Si or Ge films but not SiO_2.[49] The second method reported in literature uses directional low-energy plasma-enhanced chemical vapor deposition (PECVD) on Si pillars.[50] While the ART technique can prevent the propagation of most threading dislocations to the film surface, when individual Ge mesas strips coalesce to form a blanket film, a high density of coalescence-induced defects (10^8 cm^{-2}) appear near coalescence fronts,[48] likely due to the imperfect registry of the relaxed Ge lattice between adjacent Ge films.[51] An example of these defects can be observed in TEM cross-section in Figure 2.10. Stacking faults may also initiate during lateral overgrowth due to a high enough surface energy of the epi-film/dielectric mask interface as has been observed for Si selectively grown on a Si_3N_4 mask.[52] All line defects originate from coalescence points, not the Ge/Si interface. Coalescence fronts can be spaced further

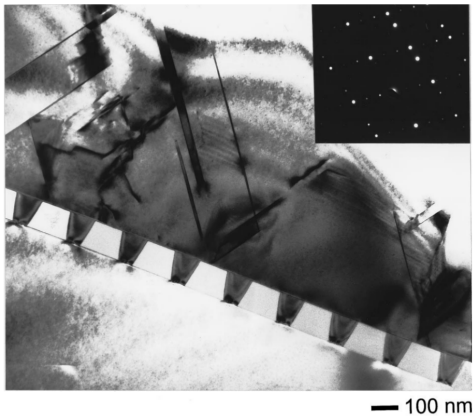
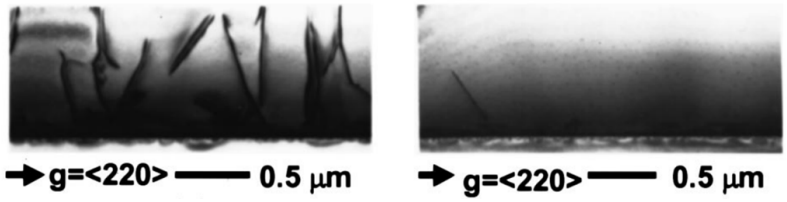


Figure 2.10: TEM cross-section of coalesced Ge film grown by aspect ratio trapping.[51]

away from each other to minimize the density of defective regions but at the cost of reduced throughput due to the increased time required for lateral overgrowth. Thicker Ge films also risk the increased probability that the film will crack after growth due to the Ge's thermal expansion coefficient mismatch with the Si substrate.[53]

Another method employs a two-step growth process, initially growing a thin, e.g. < 100 nm, pure Ge buffer layer at low temperature. Growth at low temperature prevents island formation due to the reduced mobility of Ge adatoms resulting from the surfactant effect of adsorbed hydrogen.[54] After depositing a thick enough buffer layer in which the majority of the strain resulting from lattice mismatch is relaxed by misfit generation, the growth temperature can be increased to deposit on the order of 1 μm of Ge of improved material quality, e.g. lower point defect concentrations, and at a greater deposition rate. The two-step growth process using a pure Ge buffer grown at a temperature from 330°C to 360°C typically achieves an as grown TDD on the order of 10^8 to 10^9 cm^{-2}. Annealing the film at a temperature around 850°C provides a resolved shear stress on glissile dislocations due to the difference in thermal expansion between the Ge film ($\alpha \approx 5.8$ ppm/K) and the Si substrate ($\alpha \approx 2.6$ ppm/K). As dislocations in Ge glide in response to the resolved shear stress, they will pass by and potentially react with each other. Threading dislocation annihilation will result between dislocations of exact opposite or identical burgers vectors. In the latter case (which is more likely as most dislocations should relax the lattice misfit at the Ge/Si interface, not increase it), the misfit components of the two threading dislocations combine while the two threading arms annihilate each other. Fusion of two dislocations into a single dislocation can occur



(a) (b)

Figure 2.11: TEM cross-section of blanket Ge-on-Si films (a) before, (b) after cyclic anneal.[55]

if the total strain energy associated with the dislocations decreases. To first order (ignoring the screw and edge character of each dislocation), this condition will be satisfied if $b_1^2 + b_2^2 > b_f^2$, where $b_{1,2}$ and b_f are the burgers vectors for the initial dislocations and final dislocation after fusion, respectively. The value of b_f is determined by the conservation of burgers vectors: $\vec{b}_f = \vec{b}_1 + \vec{b}_2$. In blanket films of Ge-on-Si, the TDD reduces to 2×10^7 cm^{-2} after cyclic annealing (see Figure 2.11).

A further reduction in the final TDD can be achieved in Ge mesas selectively grown on Si between SiO_2 sidewalls. Because the furthest distance between any point in the mesa interior and the mesa perimeter is half the mesa width, dislocation glide due to the resolved shear stress can cause dislocations to move to the mesa edges without requiring glide distances greater than the mesa width (e.g. 10 μm). TDD levels of 2×10^6 cm^{-2} have previously been observed in the planar centers of 10 μm × 10 μm Ge mesas after 10 cycles of annealing between 900°C and 100°C.[55] Figure 2.12 shows the locations of threading dislocations revealed by a defect etch decoration after single annealing and cyclic annealing.

A diverse collection of other growth methods are also reported in literature, each with moderate success. Surfactant mediated growth provides an alternative method to prevent Stranki-Krastanov Ge-on-Si growth by using submonolayers of antimony and tellurium to suppress Ge adatom

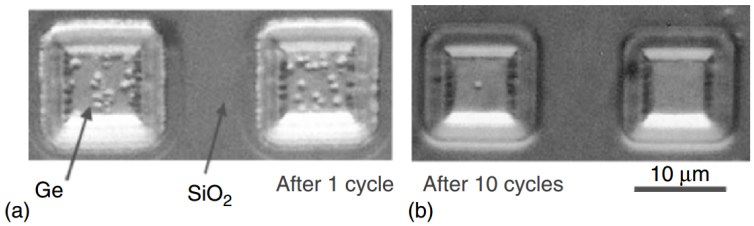


Figure 2.12: Defect etch resolved threading dislocations in 10 μm × 10 μm mesas (a) after single anneal, (b) after cyclic anneal.[56]

mobility.[57] While film islanding is prevented, TDD values of 1.6×10^8 cm^{-2} as measured by TEM plan-view result.[58] Use of thin (e.g. 100 nm) $Si_{1-x}Ge_x$ buffer layers[59–61] without annealing has produced moderate TDD results for blanket films, e.g. 6×10^6 cm^{-2}, but these results have not been widely demonstrated by others. The multistep lateral overgrowth with hydrogen annealing technique borrows aspects from ART, two-step growth, and annealing, but achieves TDD values around 4×10^6 cm^{-2}.[62] Elastic stress relaxation has been attempted through the growth and coalescence of Ge in small islands ($\approx$ 6 nm in diameter) on a partially oxidized Si surface. The nanostructure is thought to prevent the formation of threading dislocations due to the stress relaxation that occurs due to the small feature size of each individual Ge island.[63] However, a high density of stacking faults is created upon island coalescence (5×10^7 cm^{-2}). Successful attempts to remove them by annealing before complete Ge island coalescence leads to the formation of threading dislocations upon continued Ge growth (5×10^7 cm^{-2}).[64]

Because the two-step selective growth and anneal technique for selectively grown Ge films has demonstrated a low TDD without requiring thick buffer layers, it is an ideal candidate for the fabrication of solar cells that require low TDD values. Initial implementation of this growth method for the fabrication of large-area Ge photovoltaic cells composed of Ge mesa arrays is presented in Chapter 3. Chapter 4 explores the parameters that control defect reduction in these structures in order to determine what modifications will allow for further reduction of the final TDD obtained by this fabrication technique.

Chapter 3

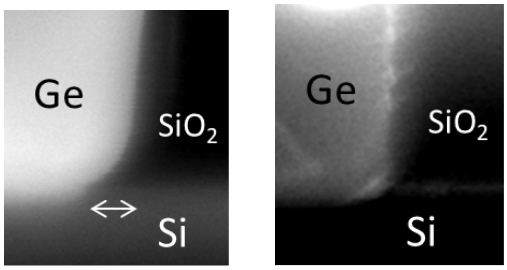
Mesa array Ge photovoltaic cells

Because selective growth of Ge on Si has previously been shown to achieve TDD values of 2×10^6 cm^{-2} without requiring 10+ μm buffer layers, this method is selected for initial Ge photovoltaic cell device fabrication. While the maximum Ge mesa width for selective growth to be sufficiently effective for dislocation reduction is 10 μm, typical photovoltaic concentrator cell sizes are on the order of 0.1 to 1 cm^2.[65] Ge mesa structures are therefore arrayed to fill 2 mm × 2 mm areas. Device fabrication, described in further detail below, electrically connects the individual Ge mesas in parallel, enabling them to collectively act as a single Ge photovoltaic cell.

3.1 SiO_2 grids for selective growth

In order to grow Ge mesas selectively on Si, a layer of SiO_2 must first be deposited and patterned. Si wafers are prepared for selective growth by growing a 0.75 μm thick SiO_2 layer via thermal oxidation of the Si substrate. Trenches in the SiO_2 layer are defined lithographically by plasma etching using a CF_4/CHF_3/Ar gas chemistry in a reactive ion etching (RIE) chamber. Initially, a dry etch - wet etch method was employed to remove 90% of the oxide thickness via RIE, while the remaining 10% of material is removed by buffered oxide etchant (BOE), composed of NH_4F:HF=7:1, after removing photoresist by oxygen plasma ashing. This two-step etching approach prevents fluoropolymer residue that forms during dry etching from accumulating directly on the Si surface. If Ge is grown on a Si surface contaminated with this residue, a high density of stacking faults will form at the surface. By using dry etching for the majority of the trench thickness, the trench retains its vertical sidewall as a result of the anisotropic nature of RIE.

While this process works well for SiO_2 thicknesses less than 300 nm, the effects of a nonuniform dry etch rate across the wafer will become more pronounced for thicker films. Because the dry etch depth varies across the wafer and wet etching is isotropic, a sloped region of variable width and depth will form at the bottom of the SiO_2 trenches, visible in Figure 3.1a indicated by an arrow

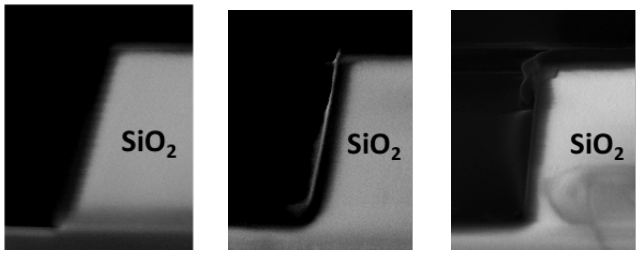


(a) (b)

Figure 3.1: SEM cross-section of SiO_2 sidewalls (and subsequent Ge growth) etched via (a) dry etch and wet etch, (b) sacrificial oxidation process flows.

approximately 100 nm in length. The low angle of the SiO_2 sidewall will induce the formation of slow growing facets during Ge growth, prolonging the time required for complete trench fill. In order to reduce the sloped region to a minimum while simultaneously improve uniformity among fabricated structures across each substrate, the entire depth of the SiO_2 layer was dry etched, exposing the Si substrate surface to fluoropolymer residue. Because of the chemical selectivity of reactive ion etching, the Si surface was not significantly etched once it was exposed. After removing the photoresist layer by ashing, a sacrificial oxidation was performed at 950°C in pure O_2 for 40 minutes to oxidize the fluoropolymer and the Si surface that was exposed to reactive ion etching damage. After removing the sacrificial oxide in BOE, a fresh Si surface became available at the surface, ready for Ge epitaxy. The modified SiO_2 trench etch process flow successfully eliminates the lip as seen in Figure 3.1b.

When Ge grows between SiO_2 sidewalls, the single crystal will form facets to reduce the SiO_2/Ge interface due to its high surface energy. Only at high non-equilibrium conditions of relatively low growth temperature, and high p_{GeH_4} supersaturation can faceting of selectively grown Ge films be avoided.[66] Because the vertical growth rates of the facets that form are reduced compared to the blanket film growth rate, the Ge mesas become bounded by these slower growing facets, prolonging the growth time required to fill the windows in the SiO_2 film with Ge. With estimated relative growth rates of {111}, {311}, and {100} assuming a ledge flow model on {111} terraces, the slowest growing facets, {111}, can be prevented from appearing only if the SiO_2 sidewall is inclined at an angle greater than 83°.[67] The dry etch process was therefore optimized to maximize the SiO_2 sidewall angle with the substrate normal, achieving a sidewall angle of 83° – 84° by increasing the CF_4/CHF_3 gas flow ratio from 1:3 to 2:1, as seen in Figure 3.2. Similar trends



(a) (b) (c)

Figure 3.2: Scanning electron microscopy (SEM) cross-section of SiO_2 films etched with a CF_4/CHF_3 chemistry in the ratio of (a) 1:3, (b) 1:1, (c) 2:1, creating sidewall angles 75°, 80°, and 83°, respectively. Unless otherwise noted, all original SEM images included in this thesis are recorded using an In-Lens detector.

have been observed in literature for etching Si_3N_4 films.[68] CF_4/CHF_3 ratios greater than 2:1 did not produce noticeably greater sidewall angles which is likely limited by the angle of the photoresist profile after development.[69] Lower F/C ratios (reduced CF_4/CHF_3) increase fluorocarbon polymer formation during the dry etch process. The polymer can be removed by ions in the plasma that directionally bombard the horizontal surface more effectively than the trench sidewalls, increasing the anisotropy of the etching process. However, less available F in the plasma reduces the chemical etch rate, making the RIE process more physical than chemical in nature. By increasing the F/C ratio of the etching gases, the chemical aspect of RIE is increased, bringing the process closer to the conditions necessary for chemical and physical synergy to achieve vertical etch profiles.[70]

3.2 Ultra-high vacuum chemical vapor deposition

After pattern transfer to the SiO_2 mask, Si substrates are wet cleaned prior to Ge film growth by the following procedure, an "RCA clean"[71] with an additional HF dip to H-terminate the Si surface:

1. Organic/particle clean (SC1): $NH_4OH:H_2O_2:H_2O$ (1:1:5) at 80°C for 10 minutes
2. Chemical oxide strip: $HF:H_2O$ (1:50) for 60 seconds
3. Ionic clean (SC2): $HCl:H_2O_2:H_2O$ (1:1:6) at 80°C for 10 minutes
4. Chemical oxide strip and passivation: $HF:H_2O$ (1:50) for 60 seconds

Both SC1 and SC2 create a chemical oxide on the Si surface which will impede epitaxial Ge growth if not removed. After the final oxide strip in dilute HF and short water rinse, the Si surface remains

terminated by hydrogen.[72] The substrates are spun dry in N_2, placed vertically into a horizontally oriented boat, and loaded into the load lock connected to the UHV-CVD process chamber. Thermal SiO_2 coated baffle wafers are placed at both ends of the wafer boat to create a more uniform environment (in terms of both temperature and gas flow) for the first and last wafers in the batch. Selectivity of Ge deposition only on Si/Ge surfaces prevents deposition on the baffle wafers.

Ge is grown on the surface of Si substrates by thermal decomposition of film precursor gases in a process called ultra-high vacuum chemical vapor deposition (UHV-CVD). In general CVD processes, thin film precursors, such as the hydrides SiH_4 and GeH_4, arrive at the substrate surface, decompose into adatoms by releasing H_2, and adsorb onto free surface sites. At favorable deposition conditions for epitaxial growth, adatoms diffuse across the surface until they find an atomic ledge at which they can deposit to lower their free energy. B_2H_6 and PH_3 can be added to the gas flow to incorporate dopants into the epitaxially growing film. With growth rates on the order of 1-10 nm/min and gas residence times of 1-10 sec, doping profiles can be controlled very accurately. During temperature ramping or annealing steps, H_2 is typically flown to maintain a reducing ambient in the chamber. While deposition occurs at chamber pressures between 10^{-3} to 10^{-1} mbar, the growth technique is referred to as UHV because the base pressure of the process chamber approaches values below 1×10^{-8} mbar. With a base pressure of that magnitude, film contamination only results from the impurity of the precursor gases. Oxygen and water vapor partial pressures are sufficiently reduced to enable oxygen-free silicon surfaces at moderate growth temperatures.[73] Due to the lower thermal budgets possible, UHV-CVD has been successfully used for the growth of SiGe layers in BiCMOS technology at temperatures below 600°C at which hydrogen passivation of the surface allows for well-controlled epitaxial growth with minimal defects. The low temperature also minimizes auto-doping and dopant diffusion during film growth.[74]

Depending on the growth conditions, a CVD process will either be surface reaction limited or mass transport limited. In the surface reaction limited regime (which occurs at lower growth temperatures), film growth is independent of the precursor partial pressure but is exponentially dependent on substrate temperature since surface reactions are activated processes. For growth of Ge and Si films, hydrogen desorption from the film surface has been identified as the rate limiting step. Adatoms can only deposit on the surface sites that are not terminated by hydrogen. Because H_2 desorption from Si and Ge surfaces has an activation energy of 2.1 eV and 1.46 eV respectively,

Ge can be grown at much lower temperatures at reasonable growth rates compared to Si.[73] When operating in the surface reaction limited regime, the thermal profile must be exceptionally uniform to maintain consistent growth rates across and between substrates.

At higher temperatures, the growth rate is instead limited by mass transport of precursors to the substrate surface, making the process weakly dependent on temperature but greatly influenced by precursor partial pressure in addition to reactor geometry and total pressure if carrier gas is used at pressures on the order of 0.1 mbar or higher. Larger precursor partial pressures will increase the precursor flux to the substrate surface, while increased total pressure will reduce the precursor diffusion coefficient, D:

$$\text{growth rate} \propto D_{\mathrm{GeH_4}} p_{\mathrm{GeH_4}} \propto \frac{T^{3/2}}{P_{\mathrm{tot}}} p_{\mathrm{GeH_4}} \tag{3.1}$$

Ge layers are grown in a Leybold Sirius CVD-300 reactor. The temperature inside the hot-walled quartz tube process chamber is controlled by three separate heater zones, calibrated to maintain a flat thermal profile across the substrate boat. Ultra-high vacuum is maintained by a turbomolecular pump backed by a combined dry rotary pump and roots pump. The chamber pressure is controlled by the position of a throttle valve between the turbomolecular pump and the process chamber while gas flow rates remain constant.

After pumping the load lock to pressures below 10^{-6} mbar using a separate turbomolecular pump and dry rotary pump, the gate valve between the load lock and process chamber opens, and the wafer boat is placed into the center of the process tube under vacuum by a hook and rod mechanically coupled to the outside magnetically. H_2 flows into the reactor during the wafer boat transfer process to minimize contamination. After the wafer boat is placed in the center of the tube, the gate valve is closed, and the base pressure of the process tube returns to below 2×10^{-8} mbar, the process chamber is initially heated to 800°C under hydrogen flow to desorb any oxide that may have formed on the Si surface since the substrates were cleaned. The tube temperature is then reduced to 350°C to grow the 80 nm thick Ge buffer at a rate of 1.0 nm/min, followed by growth of intrinsic Ge at 730°C (at a blanket growth rate of 6-10 nm/min) for a variable time, typically of a few hours. Both the buffer and high temperature layers are deposited using pure GeH_4 at pressures up to 2.0×10^{-2} mbar without carrier gas. The process chamber is then raised to 850°C to anneal

the material to reduce the threading dislocation density. See Section A.2 for a detailed typical Ge growth and anneal process recipe.

Ge's transition between the two regimes for deposition conditions used in UHV-CVD (no carrier gas and $p_{GeH_4} \approx 10^{-2}$ mbar) occurs between 400°C and 450°C.[75] Thus, control of the wafer temperature during the low temperature buffer is crucial for the two-step growth process. During high temperature growth, uniform growth rates across and between different substrates is still achievable due to the low total pressure in the reactor chamber. The mean free path, the average distance traveled by molecules between intermolecular collisions, can be estimated for an ideal gas:[76]

$$\lambda = \frac{k_B T}{\sqrt{2}\pi P_{tot} d^2} \tag{3.2}$$

where d is the effective molecular diameter (e.g. 3 Å). At a pressure of 10^{-2} mbar and temperature of 730°C, the mean free path for GeH_4 molecules in the gas phase, λ_{GeH_4}, is estimated to be 3.5 cm. Because this distance is greater than the spacing between adjacent substrates (1 cm) as well between the substrates and the quartz chamber walls (1.8 cm), the Knudsen number, $K_n = \lambda/d_B > 1$, where d_B is a characteristic distance between boundaries. For $K_n > 1$, collisions between gas molecules become negligible compared to collisions between single gas molecules and solid surfaces. Viscous flow is therefore not the dominant mechanism controlling gas flow, and stagnant boundary layers do not form at the substrate surfaces (in contrast to other CVD techniques that operate at higher total pressures). Because the sticking coefficient, the probability that an adsorbed precursor molecule successfully decomposes before desorbing, for GeH_4 is approximately 0.02 on Si surfaces[77] (and is observed to decrease as the Ge composition of the surface increases[78]), GeH_4 molecules will adsorb and desorb from the substrate surfaces many times before depositing, leading to improved uniformity across each wafer.

Even with perfectly vertical SiO_2 trench sidewalls, relatively slow-growing {311} facets will still appear for the growth conditions mentioned earlier. Once the mesa's (001) facet disappears, further increase of the Ge film thickness at the mesa center is determined by the vertical growth rate of the {311} facet. Because the (001) facet disappears earlier in the growth process for mesas with narrower widths, mesas with smaller widths will also be thinner once {311} facets dominate the mesa surface. The effect of mesa width on the mesa profile is clearly seen in the comparison of

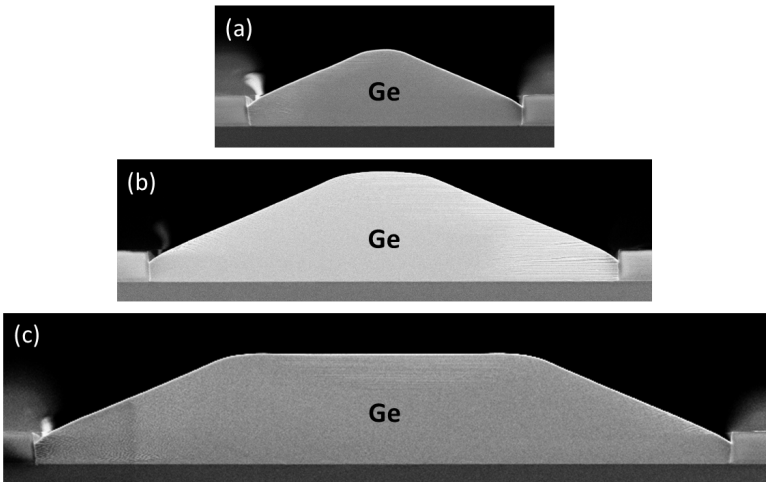


Figure 3.3: SEM cross-sections of Ge mesas with widths (a) 6 μm, (b) 10 μm, (c) 15 μm.

6 μm, 10 μm, and 15 μm wide mesas in Figure 3.3, each grown under identical process conditions with a high temperature growth time of 3 hours. Only the 15 μm wide mesa retains its (001) facet after the same duration of high temperature growth. While both the 6 μm and 10 μm mesas have lost their (001) facet, the 10 μm is thicker in its center because its (001) facet was consumed later in the growth process.

3.3 Device Fabrication

Using the optimized process flow that creates uniform, vertical SiO_2 sidewalls discussed above, arrays of Ge mesas in close proximity to each other (e.g. 2 μm) are fabricated and combined in parallel to create large area Ge photovoltaic cells. Simultaneously during the process flow described below, separate Si solar cells are also fabricated, separated by trenches etched into the wafer before Ge growth is conducted. A schematic of the intended final structure of a Si cell next to a Ge mesa array cell is displayed in Figure 3.4. The Si cells serve as useful monitors for later electrical characterization. Initially, Si substrates were blanket implanted with boron on the front side to form a p^+ region on the Si surface with an expected junction depth of 0.7 μm after all later thermal treatments in the process flow. This layer serves as both the emitter for the Si photovoltaic cells and in other regions, as the surface to initiate Ge epitaxy. During post-growth annealing of the Ge films, boron will diffuse into the low-temperature Ge buffer layer, causing the highly defective region to serve merely as a conduction path sufficiently away from the Si/Ge heterojunction.

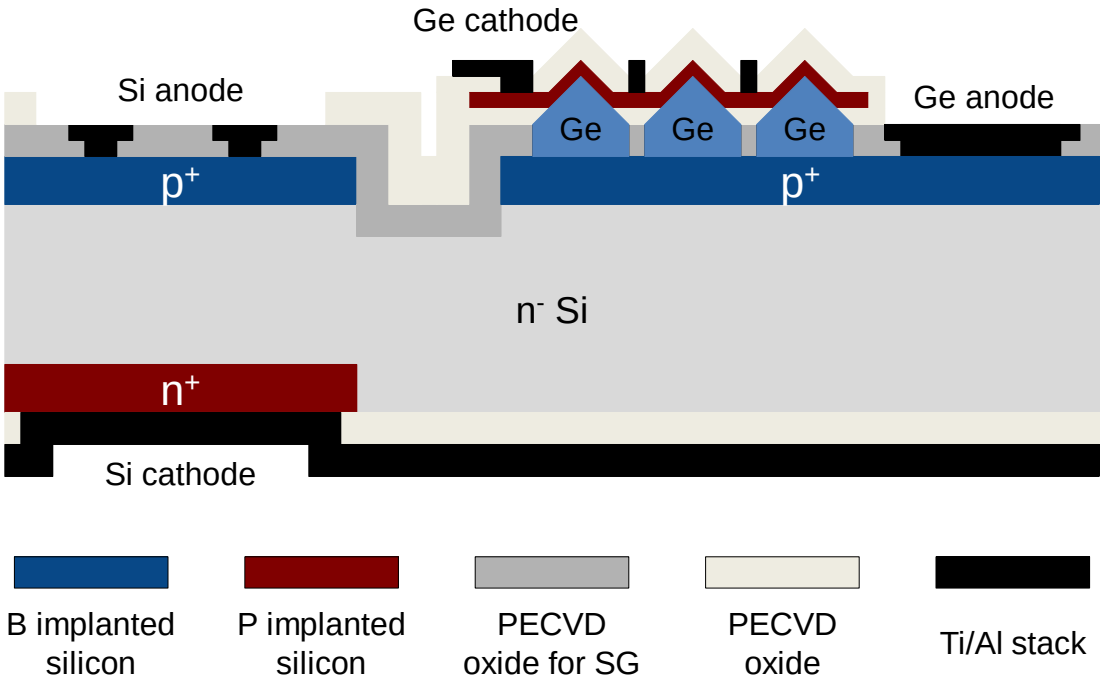


Figure 3.4: Schematic of Si and Ge mesa array solar cell structures.

Positive i-line photoresists OCG-825 and SPR-700, each spun to create ≈ 1 μm thick films, are used for all pattern transfer steps. For front side patterning, photoresist exposure is conducted using Nikon's NSR2005i9C i-line stepper, while backside exposure relies on an EVG EV620 Mask Aligner which has frontside/backside alignment capability. Trenches, 0.5 μm to 3 μm deep and approximately 100 μm wide are first dry etched on the substrate's front side to separate regions intended for Si and Ge mesa array cells (see Section A.1 for the dry etch recipe). On the substrate backside, phosphorus is implanted selectively using photoresist as a hard mask in regions directly below the areas intended for individual Si cells. Dopant activation and sacrificial oxidation of the Si trench sidewalls are conducted simultaneously by a wet thermal oxidation at 900°C, growing 100 nm of thermal SiO_2 in 45 minutes. The thermal oxide is then striped in BOE, and 1 μm of SiO_2 is deposited by PECVD (see Section A.1 for the SiO_2 deposition recipe). SiO_2 patterning for selective growth is then conducted using the dry etch, sacrifical oxidation process described in Section 3.1, followed by Ge growth and post-growth anneal as described in Section 3.2. The sacrificial oxidation step also serves to outgas the PECVD SiO_2 film of any hydrogen incorporated

into the film during deposition. If hydrogen is not sufficiently outgassed prior to Ge growth, the resulting Ge films will be extremely rough due to the local high pressure in the viscinity of the substrate surface during film growth.[79]

After Ge epitaxy with high temperature growth time of 4.5 hours to completely fill a 750 nm thick SiO_2 trench, a stack of 15 nm of SiO_2 followed by 20 nm of a-Si and an additional 100 nm thick layer of SiO_2 is deposited by PECVD to passivate the Ge around the mesa perimeters. Due to the high hydrogen content of PECVD a-Si, hydrogen from the a-Si layer can diffuse to the Ge surface through the thin oxide and passivate the dangling bonds of the Ge surface. The thin SiO_2 layer in direct contact with Ge prevents hydrogen bubble formation that occurs if the a-Si is in direct contact with the Ge surface. The final layer of SiO_2 is deposited to prevent a-Si from outgassing its hydrogen directly to the ambient. This stacked multilayer has previously been shown to be a promising passivation technique for Ge devices.[79] After opening windows in the passivation stack by dry etching through the top SiO_2 and a-Si layers, wet etching in BOE removes the final 15 nm of SiO_2. To prevent any RIE damage to the Ge surface, a highly selective Si etch recipe (with an estimated selectivity for Si:SiO_2 of 100:1) is used so that the thin SiO_2 layer in direct contact with Ge is not removed during the a-Si etch.

After opening windows in the stack passivation, the substrates are cleaned in preparation for deposition of a 160 nm thick layer of a-Si deposited by low-pressure CVD (LPCVD) at 560°C. Because the SC1 step in the standard RCA cleaning process will etch Ge at 275 nm/min even at room temperature,[80] a modified room temperature cleaning procedure is adopted specifically for compatibility with Ge:

1. Native oxide strip: NH_4OH:H_2O (1:4) for 5 minutes
2. Organic clean: H_2O_2:H_2O (1:6) for 15 seconds
3. Chemical oxide strip: HF:H_2O (1:50) for 30 seconds
4. Ionic clean and passivation: HCl:H_2O (1:4) for 30 seconds

Unlike UHV-CVD Ge growth, LPCVD Si deposits nonselectively on both the Ge mesas and SiO_2, connecting adjacent mesas electrically in parallel as intended. To form the pn heterojunction for the Ge cells, the LPCVD a-Si is implanted with phosphorus at a dose of 5×10^{15} cm^{-2}, implant energy of 70 keV, and angle offset of 7°, designed to stop mostly in the a-Si to not excessively

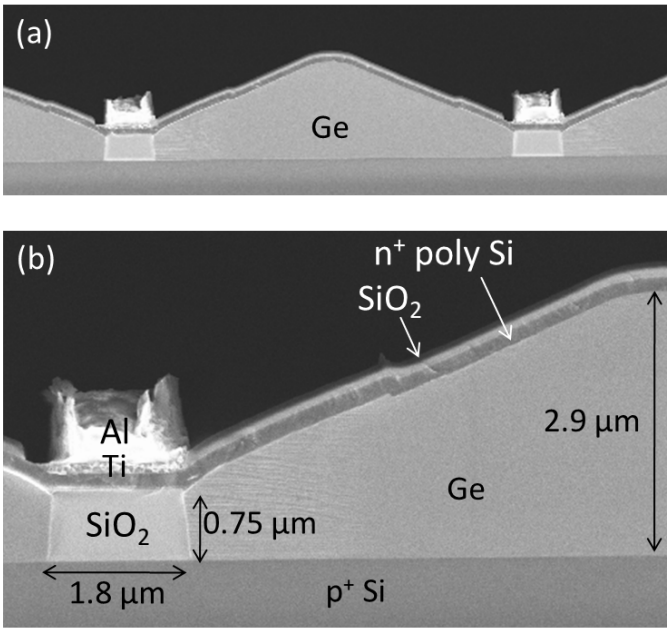


Figure 3.5: SEM cross-section of Ge mesa array showing (a) adjacent Ge mesas and (b) structural detail.

introduce implantation damage the top of the Ge mesas. Because the a-Si layer to be implanted is oriented 25° away from the substrate plane due to the {311} mesa facets, substrates are continuously rotated during the implantation to minimize non-uniformity between the effective implant angle for different {311} facets. During dopant activation at 700°C for 60 seconds by rapid thermal activation (RTA), the LPCVD a-Si layer transforms into poly-Si, and some implanted phosphorus diffuses into the top of the Ge mesas. Because the Ge is grown without dopant precursors, the resulting structure is expected to be a p-i-n Si/Ge heterojunction, an ideal structure to maximize carrier collection by drift. The poly-Si layer is patterned to isolate different mesa arrays from each other and cleared in regions intended for separate c-Si cells. After depositing and patterning a 200 nm layer of PECVD SiO_2 to passivate the poly-Si, 100 nm Ti followed by 1 µm of Al metal are sputtered deposited and patterned to serve as separate electrical contacts to both the n^+ poly-Si and p^+ c-Si layer. A similar process is used to open contact holes and deposit the metal stack on the wafer backside for the n^+ c-Si contact for the c-Si cells. As a final step, the metal films are sintered in 5% H_2/95% N_2 at 400°C for 1 hour. The final structure of the Ge-Si heterojunction device can be seen in Figure 3.5.

3.4 Characterization and Discussion

After fabrication, the Ge photovoltaic cells were evaluated by current-voltage (I-V) sweeps both in the dark and under illumination. Ge/Si heterojunction test structures consisting of large, unpatterned Ge regions were also fabricated to measure the dark current of the Ge films not affected by selective growth. To minimize any potential resistive losses, the metal contact is placed in the center of the large mesa as these structures were not intended to be measured under illumination. While a 200 μm × 200 μm single Ge mesa clearly shows diode rectification, all Ge mesa array heterojunctions cells fail to do so (see Figure 3.6).

The mesa array I-V characteristics correspond more closely to that of a resistor than a diode. The differential specific resistance, $dr \equiv dV/dJ$, is approximately constant at large forward (2 Ω cm^2) and reverse biases (3 Ω cm^2) with a gradual transition near zero bias. This I-V response can be closely modeled as a resistor, R_1, connected in series to a diode and resistor, R_2, connected to each other in parallel (see Figure 3.7). In forward bias, the current preferentially conducts through the diode with little resistance, so that the apparent resistance is just R_1. In reverse bias, current instead travels through R_2, making the apparent resistance equal to $R_1 + R_2$. In addition to recombination-generation centers, dislocations also serve as conductive leakage current paths,[81]

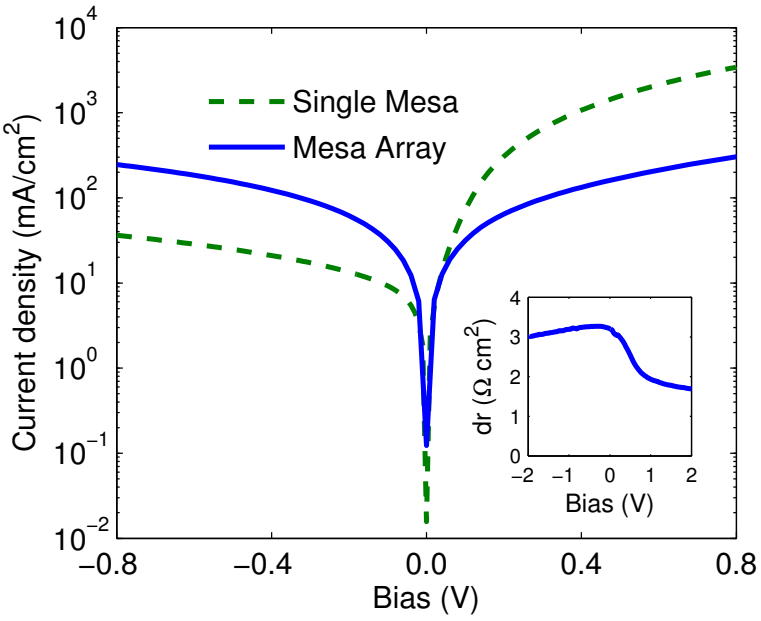


Figure 3.6: Dark I-V for Ge mesa and mesa array. An inset includes a Ge mesa array's differential specific resistance.

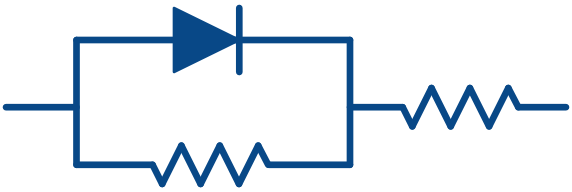


Figure 3.7: Equivalent circuit diagram for a diode with a low resistance shunt in series with other internal device resistances.

that can be modeled as a resistor in parallel with a diode, i.e. the junction the threading dislocations shunt. With consideration of other internal resistances in series with the junction, the previously discussed electrical circuit model is found to be directly applicable to the Ge/Si heterojunction devices.[67]

Several unintentional processing oversights were identified after device fabrication was completed. While the Ge growth rate at mesa edges was calibrated so that mesas completely filled the windows in the SiO_2 film after growth at mesa edges, trench fill was severely limited at mesa corners. The conventional secondary electron detector image in Figure 3.8 highlights the reduced film thickness at mesa corners after Ge growth. Other implications of the limited film growth in mesa corners is more closely examined in Chapter 5. A second unintended effect during processing occurred to the Ge passivation layers. After the first two layers (SiO_2 and a-Si) were dry etched, the photoresist was removed by oxygen plasma ashing before wet etching of the thin SiO_2 layer in direct contact with Ge. By arranging the process flow in this order, the Ge surface would not be directly exposed to O_2 plasma for any amount of time. However, because the PECVD films were not annealed, the top SiO_2 layer was completely removed during the 30 second wet etch. Unannealed PECVD SiO_2 films are removed approximately 5× as quickly as thermal SiO_2 in HF containing solutions.[82]

The first step of the Ge RCA cleaning procedure is a 5 minute etch in diluted NH_4OH. While the Ge etch rate in this solution is less than 1 nm/min,[83] without the top layer of PECVD SiO_2 or any H_2O_2 in solution to create a protective chemical oxide on Si,[84] NH_4OH is expected to completely etch away the 20 nm thick a-Si film well within 5 minutes,[85] especially due to its high hydrogen content, known to increase the wet etching rate of Si films in caustic solutions.[86] Subsequently during the 30 second immersion in dilute HF, the newly exposed 15 nm unannealed PECVD SiO_2 layer is likely etched away completely as well. Thus, all three passivation layers were

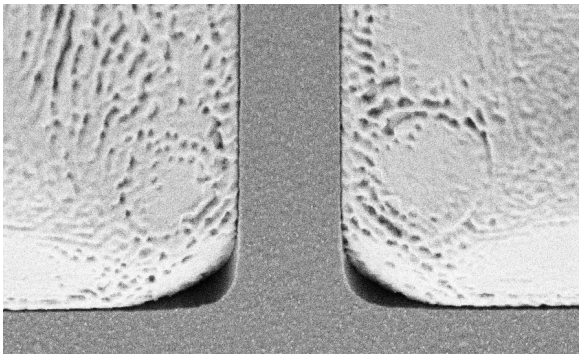
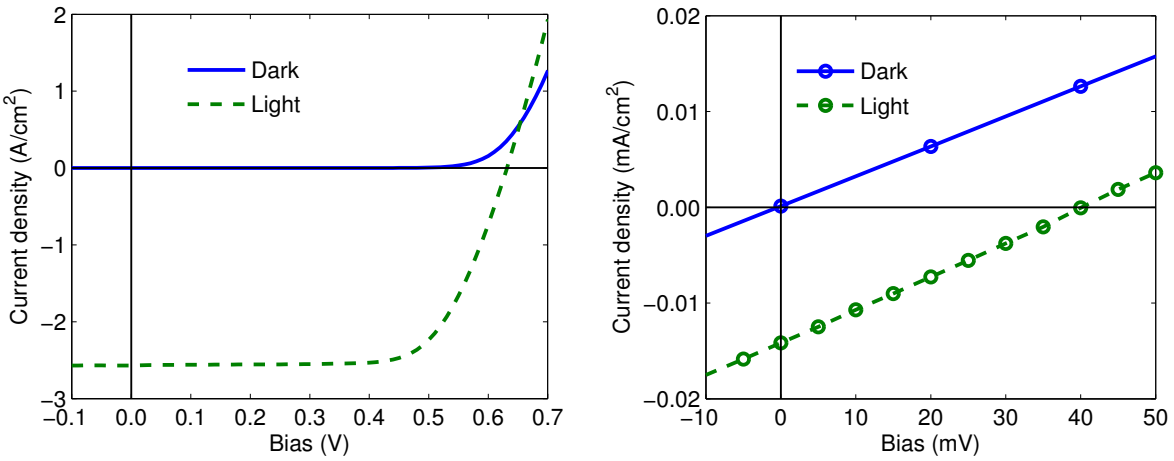


Figure 3.8: SEM conventional secondary electron top view of unfilled Ge mesa corners. The rough surface is due to in-situ deposition of a thin Si cap layer.

etched away, leaving the entire Ge mesa surface unpassivated and directly exposed to ambient prior to LPCVD a-Si deposition. In Figure 3.5, a slightly depressed portion of the Ge mesa center can be seen, starting approximately 3 μm away from the SiO_2 trench walls. The cause for the slight step is likely due to the non-zero etch rate of Ge in H_2O_2 before the final passivation layer is removed. The raised areas of Ge were still protected during the H_2O_2 cleaning step.

The relevant differences between the single mesa and mesa array devices is the proximity of the metal contact and poly-Si layers to the mesa corners. The mesa array contains at least hundreds of individual mesa corners that are directly next to the metal contact grid. In contrast, the single mesa structure has only four mesa corners, and the metal contact and poly-Si layers are defined at least 60 μm away from the Ge mesa perimeter. In the absence of Ge passivation layers, the remaining poly-Si after patterning will be in direct contact with the entire Ge surface, including at mesa corners for the mesa array. Because Ge growth at mesa corners is limited, it is likely that the poly-Si layer deposited directly on the Ge buffer layer. While post-growth anneal reduces the dislocation density in the Ge films, the dislocation density remains high in the low temperature buffer (e.g. 10^9 cm^{-2}).[69]

Any direct electrical connection between the n^+ poly-Si and the p^+ Ge buffer will provide a very low resistance shunt of the junction. Because of reduced Ge growth at mesa corners and the lack of passivation layers to physically separate the poly-Si layer from mesa perimeters, the mesa array structures were likely shunted at every mesa corner location. Rectification was still observed, however, for the single mesa structure because even with a lack of passivation layers, any poly-Si initially deposited on the four mesa corners was subsequently removed during patterning and thus,



(a) (b)

Figure 3.9: Dark and light I-V for (a) Si cell monitor, (b) Ge mesa array.

not part of the final device. The exposed mesa corners, in the case of the single large mesa diode, would be later covered by the SiO_2 layer that passivates the n^+ poly-Si. Thus, the Ge material itself is not to blame for the poor rectification but rather the unintentional fabricated structure in mesa array devices.

Despite the weak rectification observed for mesa arrays, their light I-V characteristics were still measured. A diode laser with an emission wavelength of 980 nm was used as an illumination source. The laser emission was coupled in free space to a multimode optical fiber and focused on the sample stage with a spot size diameter of approximately 5 mm. Light I-V measurements of adjacent Si solar cells provided an estimate of the diode laser power intensity on the substrate. The illumination was approximately 75× concentration of equivalent sunlight, determined by comparison of the J_{sc} measured (see Figure 3.9a) and the typical J_{sc} for Si cells with an anti-reflection coating but no texturing, illuminated with the standard AM1.5D spectrum: 35 mA/cm^2.[87] Due to the poor rectification of Ge devices that could receive illumination while tested, the structures exhibited a poor photovoltaic effect, achieving a V_{oc} of only 40 mV under 75× Sun concentration illumination (see Figure 3.9b), a fraction of the ideal V_{oc} for Ge solar cells, 300 mV.

While careful consideration of substrate patterning and cleaning immediately prior to LPCVD of a-Si can be employed to prevent some of the issues discussed above, other options for surface passivation are likely to be more successful. One potentially favorable option involves in-situ deposition of poly Si on the Ge surface immediately after post-growth anneal in the same UHV-CVD

reactor. After the SiO_2 grid is exposed to GeH_4 for several hours of growth, the SiO_2 becomes oxide-poor due the repeated formation and sublimation of GeO.[88] At certain favorable deposition conditions, Si deposits non-selectively on both Ge and the SiO_2 trench lines with introduction of SiH_4 into the reactor. With in-situ deposition, the Ge surface is never exposed to ambient, avoiding any potential sources of surface contamination. While the Si film will deposit planar layers with low surface roughness on SiO_2 and Ge (001) surfaces, film islanding in long streaks is observed for thin enough films on non-(001) facets. The undulated top layer in Figure 3.8 is composed of a poly-Si in-situ deposited layer with an average thickness of 25 nm but with individual streaked lines 50 nm high. In order for this passivation technique to be successful, the Si films must not only be thick enough to be continuous but the reduced growth of mesa corners must also be eliminated to prevent junction shunting. Chapter 5 provides an alternative selective growth pattern solution that overcomes the limitation of mesa corners but is still expected to filter threading dislocations.

Several trade-offs were encountered during the design and fabrication of the Ge mesa array photovoltaic cells. In order to create adjacent Ge mesas, SiO_2 lines were required to be placed in the region of illumination. Any light that illuminates areas not covered with Ge is lost, degrading device conversion efficiency. This effective "shadowing" was partially resolved by placing the metal contact grid directly above the SiO_2 walls (see Figure 3.5) so as to not introduce additional shadowing from the metal. Due to the resolution and alignment limitations of i-stepper lithography used for patterning, line widths of at least 1 μm were required for the metal grid lines. When overlaid with the 2 μm wide SiO_2 lines (to allow for alignment tolerances of 0.5 μm), the effective shadowed area would be 30% for arrays composed of 10 μm wide mesa squares. Shadowing of the SiO_2 grid can be reduced by extending the mesa squares into long strips but doing so will require rotation of the grid so that the effective glide distance for all slip systems stays within 10 μm. With 7 μm wide mesa strips aligned with ⟨010⟩ directions, separated by 2 μm wide SiO_2 lines, the effective shadowing only reduces to 22%. A variation of the process flow described above that only fabricates the large area Ge photovoltaic cells still consists of five pattern transfer steps (selective growth SiO_2, Ge passivation, poly-Si, poly-Si passivation, and metal). These patterns must have small line widths (<1 μm) and therefore tight alignment tolerances (<0.5 μm) if SiO_2 grid shadowing is to be minimized. Reliance on i-stepper lithography for five separate patterning techniques is expected to increase the costs of cell fabrication significantly.

For solar cells, the device must be thick enough to absorb the majority of the photons it receives to maximize its photocurrent. Ideally, the active layer region is of uniform thickness so that absorption is optimized in all regions of the cell. However, when using Ge mesa arrays as the active layer, the cell thickness will significantly vary due to facet formation at mesa sidewalls. While Ge mesas can be grown in thicker SiO_2 films (and subsequently planarized by chemical mechanical polishing (CMP) after growth) such that the optimal thickness of Ge existed throughout the mesa, the required growth time for complete trench fill would be severely increased, increasing costs of fabrication. Further, if the Ge mesas were then used as a substrate for III-V layer growth whether for single-junction III-V cells or as the the low E_g junction in a tandem cell, the III-V deposited films will form facets during growth even if the Ge was previously planarized. III-V solar cell junctions are composed of many individual layers (see Section 5.2.2) for which thickness control is crucial for efficient device operation. If grown directly on Ge mesas, layer thicknesses will be difficult to control because the existence of different facets will cause non-uniform growth rates for each layer. Further, if after III-V growth, the III-V layers still consist of isolated mesas, a fabrication scheme that electrically connects the III-V mesas in parallel must be developed.

3.5 Summary

Large area Ge photovoltaic cells were fabricated from Ge mesa arrays grown selectively in SiO_2 grids. While the mesa array cells exhibited poor V_{oc} performance, several steps of the process flow were identified as the likely causes for which modifications are suggested to avoid junction shunting. The observed reduction of film growth rates in the vicinity of mesa corners was identified as a fabrication issue needing further attention and becomes the initial focus of Chapter 5. The drawbacks for photovoltaic devices of the device structure and film growth that results from selective growth were also discussed.

Low band gap materials, such as Ge, do not fare well when compared to non-photovoltaic conversion technologies for photon energies even at E_g, an effect that is studied in more detail in Chapter 6. However, despite Ge's suboptimal E_g for photovoltaic energy conversion, Ge's lattice constant still makes it the ideal material as a virtual substrate for other materials. The focus of this thesis will shift to using Ge not as an active material but as a virtual substrate for lattice-matched

III-V such as GaAs and $In_{0.49}Ga_{0.51}P$ (E_g = 1.85 eV) that take better advantage of the photovoltaic effect for solar energy conversion. Chapter 4 will examine through simulation both the SiO_2 sidewall/Ge mesa structure described in this chapter as well as alternative structures potentially superior in regards to their ability to reach lower TDDs. Chapter 5 will begin with an investigation of modifying the selective growth process to the formation of blanket Ge films so as to prevent faceting of the III-V layers when they are deposited.

Chapter 4

Threading dislocation reduction

Although previous experimental results indicate low TDD values of 2.3×10^6 cm^{-2} for selectively grown mesas after cyclic annealing, these measurements were obtained by probing only the centers of 10 μm wide mesas due to the etch pit technique that is discussed in Section 5.2.1. It was assumed that the measured TDD in the mesa center was equivalent throughout the mesa, including around the mesa perimeter. Upon closer examination of Figure 2.12, several etch pits are observed along the perimeter of the mesas' flat regions. Application of Ge selectively grown structures to blanket, large area films will require elimination of threading dislocations throughout the entire film region. Further, achieving dislocation densities below 10^6 cm^{-2} will be necessary for Ge virtual substrates if they are to be used for GaAs n^+/p solar cells or lower E_g materials such as Ge or InGaAs.

The mesa square structure is first analyzed in Section 4.1 by thermal stress simulations to identify potential roadblocks to complete dislocation glide due to deviations from the simpler plane-stress result of blanket Ge-on-Si. Analysis is not limited to Ge mesas grown in SiO_2 trenches as alternative structures are found to provide more favorable conditions for complete threading dislocation removal. Section 4.2 investigates the effects of strain fields from other dislocations on dislocation pinning and the necessary selection of sidewall and Ge film parameters to avoid incomplete threading dislocation removal when applying an otherwise sufficient thermal stress.

4.1 Stress simulations

In order for all glissile dislocations to glide to the edges of the Ge crystal during the annealing process, the resolved shear stress from thermal strain on threading dislocations in eight of the twelve $\{111\}\langle110\rangle$ slip systems must provide favorable conditions. Because $\{111\}$ planes intersect a (001) oriented substrate along lines parallel to $\langle110\rangle$, threading dislocations will glide in $\langle110\rangle$ directions. These eight slip systems are defined by burgers vectors with an out-of-plane component relative to

the Si/Ge (001) interface:

$$n = (111) : \vec{b} = [\bar{1}01], [0\bar{1}1] \qquad n = (1\bar{1}1) : \vec{b} = [\bar{1}01], [011]$$
$$n = (\bar{1}11) : \vec{b} = [101], [0\bar{1}1] \qquad n = (\bar{1}\bar{1}1) : \vec{b} = [101], [011]$$

The four remaining slip systems have burgers vectors that lie entirely in the interface plane, e.g. $\vec{b} = [110]$. Threading dislocations with these burgers vectors are ignored in the following analysis because for a biaxial stress state, the resolved shear stress will be zero (see Eq. 4.1) and therefore sessile, i.e. these dislocations are unable to glide. Luan noted that although before cyclic annealing, both glissile and sessile threading dislocations were observed in the as-grown blanket Ge films, once they were annealed, only glissile threading dislocations were observed in cross-sectional TEM. This result was attributed to the fact that of the 144 possible interactions between two dislocations of any burgers vector combination, 4.5× as many reactions result in a glissile dislocation versus a sessile dislocation.[89] For systems of large lattice misfit, such as Si/Ge, many of the misfit dislocations that form are pure edge dislocations with their burgers vectors entirely in the interface plane, exactly 90° away from the misfit dislocation line direction.[40] These 90° dislocations are most efficient at plastically relaxing lattice misfit strain. The other dislocation type, the "60° dislocation," has its burgers vector directed partially out of the plane (e.g. $\vec{b} = [101]$), thus, less effective at strain relaxation, but these dislocations are glissile. The latter type are the dislocations on which the simulations below focus.

4.1.1 Thermal cycling

Geometries of different trench structures and the Ge films within those trenches are modeled using COMSOL Multiphysics. Each geometry is assumed elastically relaxed at 850°C, and after reducing the temperature to 650°C without allowing any plastic relaxation (generation or glide of dislocations), the thermal stress σ is recorded. To determine the resolved shear stress τ_{res} from an arbitrary stress state σ on a slip system defined by slip plane $\hat{n}$ and burgers vector $\vec{b}$, the fraction of each component of σ that contributes to shear stress on the glide plane in the slip direction are

summed:

$$\tau_{\hat{n},\hat{b}} = \sum_{i,j} \hat{b}_i \hat{n}_j \sigma_{ij} \tag{4.1}$$

The direction of the calculated resolved stress is in the glide plane and perpendicular to the dislocation line direction (despite not appearing in Eq. 4.1). Material properties included in the model include coefficients of thermal expansion (CTE) as a function of temperature ($T\,[=]\,K$) as well as elastic stiffness tensor constants listed in Table 4.1:

$$\begin{aligned}
\alpha_{\mathrm{Si}}{}^{[90]} &= 3.725 \times \left[1 - \exp\left(-5.88 \times 10^{-3}\,(T-124)\right) + 5.548 \times 10^{-4} T\right] && \text{ppm/K} \\
\alpha_{\mathrm{Ge}}{}^{[91]} &= 5.04 + 3.79 \times 10^{-3} T - 3.50 \times 10^{-7} T^2 && \text{ppm/K} \\
\alpha_{\mathrm{SiO2}}{}^{[92]} &= 0.55 && \text{ppm/K}
\end{aligned}$$

(GPa)	C_{11}	C_{12}	C_{44}	$C'_{\{11,22\}}$	C'_{12}	C'_{66}
Si[93]	166	63.9	79.6	194	35.4	51.1
Ge[93]	129	47.9	67.0	155	21.5	40.6
SiO_2[92]	70.9	14.5	28.2			

Table 4.1: Elastic constants of Si, Ge, and SiO_2

The structural model in COMSOL is oriented so that $\vec{x}$ and $\vec{y}$ correspond to in-plane $\langle 110 \rangle$ directions, the horizontal threading dislocation glide directions for (100) substrates. Values of C_{11}, C_{12}, C_{22}, and C_{66} are therefore modified for Si and Ge by a 45° rotation in the x-y plane[94] due to their anisotropic stiffness tensors, the results of which are also included in Table 4.1 for reference. Similarly, the components σ_{ij} must also undergo a 45° rotation in order to correctly calculate the resolved shear stress for each slip system. As an example, the explicit expressions for τ_{res} of the slip system, $\hat{n} = \left(111\right) : \hat{b} = \left[\bar{1}01\right]$, before and after rotation are as follows:

$$\tau_{\mathrm{res}} = \frac{-\sigma_{xx} + \sigma_{zz} - \sigma_{xy} + \sigma_{yz}}{\sqrt{6}} \tag{4.2a}$$

$$\tau'_{\mathrm{res}} = \frac{\sqrt{2}\left(-\sigma_{x'x'} + \sigma_{zz} + \sigma_{x'y'}\right) + \sigma_{x'z} + \sigma_{y'z}}{\sqrt{12}} \tag{4.2b}$$

where x', y' are the lateral directions as defined in the simulation.

Although there are eight relevant slip systems, the symmetry of the mesa square structure allows for an initial reduction to four in the analysis. Further, because the two relevant burgers vectors for

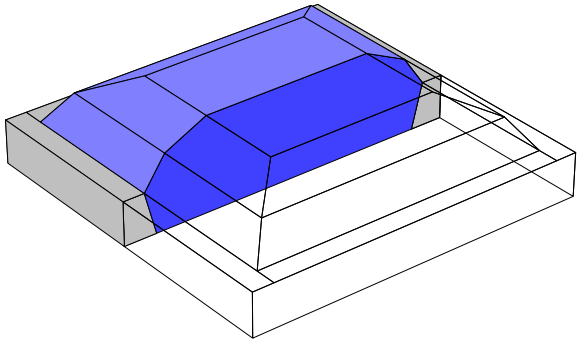


Figure 4.1: Cross-section taken through the center of a Ge mesa (blue) surrounded by SiO_2 sidewalls (grey).

each slip plane produce reflected results of the resolved shear stress, the number of slip systems to observe can be further reduced to two when only considering thermal stresses. Vertical slices are made in the simulation results that run parallel to the ⟨110⟩ glide directions of these slip planes to observe the resolved shear stress experienced by threading dislocation segments as they glide. The existence of favorable conditions for complete dislocation glide out of the selectively-grown Ge films can be identified by analysis of the resolved shear stress fields observed in these vertical slices. An example of a slice through a Ge mesa surrounded by SiO_2 trench walls is schematically shown in Figure 4.1.

In order to simulate the Ge mesa/trench on a thick Si substrate, a 50 μm thick slab of Si is included under the structure. Due to the repetitive structure of Ge/trench structure and because an actual wafer's lateral dimensions are much larger than its thickness, the boundary conditions are defined as follows: The $-x$ and $-y$ boundary planes of the unit cell are fixed in x and y, respectively, while the $+x$ and $+y$ boundaries are free to move but must maintain their original surface normals. These "straight-face" boundary conditions allow for free movement in the direction normal to the boundary planes but only by a uniform displacement across the plane, allowing the substrate to expand or contract to its relaxed state by elastic deformation.[95] The combination of the fixed and straight-face boundary conditions are equivalent to forcing symmetry at all four lateral boundaries, enabling the result to accurately describe the stress conditions in a repetitive structure containing the unit cell simulated. The bottom boundary plane is constrained in z while free boundary conditions are defined for the top boundaries planes. Lastly, displacement continuity is held across internal boundaries shared between the Ge film, the Si substrate, and the surrounding sidewall material.

4.1.2 Thermal stresses in mesa arrays

Thermal stresses due to cyclic annealing is first investigated for a baseline structure composed of 7 μm wide mesas grown between 1 μm thick and 1 μm wide SiO_2 lines with a sidewall angle of 75° (the line width is always defined by the width at the line's top surface). The mesas have completely filled the trench and have formed {111} and {311} facets at the edges while retaining a (001) facet in the center. After defining the structure to be perfectly relaxed at 850°C, the temperature is reduced to 650°C, and the resulting thermal stress, σ_{ij}, without allowing for plastic relaxation to occur, is recorded. Several vertical slices are made through the mesa at various distances between the mesa center and edge to observe how the resolved stress fields evolve. Three slices are included for the $\left(\bar{1}11\right)\left[101\right]$ and $\left(1\bar{1}1\right)\left[011\right]$ slip systems in Figure 4.2. Both of these slip planes are vertically tilted from the plane of the page so that one can visualize the resolved stress across the horizontal path that a dislocation section glides across in its slip plane.

For slices near the mesa center, the resolved shear stress is relatively uniform except near the center facet and the boundaries with the SiO_2 lines. The resolved shear stress near the mesa edge approaches zero due to thermal expansion compensation between the Ge mesa and SiO_2 lines. While the Ge mesa and the Si substrate have a CTE mismatch of approximately 2.7 ppm/K in the temperature range of interest, SiO_2 has a similar mismatch but of opposite sign (relative to Si) of approximately 2.1 ppm/K. As the structure's temperature is reduced, Ge elastically contracts while the SiO_2 lines expand relative to the Si substrate. The creation of a thermal stress is thus made difficult near the interface for the material combination in this structure. In the absence of other stresses, the lack of a resolved shear stress in Ge near the mesa edge can prevent dislocations from gliding once they approach the mesa edge. During the next annealing half cycle in which the temperature change is reversed in sign, those same dislocations will experience a resolved shear stress in the opposite direction, moving them away from the Ge film edge. Ideally, the resolved shear stress should be of maximum magnitude and uniform in sign throughout the mesa.

As the slice extends closer to the mesa edge, the sign of the resolved shear stress inverts at the top surface of the mesa and eventually at the bottom as well while the resolved shear stress at the center remains unchanged in sign. These regions of inverted shear will cause different segments of a given threading dislocation to experience resolved shear stresses in opposing directions. Dislocations do

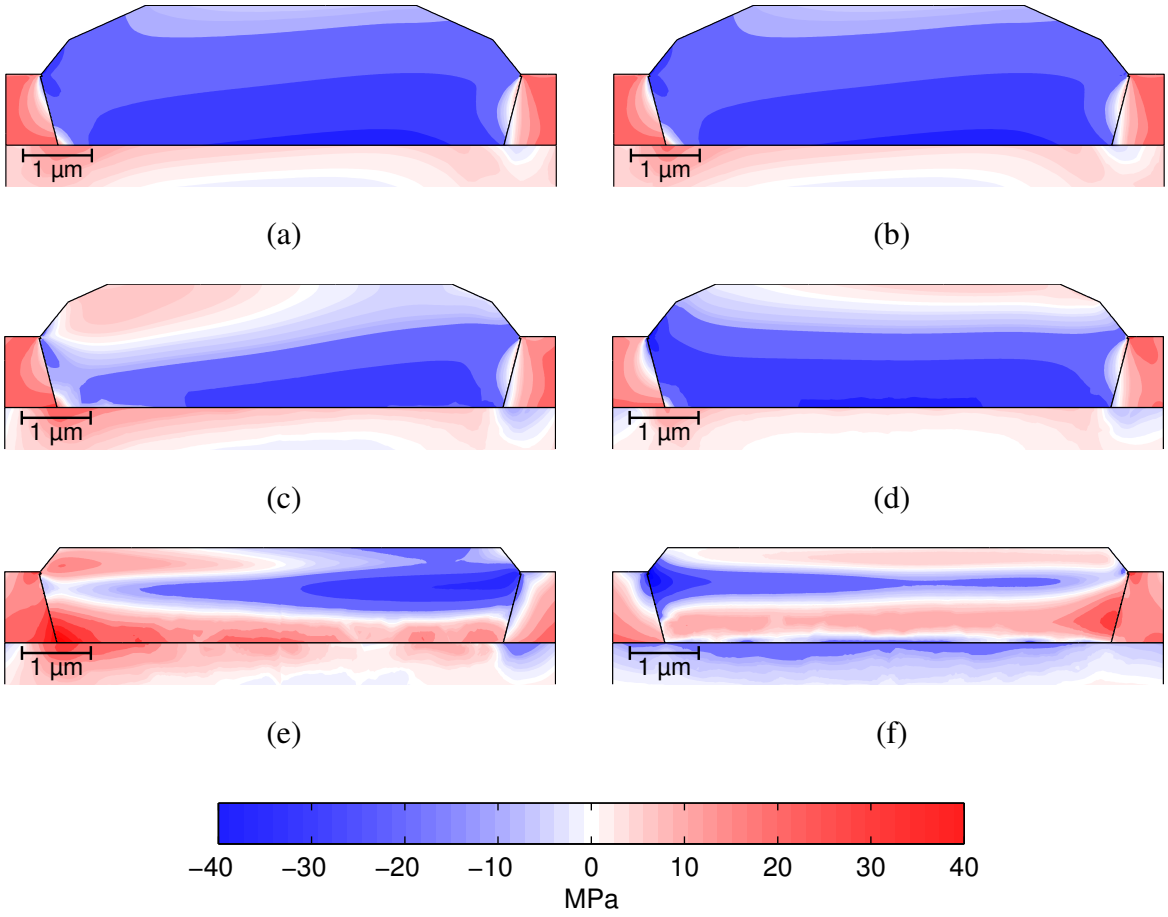


Figure 4.2: Thermal resolved shear stress in 7 μm mesa in vertical slices (a,b) through the mesa center, (c,d) 1 μm from the mesa edge, and (e,f) 300 nm from the mesa edge for slip systems (a,c,e) $\left(\bar{1}11\right)\left[101\right]$ and (b,d,f) $\left(1\bar{1}1\right)\left[011\right]$.

not prefer to extend their line length (as doing so increases the crystal's elastic energy) unless the extension leads to a reduction in another form of energy (e.g. misfit extension to reduce strain due to lattice mismatch). Extension of the threading component of a dislocation within the mesa does not relax thermal strain at the Si/Ge interface, and thus, the inverted shear regions are expected to prevent dislocations from completely gliding to the mesa's edges. For smaller mesas for which non-(100) facets compose a greater fraction of the surface, inverted shear regions are observed throughout the mesa, even in slices through the mesa center (see Figure 4.3). The increased relaxation of biaxial strain in Ge-on-Si mesas due to CTE mismatch has previously been observed for mesas with smaller widths, particularly when below 10 μm.[96]

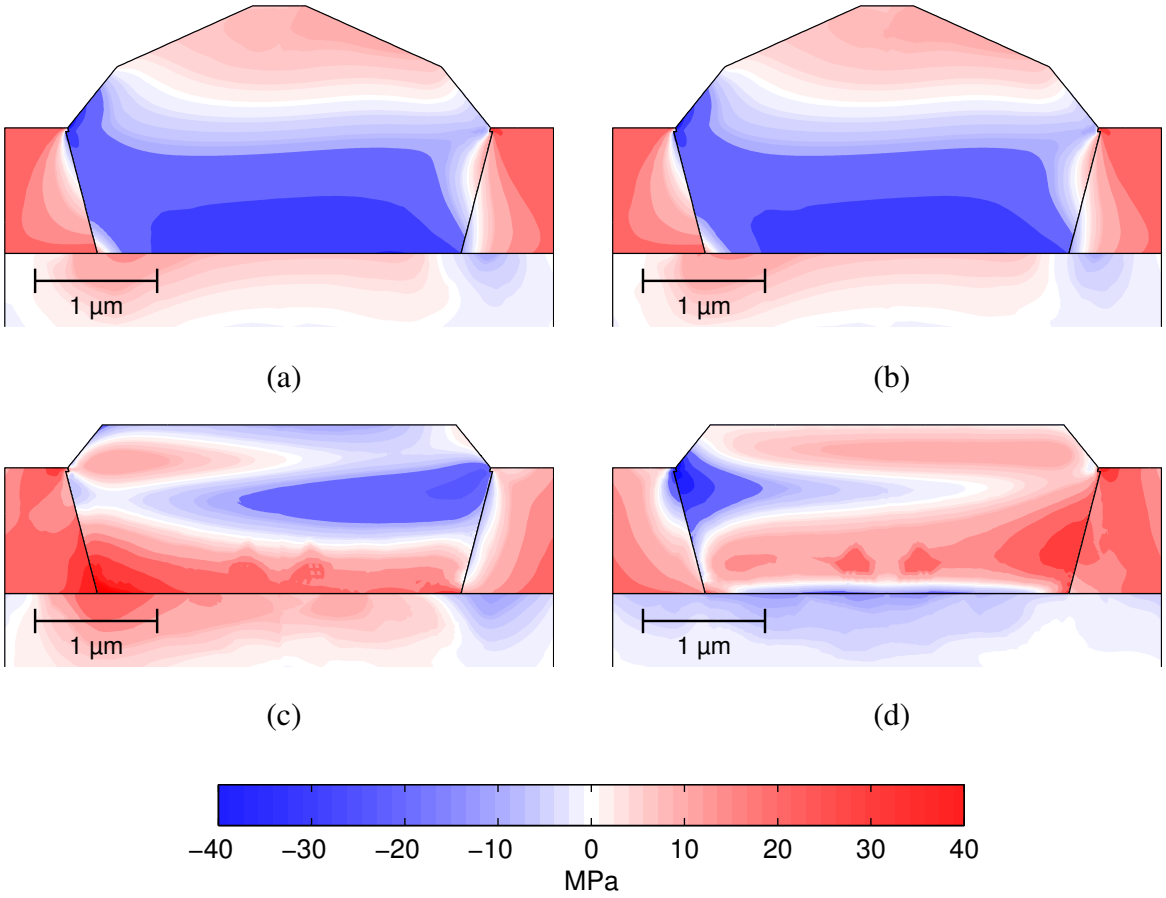


Figure 4.3: Thermal resolved shear stress in a 3 μm mesa in vertical slices (a,b) through the mesa center and (c,d) 300 nm from the mesa edge for slip systems (a,c) $\left(\bar{1}11\right)\left[101\right]$ and (b,d) $\left(1\bar{1}1\right)\left[011\right]$.

The faceted surface of a mesa enables elastic relaxation of the thermal strain, $\epsilon^{\mathrm{t}} = \Delta\alpha\Delta T$, reducing the final stresses due to thermal cycling.[67,97] In the limit of a faceted mesa strip infinitely long in the $\vec{y}$ direction, strain relaxation can only occur in directions with non-zero components in the facets' surface normals, i.e. the transverse direction and the vertical direction, $\vec{x}$ and $\vec{z}$. The mesa will adjust the overall strain components, $\epsilon_{ij} = \epsilon^{\mathrm{e}}_{ij} + \delta_{ij}\epsilon^{\mathrm{t}}$, to minimize its elastic energy:

$$E_{\mathrm{e}} = \frac{1}{2}\int_V \sum_{i,j} \sigma_{ij}\epsilon^{\mathrm{e}}_{ij} = \frac{1}{2}\int_V \sum_{i,j} \sigma_{ij}\left(\epsilon_{ij} - \delta_{ij}\epsilon^{\mathrm{t}}\right) \tag{4.3}$$

where ϵ^{e} is the elastic strain and $\delta_{ij} = 1$ for $i = j$ and 0 otherwise. For complete relaxation throughout an arbitrary volume, assuming strain components can be modified independently of each other (an incorrect assumption made only for purposes of the simplified calculation below), $\epsilon_{\mathrm{xx}} = \epsilon_{\mathrm{zz}} = \Delta\alpha\Delta T$,

while the remaining ϵ_{ij} components are zero. Changes in σ_{ii} due to thermal strain relaxation can be estimated for a material with isotropic elastic constants (shear stresses are ignored as ϵ^{t} does not directly affect them):

$$\begin{bmatrix}\sigma_{xx}\\ \sigma_{yy}\\ \sigma_{zz}\end{bmatrix} = \begin{bmatrix}C_{11} & C_{12} & C_{12}\\ C_{12} & C_{11} & C_{12}\\ C_{12} & C_{12} & C_{11}\end{bmatrix}\begin{bmatrix}\epsilon_{\mathrm{xx}}\\ 0\\ \epsilon_{\mathrm{zz}}\end{bmatrix} - (C_{11}+2C_{12})\,\Delta\alpha\Delta T\begin{bmatrix}1\\ 1\\ 1\end{bmatrix} \tag{4.4}$$

Compared to the thermal stresses obtained in a blanket film for which only $\epsilon_{\mathrm{zz}} = \Delta\alpha\Delta T$, in a mesa strip infinitely long in the $\vec{y}$ direction, σ_{xx}, σ_{yy}, and σ_{zz} are reduced by factors ν, $1-\nu$, and 0.5, respectively, if faceting allows for complete thermal strain relaxation in $\vec{x}$ and $\vec{z}$ throughout the mesa strip.

In an actual structure defined by mesa facets, strain relaxation is determined by satisfaction of boundary conditions on the free surfaces of the mesa facets and the internal boundaries of the mesa with the sidewall material and substrate. All tractions on free surfaces (e.g. Ge mesa facets) must tend to zero, and tractions must be continuous across all boundaries between the Ge mesa, SiO_2 sidewalls, and Si substrate:

$$\sigma^{(1)}\cdot\hat{n}_k = 0 \tag{4.5a}$$

$$\sigma^{(1)}\cdot\hat{n}_k = \sigma^{(2)}\cdot\hat{n}_k \tag{4.5b}$$

$\sigma^{(1)}$ and $\sigma^{(2)}$ are the stress states on the surface of both materials that share boundary k. The normal components σ_{ii} that factor into the resolved shear stress for the two slip systems shown in Figures 4.2 and 4.3 include only σ_{yy} and σ_{zz} since $\hat{n}_x = 0$ for the glide planes (see Eq. 4.1). For the facets visible in the cross-section, $\hat{n}_{\mathrm{y}} = 0$, and thus, no first order effect from the facet surfaces exists for ϵ_{yy}. Minor reduction to σ_{yy} occurs due to the non-zero value of ν. The changes to σ_{zz} are primarily due to the SiO_2 sidewalls that both prevent the film from completing relaxing in $\hat{z}$ but allow local relaxation in $\hat{x}$, the exact opposite of what occurs in blanket Ge-on-Si films. The only non-zero shear stress is σ_{xz} which arises from both the facets and the SiO_2 sidewall interface whose surface normals have both $\hat{x}$ and $\hat{z}$ components. The non-zero components of σ_{ij} are shown in Figure 4.4 for a vertical slice through the center of a 7 μm mesa (an approximation for an infinitely

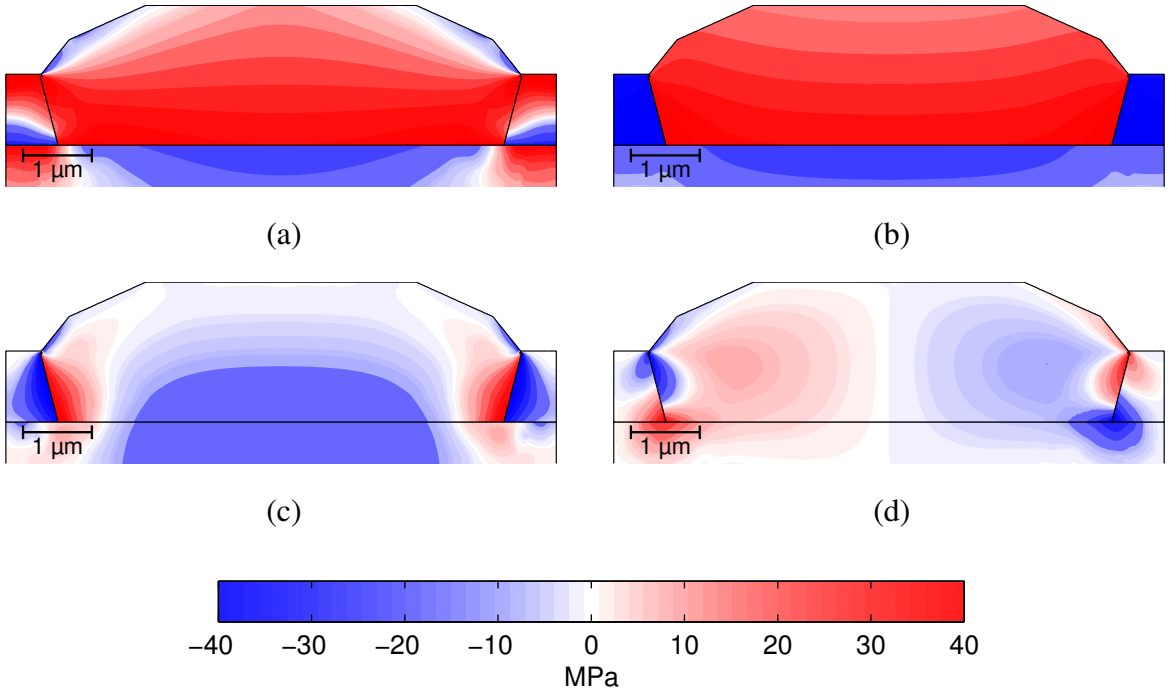


Figure 4.4: Thermal stress components in a vertical slice through the center of a 7 μm mesa: (a) σ_{xx}, (b) σ_{yy}, (c) σ_{zz}, (d) σ_{xz}.

long mesa strip). While the boundary conditions only define the stress conditions at the mesa facet surfaces and internal boundaries, stress and displacement continuity propagates the relaxation (and lack thereof) into the mesa interior. For the vertical slice through the center of the 7 μm wide mesa, the stress relaxation is limited in extent of the mesa's cross-section. However, for the smaller 3 μm mesa in Figure 4.3, the effects on τ_{res} from the facets and SiO_2 sidewalls are more pronounced throughout the mesa as a greater portion of the structure is defined by facets, and the facets and SiO_2 lines on opposing sides are also in closer proximity to each other.

While the tractions on the mesa facets will tend to zero, the individual stress components as defined in the $\vec{x}$, $\vec{y}$, $\vec{z}$ coordinate system will not necessarily be zero unless the facet's normal is exactly aligned to $\vec{x}$, $\vec{y}$, or $\vec{z}$. Thus, stress components such as σ_{xx} in Figure 4.4a, may even reverse in sign on the facets to simultaneously satisfy the boundary conditions of that facet while maintaining displacement and stress continuity with the rest of the structure. As the slice through the mesa is taken closer to the mesa edge, facets in and out of the page with non-zero $\hat{n}_{\text{y}}$ components allow for strain relaxation in all directions, including $\vec{y}$. The effect on σ_{yy} is identical to that of σ_{xx} in

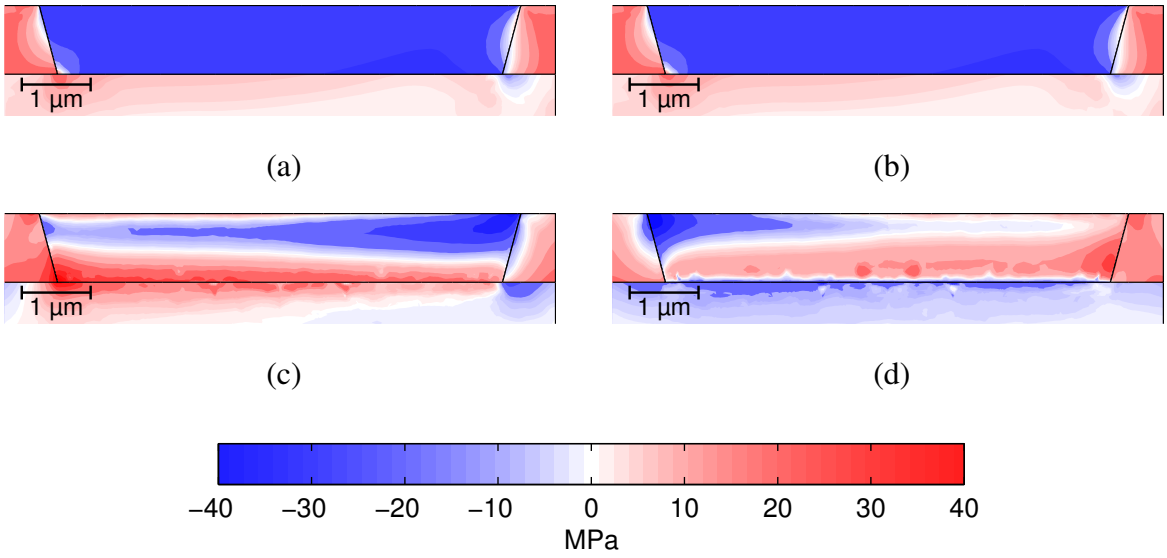


Figure 4.5: Thermal resolved shear stress in a facet-free 7 µm mesa with a SiO_2 sidewall angle of 75° in vertical slices (a,b) through the mesa center and (c,d) 300 nm from the mesa edge for slip systems (a,c) $\left(\bar{1}11\right)\left[101\right]$ and (b,d) $\left(1\bar{1}1\right)\left[011\right]$.

Figure 4.4a, just rotated 90° in the xy plane. Shear stresses σ_{xy} and σ_{yz} also appear as the vertical slice approaches the mesa edge.

The resolved shear stress for each slip system is determined by all components of σ_{ij} except for either σ_{xx} or σ_{yy} (depending on the slip system) in the simulation's coordinate system. The components' varied magnitude and sign, when combined, are responsible for the regions of inverted resolved shear stress earlier observed. While the effect of the shear stress components on τ_{res} is different for each slip system, a common factor, $\sigma_{zz} - \sigma_{ii}$ (where $i = x$ or y) is shared in the calculation for all slip systems. Thus, to prevent inverted shear regions on all slip systems simultaneously, all shear components of σ_{ij} must be minimized in magnitude, and the difference between σ_{zz} and σ_{ii} must be both maximized and of the same sign throughout the Ge mesa, i.e. the biaxial stress state of a blanket film of Ge on Si.

To observe the effects of faceting, the baseline 7 µm mesa structure is modified by removing any Ge above the SiO_2 lines while maintaining complete trench fill. By removing the facets, the free surface boundary condition can no longer directly constrain σ_{xx}, σ_{yy}, or σ_{xy}. The resulting resolved shear stress for the two slips systems tilted vertically from the page plane are shown for vertical slices through the mesa center and near the mesa edge in Figure 4.5. While removing the

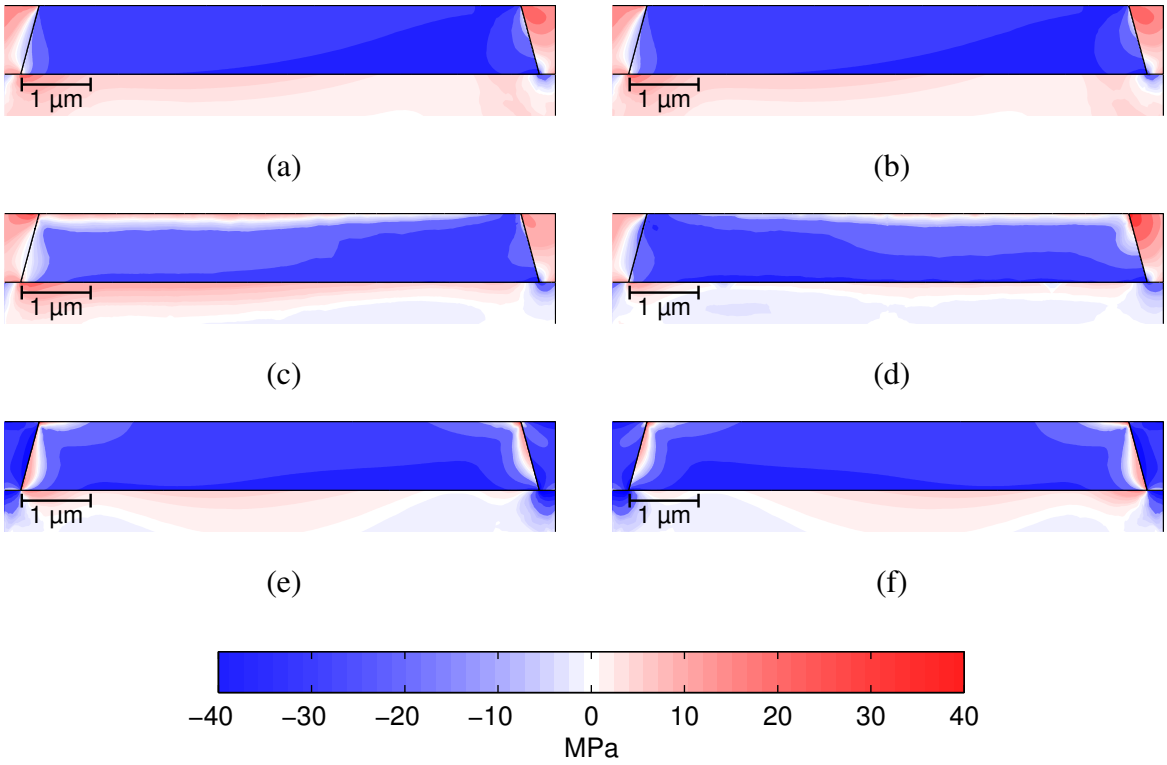


Figure 4.6: Thermal resolved shear stress in a facet-free 7 µm mesa with a SiO_2 sidewall angle of 105° in vertical slices (a,b) through the mesa center, (c,d) 300 nm from the mesa edge, and (e,f) through the mesa center but perpendicular to the slices in (a-d) for slip systems (a,c,e) $\left(\bar{1}11\right)\left[101\right]$ and (b,d,f) $\left(1\bar{1}1\right)\left[011\right]$.

mesa facets segregates the inverted shear region to smaller regions within the vicinity of the SiO_2 lines, these potentially problematic stress fields are not eliminated. The SiO_2 lines are then modified by inverting the sidewall angle from 75° to 105°. With this additional alteration, the inverted shear regions further reduce in size as shown in Figure 4.6.

The component of σ_{ij} primarily responsible for the change in the resolved stress field between the two facet-free structures with different sidewall angles is σ_{yz}, compared in Figure 4.7a,c in a cross-section 300 nm from the mesa edge. The cause for this difference in stress fields can more easily be elucidated by looking at a cross-section of the structure perpendicular to the those shown in Figure 4.7ac. Perpendicular, vertical slices of the stress fields made through the mesa center are displayed in Figure 4.7b,d for both facet-free structures. The section of the Ge mesa in Figure 4.7b bordering the left SiO_2 line that appears in dark blue (large negative stress values) corresponds to the cross-section in Figure 4.7a. When the sidewall's angle is inverted from 75° to 105°, the region

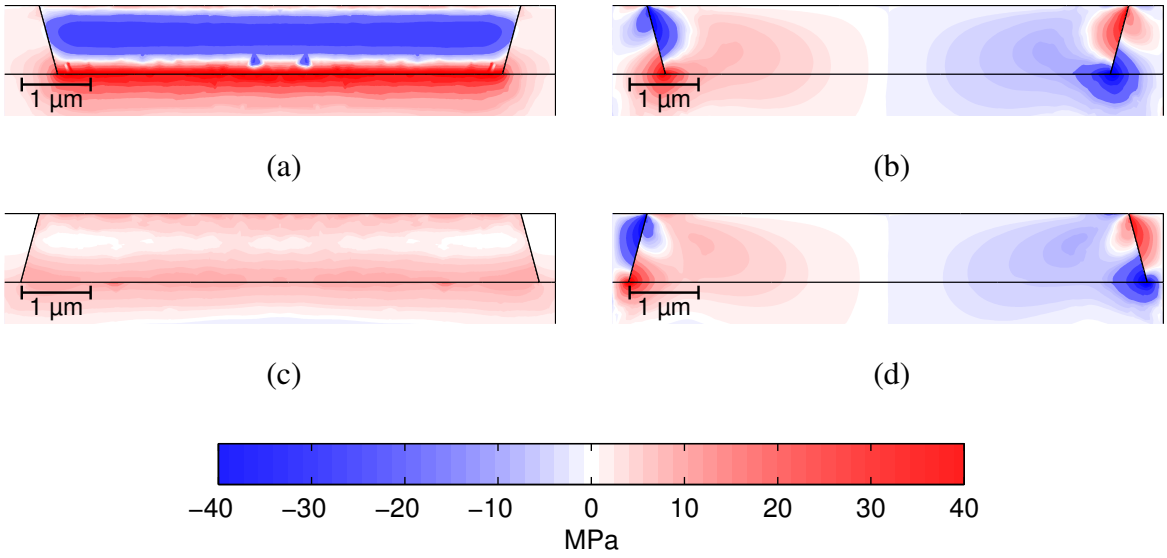


Figure 4.7: Thermal stress component σ_{yz} in a facet-free 7 µm mesa with a SiO_2 sidewall angle of (a,b) 75° and (c,d) 105° in vertical slices (a,c) 300 nm away from the mesa edge and (b,d) through the mesa center in slices perpendicular to those in (a,c).

of large negative values for the shear stress shifts to the SiO_2 line, leaving behind only a minor area in the Ge mesa still affected. The origin of this shear stress can be understood by the displacement directions in the Ge mesa and the SiO_2 line that would elastically relax the thermal strain in the two materials when the temperature is reduced. At the Ge/SiO_2 interface on the left SiO_2 line in Figure 4.7, the SiO_2 will expand up and to the right while the Ge mesa will contract down and to the right (the SiO_2 line cannot expand to the left due to symmetry at the boundaries). For the 75° sidewall structure, Ge's relaxation displacement direction has a large component in-plane of the Ge/SiO_2 interface, inducing the shear strain ϵ_{xz} which directly leads to σ_{xz}. When the SiO_2 sidewall angle is inverted to 105°, the relaxation direction in Ge becomes closer to the normal of the Ge/SiO_2 interface, reducing the shear strain produced in the mesa and thus, limiting any contribution to σ_{xz}. Similar reasoning can explain the reverse change in the shear stress in the SiO_2 line after inverting the sidewall angle.

In order to maximize the fraction of a Ge mesa defined by SiO_2 lines in which threading dislocations experience thermal resolved shear stress fields uniform in sign, the mesa must be facet-free during application of the thermal stress, and the SiO_2 sidewall angle must be greater than 90°. However, as one can see in Figure 4.6e,f, even with an optimum SiO_2 line structure, threading

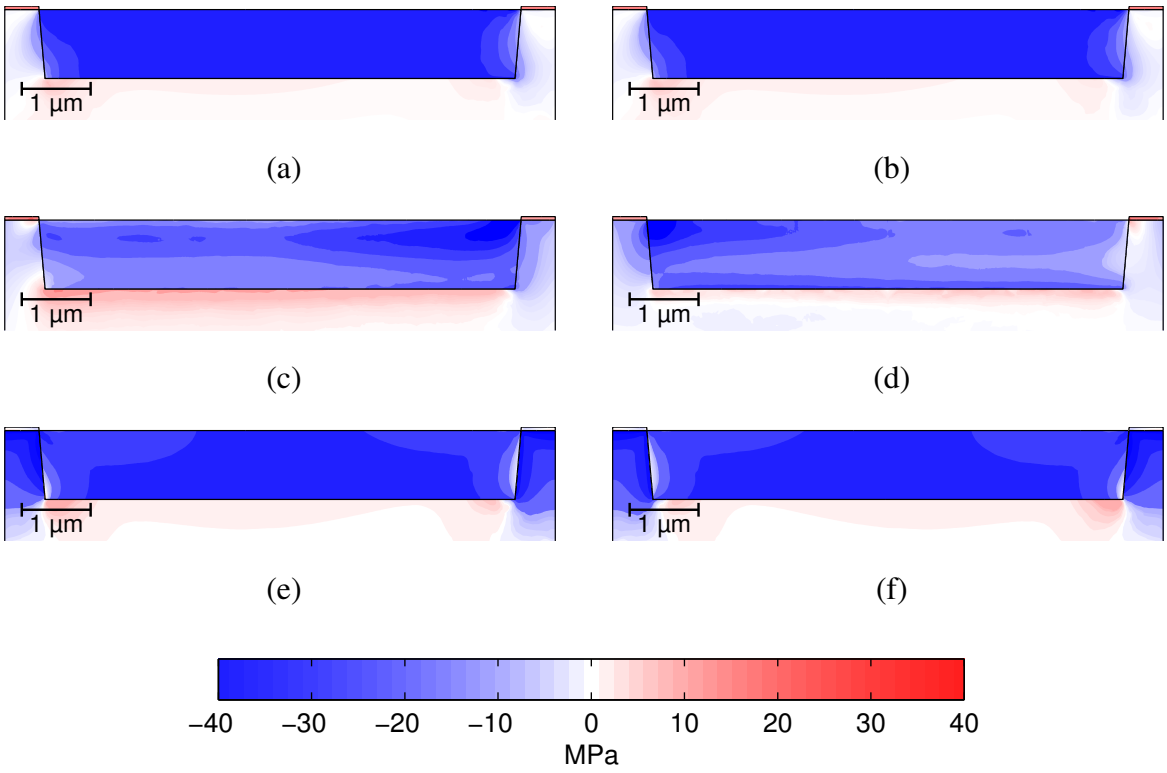


Figure 4.8: Thermal resolved shear stress in a facet-free 7 µm mesa with a Si sidewall angle of 85° in vertical slices through the (a,b) center of the mesa, (c,d) 300 nm from the mesa edge, and (e,f) center of the mesa but perpendicular to the slices in (a-d) for slip systems (a,c,e) $\left(\bar{1}11\right)\left[101\right]$ and (b,d,f) $\left(1\bar{1}1\right)\left[011\right]$.

dislocations that are within ~ 250 nm of a mesa edge that is parallel to their slip plane will still encounter regions of inverted shear.

Because CTE mismatch between Ge and SiO_2 sidewalls is the primary cause of the non-ideal stress state, a sidewall material with CTE close to that of Ge should provide better more uniform resolved shear stress fields. One immediately available option is Si as trenches can be etched directly into the Si substrate. Si's CTE mismatch with Ge is approximately half that of SiO_2. For the same sidewall angles, the resulting shear strain from the sidewalls is reduced, decreasing the regions of inverted shear. Even with a sidewall angle of 85°, the Si trench structure is able to better maintain a resolved shear stress of uniform sign as shown in Figure 4.8 across the majority of the mesa compared to SiO_2 lines with sidewall angles above 90°. Only within 200 nm of mesa corners does the shear stress begin to invert (not shown). Inclusion of a 25 nm thick SiO_2 film at the top of the Si trenches (shown in Figure 4.8) to maintain selective Ge growth between mesas does not

effectively alter the stress state in the Ge mesa. Brief discussion of the fabrication of both SiO_2 and Si sidewalls with inverted sidewall angles as well as the growth of facet-free Ge is included in Section 7.2.

4.1.3 Image stresses

The stress analysis presented thus far has only concerned the resolved shear stresses due to thermal cycling without consideration of stress fields due to the existence of the dislocation itself. These stresses are "image forces" caused by the proximity of free surfaces as well as internal boundaries between materials of different shear moduli (for which free surfaces are an extreme example). Effectively, the image force represents the reduction of elastic energy of the crystal when a dislocation moves in response to it.[39] A dislocation generates a stress field depending on its edge, screw, or mixed character. For the screw component of a dislocation with line direction, $\vec{\xi} = \vec{z}$, only shear stresses are produced:

$$\sigma_{\theta z} = \frac{Gb_{\text{screw}}}{2\pi r} \tag{4.6}$$

where r is the distance from the dislocation line, and $b_{\text{screw}} = \vec{b} \cdot \vec{\xi}$ is the screw component of the dislocation. This stress field will be non-zero at free surfaces and will create traction discontinuities at internal boundaries if not properly modified. In order to maintain the boundary conditions of Eq. 4.5a for a free surface, a virtual screw dislocation with a burgers vector equal in magnitude but opposite in sign is placed across the free surface by the same distance between the physical dislocation and the free surface. Adding the resulting stress field from the virtual dislocation satisfies the boundary condition everywhere on the free surface as well as applies an image force per unit length on the original dislocation:

$$\delta F_{\text{image}} = \frac{-Gb_{\text{screw}}^2}{4\pi r}\delta l \tag{4.7}$$

Because the two dislocations' burgers vectors are of opposite signs, the physical dislocation is always attracted to the free surface. As a dislocation approaches a free surface, half of its stress field begins to exit the crystal, lowering the crystal's elastic energy as expected. For internal boundaries

between two materials, the burgers vector for a virtual dislocation is equal to $\gamma_k \vec{b}$ where:

$$\gamma_k = \frac{G_k - G_0}{G_0 + G_k} \tag{4.8}$$

G_0 and G_k are the shear moduli of the material containing the dislocation and the material on the other side of boundary k, respectively. This adjustment guarantees continuity of stresses across the interface for pure screw dislocations and a close approximation for dislocations with an edge component.[98] The value of γ_k reduces to -1 in the case of a free surface ($G_k = 0$) but becomes positive (and thus, the image force is repulsive) if $G_k > G_0$.

In the case of a dislocation with an edge component, normal and shear stresses are both present. For a dislocation with $b_{\text{edge}} = \vec{x}$ and $\vec{\xi} = \vec{z}$:

$$\sigma_{xx} = -G' y \left(3x^2 + y^2\right) / r^4 \tag{4.9a}$$

$$\sigma_{yy} = G' y \left(x^2 - y^2\right) / r^4 \tag{4.9b}$$

$$\sigma_{zz} = -2\nu G' y / r^2 \tag{4.9c}$$

$$\sigma_{xy} = G' x \left(x^2 - y^2\right) / r^4 \tag{4.9d}$$

where $G' = \frac{G b_{\text{edge}}}{2\pi(1-\nu)}$ and $r^2 = x^2 + y^2$. Placing a virtual dislocation in the same manner used for screw dislocations will not cancel all tractions from the two superimposed stress fields on a free surface, regardless of its orientation. If $\vec{\xi} \perp \vec{n}_{\text{surface}}$, the image force will be calculated correctly at the location of the dislocation itself but not elsewhere. To determine the correct stress field at other points in space in the event one wants to calculate the effect of free surfaces on one dislocation to the forces felt by other dislocations nearby, a solution to the Airy stress function, $\psi = X(x)Y(y)$, must be found that satisfies the boundary condition of the free surface by exactly canceling the residual stress fields from the physical and virtual dislocations. Second derivatives of ψ will generate the stress field components which will be zero at the dislocation position.[39] For the purposes of thermal stress simulations in this thesis, gliding dislocations are assumed to be on average 1 μm or more away from each other, a distance great enough that both image forces from other dislocations can safely be neglected, especially when comparison is made to relevant thermal stresses. Interaction between dislocations due to their stress fields are considered later in Section 4.2.

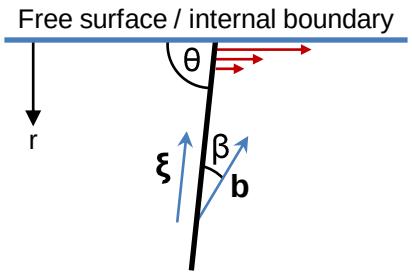


Figure 4.9: Schematic of a mixed dislocation inclined to a free surface or internal boundary. Red arrows indicate the relative magnitude of the self-force for different segments of the dislocation drawn.

For mixed or edge dislocations with line directions inclined to a free surface, the simple image force method is not sufficient to yield an accurate result even at the location of the dislocation itself. There is, however, an additional force associated with free surfaces that correctly determines the total image force felt by the dislocation. While an infinitely long, straight dislocation will not exert a force on itself, once truncated, a self-force per unit length will arise if the dislocation contains an edge component:

$$\delta F_{\text{self}} = \frac{\nu G b^2 \sin 2\beta \sin\theta}{4\pi(1-\nu)r}\delta l \tag{4.10}$$

where β is the angle between $\vec{\xi}$ and $\vec{b}$, θ is the angle between $\vec{\xi}$ and a line parallel to the free surface that is coplanar with both $\vec{\xi}$ and $\vec{b}$, i.e. the glide plane, and r is the closest distance to the free surface from the point of interest on the dislocation (see Figure 4.9).

The self-force can also be understood as a second way a free surface encourages dislocations to lower the elastic energy of the crystal. A dislocation has an associated line energy per unit length that results from the elastic work required to create the deformation:

$$\delta W_{\text{e}} = \frac{Gb^2}{4\pi}\left(\cos^2\beta + \frac{\sin^2\beta}{1-\nu}\right)\ln\left(\frac{R}{r_0}\right)\delta l \tag{4.11}$$

where R is the minimum of the distance to adjacent dislocations and the crystal's finite dimensions, and r_0 is the inner cutoff radius $\approx b$. As ν is positive for relevant materials, the energy per unit length of the dislocation reduces as the dislocation adopts a more screw character. Because the strength of the self-force is proportional to the inverse of the distance from the boundary, if this force alone determined the extent of dislocation glide, sections of the dislocation line closer to the boundary

would glide faster until the line direction of the dislocation aligned to its burgers vector, forming a pure screw ($\beta = 0°$). In practice, this additional force acts in concert with the force calculated from the virtual dislocation concept to give the overall image+self force experienced by any straight dislocation whether or not it is inclined to the boundary:

$$\delta F_{\mathrm{I+S}} = \frac{-Gb^2}{4\pi r}\left[\left(\cos^2\beta + \frac{\sin^2\beta}{1-\nu}\right)\gamma\cos\theta + \frac{\nu\sin 2\beta}{1-\nu}\sin\theta\right]\delta l \tag{4.12}$$

Due to the breakdown of elastic theory within the dislocation core, for distances $r < 4b$, the effective value for r remains equal to $4b$ for the purposes of calculation.

Similarly to the value of τ_{res}, the resolved shear stress due to thermal stresses, $\delta F_{\mathrm{I+S}}$ is also oriented perpendicular to $\vec{\xi}$ and in the glide plane. The associated image stress is:

$$\sigma_{I+S} = \frac{\delta F_{\mathrm{I+S}}}{b\,(\delta L)} \tag{4.13}$$

Thus, the total calculated resolved stress on a dislocation is determined:

$$\tau^{\mathrm{tot}}_{\hat{n},\hat{b}} = \sum_{i,j}\hat{b}_i\hat{n}_j\sigma_{ij} + \sum_k \frac{Gb}{4\pi r_k}\left[\left(\cos^2\beta + \frac{\sin^2\beta}{1-\nu}\right)\gamma_k\cos\theta_k + \frac{\nu\sin 2\beta}{1-\nu}\sin\theta_k\right] \tag{4.14}$$

where the second summation is over each boundary or free surface, k. While Eq. 4.12 provides an exact solution for the image forces on a dislocation in the presence of one free surface, the method's calculation provides only an approximation when multiple free surfaces are concerned as in Eq. 4.14. Although at least for the screw component, each image force will correctly satisfy the boundary conditions on the interface for which it is calculated, the stress field of one virtual dislocation will conflict with independently satisfying the boundary conditions on other free surfaces in the structure. Although image forces that correspond to additional virtual dislocations can be added in a recursive manner until the boundary conditions on all free surfaces are satisfied within acceptable error ranges, convergence is only guaranteed for cases with parallel sets of boundaries.[39] For two parallel surfaces, the error associated with neglecting this second order effect is at most 20% which occurs for points where image forces are minimal in magnitude. Therefore, the following simulation results neglect accounting for the minor error.

For each boundary, the image and self stress terms are first calculated separately. Each is independently reversed in sign if necessary so that positive (negative) values correspond to stresses with a positive (negative) $\vec{x}$ component, i.e. stresses that should induce dislocation glide to the right (left). While the resolved shear stress induced thermally is solely determined by the slip system of the dislocation, the image and self forces directly depend on the dislocation line direction, $\vec{\xi}$. While in practice, dislocation line directions may vary within the glide plane, for the purposes of simulation, threading dislocations are assumed to be oriented in the configuration that minimizes the elastic free energy of the Ge film. While a pure screw orientation will minimize the elastic energy per unit length (see Eq. 4.11), the dislocation can orient away from pure screw to reduce its total length. The thermodynamic equilibrium orientation is 17° away from the burgers vector, rotated towards the most vertical orientation within the glide plane,[99] similar to that drawn in Figure 4.9. When calculating the image and self stresses in each slip systems, $\vec{\xi}$ is always defined by this equilibrium orientation relative to the burgers vector of the slip system considered.

Based on the conclusions from the analysis of thermal stresses, surrounding the Ge film with a material with the identical thermal expansion coefficient as Ge would completely remove all non-uniformities in the resolved thermal shear stress fields. Si_3N_4 lines would provide a moderate improvement over Si trenches, $\alpha_{Si_3N_4}$ = 3.1 ppm/K, but both Si_3N_4 and Si have unfavorable values compared to SiO_2 for another relevant material property, the shear modulus. When the shear modulus of these two materials is greater than that of Ge (see C_{44} in Table 4.1 for Si and Ge; G for Si_3N_4 can vary between 55 and 120 GPa based on its density[100]), the image forces will repel dislocations from the mesa sidewalls, overtaking thermal stresses at close enough distances. In contrast, the low value of 30 GPa for SiO_2 makes the material preferable when considering image forces. Because $\gamma_{SiO_2} = -0.4$, the image force from an internal boundary between Ge and SiO_2 will produce an attractive force approximately 40% as strong as that from a free surface. In Figure 4.10, the image forces are shown by themselves as well as superimposed on both the thermal stress resulting from a temperature ramp from 850°C to 650°C and vice versa for a Ge mesa surrounded by SiO_2 sidewalls. During cyclic annealing, it is assumed that the structure relaxes at both temperatures after waiting a sufficient time only after which the next thermal half-cycle will begin. The apparent weak dependence of the total image+self stress from the free surface occurs because the image force and self force are always in opposing directions for the specific directions assumed for $\vec{\xi}$ in

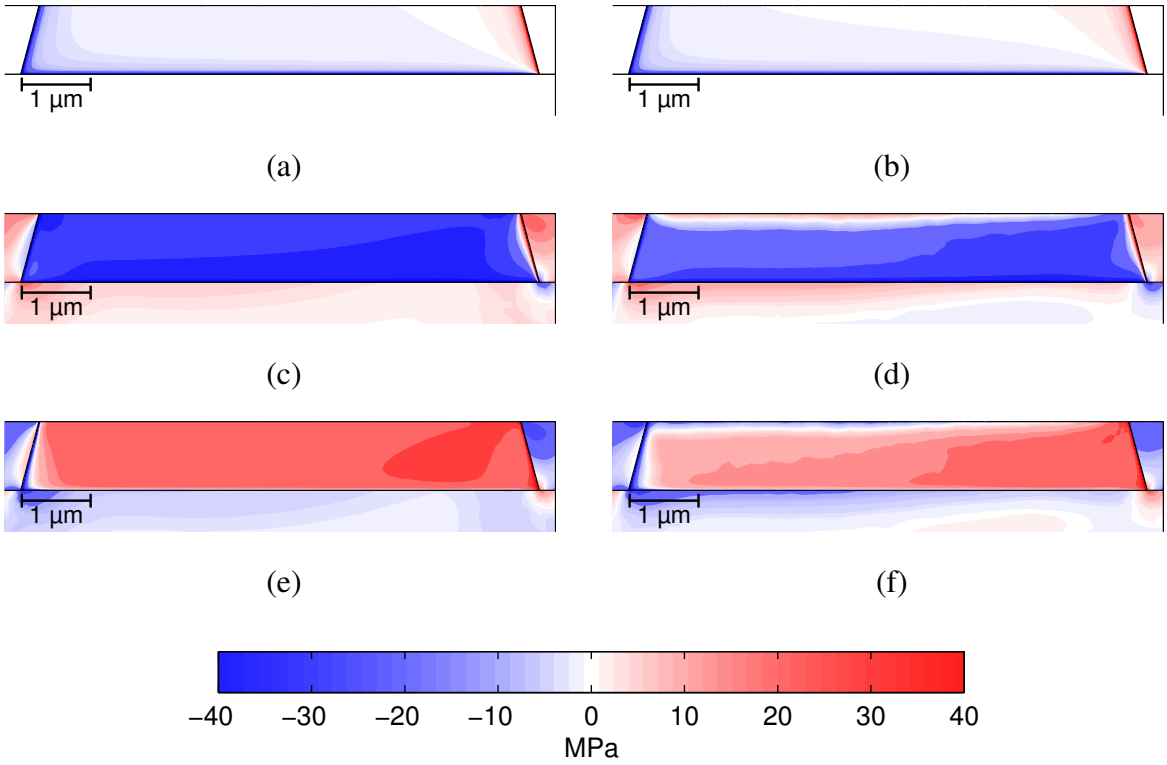


Figure 4.10: Image stress fields in a facet-free 7 µm Ge mesa surrounded by SiO_2 lines with a sidewall angle of 105° for the $\left(\bar{1}11\right)\left[101\right]$ slip system: (a,b) by itself, (c,d) super-imposed on a $\Delta T = -200°C$, (e,f) $\Delta T = +200°C$ in vertical slices (a,c,e) through the mesa center and (b,d,f) 300 nm away from the mesa edge.

the simulation. Despite $\gamma_{Si} = 0.09$, the noticeable total force from the substrate occurs because the two stress components do not counteract each other at this interface.

Regardless of the magnitude of the resolved thermal shear stresses, image and self stresses dominate the sign of the overall stress as a dislocation approaches the film edge due to the inverse relationship between distance and stress magnitude. If the walls surrounding the mesa are composed of a material with $G_k > G_{Ge}$, some dislocations (depending on their specific value of $\vec{\xi}$) will be repelled as they approach Ge film edges, irrespective of the applied thermal stress, preventing threading dislocation removal. The ideal material is therefore one that shares Ge's CTE but also has a minimal shear modulus that cannot exceed that of Ge. Many CTE matching materials have greater shear moduli (e.g. GaAs, alumina) and cannot be used based on their mechanical properties alone. One candidate material based on its CTE (5.9 ppm/K) and shear modulus (34 GPa)[101] is Zirconium metal but questions remain on material compatibility, e.g. Zr film patterning and

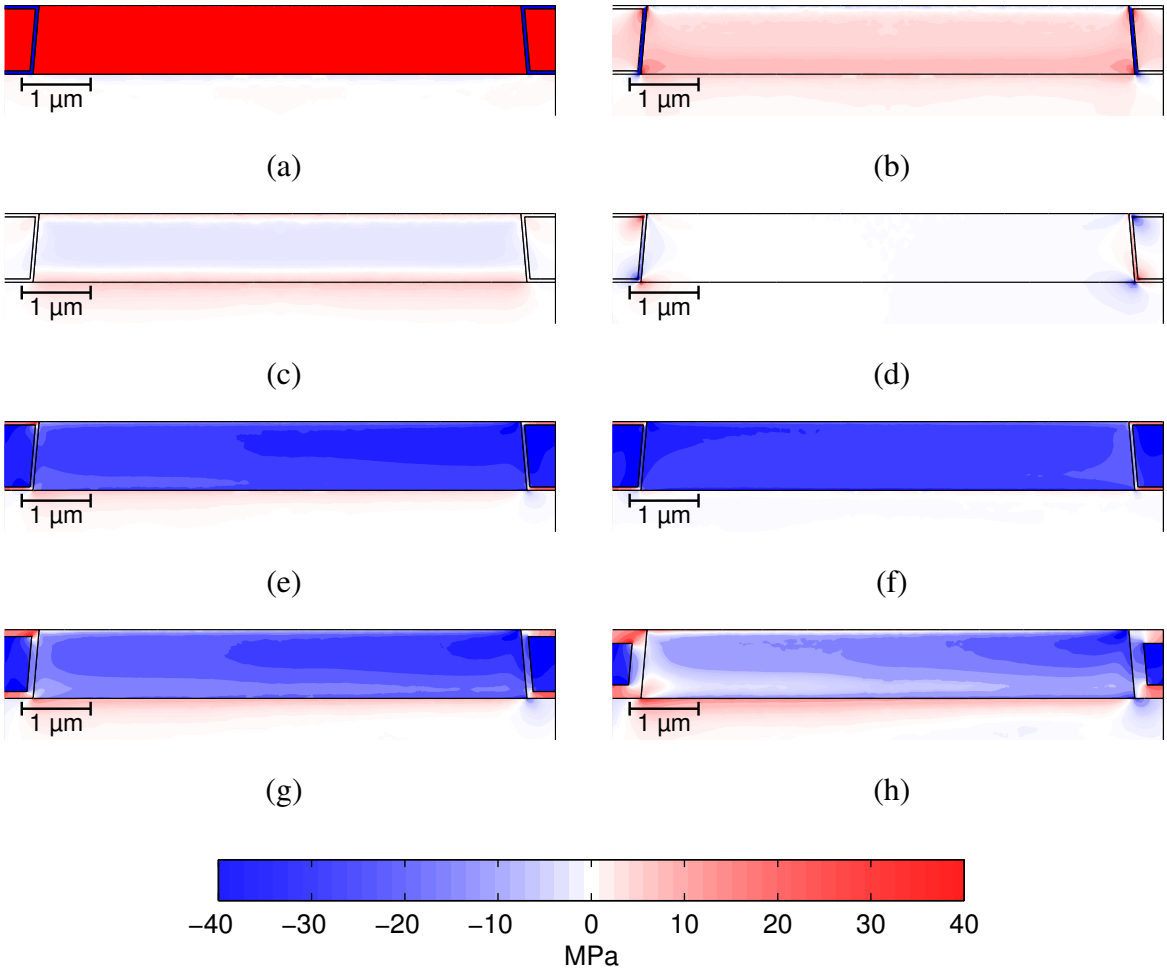


Figure 4.11: Thermal stress fields 100 nm away from the mesa edge in a facet-free 7 µm Ge mesa surrounded by SiO_2 coated Ge lines with a sidewall angle of 95° including (a) σ_{yy}, (b) σ_{zz}, (c) σ_{yz}, (d) σ_{xz} and resolved shear stress for the (e,g,h) $\left(\bar{1}11\right)\left[101\right]$ and (f) $\left(1\bar{1}1\right)\left[011\right]$ slip systems. The SiO_2 thickness is (a-f) 50 nm, (g) 100 nm, (h) 200 nm.

selective Ge growth. A thin oxide layer could potentially be deposited over the Zr lines to maintain selective growth. Another option that certainly ensures material compatibility is Ge itself coated with a thin layer of SiO_2, the fabrication of which is described in Chapter 7.

SiO_2 coated Ge lines will provide a better effective CTE compared to that of a pure SiO_2 line. While the image+self forces will be reduced as the shear moduli of the Ge film and surrounding lines are very similar, for distances within the thickness of the SiO_2 layer, these forces will be determined by the shear modulus of the SiO_2 as most of the strain field from the virtual dislocation will exist within that layer.[98] The thermal resolved shear stresses are shown for Ge lines surrounded and

capped by layers of SiO_2 in Figure 7.2 in slices only 100 nm away from the mesa edge (where problematic stress fields are most likely to appear). σ_{yz} and σ_{xz} are included for comparison with Figure 4.7.

Ge lines coated with a thin layer of SiO_2 is clearly seen as the optimal structure in the τ_{res} stress fields of Figure 7.2e,f. σ_{xx} and σ_{yy} are greater in magnitude than all other stress components, and σ_{zz} is near zero except very close to the mesa perimeter. Due to the inverted sidewall angle, the remnant inverted shear regions (as shown in Figure 7.2d) are contained within the SiO_2 coated Ge, not the Ge of the mesa. The glide and reduction of dislocations within the Ge lines are not relevant to the final TDD as this Ge only serves as a structural component to reduce the TDD in the primary Ge films. The effectiveness of the surrounding Ge lines in maintaining resolved shear stresses of uniform sign reduces as the thickness of the SiO_2 layer used to encapsulate the Ge lines is increased from 50 nm to 200 nm as seen in comparison of Figure 7.2e,g,h. Thus, a trade-off may occur between the fabrication process tolerances of the SiO_2 film thickness and the effectiveness of the SiO_2 coated Ge lines in reducing the TDD.

4.2 Dislocation glide simulations

With the optimal patterned structures determined in Section 4.1, the analysis now shifts to the effect of the mesa geometry on dislocation glide, assuming a uniform stress field provided by thermal strain during cyclic annealing. Several models have previously been developed to predict the effects of blanket film thickness and mesa lateral dimensions in the case of selectively grown films for threading dislocation reduction. These models are first reviewed, and a subset is unified to build a predictive model relevant to selectively grown Ge-on-Si films.

4.2.1 Effect of film thickness

During the initial growth of Ge on Si, threading dislocations on the order of 10^9 cm^{-2} are formed from an array of misfit dislocations at the Si/Ge interface. Because {111} glide planes are not strictly vertical on (100) substrates, the line directions of threading dislocations do not align with the surface normal. Thus, as the Ge film grows, the location of each threading dislocation at the film surface moves, even in the absence of dislocation glide (see Figure 4.12). If two threading

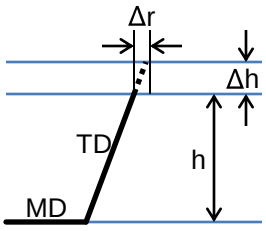


Figure 4.12: Schematic of the changing position of a threading dislocation at the film surface during film growth in the absence of threading dislocation glide.

dislocations approach each other within a specified "reaction radius", r_a, attractive forces between the dislocations can induce dislocation fusion or annihilation reactions. Over the course of film growth, each threading dislocation sweeps out an area within a reaction radius, $2r_a dr$, where dr is the horizontal glide of the dislocation solely due to film growth, thus coming in contact with $dN = \rho_{TD}(2r_a dr)$ adjacent dislocations with which dislocation interaction is possible. Assuming a linear relationship between the dislocation lateral movement and film growth, $dr/dh = c$, a reaction radius independent of film thickness, and the reduction in ρ_{TD} to be a second order reaction between adjacent dislocations, the functional dependence of the TDD with film thickness can be determined:[102]

$$d\rho = -\rho dN = -K\rho^2 dh \tag{4.15a}$$

$$\rho = \frac{\rho_0}{\rho_0 K(h-h_0)+1} \approx \frac{1}{Kh} \tag{4.15b}$$

where $K = 2r_a c$ and ρ_0 is the initial TDD at initial film thickness, h_0. With typical values of $r_a = 0.5$ μm, $c \sim 1$, $\rho_0 = 10^9$ cm^{-2}, $h = 1$ μm, $\rho_0 Kh \gg 1$, and so the effect of film thickness on the TDD can be estimated by the simple inverse relationship, $\rho \propto h^{-1}$.

Other models focus on the determination of the reaction radius to predict the effect of film thickness on TDD. Two adjacent dislocations on the same glide plane, if of opposite burgers vectors, will experience an attractive force (assuming perfect screw orientation):

$$F_a = \frac{Gb^2 h}{2\pi r \cos\phi} \tag{4.16}$$

where ϕ is the angle between the threading dislocation and the film surface normal. Opposing this force will be the line tension of the dislocation:

$$F_{\mathrm{L}} = \frac{Gb^2}{4\pi}\left(\cos^2\beta_{\mathrm{MD}} + \frac{\sin^2\beta_{\mathrm{MD}}}{\delta(1-\nu)}\right)\ln\left(\frac{R}{4b}\right) \tag{4.17}$$

where β_{MD} is specifically for the misfit section of the dislocation and $\delta \geq 1$ reduces the line tension of the edge component if its extension reduces the misfit strain due to lattice mismatch between the film and substrate (one estimate suggests $\delta = 4$).[103] The maximum distance between which dislocations will glide to each other due to the attractive force F_{a} and react is determined by the distance r_{FB} for which these two forces balance. Misfits segments are, on average, expected to separated by a distance:

$$R = \frac{b\sin\beta_{\mathrm{MD}}\cos\alpha}{f} \tag{4.18}$$

where α is the angle between the burgers vector and the film surface and f is the lattice misfit between the film and substrate. Assuming lack of interaction between dislocations on different glide planes and an average threading dislocation separation on a given glide plane to be twice the distance at which $F_{\mathrm{a}} = F_{\mathrm{L}}$,[103] the TDD as a function of thickness can be estimated (for the simple case of a pure edge oriented in the film plane, $\beta_{\mathrm{MD}} = 90°$, $\alpha = 0°$):

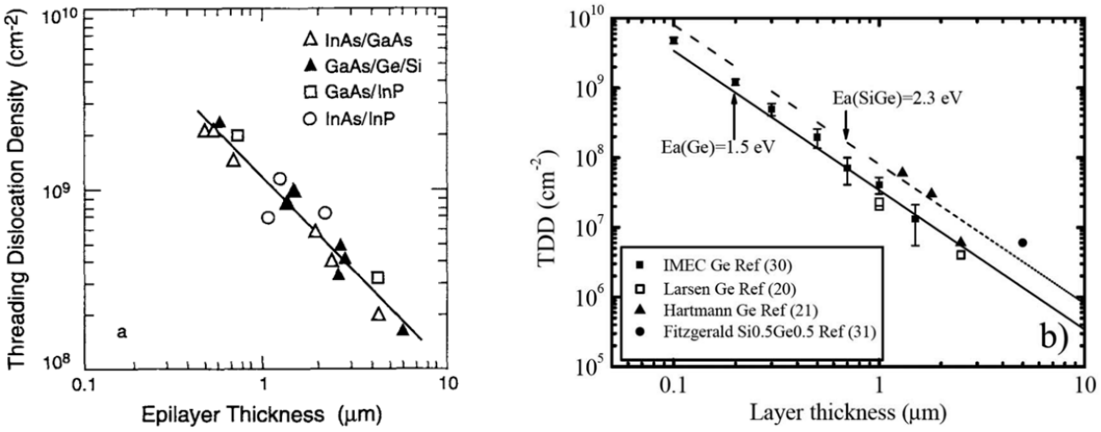
$$\rho = \frac{2}{R(2r_{\mathrm{FB}})} = \frac{f\cos\phi}{2\delta hb(1-\nu)}\ln\left(\frac{1}{4f}\right) \tag{4.19}$$

thus, also predicting a $\rho \propto h^{-1}$ dependence.

Alternatively, the balance of forces can be substituted with a balance of the activation energy for dislocation glide, E_{a}, with the dislocation interaction energy, W, associated with glide across a distance of $b/2$, the distance over which the dislocation experiences the energy barrier, E_{a}:

$$\Delta W = \frac{hGb^2\left(1-\nu\cos^2\beta_{\mathrm{TD}}\right)}{2\pi(1-\nu)\cos\phi}\ln\left(\frac{r}{r-b/2}\right) \tag{4.20}$$

While the glide activation energy can be a function of the stress condition,[39] constant values can still be used for general predictive purposes, e.g. 1.6 eV for Ge.[104] With the approximation $\ln\left(\frac{r}{r-b/2}\right) \approx \frac{b/2}{r}$ and equating $W = E_{\mathrm{a}}$, the equilibrium spacing between dislocations, r_{EB}, based on



(a) (b)

Figure 4.13: Effect of film thickness on TDD for (a) as grown films,[102] (b) films after annealing.[66]

an energy balance, can be estimated:[105]

$$r_{\mathrm{EB}} \approx \frac{hGb^3\left(1-\nu\cos^2\beta_{\mathrm{TD}}\right)}{4\pi(1-\nu)E_{\mathrm{a}}\cos\phi} \tag{4.21}$$

Even in the absence of cross-slip, both the force balance and energy balance models can allow for interaction between dislocations on intersecting glide planes. Two glissile dislocations of the 60° type that exist on orthogonal glide planes can share identical burgers vectors, allowing their threading segments to annihilate each other (due to their opposite sense of $\vec{\xi}$). Only the dislocations' now combined misfit segments near the substrate/film interface will remain. Annealing at temperatures greater than $0.6T_{\mathrm{m}}$ for Ge (and above $0.7T_{\mathrm{m}}$ for Si)[106] and the expected strong screw character of threading dislocations make cross-slip all the more likely to occur. With the inclusion of high temperature annealing, the expected final TDD should vary with the square of the dislocation spacing, providing a stronger $\rho \propto h^{-2}$ relationship between TDD and film thickness. Across multiple materials systems, the $\rho \propto h^{-1}$ relationship is observed for as grown films by molecular beam epitaxy (MBE) (see Figure 4.13a) while the trend $\rho \propto h^{-2}$ describes films after post-growth annealing at high temperatures for the Ge/Si materials system (see Figure 4.13b).

4.2.2 Selective growth models

Misfit dislocations relax the lattice mismatch at an interface by extending their line lengths. In order for a dislocation's threading component to leave the film surface, the end of the dislocation must reach the edge of the substrate. However, once the lattice mismatch is completely relaxed, dislocations will effectively stop gliding as doing so will not lead to a further decrease in elastic energy (unless already in close proximity, e.g. < 1 μm, to the film edge due to image forces). Assuming each dislocation extends by roughly the same amount, the maximum TDD to be filtered from a substrate of lateral dimension, w, can be estimated by setting the final elastic strain, ϵ, to zero:

$$\epsilon = f - \frac{\rho w b_{\text{eff}}}{4} \tag{4.22a}$$

$$\rho_{\text{max}} = \frac{4f}{w b_{\text{eff}}} \tag{4.22b}$$

where the factor of 4 results from two threading components per misfit and relaxation in both lateral directions, and b_{eff} is the edge component of the burgers vector in the interface plane, i.e. the component that relaxes the lattice misfit. For glissile 60° dislocations, $b_{\text{eff}} = |\vec{b}|/2$. Because $\rho_{\text{max}} \propto w^{-1}$, smaller substrates or equivalently, selectively grown films, can filter a greater level of threading dislocations. The benefit of selective growth was first demonstrated to decrease the final TDD in low misfit systems ($f < 1\%$), e.g. $In_{0.05}Ga_{0.95}As$ on GaAs ($f = 0.36\%$) and $Si_{0.85}Ge_{0.15}$ on Si ($f = 0.63\%$).[107] In these systems, homogeneous nucleation of dislocations is avoided,[40] allowing for a controlled introduction of dislocations into the epitaxial film. However, in the Ge-on-Si materials system, the large misfit allows for homogeneous nucleation, allowing for strain relaxation without significant dislocation glide. In order for Eq. 4.22b to be valid, the TDD must never exceed ρ_{max} during the film growth process. The initial TDD is likely greater than 10^{10} cm^{-2} in Ge-on-Si films due to homogeneous nucleation, and with $f = 4.2\%$ and $b_{\text{eff}} = 2.0$ Å, the above model is unlikely to be applicable.

Selective growth has also been evaluated from the perspective of attractive image forces due to free surfaces in a mesa. Within distances less than d from the mesa surface where F_{a} (Eq. 4.16 where $r = 2d$) is greater than F_{L} (Eq. 4.17), dislocations are expected to glide to the mesa edge. With independent consideration of individual image forces, mesas with widths $w < 16h/\ln(1/4f)$

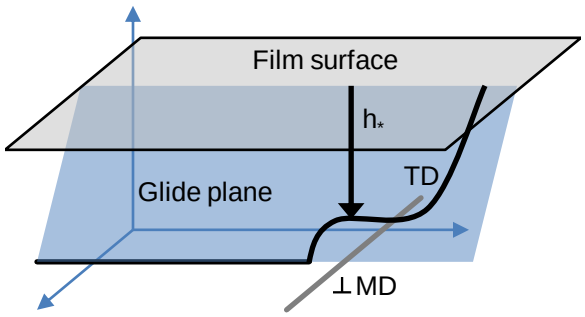


Figure 4.14: Schematic of an orthogonal misfit dislocation intersecting the glide plane of the threading component of another dislocation.

should be completely defect free.[108] However, image forces from opposite edges of the mesa will attract a given dislocation in opposing directions. Within a section of the mesa center, the sum of images forces will always reduce to a value below the magnitude of the line tension force, preventing complete dislocation removal.

4.2.3 Dislocation pinning

As threading dislocations glide due to an applied stress, their line length must extend across the glide distance. When Ge is deposited directly on Si, a dense network of misfit dislocations forms at the Ge/Si interface. This network induces a local stress field that only extends approximately the distance between adjacent dislocations (10 nm) since they cancel the lattice misfit strain throughout the thin film. As dislocations glide, the misfits corresponding to the gliding threading components are repelled from the actual Ge/Si interface and instead laterally extend above the Ge/Si interface by up to 50 nm, forming a separate confined dislocation network (CDN).[66] Unlike the misfit dislocations at the interface, the dislocations in the CDN are further spaced from each other, allowing their stress fields to extend further into the Ge film above.

A misfit segment from one dislocation can block a threading dislocation gliding in a direction orthogonal to the misfit segment line length (see Figure 4.14). For sufficiently thin films and low applied stresses, the misfit's stress field, σ_*, can cancel the applied stress field σ_0 due to thermal cycling. To determine whether blocking will occur, the film thickness, glide plane, and the burgers vector of the orthogonal misfit and the gliding thread are used to calculate the minimum required applied strain field, ϵ_0, that will allow a threading dislocation to glide above the misfit dislocation

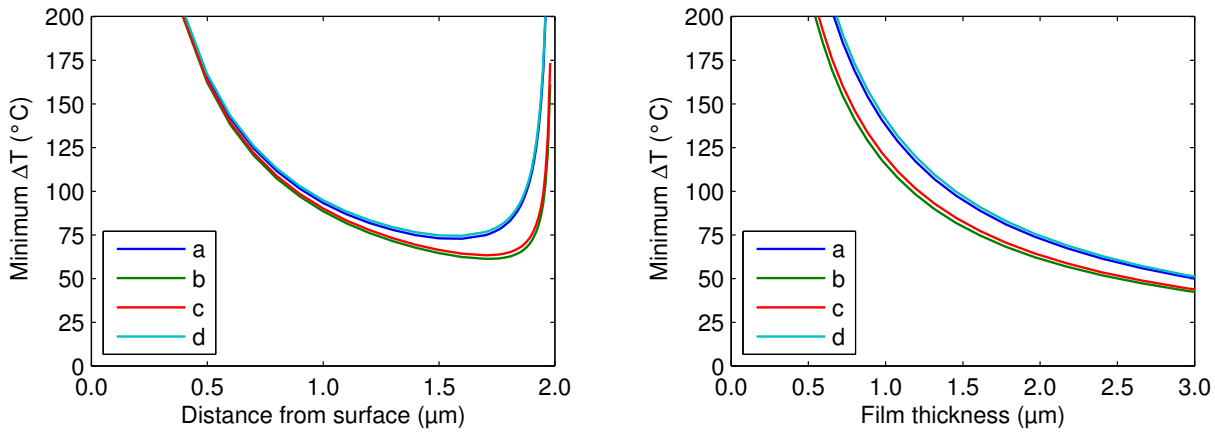


Figure 4.15: Minimum ΔT required to generate a thermal strain sufficient for threading dislocation glide over orthogonal misfit dislocations as a function of (a) extent of the gliding component from the surface into a 2 μm thick Ge film (h_*), (b) total film thickness.

through a distance h_* from the film top surface:[109]

$$\epsilon_0 = \epsilon_* + \frac{3b}{16\pi h_* (1+\nu)}\left[\frac{4-\nu}{3}\ln\left(\frac{8h_*}{b}\right) - \frac{\cos 2\alpha}{2} - \frac{1-2\nu}{4(1-\nu)}\right] \tag{4.23}$$

where ϵ_*, the reduction in strain due to the misfit, is a function of h_* dependent on the specific combination of burgers vectors of the two dislocations. At one value of h_*, a minimum value for ϵ_0 will exist where there is a balance between the image force, the applied stress, and the orthogonal dislocation stress on the segment of the threading dislocation within h_* of the film surface.

The temperature changes corresponding to minimum thermal strains of Ge-on-Si required for dislocation glide are plotted in Figure 4.15a as a function of h_* for a 2 μm thick Ge film. Curves are included for the four different combinations of burgers vectors (which are assumed to all be of the glissile type) for the threading component and orthogonal misfit segment as defined by Freund.[109] For combinations a and d, the minimum thermal strain corresponds to a temperature of 75°C. With the corresponding thermal strain applied, only the section of the threading component within 1.5 μm of the film surface will glide above the orthogonal misfit line. If however, the applied strain corresponds to less than 60°C, dislocation pairs of all combinations are expected to be blocked in the 2 μm thick Ge film. The minimum temperature change required to avoid dislocation pinning by an isolated orthogonal misfit line as a function of total film thickness is shown in Figure 4.15b.

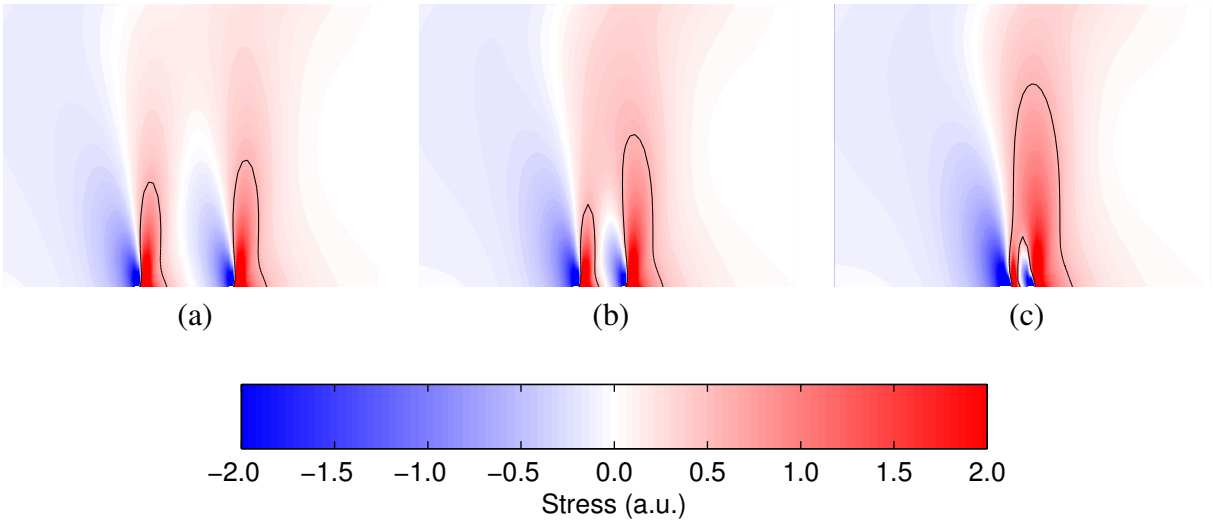


Figure 4.16: Resolved shear stress fields on a threading dislocation from two adjacent misfit segments of identical line directions and burgers vectors (type a) separated from each other by a distance, d, (a) $d/h = 1$, (b) $d/h = 0.5$, (c) $d/h = 0.25$. The two misfit segments extend through the page at the bottom of each contour plot, i.e. the substrate/film interface. The top boundary of each contour plot is the thin film surface. Positive stress values correspond to elastic energy barriers for glide.

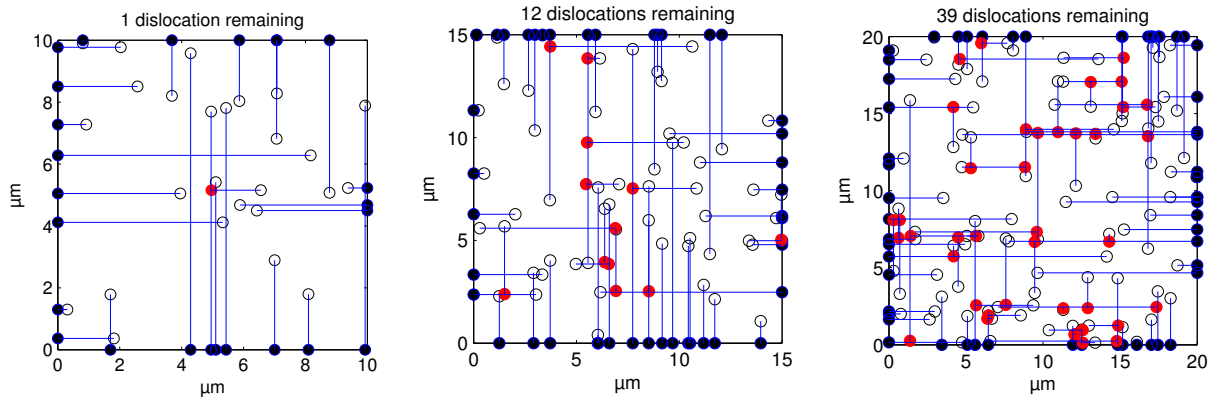
In order to maintain dislocation glide during annealing, the thermal strain must remain above the required minimum. As the structure plastically relaxes by dislocation glide, the effective temperature change is reduced. Once it drops below the value prescribed in Figure 4.15b, dislocation pinning will become operative until the temperature is changed again. Thus, the effective ΔT for dislocation glide is equal to the minimum ΔT for the particular film thickness subtracted from the applied ΔT. It should be noted that the minimum ΔT calculated is only an approximation, even for isolated misfit segments. If multiple misfit segments are in close proximity of distance d from each other, e.g. $d < h$, their stress fields begin to overlap as shown in Figure 4.16. One common stress contour level is shown in each figure, the vertical maximum of which increases as the misfits approach each other, $d/h \rightarrow 0$. A larger minimum applied stress is required for a gliding thread to pass over the two misfit dislocations as they are brought closer together. Thus, even when the applied stress is above the minimum required ΔT for an isolated lateral dislocation segment, there exists a non-zero probability that dislocation blocking will still occur due to lateral dislocation segments in close proximity to each other. For dislocation densities of 10^8 cm^{-2}, the average threading dislocation spacing is 1 μm, a value on the order of typical film thicknesses.

4.2.4 Dislocation glide in mesas

Monte Carlo simulations conducted in Matlab are used to evaluate the effect of dislocation pinning on dislocation glide in patterned Ge-on-Si films. The simulation first assumes that the film has reduced its TDD after a single high-temperature anneal to a level based solely on its thickness. An initial TDD value is determined by the inverse square of r_{EB} from Eq. 4.21, which accurately predicts Ge-on-Si experimental data as shown in Figure 4.13b. To place dislocations randomly throughout the simulation area of a single mesa, dislocation are first placed in a 100 μm square region according to the prescribed TDD, and the dislocations within a randomly selected region with the dimensions of the mesa to be analyzed are kept. This method allows for a random number of initial threading dislocations, centered around an average that need not be an integer number of threading dislocations per mesa (this point is particularly important for mesas of small widths). For each threading dislocation, an initial ⟨110⟩ glide direction is assigned which can only change in sign (no cross-slip is allowed). Dislocation reactions are also ignored.

After placing the dislocations, they are allowed to move in ⟨110⟩ directions, extending their horizontal segments between the threads' initial and current positions. When a threading dislocation arrives at a mesa edge, the threading dislocation is permanently removed from the simulation. If a threading dislocation approaches an orthogonal misfit segment from another dislocation, an adjustable probability determines whether the threading dislocation becomes blocked. This probability serves as a lumped parameter to account for the variability of the distance between adjacent misfit segments, their burgers vectors' combination, and the remaining applied thermal stress. Alternatively, the result of the dislocation interaction can be determined by assigning burgers vectors to individual dislocations in the simulation, calculating the overall stress field for that specific dislocation configuration as well as taking into consideration the overlap of stress fields from other misfit dislocations within a certain distance. In the case of the simple lumped parameter approach, the probability should scale monotonically with the initial TDD as a higher TDD will reduce the average spacing between extended misfit segments.

If the orthogonal misfit blocks the thread, the thread remains immobile for the duration of the current thermal half-cycle. Remaining unpinned dislocations are allowed to glide until they either become pinned or glide to the mesa edge. Once all remaining dislocations become immobile, the

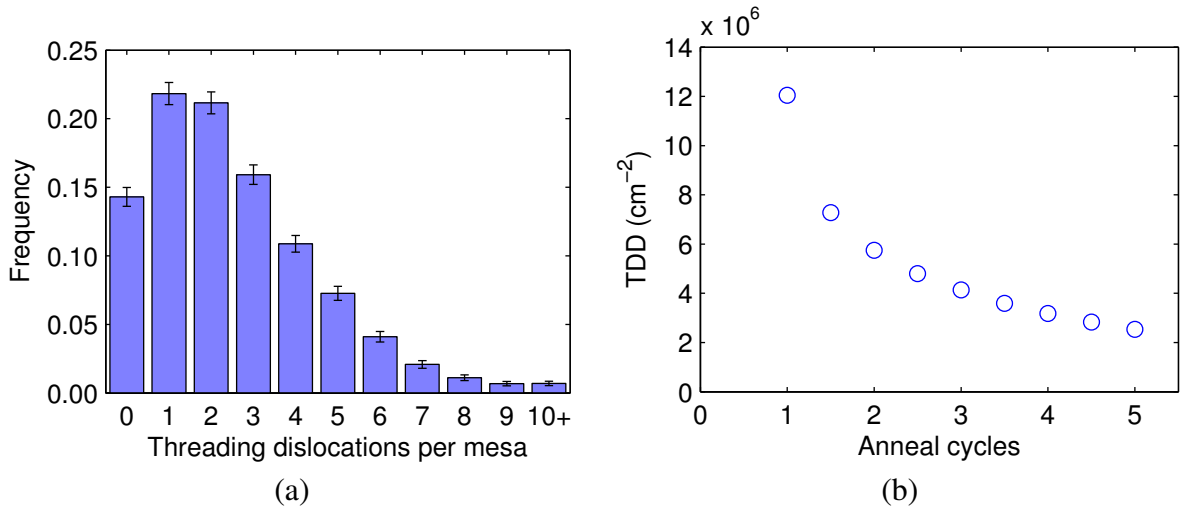


(a) (b) (c)

Figure 4.17: Results of dislocation pinning model for (a) 10 μm, (b) 15 μm, (c) 20 μm mesas. Unfilled circles are initial locations of TDs, blue circles are TDs that reached the mesa edges, and red circles are threading dislocations that remained within the mesa after five annealing cycles.

thermal stress is reversed, allowing initially pinned dislocations to glide in the opposite direction. The process is repeated several times until the simulation is terminated. As representative examples of the model's results, the final threading dislocation locations after five annealing cycles for mesas of widths 10, 15, and 20 μm, with an initial TDD of 2.5×10^7 cm^{-2} and a pinning probability of 2/3, is shown in Figure 4.17. The final TDD values for these particular mesas are 1×10^6 cm^{-2}, 5.3×10^6 cm^{-2}, and 9.8×10^6 cm^{-2}, respectively, roughly corresponding with trends observed experimentally.[89] As the mesa width is increased, both the mean required distance for glide to a mesa edge and the linear density of orthogonal misfits increase, favoring the probability of dislocation pinning.

Because of the inherently random nature of dislocation configurations, the simulation is repeated for the same mesa structure many times and averaged. The TDD distribution for 10 μm mesa squares after five annealing cycles can be seen in Figure 4.18a while the average reduction per half-cycle can be seen in Figure 4.18b. Instead of approaching a constant value, the TDD continuously decreases exponentially with annealing cycles. This result is due to the random determination of the outcome of a potential pinning event between the same two dislocations each time they interact. In a more accurate simulation, the determination of dislocation pinning based upon the remaining applied thermal stress, burgers vectors of the gliding dislocation and orthogonal misfits, and the relative positions of orthogonal misfits in close proximity should result in the approach of the TDD to a final finite average value.



(a) (b)

Figure 4.18: (a) TDD distribution in 10 μm mesa squares, (b) Average TDD in 10 μm mesa squares as a function of annealing cycles (a full cycle is $T_1 \rightarrow T_2 \rightarrow T_1$).

It should also be noted that a finite applied thermal strain will relax after sufficient dislocation glide. With a density of gliding dislocations ρ_{TD}, the average glide distance per dislocation to completely relieve the thermal strain $\epsilon^{\mathrm{t}} = \Delta\alpha\Delta T$ is:

$$l = \frac{2\Delta\alpha\Delta T}{b_{\mathrm{eff}}\rho_{\mathrm{TD}}} \tag{4.24}$$

where the factor of 2 is included for the two orthogonal sets of misfits dislocations. For $\rho = 2.5\times10^7$ cm^{-2}, $b_{\mathrm{eff}} = 2.0$ Å, and $\Delta T = 200$°C, $l \approx 20$ μm. In the case of even a 20 μm wide mesa with these initial conditions, this effect can be safely overlooked. As more dislocations are removed from the mesa, the required glide distance per dislocation for complete relaxation increases, further making this aspect less relevant. The primary result of ignoring this effect is the overestimation of the dislocation reduction in the first few thermal cycles.

The spatial distribution of pinned dislocations can be seen by the averaged local TDD in 10 μm and 20 μm mesa squares in Figure 4.19. For Figure 4.19a,b, the mesas are of uniform thickness and have an initial TDD of 2.5×10^7 cm^{-2}. In 10 μm mesas, the distribution of pinned threads is not a strong function of position. However, for larger mesas, e.g. 20 μm, the local TDD is noticeably greatest in the mesa center and diminishes as a function of radial distance. Dislocations further away from mesa sidewalls must glide past an increasing number of orthogonal misfit segments and are most likely to get pinned.

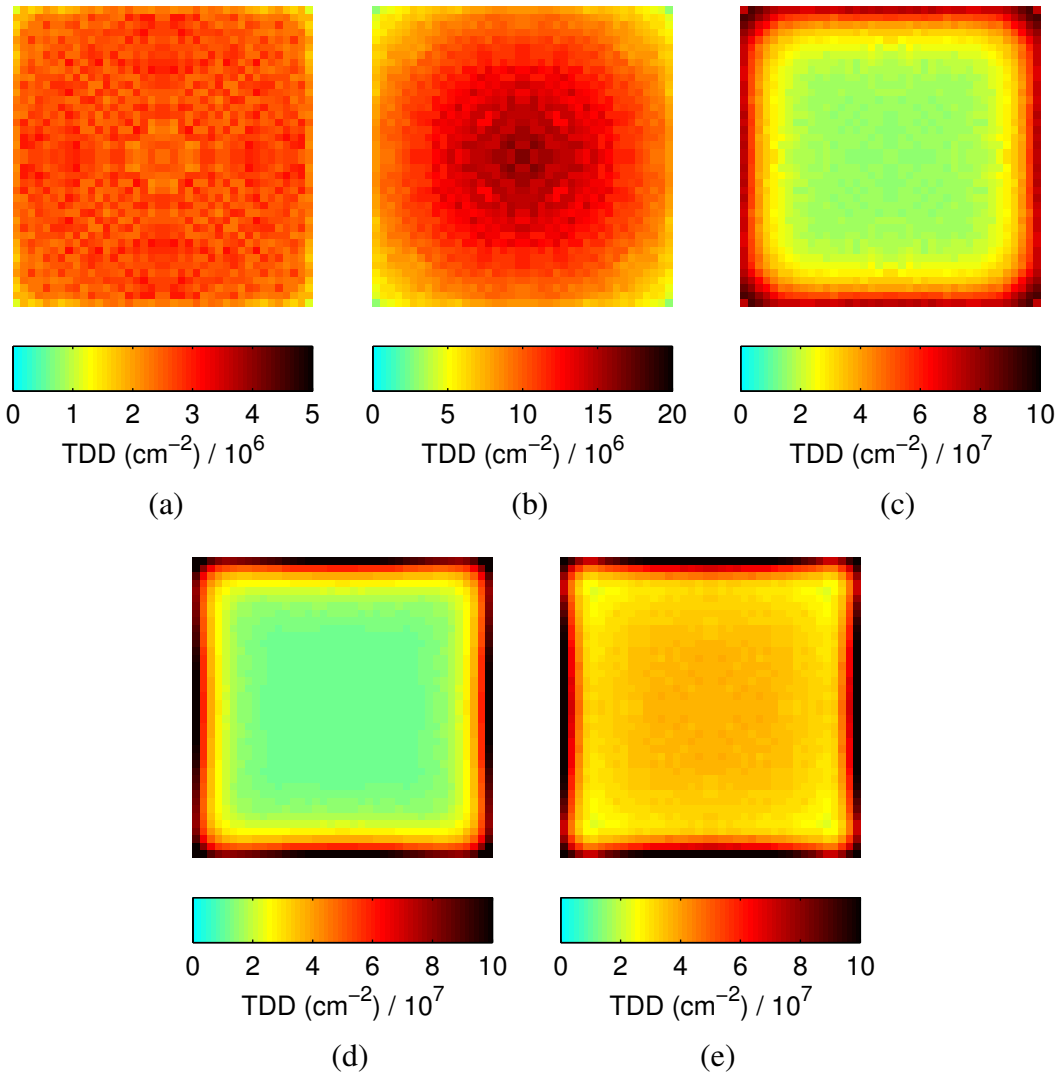


Figure 4.19: (a) Local TDD in mesa squares with widths (a,c) 10 μm, (b,d,e) 20 μm. Mesa film thicknesses are uniform in (a,b) while the mesas in (c,d,e) are relatively thicker in the center and thinner at its edges, representing a mesa that has developed facets during growth. Mesas in (c,d) are completely defined by {311} facets while the mesa in (e) is defined by both {311} facets and a (001) facet in a 10 μm square in the mesa center.

The 10 μm mesa represented in Figure 4.19c is five times as thick at its center as it is at its sidewalls, representing a mesa that developed {311} facets during film growth (for reference, see Figure 3.3b). The corresponding initial local TDD ranged from 1×10^7 cm^{-2} in the mesa center to 2.5×10^8 cm^{-2} at the mesa edge, with the local TDD value determined by the inverse square of the local thickness. For a square mesa completely defined by four symmetrical facets, the local thickness is a linear function of the closest distance to a mesa sidewall. Threading dislocations

were first placed in the entire mesa region at a density corresponding to the highest possible TDD level in the mesa, i.e. the TDD at the mesa sidewalls. For each resulting threading dislocation, a random number was generated and compared to the square of the ratio of the sidewall thickness to the local thickness, i.e. a ratio of the TDD values. If the random number was less than the ratio, the dislocation was kept. Otherwise, it was discarded from the initial conditions of the simulation.

Although the 10 μm mesa in Figure 4.19c is thicker in its center than the 10 μm mesa in Figure 4.19a, the excessive number of dislocations around the former mesa's perimeter effectively prevents the dislocations in the center of the mesa from gliding to the mesa sidewalls. The average local TDD in the mesa center, $\approx 2\times10^7$ cm^{-2}, is above the initial level of 1×10^7 cm^{-2}, while the average TDD in the faceted 10 μm mesa is 3.2×10^7 cm^{-2}. Similar results are observed for 20 μm wide mesas completely defined by {311} facets as shown in Figure 4.19d. For relevant growth times, a 20 μm wide mesa may instead be partially defined by {311} facets while retaining a (001) facet within a certain square area in the mesa center (see Figure 3.3c). When dislocation glide is simulated in this structure type with a 10 μm wide square (001) facet, a local maximum in the spatially local TDD is observed at the mesa center as shown in Figure 4.19e. Because dislocations in the mesa center must travel a longer lateral distance to exit the film, they are more likely to be pinned during glide. Dislocations approximately 5 μm away from the mesa perimeter can glide out of their initial positions but will become pinned either in the mesa center or near the mesa sidewalls. For mesas completely defined by {311} facets, the increased thickness in the mesa center that defines a lower initial TDD there hides the local maximum that was only observed in the case of Figure 4.19e.

To determine the overall trends of the simulation parameters on the final average TDD, mesas of different widths with corresponding initial uniform TDD levels and different pinning probabilities were simulated, the results of which are summarized in the log-log plot of Figure 4.20. In the regime of low mesa widths, the final TDD scales depends strongly on the mesa width in a power-law relationship. The exponent is dependent on the pinning probability, ranging from 3.7 for $P_{\text{pin}} = 0.5$ to 5.0 for $P_{\text{pin}} = 0.9$. The effect of changing the mesa width in the regime of larger mesas is reduced as dislocation pinning events become more common an interaction than dislocations arriving at the mesa sidewalls. The power-law regime does not end abruptly as it approaches the initial TDD level as the dislocations initially near the mesa perimeter can still reach the nearby mesa edge independent

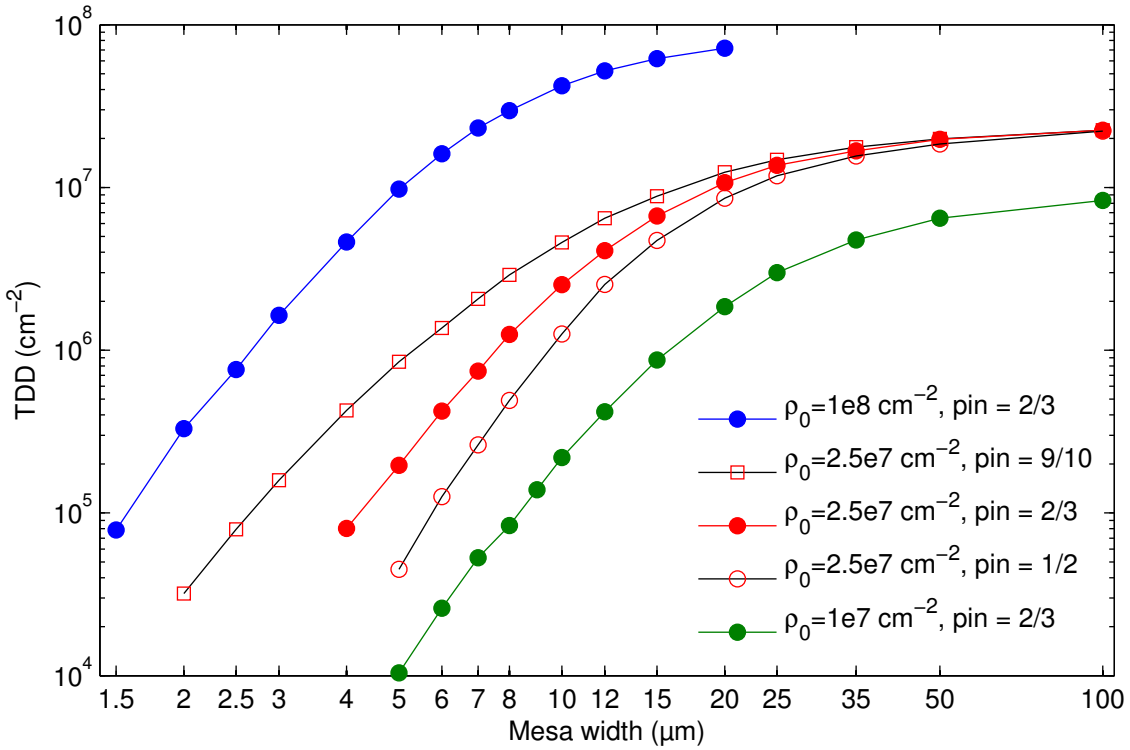


Figure 4.20: Final TDD after five annealing cycles for various combinations of mesa widths, initial TDDs, ρ_0, and pinning probabilities.

of the mesa's full dimensions. The initial TDD in the mesa, dependent on the mesa thickness, determines below which mesa width the power-law regime effectively begins. Thus, a trade-off can be made between increasing the mesa thickness and decreasing the mesa width to obtain the same final TDD level. For mesa dimensions that on average have less than 5 initial threading dislocations, almost all simulated mesas are dislocation free. Even with a pinning probability of 1, at least three threading dislocations are required to leave one thread pinned between the horizontal segments of the two dislocations that successfully left the Ge film. Because the starting number of dislocations is randomly generated, it is only the few mesas that initially contain a significantly above average number of threading dislocations that are responsible for the non-zero average TDD values reported for the smallest mesas. Figure 4.20 can be used as a guideline to determine the maximum mesa width allowed to achieve a certain TDD given the initial TDD (determined by film thickness). Future work and experimental verification can determine the effective pinning probability for a given initial TDD, remaining thermal strain, and the specific misfit dislocation configuration.

4.3 Summary

This chapter focused on the creation of a set of guidelines that can direct the design of a selectively-grown Ge-on-Si structure that will enable low TDD films. The results of structural mechanics modeling in Section 4.1 indicated that Ge mesa facets and thermal expansion compensation by SiO_2 sidewalls will create undesirable inverted shear regions. Identified potential solutions, in order of increasing predicted efficacy, included SiO_2 sidewalls with inverted angles and Ge lines thinly coated with SiO_2. The effects of mesa thickness and width on dislocation reduction and glide were studied in Section 4.2. Initial TDD reduction due to dislocation reactions has been observed to follow the scaling law with film thickness $\rho_{TD} \propto h^{-2}$. Monte Carlo simulations of dislocation glide and pinning has suggested the scaling law for mesa width in the small width regime: $\rho_{TD} \propto w^a$ where $a \sim 4$, a regime which expands increasing film thickness, corresponding to lower initial TDD levels. Faceting is identified as potential bottleneck to dislocation reduction by glide (due to reduced mesa thickness at edges) in addition to its adverse effects on the applied thermal stress and should therefore be avoided (or eliminated) during fabrication.

Chapter 5

Ge-on-Si Virtual Substrates

A Ge virtual substrate on Si should consist of a Ge surface that is planar and continuous. While patterned Ge mesas can likely achieve threading dislocation free material by following the design rules specified in Chapter 4, these structures do not meet the requirements for a virtual substrate and must therefore be modified. After lowering the defect density in patterned Ge films, Ge growth can be resumed, relying on epitaxial lateral overgrowth (ELO) over areas where the substrate surface is not directly exposed to allow adjacent mesas to coalesce. After growth, the coalesced film can be planarized by chemical mechanical polishing (CMP) using a low pH silica slurry, previously developed for Tungsten CMP.[110]

By following the process described above, blanket films with the TDD of the selectively grown material can be created, provided that ELO and coalescence-induced defects are minimized. The process is similar to ART (discussed earlier in Section 2.4) with the exception that increasing the spacing between coalescence fronts does not affect the required lateral overgrowth distance. Instead of relying on vertical geometric confinement of threading dislocations, this proposed process takes advantage of dislocation glide across laterally confined areas to reduce the TDD. In order for the proposed process to be successful, several key requirements must be satisfied:

1. Cyclic annealing produces resolved shear stresses of uniform sign throughout the Ge film
2. Dislocation pinning by other dislocations is minimized
3. Patterned Ge films completely coalesce during lateral overgrowth after defect removal
4. Defect generation after coalescence is avoided

While Chapter 4 provided solutions to the first two requirements, this chapter focuses on the third one and touches on the last. Section 5.1 investigates several aspects of overgrowth and coalescence and identifies design rules for a film patterning structure that optimizes the film growth process, primarily from the perspectives of control and throughput. Section 5.2 discusses complementary methods for identifying threading dislocations at the surface of Ge films and the possible implications

of experimental results obtained. A patterning solution for the third requirement is expected to simultaneously serve as a solution for the fourth based on preliminary observations.

5.1 Overgrowth and Coalescence

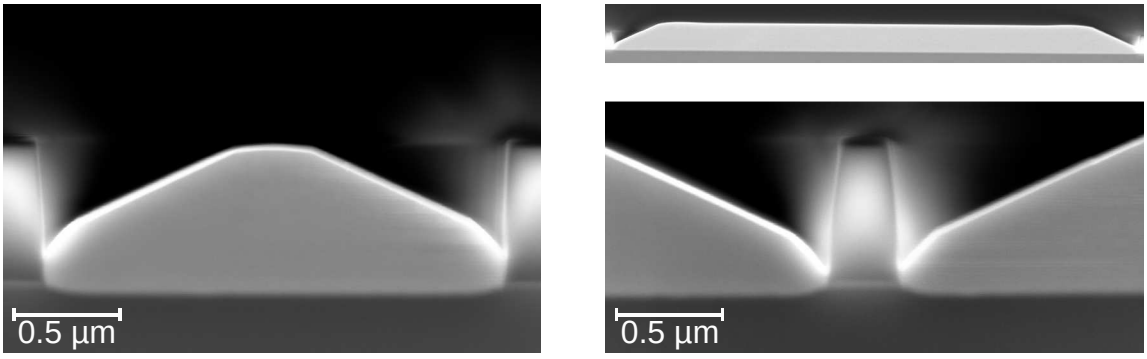
While the high surface energy of a Ge/SiO_2 interface enables selective Ge growth, it is also responsible for the reluctance of Ge to laterally overgrow across the SiO_2 sidewalls after trench fill is complete. Ge will initially create additional facets to avoid extension of the Ge/SiO_2 interface but will eventually initiate lateral overgrowth after sufficient growth time. After adjacent Ge lateral growth fronts impinge on each other, the two films will coalesce. Previous work on Ge lateral overgrowth and coalescence primarily involved ART experiments and is limited to long, parallel lines oriented in ⟨110⟩ directions. Two-dimensional patterns are more suitable for the proposed approach as they can reduce the required glide distances for all glissile dislocations without resorting to lines oriented in ⟨100⟩ directions. Different selective growth geometries are investigated to determine what general structure types enable complete film coalescence and how specific geometric parameters affect the rate of the overgrowth. Obtaining this information is crucial for the successful implementation of the proposed defect reduction process.

For the majority of overgrowth and coalescence experiments, Ge films were grown in a thin patterned SiO_2 layer so that trench fill would occur relatively quickly. The only constraint for the initial SiO_2 thickness was that the layer needed to be thick enough so that it would not be removed after the SiO_2 etching steps in the pre-epitaxial growth chemical cleaning process. Thin SiO_2 layers were grown by dry oxidation at 1000°C for 2 hours, obtaining an SiO_2 thickness of approximately 85 nm. Because the SiO_2 layer was less than 100 nm thick (unlike the SiO_2 trenches used in Chapter 3), the dry etch, wet etch process without sacrificial oxidation could be used while still maintaining good uniformity of the SiO_2 sidewall profile across the substrate. After photoresist patterning, SiO_2 dry etching removed approximately 65 nm, while the remaining 20 nm was removed in BOE after the photoresist was ashed. Due to overetching in BOE and the wet cleaning steps before Ge epitaxy, the final SiO_2 thickness was approximately 40 nm.

5.1.1 Trench fill and overgrowth

For mesa trenches aligned exactly to ⟨110⟩ directions and with sidewall angles greater than 83°, film faceting is primarily defined by {113} surfaces during trench fill. However, at the beginning of Ge film growth, {111} oriented planes temporarily appear at the trench sidewalls as seen in Figure 5.1a,b. The growth rate of the temporary {111} facets (before they are consumed by adjacent {311} facets) is greater for mesas of smaller widths. While the {111} facets are observed to vertically grow 200 nm for 2 μm wide mesas, these facets grow less than 100 nm in 20 μm wide mesas for the same growth time and under identical growth conditions. The primary difference between narrow and wide mesas is the growth time until the (001) plane in the mesa center disappears due to growth rate anisotropy. While the flux of available Ge from the vapor will be relatively uniform between facets, because Ge adatoms will preferentially deposit on fast growing planes such as the (001), there will exist a net flux of adatoms to the (001) plane from other facets.[111,112] Because growth is mass transport limited at high temperature, the growth rates of the {111} and {311} planes will be temporarily reduced until the (001) facet disappears.

The condition that a trench sidewall angle of 83° prevents the formation of {111} facets is derived from specific relative growth rates of the {111} and {311} planes: $v_{\{111\}} = 0.71 v_{\{311\}}$. This ratio was determined from the growth of mesa structures after the (001) plane disappears.[67] If while the (001) still exists, the growth rates of {111} and {311} planes are both reduced such that the relative growth rate ratio $v_{\{111\}}/v_{\{311\}}$ decreases by 10%, {111} facet formation will only be prevented for sidewall angles greater than 90°. Because the (001) facet was removed in the 2 μm



(a) (b)

Figure 5.1: SEM cross-sections of initial Ge film growth for mesa strips with widths (a) 2 μm and (b) 20 μm. The top inset in (b) is a lower magnification showing the entire 20 μm wide strip.

wide mesa (but not the 20 μm wide mesa - see the top inset of Figure 5.1b) before growth was terminated for these structures, the {111} facet growth rate had increased, leading to increased growth at the SiO_2 sidewall. The effects of mass transport between facets has also been observed in the enhanced growth rate of the (001) plane in selectively grown films. In Ge growths that were extended so that the (001) facet disappeared in 20 μm wide mesas just before growth was terminated, the film thickness at the mesa center, 4.6 μm, was greater than the thickness of blanket films, 3.2 μm. Because the mesa's surface was increasingly defined by relatively slow growing facets, the mesa's (001) facet experienced an ever increasing flux of adatoms, locally enhancing the deposition rate on the (001) surface. The average film thickness of the 20 μm mesa, 2.6 μm, is of course less than the blanket film's thickness due to the faceted structure.

Because mesa strips with smaller widths lose their (001) facet earlier in the growth process, these mesas also complete trench fill and begin overgrowth sooner. The time required for trench fill additionally depends on the orientation of the mesa strip due to the different facets that develop at the mesa edge. While the 10 μm wide mesa strip oriented exactly in ⟨110⟩ directions shown in Figure 3.3 only partially fills the trench, rotating the trench orientation by 15° enables complete trench fill under identical growth conditions as seen in Figure 5.2a. The effects of mesa width and trench orientation are also apparent in Ge films grown in very thin SiO_2 films. As a 4.5 μm wide mesa is oriented away from ⟨110⟩ directions by 7.5°, the maximum thickness in the mesa center increases from 1.36 μm to 1.74 μm and the angle of the Ge facet in contact with the SiO_2 layer also increases as seen in Figure 5.2b,c. Because the facets that form in slightly misoriented Ge mesas grow more quickly than {311} facets, the (001) disappears later in the growth process, allowing it to grow for a longer period of time. Trench filling has been observed to reach a local maximum for 7.5° among the offsets tested (0 to 45° with 7.5° resolution). Reducing the width of the misoriented mesa to 2 μm causes the facet angle at the SiO_2/Ge interface to invert after the same growth time (see Figure 5.2d), indicating the commencement of lateral overgrowth.

Faceting, trench fill, and lateral overgrowth were also observed on vicinal substrates, nominally oriented to (001) but miscut to (111) by 6°. When the Ge virtual substrate is to be used for III-V integration, Ge must be grown on off-cut Si substrates to prevent the formation of anti-phase domains in the compound semiconductor films. Only when the surface of a film is off-cut by several degrees will double atomic steps appear in the surface reconstruction at typical growth

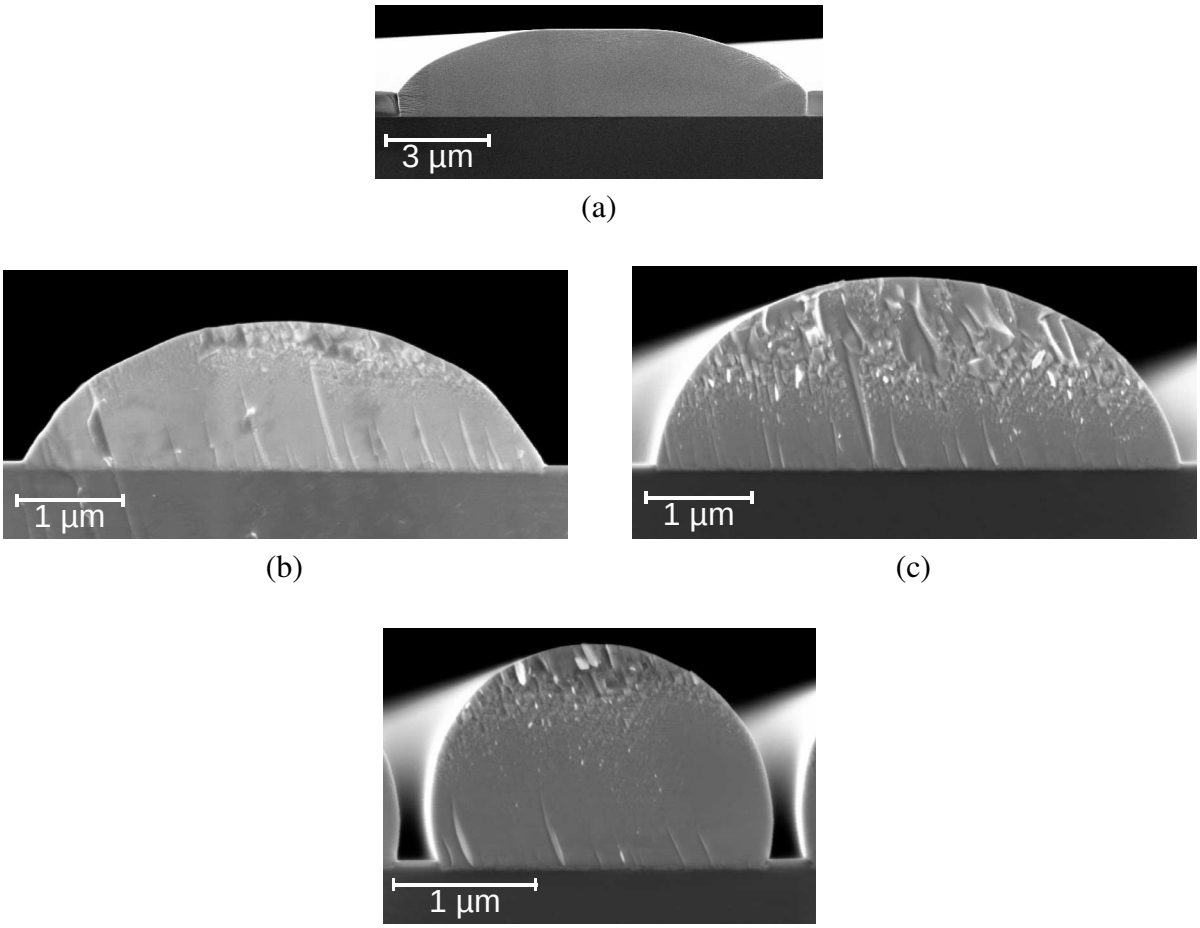


Figure 5.2: SEM cross-sections of mesa strips with widths of (a) 10 µm, (b,c) 4.5 µm, (d) 2 µm. SiO_2 lines are oriented relative to ⟨110⟩ directions by (b) 0°, (c,d) 7.5°, or (a) 15°. SiO_2 line thicknesses are either (a) 750 nm or (b,c,d) 40 nm. All cross-sections are taken along ⟨110⟩ directions.

temperatures.[113] The mesa strips are patterned such that the reference ⟨110⟩ direction for their longitudinal dimension is aligned to the rotation axis used to apply the miscut. Thus, the rotation of the wafer's crystal orientation is in the plane of the mesa cross-sections shown in Figure 5.3, specifically, 6° counter-clockwise. Two {311} facets can be observed in the 4.7 µm wide mesas oriented exactly to ⟨110⟩, at angles of 31° and 19° away from the substrate surface (corresponding to 25°± 6°). The asymmetry resembles similar faceting observed for Si selectively grown on vicinal substrates.[114] As demonstrated by comparison of the 2 µm wide mesa oriented by 7.5° to the other structures in Figure 5.3, the effects of smaller mesa widths and slight in-plane trench rotation, e.g. 7.5°, are observed to reduce the time before overgrowth on vicinal substrates as previously described for films grown on on-axis substrates. Lateral overgrowth is asymmetrical due the different facets

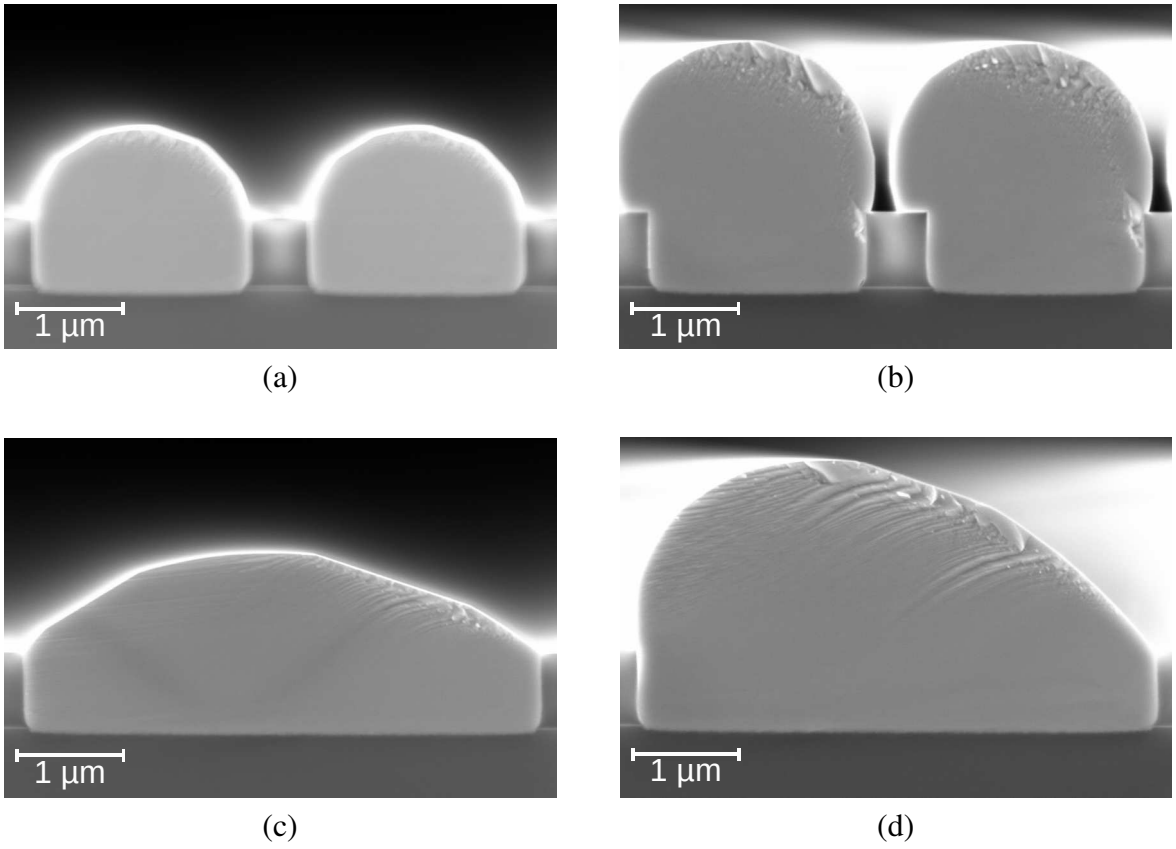
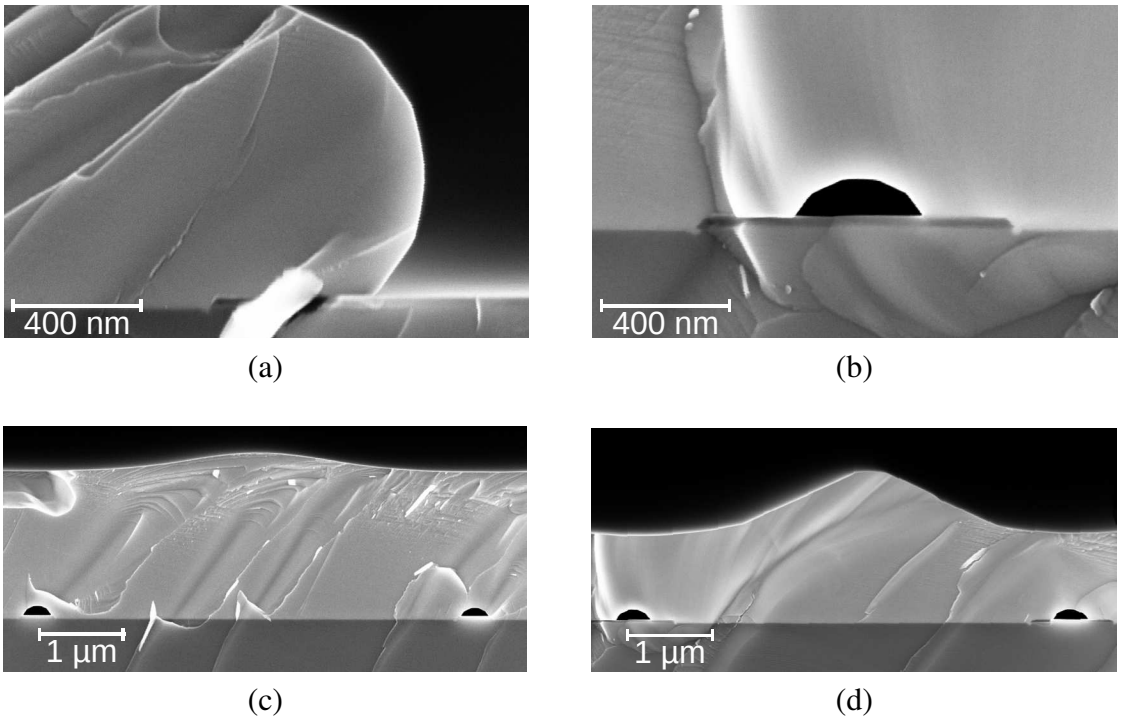


Figure 5.3: SEM cross-sections of mesa strips with widths of (a,b) 2 µm and (c,d) 4.7 µm. SiO_2 lines are oriented relative to ⟨110⟩ directions by either (a,c) 0° or (b,d) 7.5°. All cross-sections are taken along ⟨110⟩ substrate surface projections.

that form on each edge of the mesa strip. Only lateral growth from one side of the mesa is necessary to later achieve coalescence.

Once the Ge film initiates lateral overgrowth, additional facets form between Ge and the top surface of the SiO_2 lines in an attempt to impede expansion of the Ge/SiO_2 interface. In Figure 5.4a, {113} and {111} facets are observed on the top of the laterally overgrown film while an additional inverted {111} facet forms at the top surface of the SiO_2. Two additional facets exist between the two {111} facets and are likely of {119} orientation, based on faceting results previously report for selectively grown Si.[115] Growth fronts from adjacent mesa rectangles eventually converge, leaving a void above the SiO_2 due to the shape of the overgrown films at the moment of coalescence (see Figure 5.4b), as similarly reported in literature.[116,117] Limited film growth after the void has formed results in additional {113} facet formation to define the void's final surface. After initial film coalescence, further growth of the newly formed concave Ge surface above the void is instead



(a) (b)

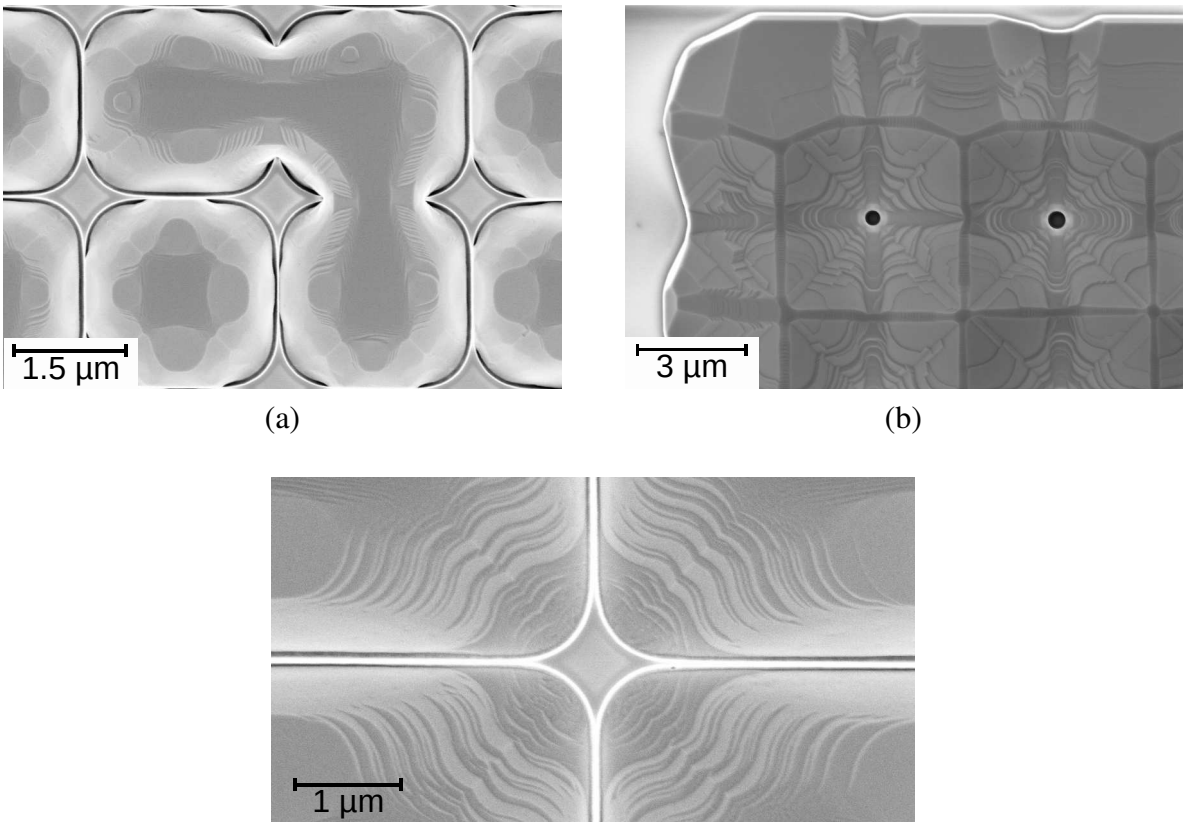
(c) (d)

Figure 5.4: SEM cross-sections of (a) faceting during lateral overgrowth, (b) remaining void after coalescence, (c,d) coalesced and partially planarized 5 μm spaced mesa strips with SiO_2 line widths of (c) 0.5 μm and (d) 1.0 μm.

determined by fast growing facets. Figure 5.4c,d compare the film profiles for Ge grown in 5 μm spaced mesa strips separated by 0.5 μm and 1.0 μm wide thin SiO_2 lines. Because coalescence occurs earlier in the growth process for mesa strips separated by narrower SiO_2 lines, the fast growing planes in these structures experience more growth time, allowing them to further planarize the structure. The peaks observed in both structures correspond to the location of peak of the original mesa strip before coalescence.

5.1.2 Grid structures

Similar to mesa strips, overgrowth and coalescence of adjacent mesa squares will readily occur between their edges as seen in Figure 5.5a. However, even after extension of the high temperature Ge growth time to twice that required for complete sidewall coalescence, overgrowth in the vicinity of mesa corners remains incomplete. In Figure 5.5b, black circles can be seen at the locations where the SiO_2 grid lines intersect, indicating a lack of Ge growth. The severe delay in overgrowth from mesa corners is due to unavoidable rounded corners in the SiO_2 grid that result from the



(a) (b)

(c)

Figure 5.5: SEM top views of (a) mesa squares partially coalescing at their sides (b) but not at their corners due to (c) slow growing facets that occur at rounded convex corners.

resolution limitations of i-stepper lithography. Complex faceted regions in the Ge mesas as shown in Figure 5.5c are created due to the rounded corners, exposing many facets with significantly reduced vertical growth rates.

To guarantee complete film coalescence, mesa squares can be replaced by long mesa strips so that corners in the Ge film only exist at the edges of the mesa array. While such a structure can completely coalesce, the ability to drive out threading dislocations is dependent on trench orientation relative to the substrate. If mesas are aligned to ⟨110⟩ directions, only dislocations in glide planes that span the transverse dimension of the mesa will be able to successfully glide to the mesa edge. Therefore, in order to minimize the glide distance required for all glissile dislocations in the mesa, the ideal trench orientation is 45° away from ⟨110⟩, i.e. in-plane ⟨100⟩ directions. As previously discussed, the rate of lateral overgrowth is affected by facet formation at mesa edges which in turn is affected by orientation of the mesa relative to the substrate. Similar to the trend

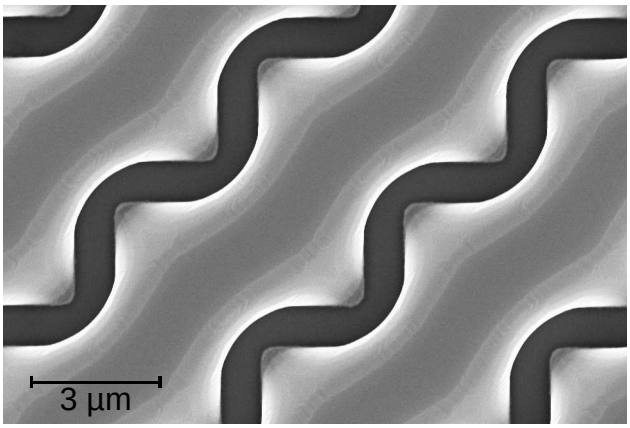
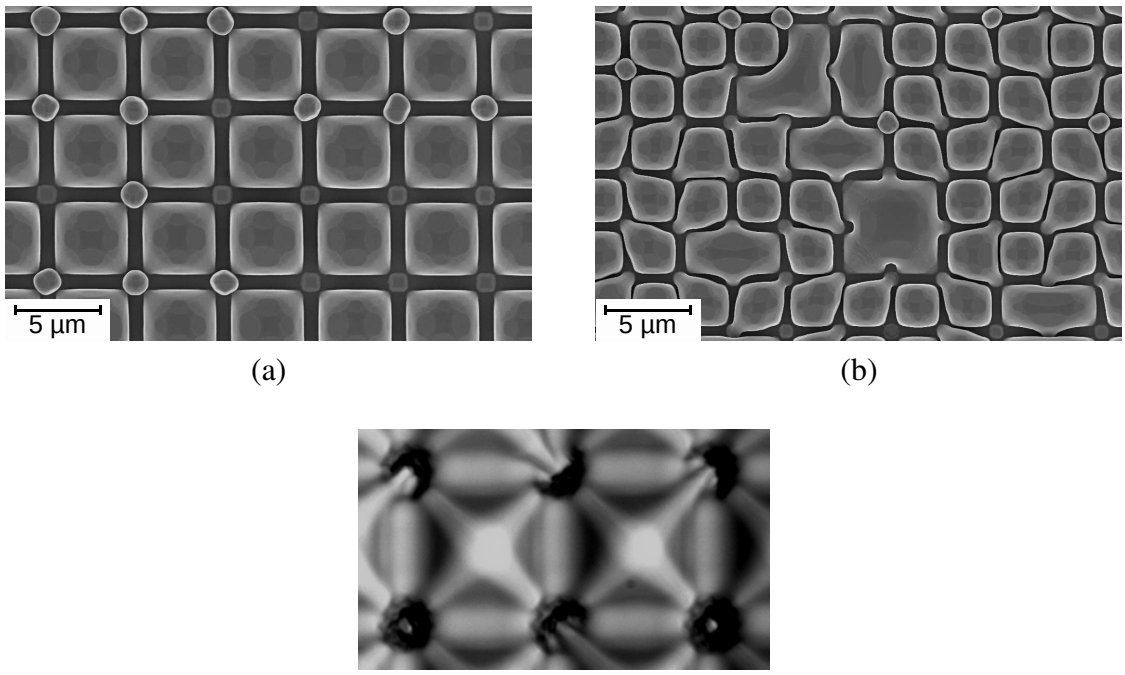


Figure 5.6: SEM top view of Ge grown in a mesa strip with a staircase edge.

observed for trench filling, the facets that form for different orientations lead to a minimum required time for lateral overgrowth for offsets between 0° and 15° away from ⟨110⟩ directions. Above 30°, trench fill and overgrowth become difficult and at 45°, coalescence has yet to observed after growth times corresponding to 7 µm of blanket film deposition. In an attempt to aid in the ability to laterally overgrow, the straight edge trenches aligned to ⟨100⟩ lateral directions can be replaced with a staircase structure composed of edges alternating in orthogonal ⟨110⟩ directions as shown in Figure 5.6. While Ge will readily overgrow the SiO_2 lines at the concave corners, the convex corners (equivalent to those in mesas) prevent further overgrowth. Thus, it is difficult to define a mesa structure that simultaneously allows for sufficient threading dislocation reduction during anneal and complete film coalescence after lateral overgrowth.

One attempt to facilitate coalescence at mesa corners in aligned mesa arrays was the inclusion of holes in the SiO_2 grid where orthogonal lines intersect. These holes would allow small Ge crystals to grow in close vicinity to the adjacent mesa corners. Because the hole size was close to the resolution of the i-stepper exposure system used, significant variation was observed in the trench fill time and shape of the small Ge islands that grew (see Figure 5.7a). These "coalescing-assistance" crystals would coalesce with at most one or two adjacent mesa corners, and only in rare cases, three or all four, single examples of which appear in Figure 5.7b. Even after excessive 7 µm of equivalent vertical blanket film growth, these additional Ge crystals failed to assist in the coalescence process (as seen in Figure 5.7c).



(a) (b)

(c)

Figure 5.7: SEM top views of nonuniform Ge fill of "coalescence-assistance" crystals (a) prior to coalescence, (b) after partial coalescence with at most one neighboring larger mesa with two visible exceptions. (c) Optical top view of unsuccessful coalescence between 20 µm Ge mesa squares separated by 1 µm SiO_2 lines with 1 µm diameter holes in the SiO_2 grid.

Instead of defining mesas by orthogonal sets of intersecting SiO_2 grid lines, square mesas can also be arranged in staggered patterns as shown in Figure 5.8. In this arrangement, only two mesa corners exist at each intersection of SiO_2 lines. Overgrowth from the middle of the center mesa at each intersection will occur similarly to a mesa strip without initial restriction. After sufficient lateral overgrowth across the SiO_2 from the center mesa, the growth front will reach the horizontal border of the two other mesas. At this point in the ELO process, partial coalescence has occurred between all three mesas. However, unlike in the perfectly gridded array, in the staggered array, Ge will readily continue to overgrow from the center mesa growth front, covering the remaining surface of the SiO_2 film. Overgrowth from the Ge ELO front to the non-center mesas will occur, possibly creating voids at the mesas' corners after complete film coalescence. The lower, planar regions observed in the optical image and AFM scan of a completely coalesced staggered mesa array in Figure 5.9 correspond to fast-growing (001) planes that emerge from the final locations of coalescence. The centers of the planar regions are offset from the exact intersection of the SiO_2 lines, appearing closer to the mesa corners than the center mesa. The observed offset indicates

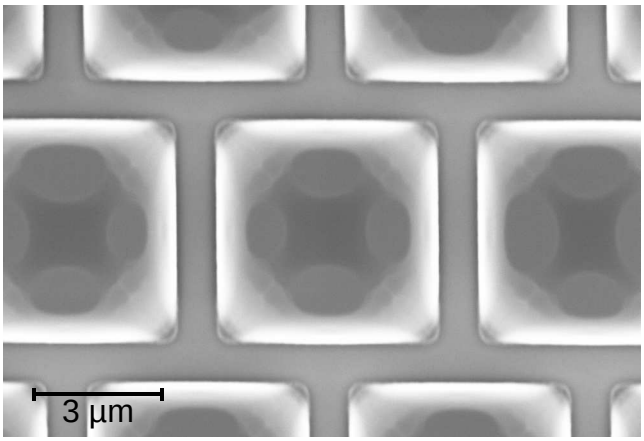


Figure 5.8: SEM top view of staggered Ge mesa array prior to overgrowth.

that significant lateral overgrowth from the center mesa occurs prior to complete coalescence as expected.

Although lateral overgrowth does not readily occur above mesa corners, the coalescence process across SiO_2 lines between mesas squares and strips is observed to initiate in close proximity to them. Despite the substantially reduced deposition rate at mesa corners, the flux of GeH_4 to the substrate area remains the same. With reduced competition from other surfaces, Ge mesa facets near mesa corners experience a local higher density of available adatoms, thus creating a slightly greater supersaturation, and a possibly greater growth rate. Rarely noticeable during the growth process, this slight advantage is likely responsible for the initiation of overgrowth in several structures.

Once the coalescence process initiates, the movement of the coalescence front across the SiO_2 line is exceedingly fast. In most experiments, overgrowth between mesa strips is typically a binary result; either no coalescence is observed or the region is completely coalesced. A select few growth

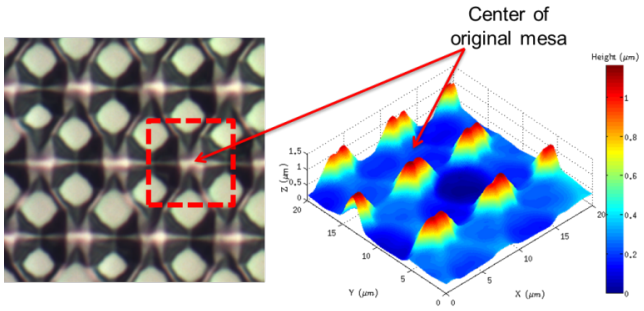
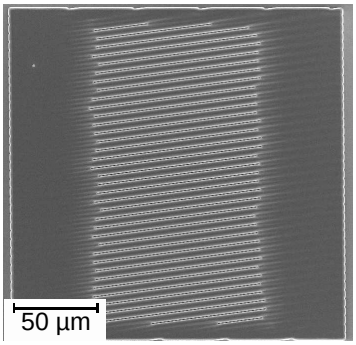
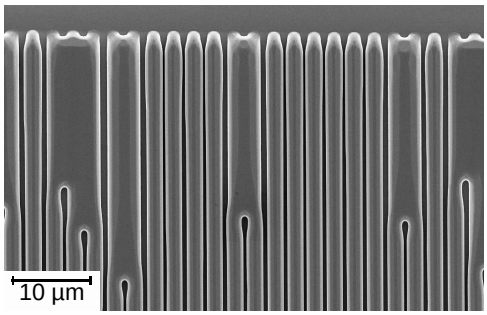


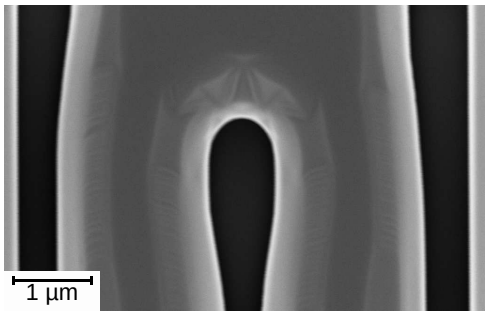
Figure 5.9: Optical image and AFM scan of a completely coalesced staggered mesa array.



(a)



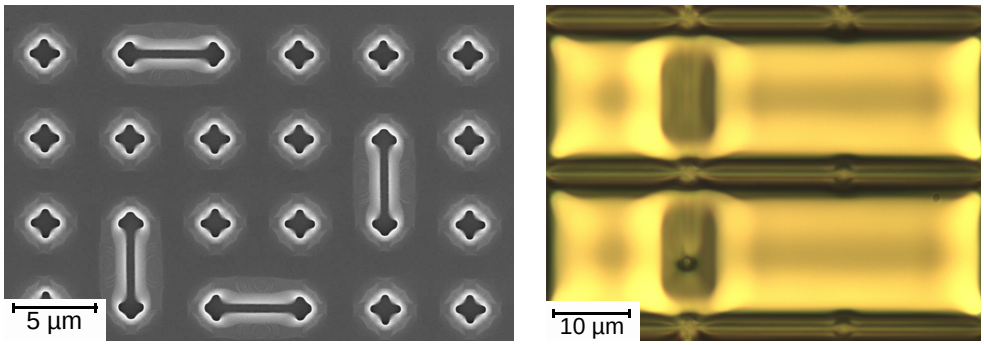
(b)



(c)

Figure 5.10: SEM top views of (a) coalescing 5 µm Ge mesa strips separated by 0.5 µm SiO_2 lines oriented 7.5° away from ⟨110⟩ directions, (b) coalescing 1.5 µm Ge mesa strips separated by 1.0 µm SiO_2 lines. (c) A single coalescence front from (b) is imaged at higher magnification.

experiments that provided non-binary results are now discussed. In Figure 5.10a, coalescence initiated from the edges of a 200 µm × 200 µm test structure of mesa strips oriented 7.5° away from ⟨110⟩ directions. The white lines indicate SiO_2 line segments which have not yet been covered by Ge, while the dark grey perimeter, approximately 50 µm wide, is Ge that has coalesced and self-planarized. In Figure 5.10b, coalescence between 1.5 µm mesa strips separated by 1.0 µm wide SiO_2 lines initiated shortly before Ge growth was terminated. While partial coalescence occurred over only a few SiO_2 lines, the extent of the coalescence distance varied by more than 10 µm. Within 2 µm of the location of the coalescence front, more clearly visible in Figure 5.10c, the overgrowth in the transverse dimension is actually reduced compared to locations further away where coalescence has not yet occurred. Additionally, overgrowth in the transverse direction above where the coalescence front has already passed is also relatively reduced. The various aspects of this curious structure can be explained as follows: Ge adatoms will preferably transfer from the slow growing facets typically associated with ELO to the concave growth front as it rapidly extends



(a) (b)

Figure 5.11: SEM top views of (a) 5 µm mesa squares separated by 0.5 µm SiO_2 lines. (b) Optical top view of 20 µm mesa squares immediately before final coalescence between adjacent mesa sides.

down the SiO_2 line. This mass transport on the surface results in a local reduction of the lateral overgrowth in the mesa strip's transverse dimension immediately in front of the concave growth front. Once the concave growth front moves past, causing the two adjacent mesa strips to coalesce, a new (001) plane emerges (this facet can be seen in Figure 5.10b,c as the darkest grey region of the Ge strips). The fast-growing (001) plane will consume most of the locally available adatoms, temporarily reducing the rate of local lateral overgrowth. Only after the coalesced mesa strips' (001) plane is overtaken by slower growing facets will more adatoms again significantly contribute to lateral overgrowth in the strips' transverse direction. The extent of the reemerged (001) plane reduces in its transverse width further away from the coalescence front since more growth time has passed after coalescence for slower growing facets to consume the (001) plane.

Sometimes in mesa arrays such as the one in Figure 5.11a, many line segments between mesas will be overgrown completely (excluding the corners) while among those that are not, the extent of partial lateral overgrowth does not appear significant. The full 0.5 µm width of the four uncovered SiO_2 line segments is completely uncovered while Ge has already planarized above the remaining segments. Once coalesce begins to occur between adjacent mesa square sides, the overgrowth typically proceeds too quickly to observe the process at an intermediate stage due to the relatively short width of the mesas. Only in the following rare occurrence has this process been observed immediately before completion. In Figure 5.11b, overgrowth had begun but did not finish above one of the SiO_2 line segments between two 20 µm wide mesas. The SiO_2 line still has an exposed point not covered by Ge. Because this point is roughly equidistant between the two mesa corners, it is likely that overgrowth initiated from both adjacent mesa corners at approximately the same time.

5.1.3 Isolated line structures

To eliminate the effects of slow growing facets due to mesa corners, the SiO_2 grid which defines mesa arrays is replaced with lines that do not connect to each other. While the Ge defined by SiO_2 grids was confined to individual, disconnected mesas prior to overgrowth, this alternative structure reverses the case as a continuous Ge film now surrounds individual SiO_2 lines. In these patterns, the isolated SiO_2 lines are oriented such that threading dislocations of all four in-plane $\langle 110 \rangle$ slip directions can glide to a Ge film edge and terminate independent of their initial position in the Ge film. A schematic of the general isolated line structure can be seen in Figure 5.12 with arrows indicating the all potential directions that the example threading dislocation ("TD") may glide to exit the film. A similar approach appears in the work of Hull *et al.*, involving pillars arranged in a staggered array but is limited to low misfit heterostructures due to the large glide distances required for threading dislocation removal.[118] By eliminating locations where orthogonal SiO_2 lines intersect, convex edge regions in the Ge film are prevented. Instead, the isolated lines create only concave and zero curvature perimeters in the continuous Ge film that surrounds the lines. Figure 5.13 provides two examples of completely coalesced Ge films grown around isolated SiO_2 lines on on-axis and vicinal substrates. In both samples, after Ge completed coalescence above the SiO_2 lines, (001) planes formed and began planarization of the local film area. These planarized areas have dark contrast, surrounded by transitional faceted regions that appear lightest in SEM top view. The coalesced films are thickest at the interconnected curved lines (also of dark contrast) that intersect each other between the isolated SiO_2 lines. These areas are furthest away from the SiO_2 lines and are the closest equivalent to the peaks of mesa squares grown in SiO_2 grids.

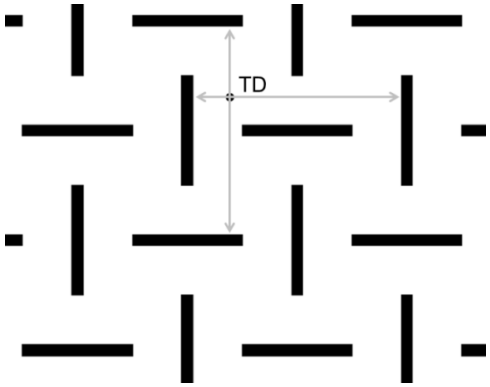
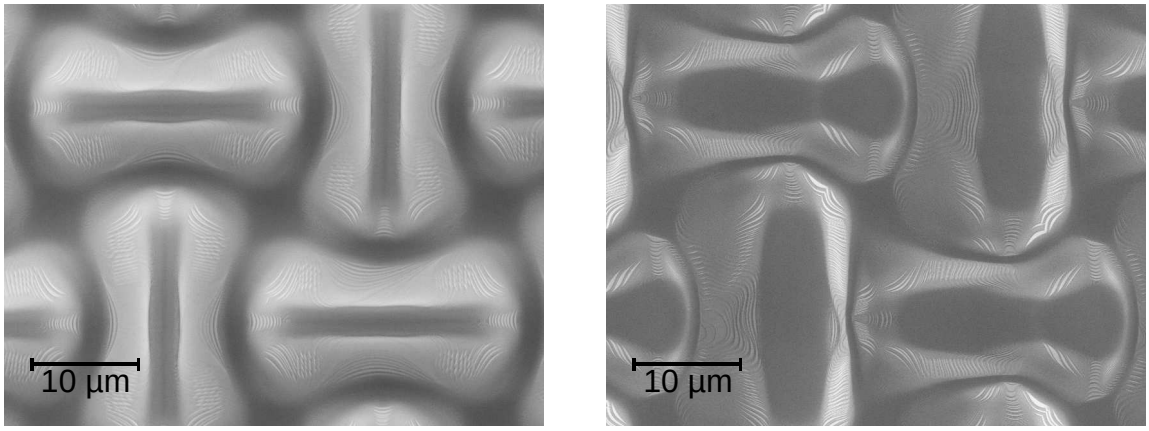


Figure 5.12: Schematic of the isolated line pattern design basis.

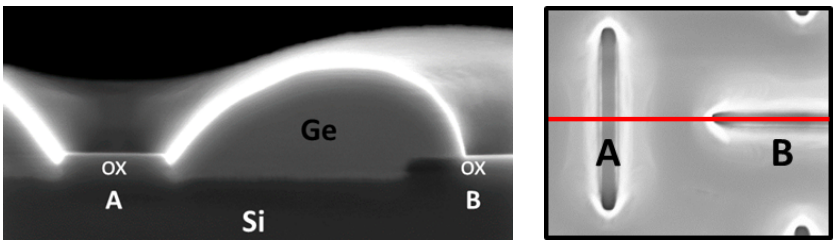


(a) (b)

Figure 5.13: SEM top view of completely coalesced Ge films grown in isolated SiO_2 line structures on (a) on-axis (001) Si and (b) vicinal (001) Si substrates, miscut by 6° to (111).

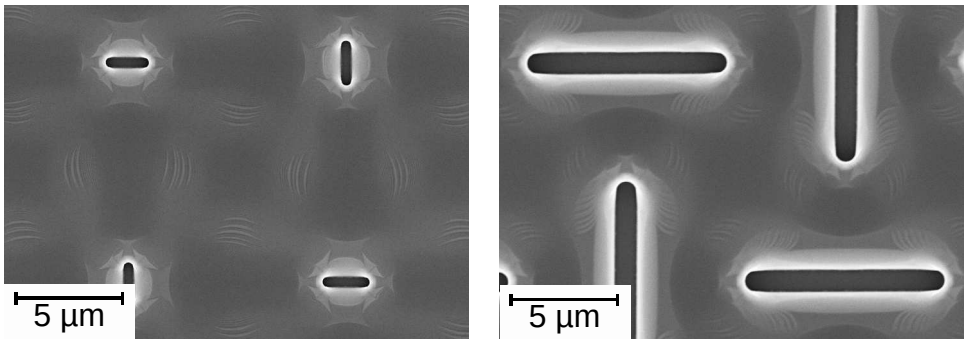
While crystal growth at convex perimeters is limited by the slowest growing facts that develop, growth at the concave corners have no restriction according to Wulff construction reasoning.[119] Ge overgrowth starting at the ends of SiO_2 lines is observed to proceed sooner than overgrowth along the length of lines with zero curvature. In the cross-section shown in Figure 5.14a, Ge has yet to overgrow the oxide at line "A" where the curvature is zero while at line "B," Ge has already begun lateral overgrowth. Unlike in overgrowth between mesa strips, no void is observed on top of the SiO_2 line. Figure 5.14b indicates the location of cross-section in (a) by a red horizontal line.

The extent of accelerated growth is found to be a strong function of SiO_2 line width. Figure 5.15 compares the extent of lateral overgrowth along 15 µm long SiO_2 lines with widths of 0.5 µm and 1.0 µm midway through the overgrowth process. Under the same growth conditions, Ge laterally overgrows the 0.5 µm wide lines by about 6 µm from both line edges while Ge only extends over the 1.0 µm wide lines by 2 µm. As the line is widened, the concave corners in the Ge film constitute



(a) (b)

Figure 5.14: (a) SEM cross-section of lateral overgrowth above isolated SiO_2 lines. (b) SEM top view for reference.



(a) (b)

Figure 5.15: SEM top views of lateral overgrowth across 15 µm long isolated SiO_2 lines of widths (a) 0.5 µm and (b) 1.0 µm.

a smaller fraction of the total edge, and for widths greater than 1.5 µm, accelerated overgrowth is absent. Although the exact facets that form in the overgrowth is unclear, the close proximity of these facets from both edges at the SiO_2 lines' ends is likely responsible for initial accelerated overgrowth. For the larger widths, the zero curvature edge geometrically separates the two concave corners in Ge sufficiently to prevent the accelerated effect.

As the Ge growth fronts approach each other from opposite ends of a SiO_2 line, the overgrowth becomes halted. Final coalescence across the center of the SiO_2 line in isolated line structures is observed to occur later during growth then for the coalescence between mesa strips with SiO_2 lines of identical width, i.e. structures with the same lateral overgrowth distance required for coalescence. In Figure 5.16, complete film coalescence is observed for mesa strips of all SiO_2 line widths up to at least 1 µm. Meanwhile, isolated SiO_2 lines of widths 0.75 µm or greater remain partially uncovered after the identical Ge growth process, appearing as dark lines in Figure 5.16b,c.

The stagnation of overgrowth late in the process on isolated SiO_2 lines can be explained by contrasting the accelerated overgrowth observed between Ge mesa strips and within Ge films with isolated SiO_2 lines. The initiation of overgrowth in mesa strips is thought to occur due to the local enhanced supply of Ge adatoms due to lack of film growth at convex mesa corners. For isolated line structures, initiation instead results from the enhanced growth rate directly due to the concave perimeter at the ends of SiO_2 lines. While initiation for mesa structures occurs later during the growth process and more randomly compared to films with isolated SiO_2 lines, the coalescence front moves much more quickly and does not appear to ever become halted. While void formation is observed over SiO_2 lines between mesa strips, voids have not been observed over isolated lines

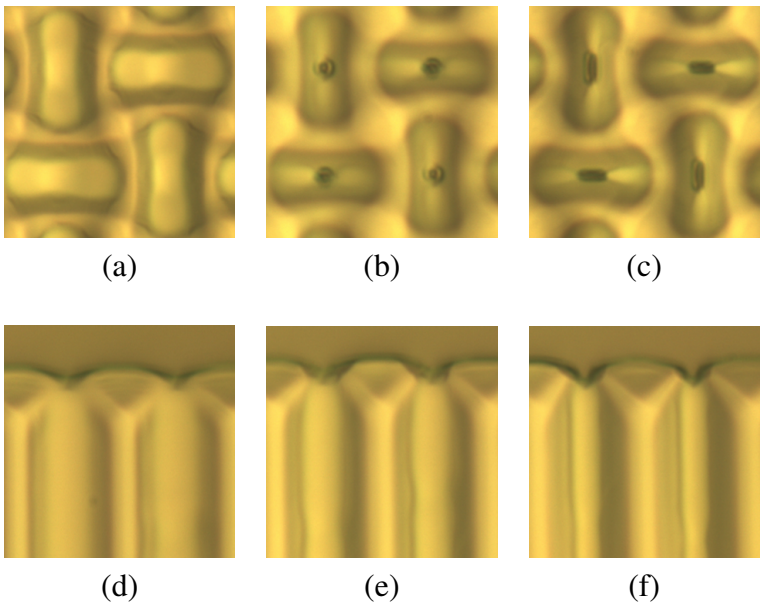
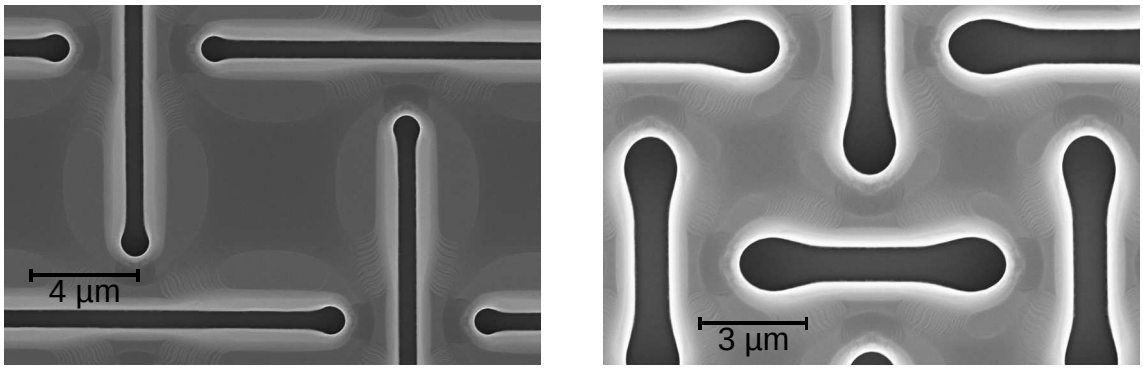


Figure 5.16: Optical top view images of (a-c) Ge films surrounding isolated SiO_2 lines and (d-f) Ge mesa strips after overgrowth across SiO_2 line widths of (a,d) 0.5 μm, (b,e) 0.75 μm, (c,f) 1.0 μm.

completely surrounded by Ge. Although void formation cannot be ruled out, assuming that the contrast to overgrowth over grid structures holds as the Ge grown on the SiO_2 labeled "B" in Figure 5.14a would suggest, the void may responsible for the difference observed in the overgrowth of concave growth fronts between mesa strips and over isolated SiO_2 lines. In an array of mesa strips, coalescence only begins once adjacent lateral overgrowth fronts impinge each other above the SiO_2. Once this crucial step occurs, continued coalescence does not require an increase of the Ge/SiO_2 interface as the void can extend further down the SiO_2 strip. However, coalescence across the isolated SiO_2 lines occurs precisely because of initial overgrowth on the SiO_2 surface, starting at the SiO_2 line ends. Because the coalescing front is in direct contact with SiO_2, further overgrowth directly leads to extension of the Ge/SiO_2 interfacial area, which the Ge adatoms have a energetic reason to avoid increasing. As the isolated SiO_2 lines are overgrown, the facets of the coalescing front evolves, eventually developing into a collection of facets which delay overgrowth instead of accelerating it.

While initial accelerated overgrowth over isolated SiO_2 lines is beneficial from the perspective of throughput, the overgrowth over lines of constant width is difficult to control. Because the initial overgrowth occurs very quickly, the Ge film overgrows across a significant length of the SiO_2 lines before Ge can completely fill the thin SiO_2 trench at zero curvature edges as seen in



(a) (b)

Figure 5.17: SEM top views of lateral overgrowth across isolated SiO_2 lines spaced (a) 10 μm and (b) 6 μm apart. The original perimeter of the SiO_2 lines was defined by (a) T-shaped ends 3 μm long and a constant line width of 0.75 μm, (b) a variable line width 1.5 μm at the ends and 0.8 μm at the center.

Figure 5.14. During post-growth annealing to drive glissile dislocations to the film sidewalls, if Ge completely overgrows a section of the SiO_2 sidewall, that section can no longer serve its purpose of providing an edge for threading dislocations to terminate. As a dislocation arrives at an SiO_2 sidewall, while one section of the threading dislocation can terminate, the component of the initial threading dislocation that extends above the SiO_2 line to the top surface of the Ge film will not automatically be eliminated.

To better control the overgrowth process over isolated SiO_2 lines, several patterning aspects were introduced. Keeping the center locations of each isolated line unchanged, the lengths of the SiO_2 lines were extended to allow for some overgrowth to occur without rendering the SiO_2 sidewalls ineffective at providing sinks for dislocations when working in concert with each other. Branched T-shape ends were also introduced as a means to extend the maximum possible length. However, because the accelerated overgrowth occurs so quickly, within the growth time required for 1.75 μm thick blanket Ge films, maximizing the SiO_2 line length extension was only effective for relatively large line spacings (see Figure 5.17a), structures that would require up to 20 μm glide distances for dislocations to create dislocation-free material.

While premature overgrowth at concave corners may be eliminated by increasing the SiO_2 mask thickness, process control is still possible for thin SiO_2 structures that require dislocation glide distances < 20 μm. By adjusting the line width as a function of distance from the line center, control of the accelerated overgrowth process is improved. By increasing the SiO_2 line width to at least

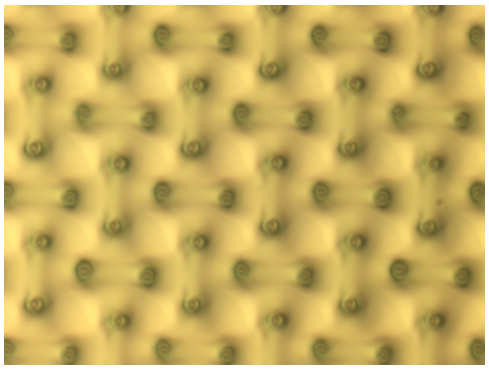


Figure 5.18: Optical top view image of partial overgrowth over isolated SiO_2 lines 2.5 μm wide at their ends and 0.75 μm at their centers, oriented 5° away from ⟨110⟩ directions.

1.5 μm at the trend line ends, accelerated overgrowth is sufficiently reduced, allowing growth of a thicker Ge film before attempting defect reduction by cyclic annealing. To minimize the growth time required for complete coalescence, the line width in the center can be reduced without affecting the initial overgrowth rate from the trench line ends. An example of Ge grown around SiO_2 lines of variable width is shown in Figure 5.17b. Even after 1.6 μm of equivalent blanket growth, the edges of the Ge film will still allow dislocations of all glide directions and initial positions to terminate. Only for extreme width variation is coalescence observed to occur at the center of the SiO_2 lines before overgrowth from the SiO_2 line ends. For lines with a width variation from 2.5 μm to 0.75 μm, two separate regions of each SiO_2 line are left uncovered at an intermediate stage of the overgrowth process, appearing as the dark circular regions above the SiO_2 in Figure 5.18. By moderate variation of the line width along its length so that the width is large at the ends and reduced at the center, the initial accelerated overgrowth can be avoided while simultaneously avoiding delayed coalescence at the end of the overgrowth process.

5.1.4 Substrate orientation

As the orientation of SiO_2 grids relative to ⟨110⟩ in-plane directions on (001) Si on-axis substrates effects the trench filling rate, a similar orientation dependence exists for the time before coalescence between mesas edges, minimizing for offsets near 7.5°. For Ge mesas grown on (001) Si wafers miscut 6° to (111), the maximum overgrowth and partial coalescence rate (excluding mesa corners) is observed for SiO_2 grids roughly parallel to the wafer flat while the familiar local

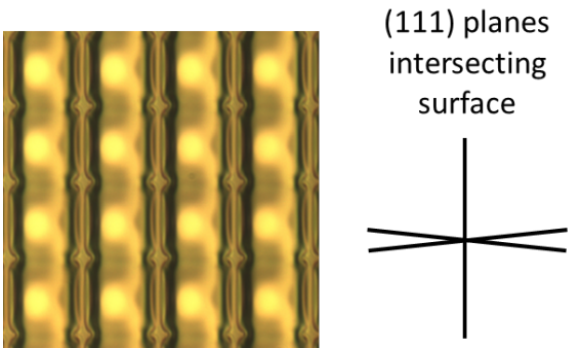
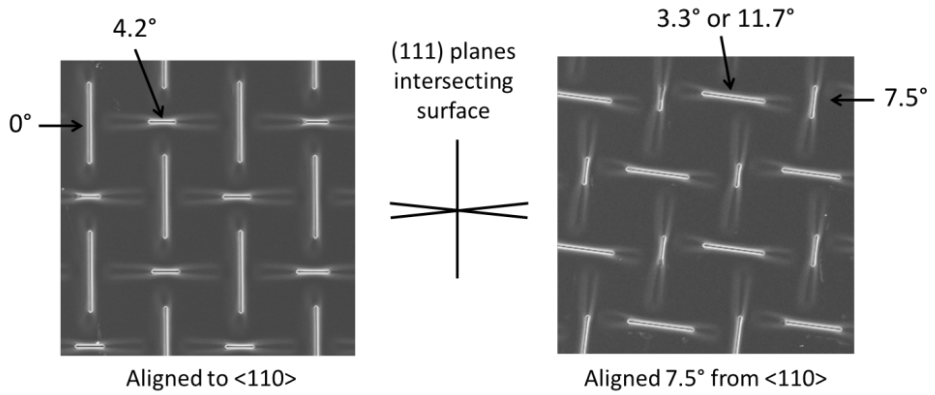


Figure 5.19: Optical top view image showing overgrowth in a mesa array on Si (100) offcut by 6° to (111).

maximum between 0 and 15° is observed in the orthogonal direction. In Figure 5.19, overgrowth has only occurred between vertically adjacent mesas. The narrow vertical lines in the image are exposed SiO_2 lines that Ge has not yet overgrown. The wafer flat is parallel to the horizontal direction. For mesas grown from on-axis wafers in the identical SiO_2 grid (mesa width = 10 μm, SiO_2 line width = 0.75 μm), overgrowth does not occur above either set of SiO_2 grid lines.

The vicinal wafers are sliced such that the wafer plane is rotated from the nominal (001) plane by rotating around the axis that passes through the wafer flat to the opposite edge of the wafer. The rotation of the substrate's crystal orientation does not alter the projection of either set of in-plane ⟨110⟩ directions. However, {111} planes nominally parallel to the wafer flat will no longer intersect the vicinal substrate's surface exactly at the horizontal ⟨110⟩ projection.[47] Instead, the intersection is rotated in-plane by ±4.2° away from the direction parallel to the wafer flat while the intersection of {111} planes perpendicular to the wafer flat remain exactly parallel to the orthogonal ⟨110⟩ directions (shown for reference in Figure 5.19). SiO_2 lines parallel to the wafer flat on vicinal substrates are therefore disoriented by 4.2°. No clear increase in the overgrowth rate occurs over lines parallel to the direction for which {111} intersections remains unchanged due to the substrate miscut. Therefore, the controlling factor for maximizing the rate of coalescence appears to be the offset between the longitudinal direction of the SiO_2 line to be overgrown and the nearest intersection direction of a {111} plane at the substrate surface.

Similar effects are observed in the overgrowth process around isolated SiO_2 lines. For on-axis substrates, maximum accelerated overgrowth occurs for isolated SiO_2 lines oriented 7.5° away from ⟨110⟩ directions. For Ge films grown between SiO_2 lines parallel to ⟨110⟩ projections on



(a) (b)

Figure 5.20: SEM top view images of overgrowth around isolated SiO_2 lines on (100) Si offcut by 6° to (111). SiO_2 lines are oriented relative to ⟨110⟩ projections on the surface by (a) 0° and (b) 7.5°. Angles indicate the offset between the SiO_2 line direction and closest {111} plane intersections on the substrate surface.

vicinal substrates, the extent of overgrowth is greater for lines oriented parallel to the wafer flat (see Figure 5.20a). In structures with both sets of orthogonal trench lines rotated by 7.5° as in Figure 5.20b, the comparative result reverses. While nominally vertically aligned SiO_2 lines remain exactly 7.5° away from {111} surface intersections, the horizontal SiO_2 lines are 3.3° and 11.7° away from the two sets of non-parallel, nominally horizontal directions. While it is not entirely clear which angle offset is relevant, the rate of overgrowth is likely determined by whichever angle corresponds to a slower overgrowth rate.

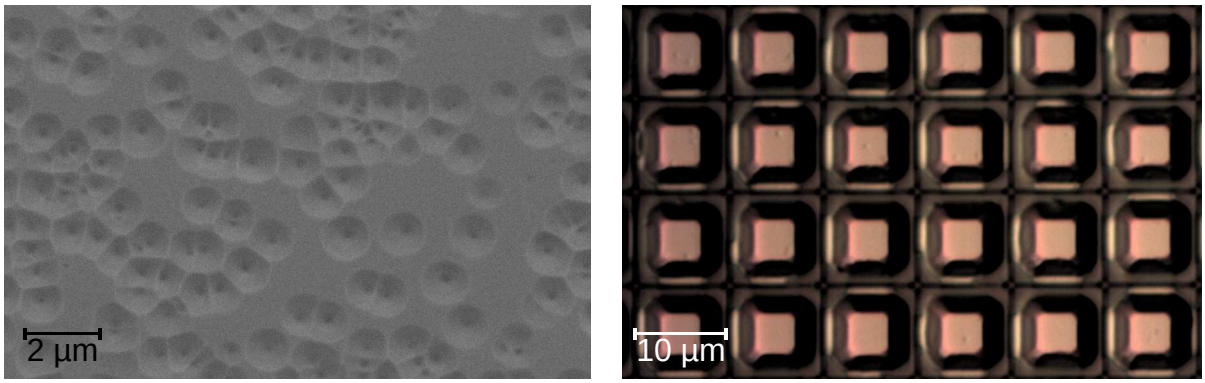
Thus, in the case of Ge films surrounding isolated SiO_2 lines as well, the accelerated overgrowth rate is maximized for SiO_2 lines with their longitudinal direction rotated away from the surface intersection of {111} planes by just a few degrees. The rate of coalescence both between mesas and around isolated SiO_2 lines is determined by the orientation parallel to the SiO_2 line's length, i.e. the direction in which the coalescence growth front quickly moves. In order to synchronize overgrowth over both orthogonal sets of trench lines in both structure types, SiO_2 lines must be oriented such that the angle offsets to nearby {111} surface intersection directions are equal.

5.2 Defect Evaluation in Ge films

Dislocations can be observed by several experimental techniques. TEM is the most accurate and detailed method of characterization as the analysis of samples allows for the determination of both burgers vectors and line directions of individual dislocations.[120] Planes of analysis include both sample cross-sections and plan-view by sample beveling. While the resolution provided by TEM is relatively high, the limited sampling rate diminishes its usefulness for most analysis in this thesis. Because typical TDD values for samples evaluated are on the order of 10^7 cm^{-2}, cross-sectional TEM would require ≈ 100 μm of lateral sample inspection to record statistically significant results. Plan view measurements increase the sampling area but still require significant sample preparation. Two other techniques exist (among others) for measuring surface TDD values: selective defect etching and electron-beam induced current (EBIC). The former method allows for large sampling with minimal preparation but reduced accuracy for high TDD values. The latter method requires device fabrication but generates more accurate results. EBIC samples can also be repurposed for further electrical characterization. Both of these methods are described below.

5.2.1 Selective defect etching

The strain fields created by dislocations distort the atomic bonds in the crystal lattice, increasing the potential energy of the chemical bonds. This increase in potential energy lowers the activation energy for etching reactions, causing etch rates to increase in the vicinity of the location where a dislocation terminates at an exposed surface.[121] By placing a film sample in a Ge etching solution, the Ge film will more quickly be etched at the locations of dislocations, forming etch pits approximately 1 μm in diameter (depending on etching time). Etch pits are clearly visible in optical microscopy, allowing for rapid analysis of large sample areas. A solution of 20% HNO_3, 10% HF, and 70% CH_3COOH with 0.3 g/L dissolved I_2 is used to evaluate the surface TDD of Ge films. In this etch chemistry, HNO_3 and I_2 oxidize the Ge film, while HF simultaneously etches the Ge oxide. The HNO_3 is kept in excess so that Ge oxide removal is the rate limiting step. Because a thin oxide always exists at the surface, film roughening (in the absence of defects at the surface) is minimized.[122] CH_3COOH serves as both a diluent and as a low dielectric alternative to water to maintain the oxidative properties of HNO_3.[123] Etching time in the solution is typically between 10



(a) (b)

Figure 5.21: (a) SEM conventional secondary electron top view of etch pits in a blanket Ge film. (b) Optical top view of etch pits in 10 µm mesa squares. Light colored areas are (001) planes in the mesa centers.

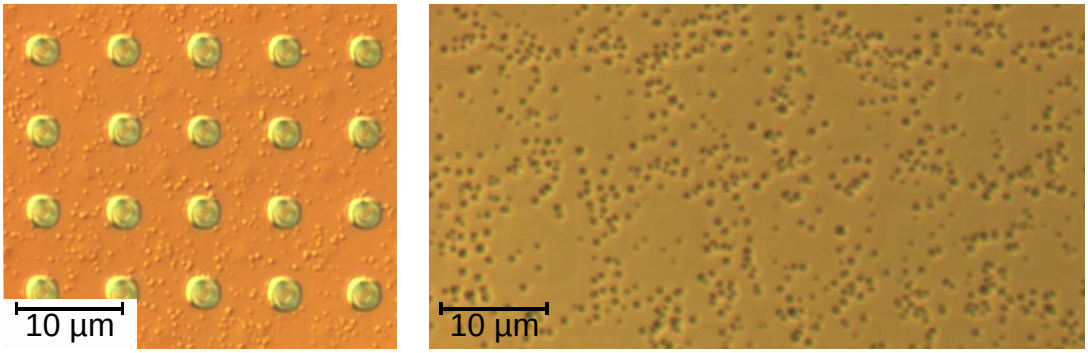
and 20 seconds, after which, the sample is submersed in two DI water baths in immediate succession and then rinsed.

There are, however, several drawbacks of the etch pit method that must be considered. Because individual etch pits are on the order of 1 µm in diameter, etch pits from adjacent dislocations can overlap with each other for large enough TDD levels. Figure 5.21a shows the resulting etch pits on a Ge film with an observed TDD of 8×10^7 cm^{-2}. While individual etch pits can be differentiated by SEM, they would not be discernible optically for this sample. Assuming a random distribution of threading dislocations at the film surface, the actual TDD, ρ_d, can be estimated from etch pit density (EPD) results:[96]

$$\frac{\text{EPD}}{\rho_d} = \frac{1 - \exp\left(-\rho_d \pi r^2\right)}{\rho_d \pi r^2} \tag{5.1}$$

where r is the estimated resolution of etch pit differentiation (e.g. 1 µm for optical measurements, 50 nm for SEM). Thus, only for TDD values 10^7 cm^{-2} and below can etch pit counting from optical images be reliable for accurate measurements.

The etch pit solution will also only provide satisfactory results on non-faceted structures.[96] In Ge mesas grown on (001) Si, the solution only provides sufficient contrast on the (001) plane. At mesa facets, the etching solution also etches the film laterally at greater rates than for the (001) planes. Therefore, before introducing samples to the etch pit solution, selectively-grown Ge films are first planarized using CMP. For Ge films grown in thin SiO_2 trenches that have not coalesced, the Ge films are only partially planarized as the defect density within the buffer thickness (80 nm)



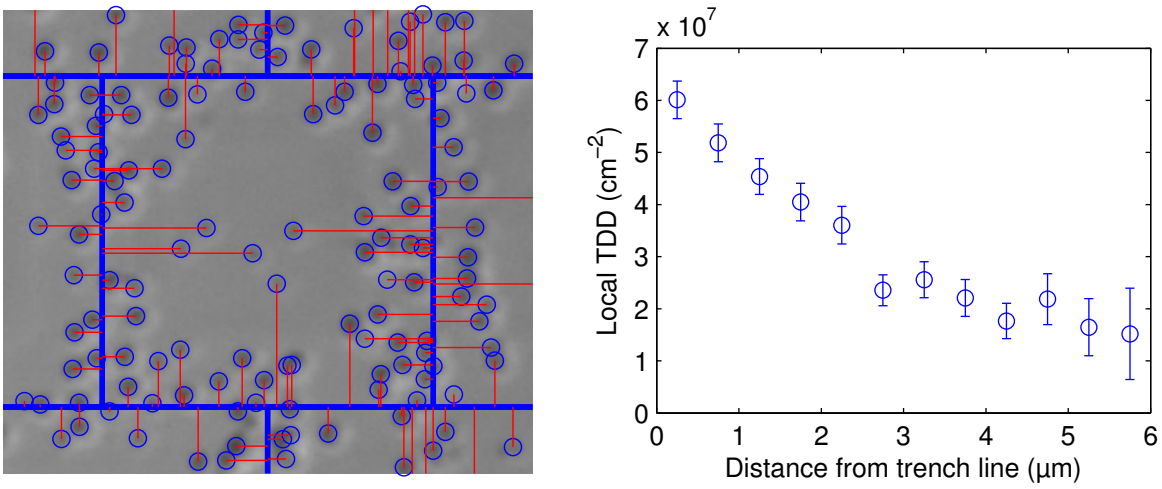
(a) (b)

Figure 5.22: Optical top views of etch pits in (a) a semi-coalesced array of 7.5 µm wide Ge mesas and (b) a fully coalesced staggered array of 12 µm wide Ge mesas. Circular holes ($\approx$ 2 µm in diameter) in (a) indicate regions of mesa corners where the SiO_2 lines are still exposed.

would not be representative of the TDD of the original Ge surface after annealing. When evaluating mesa squares with a sidewall length of 10 µm as shown in Figure 5.21b, only 25% of the total mesa area is evaluated. One cannot conclude from an observation of even zero etch pits within this region for every mesa that the entire mesa structure is free of threading dislocations protruding at its original surface.

One method to probe the surface TDD in the vicinity of the Ge film edges using the etch pit approach involves first laterally overgrowing the selectively-grown Ge film so that it coalesces, followed by CMP for planarization. At this point, if coalescence is complete, no edge in the Ge film will be available for the etch pit solution to aggressively attack, allowing for a full evaluation of the TDD on the Ge surface. Although mesas defined by an regular SiO_2 grid do not coalesce completely at mesa corners (see Section 5.1.2), obtaining EPD results for partially coalesced samples is still possible. In Figure 5.22a, etch pits in the array of semi-coalesced 7.5 µm wide Ge mesas are observed at a density of approximately 5×10^7 cm^{-2}, a value much greater than that observed in isolated mesas after cyclic annealing. The actual TDD in the partially coalesced film is underestimated by the measured EPD for both the resolution reason discussed earlier as well as the lack of Ge material above the SiO_2 grid surrounded by mesa corners.

For mesa arrays with a staggered orientation, coalescence is complete, and the EPD can be observed across the blanket Ge film as seen in Figure 5.22b. For mesas with widths of 12 µm (the largest width included in staggered mesa experiments), the EPD visually appears to be a function of distance to the nearest SiO_2 sidewall. To quantify the EPD dependence, etch pit images such

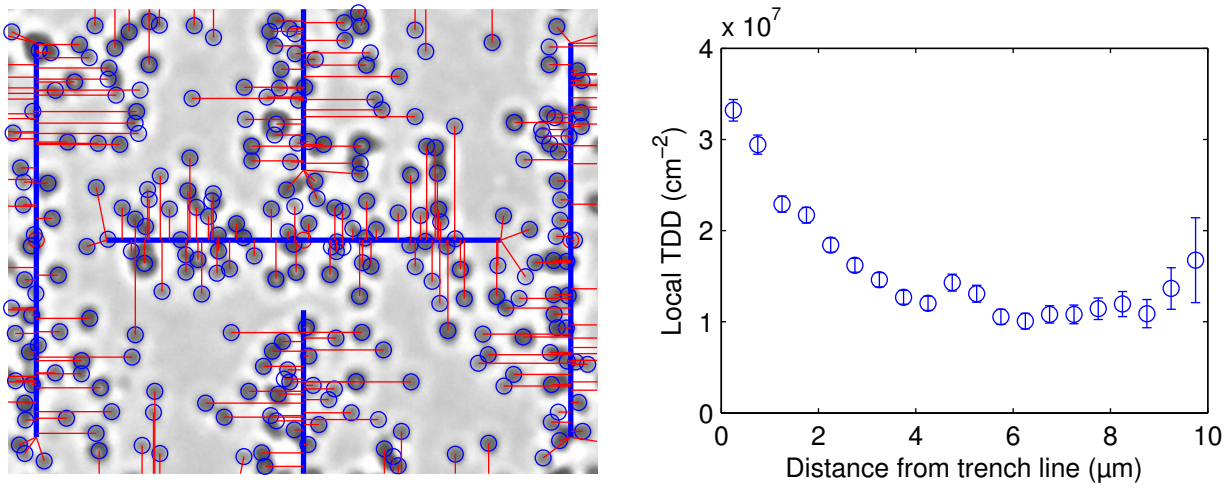


(a) (b)

Figure 5.23: (a) Image processing of EPD optical top view of coalesced 12 μm staggered mesas. Red lines are the shortest distance between each dislocation (blue circles) and underlying SiO_2 grid (blue lines). (b) Local EPD as a function of distance from the closest SiO_2 line determined by results partially displayed in (a).

as Figure 5.22b are fed through an image recognition script in Matlab to determine the location of each etch pit in the field of view and the distance between each pit and the closest SiO_2 line. The staggered SiO_2 grid structure is superimposed on the EPD image to conduct the analysis for which a partial section is shown in Figure 5.23a. The local EPD is determined by first defining distance bins and calculating the area in the image that corresponds to each bin. The number of etch pits in each distance bin is then divided by the area of the image within that bin. Results are shown in Figure 5.23b. The error bars correspond to the calculated value divided by the square root of the number of dislocations (observed "events") within that bin.

The observed EPD clearly decreases with increased distance from the mesa sidewalls. Multiple potentially valid hypotheses can predict this result but none can be completely ruled out without complimentary characterization. Based on the discussion in Section 4.2, the reduced thickness at the mesa edges during the annealing stage in the growth process could be responsible for a high TDD surrounding the mesa perimeter, preventing dislocation glide due to substantial dislocation pinning. In fact, the experimental results of Figure 5.23 roughly correspond with the theoretical predictions for faceted 10 μm mesas in Figure 4.19c in which the center of the mesa has a local TDD $\approx 10^7$ cm^{-2} while the TDD at the edge approaches 8×10^7 cm^{-2}. However, because the defect etch is conducted after coalescence, EPD results from coalesced films will not necessarily be



(a) (b)

Figure 5.24: (a) Image processing of EPD optical top view of coalesced Ge films surrounding isolated SiO_2 lines spaced 20 μm apart. (b) Local EPD as a function of distance from SiO_2 lines determined by results partially displayed in (a).

representative of the density and location of defects before lateral overgrowth. Additional defects may be generated as a result of overgrowth and coalescence that occur after the cyclic anneal or by thermal strain that builds while cooling down from the growth temperature. Based on these experimental results alone, one cannot conclude whether the observed threading dislocations could not be removed during the cyclic anneal or whether they were introduced after annealing.

Similar analysis can also be conducted for structures containing isolated SiO_2 lines. A dependence of the EPD on location is also observed around isolated lines separated by 20 μm (the largest separation included in experiments) as shown in Figure 5.24. For threading dislocations near the longitudinal edges of SiO_2 lines, the minimum distance is occasionally determined by a non-orthogonal distance to a SiO_2 line edge. For distances greater than approximately 5 μm, the local TDD becomes independent of distance until distances near 10 μm, at which the local TDD begins to increase with distance. While this increase could be within error due to the smaller area sampled for the largest distance bins, if real, this trend suggests an effective maximum glide distance of 5 to 10 μm. The main TDD reduction observed in isolated mesas of different sizes occurs for mesas with widths ranging from 10 to 20 μm,[89] corresponding closely to the average glide distances of 5 to 10 μm. Glide simulations for faceted 20 μm wide mesas partially defined by (001) facets in the mesa center also predict a weak local TDD increase at the mesa center (see

Figure 4.19e), providing another qualitative agreement between the glide simulations in Section 4.2 and experimental results presented here.

It is interesting to note that the TDD observed in locations closest to SiO_2 lines in mesas defined by a staggered grid approaches a value approximately twice that observed for structures with isolated SiO_2 lines. While not conclusive, this difference may be attributed to coalescence-induced defects that are avoided in the latter structure type. In work reported in literature on ART overgrowth, defects were observed to occur as a result of coalescence fronts impinging each other.[48] One explanation identified the cause as the strain resulting from a non-integer number of Ge lattice sites between the two fronts.[51] While impinging ELO growth fronts do not share a continuous, relaxed Ge lattice in mesas grown in SiO_2 grids and ART structures (which are effectively mesa strip arrays) prior to coalescence, the situation is quite different for structures with isolated SiO_2 lines surrounding a single Ge film. Even before overgrowth begins, the two Ge growth fronts along each isolated line's longitudinal direction are already connected at the line ends and thus share a well-defined Ge lattice. Assuming that the Ge film is relaxed, further extension of these smaller fronts across the trench lines during ELO is not expected to cause additional strain in the Ge film, reducing the likelihood that additional defects are generated.

5.2.2 III-V photovoltaic cells

Dislocations can also be spatially identified by electrical excitation techniques. One such method, cathodoluminescence (CL), raster scans an electron beam in an SEM, creating excited electron-hole pairs. Depending on the material, a portion of excited carriers radiatively recombine and the sample's photon emission can be recorded as a spatial function of the electron beam on the sample. If the area exposed to the electron beam is within a minority carrier diffusion length of a dislocation, a portion of the excited carriers will recombine nonradiatively due to the defect state of the dislocation, decreasing the observed photoemission signal. The effects of individual dislocations can be observed as dark spots in a spatial map of the CL signal. While CL does not inherently require fabrication of a multilayered device structure, it is only useful for materials with strong radiative recombination. Ge, having an indirect band gap, will not produce a strong signal using

this technique. Furthermore, the CL method in general can also suffer from poor light extraction due to the relatively high refractive indexes of the materials typically probed.

A similar measurement technique that also rasters an electron beam is electron beam induced current (EBIC). In this method, instead of collecting photons resulting from the radiative recombination of excited carriers, the excited carriers are electrically collected by contact probes. Excess carrier collection requires rectification by either a p-n junction or Schottky contact in part of the device containing the layer to be probed for its TDD. Similar to CL, in devices containing a p-n junction, the collected current from excited carriers will be reduced in the vicinity of a dislocation due to recombination. The resolution of both CL and EBIC is a function of the minority carrier diffusion length in the material containing dislocations.[124] Due to the relatively large diffusion length in Ge, the dark spots in an EBIC image are typically 5 μm in diameter.[125] With a resolution of this magnitude, identifying individual threading dislocations is difficult for TDD values above 10^6 cm^{-2}.

However, because threading dislocations that terminate on the Ge film surface will continue into epitaxially deposited direct-gap III-V layers, the TDD in the III-V layers can provide a proxy for measurement of the TDD on the Ge film surface. While additional dislocations can be introduced at the Ge/III-V interface, thus causing the indirect measurement to serve as an upper bound, increases to the TDD can be avoided by depositing lattice-matched III-V compositions. Because the minority carrier diffusion lengths (in dislocation-free material) are smaller in the nearly-latticed match III-V materials, GaAs and $In_{0.49}Ga_{0.51}P$, compared to in Ge, dark spots in EBIC maps will be smaller, enabling identification of single dislocations for dislocation densities on the order of 10^7 cm^{-2}. Because $In_{0.49}Ga_{0.51}P$ has smaller diffusion lengths compared to those in GaAs, $In_{0.49}Ga_{0.51}P$ single-junction solar cells were grown on Ge films to both characterize the TDD in the Ge films grown on Si as well as to experimentally determine the TDD threshold for efficiency reductions of $In_{0.49}Ga_{0.51}P$ cells. This threshold determined experimentally has primarily been reported in literature for GaAs junction cells.

Photovoltaic cells were grown on three different substrates: bulk Ge, blanket Ge-on-Si, and coalesced selectively grown Ge-on-Si. All substrates were n-type and nominally (001) oriented, miscut by 6° to (111). Bulk Ge wafers were used as received due to their epi-ready surface. Ge-on-Si samples were cleaned in diluted H_2SO_4 (1:4) to remove organics without severely etching

the Ge, followed by HF to dissolve any GeO_x that formed on the surface, and finishing with diluted HCl (1:4) to passivate the Ge surface with Cl prior to epitaxy.[126] The solar cell structure consisted of the following layers in order of deposition:

III-V initiation:	n^+ GaAs	600 nm
Back-surface field (BSF):	n^+ $Al_{0.5}In_{0.5}P$	50 nm
Base:	n^- $In_{0.49}Ga_{0.51}P$	1000 nm
Emitter:	p^+ $In_{0.49}Ga_{0.51}P$	100 nm
Window layer:	p^+ $Al_{0.5}In_{0.5}P$	30 nm
Contact layer:	p^+ GaAs	300 nm

III-V epitaxy was conducted in a Thomas Swan MOCVD reactor at growth conditions of 650°C, 100 Torr, a V/III ratio of 77, and N_2 carrier gas. Si and Zn were used as n-type and p-type dopants, respectively. To create a sample with a moderate TDD level, e.g. 10^6 cm^{-2}, a similar device structure was grown on a Ge substrate by growing a lattice-mismatched In-poor $Al_{1-x}In_xP$ BSF layer to generate dislocations with the rest of the device layers unchanged. The top metal contact stack of Ti/Pt/Au (10/20/300 nm) was deposited by electron beam evaporation and patterned by lift-off. In order for photons to reach the $In_{0.49}Ga_{0.51}P$ layers, the highly-doped GaAs contact layer was removed by wet etching in a solution of H_3PO_4:H_2O_2:H_2O (3:1:25), using the metal contact layers as a hard mask. The etch chemistry is selective to arsenides over phosphides and thus stops on the $Al_{0.5}In_{0.5}P$ window layer. Individual cells are then defined by a mesa etch, using photoresist as a mask and a solution of HCl:H_2O (2:1) to selectively remove the $In_{0.49}Ga_{0.51}P$ and $Al_{0.5}In_{0.5}P$ layers while stopping on the GaAs III-V initiation layer.

For backside contacting of bulk Ge wafers, a 5 nm layer of TiO_2 followed by 300 nm of Ti were sputter deposited from TiO_2 and Ti targets, respectively. Schottky contacts typically form on n-type Ge due to Fermi-level pinning from Ge/metal surface defect states whose energy levels are situated closer to Ge's valence band. Placing a thin layer of TiO_2 between n-type Ge and metal layers has been found to substantially reduce Fermi-level pinning without introducing a large tunneling resistance.[127] For cells grown on Ge layers deposited on Si, the n-type contact was instead made on the GaAs III-V initiation layer because the Ge films were grown without intentional doping. A

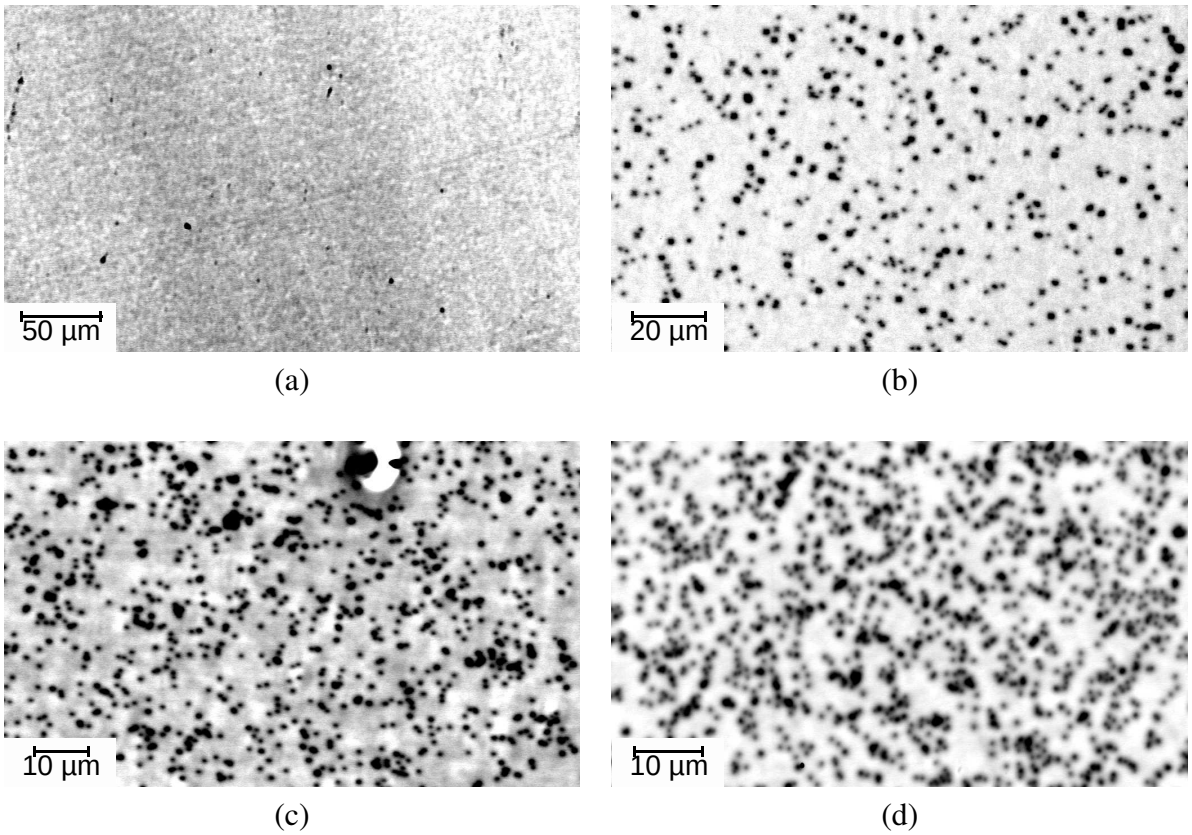


Figure 5.25: EBIC images of $In_{0.49}Ga_{0.51}P$ solar cells (courtesy of Dr. Tae-Wan Kim) fabricated on the following substrates: (a) Ge, (b) Ge with a lattice-mismatched BSF, (c) Ge-on-Si blanket film, and (d) Ge-on-Si selectively grown and coalesced film. The corresponding TDD values are 4.1×10^4, 3.6×10^6, 1.5×10^7, and 2.5×10^7 cm^{-2}.

metal stack of Ni/Ge/Au (5/20/300 nm) was deposited on the n^+ GaAs by electron beam evaporation using a lift-off process for patterning.

EBIC measurements, conducted using a Helios NanoLab SEM and a SRS SR-570 low-noise current amplifier, are shown in Figure 5.25 for cells grown on Ge bulk wafers and Ge-on-Si virtual substrates. The same solar cell structures were also evaluated for their photovoltaic device performance. Each cell was illuminated with a Xenon arc lamp, producing a flux equivalent to approximately $0.5-1.0\times$ Sun irradiation. Current-voltage sweeps were obtained using a HP4145A semiconductor analyzer. In Figure 5.26, the V_{oc} of $In_{0.49}Ga_{0.51}P$ cells grown on the four different substrates are plotted as a function of the TDD as determined by EBIC.

Because $In_{0.49}Ga_{0.51}P$ is a high E_g material, it is expected that the junction dark current will be dominated by recombination in the depletion region. Therefore, the V_{oc} is related primarily to the

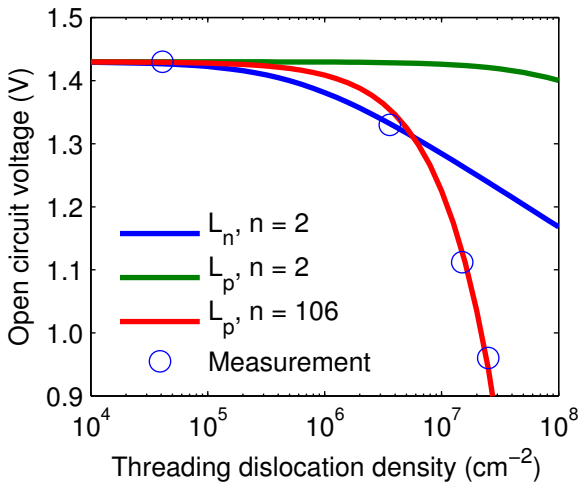


Figure 5.26: V_{oc} vs. TDD as determined by EBIC for $In_{0.49}Ga_{0.51}P$ solar cells.

dark current $J_{0,2}$ of the junction:[21]

$$V_{oc} = \frac{2kT}{q} \ln\left(\frac{J_{sc}}{J_{0,2}} + 1\right) \tag{5.2}$$

where $J_{0,2} \propto \tau^{-1}$ (see Eq. 2.15). The V_{oc} can be estimated as a function of minority carrier lifetime which itself depends on the TDD according to Eq. 2.10 with possible modification based on Eq. 2.9. Assuming minority carrier diffusivity determines the effect of dislocations on carrier lifetime, for a p-n junction with threading dislocations, V_{oc} can be written as an explicit function of the TDD, ρ:

$$V_{oc} \approx \frac{2kT}{q} \ln\left(\frac{2J_{sc}}{qn_i W}\right) - \frac{2kT}{q} \ln\left(\frac{1}{\tau_0} + \frac{\rho\pi^3 D_{n,p}}{4}\right) \tag{5.3}$$

The first term in Eq. 5.3 is independent of the TDD and is only weakly dependent on the illumination conditions and junction characteristics. An initial value for the first term was determined from $J_{sc} = 10$ mA/cm^2, $n_i = 2 \times 10^3$ cm^{-3}, and $W = 50$ nm. Due to some uncertainty for several of these variables, including temperature, this term is modified during regression so that V_{oc} approaches 1.43 V (the value measured from lattice-matched cells on Ge substrates) as $\rho \to 0$.

A regression is performed using a modified form of Eq. 5.3 in which only the TD-free diffusion length, L_0, is explicitly defined while the minority carrier diffusivity gets absorbed into the first

term:

$$V_{oc} \approx \frac{2kT}{q} \ln\left(\frac{2J_{sc}}{qn_i W D_{n,p}}\right) - \frac{2kT}{q} \ln\left(\frac{1}{L_0^2} + \frac{\rho\pi^3}{4}\right) \tag{5.4}$$

Using the two data points from cells grown on bulk Ge substrates, regression results in a diffusion length of $L_0 = 4.6$ μm, corresponding closely to that derived from $D_n = 27$ cm^2/s[128] and $\tau_{0,n} = 7.8$ ns[27] despite the n-type base of the fabricated $In_{0.49}Ga_{0.51}P$ cells. Previous experimental work reported in literature on GaAs cells discussed in Section 2.2 would suggest that the shorter diffusion length of holes, $L_{0,p} \approx 300$ nm, should instead determine the effect of TDD on V_{oc} for p^+/n cells. The curve corresponding to the diffusion length of holes in TD-free n-type $In_{0.49}Ga_{0.51}P$ in Figure 5.26 significantly overestimates the expected V_{oc} for TDD > 10^6 cm^{-2}. Simulations of GaAs cells presented in Section 2.2 indicated that while the lifetime of the minority carrier in the base more strongly effects the V_{oc} of a solar cell, the distinction reduces as the geometric mean of both carrier lifetimes, and by extension, diffusion lengths, decreases. As the TD-free diffusion lengths of both electrons and holes in $In_{0.49}Ga_{0.51}P$ are smaller than those in GaAs, $In_{0.49}Ga_{0.51}P$ cells are less likely to show a differentiation in the V_{oc} response between changes to the lifetime of one carrier over the other.

Only by adjusting the diode ideality constant from 2 to over 100 can all four data points be simultaneously fit. As this ideality constant is non-physical, other factors not directly related to the existence of TDs are likely responsible for the lower than expected V_{oc} values for the TDD cells grown on Ge virtual substrates. Reductions in the V_{oc} could be due to post-epitaxy device fabrication. No surface passivation layers were deposited on the sidewalls of the cells after mesa etching, causing significant variation in cell performance from the same substrate sample. The data presented in Figure 5.26 represents the highest measured V_{oc} values from cells fabricated on each substrate type.

Another potential cause for the underestimated reductions in V_{oc} is insufficient substrate cleaning before III-V epitaxy. Surface cleaning is not an issue for cells deposited on bulk Ge substrates as they come prepared as epi-ready, having a thin protective oxide film on the surface that can easily be desorbed under vacuum at high temperature immediately before epitaxy begins.[129] Unlike the specular films deposited on bulk Ge substrates, the deposition on the Ge virtual substrates in the same growth run appeared slightly hazy, indicating non-ideal surface conditions.

Future work involving III-V growth on Ge-on-Si virtual substrates should include proper surface cleaning of the Ge substrate prior to epitaxy. One promising option for Ge surfaces is the formation of a thin Ge oxide created by immersion in H_2O_2 after prior cleans that result in a fresh Ge surface.[130] Because GeO_2 will dissolve in water, the H_2O_2 bath must be followed by careful drying in N_2 without rinsing. With proper surface cleaning prior to to epitaxy and an improved process flow that includes mesa sidewall passivation, the trend between TDD and V_{oc} will be made easier to ascertain.

More experiments involving both n^+/p and p^+/n $In_{0.49}Ga_{0.51}P$ cell structures should be conducted before any concrete conclusion can be made as to the fundamental mechanism that determines the V_{oc} of $In_{0.49}Ga_{0.51}P$ cells. Additional experiments for which n-type dopants are changed from Si to Te may shed light on whether the dopants themselves have a significant effect on the passivation of trap states in threading dislocations in either $In_{0.49}Ga_{0.51}P$ or GaAs cells. However, based on the data presented, one can conclude that for a $In_{0.49}Ga_{0.51}P$ cell with a TDD level at or below 4×10^6 cm^{-2}, the V_{oc} is expected to retain at least 94% of its TD-free value due to the presence of threading dislocations. If the TDD can be reduced further in Ge films grown on Si, GaAs solar cells additionally become available as potential applications for these Ge virtual substrates.

5.3 Summary

This chapter began with developing an understanding of the Ge growth process, from the early stages of trench fill, to the initiation of lateral overgrowth, and to the completion of film coalescence. Trench parameters including the SiO_2 lines' spacing, width, concavity, and relative orientation to the crystal structure of the substrate affect the various stages of film growth. Competition for adatoms between different facets and the high surface energy of the Ge/SiO_2 interface are the primary physical mechanisms that bring about the trends observed and discussed. While the initially proposed structure of a regularly arranged mesa array defined by an SiO_2 grid was not found to easily coalesce completely, two successful alternatives were identified: staggered mesa arrays and film growth around isolated SiO_2 lines.

Two methods of identifying threading dislocations were then discussed: selective defect etching and EBIC. Results from etch pit experiments of completely coalesced films indicated that the local

TDD was greatest at the trench sidewalls, qualitatively agreeing with the simulated predictions of strong dislocation pinning at the edges of faceted Ge films due to the reduce local film thickness as discussed in Section 4.2. The fabrication of $In_{0.49}Ga_{0.51}P$ solar cells on Ge virtual substrates allowed for indirect determination of the TDD for Ge films by EBIC measurements. Comparison of the V_{oc} for cells grown on bulk Ge substrates at low and moderate TDD levels (4.1×10^4 cm^{-2} and 3.6×10^6 cm^{-2}, respectively) provided an upper bound for the expected V_{oc} reduction if the TDD of large area Ge virtual substrates can be reduced below the latter TDD level. One potential application of single junction $In_{0.49}Ga_{0.51}P$ solar cells in high efficiency, hybrid solar energy conversion systems is next described in Chapter 6.

Chapter 6

Broadband photonic design

6.1 Efficient broadband energy conversion

As briefly mentioned in Section 2.1, the efficiency of a single-junction solar cell of increasing E_g is limited by the balance of decreasing carrier thermalization (higher V) and increasing photon transparency (lower I). One proposed method to overcome this limitation for single-junction solar cells is the inclusion of an intermediate band (IB) in the semiconductor band gap created by the introduction of impurities such as quantum dots in order to increase photon absorption. Pairs of photons below E_g could excite carriers first from the valence band to the IB, and then from the IB to the conduction band. Without the IB, neither photon would otherwise contribute to the photocurrent. While IB solar cells have been predicted to promise energy conversion efficiencies above 60%,[131] in implementation, quantum efficiencies below E_g are typically at most 5% and quickly drop by orders of magnitude at lower photon energies.[132] Complications caused by introduction of the IB also decrease the V_{oc} of the cell, thus providing a net disadvantage to conversion efficiency. The complimentary method of single exciton fission uses materials such as tetracene to convert energy absorbed from one photon to generate two excitons, reducing thermalization losses by increasing the photogenerated current. By transfer of each exciton to a lower E_g solar cell, thermalization losses are reduced for the single high energy photon initially absorbed. While internal quantum efficiencies above 125% have been reached,[133] the approach still needs significant refinement to create a net efficiency benefit to compete with single junction cells with efficiencies already close to their Shockley-Quissier limit.

6.1.1 Multijunction Photovoltaics

The efficiency limitation for single-junction solar cells have however been successfully surpassed by multijunction approaches. By apportioning different sections of the solar spectrum to specific E_g

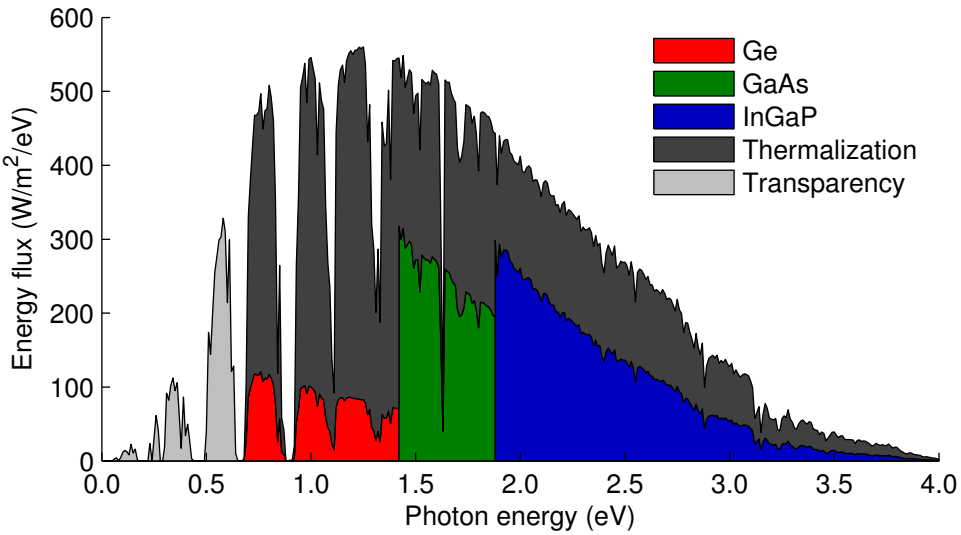


Figure 6.1: AM1.5D spectrum partitioned to three separate solar cells. Losses due to thermalization and material transparency are also indicated.

materials, thermalization and transparency losses can be simultaneously reduced. Low E_g materials such as Ge can collect low energy photons that would otherwise not be absorbed while very high E_g materials such as InGaP can absorb high energy photons with reduced thermalization losses. Figure 6.1 visualizes the fraction of the AM1.5D spectrum that can theoretically be captured by Ge, GaAs, and $In_{0.49}Ga_{0.51}P$ solar cells, each operating at their maximum power point at $T = 25°C$ and 1× Sun concentration. Assumptions in the calculation include independent cell operation, no series resistance losses, $V_{oc} = E_g - 0.4$ V,[134] fill factors determined by Eq. 2.3, and 100% external quantum efficiency for photons above E_g when the cell is operated in reverse bias.

Multijunction photovoltaic cells have primarily been realized in practice in a tandem architecture in which individual junctions are stacked directly on top of each other, electrically connected in series. The highest E_g material is placed at the top of the structure to absorb the high energy photons, allowing photons with energies below the band gap to transmit through the first cell to the next cell with a lower E_g. Ideally, this structure passively splits the incoming spectrum such that each photon is absorbed in the highest band gap material that can absorb it. Tunnel junctions are placed between sub-cells to enable recombination of extracted carriers, completing the internal circuit of the device. Operation of tandem cells enforces a current matching constraint between junctions due to the series arrangement of the cells. Any excess photocurrent that is generated in a given sub-cell will be lost via recombination within the sub-cell since complimentary excited carriers

from adjacent sub-cells will not be available at tunnel junctions. For some combinations of band gaps, layer thicknesses can be adjusted to minimize efficiency losses due to current mismatch. In the InGaP/GaAs/Ge materials set, the ideal photocurrent from each junction follows the trend: $I_{Ge} > I_{InGaP} > I_{GaAs}$. Current matching between InGaP and GaAs is achieved by thinning the InGaP cell to allow some photons with $h\nu > E_{g\text{-}InGaP}$ to instead be absorbed in the GaAs layer. Since no lower E_g junction exists below Ge, the excess photocurrent in Ge is lost. While more optimal E_g current matching combinations exist such as lattice-matched InGaP/GaAsP/SiGe,[135] current matching will only be achieved for a specific reference spectrum. In practice, the spectrum a solar cell receives not only varies by location but by the seasonal/hourly position of the sun and changing weather conditions.[136]

An alternative multijunction approach involves one or more individual sub-cells connected in parallel rather than series, reducing (or eliminating) current matching constraints and in some cases, lattice-matching constraints as well. Individual sub-cells may still be multijunction tandems but are only tasked with collecting a limited portion of the solar spectrum. Some designs maintain mechanical stacking of individual sub-cells to take advantage of passive filtering by placing the highest band gap junction on the top of the stack and aligning top and bottom contact grids to allow unabsorbed light to reach the next cell below.[137] Active splitting systems typically employ prisms[138] or diffractive optical elements[139] to separate light into a continuously changing spectrum on a receiver plane containing a discrete set of solar cells closely spaced to each other. Alternatively, a spectral filter can be used to separate light into two discrete photon energy ranges, one reflected by the filter and the other transmitted.[140] Cascades of filters can further split the spectrum into more than two components if many individual junctions are included in the overall device.[141]

Obtaining near perfect spectrum splitting by any of these optical methods has not yet been demonstrated, reducing the overall benefit compared to the more prevalent tandem cell architecture. Because different solar cells will operate at different voltages per cell, additional considerations must be made for power electronics.[142] Because multijunction cells typically use concentrated sunlight, efficiency reductions due to series resistance cannot be understated for concentrated photovoltaic (CPV) cells. Series resistance results from resistive losses in the metal grid and the lateral conduction in the top junction emitter. Because for the same set of materials, the current

level in both lateral multijunctions and tandem cells will be equivalent to first order, the voltage drop across the resistive emitter and and metal grid will be approximately the same. However, while a tandem cell only experiences this voltage reduction once, a parallel configuration experiences the same voltage drop for each independently operating cell, unintentionally multiplying the loss mechanism.[136] One partial solution around this limitation is significant reduction of the area per cell but increased costs in doing so must also be taken into consideration to determine the optimal cell size depending on the relevant metric (e.g. $/W).[143]

6.1.2 Hybrid Conversion Systems

Solar energy can also be utilized by concentrating sunlight on thermal receivers and collecting the thermal energy in heat transfer fluids. The captured energy can either be temporarily stored for several hours or used immediately to run a steam turbine and generate electricity.[144,145] The primary advantage of these concentrator solar power (CSP) systems is the relatively inexpensive means for energy storage, reducing the intermittent aspect of solar energy that will otherwise prevent greater penetration of solar into the electrical grid.[146] For a typical CSP plant operating at a high temperature $T_H = 565°C$ (the creep temperature of steel) and using molten salts such as sodium and potassium nitrate as the heat transfer fluid,[147] practical conversion efficiencies for generating electricity (ignoring collection efficiency) are approximately 40%, estimated using the semi-empirical endoreversible efficiency calculation: $\eta_{th} = 1 - \sqrt{T_L/T_H}$.[148] The efficiency of CSP is fairly insensitive to photon energy as the thermal receivers use materials that strongly absorb photons across a broad range of the solar spectrum.[149]

In contrast, the efficiency of photovoltaic cells is a strong function of photon energy, dependent on the semiconductor(s) employed. Even if every photon were absorbed in a material with a band gap just below the photon's energy, thermalization losses will still exist due to the approximate 0.4 eV difference between E_g and V_{oc} that naturally arises due to the intensity magnitude of the solar irradiation and the relationship between cell E_g and radiative recombination. Because this offset is, to first order, independent of E_g, fractional energy losses due to thermalization will be most severe for photons absorbed in low E_g junctions. As illustrated in terms of power in Figure 6.1, the photon specific efficiency, $\eta_{h\nu}$, for Ge ranges from 28% at its own E_g down to 13% at the E_g

of GaAs. Even for the relatively high E_g material GaAs operating at room temperature, $\eta_{h\nu}$ will approach 44% near E_g of $In_{0.49}Ga_{0.51}P$ and will further reduce to 42% if the cell instead operates at 60°C (a typical operating temperature for passively cooled concentrator solar cells).[136] With consideration of other loss sources relevant to practical devices including contact shadowing, series resistance, and quantum efficiencies below 100%, $\eta_{h\nu}$ of GaAs can approach or drop below the expected efficiency of solar thermal energy conversion systems operating at high T_H. For low E_g materials such as Ge, $\eta_{h\nu}$ will always be surpassed by the conversion efficiency of a thermal cycle unless T_H is relatively low (e.g. $T_H \approx 150°C$). While $\eta_{h\nu}$ for $In_{0.49}Ga_{0.51}P$ is estimated to surpass 65% near its E_g, if no materials with a greater E_g are included, the ideal device $\eta_{h\nu}$ will drop to 45% for photon energies greater than 2.8 eV.

For a given set of photovoltaic materials available, $\eta_{h\nu} > \eta_{th}$ will hold for a limited range of high photon energies primarily focused in the visible portion of the spectrum, while the reverse will be true for low energy photons as well as for photons with $h\nu \gg {E_g}^{max}$. Because photovoltaic and solar thermal approaches are favored for different portions of the solar spectrum, higher overall conversion efficiencies should be possible by developing a hybrid system that incorporates elements of both CPV and CSP. By apportioning a limited range of high energy photons to photovoltaics cells and the remaining photons to thermal receivers, the energy of each photon can be collected using the method that is most efficient for it. Many different arrangements of hybrid systems have previously been explored.[150]

In one conceptual example, photon energies between 1.9 and 2.5 eV can be sent to single junction $In_{0.49}Ga_{0.51}P$ photovoltaic cells. All photons with $h\nu < 1.9$ eV and some photons with $h\nu > 2.5$ eV are sent to a thermal receiver. In addition to increased thermalization losses for higher energy photons, external quantum efficiencies also tend to decrease above 2.5 eV due to increased effects of recombination as a greater portion of these photons are absorbed in the emitter layer.[151] While a GaAs junction could also be included, limitation to a single junction prevents any further reduction in efficiency due to a current matching constraint if a tandem cell is used or additional series resistance and optical losses if a parallel junction approach is adopted. Spectrum partitioning can be achieved by the design of a dichroic filter. Section 6.1.3 will first describe the physics behind the operation of a filter and the methodology developed to design a dichroic filter specifically for the hybrid system described above.

6.1.3 Distributed Bragg Reflectors

Theory

Much like the periodic potential resulting from the repeated atomic nuclei that is responsible for electronic band gaps in $E-k$ dispersion relations (as illustrated in Figure 2.3), photonic crystals, repetitive structures with a periodic refractive index, similarly create optical band gaps, preventing the propagation of certain photon frequencies. A one-dimensional photonic crystal, or distributed Bragg reflector (DBR), is composed of a multilayer stack of materials with alternating high and low refractive indexes. Applications for these structures have included lasers,[38] waveguides,[152] optical fibers,[153] and back-side reflectors for photovoltaic cells[154] and light-emitting diodes.[155]

To create a stop band centered around a particular photon frequency, layer thicknesses d_i are initially determined by the "quarter-wave" criterion: $d_i = \lambda_0/4n\cos\theta_i$, where λ_0 is the corresponding wavelength in the stop band center, n is the real component of the refractive index for the layer material at the center wavelength, and θ_i is the angle between the wave's propagation direction and the film surface normal. This criterion maximizes the constructive interference of waves reflected off each interface in the opposite direction of the incident light. Multilayer structures composed of transparent materials can easily achieve near 100% reflectivity at the desired photon frequency ν_0. Typically, DBR multilayers are composed of two optical materials with frequency dependent refractive indexes n_1 and n_2. In the limit of an infinite number of layers, the bandwidth of the stop band $\Delta\nu$ at normal incidence can be calculated:

$$\frac{\Delta\nu}{\nu_0} = \frac{4}{\pi}\left[\arcsin\left(\frac{n_2 - n_1}{n_2 + n_1}\right)\right] \tag{6.1}$$

Higher index contrast, $\Delta n = n_2 - n_1$, increases the stop band width normalized to the center frequency. Because greater index contrast also increases the reflection coefficients at each interface, larger Δn allows a finite structure to approach the infinite result with less layers. The calculated normal incidence reflection spectrum for a DBR composed of 29 alternating layers with n_L = 1.5 and n_H = 2.0 surrounded be a medium of n_0 = 1.5 (to minimize additional reflections) is shown in Figure 6.2.

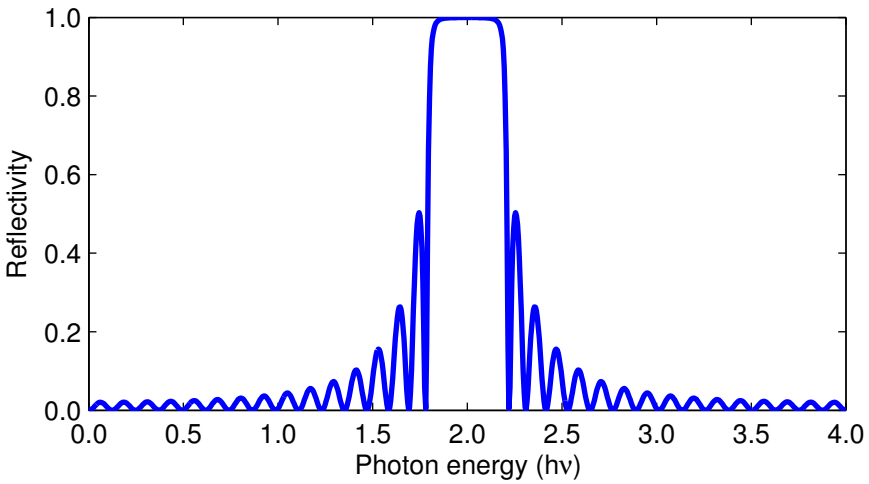


Figure 6.2: Reflection spectrum of a 29-layered DBR with $n_{\mathrm{L}} = 1.5$ and $n_{\mathrm{H}} = 2.0$.

The transfer matrix method allows for the determination of the expected reflection spectrum from an arbitrary multilayered structure. For each layer and interface between layers, matrices are defined that relate forward, $E^{+}(x)$, and backward, $E^{-}(x)$, waves on both sides of the component:[156]

$$\begin{bmatrix} E_i^+ \\ E_i^- \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} E_j^+ \\ E_j^- \end{bmatrix} \tag{6.2}$$

These matrices each account for propagation/absorption (in layers) and reflection/transmission (at interfaces) based on the Fresnel relations. Propagation matrices, $\mathbf{P}_j$, and interface boundary matrices, $\mathbf{B}_{i,j}$, are calculated as follows:

$$\mathbf{P}_j = \begin{bmatrix} \exp\left(ik_j d_j\right) & 0 \\ 0 & \exp\left(-ik_j d_j\right) \end{bmatrix} \tag{6.3a}$$

$$\mathbf{B}_{i,j} = \frac{1}{2a_{ij}\tilde{n}_j} \begin{bmatrix} \tilde{n}_i + \tilde{n}_j & \tilde{n}_i - \tilde{n}_j \\ \tilde{n}_i - \tilde{n}_j & \tilde{n}_i + \tilde{n}_j \end{bmatrix} \tag{6.3b}$$

where $k_j = 2\pi n_j \cos\theta_j / \lambda_0$, n_j is the complex refractive index for layer j, $\tilde{n}_j = n_j \cos\theta_j$ and $a_{ij} = 1$ for TE polarized waves and $\tilde{n}_j = n_j \sec\theta_j$ and $a_{ij} = \cos\theta_i / \cos\theta_j$ for TM polarized waves. If one or more of the materials are lossy ($\kappa > 0$) in the photon range of interest, θ_j will take on complex values as determined by Snell's law: $n_i \sin\theta_i = n_j \sin\theta_j$. The overall transfer matrix that relates the

electric fields of forward and backward traveling waves at the incident surface of the multilayer, $E_0^{\pm}$, to those on the other side, $E_s^{\pm}$, is determined by the product of the matrices:

$$\mathbf{M} = \mathbf{B}_{0,1}\mathbf{P}_1\mathbf{B}_{1,2}\mathbf{P}_2\cdots\mathbf{P}_N\mathbf{B}_{N,s} \tag{6.4}$$

where 0 represents the incident medium (e.g. air) and s represents the substrate (or medium) behind the last layer. To determine the reflection and transmission electric fields, the ratios $r = E_0^-/E_0^+$ and $t = E_s^+/E_0^+$ are found from the matrix components of M by setting E_s^- to zero (see Eq. 6.2). The electric fields are squared to obtain reflection and transmission in terms of power:

$$R = \left|\frac{M_{21}}{M_{11}}\right|^2 \qquad T = \frac{n_s}{n_0}\left|\frac{1}{M_{11}}\right|^2 \tag{6.5}$$

For an angle of incidence greater than zero, the calculation outlined above is conducted separately for both TE and TM polarized light. The reflected and transmitted power calculated by Eq. 6.5 for both polarizations are averaged based on the polarization of the relevant illumination source.

Dichroic mirror design

Because a DBR naturally produces a stop band typically spanning up to 20% of the central frequency on either side (depending on the index contrast), it is more straightforward to design the dichroic to reflect photon energies between 1.9 eV and 2.5 eV to the photovoltaic array than attempt to transmit within this range and reflect all remaining frequencies. The design of a reversed filter is not impossible but requires multiple DBR structures with different stop bands that partially overlap with each other, thus requiring significantly thicker multilayers. Because the reflected and transmitted light paths of the filter must be different, an angle of incidence of 45° of unpolarized light from air is assumed for the purposes of design. The alternating materials that compose the DBR must both be of sufficient index contrast to create a wide enough stop band and simultaneously be transparent across the majority of the solar spectrum. The combination of SiO_2 ($n_L \approx 1.45$) and TiO_2 ($n_H \approx 2.4$) has sufficient contrast, and TiO_2 only begins to weakly absorb above its E_g of 3.0 eV[157] (which constitutes a small fraction of the solar spectrum). A DBR structure is initially designed using the quarter-wave criterion for a center photon energy of 2.2 eV, using a 16-layer

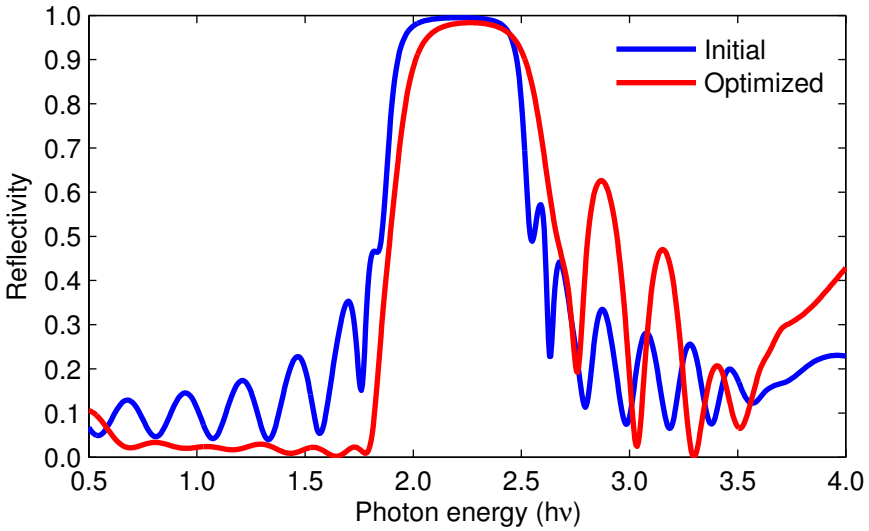


Figure 6.3: Reflection spectra of SiO_2/TiO_2 16-layer DBR designs deposited on a quartz substrate.

design alternating between SiO_2 and TiO_2 deposited on a quartz substrate ($n_s \approx 1.5$), for a total film thickness of 1.4 μm. An average reflectivity above 95% is predicted in the intended stop band for the initial design as shown in Figure 6.3. It should be noted that both reflection spectra in Figure 6.3 do not account for reflections from the quartz/air interface on the backside of the substrate. With an estimated 4% reflectivity if the interface is left unmodified, reflection at this interface can be sufficiently reduced by application of a multilayer, broadband anti-reflection coating (ARC). Because the substrate thickness for the dichroic mirror, e.g. 1 mm, is much greater than the coherence length of sunlight, e.g. 1 μm,[156] the design of the DBR and the ARC can be conducted independently.

Because a dichroic mirror to be used as a spectrum splitter must operate across a broad range of photon energies, additional aspects of DBR design must be considered. While for an ideal spectrum splitter, reflectivity outside of the intended stop band is zero, side bands will appear in the reflection spectrum. These unintended reflections occur due to the discontinuous changes in the refractive index at each layer interface, creating higher-order frequency components in the Fourier transform of the structure. The alternative rugate filter, composed of a film with a continuously changing sinusoidal refractive index, reduces this effect but does not immediately eliminate it.[158] Both DBR and rugate filters still require modification to reduce impedance mismatch at their front and back surfaces with the outside media. Anti-reflection techniques exist for both class of filters but typically

require several additional layers for a DBR[159] or an increased range of the continuously varying refractive index for a rugate filter.[160] To create a filter with a prespecified reflection spectrum, the truncated spectrum's Fourier transform can determine the ideal inhomogeneous film (rugate-like) design which can then be approximated by a stack composed of homogeneous layers.[161] Often, these solutions require a total film thickness of several microns as well as very high refractive indexes which do not correspond to any known materials transparent in the visible spectrum. If additional limitations are placed on the number of distinct refractive indexes in the structure or total film thickness, the resulting spectrum increasingly deviates from the original design.[162]

In an effort to design a filter that is relatively thin but still serves its function well as a spectrum splitter, the general structure of the initial 16-layer design is maintained. Through an iterative apodization process of making small, randomly generated changes to individual layer thicknesses, the resulting reflection spectrum can be improved. To determine whether to keep a layer modification, an objective function f that is proportional to the efficiency of an $In_{0.49}Ga_{0.51}P$ solar cell for the spectrum it receives is first defined:

$$f = \frac{\int \alpha(\nu) I(\nu) \, \mathrm{d}\nu}{\int \nu I(\nu) \, \mathrm{d}\nu} \tag{6.6}$$

where $\alpha(\nu)$ is the absorption of the solar cell (taking values between 0 and 1) and $I(\nu)$ is the intensity of the light reflected from the dichroic, a function of the dichroic's reflectivity and the solar spectrum. For the purposes of optimization, $\alpha(\nu)$ is estimated by the absorption of a 2 μm thick $In_{0.49}Ga_{0.51}P$ layer with two-layer ZnS/MgF_2 ARC.[163] The internal quantum efficiency is assumed to be unity for all photons $> E_g$. For photons below E_g, $\alpha(\nu) = 0$, while for very high energy photons, the ν term in the denominator accounts for increased thermalization losses. Once the initial layer thicknesses and the objection function are defined, no further input is required. During the optimization process, the multilayer will depart from the exact quarter-wave design but only if the modification produces an improved spectrum for the $In_{0.49}Ga_{0.51}P$ cell. After the value of f appears to plateau after several successful iterations, the optimization process is terminated. The layer thicknesses in the initial and optimized structures are compared in Figure 6.4. The layers in contact with air are located at the left end of the diagram while the quartz substrate appears on the right.

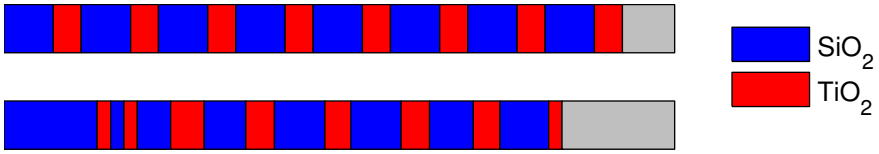


Figure 6.4: Relative layer thicknesses in the initial quarter-wave and optimized designs.

Several aspects of the dichroic filter's reflection spectrum change during the optimization process. The calculated reflection spectrum for the initial design is relatively symmetric above and below the stop band (slight asymmetry is due to dispersion and different media at the front and back of the multilayer). However, after several hundred rounds of iteration, a highly asymmetric spectrum develops for an optimized selection of layer thicknesses (see Figure 6.3). While for the majority of photon energies below E_g, the optimized filter's reflectivity decreases to 5% or below, the expected reflectivity increases for photon energies above 2.5 eV. Because the optimized layer thicknesses shift away from the exact quarter-wave design, the reflectivity in the stop band is slightly reduced. Additionally, it is noted that the center of the stop band moves to a slightly higher photon energy.

The energy of any photon below E_g that is sent to the photovoltaic cell will be entirely wasted. In order to maximize the limited spectrum cell efficiency, it is crucial that all sub-E_g photons are sent to the thermal receiver. In contrast, any photon greater than E_g can still contribute to photovoltaic energy conversion. Thus, during the optimization, if a proposed layer thickness change trades one sub-E_g photon for one photon with $h\nu \geq 2E_g$, the algorithm is likely to accept the change. Since the spectral energy density of the AM1.5D spectrum is greater below the stop band, a 1% decrease in reflectivity below E_g has a significantly greater effect on total device efficiency than the same decrease in reflectivity > 0.6 eV above E_g. The noticeable shift of the stop band to higher photon energies is due to the finite change in reflectivity as a function of photon energy at the stop band's edge. Higher refractive index contrast and an increase in the number of layers would reduce the photon energy range of the transition. Because the number of layers and the material refractive indexes are held constant during optimization, the simulation instead shifts the stop band to minimize the number of reflected below-E_g photons that fall within the transitional region.

While this randomized process does not guarantee finding the global optimized selection of layer thicknesses given the constraints of the number of layers and materials available, it successfully

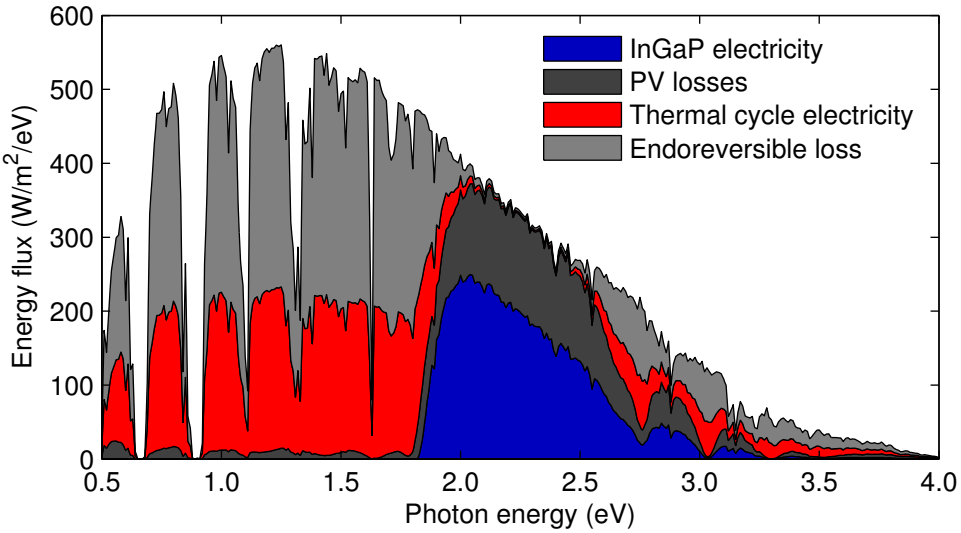


Figure 6.5: AM1.5D spectrum split between InGaP cells and a thermal receiver using the optimized dichroic mirror design. Thermal cycle and PV losses at 500× concentration and $T_{pv} = 60°C$ are also indicated.

removes the majority of reflection for sub-E_g photons while maintaining high reflectivity for photons just above E_g, improving the efficiency of the InGaP solar cells for the spectrum they receive. With a 16-layer design of alternating layers of SiO_2 and TiO_2 (for a total thickness of 1.24 μm) deposited on a quartz substrate, the idealized estimated efficiency of the $In_{0.49}Ga_{0.51}P$ cell for the spectrum it receives is 50% at 1× concentration and $T = 25°C$ (presuming a sufficiently designed ARC is applied to the quartz substrate backside). At a concentration of 500× and $T = 60°C$, the limited-spectrum efficiency of the cells increases to 56%. Assuming $\eta_{th} = 40\%$ for the thermal cycle, the total ideal expected system efficiency is 45%, of which electricity from the photovoltaic cells and the thermal cycle constitute fractions: $x_{pv} = 0.39$ and $x_{th} = 0.61$. Splitting of the AM1.5D spectrum and the energy contributions from the photovoltaic cells and the thermal cycle are plotted as a function of photon energy in Figure 6.5. While this ideal system efficiency does not drastically surpass the efficiency expected from a CPV system alone (after practical losses are considered), more than half of the electricity generated can temporarily be stored as thermal energy, mitigating the intermittent aspect of the original energy source. The hybrid CPV-CSP system described above has the advantages of both thermal storage and photovoltaic efficiencies potentially above 50% that cannot be matched by any current CPV system.

6.2 Structural color

While the objective function guiding the optimization of a DBR design described in Section 6.1.3 was specific to the operation of a photovoltaic cell, the method of design can be generalized for any application that can be quantitatively described. The following section describes the design of a DBR with the purpose of imitating the visual appearance of saturated pigments. By describing color numerically, objective functions specific to these artificial pigments are established to aid in their design.

6.2.1 Background

Pigments can be organized into two general categories based on the relevant method of color generation: by absorption and by structure. Most color is generated by selective absorption of visible wavelengths, determined by the intrinsic chemical properties of the material, whether they be delocalized orbitals of carbon rings in aryl azo compounds[164] or the ions in inorganic metal oxides.[165] Structural color pigments instead rely on wavelength specific interference, diffraction, and scattering effects due to periodic refractive index variation on the order of visible wavelengths within the pigment structure. Examples of naturally occurring structural color pigments include the well-known blue Morpho butterfly[166] as well as small-scale structures found in other insects, fish, birds, and even algae.[167] Artificial structural color pigments have recently been manufactured for cosmetic and paint applications, specifically to achieve iridescent effects (angular dependent color) that cannot be reproduced by traditional dyes and pigments.[168] Most structural colors used in cosmetics are based on the interference effect from single films of TiO_2 and iron oxide thin films deposited on mica flakes using a controlled precipitation fabrication method.[168,169] However, these synthetic structures do not readily replicate the visual appearance of organic pigments, specifically, a diffuse, color saturated reflectance independent of viewing angle.

Artificial structural color pigments may also serve as potential substitutes for organic pigments that become unavailable to consumers. One of the most widely used absorptive pigments in cosmetics, Carmine red, or "Natural red 4," is an extract from cochineal bugs found in South America. Now known to cause severe allergic reactions, products containing Carmine red must be labeled as "cochineal" in the United States and Europe, potentially reducing the pigment's

marketability. Additionally, its recent limited supply has caused significant price increases,[170] prompting a search for alternatives. Azo dyes such as Tartrazine, or "Yellow 5," have also been shown to cause allergic and intolerance reactions, and until recently, had been banned in several European countries.[171] If photonic crystals composed of stable and nonhazardous materials can be designed to generate high color saturation, low angular dependence, and diffuse reflectivity, many potential color applications become available, including paints, plastics, printing inks,[172] cosmetics, decorative coatings, sensors,[173] and color displays.[174]

Recent demonstrations of artificial structural colors have been conducted using DBR structures deposited on planar substrates. Yasuda *et al.* fabricated a multilayer film composed of a seven layer SiO_2/TiO_2 stack deposited on a sacrificial substrate by the sol-gel spin-on technique, achieving peak reflectivity of above 75% at the intended design wavelength.[175] After depositing the multilayer stack, the films were particularized into individual pigment particles by removing the sacrificial layer between the substrate and the multilayer. Because both layer materials are transparent in the visible spectrum, visible light that transmits through one particle will partially reflect off a particle behind it. While the reflectivity in the green ($\approx$ 550 nm) is below 20% for a single flake designed to appear red, the reflectivity from a sample of multiple flakes overlaid on top of each other corresponds to an unsaturated color due to increased effective reflection for wavelengths primarily transmitted by an individual pigment platelet.

In order to achieve color with higher saturation and lower angular dependence, Banerjee and Zhang incorporated higher index materials and more layers, fabricating a 31-layer TiO_2/HfO_2 ($n \approx 2.0$) stack by using electron beam evaporation to deposit the multilayer.[176] To achieve saturated color, a very thick stack (2.3 μm) was required due to the small refractive index contrast between TiO_2 and HfO_2. Theoretical modeling only considered the angular dependence of the central wavelength and the wavelength range of high reflectivity, not the actual shift in color as the human eye would perceive. The fabricated structure was never removed from the substrate used for deposition, preventing evaluation of the effective reflectance from the particularized film.

In order to properly design an artificial pigment, color must first be described numerically. Color is typically defined using three-dimensional color spaces for which standards are set by the Commission Internationale de l'Eclairage (CIE).[177] To convert a reflection spectrum to a point in a color space, three experimentally determined observer functions, $\bar{x}$, $\bar{y}$, and $\bar{z}$, are used. Each observer

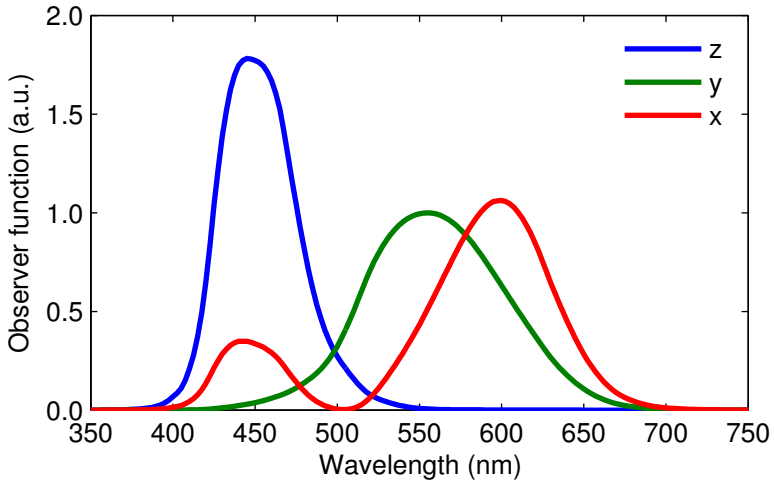


Figure 6.6: Observer functions used to calculate the tristimulus values X, Y, Z from a spectrum.

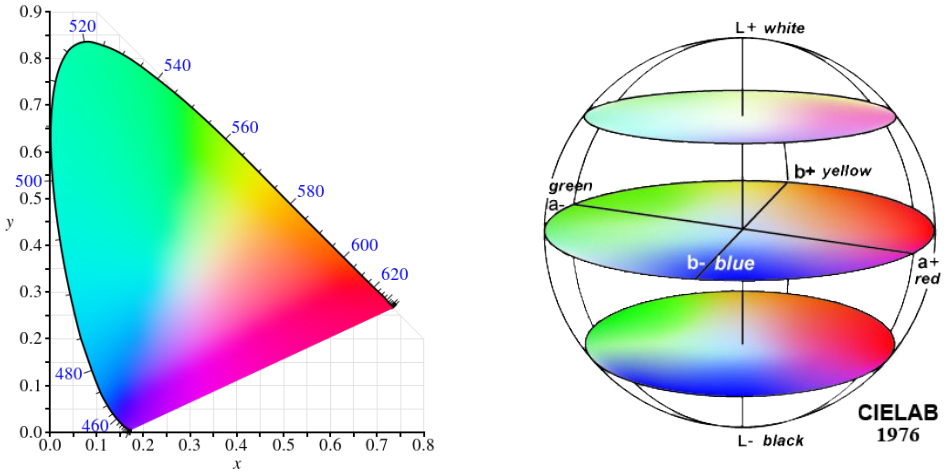
function is defined such that the effect on the perceived color from each wavelength in the spectrum is decomposed by the equivalent effects of monochromatic red (700 nm), green (546.1 nm), and blue (435.8 nm) light sources used in color-matching experiments[178] after some modification.[179] Tristimulus values, X, Y, and Z, are determined by taking the inner product of the observer functions (see Figure 6.6) with the reflection spectrum, $R(\lambda)$, and the illumination spectrum, $I(\lambda)$:

$$X = k\int \bar{x}(\lambda)I(\lambda)R(\lambda)\ \mathrm{d}\lambda \qquad Y = k\int \bar{y}(\lambda)I(\lambda)R(\lambda)\ \mathrm{d}\lambda \qquad Z = k\int \bar{z}(\lambda)I(\lambda)R(\lambda)\ \mathrm{d}\lambda \quad (6.7)$$

where the illumination normalization factor $k = 100\left(\int \bar{y}(\lambda)I(\lambda)\ \mathrm{d}\lambda\right)^{-1}$. The typical illumination condition employed is the "D65" spectrum which simulates sunlights at noon (based on the spectrum of a blackbody at 6500 K). (X, Y, Z) points can be further normalized by the brightness level, Y, to transfer to the xyY color space. A slice in the xyY space for a particular brightness value is shown in Figure 6.7a. The relationship between a particular distance between points in this colorspace to the difference in perceived color greatly depends on position, an aspect that is not ideal when attempting to compare colors numerically.

In order to create a more linear colorspace in which numerical color comparisons better represent optically perceived differences, the CIE:xyY space is converted to the CIE:Lab color space by the following set of equations:

$$L^* = 116f(Y/Y_\mathrm{n}) - 16 \quad (6.8a)$$



(a) (b)

Figure 6.7: Horizontal slices in the (a) CIE-xyY and (b) CIE-Lab color spaces. Values listed on the perimeter in (a) are the corresponding wavelengths (in nm) of monochromatic light.

$$a^* = 500\left[f\left(X/X_\mathrm{n}\right) - f\left(Y/Y_\mathrm{n}\right)\right] \tag{6.8b}$$

$$b^* = 200\left[f\left(Y/Y_\mathrm{n}\right) - f\left(Z/Z_\mathrm{n}\right)\right] \tag{6.8c}$$

$$f(t) = \begin{cases} t^{1/3} & \text{for } t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29} & \text{otherwise} \end{cases} \tag{6.8d}$$

where (X_n, Y_n, Z_n) is the white point defined by the illumination condition. In the CIE:Lab space, L is a measure of brightness, while a measures the relative amount of magenta and green, and b measures the relative amount of yellow and blue. Several horizontal slices through the CIE:Lab colorspace are shown in Figure 6.7b. A spectrum's a and b coordinates can alternatively be described by C, the chroma or saturation, and h, the hue angle:

$$C = \sqrt{a^2 + b^2} \qquad h = \arctan\left(b/a\right) \tag{6.9}$$

Because the CIE:Lab color space is designed to be linear with color, the Euclidean distance ΔE can serve as an effective measure of the extent an observer can distinguish two colors:

$$\Delta E = \sqrt{(\Delta L)^2 + (\Delta a)^2 + (\Delta b)^2} \tag{6.10}$$

Two colors are assumed to be indistinguishable to the human eye if $\Delta E < 0.7$.

6.2.2 Design

For the design of structural color pigments, minimization of ΔE between the target color and the designed color is the natural choice for the objective function during the DBR design process (assuming a target is specified). However, before picking a target color, design rules are first established for achieving saturated color in general. According to Eq. 6.9, highly saturated colors are generated when the magnitudes of a and/or b are large. Using the relationships specified in Eq. 6.8, the effects of a spectrum on a and b as a function of wavelength are shown in Figure 6.8 for a uniform light source with a reference white point $X_n = Y_n = Z_n = 100$. To create a saturated magenta ($a \gg 0$), reflectivity must be maximized in the ranges of $\lambda < 475$ nm and $\lambda > 575$ nm while minimized elsewhere. A saturated yellow ($b \gg 0$) is obtained with high reflectivity for $\lambda > 500$ nm. The requirements for a saturated red are a bit more involved: $a \gg 0$, $b \gg 0$, and $b/a \approx 0.8$. This combination of constraints requires high reflectivity similar to that for magenta with the added caveat that $a > b$. While $\lambda > 575$ nm photons will contribute to $+a$, the contribution to b is always greater, creating a trade-off between color saturation and the red hue. Unexpectedly, a saturated red is only possible with some reflectivity in the blue part of the visible spectrum as these wavelengths will increase a while simultaneously reduce b.

In a reflection spectrum corresponding to high color saturation, the transition between high and low reflectivity regions must be as sharp as design allows, similar to the requirements for a dichroic filter. Therefore, for a given incidence angle, color saturation will be maximized for a DBR

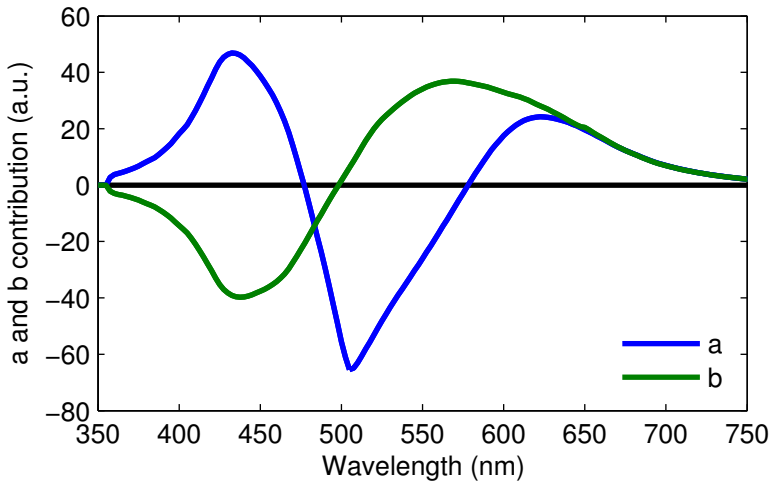


Figure 6.8: Relative contribution to a and b as a function of wavelength for a uniform light source.

composed of many layers with a large Δn. A 17-layer structure containing alternating layers of SiO_2 and TiO_2 is initially chosen. In order to minimize impedance mismatch at the DBR's outer surfaces, the lower index SiO_2 layers are chosen for the first and last layers of the structure. Unlike the design of the dichroic mirror, the artificial pigment DBR design must be symmetric. While the structure may be initially fabricated on a substrate, the end product is a free standing, substrate-less particle, just as likely to orient on one side as it is on the other when applied to a surface. Assuming normal incidence light and air as the surrounding medium, the layers are modified in a similar process used for the dichroic filter with the only modification being the objective function.

For the purposes of demonstration, an initial target color of Lab = [53, 83, 67] is chosen, corresponding to a saturated red. A stop band center wavelength of λ_0 = 690 nm defines the initial film thicknesses. During each iteration of the DBR structure, ΔE is calculated between the target color and color of the current design. When a lower ΔE is obtained, the recent modification to the layer thicknesses is kept. After several successful iterations, the Lab value of the design closely matched the target. The initial and final film thickness are compared in Figure 6.9 while the reflection spectrum and expected color are shown in Figure 6.10. Simply by specifying the objective function ΔE and providing the simulation with the tools to calculate Lab, the simulation automatically chooses a structure which not only produces a sharp edge just above 575 nm but maintains near zero reflectivity between 450 nm and 575 nm and allows for some reflection in the blue, the three requirements mentioned earlier in the discussion on obtaining a saturated red.

The reflection spectrum of a DBR is dependent on the angle of illumination because the phase shift due to propagation through each layer is proportional to $\cos\theta_i$ (see Eq. 6.3a). At greater incidence angles, the effective thickness of each layer is reduced, causing the stop band to shift toward smaller wavelengths. While θ_i in high refractive index materials will be less sensitive to the incidence angle from air (due to Snell's law), materials with low refractive indexes are less

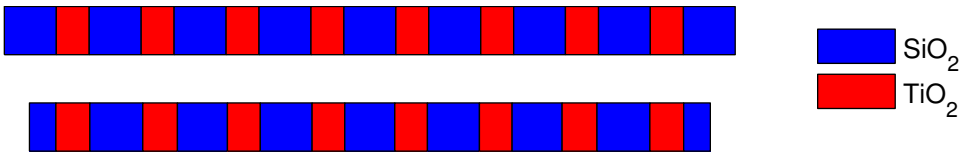


Figure 6.9: Comparison of the relative film thicknesses of the DBR designed for a saturated red before and after optimization.

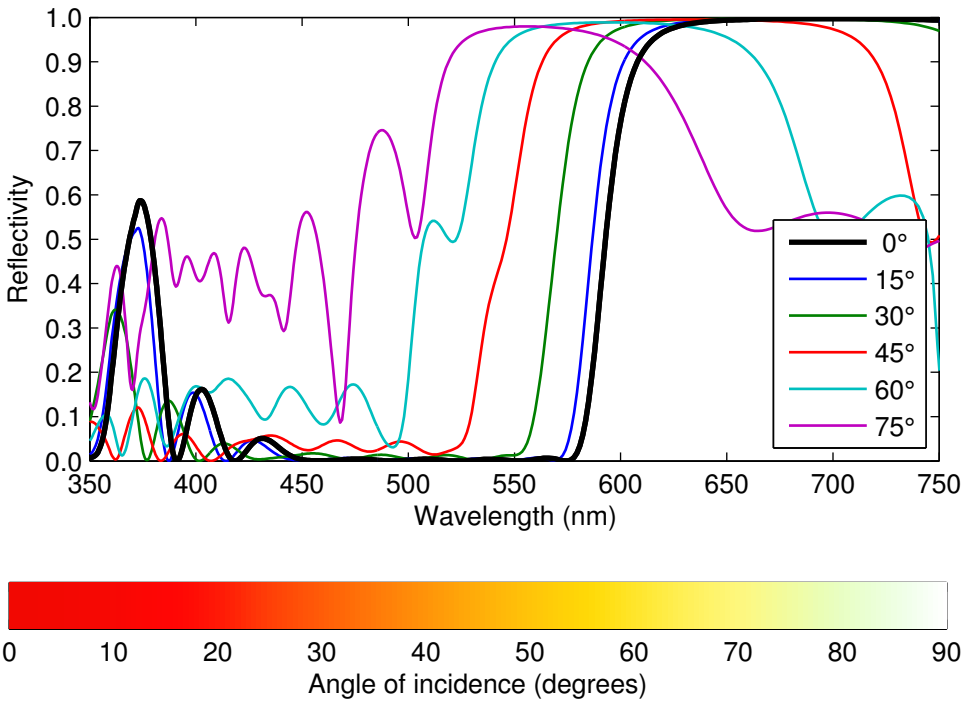


Figure 6.10: Reflection spectra at different angles of incidence for a SiO_2/TiO_2 DBR designed to reflect saturated red at normal incidence. The expected reflected color of the DBR as a function of incidence angle is also shown.

immune. If the DBR contains a low refractive index material such as SiO_2, the perceived color will be a strong function of the incidence angle of light. Figure 6.10 illustrates the angular dependent reflection spectrum and expected color from unpolarized light incident on the SiO_2/TiO_2 DBR designed to reflect a saturated red for normal incidence (Lab = [50.3, 82.5, 67.0]). As θ_i increases, the stop band progressively moves further into the visible spectrum. For $\theta_i > 25°$, the expected color has noticeably changed from a red hue.

The highly saturated red at normal incidence is easily designed with only 17 layers due to the high index contrast of the material combination used. To achieve such high contrast, however, the design resorts to using one layer material with the relatively low $n \approx 1.45$. Thus, there is a trade-off between obtaining a highly saturated color and an angular-independent color. Both materials in the DBR must have relatively large refractive indexes (e.g. $n > 2.0$) but simultaneously require moderate contrast between them. Ideally, both high index contrast and average high index would be possible if materials with high refractive indexes, e.g. $n > 3.0$, such as Ge, GaAs, and chalcogenides did not

also absorb greatly in the visible spectrum. Many high index materials are also toxic, preventing their use in a consumer product.

TiO_2 and SiC are two materials with $n \approx 2.4$ with minimal absorption for wavelengths above 400 nm depending on how they are deposited. While the refractive index of Si_3N_4 is adjustable by changing the Si:N stoichiometry during film growth,[180] it must be limited to $n \approx 1.95$ to remain transparent within the visible wavelength range. By replacing SiO_2 layers with Si_3N_4, the reflection spectrum will shift less with changes in incidence angle. Due to the diminished Δn, however, a saturated red can no longer be as easily created. Replicating a hue of Carmine, $h = +22°$, is a more attainable target for a Si_3N_4/TiO_2 DBR. In the optimization process, multiple objective functions can be defined. In order to find the greatest saturation at the specified hue h, the iterative process is modified so that it only accepts a thickness modification if it both increases color saturation and simultaneously reduces Δh from the target unless Δh is within an arbitrarily chosen tolerance of 3°. Due to the lower Δn, the DBR must be composed of 25 layers to obtain sufficient reflectivity in the stop band. The reflection spectra of the optimized structure and the associated colors as a function of incidence angle are shown in Figure 6.11. While the angular shift of the reflection spectrum is reduced as expected, the bandwidth of the DBR's stop band is also diminished due to the lower Δn. Unlike for the SiO_2/TiO_2 DBR, the high wavelength edge of the stop band falls in the visible portion of the spectrum, and increasingly so at greater values of θ_i. With simultaneously reduced reflection above 600 nm as reflection increases below 575 nm at larger values of θ_i, the effective angular dependence of the observed color is similar to that observed in Figure 6.10 despite the higher mean refractive index of the DBR. Without materials with refractive indexes greater than 2.4, DBR designs will fail to produce angular independent saturated reflectors.

One material that with a very high refractive index ($n_{vis} > 3.5$) and absorption length greater than 1 μm for $\lambda > 500$ nm is single-crystalline Si. The gradual absorption edge due to Si's indirect band gap can effectively be sharpened using a DBR design, creating a saturated red reflector. By placing the stop band in the red portion of the visible spectrum, the effective thickness of the Si layers for longer wavelengths is reduced while shorter wavelengths will transmit through the entire structure and will more strongly be absorbed in the process. While c-Si cannot be deposited in a multilayer with amorphous films, poly-Si and a-Si are potential alternatives, depending on the

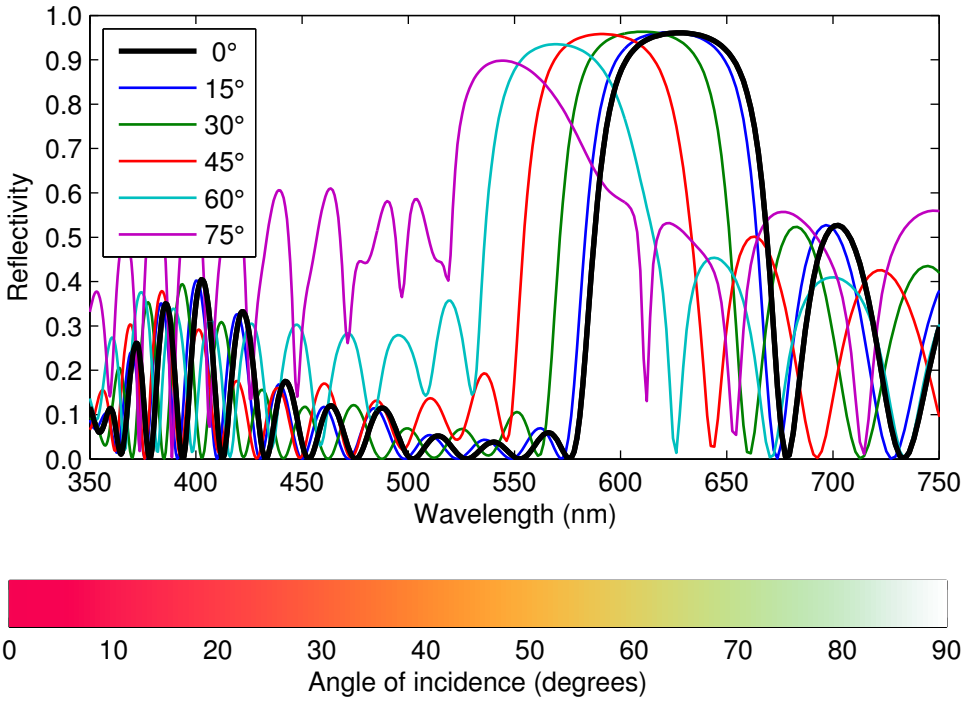


Figure 6.11: Reflection spectra and expected reflector color at different angles of incidence for a Si_3N_4/TiO_2 DBR designed to reflect a saturated carmine hue at normal incidence.

available deposition methods. The absorption tail in the visible spectrum for either alternative form of Si is dependent on how the thin film is deposited and post-processed.

For the purposes of simulation, the measured complex refractive index of boron-doped a-Si deposited by PECVD is chosen from literature.[181] Because of the high index of a-Si, TiO_2 can be used as the "low" index material in a 13-layer structure. The outer two TiO_2 layers of the DBR are replaced with layers of Si_3N_4 in order to reduce impedance mismatch at the film/air interfaces, and thus reduce the side bands below 575 nm for $\theta_i < 45°$. To guide the optimization of this structure, the simulation sought to maximize saturation for a red hue ($h = +38°$) while maintaining the same tolerance of 3° for Δh once it is reached during the iterative process. The optimized structure's layer thicknesses as well as its simulated reflection spectra and corresponding colors for different incidence angles are shown in Figure 6.12. By incorporating a-Si into the 700 nm thick DBR, angular dependence of the observed color is greatly reduced compared to the previously presented designs. The resulting averaged spectrum generated by a collection of pseudo-randomly oriented particles is expected to maintain its saturation to a greater extent for designs containing layers of a-Si or other weakly absorbing films such as iron oxide.

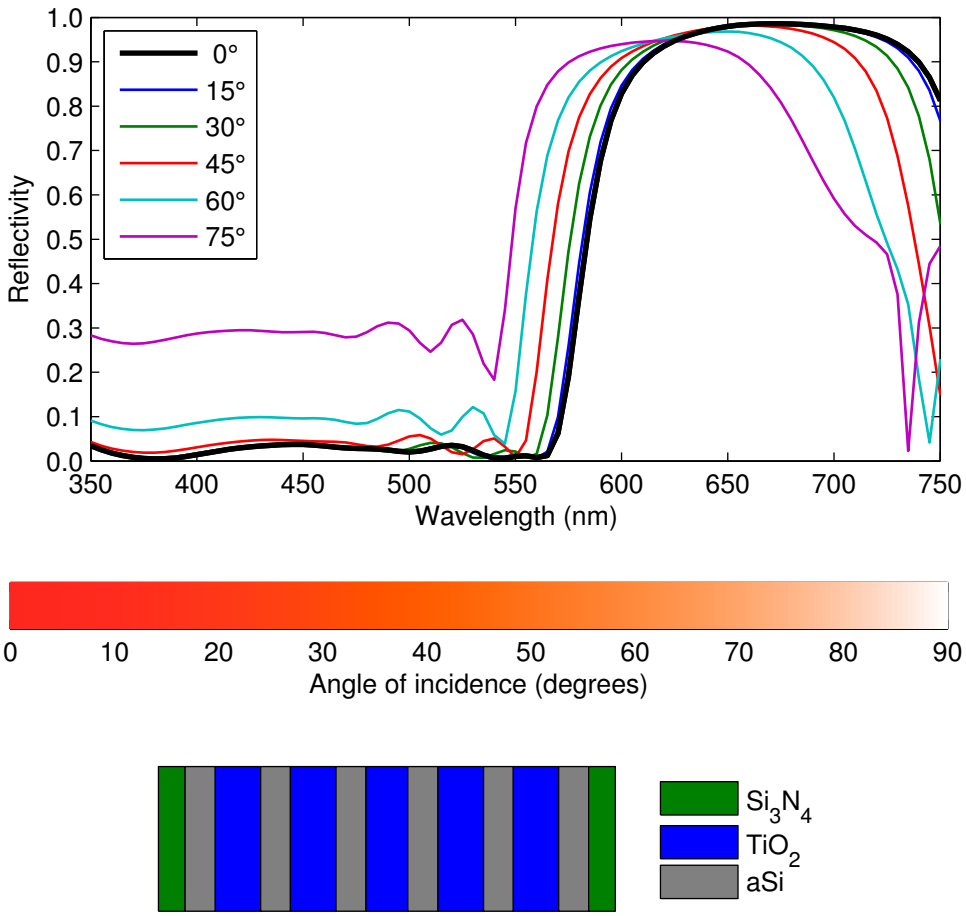


Figure 6.12: Reflection spectra and expected reflected color at different angles of incidence for a Si_3N_4/TiO_2/a-Si DBR designed to reflect a saturated red hue at normal incidence. Relative layer thicknesses of the optimized DBR structure are also shown.

6.2.3 Fabrication

The expected color saturation and minimal angular dependence of a DBR composed of alternating high refractive index materials that still maintain high contrast is validated by fabrication of multilayer structures by PECVD. In order to deposit a multilayer that could easily be removed from the substrate on which it was deposited, a 1 μm thick sacrificial SiO_2 film was first deposited on a Si substrate, followed by the alternating layers of the DBR. After deposition, the substrate was placed in a solution of 5% HF to etch away the SiO_2 film between the substrate and the multilayer. Because the removal of the SiO_2 layer relied on lateral etching, samples were required to be submerged in solution for several hours. Once the SiO_2 was removed, the multilayer would

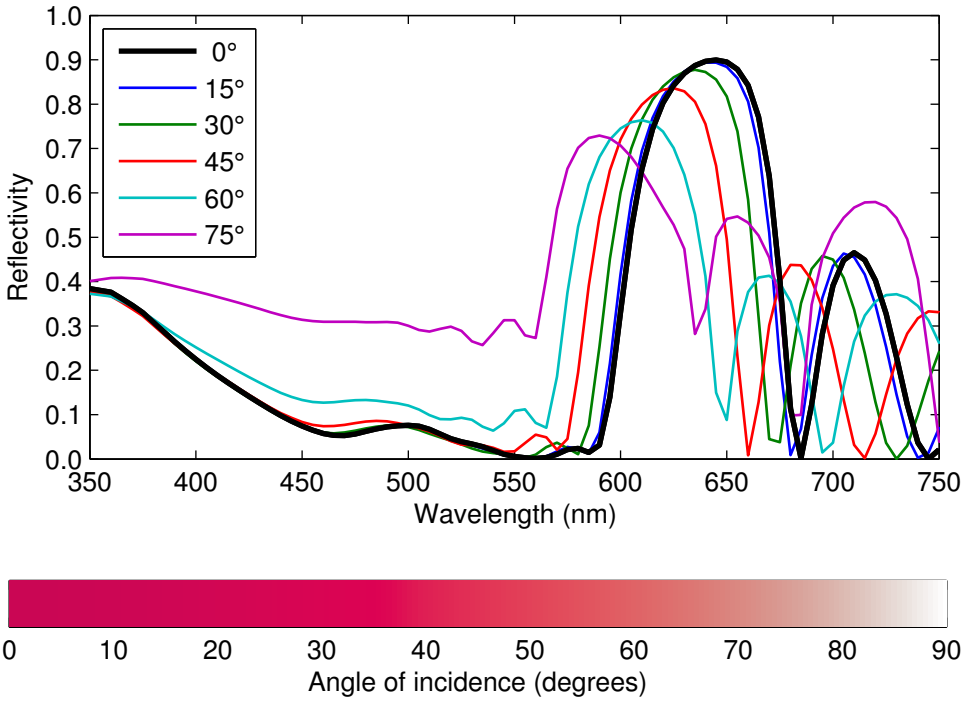


Figure 6.13: Reflection spectra and expected reflected color at different angles of incidence for a SiC/a-Si DBR designed to reflect a saturated carmine hue at normal incidence.

detach from the substrate after immersion in a water bath. After successive water baths to remove remnant HF, the multilayer film was collected by filtration and particularized by ultrasonication.

Si-poor PECVD Si_3N_4 (low Si-content would be required for transparent films) would be etched away in HF solutions.[82] While one promising demonstration of an artificial pigment would contain Si_3N_4, TiO_2, and a-Si, Si_3N_4 layers could not be included for reasons of chemical compatibility during the long lift-off process. Additionally, because the available materials to deposit by the PECVD technique were limited to SiO_2, Si_3N_4, SiC, and a-Si, the role of TiO_2 as the "low" index material was given to SiC due to the two materials' similar refractive indexes. Therefore, an 11-layer SiC/a-Si multilayer was chosen for demonstration.

While an initial quarter-wave design will maximize constructive interference for reflection, the same favorable π round-trip phase shift can be created by any odd quarter-wave film thickness. By beginning with three quarter-wave thicknesses for a-Si layers while retaining single quarter-wave thicknesses for SiC, a stop band centered at the design wavelength of 645 nm can still be created. The expected reflection spectra and observed color for the design is displayed in Figure 6.13. The benefit of the increased thickness of a-Si in the DBR is primarily for the end use of the multilayer:

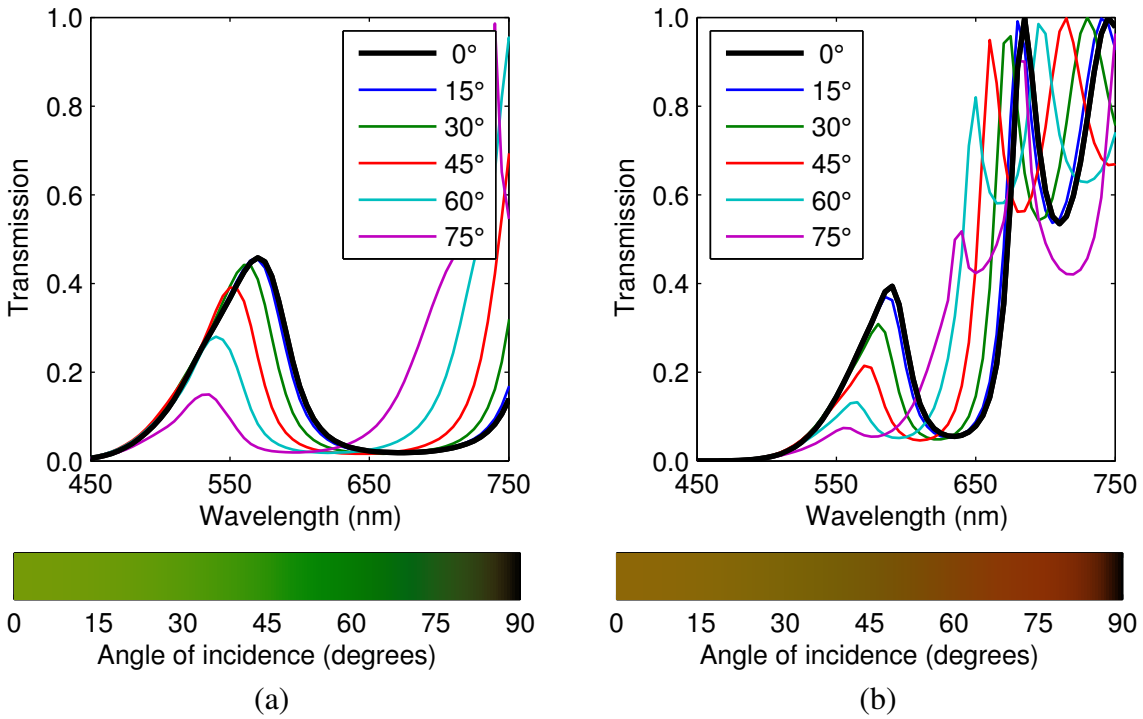


Figure 6.14: Transmission spectra and expected transmitted color at different angles of incidence for the (a) Si_3N_4/TiO_2/a-Si and (b) SiC/a-Si multilayer designs.

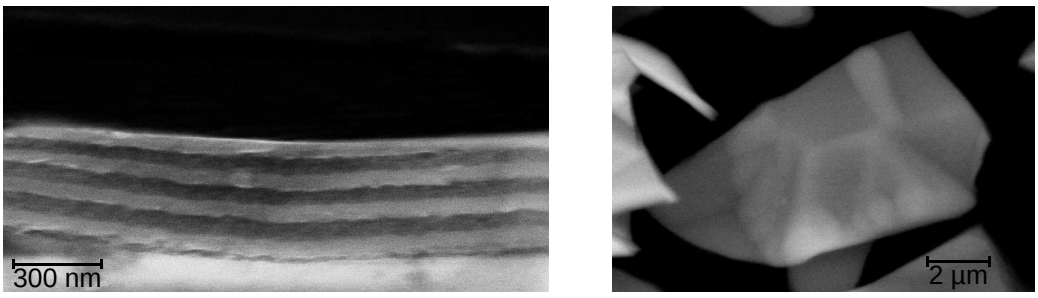
multiple platelets laid on top of each other. While the reflection spectrum of the multilayer is in no doubt important, transmission through the multilayer cannot be ignored. Any light that transmits through one pigment particle will travel to the next platelet below it. Unless reflectivity is zero for most incidence angles, after several layers of particles, a significant portion of light primarily transmitted by a single multilayer will be collectively reflected by the platelets as a whole. Including a-Si not only provides high Δn and minimal angular dependence but also reduces transmission of unwanted light through the multilayer. While the design containing single quarter wave thicknesses for a-Si layers with reflection spectra shown in Figure 6.12 significantly transmits wavelengths corresponding to green, transmission through the multilayer with the three-quarter wave thick a-Si layers is significantly reduced at wavelengths that will desaturate the reflected carmine hue. Figure 6.14 compares the two sets of spectra and expected colors of transmitted light.

Figure 6.15a shows the visual appearance of the SiC/a-Si DBR after lift-off and particularization by ultrasonication. While the saturated carmine hue is achieved, the artificial pigment lacks the diffuse reflection typical of organic pigments. The initial batches of the SiC/a-Si multilayers were deposited on the polished side of Si substrates. By depositing the same multilayers on the

(a) (b)

Figure 6.15: Optical images of particularized DBR multilayers deposited on (a) specular and (b) unpolished Si substrate surfaces.

unpolished, backside of the substrate, the specular reflection of the multilayer pigments can be significantly reduced as shown in Figure 6.15b. Because the multilayer deposition on the substrate is conformal, the slight roughness of the substrate will translate to roughness in the multilayer that deposits. Multilayers deposited on the backside of a Si substrate are shown before lift-off and after particularization in Figure 6.16. The effect of roughness of the Si substrate can clearly be seen in the surface of the particularized pigment.



(a) (b)

Figure 6.16: (a) SEM cross-section of a SiC/a-Si PECVD multilayer deposited on the unpolished side of a Si substrate covered. (b) SEM top view of a freestanding, particularized multilayer.

6.3 Summary

This chapter focused on the application of one-dimensional photonic crystals to provide design solutions for two separate applications that, at first glance, may have initially seemed unrelated. However, both solar energy conversion and artificial pigments require careful consideration of a broad range of photon energies (or wavelengths) in optical design to meet targeted design goals. In both applications, a sharp edge of the stop band of maximum reflectivity was required along with ranges of near zero reflectivity, leading to the choice of apodized quarter-wave designs for both applications. Complete transparency across the entire spectrum was required for dichroic mirrors while in the case of structural color pigments, selective absorption was identified as necessary for success. The approach of defining an objective function, whether it depends on the efficiency of a single junction solar cell or observer functions mapping a spectrum to a point in a color space, to reach a design target through iteration is general and can be applied to any other broadband spectrum application.

Chapter 7

Summary and Future work

7.1 Summary

The focus of this thesis revolved primarily around the identification of solutions to create a large area, thin Ge film on Si of minimal TDD to serve as a virtual substrate for high efficiency Ge lattice matched III-V solar cells. Discussion of the fabrication of large-area Ge mesa array phovoltaic cells in Chapter 3 identified the shortcomings of selectively grown Ge films specific to photovoltaic devices which are several orders of magnitude larger than the dimensions required for effective defect reduction by selective growth. In order to create a more cohesive understanding of the Ge film patterning parameters that determine the final threading dislocation after annealing, simulations in Chapter 4 highlighted potential problems that may arise during dislocation glide in Ge films previously grown selectively in SiO_2 trenches. Elimination of Ge film facets, inversion of the SiO_2 sidewall angle and changing the sidewall material to Ge were identified as promising methods to retain uniform resolved shear stresses on threading dislocations during cyclic annealing. The latter half of Chapter 4 identified the mechanism of dislocation pinning during glide and provided estimates through Monte Carlo simulations of the effects of the film thickness, mesa width, and applied ΔT during cyclic annealing on the final dislocation density.

Chapter 5 explored various aspects of trench fill, lateral overgrowth, and coalescence. While the basic structure of Ge mesas defined by a regular grid were not observed to coalesce within reasonable growth times for the growth method and conditions employed, overgrowth within staggered grids and surrounded isolated SiO_2 lines exhibited complete coalescence. Defect etch pits observed on coalesced Ge films provided support of the theoretical results presented on the detrimental effects of film faceting on dislocation pinning in Section 4.2. EBIC studies of $In_{0.49}Ga_{0.51}P$ cells grown on Ge surfaces of various TDD levels provided some preliminary data on the tolerance of $In_{0.49}Ga_{0.51}P$ device efficiency on the substrate's TDD. Chapter 6 explored one potential application of $In_{0.49}Ga_{0.51}P$ single junction cells for a high efficiency, CPV-CSP hybrid energy conversion

system that relied on spectrum splitting provided by a dichroic mirror. The photovoltaic component of such a system has the potential to see significant cost reductions if the photovoltaic cells for the system can be fabricated on Ge-on-Si virtual substrates. The design methodology developed for the dichroic mirror was also adapted for another broadband application concerning the visible spectrum, artificial structural color pigments.

7.2 Future Work

While Chapter 4 identified the ideal patterned structures that will maximize threading dislocation glide to the Ge film edge, these structures have not yet been thoroughly tested experimentally. All structures require that the Ge films be facet-free. Although Ge films generally facet when grown selectively, facet-free Ge can be obtained by growth at low temperature and high p_{GeH_4}.[66] If changing Ge CVD growth conditions is not an option, Ge mesas can be planarized by CMP after mesa growth but before cyclic annealing. To create trench structures with inverted sidewall angles ($\theta > 90°$) as would be required for SiO_2, Ge blanket films can instead be initially grown. Trenches can then be defined post-growth by Ge RIE and trench filling with SiO_2 by a CVD process. Combinations of patterned etch and CMP of the trench fill layer will produce the final trench structure to be cyclically annealed and subsequently, laterally overgrown by Ge after defect reduction by dislocation glide to dislocation sinks.

For trenches composed of c-Si coated with a thin layer of SiO_2 deposited on top of the Si to prevent direct deposition on top of the trench, facet-free Ge can automatically be obtained in Si trenches without modifying Ge growth process parameters. Because Ge deposits directly on Si, the Si trench sidewalls will immediately be coated with Ge. Due to the concave exposed Si surface (as observed in cross-section), the trench fill rate is determined by the fastest growing facet, the (001). The Ge mesa will only begin to develop {311} and {111} facets when the (001) facet surpasses the height of the Si trench. In the fabrication of a Si trench structure, processing is identical to that of a thin SiO_2 trench with an additional etch into the Si substrate immediately after the SiO_2 dry etch. The Si etch depth is on the order of 1 μm, conducted using a Cl_2/HBr etch chemistry (see Appendix A). To recover a fresh Si surface for epitaxy, the substrate undergoes a sacrificial oxidation and wet etch. Initial fabrication of Ge mesas grown in c-Si trenches is shown in Figure 7.1.

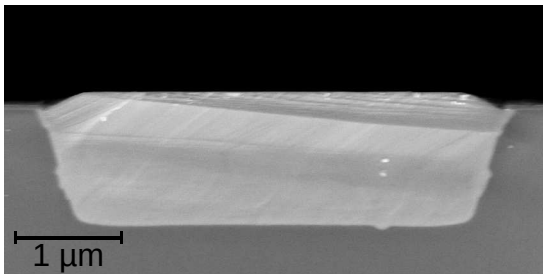
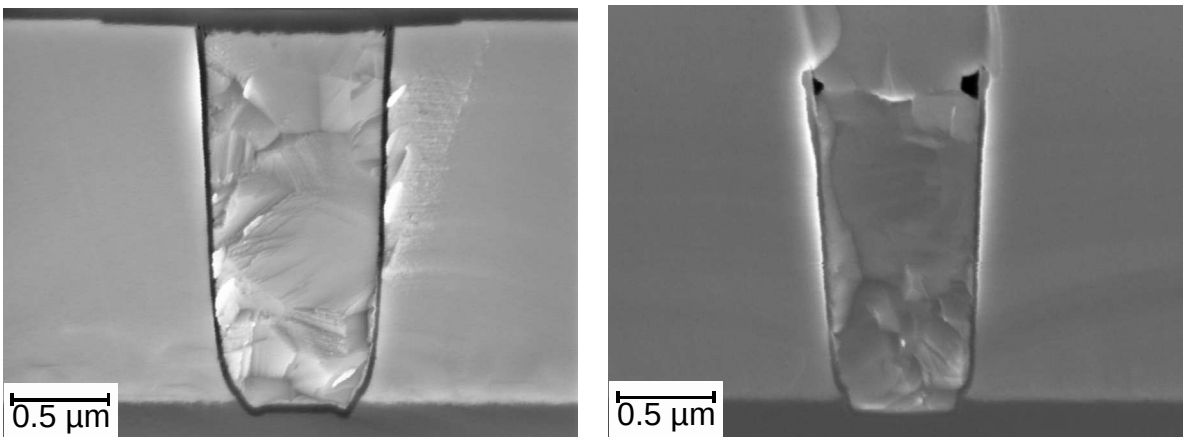


Figure 7.1: SEM cross-section of a Ge mesa selectively grown in a c-Si trench. A thin layer of SiO_2 covers the Si substrate, enabling selective growth.

There are however some concerns for the effectiveness of a Si trench structure. Because the trench is composed of c-Si, when dislocations glide to the Ge/Si trench sidewall, they will not immediately terminate due to the continuation of the crystal lattice. For this structure to be successful at dislocation reduction, dislocations must be directed to stop gliding once they glide into the Si trench and have their termination point at the interface between the SiO_2 layer and the Si trench or Ge film underneath. It also remains to be observed experimentally whether the slightly repulsive image forces at the Ge/Si trench interface will prevent complete dislocation glide out of the Ge film. It should be noted, however, that for some combinations of slip systems and the specific dislocation line direction, the self-stress force will be directed toward the film edge, sometimes overtaking the repulsive image force, creating a favorable condition for threading dislocation reduction.

While etched trenches filled with Ge coated with a thin layer of SiO_2 on all sides is the most promising structure based on the analysis presented in Chapter 4, the process flow is more involved compared to previous ones described. A general process flow is listed below:

1. Blanket growth of Ge
2. CVD of SiO_2 to protect the Ge surface during patterning
3. SiO_2 and Ge dry etching using an inverted trench pattern
4. Wet etch removal of the temporary SiO_2 film
5. CVD of SiO_2 and a thin a-Si layer for nucleating poly-Ge
6. a-Si patterning to restrict poly-Ge growth to trenches (optional)
7. CVD of poly-Ge on a-Si to fill the trench
8. CMP of poly-Ge
9. Etch of remaining a-Si and/or partial etch of poly-Ge (optional)



(a) (b)

Figure 7.2: SEM cross-sections of (a) a poly-Ge trench after SiO_2 encapsulation and (b) a poly-Ge trench without encapsulation after ELO.

10. CVD of SiO_2 to cap the Ge trench (optional)
11. CMP or patterned etch of the capping SiO_2 layer for planarization (optional)
12. SiO_2 wet etch to expose the single-crystal Ge surfaces but not the poly-Ge (optional)
13. Cyclic anneal and Ge ELO

Initial fabrication of the structure described above employed permanent layers of 100 nm thick PECVD SiO_2 and 50 nm thick PECVD a-Si. Due to the high sticking coefficient of SiH_4 at the PECVD process conditions, the resulting film thicknesses are reduced by approximately a factor of 2 when deposited in the trenches (with an aspect ratio > 1) etched into the blanket Ge film. Figure 7.2a shows a SEM cross-section of the resulting structure after poly-Ge CMP and encapsulation by an additional 200 nm thick PECVD SiO_2 film. The poly-crystalline nature of the Ge deposited in the trench surrounded by SiO_2 on all four sides is clearly visible. SiO_2 encapsulation on the top surface may be avoided if lateral overgrowth from the single crystalline Ge outcompetes the vertical growth of poly-Ge from the trench. Figure 7.2b displays a SEM cross-section of a poly-Ge trench immediately surrounded by SiO_2 on three sides. The top surface is instead covered by single-crystalline Ge resulting from overgrowth over the 50 nm wide SiO_2 separation layer from both sides as suggested by the existence of voids at the upper corners of the poly-Ge filled trench.

Whether blanket films of Ge are grown, followed by trench etch and fill with SiO_2 and/or poly-Ge, or Ge is grown selectively in c-Si trenches coated with SiO_2 on the top surface, premature overgrowth above isolated trench lines will either not occur (in the case of initial blanket film

growth) or can easily be avoided (for growth in c-Si trenches). Therefore, isolated lines of minimal thickness can be used to define the edges of the Ge films to maximize throughput. For the creation of a Ge virtual substrate for III-V epitaxy from a vicinal (001) Si wafer miscut by 6°, isolated lines perpendicular to the axis of miscut rotation should remained aligned with the ⟨110⟩ projection on the substrate while lines parallel to the axis of rotation should be rotated by $5°-10°$. This arrangement will minimize the time of overgrowth for both sets of nominally orthogonal Ge film edges as inferred from Figure 5.20. The selection of isolated line edges in place of staggered grids is also expected to reduce the number of defects reintroduced by coalescence between adjacent Ge film edges. While staggered grids consist of initially discontinuous Ge films which do not share a relaxed Ge lattice, Ge films surrounding isolated lines do not suffer from this problem. If coalescence-induced defects are avoided altogether, the distance between isolated lines should be minimized until the non-ideal resolved shear stress conditions become appreciable near Ge film edges. Reduced distances between isolated lines decrease the glide distances required for threading dislocation removal, allowing reductions of the Ge film thickness to achieve the same TDD level. Additionally, the time required for overgrowth is lessened since the isolated lines' length scales with the distance between adjacent isolated lines.

By selecting one or more of the options described above for the materials surrounding patterned Ge films, the isolated line structure that optimizes the film coalescence process, and film dimensions corresponding to a sufficiently low predicted TDD, a thin Ge-on-Si virtual substrate nearly free of threading dislocations has the potential to be experimentally demonstrated.

Appendix A

Process recipes

This appendix includes the detailed process conditions for several fabrication steps referred to throughout the thesis.

A.1 Plasma processing

Dry etching

Silicon dioxide, silicon, and germanium are dry etched using a magnetically-enhanced reactive ion etching tool, Applied Material's Precision 5000. The process conditions are listed below for SiO_2, Si, and Ge etch recipes. An additional recipe, Si*, is included for Si etching that is highly selective to SiO_2. While the etch recipe produces an extremely rough Si surface, use of an SiO_2 layer etch stop leads to a very smooth surface once the Si layer is entirely removed due to the etch's high selectivity of 100:1. If a Si etch is not immediately proceeded by dry etching of a SiO_2 layer, a short "break-through" etch is conducted using CF_4 to remove any residual SiO_2 on the surface. Separate etching chambers are dedicated to F and Cl/Br gas chemistries.

Material	SiO_2	Si	Si*	Ge	
Etch rate	40	67	slow	110	Å/s
Pressure	250	100	100	100	mTorr
CF_4	36	0	0	0	sccm
CHF_3	18	0	0	0	sccm
Ar	100	0	0	0	sccm
Cl_2	0	60	0	40	sccm
HBr	0	40	40	40	sccm
RF power	600	400	75	400	W
Magnetic field	30	60	50	60	G

PECVD

Silicon dioxide and amorphous Si films are deposited by PECVD using the deposition tool, Applied Material's Centura 5200. The process conditions are listed below for SiO_2 deposition for thin (e.g. < 0.5 μm) and thick film thicknesses as well as for a-Si:

Material	SiO_2	SiO_2	a-Si	
Deposition rate	36	170	34	Å/s
Temperature	400	400	350	°C
Pressure	2.7	2.7	4.5	Torr
SiH_4	50	115	165	sccm
N_2O	800	2000	0	sccm
Ar	0	0	1450	sccm
RF power	255	275	180	W

A.2 UHV-CVD

Germanium films are grown by the UHV-CVD system: Leybold Sirius CVD-300. The three zone heater wall temperatures are set to achieve a uniform thermal profile across the wafer boat. A throttle valve is adjusted to maintain the specified chamber pressure. The following is a typical process recipe for two-step Ge growth on Si followed by an in-situ anneal without cycling:

Step	**GeH_4 (sccm)**	**H_2 (sccm)**	**Temperature (°C)**	**Pressure (mbar)**	**Time (min)**
Boat in	0	10	650	$< 10^{-4}$	5
Bake	0	5	800	2×10^{-3}	60
Ramp	0	5	350	2×10^{-3}	180
Buffer	10	0	350	1.2×10^{-2}	80
Ramp	0	5	730	2×10^{-3}	60
Growth	7.5	0	730	1.0×10^{-2}	240
Anneal	0	5	850	2×10^{-3}	80
Ramp	0	5	650	2×10^{-3}	50
Boat out	0	10	650	$< 10^{-4}$	5

The following is a process recipe for poly-Ge growth on a layer of PECVD a-Si. After loading the wafers, a residual gas analyzer (RGA) monitors the H_2 that outgasses from the PECVD films. Once the H_2 partial pressure stabilizes after 1 to 2 hours, poly-Ge growth is initiated. The chamber temperature is not increased above 450°C to maintain a continuous a-Si seed layer before and during poly-Ge growth. The growth rate of poly-Ge at 450°C is approximately 150 nm/hr.

Step	**GeH_4 (sccm)**	**H_2 (sccm)**	**Temperature (°C)**	**Pressure (mbar)**	**Time (min)**
Boat in	0	10	450	$< 10^{-4}$	5
Bake	0	5	450	2×10^{-3}	120
Growth	7.5	0	450	1.0×10^{-2}	300
Boat out	0	10	450	$< 10^{-4}$	5

A.3 Photolithography

Positive resist SPR-700

Spin coating and development is conducted by an automated wafer track system: SSI150 and exposed by a Nikon NSR2005i9C i-stepper. For 1 μm thick resist, the following recipe is followed:

Step	**Time (sec)**	**Conditions**
HMDS bake	60	130°C
Resist dispense	1	0 rpm
Spread	6	500 rpm
Spin	30	5000 rpm
Edge-bead removal	5	500 rpm
Final spin	15	1500 rpm
Soft bake	30	95°C
Exposure	~ 0.175	(per die)
Post-exposure bake	30	115°C
MF CD-26 dispense	60	100 rpm
DI water rinse	20	500 rpm
Spin dry	40	2500 rpm
Hard bake	60	120°C

Image reveral resist AZ5214E

Image reversal resist finds use in multilevel patterning process flows that begin with trenches etched into blanket Ge. Spin coating is conducted manually, while film baking, exposure, and development use the same equipment used for SPR-700. As received, the following recipe will create 1.6 μm thick photoresist films prior to initial baking. For 1.0 μm thick resist, AZ5214E solution is diluted with its solvent, PGMEA: 3 parts solution, 1 part solvent. Thinner resist films are required to resolve 0.5 μm wide lines of photoresist after patterning. Substrate heating between flood exposure and development must be minimized.

Step	**Time (sec)**	**Conditions**
Resist dispense	6	500 rpm
Spread	6	750 rpm
Spin	40	4000 rpm
Soft bake	60	95°C
Exposure (1.6 μm)	0.100	(per die)
Exposure (1.0 μm)	0.080	(per die)
Image-reversal bake	120	120°C
Flood exposure	0.600	(per die)
MF CD-26 dispense	60	100 rpm
DI water rinse	20	500 rpm
Spin dry	40	2500 rpm
Hard bake	60	120°C

Bibliography

[1] V. Smil, "Energy at the crossroads", http://ec.europa.eu/clima/consultations/docs/0005/registered/85246854723-63_confederation_of_industry_czech_republic_en.pdf (2006). Accessed: 2015-09-04.

[2] "BP statistical review of world energy", http://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html (2015). Accessed: 2015-09-04.

[3] T. A. Boden, G. Marland, and R. J. Andres, "Global, regional, and national fossil-fuel CO2 emissions", doi:10.3334/CDIAC/00001_V2013.

[4] G. S. Callendar, *Q.J.R. Metrological Society* **64**, 223 (1938). doi:10.1002/qj.49706427503.

[5] M. Rekinger, F. Thies, G. Masson, and S. Orlandi, "Global market outlook for solar power, 2015-2019", http://www.solarpowereurope.org/. Accessed: 2015-09-07.

[6] "Sunny uplands: Alternative energy will no longer be alternative", The Economist (2012). http://www.economist.com/node/21566414/print.

[7] K. Masuko *et al.*, *IEEE Journal of Photovoltaics* **4**, 1433 (2014). doi:10.1109/JPHOTOV.2014.2352151.

[8] W. Shockley and H. J. Queisser, *Journal of Applied Physics* **32**, 510 (1961). doi:10.1063/1.1736034.

[9] X. Sheng, "Thin-film Silicon Solar Cells - Photonic Design, Process and Fundamentals", Ph.D. thesis, Massachusetts Institute of Technology (2012).

[10] M. A. Green, K. Emery, Y. Hishikawa, W. Warta, and E. D. Dunlop, *Progress in Photovoltaics: Research and Applications* **23**, 805 (2015). doi:10.1002/pip.2637.

[11] H. Cotal *et al.*, *Energy & Environmental Science* **2**, 174 (2009). doi:10.1039/B809257E.

[12] "CPV point focus solar cells", http://www.spectrolab.com/DataSheets/PV/CPV/C4MJ_40_Percent_Solar_Cell.pdf (2011). Accessed: 2015-09-08.

[13] I. García, I. Rey-Stolle, B. Galiana, and C. Algora, *Applied Physics Letters* **94**, 053509 (2009). doi:10.1063/1.3078817.

[14] Z. S. Judkins, "A market analysis for high efficiency Multi-Junction solar cells grown on SiGe", Master's thesis, Massachusetts Institute of Technology (2007). http://hdl.handle.net/1721.1/42143.

[15] "ASTM G173-03(2012), Standard Tables for Reference Solar Spectral Irradiances: Direct Normal and Hemispherical on 37° Tilted Surface", Tech. rep., ASTM International (2012). doi:10.1520/G0173-03R12.

[16] S. D. Lord, "A new software tool for computing Earth's atmospheric transmission of near- and far-infrared radiation, NASA Technical Memorandum 103957", Tech. rep., NASA Ames Research Center (1992). http://ntrs.nasa.gov/search.jsp?R=19930010877.

[17] M. A. Green, *Solar Cells: Operating Principles, Technology and Systems Applications* (Prentice-Hall, 1982).

[18] J. R. Chelikowsky and M. L. Cohen, *Physical Review B* **14**, 556 (1976). doi:10.1103/PhysRevB.14.556.

[19] J. E. Parrott, *Solar Energy Materials and Solar Cells* **30**, 221 (1993). doi:10.1016/0927-0248(93)90142-P.

[20] L. M. Giovane, H.-C. Luan, A. M. Agarwal, and L. C. Kimerling, *Applied Physics Letters* **78**, 541 (2001). doi:10.1063/1.1341230.

[21] M. Yamaguchi and C. Amano, *Journal of Applied Physics* **58**, 3601 (1985). doi:10.1063/1.335737.

[22] R. F. Pierret, *Semiconductor Device Fundamentals* (Addison Wesley, 1996).

[23] S. S. Li, *Semiconductor Physical Electronics* (Springer New York, 2006).

[24] S. C. Choo, *Solid-State Electronics* **11**, 1069 (1968). doi:10.1016/0038-1101(68)90129-9.

[25] C. L. Andre, D. M. Wilt, A. J. Pitera, M. L. Lee, E. A. Fitzgerald, and S. A. Ringel, *Journal of Applied Physics* **98**, 014502 (2005). doi:10.1063/1.1946194.

[26] C. L. Andre, J. J. Boeckl, D. M. Wilt, A. J. Pitera, M. L. Lee, E. A. Fitzgerald, B. M. Keyes, and S. A. Ringel, *Applied Physics Letters* **84**, 3447 (2004). doi:10.1063/1.1736318.

[27] N. Jain and M. Hudait, *IEEE Journal of Photovoltaics* **3**, 528 (2013). doi:10.1109/JPHOTOV.2012.2213073.

[28] D. A. Clugston and P. A. Basore (Conference Record of the Twenty-Sixth IEEE Photovoltaic Specialists Conference, 1997), p. 207. doi:10.1109/PVSC.1997.654065.

[29] C. Yan, R. Camacho-Aguilera, J. T. Bassette, L. C. Kimerling, and J. Michel, *Journal of Applied Physics* **112**, 034509 (2012). doi:10.1063/1.4745020.

[30] K. Sumino, *Physica Status Solidi (A)* **171**, 111 (1999). doi:10.1002/(SICI)1521-396X(199901)171:1<111::AID-PSSA111>3.0.CO;2-T.

[31] T.-L. Chan, D. West, and S. B. Zhang, *Physical Review Letters* **107**, 035503 (2011). doi:10.1103/PhysRevLett.107.035503.

[32] L. Zhang, W. E. McMahon, and S.-H. Wei, *Applied Physics Letters* **96**, 121912 (2010). doi:10.1063/1.3364140.

[33] G. Wang, T. Ogawa, T. Soga, T. Jimbo, and M. Umeno, *Applied Physics Letters* **78**, 3463 (2001). doi:10.1063/1.1376433.

[34] P. Sitch, R. Jones, S. Öberg, and M. I. Heggie, *Physical Review B* **50**, 17717 (1994). doi:10.1103/PhysRevB.50.17717.

[35] T. Ghani *et al.* (IEEE Internaional, IEDM '03 Technical Digest, 2003). doi:10.1109/IEDM.2003.1269442.

[36] J. Liu, M. Beals, A. Pomerene, S. Bernardis, R. Sun, J. Cheng, L. C. Kimerling, and J. Michel, *Nature Photonics* **2**, 433 (2008). doi:10.1038/nphoton.2008.99.

[37] G. Packeiser, *Advanced Materials* **2**, 464 (1990). doi:10.1002/adma.19900021005.

[38] K. Iga, *IEEE Journal on Selected Topics in Quantum Electronics* **6**, 1201 (2000). doi:10.1109/2944.902168.

[39] J. P. Hirth and J. Lothe, *Theory of Dislocations* (John Wiley & Sons, 1982).

[40] E. A. Fitzgerald, *Materials Science Reports* **7**, 87 (1991). doi:10.1016/0920-2307(91)90006-9.

[41] J. W. Matthews, *Journal of Vacuum Science & Technology* **12**, 126 (1975). doi:10.1116/1.568741.

[42] R. People and J. C. Bean, *Applied Physics Letters* **47**, 322 (1985). doi:10.1063/1.96206.

[43] E. Kasper, H. J. Herzog, and H. Kibbel, *Applied Physics* **8**, 199 (1975). doi:10.1007/BF00896611.

[44] B. Cunningham, J. O. Chu, and S. Akbar, *Applied Physics Letters* **59**, 3574 (1991). doi:10.1063/1.105636.

[45] C. Claeys and E. Simoen, *Fundamental and technological aspects of extended defects in Germanium* (Springer, 2008), p. 194.

[46] M. T. Currie, S. B. Samavedam, T. A. Langdo, C. W. Leitz, and E. A. Fitzgerald, *Applied Physics Letters* **72**, 1718 (1998). doi:10.1063/1.121162.

[47] S. B. Samavedam and E. A. Fitzgerald, *Journal of Applied Physics* **81**, 3108 (1997). doi:10.1063/1.364345.

[48] J.-S. Park, M. Curtin, J. M. Hydrick, J. Bai, J.-T. Li, Z. Cheng, M. Carroll, J. G. Fiorenza, and A. Lochtefeld, *Electrochemical and Solid-State Letters* **12**, H142 (2009). doi:10.1149/1.3077178.

[49] S.-I. Kobayashi, M.-L. Cheng, A. Kohlhase, T. Sato, J. Murota, and N. Mikoshiba, *Journal of Crystal Growth* **99**, 259 (1990). doi:10.1016/0022-0248(90)90523-N.

[50] C. V. Falub *et al.*, *Science* **335**, 1330 (2012). doi:10.1126/science.1217666.

[51] T. A. Langdo, C. W. Leitz, M. T. Currie, E. A. Fitzgerald, A. Lochtefeld, and D. A. Antoniadis, *Applied Physics Letters* **76**, 3700 (2000). doi:10.1063/1.126754.

[52] C. Pribat, G. Servanton, L. Depoyan, and D. Dutartre, *Solid-State Electronics* **53**, 865 (2009). doi:10.1016/j.sse.2009.04.038.

[53] J. W. Matthews and E. Klokholm, *Materials Research Bulletin* **7**, 213 (1972). doi:10.1016/0025-5408(72)90239-5.

[54] L. Colace *et al.*, *Applied Physics Letters* **72**, 3175 (1998). doi:10.1063/1.121584.

[55] H.-C. Luan, D. R. Lim, K. K. Lee, K. M. Chen, J. G. Sandland, K. Wada, and L. C. Kimerling, *Applied Physics Letters* **75**, 2909 (1999). doi:10.1063/1.125187.

[56] K. Wada *et al.*, *Optical interconnects* (Springer, 2006), chap. 11, pp. 291–310. doi:10.1007/978-3-540-28912-8_11.

[57] H. J. Osten, J. Klatt, G. Lippert, E. Bugiel, and S. Higuchi, *Journal of Applied Physics* **74**, 2507 (1993). doi:10.1063/1.354690.

[58] T. F. Wietler, E. Bugiel, and K. R. Hofmann, *Applied Physics Letters* **87**, 182102 (2005). doi:10.1063/1.2120900.

[59] K.-W. Ang, M.-B. Yu, G.-Q. Lo, and D.-L. Kwong, *IEEE Electron Device Letters* **29**, 1124 (2008). doi:10.1109/LED.2008.2004469.

[60] T. H. Loh *et al.*, *Applied Physics Letters* **90**, 092108 (2007). doi:10.1063/1.2709993.

[61] J. Nakatsuru, H. Date, S. Mashiro, and M. Ikemoto (2006), vol. 891 of *MRS Proceedings*. doi:10.1557/PROC-0891-EE07-24.

[62] H.-Y. Yu *et al.*, *Applied Physics Letters* **97**, 063503 (2010). doi:10.1063/1.3478242.

[63] Q. Li, Y.-B. Kiang, H. Xu, S. Hersee, and S. M. Han, *Applied Physics Letters* **85**, 1928 (2004). doi:10.1063/1.1790027.

[64] D. Leonhardt *et al.*, *Thin Solid Films* **519**, 7664 (2011). doi:10.1016/j.tsf.2011.05.044.

[65] "Concentrator photovoltaics: Frequently asked questions", http://www.spectrolab.com/faqs-terrestrial.htm. Accessed: 2015-08-24.

[66] R. Loo *et al.*, *Journal of The Electrochemical Society* **157**, H13 (2010). doi:10.1149/1.3244564.

[67] J. Liu, "GeSi Photodetectors and Electro-absorption Modulators for Si Electronic-photonic Integrated Circuits", Ph.D. thesis, Massachusetts Institute of Technology (2007). http://hdl.handle.net/1721.1/38582.

[68] J. H. Ye, P.-C. Lee, and M. S. Zhou (Plasma Processing XII, 1998), vol. 98, p. 189.

[69] Y. Cai, "Materials Science and Design for Germanium Monolithic Light Source on Silicon", Ph.D. thesis, Massachusetts Institute of Technology (2014). http://hdl.handle.net/1721.1/92958.

[70] H. Jansen, H. Gardeniers, M. de Boer, M. Elwenspoek, and J. Fluitman, *Journal of Micromechanics and Microengineering* **6**, 14 (1996). doi:10.1088/0960-1317/6/1/002.

[71] W. Kern, *Journal of the Electrochemical Society* **137**, 1887 (1990). doi:10.1149/1.2086825.

[72] E. Yablonovich, D. L. Allara, C. C. Chang, T. Gmitter, and T. B. Bright, *Physical Review Letters* **57**, 249 (1986). doi:10.1103/PhysRevLett.57.249.

[73] D. W. Greeve, *Properties of Crystalline Silicon* (Institution of Engineering and Technology, 1998), chap. 2.

[74] R. Singh, D. L. Harame, and M. M. Oprysko, *Silicon germanium: technology, modeling, and design* (IEEE Press, 2004).

[75] X. Sun, "Ge-on-Si light-emitting materials and devices for silicon photonics", Ph.D. thesis, Massachusetts Institute of Technology (2009). http://hdl.handle.net/1721.1/53254.

[76] W. He, W. Lv, and J. H. Dickerson, *Gas Transport in Solid Oxide Fuel Cells* (Springer International Publishing, 2014). doi:10.1007/978-3-319-09737-4_2.

[77] D. W. Greve, *Thin Films: Advances in Research and Development* (Academic Press, 1998), vol. 23.

[78] D. J. Robbins, J. L. Glasper, A. G. Cullis, and W. Y. Leong, *Journal of Applied Physics* **69**, 3729 (1991). doi:10.1063/1.348466.

[79] J. Cheng, "Large area Ge-on-Si heterostructures for power generation and photon detection", Ph.D. thesis, Massachusetts Institute of Technology (2011).

[80] S. Sioncke, D. P. Brunco, M. Meuris, O. Uwamahoro, J. Van Steenbergen, E. Vrancken, and M. M. Heyns, *Solid State Phenomena* **145-146**, 203 (2009). doi:10.4028/www.scientific.net/SSP.145-146.203.

[81] S. A. Shevchenko, *Semiconductors* **34**, 527 (2000). doi:10.1134/1.1188021.

[82] K. R. Williams, K. Gupta, and M. Wasilik, *Journal of Microelectromechanical Systems* **12**, 761 (2003). doi:10.1109/JMEMS.2003.820936.

[83] S. Sioncke, D. P. Brunco, M. Meuris, O. Uwamahoro, J. Van Steenbergen, E. Vrancken, and M. M. Heyns, *ECS Transactions* **16**, 451 (2008). doi:10.1149/1.2986802.

[84] K. Yamamoto, A. Nakamura, and U. Hase, *IEEE Transactions on Semiconductor Manufacturing* **12**, 288 (1999). doi:10.1109/66.778192.

[85] U. Schnakenberg, W. Benecke, and B. Löchel, *Sensors and Actuators* **A21-A23**, 1031 (1990). doi:10.1016/0924-4247(90)87084-V.

[86] I. Haller, Y. H. Lee, J. J. Nocera Jr., and M. A. Jaso, *Journal of the Electrochemical Society* **135**, 2042 (1988). doi:10.1149/1.2096204.

[87] J. Zhao, A. Wang, M. A. Green, and F. Ferrazza, *Applied Physics Letters* **73**, 1991 (1998). doi:10.1063/1.122345.

[88] S. K. Stanley, S. S. Coffee, and J. G. Ekerdt, *Applied Surface Science* **252**, 878 (2005). doi:10.1016/j.apsusc.2005.01.149.

[89] A. Luan, "Ge Photodetectors for Si Microphotonics", Ph.D. thesis, Massachusetts Institute of Technology (2001). http://hdl.handle.net/1721.1/8782.

[90] Y. Okada and Y. Tokumaru, *Journal of Applied Physics* **56**, 314 (1984). doi:10.1063/1.333965.

[91] H. P. Singh, *Acta Crystallographica* **A24**, 469 (1968). doi:10.1107/S056773946800094X.

[92] B. El-Kareh, *Fundamentals of Semiconductor Processing Technology* (Springer US, 1995).

[93] W. P. Mason, *Physical Acoustics and the Properties of Solids* (D. Van Nostrand Company, 1958).

[94] J. J. Wortman and R. A. Evans, *Journal of Applied Physics* **36**, 153 (1965). doi:10.1063/1.1713863.

[95] B. Patham and X. Huang, *Journal of Composites* **2014**, 172560 (2014). doi:10.1155/2014/172560.

[96] J. Lu, J. Liu, G. Rozgonyi, L. Kimerling, and J. Michel, *ECS Transactions* **3**, 325 (2006). doi:10.1149/1.2355830.

[97] Y. Takada, J. Osaka, Y. Ishikawa, and K. Wada, *Japanese Journal of Applied Physics* **49**, 04DG23 (2010). doi:10.1143/JJAP.49.04DG23.

[98] A. K. Head, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **44**, 92 (1953). doi:10.1080/14786440108520278.

[99] J. S. Speck, M. A. Brewer, G. Beltz, A. E. Romanov, and W. Pompe, *Journal of Applied Physics* **80**, 3808 (1996). doi:10.1063/1.363334.

[100] T. S. Hickernell, F. M. Fliegel, and F. S. Hickernell, *Ultrasonics Symposium, IEEE Proceedings* **1**, 445 (1990). doi:10.1109/ULTSYM.1990.171405.

[101] A. Singh, R. P. S. Rathore, and R. M. Agarwal, *Acta Physica Hungarica* **72**, 133 (1992). http://link.springer.com/article/10.1007/BF03054157.

[102] A. E. Romanov, W. Pompe, G. Beltz, and J. S. Speck, *Physica Status Solidi (B)* **198**, 599 (1996). doi:10.1002/pssb.2221980205.

[103] J. E. Ayers, *Journal of Applied Physics* **78**, 3724 (1995). doi:10.1063/1.359952.

[104] A. R. Chaudhuri, J. R. Patel, and L. G. Rubin, *Journal of Applied Physics* **33**, 2736 (1962). doi:10.1063/1.1702540.

[105] G. Wang, R. Loo, E. Simoen, L. Souriau, M. Caymax, M. M. Heyns, and B. Blanpain, *Applied Physics Letters* **94**, 102115 (2009). doi:10.1063/1.3097245.

[106] H.-J. Möller, H. Ewaldt, and P. Haasen, *Physica Status Solidi (A)* **55**, 469 (1979). doi:10.1002/pssa.2210550214.

[107] E. A. Fitzgerald, *Journal of Vacuum Science & Technology B* **7**, 782 (1989). doi:10.1116/1.584600.

[108] X. G. Zhang, P. Li, G. Zhao, D. W. Parent, F. C. Jain, and J. E. Ayers, *Journal of Electronic Materials* **27**, 1248 (1998). doi:10.1007/s11664-998-0078-3.

[109] L. B. Freund, *Journal of Applied Physics* **68**, 2073 (1990). doi:10.1063/1.346560.

[110] Y. Cai, W. Yu, L. C. Kimerling, and J. Michel, *ECS Journal of Solid State Science and Technology* **3**, P5 (2014). doi:10.1149/2.003401jss.

[111] Q. Xiang, S. Li, D. Wang, K. L. Wang, J. G. Couillard, and H. G. Craighead, *Journal of Vacuum Science & Technology B* **14**, 2381 (1996). doi:10.1116/1.588866.

[112] S.-H. Lim, S. Song, G.-D. Lee, E. Yoon, and J.-H. Lee, *Journal of Vacuum Science & Technology B* **22**, 682 (2004). doi:10.1116/1.1676595.

[113] R. M. Sieg, S. A. Ringel, S. M. Ting, E. A. Fitzgerald, and R. N. Sacks, *Journal of Electronic Materials* **27**, 900 (1998). doi:10.1007/s11664-998-0116-1.

[114] H. Hirayama, M. Hiroi, and T. Ide, *Physical Review B* **48**, 17331 (1993). doi:10.1103/PhysRevB.48.17331.

[115] L. Vescan, K. Grimm, and C. Dieker, *Journal of Vacuum Science & Technology B* **16**, 1549 (1998). doi:10.1116/1.589937.

[116] J.-S. Park, J. Bai, M. Curtin, M. Carroll, and A. Lochtefeld, *Journal of Vacuum Science & Technology B* **26**, 117 (2008). doi:10.1116/1.2825165.

[117] J. H. Nam *et al.*, *Journal of Crystal Growth* **416**, 21 (2015). doi:10.1016/j.jcrysgro.2014.11.004.

[118] R. Hull, J. C. Bean, G. S. Higashi, M. L. Green, L. Peticolas, D. Bahnck, and D. Brasen, *Applied Physics Letters* **60**, 1468 (1992). doi:10.1063/1.107273.

[119] S. H. Jones, L. K. Seidel, K. M. Lau, and M. Harold, *Journal of Crystal Growth* **108**, 73 (1991). doi:10.1016/0022-0248(91)90355-9.

[120] D. B. Williams and C. B. Carter, *Transmission electron microscopy: a textbook for materials science* (Springer, 2008), 2nd ed. doi:10.1007/978-0-387-76501-3.

[121] G. Kissinger and S. Pizzini, *Silicon, Germanium, and Their Alloys: Growth, Defects, Impurities, and Nanocrystals* (Springer International Publishing, 2014).

[122] B. Schwartz and H. Robbins, *Journal of the Electrochemical Society* **123**, 1903 (1976). doi:10.1149/1.2132721.

[123] H. Robbins and B. Schwartz, *Journal of the Electrochemical Society* **107**, 108 (1960). doi:10.1149/1.2427617.

[124] L. Pasemann, H. Blumtritt, and R. Gleichmann, *Physica Status Solidi (A)* **70**, 197 (1982). doi:10.1002/pssa.2210700125.

[125] J. Hess, J. Schreiber, S. Hildebrandt, and R. Labusch, *Physica Status Solidi (B)* **172**, 225 (1992). doi:10.1002/pssb.2221720121.

[126] S. Sun, Y. Sun, Z. Liu, D.-I. Lee, S. Peterson, and P. Pianetta, *Applied Physics Letters* **88**, 021903 (2006). doi:10.1063/1.2162699.

[127] J.-Y. J. Lin, A. M. Roy, A. Nainani, Y. Sun, and K. C. Saraswat, *Applied Physics Letters* **98**, 092113 (2011). doi:10.1063/1.3562305.

[128] M. Ikeda and K. Kaneko, *Journal of Applied Physics* **66**, 5285 (1989). doi:10.1063/1.343718.

[129] I. Rey-Stolle, E. Barrigón, B. Galiana, and C. Algora, *Journal of Crystal Growth* **310**, 4803 (2008). doi:10.1016/j.jcrysgro.2008.07.116.

[130] T. Akane, H. Okumura, J. Tanaka, and S. Matsumoto, *Thin Solid Films* **294**, 153 (1997). doi:10.1016/S0040-6090(96)09258-9.

[131] A. Luque and A. Martí, *Physical Review Letters* **78**, 5014 (1997). doi:10.1103/PhysRevLett.78.5014.

[132] A. Martí, E. Antolín, E. Cánovas, N. López, P. G. Linares, A. Luque, C. R. Stanley, and C. D. Farmer, *Thin Solid Films* **516**, 6716 (2008). doi:10.1016/j.tsf.2007.12.064.

[133] T. C. Wu, N. J. Thompson, D. N. Congreve, E. Hontz, S. R. Yost, T. Van Voorhis, and M. A. Baldo, *Applied Physics Letters* **104**, 193901 (2014). doi:10.1063/1.4876600.

[134] J. R. Sites, *Solar Cells* **25**, 163 (1988). doi:10.1016/0379-6787(88)90020-8.

[135] A. M. Carlin, E. A. Fitzgerald, and S. A. Ringel, in *Photovoltaic Specialist Conference (PVSC), 2015 IEEE 42nd* (2015), pp. 1–4. doi:10.1109/PVSC.2015.7355600.

[136] L. Z. Broderick, B. R. Albert, B. S. Pearson, L. C. Kimerling, and J. Michel, *Solar Energy Materials and Solar Cells* **136**, 48 (2015). doi:10.1016/j.solmat.2014.12.034.

[137] X. Sheng *et al.*, *Nature Materials* **13**, 593 (2014). doi:10.1038/nmat3946.

[138] M. Stefancich, A. Zayan, M. Chiesa, S. Rampino, D. Roncati, L. Kimerling, and J. Michel, *Optics Express* **20**, 9004 (2012). doi:10.1364/OE.20.009004.

[139] Q. Huang *et al.*, *Applied Optics* **52**, 2312 (2013). doi:10.1364/AO.52.002312.

[140] A. Barnett *et al.*, *Progress in Photovoltaics: Research and Applications* **17**, 75 (2009). doi:10.1002/pip.852.

[141] C. N. Eisler, E. D. Kosten, E. C. Warmann, and H. A. Atwater, in *Photovoltaic Specialists Conference (PVSC), 2013 IEEE 39th* (2013), pp. 1848–1851. doi:10.1109/PVSC.2013.6744502.

[142] M. K. Alam, F. H. Khan, and A. S. Imtiaz, in *IECON 2011 - 37th Annual Conference on IEEE Industrial Electronics Society* (2011), pp. 4373–4378. doi:10.1109/IECON.2011.6120028.

[143] G. Sala and I. Antón, *Handbook of Photovoltaic Science and Engineering* (John Wiley & Sons, Ltd, 2010), chap. 10. doi:10.1002/9780470974704.ch10.

[144] S. A. Kalogirou, *Progress in Energy and Combustion Science* **30**, 231 (2004). doi:10.1016/j.pecs.2004.02.001.

[145] S. Kuravi, J. Trahan, D. Y. Goswami, M. M. Rahman, and E. K. Stefanakos, *Progress in Energy and Combustion Science* **39**, 285 (2013). doi:10.1016/j.pecs.2013.02.001.

[146] P. Denholm and R. M. Margolis, *Energy Policy* **35**, 2852 (2007). doi:10.1016/j.enpol.2006.10.014.

[147] National Renewable Energy Laboratory, "Gemasolar Thermosolar Plant", http://www.nrel.gov/csp/solarpaces/project_detail.cfm/projectID=40 (2011). Accessed: 2016-01-04.

[148] A. Bejan, *Journal of Applied Physics* **79**, 1191 (1996). doi:10.1063/1.362674.

[149] C. K. Ho and B. D. Iverson, *Renewable and Sustainable Energy Reviews* **29**, 835 (2014). doi:10.1016/j.rser.2013.08.099.

[150] A. G. Imenes and D. R. Mills, *Solar Energy Materials and Solar Cells* **84**, 19 (2004). doi:10.1016/j.solmat.2004.01.038.

[151] T. Takamoto, E. Ikeda, H. Kurita, and M. Ohmori, *Solar Energy Materials and Solar Cells* **35**, 25 (1994). doi:10.1016/0927-0248(94)90118-X.

[152] Y. Yi, S. Akiyama, P. Bermel, X. Duan, and L. C. Kimerling, *Optics Express* **12**, 4775 (2004). doi:10.1364/OPEX.12.004775.

[153] B. Temelkuran, S. D. Hart, G. Benoit, J. D. Joannopoulos, and Y. Fink, *Nature* **420**, 650 (2002). doi:10.1038/nature01275.

[154] L. Zeng *et al.*, *Applied Physics Letters* **93**, 221105 (2008). doi:10.1063/1.3039787.

[155] X. Sheng, L. Z. Broderick, J. Hu, L. Yang, A. Eshed, E. A. Fitzgerald, J. Michel, and L. C. Kimerling, *Optics Express* **19**, A701 (2011). doi:10.1364/OE.19.00A701.

[156] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics* (John Wiley & Sons, 2007), chap. 7.

[157] J. Tao, M. Yang, J. W. Chai, J. S. Pan, Y. P. Feng, and S. J. Wang, *Journal of Physical Chemistry C* **118**, 994 (2014). doi:10.1021/jp408798f.

[158] W. H. Southwell, *Applied Optics* **28**, 5091 (1989). doi:10.1364/AO.28.005091.

[159] A. Thelen, *Journal of the Optical Society of America* **61**, 365 (1971). doi:10.1364/JOSA.61.000365.

[160] W. H. Southwell and R. L. Hall, *Applied Optics* **28**, 2949 (1989). doi:10.1364/AO.28.002949.

[161] A. V. Tikhonravov and J. A. Dobrowolski, *Applied Optics* **32**, 4265 (1993). doi:10.1364/AO.32.004265.

[162] J. A. Dobrowolski and D. Lowe, *Applied Optics* **17**, 3039 (1978). doi:10.1364/AO.17.003039.

[163] S. Liu and Y.-H. Zhang, in *Optics for Solar Energy* (2013), pp. RM4D–5. doi:10.1364/OSE.2013.RM4D.5.

[164] K. Hunger and W. Herbst, *Pigments, Organic* (Wiley-VCH Verlag GmbH & Co. KGaA, 2000). doi:10.1002/14356007.a20_371.

[165] G. Buxbaum *et al.*, *Pigments, Inorganic, 3. Colored Pigments* (Wiley-VCH Verlag GmbH & Co. KGaA, 2000). doi:10.1002/14356007.n20_n02.

[166] P. Vukusic, J. R. Sambles, C. R. Lawrence, and R. J. Wootton, *Proceedings of the Royal Society B: Biological Sciences* **266**, 1403 (1999). doi:10.1098/rspb.1999.0794.

[167] Y. Zhao, Z. Xie, H. Gu, C. Zhu, and Z. Gu, *Chemical Society Reviews* **41**, 3297 (2012). doi:10.1039/C2CS15267C.

[168] A. Saito, *Science and Technology of Advanced Materials* **12**, 064709 (2011). doi:10.1088/1468-6996/12/6/064709.

[169] H. Shiomi, E. Misaki, M. Adachi, and F. Suzuki, *Journal of Coatings Technology and Research* **5**, 455 (2008). doi:10.1007/s11998-008-9085-9.

[170] "High carmine prices boost search for alternative colours", http://www.foodnavigator.com/content/view/print/317236. Accessed: 2015-09-01.

[171] European Food Safety Authority, *EFSA Journal* **7**, 1331 (2009). doi:10.2903/j.efsa.2009.1331.

[172] H. Kim *et al.*, *Nature Photonics* **3**, 534 (2009). doi:10.1038/nphoton.2009.141.

[173] K. Suzumori, M. Mihara, and S. Wakimoto, in *IEEE International Conference on Robotics and Automation (ICRA)* (2011), pp. 2771–2776. doi:10.1109/ICRA.2011.5979945.

[174] C. Chui *et al.*, *Proceedings of SPIE* **6466**, 646609 (2007). doi:10.1117/12.713949.

[175] T. Yasuda, K. Nishikawa, and S. Furukawa, *Dyes and Pigments* **92**, 1122 (2012). doi:10.1016/j.dyepig.2011.08.006.

[176] D. Banerjee and M. Zhang, *Journal of Modern Optics* **57**, 1180 (2010). doi:10.1080/09500340.2010.506956.

[177] "Colorimetry", Tech. Rep. 15, Commission Internationale de l'Eclairage, Vienna (2004).

[178] W. D. Wright, *Transactions of the Optical Society* **31**, 201 (1930). doi:10.1088/1475-4878/31/4/303.

[179] H. S. Fairman, M. H. Brill, and H. Hemmendinger, *Color Research & Application* **22**, 11 (1997). doi:10.1002/(SICI)1520-6378(199702)22:1<11::AID-COL4>3.0.CO;2-7.

[180] V. T. Nguyen, "Efficient power coupling to waveguides in high index contrast systems", Ph.D. thesis, Massachusetts Institute of Technology (2006). http://hdl.handle.net/1721.1/36211.

[181] Z. C. Holman, A. Descoeudres, L. Barraud, F. Z. Fernandez, J. P. Seif, S. De Wolf, and C. Ballif, *IEEE Journal of Photovoltaics* **2**, 7 (2012). doi:10.1109/JPHOTOV.2011.2174967.