DIODES FOR OPTICAL RECTENNAS

by

SACHIT GROVER

B.Tech., Electrical Engineering,
Indian Institute of Technology Delhi, 2006

M.S., Electrical Engineering,
University of Colorado Boulder, 2009

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
Department of Electrical Engineering
2011

This thesis entitled:

Diodes for Optical Rectennas

written by Sachit Grover

has been approved for the Department of Electrical Engineering

_____

Prof. Garret Moddel

_____

Prof. Charles Rogers

Date_____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline

Grover, Sachit (Ph.D., Electrical Engineering)

Diodes for Optical Rectennas

Thesis directed by Prof. Garret Moddel

Two types of ultra-fast diode are fabricated, characterized, and simulated for use in optical rectennas. A rectenna consists of an antenna connected to a diode in which the electromagnetic radiation received by the antenna is rectified in the diode. I have investigated metal/insulator/metal (MIM) tunnel diodes and a new, geometric diode for use in rectenna-based infrared detectors and solar cells. Factors influencing the performance of a rectenna are analyzed. These include DC and optical-frequency diode-characteristics, circuit parameters, signal amplitude, and coherence of incoming radiation.

To understand and increase the rectification response of MIM-based rectennas, I carry out an in-depth, simulation-based analysis of MIM diodes and design improved multi-insulator tunnel barriers. MIM diodes are fundamentally fast. However, from a small-signal circuit model the operating frequency of a rectenna is found to be limited by the diode's RC time constant. To overcome this limitation, I have designed and simulated a distributed rectifier that uses the MIM diode in a traveling-wave configuration. High-frequency characteristics of MIM diodes are obtained from a semiclassical theory for photon-assisted tunneling. Using this theory, the dependence of rectenna efficiency on diode characteristics and signal amplitude is evaluated along with the maximum achievable efficiency. A correspondence is established between the first-order semiclassical theory and the small-signal circuit model.

The RC time constant of MIM diodes is too large for efficient operation at near-infrared-to-visible frequencies. To this end, a new, planar rectifier that consists of an asymmetrically-patterned thin-film, is developed. The diode behavior in this device is attributed to the geometric asymmetry of the conductor. Geometric diodes are fabricated using graphene and measured for response to infrared illumination. To model the $I(V)$ curve of geometric diodes, I have implemented a quantum mechanical simulation based on the tight-binding Hamiltonian. The simulated and the measured current-voltage characteristics are consistent with each other. I have also derived a semiclassical theory, analogous to the one for MIM diodes, for analyzing the optical response of geometric diodes.

Dedicated to my parents

Kamlesh and Anand Grover

&

my wife

Ginni

ACKNOWLEDGMENTS

The last five years of my life have been enriched in many ways through my interaction with Prof. Garret Moddel. I wish to extend my sincere thanks to him for not just being an excellent advisor but a great mentor as well.

I thank the faculty members in my committee, Profs. Robert McLeod, Won Park, Charles Rogers, and Bart Van Zeghbroeck, for sharing their expertise on various aspects of my thesis. I also thank Profs. E. F. Kuester, Kelvin Wagner, and Leo Radzihovsky for helpful conversations. I sincerely thank Mike Estes for our collaborations, and Mario Scurati for his questions related to rectenna design.

I gratefully acknowledge Dave Powell of Abengoa Solar for financial support. Thanks is due to the never tiring team at CNL, including Jan Van Zeghbroeck, Tomoko Borsa, and Vincenzo LaSalvia, and to Bill Mitchell at UCSB for process support. I also thank former Phiar Corporation employees for their help and guidance on various fronts related to the development of the MIMFET.

For making grad school an enjoyable experience and for being partners in procrastination, I wish to thank current and former members of the Quantum Engr. Lab. including Olga Dmitriyeva, Rahul Trivedi, David Doroski, James Zhu, Kendra Krueger, Saumil Joshi, Xi Chen, M. A. Mohamed, S. Tigrek, and Ben McPheron. In the same spirit, I also thank my friends outside the lab including Venkata Tamma, Parag Shah, Gurpreet Singh, and Krishna Ramadurai.

I am eternally grateful to my parents, Kamlesh and Anand, for the strength and encouragement they provided throughout this journey. I thank my sisters, Kanika and Shalini, and their families for constant love and support, with a high five to Sumira, Arnav, Neal, and Kabir for enlivening my vacations. Finally, using the clichéd adage, behind every successful man there is a woman, I thank my wife Ginni for her love, patience, support, and understanding.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

A. Rectennas

Extensive research is being conducted on new materials and concepts for the next generation of energy conversion and sensing devices (Luryi, 2010). In order to make solar energy as cost effective as coal, the requirements for a third generation solar cell are low-cost and high power conversion efficiency (Green, 2001). Similar cost and efficiency requirements are imposed on infrared detector technologies for applications like automotive night vision. The rectenna is a device that has the potential to deliver the desired performance at low-cost for both photovoltaics and detection. It is essentially an antenna plus rectifier, which in-principle can operate at any frequency.

The basic rectenna circuit comprises an antenna connected to a diode in which the electromagnetic radiation received by the antenna is converted to a DC signal by the diode. The radiation induces an AC signal on the antenna arms, which gets channeled into a region, called the antenna feedpoint, where the diode is connected. Such an arrangement is shown in Figure I-1. The conversion from AC to DC occurs due to the difference in resistance of the diode for the positive and the negative cycles of the oscillating current induced on the antenna. Depending on whether the DC signal is sensed by an amplifier or applied across a load resistance, the rectenna can be configured as a detector or as a photovoltaic rectifier.

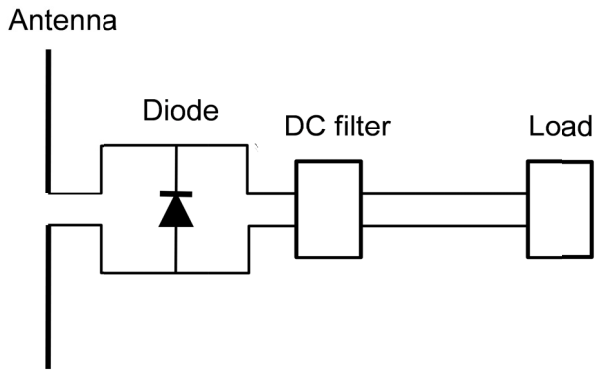


Figure I-1 Schematic of an antenna-coupled diode rectifier, also known as a rectenna.

Significant research was conducted in the 1960s and 70s towards the use of rectennas for microwave powered helicopters and airplanes (Brown, 1984). At low frequencies, rectennas with power-conversion efficiencies greater than 90% (Brown, 1976) have been demonstrated and are used in a variety of energy transmission (Shinohara, 1998) and harvesting (Hagerty, 2004), (Singh, 2006) applications.

At higher frequencies, detectors based on rectennas have been widely investigated (Fumeaux, 1998). Infrared rectenna-detectors have several advantages over narrowband semiconductor photodetectors (Rogalski, 2003). Rectennas do not require cooling and there is no lower limit on the frequency of the signal that can be detected. Moreover, rectennas can detect the envelope of a carrier signal modulated at an extremely high frequency. For infrared carrier wavelengths, this allows rectennas to have modulation bandwidths several orders of magnitude higher than those of competing bolometers (Richards, 1997).

Photovoltaic rectification is another application where rectennas can potentially form high efficiency solar cells (Berland, 2003). This idea was originally proposed (Bailey, 1972) and patented (Marks, 1984) several decades ago, and has recently gained significant attention. Here again, the absence of a bandgap implies that the efficiency limitations of a semiconductor solar cell are not applicable.

Despite the decades of work done in this field, infrared and visible applications of the rectenna are still in the research phase. On the antenna side, historically, planar metal antennas have been used in infrared detectors (Fumeaux, 1998) and work is underway for their large scale integration into flexible substrates for thermal photovoltaics (Kotter, 2008). However, even at 10.6 μm wavelength, the resistive losses in the antenna-metal limit the antenna efficiency to below 50% (González, 2005). A proposed alternative is dielectric antennas that collect and channel radiation by means of a dielectric rod (Sarehraz, 2005). Such antennas can potentially achieve higher efficiencies at visible wavelengths. However, considerable research and innovation is required to efficiently interface them with diodes.

The choice of a suitable diode for a rectenna is based on its operating frequency. The transit time of charges in semiconductor p-n junction diodes limits their frequency of operation to the gigahertz range (Sedra, 1997). At 35 GHz, rectennas using GaAs Schottky diodes have been designed (Yoo, 1992). Schottky diodes are also used at terahertz and far-infrared frequencies (Brown, 2004), (Kazemi, 2007). However, beyond 12 THz the metal/insulator/metal (MIM) tunnel diode is deemed more suitable for rectennas (Hübers, 1994).

In this thesis, I investigate two diode technologies for rectennas operating at optical frequencies. These include the MIM diode and a new diode, referred to as geometric diode. In conjunction with my work on the diodes, I lay emphasis on the analysis of the diode properties, leading to the projected performance of a detector or a solar cell. In the next section, I provide a background on MIM diodes, which have been extensively used in optical rectennas. I also explain the operation of geometric diodes (Moddel, 2009), which are currently under investigation as potential rectifiers for optical rectennas.

B. High-speed diodes

1. MIM diode

A metal/insulator/metal tunnel diode is made of two metallic electrodes spaced apart by an extremely thin (few nanometers) insulator. In the MIM diode, the metals have a higher work function than the electron-affinity of the insulator producing a barrier at the metal-insulator interfaces as shown in the energy-band profile in Figure I-2.

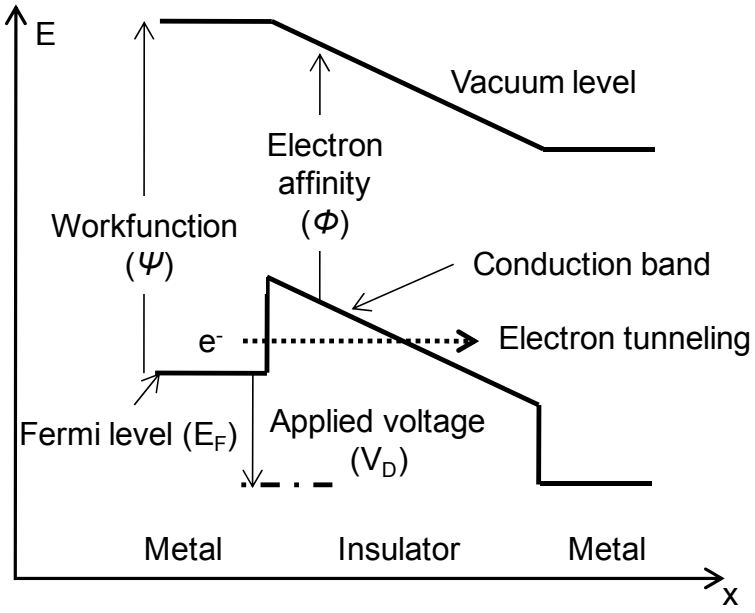


Figure I-2 Energy-band profile of an MIM diode.

Charge transport across the insulator occurs due to the quantum-mechanical tunneling (Kroemer, 1994, p.148) of electrons. A transmission probability is associated with the likelihood of an electron tunneling through the classically forbidden region of an insulator bandgap. This probability depends exponentially on the thickness and height of the barrier (Stratton, 1962), which changes with the voltage across the diode. This gives rise to the nonlinear dependence of the tunnel current on the applied voltage and hence the diode characteristics.

Electron tunneling in MIM junctions occurs on a femtosecond timescale (Nagae, 1972). This inherently fast charge transport across the tunnel barrier allows MIM diodes to operate at optical frequencies. To ensure that tunneling is the dominant conduction mechanism, the thickness of the insulating layer should not be more than a few nanometers (Simmons, 1971).

Early MIM point-contact diodes were made by pressing a thin metal-wire against an oxidized sheet of metal (Riccius, 1978). They are also called cat's-whisker diodes, and achieve small junction areas without requiring fine lithography. Point-contact diodes use a simple fabrication technique that can test several metal wires for the same oxidized metal sheet at a high throughput (Periasamy, 2010). However, such diodes are unreliable due to the mechanically unstable nature of the junction and it is difficult to reproduce their characteristics. Moreover, an unintended native oxide layer or impurities can significantly alter the diode characteristics. Nevertheless, point-contact MIM diodes have been successfully used in experiments on the detection and mixing of infrared radiation (Riccius, 1984) and in the frequency characterization of laser lines in the infrared (Evenson, 1970).

Significant progress in lithography has allowed small area MIM diodes to be made more reliably. Moreover, the diodes can be made using thin-film materials. The insulator can be a grown oxide obtained by oxidizing a metal film to the desired thickness, followed by the deposition of a second metal. Alternatively, depositing a stack of metals and insulators provides the freedom to choose the barrier materials independent of the metals. These techniques allow the MIM tunnel barrier to be formed without breaking vacuum, preventing contamination at the metal-insulator interfaces.

Despite the progress in the fabrication methods, rectennas based on MIM diode lack the efficiency required for successful application in detectors or solar cells. A major factor for the inefficiency is the large RC time constant of the rectenna circuit

that uses MIM diodes (Sanchez, 1978). A proposed solution for reducing the RC value is to reduce the junction capacitance by making smaller diodes. However, this increases the diode resistance leading to an impedance mismatch between the antenna and the diode.

An alternative, approach is to use a diode that minimizes the capacitance due to the absence of a parallel-plate structure. Such a diode is discussed in the next section.

2. Geometric diode

A mesoscopic junction is categorized as having length-scales comparable to the electron phase coherence length (l_c) (Fal'ko, 1989). An asymmetric mesoscopic junction can show rectification due to interaction of carriers with the conductor boundaries. Rectification in such junctions was predicted to occur due to an asymmetric conductor or an asymmetric illumination in a symmetric conductor (Datta, 1992). At least one of these two conditions is necessary. A photovoltaic effect has been observed in small conductors having geometric asymmetry due to disorder (Liu, 1990).

A previously reported device based on the concept of geometric rectification in asymmetric junctions imposes the requirement of ballistic transport of charges through the device (Song, 1998). Song's device uses a four terminal configuration, similar to a bridge rectifier, in which the AC signal is applied across two opposite terminals and the DC voltage is tapped from an orthogonal pair of terminals. The geometric diode reported here is a two terminal device and does not require ballistic transport.

The geometric diode consists of a patterned thin-film that allows a preferential motion of charge carriers in a direction defined by its geometry (Moddel, 2009). As

explained later, the relevant size requirement for an asymmetric-hourglass-shaped geometric diode shown in Figure I-3 is that the neck-size should be smaller than l_c.

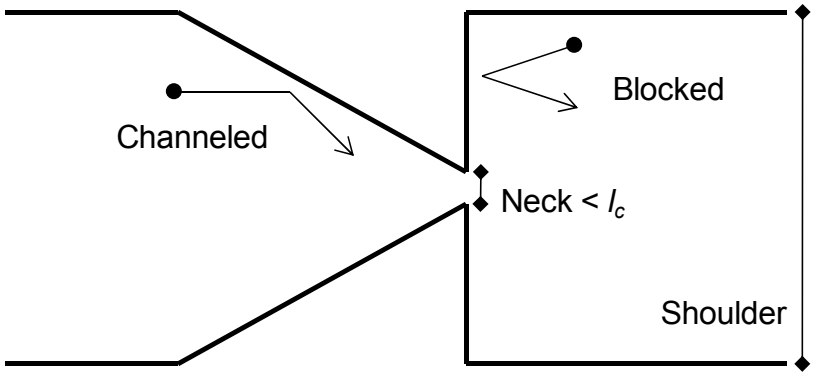


Figure I-3 Top view of a geometric diode. The constriction labelled as the neck is comparable to or smaller than the electron mean-free-path. The geometry channels charges towards the right and blocks their flow towards the left.

A classical explanation for the rectification property of the geometric diode is based on the Drude model (Ashcroft, 1976, p.1). Free charge carriers in a conductor move randomly at a velocity equal to the Fermi velocity. Assuming the conductor to have perfect edges from which the charges can reflect specularly, on striking an edge, the direction of motion for the charge is changed while maintaining the speed. In the region to the right of the neck charges moving leftwards are deflected in the opposite direction due to the vertical edge, while charges in the left region moving rightwards collide with the slanting edge, and funnel through to the right of the neck.

Inelastic scattering resets the state-of-motion of the charges, thereby reducing the effect of the geometry. Therefore, only on length-scale smaller than the inelastic collision length (l_{in}), a charge near the neck senses the geometric asymmetry leading to the net flow in one direction.

In the absence of an external bias, the net flow is compensated by a self-bias that occurs due to the local redistribution of the charge. However, on applying a

voltage, the magnitude of current varies depending on the direction of bias due to the influence of the edges.

Consistent with the classical description, the assumption of specular reflection from the edges, and the motion of charges through the neck without losing directionality, can be expressed in terms of the wave nature of the charge particles (Thouless, 1980). The quantum equivalent of the specular reflections is that the edges define the shape of an electronic wavepacket and influence its diffusion. The distance over which the wavepacket is able to diffuse elastically before undergoing an inelastic collision is controlled by l_{in} and the elastic diffusion length l_e. This distance is the electron phase-coherence length (l_c). Under the condition that the charges have more wave-character than corpuscular $l_{in} >> l_e$, the l_c is obtained as (Nimtz, 1988)

$$l_c = \sqrt{l_e l_{in}/3}$$

Eq. I-1

To observe the effect of geometry on charge transport, the critical dimensions of the conductor must be smaller than l_c. As explained in chapter VI, it is difficult to meet this requirement using metal thin-films due to a small l_{in} and thereby small l_c. Asymmetric characteristics of geometric diodes have been demonstrated using graphene as the conductor (Zhu, 2011). Graphene is a 2D sheet of carbon atoms arranged in a hexagonal lattice. It has the advantage of having an l_c of the order of 100 nm even at room temperature (Berger, 2006).

I analyze the characteristics of graphene-based geometric diodes and examine their applicability for rectenna-based optical detectors and solar cells. For high-frequency applications, a geometric diode offers two distinct advantages. First, unlike the parallel plate structure of the MIM diode, the geometric diode being planar does not have a large capacitance. Second, the absence of a tunnel junction

results in a lower resistance than MIM diodes. These factors facilitate a lower RC time constant for the rectenna circuit and a better impedance matching between the antenna and the diode (Zhu, 2011).

C. Thesis outline

In chapter II, I develop a basic understanding for the properties of MIM diodes and explain the limitations of single-insulator diodes.

MIM diodes are characterized by their resistance and the nonlinearity in their current-voltage $I(V)$ curve. In chapter III, I present a detailed analysis of two mechanisms that facilitate increased nonlinearity through the design of multi-insulator tunnel barriers.

In chapter IV, I analyze the efficiency of the rectenna based on a classical (circuit) model. An in-depth investigation of several issues that influence the efficiency is carried out. These include the RC time constant for the rectenna, the impedance match between the antenna and the diode, and the area of spatially coherent radiation that gets collected by one rectenna element. I estimate the performance of rectenna-based detectors and solar cells using MIM and geometric diodes. Design guidelines and the ultimate (thermodynamically limited) conversion efficiency for rectenna solar cells are also examined.

At optical frequencies, diodes are no longer classical rectifiers. The interaction of the electron with the electromagnetic-wave is quantized at the energy of the photons forming the wave. As presented in chapter V, this requires a modified (semiclassical) theory for investigating the properties of MIM diodes in an optical rectenna. Based on the semiclassical theory, an operating-point analysis of rectenna solar cell is developed and correspondence between the semiclassical and classical theories is established.

In chapter VI, I analyze a technique for improving the infrared bandwidth and efficiency of detectors based on MIM rectennas. This is achieved through an MIM diode-cum-waveguide structure which operates as a distributed rectifier, removing the RC bandwidth limitation.

Fabrication and characterization of graphene geometric diodes is given in chapter VII. Experimentally-measured infrared-characteristics of graphene based antenna-coupled geometric diodes are also presented.

In chapter VIII, I develop a quantum simulation to model the geometric rectifiers made from graphene. The simulations are based on the non-equilibrium Green's function technique (Datta, 2005).

In chapter IX, I carry out a derivation that extends the semiclassical theory presented in chapter V to the domain of non-tunneling based diodes.

Conclusions and directions for future work are discussed in chapter X.

In appendix A, I describe my work on the development of a field effect transistor based on the MIM diode.

CHAPTER II

METAL-INSULATOR TUNNEL DIODES

Metal-insulator-metal (MIM) diodes have shown promise in a variety of high frequency applications, such as frequency measurement in the infrared (Daneu, 1969), as infrared detectors (Grover, 2010), and in the emerging field of terahertz electronics (Estes, 2005). MIM diodes are also used for detection and mixing of radiation in millimeter wave and sub-millimeter wave bands (Fumeaux, 1998), (Abdel-Rahman, 2004), (Rockwell, 2007). These fast advancing technologies require a precise understanding of the operating principles of MIM diodes to accurately predict and improve their characteristics.

The operation of the MIM diode is based on the quantum tunneling of electrons between two metal electrodes that are spaced apart by several nanometers of insulator or a stack of insulators. Tunneling leads to nonlinear current-voltage $I(V)$ characteristics that depend on the shape of the barrier. To estimate this nonlinear $I(V)$, several analyses of MIM diodes have been carried out. To simplify analysis, most of these calculate the tunnel probability using the WKB approximation (Simmons, 1963), (Chapline, 2007) and/or assume the temperature to be 0 K to simplify the Fermi distribution of electrons in the metals (Chapline, 2007), (Guo, 1998), (Hegyi, 2007). These simplifications allow a straightforward and intuitive evaluation of the $I(V)$ characteristics.

As explained in chapter IV, rectennas for high frequency applications require low-resistance diodes. This is achieved by having low barrier-height and thin insulators that allow high-currents to flow through. The approximate methods mentioned above lead to a significant deviation in the predicted characteristics of

low-barrier diodes as compared to a more rigorous solution. Here, I use the transfer-matrix method (TMM) to model the MIM diode and explain the barrier height- and temperature- dependence of tunnel current.

A. MIM diode simulations

1. Methodology

I assume a charge-free oxide region to determine the shape of the tunnel barrier and modify it by image-force barrier lowering. The validity of the image-force lowering is outlined later. The resulting barrier shape is used to calculate the tunnel probability, which along with the Fermi distribution of electrons provides the tunnel current. Consider the barrier shown in Figure II-1 with an arbitrary potential profile.

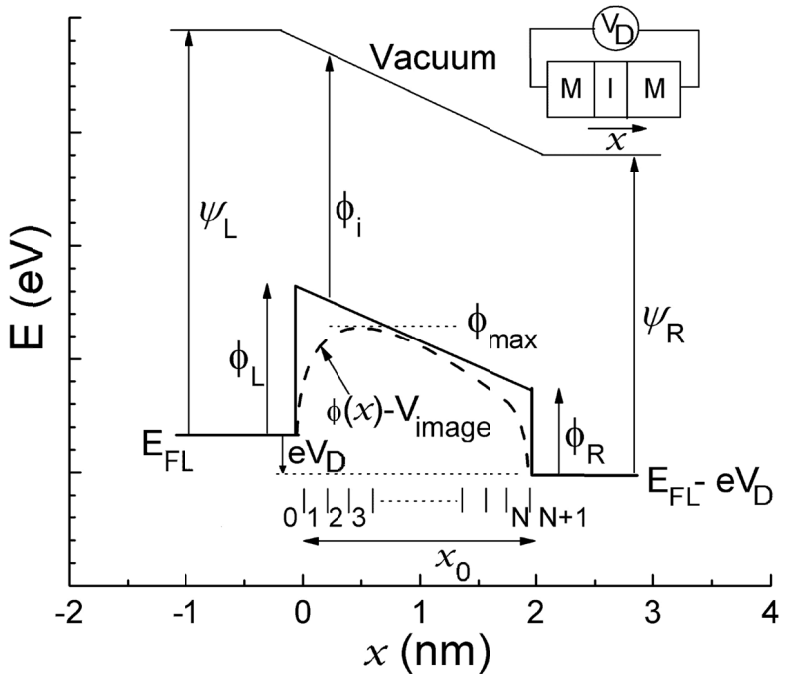


Figure II-1 Energy-band diagram for an asymmetric ($\Phi_L \neq \Phi_R$) tunnel barrier. Here, Φ_i is the electron affinity of the insulator, ψ is the metal work function, and V_D is the voltage applied across the diode. The Fermi level of the left metal electrode (E_{FL}) is held fixed while that of the right electrode varies with applied voltage across the diode. The rectangular barrier is modified by the image-force barrier lowering to give the effective profile (dashed).

An electron with total energy E, has an x-directed component of energy E_x and a transmission probability $T(E_x)$. Assuming an isotropic distribution of electron velocities in the metal electrodes, the formula for the tunnel current from the left (cathode) to the right (anode) electrode can be written as (Simmons, 1963)

$$J_{L\to R}(V_D) = \frac{4\pi m_L e}{h^3}\int_0^\infty T(E_x)dE_x \int_{E_x}^\infty f_L(E)\{1 - f_R(E + eV_D)\}dE$$

Eq. II-1

The Fermi-Dirac distribution functions in the left (f_L) and the right (f_R) metal electrodes are given by

$$f_L(E) = \frac{1}{1 + exp\left(\frac{E - E_{FL}}{kT}\right)}$$

Eq. II-2

$$f_R(E + eV_D) = \frac{1}{1 + exp\left(\frac{E - (E_{FL} - eV_D)}{kT}\right)}$$

Eq. II-3

In Eq. II-1 the inner integral is over all possible total energies E with incident energy E_x for which there are filled states on the left and empty states on the right. The outer integral then multiplies this total number of electrons with the transmission probability $T(E_x)$ and sums the product over all E_x. The net tunnel current is the difference between the currents from the left to the right electrode ($J_{L\to R}$) and from the right to the left electrode ($J_{R\to L}$), where $J_{R\to L}$ can be written in a form similar to that of Eq. II-1. Assuming effective masses in each metal region, $m_L = m_R = m_0$, where m_0 is the electron rest mass, the net current is given by

$$J(V_D) = J_{L\to R} - J_{R\to L} = \frac{4\pi m_0 e}{h^3}\int_0^\infty T(E_x)dE_x \int_{E_x}^\infty \{f_L(E) - f_R(E + eV_D)\}dE$$

Eq. II-4

To calculate the transmission probability I find a plane-wave solution for the Schrödinger equation using the transfer-matrix method (Jonsson, 1990). I divide the tunnel barrier into N steps where $N \sim dV/dx$ and the continuity condition for the wavefunction and its first derivative is applied at each interface. Unlike the basic version of a transfer-matrix method (Eliasson, 2001, p.63), where a large number of 2x2 matrices need to be multiplied, the approach I use combines all the continuity equations into a $(2N+2)$ by $(2N+2)$ near-diagonal square matrix (Probst, 2002). This method prevents round-off errors and provides numerical stability.

Applying the conditions that the amplitude of the incoming wave is unity ($A_0^+ = 1$) and that there is no reflected component in the N+1th region ($A_{N+1}^- = 0$), gives the following relation for the tunnel probability

$$T(E_X) = \frac{k_{N+1}}{k_0} \frac{|A_{N+1}^+|^2}{|A_0^+|^2}$$

Eq. II-5

where, $k_0 = \sqrt{2m_e m_L eE_x}/\hbar$ and $k_{N+1} = \sqrt{2m_e m_R eE_x}/\hbar$.

For calculating $T(E_x)$, I use an adaptive step size for E_x based on the slope of $T(E_x)$. This helps preserve accuracy when required without having a very fine grid throughout. Resonance peaks are accurately tracked with this implementation.

The effective mass of the electron in the insulator (m_e) is assumed to be equal to the rest mass (m_0). This assumption is made in the absence of a more accurate estimate. For crystalline semiconductors, the m_e can be obtained from the band structure (Kittel, 1996, p.209). However, for the grown or deposited amorphous insulators under consideration, a direct experimental measurement is required to determine the effective mass (Solymar, 2010, p.143).

The shape of the potential barrier is determined by the work function of the metals, the electron affinity of the insulators, and the applied voltage. In addition, an electron in the vicinity of a metal experiences an image potential that causes barrier lowering, as given by Eq. II-6 (Sze, 2006, p.146).

$$V_{image}(x) = -\frac{e^2}{16\pi\varepsilon_i\varepsilon_0}\left(\frac{1}{x} + \frac{1}{x_0 - x}\right)$$

Eq. II-6

where x_0 is defined in Figure II-1. The magnitude of lowering is inversely proportional to the distance from the metal surfaces and the insulator dielectric constant ε_i. Eq. II-6 represents a classical concept that is correct for electrons

moving near a metal plane but its application to tunneling, which is a quantum mechanical phenomenon, has been questioned (Hartstein, 1978). For the quantum treatment of image potential, a number of theories have been proposed (Puri, 1983), (Šunjić, 1991). Due to the lack of consensus in what has been proposed in these models, I chose to use the classical result of Eq. II-6. From the results reported by Šunjić it can be seen that the barrier lowering obtained from the quantum mechanical image potential is smaller than that from the classical image potential. For tunnel barriers with a high dielectric constant (ε_i), V_{image} is small, which means that the difference between classical and quantum barrier lowering is minor. Hence the choice of using the classical result for the image force is reasonable.

In the simulations I assume a perfect insulator for a solely tunneling-based analysis of the $I(V)$ characteristics. In an experimental diode, there also can be conduction through defects, surface states and charge build-up at the interfaces that can affect the current. Scattering of the electrons in the insulator also needs to be considered for the thicker diodes.

2. Comparison with other simulation techniques

I now compare the transmission probability calculated by the transfer matrix method (TMM), with those obtained from the WKB approximation, which is dated but has been used recently (Chapline, 2007), and the more current quantum transmitting boundary method (QTBM) (Lent, 1990). Consider a symmetric tunnel barrier with insulator thickness x_0 = 2 nm, $\Phi_L = \Phi_R$ = 0.5 eV and E_{FL}=10 eV, where E_{FL} is the Fermi level referenced to the bottom of the conduction band in the left metal. The electron wavefunction for $E_x < \Phi_{max}$, where Φ_{max} is the highest potential on the modified barrier as shown in Figure II-1, decays with increasing x inside the barrier. The $T(E_x)$ rises sharply with increasing energy, as seen in the plot of Figure II-2(a). When the electron energy (E_x) rises above the barrier ($\Phi_{\max}$), the

wavefunction becomes oscillatory and the transmission probability stays close to unity. For $E_x > \Phi_{max}$, there are resonances in $T(E_x)$ due to interference of the wavefunction inside the insulator.

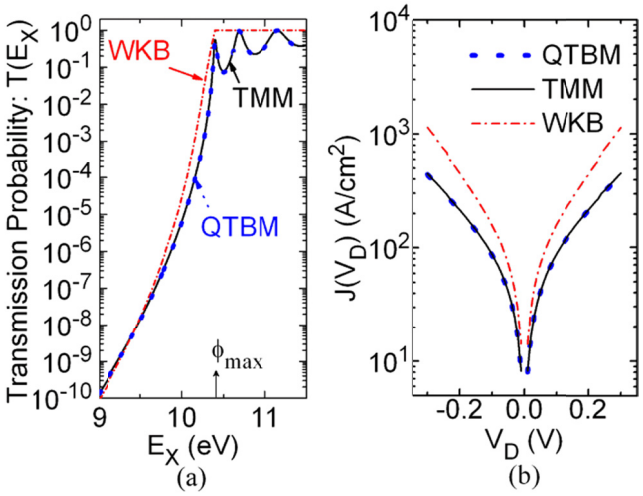


Figure II-2(a) Log-scale plot of the transmission probability $T(E_x)$ vs. the x-directed energy of the electron (E_x) obtained from the transfer matrix (TMM) (solid), the WKB (dot-dash), and the quantum transmitting boundary (QTBM) (dot) methods. The simulated diode is symmetric with barrier heights of 0.5 eV and an insulator thickness of 2 nm. The bias voltage on the diode is 0.3 V. The TMM and the QTBM give an accurate value for $T(E_x)$, while the WKB overestimates $T(E_x)$ for electrons energies near the peak of the barrier (Φ_{max}). (b): Log-scale plot of the simulated $I(V)$ characteristics of the symmetric diode. The TMM and the QTBM results match well, while the WKB predicts a higher current with a larger slope.

The TMM calculation of $T(E_x)$ is in close agreement with the WKB approximation for $E_x < E_{FL}$ (10 eV). However, at higher energy the WKB overestimates the transmission probability and gives $T(E_x) = 1$ for $E_x > \Phi_{max}$. As shown in Figure II-2(b), this results in a significant deviation of $I(V)$ characteristics on using the WKB method. On the other hand, in both the figures, there is no difference between the results obtained from the TMM and the QTBM, both of which give solutions to the Schrödinger equation.

3. Comparison with experimental characteristics

Simulated and experimental characteristics of two asymmetric MIM diodes are compared in Figure II-3. The diodes are made from sputtered insulator and metal layers and the dimensions shown are the targeted thickness of the insulators. The simulated $I(V)$ curves are in good agreement with the measured characteristics.

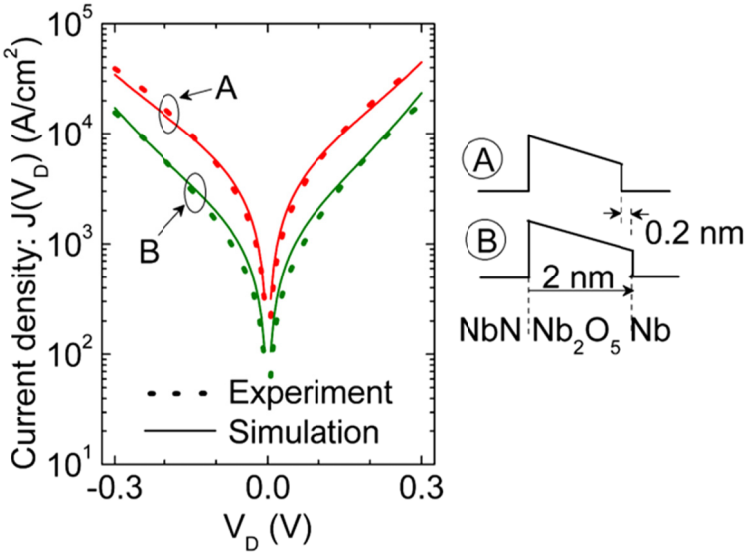


Figure II-3 Comparison of simulated and experimental $I(V)$ characteristics for two MIM diodes. The simulated curves obtained from the transfer matrix method are in close agreement with experimental characteristics. The insulator widths used in the diode simulations are as targeted during deposition.

The simulated and the experimental curves match well without the use of any fitting parameters besides the choice of effective mass equal to the rest mass. The parameters for the materials, used in the simulation and given in the table below, are either book values or extracted from measurements (Phiar, 2007).

Metal	Workfunction (eV)
Nb	4.33
NbN	4.7

Insulator	Electron affinity (eV)	Dielectric constant
Nb2O5	4.23	25
Ta2O5	3.83	20

B. Single-insulator (MIM) diodes

Using the simulation methodology described above, I now analyze the properties of a single-insulator diode in more detail. A closer look at how the barrier-height affects the tunnel current is facilitated by comparing two diodes with different barrier-heights. An understanding for the temperature dependence of tunnel current that arises due to the Fermi distribution is also developed. I extend this analysis to show how the variation of current with temperature decreases for large barrier-heights.

1. Barrier-height dependence of tunnel current

As given by Eq. II-4, the Fermi distribution and the transmission probability of electrons tunneling across the barrier together determine the tunnel current. For two diodes that differ only in their barrier heights, the Fermi distribution is identical but the transmission probability is different. To explain the dependence of tunnel current on the barrier-height, consider a low-barrier (0.5 eV) and a high-barrier (1 eV) diode. Referring to Figure II-1, Φ_{max} is closer to E_{FL} in the low-barrier case. Therefore, as seen in Figure II-2(a), the sharp peak in $T(E_x)$ near $E_x = \Phi_{max}$, adds significantly to the tunnel current only in the low-barrier diode. In the high-barrier diode, this rise in $T(E_x)$ is insignificant, as the Fermi distribution results in a concentration of tunneling electrons that is several orders of magnitude smaller than for the low-barrier. This leads to a smaller tunnel current in the high-barrier diode.

To show that the contribution of electrons near Φ_{max} is dominant only in the low-barrier case, I compare the tunnel currents calculated using the $T(E_x)$ obtained from the plane-wave solution and the WKB approximation. This comparison provides physical insights regarding the sensitivity of the tunnel current to an

inaccurate estimate of $T(E_x)$ near the peak of the barrier. The sensitivity decreases with increasing barrier height as the contribution of electrons near the peak of the barrier reduces. As shown in Figure II-2(a), the WKB overestimates the transmission probability around Φ_{max}, and hence it should give a higher estimate for the tunnel current than the TMM. In Figure II-4 the diode $I(V)$ for the two barrier heights are plotted using the plane-wave (TMM) and WKB methods. In the low-barrier (0.5 eV) diode, WKB predicts higher current than the plane-wave method. However, for the high-barrier diode (1 eV) the WKB and plane-wave results are in agreement.

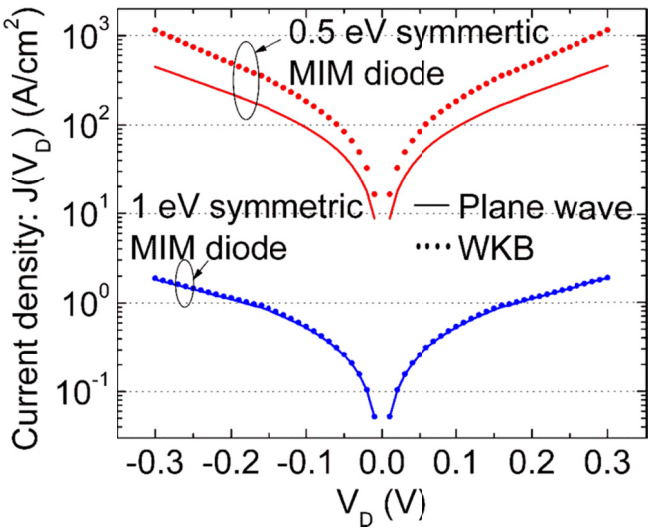


Figure II-4 Current-voltage characteristics of two symmetric MIM diodes having different barrier heights (0.5 eV and 1 eV). The insulator thickness is 2 nm for both diodes. The curves were computed at 300 K using the plane wave method (line) and WKB approximation (dots). The tunnel current in the low-barrier diode has a significant contribution from electrons near the top of the barrier. The WKB, which overestimates $T(E_x)$ in this range, predicts a higher current than the plane wave method. For the high-barrier diode, the two methods give well matched results.

The above results show that in the high-barrier case, the electrons near Φ_{max} have a smaller contribution to the total current. They also show the limited validity of the WKB method, which is accurate only for high-barrier diodes.

2. Temperature dependence of tunnel current

In addition to determining the energy range of electrons that contribute to the tunnel current, the Fermi distribution also determines the temperature dependence of the tunnel current. To explain the variation in tunnel current with temperature, I analyze the 0.5 eV symmetric tunnel barrier of Figure II-4, biased at 0.3 V. The tunnel current at any E_x is the product of the transmission probability and the integral of the difference in Fermi distributions in the two metal electrodes. In Figure II-5, I plot the transmission probability $T(E_x)$, the Fermi distribution $f_L(E_x)$ and the argument of the outer integral in Eq. II-4, $J(E_x)$, calculated using TMM for 77 and 300 K.

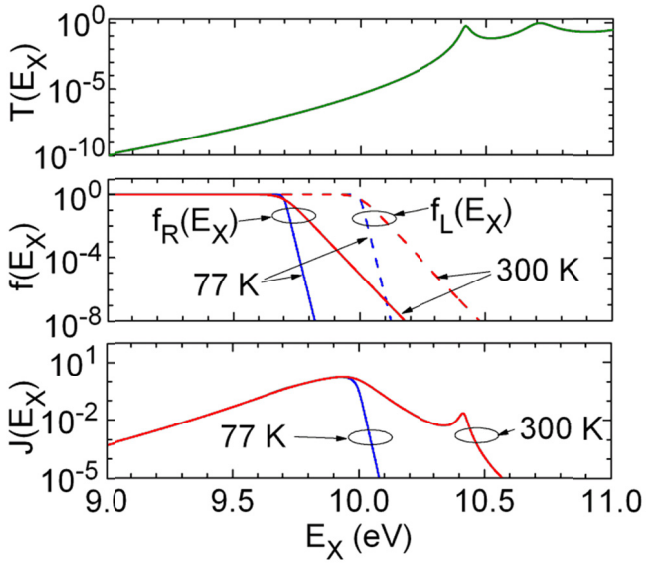


Figure II-5 Contribution to total current density by electrons at different energies $J(E_x)$ (A/cm^2/eV) for a symmetric diode. Transmission probability, $T(E_x)$, is obtained from the plane-wave solution of the Schrödinger equation. Fermi distribution on the left and right metal electrodes, $f_L(E_x)$ and $f_R(E_x)$, respectively, is needed to determine $J(E_x)$. The simulated diode has a barrier height of 0.5 eV and a thickness of 2 nm, and is biased at 0.3 V. Since f_L at 77 K drops faster with energy compared to f_L at 300 K, $J(E_x)$ follows the same trend. A second peak is seen in $J(E_x)$ at 300 K as $T(E_x)$ rises sharply near the top of the barrier, briefly compensating the effect of decreasing f_L. The increase in area under $J(E_x)$, from 77 to 300 K, leads to the temperature dependence of tunnel current.

The $J(E_x)$ is the current density as a function of the x-directed electron energy and its integral gives the total current density. The sharper drop of the $f_L(E_x)$ at 77 K as compared to 300 K is evident. The $f_R(E_x)$ is a left-shifted version of $f_L(E_x)$ on the energy scale by 0.3 eV because of the applied bias of 0.3 V. The $J(E_x)$ is calculated using the corresponding f_L and f_R at both 77 K and 300 K.

The $J(E_x)$ curve at 77 K indicates that a significant contribution to the tunnel current occurs from electrons in the range of $E_{FR} < E_x < E_{FL}$ (9.7 < E_x < 10 eV). However, at room temperature the Fermi function has a larger spread with smaller slope, and so a sharp rise in $T(E_x)$ near the peak of the barrier (10 < E_x < 10.4 eV) adds to the total current. As the temperature rises, the spread in the Fermi

distribution causes a larger contribution from the high (above E_{FL}) energy electrons. Electrons at higher energies have a greater probability of tunneling across the barrier. Therefore, at higher temperature, the increased concentration of high energy electrons and their greater probability of tunneling result in a larger tunnel current.

As the barrier-height of the diode is increased, the peak of the barrier (Φ_{max}) shifts away from E_{FL}. This reduces the Fermi distribution near Φ_{max}, which decreases the contribution of the second peak in $J(E_x)$. Hence the temperature dependence of tunnel current is expected to be smaller for high as opposed to low barrier diodes. To demonstrate this trend, I calculate the tunnel current at various temperatures, and for a range of barrier-heights.

Consider symmetric tunnel junctions ($\Phi_L = \Phi_R = \Phi$) of barrier heights varying from 0.2 to 2 eV. The insulators are 2 nm thick and the temperature is varied from 50 K to 400 K. To compare all barriers on a common scale, the current density is normalized by its value at 50 K for each barrier height and is plotted in Figure II-6. The variation with temperature is larger for smaller barriers. As shown in Figure II-5, this happens because the tail of the Fermi distribution is significant near the top of the barrier, where the transmission probability is also high.

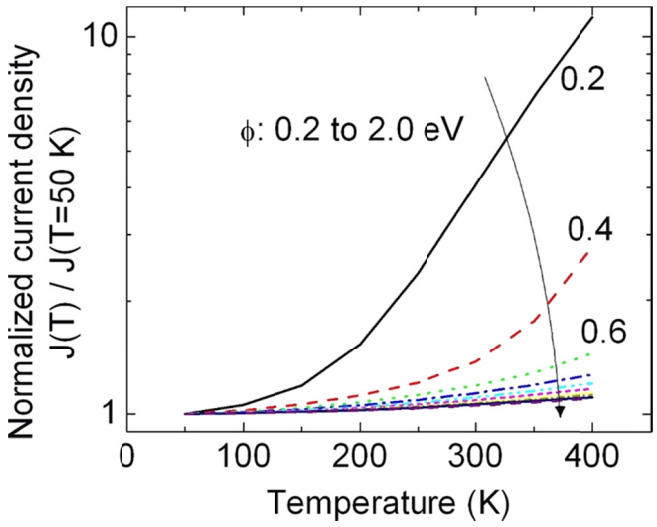


Figure II-6 Normalized current density vs. temperature for a range of barrier heights in a symmetric MIM tunnel diode biased at 0.3 V. The thickness of the insulator is 2 nm. The temperature dependence is larger for smaller barriers. The change in current with barrier-height increases as the temperature rises.

The temperature dependence of current can be reduced to a power law relation of the type

$$y = ax^b + c$$

Eq. II-7

The data of Figure II-6 is fitted to Eq. II-7 and compared with the analytically obtained temperature dependence of the form

$$\frac{J(T \neq 0)}{J(T = 0)} = \frac{\pi c_1 kT}{\sin(\pi c_1 kT)} = 1 + \frac{1}{6}(\pi c_1 kT)^2 + \ldots$$

Eq. II-8

where c_1 is a function of the barrier shape. In Eq. II-8 the coefficient 'b' is 2. The actual power-law temperature dependence of current at low barrier heights is greater than 2 as seen in Figure II-7.

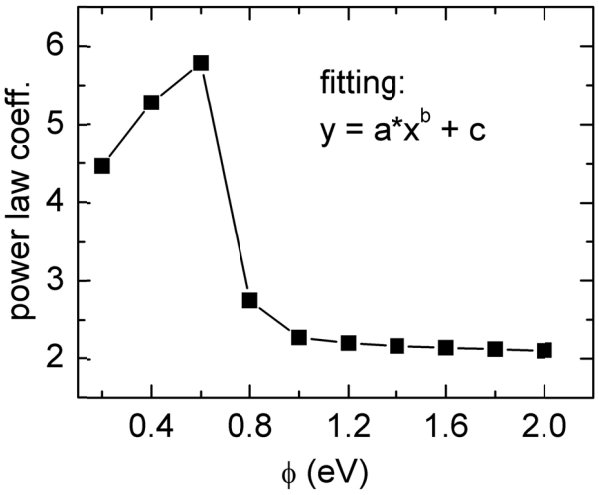


Figure II-7 Power-law coefficient *b* vs. barrier-height from the curve fit for temperature dependence of tunnel current. For the 2 nm diode under consideration, only for barrier heights greater than 0.8 eV does the temperature dependence become quadratic as predicted by the analytical formula in given in Eq. II-8.

The coefficient *c* in Eq. II-7 is 1 within a 5% margin, indicating the accuracy of the curve fitting. Diodes with low barriers have highly temperature dependent tunnel currents. Only for barrier heights greater than 0.8 eV does the behavior start to exhibit the quadratic temperature dependence given in Eq. II-8. This observation is dependent on several variables including the temperature range in consideration and the width of the barrier. Nonetheless, it signifies that the analytical formula does not provide the correct temperature dependence for low-barrier diodes.

C. Shortcomings of MIM diodes

Eliasson (Eliasson, 2001) has extensively analyzed the possible variations of a single-insulator MIM diode. The I(V) characteristics of the diode depend on the shape of the tunnel barrier, which is determined by the metals and insulators used to form the diode. To optimize the MIM diode, the variable parameters are the barrier heights Φ_L, Φ_R and the insulator thickness. The characteristics that need to

be optimized are the differential resistance and responsivity of the diode. The current responsivity is a measure of the diode nonlinearity and is defined at the operating voltage of interest as $\beta_i = I''/(2I')$ (Sanchez, 1978). It signifies the DC current generated in the diode per unit of AC power incident.

A low resistance is necessary to ensure efficient coupling of the diode to the antenna, and is achieved by keeping the barrier heights low (Eliasson, 2001). A high responsivity is required for efficient square-law (small signal) rectification (Sanchez, 1978). Here I analyze these characteristics at zero bias, which reduces the complexity of comparing the resistance and responsivity of several diodes. At zero bias, the responsivity is determined by the degree of asymmetry in the shape of the tunnel barrier, which causes the asymmetry in the $I(V)$ curve. In Figure II-8(a)&(b), I plot the responsivity and resistance vs. the difference in barrier height on the left (Φ_L)and the right (Φ_R). Experimentally, this can be achieved by varying the metal on the left while keeping the insulator and the metal on the right fixed.

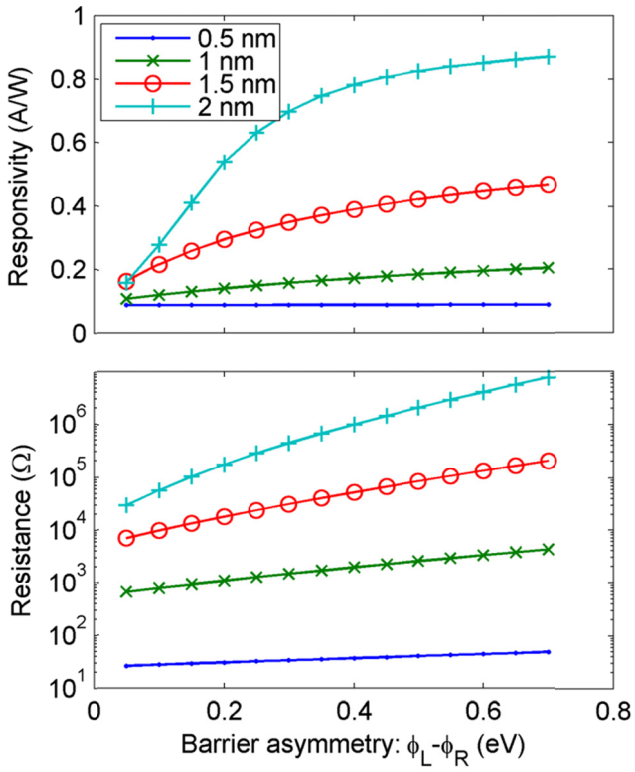


Figure II-8 (a) Responsivity and (b) resistance vs. barrier asymmetry for single-insulator diodes. The diode thickness and the left barrier height (Φ_L) are varied while the right barrier height (Φ_R) is kept fixed at 0.1 eV. The responsivity increases with increase in asymmetry but saturates for high Φ_L-Φ_R. For the same asymmetry, the responsivity is larger for thicker diodes. Increasing asymmetry and increasing thickness, both lead to larger resistance.

As explained in chapter V, an asymmetric $I(V)$ is necessary for self-bias generation and efficient rectification. The zero-bias responsivity is an indicator of this asymmetry. As seen in Figure II-8(a), for a fixed asymmetry, the responsivity is higher for thicker barriers. Also, the thicker barriers show a greater change in responsivity with increasing asymmetry. However, the responsivity saturates at large asymmetry. As the responsivity increases with increasing asymmetry or increasing thickness, so does the resistance as shown in Figure II-8(b). In a rectenna, this negates the improvement in responsivity as the impedance match between the antenna and the diode becomes worse.

The above characteristics are not representative of trends at a non-zero bias. The non-zero bias responsivity of the diodes described above may be sufficient for the operation of a biased detector. Another possibility is that the asymmetry in the $I(V)$ may occur at higher bias voltages instead of zero-bias.

Characteristics of MIM diodes can be improved through the design of multi-insulator tunnel diodes, as discussed in the next chapter.

CHAPTER III

DOUBLE-INSULATOR (MIIM) TUNNEL DIODES

Based on the application, a diode with a high forward-to-reverse current ratio (asymmetry) or a sharp turn-on (nonlinearity) may be required. As analyzed in the previous chapter, low-resistance metal-insulator-metal (MIM) tunnel diodes fail to achieve these requirements. Well engineered multi-insulator diodes can have improved current-voltage $I(V)$ characteristics satisfying both these requirements. In this chapter, I analyze two mechanisms that can improve the performance of multi-insulator diodes. Either of these mechanisms can be made to dominate by the appropriate choice of insulators and variation of barrier thicknesses. Two double-insulator (MIIM) diodes based on these mechanisms are simulated and their characteristics are compared with MIM diodes.

Hegyi et. al. (Hegyi, 2007) conducted a simulation based investigation of parameters for an optimized double-insulator (MIIM) diode. However, their implementation fails to capture the effect of resonant tunneling (Eliasson, 2001), which may significantly alter the diode behavior. In another MIIM configuration (Matsumoto, 1996), an abrupt change in tunnel distance with increasing bias voltage leads to a high forward-to-reverse current ratio. I develop an in-depth understanding of these effects and use them to design MIIM diodes with improved characteristics for high frequency rectennas.

A. Simulation methodology

In chapter II, I gave the framework for simulating MIM diodes using the transfer-matrix method. The same methodology is applicable to a multi-insulator

barrier profile. However, in a multi-insulator diode, the dielectric constants of the insulators play an important role in determining the voltage drop across each insulator layer. To determine the energy-band profile at a certain bias (V_D), I apply the condition for continuity of the electric displacement vector at each insulator interface and obtain the voltage drop across each layer

$$\Delta V_j = (V_D - V_{bi}) \frac{x_j / \varepsilon_j}{\sum x_j / \varepsilon_j}$$

Eq. III-1

where, x_j and ε_j represent the thickness and dielectric constant, respectively, of the j^{th} layer, and $eV_{bi}(=\psi_L - \psi_R)$ is the built-in potential.

In a multi-insulator diode, the effect of the image force is calculated as

$$V_{image}(x) = -\frac{e^2}{16\pi\varepsilon_0}\left(\frac{1}{\int_0^x \varepsilon(x')dx'} + \frac{1}{\int_x^L \varepsilon(x')dx'}\right)$$

Eq. III-2

where, $L = \sum_1^K x_j$ with K being the number of insulator layers. The integrals in the denominator represent the effective distance of an electron from the left or the right metal electrode, while accounting for the changing dielectric constant.

A key effect of interest in multi-insulator diodes is that of resonant tunneling of electrons through a quantum well. The transmission probability calculated here accounts for the resonant tunneling but assumes an empty well. However the resonant level has a certain probability of being filled and is likely to be occupied if its energy is close to or less than the higher metal-Fermi level. This correction is not included in the simulation.

B. Double-insulator configurations

To obtain a high responsivity and low resistance diode, one can design an MIIM barrier with resonant tunneling (Eliasson, 2001), (Moddel, 2004). Alternatively, an MIIM configuration can be designed to have a step-change in tunnel distance (with increasing voltage) for electrons tunneling from the higher Fermi level (Matsumoto, 1996). Both these mechanisms can occur in the same diode and the overall asymmetry of the $I(V)$ curve is regulated by the one that dominates. I examine these effects through the simulation of two double-insulator tunnel diodes.

Consider two MIIM diodes that have the same material configuration but different insulator thicknesses. Diode MIIM1 consists of W-Nb_2O_5(3 nm)-Ta_2O_5(1 nm)-W, and MIIM2 consists of W-Nb_2O_5(1 nm)-Ta_2O_5(1 nm)-W. The material parameters for the oxide were listed in chapter II. The work function of tungsten is 4.55 eV (Camp, 1965). This choice of materials and dimensions is not optimized for maximum nonlinearity or current but rather is chosen to demonstrate the difference between the resonant tunneling dominant in MIIM1 and the step change dominant in MIIM2.

For the two diodes, the conduction band profiles under positive and negative bias are shown in Figure III-1. A quantum well is formed in both the MIIM diodes under positive bias (a) & (b). However, only in the MIIM1, the quantum well is wide enough to have a resonant energy level. On the other hand, under negative bias (c) & (d), the step barrier-profile leads to an abrupt change in the tunneling distance for the electrons near the Fermi level on the right metal-electrode. The Fermi level on the left metal electrode is fixed at 10 eV.

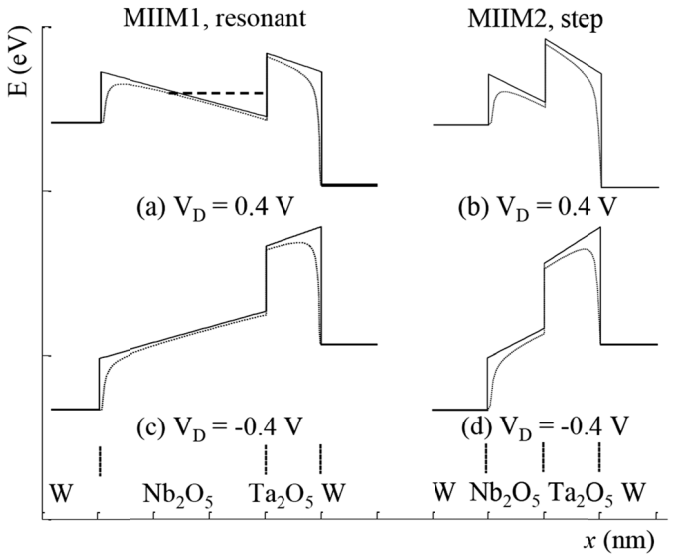


Figure III-1 Energy-band profiles for the resonant and step MIIM diodes. Forward and reverse bias profiles are shown respectively in (a) and (c) for the resonant, and in (b) and (d) for the step diode. The dotted lines show the profiles with barrier lowering. The diode parameters are given in Table I. The thickness of the Nb_2O_5 layer is the only difference between the two diodes.

In Figure III-2, I plot the transmission probability $T(E_x)$ and the current density $J(E_x)$ for the four barrier profiles of Figure III-1. I first discuss $T(E_x)$ comparing the curves on the basis of the barrier shapes and the applied voltages. Due to the magnitude of the bias voltages, the negative bias curves are offset along the energy axis from those at positive bias by approximately 0.4 eV. The fact that the offset is 0.4 eV instead of the difference between the biases (0.8 eV) is explained by the reciprocity of $T(E_x)$. In Figure III-1, if the barrier profiles at negative voltages are mirrored along the vertical, the difference in Fermi levels between the barrier profiles of opposite bias is 0.4 eV and hence the offset. For low electron energies, the transmission probability $T(E_x)$ for the step (MIIM2) diode, represented by curves (b) & (d), is higher than for the resonant (MIIM1) diode, represented by (a) & (c). This is because it is easier to tunnel across a thin barrier as compared to a thick one.

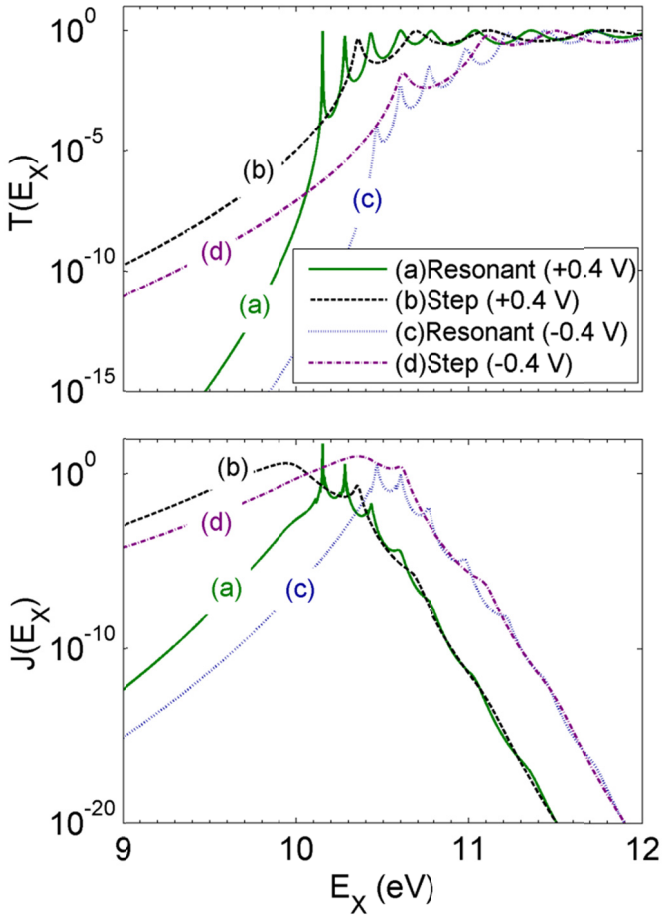


Figure III-2 Electron transmission probability $T(E_x)$ and current density $J(E_x)$ vs. energy for the resonant and step MIIM diodes of Figure III-1. A sharp resonance peak is observed in the resonant diode under forward bias due to the formation of a quantum well.

The resonant diode under positive bias (a) has a sharp rise in $T(E_x)$ exceeding (b) at the resonance peak. Except for the resonance peak, the $T(E_x)$ near the top of the barriers, $E_x > 10$ eV for (a) & (b) and $E_x > 10.4$ eV for (c) & (d), is almost equal for the two diodes. As E_x rises above the highest potential on the low-barrier insulator, the transmission probability exhibits oscillatory behavior for all four cases. In this energy range, the electron wavefunction exponentially decays while tunneling and is sinusoidal through the remaining region of transmission above the conduction band. The sinusoidal wave function resonates with the edges of the

tunnel barrier causing the oscillations in $T(E_x)$. These oscillations modify the probability of tunneling through the higher barrier to give the net $T(E_x)$.

The above trends in $T(E_x)$ influence the current density $J(E_x)$. The area under the resonant diode $J(E_x)$ at positive bias (a) is greater than at negative bias (c). Therefore the current in the resonant diode is greater at positive bias as shown in the $I(V)$ curve of Figure III-3 (a). For the step diode, the area under the $J(E_x)$ curve at negative bias (d) is greater than under positive bias (b). This asymmetry is also seen in the $I(V)$ curve shown in Figure III-3(a). Comparing the resonant and the step diode $J(E_x)$ curves, the narrow resonance peak in (a) is large enough to give a current greater than that in the step diode under positive bias (b) but not enough to exceed the current in the step diode under negative bias (d).

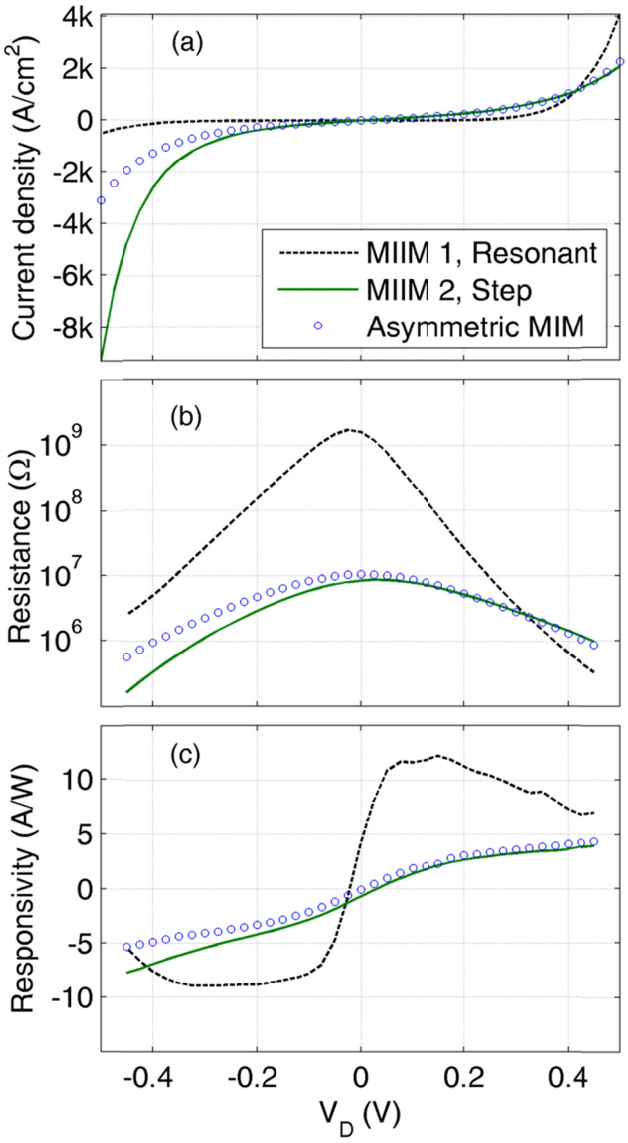


Figure III-3 (a) Current density vs. voltage for the MIIM diodes shown in Figure III-1, and a comparable asymmetric-MIM diode. The step MIIM diode has higher current magnitude under negative bias due to the direct tunneling of electrons across the high-barrier. The resonant MIIM diode has the opposite asymmetry in its I(V) characteristic, due to the formation of resonant quantum well under positive bias. Comparing these with the asymmetric-MIM diode we see that both the MIIM diodes have a smaller resistance (b) and larger nonlinearity (c) in their preferred direction of conduction.

In Figure III-3(a), I also compare the MIIM diodes with an asymmetric-MIM diode that has the barrier heights corresponding to the W-Nb_2O_5 interface on the

left and the Ta_2O_5-W interface on the right and an insulator thickness of 2 nm. The asymmetric-MIM diode is essentially the MIIM2 diode without the abrupt step in the conduction band profile. This is confirmed by their similar current densities under positive bias. However, under negative bias, the step change in tunnel distance in MIIM2 causes a sharp increase in tunnel current. This difference is also evident in the resistance and responsivity curves in Figure III-3(b) and (c) where, under negative bias, the sharp increase in current for MIIM2 leads to a lower resistance and a higher responsivity. The resistance of the resonant diode is significantly higher at zero bias but becomes comparable to the thinner diodes near $V_D = 0.4$ V. The large change in resistance also accounts for the higher magnitude of responsivity. Thus, the nonlinearity improving mechanisms enable MIIM diodes with higher responsivity and lower differential resistance than an equivalent MIM diode of comparable current density.

C. Comparison of MIM and MIIM diodes

The above example shows that just changing the thickness of an insulator in an MIIM diode made with the same pair of materials, can lead to different asymmetry and nonlinearity. It does not suggest which of the mechanisms for achieving larger nonlinearity is preferable. I have analyzed several MIIM diodes designed for implementing these mechanisms and the performance improvement over MIM diodes is observed consistently. The mechanisms exemplified in MIIM diodes can also be applied to barriers with more than two insulators (Korotkov, 1999).

The comparison of thick and thin double-insulator diodes shows that the bias direction causing higher current depends on the electron-transmission-limiting mechanism. Defining positive bias as that which produces or augments a quantum well at the interface between the insulators, if a resonant energy-level is achievable,

the current for this polarity can become larger than that under negative bias. In the absence of a resonant level, the step change in tunneling distance under negative bias causes a larger current than under positive bias. Compared to single-insulator diodes, the resonance and the step mechanisms in double-insulator diodes result in a larger responsivity and a smaller resistance.

In Figure III-4, I compare the resistance vs. responsivity at zero bias for several single- and double-insulator diodes. The simulations show that for comparable resistance values, the responsivity of double-insulator diodes is larger than that of single-insulator diodes. Therefore double-insulator devices are able to achieve both the desirable characteristics, whereas single-insulator diodes are limited in their responsivity.

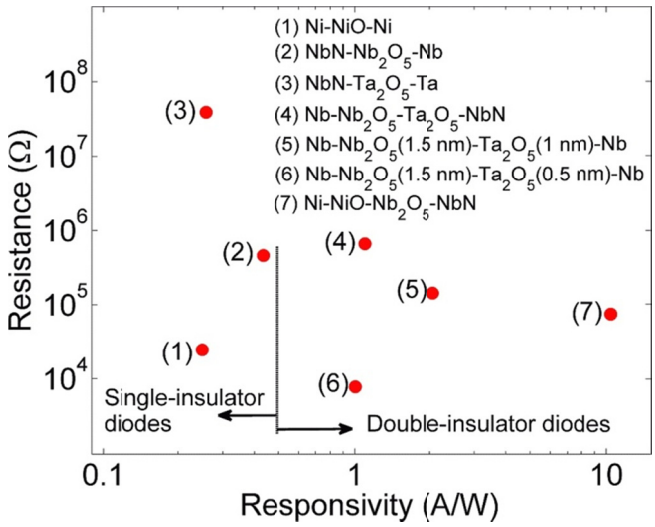


Figure III-4 Resistance vs. responsivity at zero bias for single- and double-insulator diodes. Antenna-coupled diode detectors require high responsivity and low resistance. The double-insulator diodes show improved performance, having smaller resistance and larger responsivity. The area for the diodes is 100x100 nm^2.

D. Interface stability

Sub-micron scale lithography and advanced deposition techniques have enabled the fabrication of metal-insulator diodes with a variety of materials and precise control over layer thicknesses. However, every combination of metals and insulators may not be stable. A thermodynamic analysis of the interface stability is required to ensure that the intended barriers are obtained in an experimental device. For the MIIM diodes discussed in Section III-B, I have carried out a Gibb's free energy analysis (Silberberg, 2007, p.666) for reaction between all interfacial pairs of materials using the FACTsage web software (Bale, 2011). I analyzed each of the pairs at room temperature and at 1000 K and confirmed that no unintended interfacial compounds are formed.

The stability analysis for the Ta_2O_5-Nb interface that occurs in diodes 5 and 6 of Figure III-4 is reproduced here. Using the FACTsage web software (Equilib-web), the equation obtained for the reaction between Nb and Ta_2O_5 at 300 K and 1 atm is

$$5Nb + 1Ta_2O_5 \rightarrow 5NbO + 2Ta$$

Eq. III-3

Nb also forms several other sub-oxides that were not considered. To confirm that the above reaction occurs, a Gibb's free energy calculation can also be made using the formula (Silberberg, 2007, p.666)

$$\Delta G_{sys} = \Delta H_{sys} - T\Delta S_{sys}$$

Eq. III-4

where *sys* denotes the system represented by Eq. III-3 and the Δ denotes change in quantities as reactant-values subtracted from product-values. The enthalpy of formation (ΔH) and the entropy (ΔS) data for the materials is obtained from the CRC handbook. The calculation for ΔG_{sys} is given in the table below.

Material	ΔH (kJ/mol)	ΔS (kJ/mol/K)	$T*\Delta S$ (kJ/mol)	ΔG (kJ/mol)	**multiplier**
Nb	0	0.036	10.8	-10.8	-5
Ta_2O_5	-2046	0.143	42.9	-2088.9	-1
NbO	-405.8	0.048	14.4	-420.2	5
Ta	0	0.042	12.6	-12.6	2

On multiplying ΔG by the multiplier and adding the values for all materials, ΔG_{sys} = 16.7 kJ/mol. For a spontaneous reaction to occur, the ΔG_{sys} should be negative. The calculation implies that no reaction occurs whereas FACTsage indicates that a reaction does occur. One reason for this discrepancy can be the difference in values of parameters used.

The value of ΔG_{sys} obtained here is relatively small compared to enthalpies of individual reactants and products. Even if a spontaneous reaction were to occur, the small ΔG_{sys} implies that it would be slow. This is understandable from the fact that in the solid-state the reactants are not mobile enough.

CHAPTER IV

RECTENNA CIRCUIT & EFFICIENCY

As described in Chapter I, the rectenna consists of an antenna that converts radiation to AC and a diode that rectifies AC to DC. The transport mechanism of the MIM diode is intrinsically fast and allows rectification at optical frequencies. In this chapter, I estimate the performance of a high-frequency rectenna using circuit analysis applicable to both MIM and geometric diodes. In the next chapter, I model the rectenna using a semiclassical quantum-mechanical approach that accounts for the interaction between an electromagnetic field and electrons in an MIM diode. Other than the signal strength, the semiclassical analysis shows that for large photon energies, the photon-assisted transport is dependent on the energy of the photons or the frequency of the incoming wave. The results for the semiclassical resistance and responsivity, from chapter V, are used here.

The performance of a rectifier can be modeled by a linear or a square-law rectification theory (Torrey, 1964). The choice of the relevant model depends on the diode $I(V)$ curve and the magnitude of the AC signal applied across the diode (Eliasson, 2001, p.10). For an AC voltage smaller than the voltage scale on which the diode nonlinearity is evident, the square-law analysis is applicable. For larger AC voltages the linear model is used, where the diode switches between high- and low-resistance states for negative- and positive-voltage cycles respectively.

In optical-rectenna applications, including detection and solar-energy harvesting, a small AC voltage is applied across the diode. This is due to the low intensity of the incoming radiation and the small receiving area of the antenna.

Therefore, the analysis presented in this chapter is based on the square-law rectification theory.

The derivation of the square-law rectification theory leading to the formula for the efficiency of a rectenna is available in several references (Sanchez, 1978), (Eliasson, 2001). Here I develop a deeper insight into the derivation and investigate the factors that contribute to the rectenna efficiency. The requirements for rectennas to operate as efficient detectors and photovoltaic (PV) rectifiers are discussed and calculations for efficiency based on MIM and geometric diodes are presented. Although these diodes are used as examples, the concepts given are more generally applicable. I also provide design guidelines for rectenna solar cells and discuss the thermodynamic limit on their efficiency.

An outline of the circuit analyses presented in this chapter and their key features are given in the list below.

- Coupling efficiency between antenna and diode (section A & C)
    - RC bandwidth limitation
    - Impedance matching requirement
    - Semiclassical diode resistance
- Infrared detector performance (section D.1)
    - Coupling efficiency
    - Semiclassical diode resistance and responsivity
    - Noise
- Power conversion efficiency for rectenna solar cell (section D.2)
    - Single-frequency analysis for AC signal amplitude (section B)
    - Coupling efficiency assumed to be unity
    - Classical diode resistance and responsivity

A. Circuit analysis

Under excitation from an AC voltage source, the instantaneous voltage across a diode is given by

$$V\{t\} = V_D + V_\omega \cos(\omega t)$$

Eq. IV-1

and the corresponding current, to second order, is given by the Taylor expansion

$$I\{V\{t\}\} = I\{V\{t\}\} + I'\{V_D\} V_\omega \cos(\omega t) + I''\{V_D\} \left(V_\omega \cos(\omega t)\right)^2 / 2$$

Eq. IV-2

The time averaged DC current at the diode terminals is given by

$$\begin{aligned} I_{diode} &= \left\langle I\{V\{t\}\} \right\rangle = I\{V_D\} + \frac{V_\omega^2 I''\{V_D\}}{4} \\ &= I\{V_D\} + \beta\{V_D\} \frac{V_\omega^2}{2R_D\{V_D\}} = I\{V_D\} - I_{DC}^{\omega}\{V_D\} \end{aligned}$$

Eq. IV-3

in which the first component corresponds to the DC resistance of the diode $R_{diode} = V_D/I\{V_D\}$ while the second component is the DC current generated due to rectification. Without a DC load across the diode, the two currents are equal and opposite and the net DC power in the diode is zero. The time averaged total power in the diode is given by

$$P_{diode} = \left\langle V\{t\} I\{V\{t\}\} \right\rangle = V_D I\{V_D\} - V_D I_{DC}^{\omega}\{V_D\} + \frac{V_\omega^2}{2R_D\{V_D\}}$$

Eq. IV-4

where the first two terms correspond to the DC current while the third term gives the AC power dissipation. In the presence to an external DC load, the power available from rectification is dissipated in the R_{diode} and R_{load}.

The efficiency (η) of a rectenna is determined by the combination of several factors (Kale, 1985) given by

$$\eta = \eta_a \eta_s \eta_c \eta_j$$

Eq. IV-5

where, η_a is the efficiency of coupling the incident EM radiation to the antenna; η_s is the efficiency of propagating the collected energy to the junction of the antenna and the diode; η_c is the coupling efficiency between the antenna and the diode; and η_j is the efficiency of rectifying the power received in the diode. The efficiency of the diode junction can be expressed in terms of its current responsivity $\eta_j = \beta_i$, which was discussed in Chapter II. This sets the overall units of η to be A/W implying the DC current produced per watt of incident radiation. I normalize the overall efficiency with respect to the antenna efficiencies (η_a and η_s).

The antenna-to-diode power-coupling efficiency (η_c) is given by the ratio of the AC power delivered to the diode resistance to the power of the antenna voltage source. This ratio can be calculated from the analysis of the small-signal circuit of the rectenna shown in Figure IV-1. The antenna is modeled by a Thévenin equivalent and the diode by the parallel combination of a capacitor and a voltage-dependent resistor.

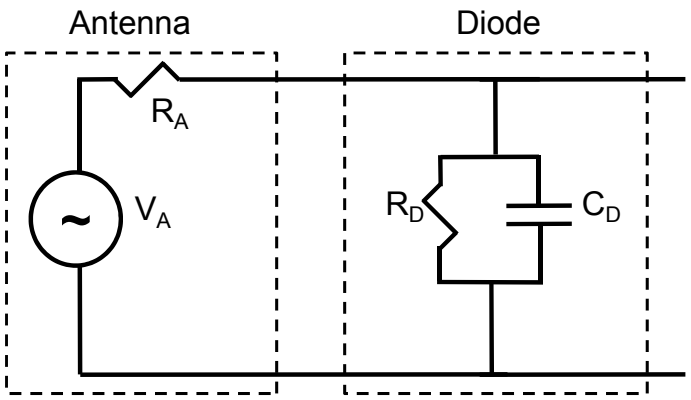


Figure IV-1 Small signal circuit model of the rectenna for determining coupling efficiency. The antenna is modeled as a voltage source in series with a resistance and the MIM diode is modeled as a resistor in parallel with a capacitor.

Assuming that the input power is delivered to the diode as AC power, the power coupling efficiency at a frequency ω is given by (Sanchez, 1978)

$$\eta_c = \frac{P_{AC,R_D}}{P_A} = \frac{4\dfrac{R_A R_D}{(R_A + R_D)^2}}{1 + \left(\omega \dfrac{R_A R_D}{(R_A + R_D)} C_D\right)^2}$$

Eq. IV-6

where, $P_A = V_A^2 / (8R_A)$. In the above equation, the numerator gives the impedance match between the antenna and the diode.

In Figure IV-1, if the capacitive branch is open-circuit due to a small capacitance or low frequency, the circuit is essentially a voltage divider between R_A and R_D. The denominator in Eq. IV-6 determines the cutoff frequency of the rectenna, which is based on the RC time constant determined by the parallel combination of the antenna and the diode resistance and the diode capacitance. Above the cutoff frequency, the capacitive impedance of the diode is smaller than the parallel resistance leading to inefficient coupling of power from the antenna to the diode resistor.

The responsivity form of the overall efficiency, given by $\eta = \eta_c \beta_i$, is a measure for the performance of a detector and indicates the current or voltage produced per watt of incident radiation.

In a PV rectifier, the performance measure of interest is the power-conversion efficiency (η_{load}) which is given by the ratio of the DC power delivered to the load and the incident AC power (Eliasson, 2001)

$$\eta_{load} = \frac{P_{load}}{P_A} = \frac{I_{DC,load}^2 R_{load}}{P_A}$$

Eq. IV-7

The $I_{DC,load}$ is calculated from the current division of $I_{DC}{}^{\omega}$ between R_{load} and R_{diode}, giving the expression for efficiency

$$\eta_{load} \propto \beta_i^2 P_A \eta_c^2$$

Eq. IV-8

From Eq. IV-8, the power-conversion efficiency depends on four factors, namely, the diode responsivity, the strength of the AC signal that depends on the power received by the rectenna, the impedance match between the antenna and the diode, and the RC time constant of the circuit. Another component that decreases the efficiency of the rectenna is the series resistance of the leads from the antenna to the diode (Yoo, 1992). However, its effect is ignored here.

B. Power received by a rectenna

In the efficiency formulas derived above, it is assumed that the radiation is at a single frequency. To use this analysis for a broadband spectrum, I approximate it by a single-frequency source of equivalent strength. Under this assumption, I calculate the strength of the voltage source equivalent to the solar spectrum (100 mW/cm^2) incident on an antenna of 1 μm^2 area. The area over which an antenna receives

radiation is proportional to the wavelength squared. Using a nominal R_A = 100 Ω, an order-of-magnitude calculation gives $V_A \approx 1$ mV. The low magnitude of the voltage source agrees with the choice of using a small-signal circuit analysis.

The assumption regarding the equivalence of a spectrum to a single-frequency source is undone in Chapter V, where the current-voltage I(V) characteristics of the diodes illuminated at optical frequencies are modeled using a semiclassical approach that accounts for photon energy of the radiation. In the case of a PV rectenna, a higher V_A leads to a greater power coupling efficiency, as given by Eq. IV-8. Also, for a detector, even though the responsivity does not change with area, the noise performance or the detectivity improves on increasing the receiving area (Grover, 2010). Possible methods of increasing the area include using concentrators or feeding the signal from an array of antennas into a diode. The second arrangement is less practical as metals at optical frequency have significant resistive losses (González, 2005). The use of dielectric antennas is proposed to overcome this limitation (Sarehraz, 2005).

However, the area from which radiation is fed into one diode cannot be made arbitrarily large. Spatial coherence of the incoming electromagnetic spectrum determines the maximum area that can be coupled to a diode. Sources of interest for rectennas are the sun for photovoltaics, recycled heat for thermo-photovoltaics, humans or animals for infrared (IR) detectors in automotive, to name a few. All these sources are approximated by a black body spectrum at the relevant temperature and the radiation emitted from different parts of the source is statistically independent implying spatial incoherence. However, radiation from an extended, incoherent source gains a certain degree of coherence as it propagates. I use the example of the sun and calculate the distance on earth over which the solar radiation is coherent.

The degree of coherence for a quasi-monochromatic source is given by the van Cittert-Zernike theorem (Born, 1999, p.572) that calculates the correlation in field oscillations (mutual intensity) at two spatially separated points illuminated by the extended source as shown in Figure IV-1. I apply this theorem to a black body spectrum, which can be treated as quasi-monochromatic (Marchenko, 2007, p.321).

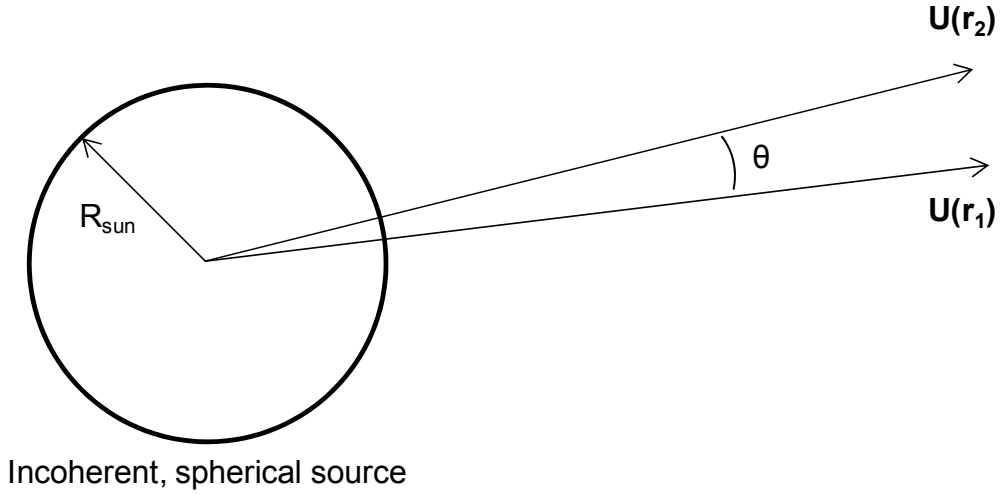


Figure IV-2 Radiation from an incoherent spherical source gains a certain degree of coherence as it propagates. This coherence is calculated from the mutual intensity of the field U at locations $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$, separated by an angle θ subtended at the center of the source.

The van Cittert-Zernike theorem is valid in the far-zone of a source, defined by

$$\frac{\pi R_{sun}^2}{\lambda r} \ll 1$$

Eq. IV-9

where λ is the wavelength of the source and r is the distance from it. The radius of the sun is approximately $7\text{x}10^5$ km and at $\lambda = 500$ nm, the far zone of the sun is at a distance greater than $3\text{x}10^{21}$ km. The distance between the sun and the earth is only $1.5\text{x}10^8$ km ($l_{s\text{-}e}$), implying that the earth is in the near-zone of the sun. However, numerical calculation of the degree of coherence without using the van-

Cittert Zernike theorem shows that the theorem is applicable even in the near field (Agarwal, 2004). Coherence in radiation emerges as close as a few wavelengths away from the sun.

Neglecting the variation in the brightness of the sun, the degree of coherence is given by

$$\mu(\theta) = \frac{2J_1\left(kR_{sun}\sin(\theta/2)\right)}{kR_{sun}\sin(\theta/2)}$$

Eq. IV-10

where, $k = 2\pi/\lambda$ and J_1 is Bessel function of first kind. The dependence of μ on θ is plotted in Figure IV-3.

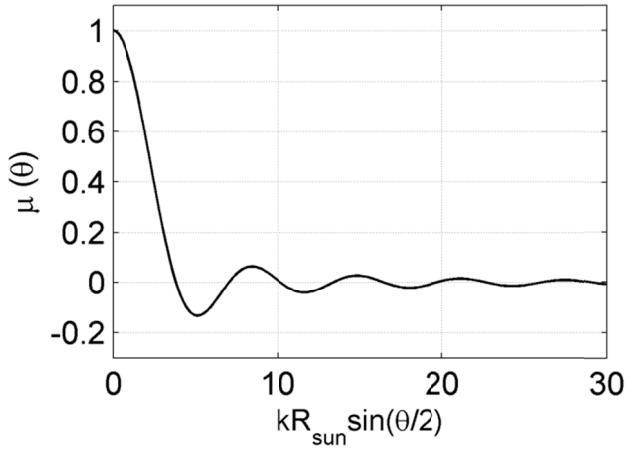


Figure IV-3 Degree of coherence plotted as a function of angular separation between two points outside the source.

The angle over which the radiation from the sun has a degree of coherence greater than 0.88 is given by

$$\theta = 2\sin^{-1}\left(\frac{1}{kR_{sun}}\right)$$

Eq. IV-11

From this the diameter of the circular area at earth over which sunlight is assumed coherent is calculated as

$$d_c = \frac{0.16 l_{s-e} \lambda}{R_{sun}}$$

Eq. IV-12

This gives d_c = 19 μm and a spatially coherent area of 283 μm^2. With the increased area, the strength of the antenna voltage source discussed earlier is increased to $V_A \approx 16$ mV, which is still within the assumption of a small signal analysis.

For IR detection of a source 1 m across, emitting a peak wavelength of 10 μm, and located 50 m away, I apply Eq. IV-12 to obtain an approximate value for the coherence distance on the detector plane. Substituting $l_{s\text{-}e}$ = 50 m and R_{sun} = 1 m, a d_c = 90 μm is obtained.

C. Impedance match and cutoff frequency

Efficient coupling of power from the antenna to the diode requires impedance matching between them. Moreover, having a small RC time constant for the circuit implies that the product of the antenna resistance (R_A) in parallel with the diode resistance (R_D) and the diode capacitance (C_D) must be smaller than the time period ($2\pi/\omega$) of radiation incident on the rectenna. This ensures that the signal from the antenna drops across the diode resistor (R_D) and is not shorted out by C_D. This condition needs to be satisfied in conjunction with $R_D = R_A$ to obtain $\eta_c = 1$.

The parameters that can be varied to achieve these conditions are the diode area, the antenna resistance, and the composition of the diode. A less resistive diode will give a higher $\eta_{coupling}$. Therefore, for this analysis I choose the Ni/NiO(1.5 nm)/Ni MIM diode, which has an extremely low resistance at zero bias and was

used in several high-frequency rectennas (Fumeaux, 1998), (Wilke, 1994), (Hobbs, 2007). The $\eta_{coupling}$ analysis for any other practical diode is likely to be worse than the present case. Here I am disregarding the fact that low resistance and/or symmetric single-insulator diodes generally have a poor responsivity at zero bias, as was shown in Chapter II.

The diode resistance is calculated classically and semiclassically, where the semiclassical resistance accounts for the energy of the optical photons. This concept is discussed in greater detail in Chapter V.

I choose a nominal antenna impedance of 377 Ω and vary the diode area. In Figure IV-4, the coupling efficiency η_c is plotted vs. the diode edge length for a classically and semiclassically calculated diode resistance.

The semiclassical resistance is lower than the classical resistance and gives a higher η_c. The peak in both the curves occurs at the same edge length, and is an outcome of the balance between the needs for impedance matching and low cutoff frequency. In either case, the maximum efficiency is much smaller than unity.

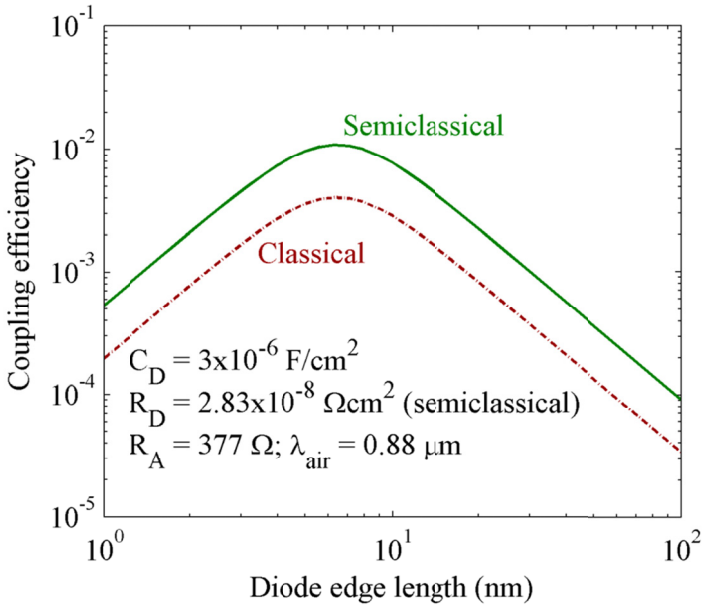


Figure IV-4 Effect of varying the diode size on the antenna to diode coupling efficiency. The peak in the efficiency occurs due to the tradeoff between impedance match and cutoff frequency. The resistance of the Ni-NiO(1.5 nm)-Ni diode is calculated from its simulated $I(V)$ curve classically and semiclassically (for a photon energy E_{ph}=1.4 eV, or λ_{air}=0.88 μm) as discussed in Chapter V. A Ni-NiO barrier height of 0.2 eV is used in the simulation (Hobbs, 2007).

To understand the coupling efficiency better, one can separate the effects of impedance matching given by the numerator (ideally R_D/R_A=1) and cutoff frequency given by the denominator (ideally $\omega(R_A||R_D)C_D$=0) in Eq. IV-6. Unity coupling efficiency under these conditions occurs for different edge lengths, as shown in Figure IV-5. The overall efficiency is limited to be smaller than the individual efficiency curves, leading to the peak in Figure IV-4.

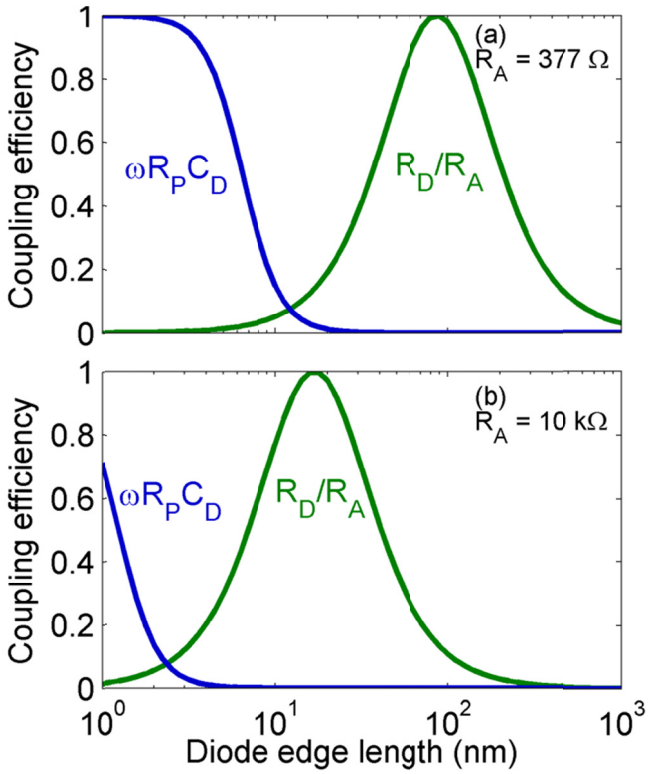


Figure IV-5 Antenna-to-diode coupling efficiency as a function of diode edge length, separating the effect of impedance match from cutoff frequency for two antenna impedance values: (a) R_A=377 Ω, and (b) R_A=10 kΩ. R_P denotes the parallel combination of R_A and R_D. The curves labeled $\omega R_P C_D$ show the coupling efficiency when only the cutoff frequency is the limiting factor, and those labeled R_D/R_A show the coupling efficiency when only the impedance match is the limiting factor. The overall efficiency remains smaller than either curve.

The tradeoff between impedance match to the antenna, for which a small R_D is desired, and a high cutoff frequency, for which a small C_D is desired, is fundamental for parallel plate devices. Even increasing the antenna impedance, if feasible, does not help. In such a case a higher R_D can be accommodated, allowing the diode area to be smaller, and resulting in a desirable smaller C_D. Unfortunately the higher R_A also increases the $(R_A||R_D)C_D$ time constant. This is shown in Figure IV-5(b), where the higher antenna resistance shifts both curves to the left with no improvement in the coupling efficiency. In other words, to improve the coupling efficiency the parameters in Eq. IV-6 would need to be adjusted so that the $\omega R_P C_D$

curve shifts to the right and R_D/R_A curve shifts to the left. The condition under which the constraints simultaneously lead to a high coupling efficiency is obtained by combining

$$\omega(R_A \| R_D)C_D << 1 \;\; \& \;\; \frac{R_D}{R_A} = 1 \;\Rightarrow\; R_D C_D << \frac{2}{\omega}$$

Eq. IV-13

For the model Ni-NiO-Ni diode discussed above, this condition is not satisfied for near-IR-light frequencies (λ = 0.88 μm), where $2/\omega$=9.4x10^{-16} s is much smaller than R_DC_D=8.5x10^{-14} s. It is satisfied for wavelengths greater than 80 μm.

Due to the parallel-plate structure of the MIM diodes, the R_DC_D time constant is independent of the diode area and is determined solely by the composition of the MIM diode. As already noted, the Ni/NiO/Ni diode is an extremely low resistance diode and NiO has a small relative dielectric constant (ε_r) of 17 at 30 THz. Even if one could substitute the oxide with a material having comparable resistance and lower capacitance (best case of ε_r=1), the R_DC_D would still be off by an order of magnitude. Putting practicality aside completely, a near-ideal resistance would result from a breakdown-level current density of 10^7 A/cm^2 at, say, 0.1 V, giving a resistance of 10^{-8} Ω-cm^2. A near-ideal capacitance would result from a vacuum dielectric separated by a relatively large 10 nm, giving a capacitance of ~10^{-7} F/cm^2. The resulting R_DC_D would be ~10^{-15} s, again too large for efficient coupling at visible wavelengths.

On the other hand, the planar structure of the geometric diode results in a very low capacitance while maintaining a low resistance. The capacitance of the geometric diode can be calculated as the parallel combination of two capacitance elements: the capacitance between the arrow-shaped conductor on one side of the neck and the square area on the other side and the quantum capacitance of the neck region. The total capacitance is calculated to be 0.06 fF (Zhu, 2011). Since the

resistance of graphene can be adjusted by doping, the impedance of the diode can be tailored to match the antenna impedance, which is about 200 ohms. For these values the time constant for the geometric diode rectenna is $R_D C_D = 7 \times 10^{-15}$ s. These calculations are based on an initial fabricated device and not for an optimized structure or the best graphene possible. Despite that, the RC time constant for the geometric diode is a factor of 10 lower than that of the Ni-NiO-Ni MIM diode, which implies that the geometric diode under consideration is not RC limited for infrared detection in the wavelength range of 8 to 14 μm.

The coupling efficiency of MIM-diode rectennas can be improved under some circumstances. They become efficient at longer wavelengths, where the condition imposed by Eq. IV-13 is easier to meet. The $R_D C_D$ can also be artificially reduced by compensating the capacitance of the MIM diode with an inductive element, but this is difficult to achieve over a broad spectrum.

In the next section, I analyze the performance of rectennas for infrared detection and photovoltaic rectification.

D. Rectenna applications

In this section, I use diode characteristics obtained from DC $I(V)$ curves, and project the performance capability of rectenna-based IR detectors and solar cells.

1. Infrared detector

A major application for a low-cost IR detector technology is night vision systems for use in automobiles. Adequate performance and low cost has led bolometers (Yon, 2003) to be popularly used in these systems. More recently, the suspended thermo-diode technology (Reinhart, 2009) has shown promise in providing a low-cost FIR detector. Rectennas are an alternative technology whose straightforward fabrication and integrability with CMOS imply low-cost of manufacturing. Here, I compare

rectennas with the currently prevalent options, and examine the noise equivalent temperature difference (NETD) of some of these detectors. A basic requirement for automotive night vision is to have a 10 μm detector with an NETD of less than 300 mK (ADOSE, 2008).

The NETD is calculated as (Rogalski, 2000)

$$NETD = \frac{I_{noise}\sqrt{BW}}{\eta \, dP/dT}$$

Eq. IV-14

where, η is the detector responsivity, I_{noise} is the noise current assumed to be 4 pA/√Hz arising due to the low-noise amplifier following the rectenna. A video bandwidth (BW) of 30 Hz is used in the calculations, and dP/dT refers to the thermal variation of the spectral emittance.

The dP/dT is calculated by subtracting the integral of Planck's law at two temperatures with ΔT = 1 K, in the bandwidth of interest. Here, the Rayleigh-Jeans limit of Planck's law cannot be used to calculate dP/dT (Brown, 2004) as that requires $\hbar\omega/kT \ll 1$. Doing so would result in a dP/dT that is two orders of magnitude larger. For a background emissivity (σ) = 1, detector area = 35x35 μm^2 and in the spectral range of 8 to 14 μm, dP/dT = 3.185 nW/K. The effective area over which the antenna receives radiation is chosen to match the area of a bolometer element.

The detector responsivity (η) accounts for the antenna efficiency assumed to be 50% ($\eta_a\eta_s$ = 0.5) (González, 2005), the antenna to diode coupling efficiency (η_c), and the current responsivity of the diode (β_i). I calculate the responsivity of the detectors based on an MIM diode and a geometric diode, at an operating wavelength of 10.6 μm. To estimate the capacitance of the MIM diode, a relative dielectric constant of 17 is used for NiO. The capacitance of the geometric diode is estimate from a strip-

line and a quantum self-capacitance (Zhu, 2011). The resistance and responsivity values are calculated from the DC $I(V)$ curves at zero bias.

Table IV-1: Estimated of detection responsivity for MIM and geometric diode.

Diode type	R_D **(Ω)**	C_D **(F)**	R_A **(Ω)**	η_c	β_i **(A/W)**	η **(A/W)**
MIM (Ni-NiO-Ni)*	100	$2.3\text{x}10^{-15}$	100	0.002	0.05	$5\text{x}10^{-5}$
Geometric diode^	200	$3.4\text{x}10^{-17}$	200	0.71	0.3	0.2

*(Fumeaux, 1998), ^ (Zhu, 2011)

The responsivity of the geometric diode is higher due to a lower capacitance and a higher current responsivity (β_i).

Using the responsivity estimate from above, I compare the performance of rectennas based on the geometric and MIM diode with the suspended thermo-diode and bolometer.

Table IV-2: Comparison of NETD for zero-bias rectenna detectors with other technologies.

	Rectenna Geometric	**Rectenna** MIM	**Suspended thermo-diode** (Reinhart, 2009)	**Bolometer** (Yon, 2003)
NETD	34 mK	137 K	< 300 mK	56 mK
Receiving area	35x35 μm^2	35x35 μm^2	225 µm (pitch)	35x35 μm^2
Detector responsivity	0.2 A/W	$5\text{x}10^{-5}$ A/W	150 V/W	
Noise current	4 pA/√Hz	4 pA/√Hz		

The above analysis for the rectennas is for zero-bias across the diode. This removes the diode shot noise and the noise due to the biasing power supply, thereby

minimizing noise current. In situations where the gain in responsivity on operating the detector at a non-zero voltage outweighs the increase in noise, a biased configuration can be advantageous.

MIM traveling-wave detector

Performance improvement of MIM based detectors is possible on using a traveling-wave structure (Estes, 2006), (Grover, 2010). This design circumvents the restrictions imposed on the MIM diode by the coupling efficiency. Akin to a transmission line where the geometry determines the impedance, the distributed RC enhances the coupling between the antenna and the traveling-wave structure. However, losses in the metallic regions of the waveguide limit its efficiency as the frequency approaches that of visible light. A detailed analysis for the traveling-wave IR detector along with a comparison to the lumped-element rectenna discussed above is given in chapter VI.

2. Solar cell

The efficiency of a rectenna solar cell given by Eq. IV-7 assumes a bias across the diode. Another approach to get a better estimate for the efficiency is to consider a self-bias voltage developed by the rectifier. The diode parameters that are used in the efficiency calculation depend on this DC voltage (V_{DC}). As with the efficiency, the DC voltage depends on the power incident on the rectenna (P_{in}).

Here I ignore the frequency response and the impedance matching requirements of the rectenna and analyze the efficiency of a rectenna solar cell from the diode perspective. The Norton equivalent circuit for the solar cell at DC is shown in Figure IV-6. Assuming that the rectenna responsivity in Eq. IV-5 depends only on the diode responsivity $\eta_j = \beta$, the current source due to rectification, for a diode voltage V_D, is given by

$$I_{DC}^{\omega} = \beta\{V_D\}\frac{V_{\omega}^2}{2R_D\{V_D\}}$$

Eq. IV-15

The operating voltage is obtained from the equivalent circuit as

$$V_D = \left(R_{diode} \parallel R_{load}\right)\beta\{V_D\}\frac{V_{\omega}^2}{2R_D\{V_D\}}$$

Eq. IV-16

The operating point for the diode is determined by the load resistance (R_{load}). For maximum transfer of DC power to the load, $R_{diode} >> R_{load}$, giving

$$V_D = R_{load}\beta\{V_D\}\frac{V_{\omega}^2}{2R_D\{V_D\}}$$

Eq. IV-17

The above equation is a transcendental equation that can be solved numerically to obtain the DC voltage for a given diode $I(V)$ characteristics and incident power.

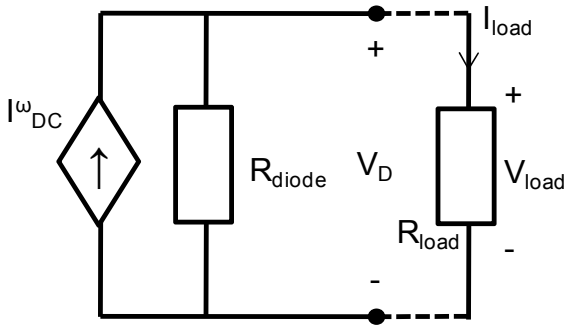


Figure IV-6 Norton equivalent of the rectenna solar cell. The diode (dark) resistance is not a part of the current source.

For the set of constraints outlined above, the power conversion efficiency for the solar cell η_{load} is given by

$$\eta_{load} = \frac{I_{load} V_{load}}{P_{in}} = \frac{\left|V_D \beta\{V_D\}\right| \frac{V_\omega^2}{2R_D\{V_D\}}}{\left(1+\left|V_D \beta\{V_D\}\right|\right) \frac{V_\omega^2}{2R_D\{V_D\}}} = \frac{\left|V_D \beta\{V_D\}\right|}{\left(1+\left|V_D \beta\{V_D\}\right|\right)}$$

Eq. IV-18

The above equation for efficiency is not independent of incident power, as there is an indirect dependence through V_D. The efficiency is limited by the maximum value of $|V_D\beta|$ which is smaller than unity as discussed from a quantum efficiency perspective in the next chapter. Therefore, the maximum efficiency for a small-signal classical rectifier is 50%.

The results presented here are obtained from a circuit analysis that uses classical diode parameters. For an exponential $I(V)$, the responsivity (β) is approximately e/kT. The maximum responsivity corresponding to unity quantum efficiency is an electron per photon, which imples that $\beta_{max}=e/\hbar\omega$. Therefore, for at high frequencies, a more accurate model for the rectenna solar cell is given by the semiclassically derived diode parameters and $I(V)$ curve under illumination. This is presented in the next chapter.

E. Rectenna design for solar energy harvesting

The above calculation for the solar cell efficiency assumed rectification of the total power available in the solar spectrum. An actual rectenna would be limited by the bandwidth received by the antenna and the central frequency it is designed to operate at. The spectral irradiance of the sun received on the earth can be written for a fractional bandwidth as (Moddel, 2001)

$$J' = \frac{2\pi\upsilon^4}{c^2\left(\exp\left(h\upsilon/kT\right)-1\right)}\left(\frac{R_{sun}}{l_{s-e}}\right)^2$$

Eq. IV-19

The intensity distribution normalized to the maximum at 633 nm is plotted in Figure IV-7. A narrowband solar rectenna designed to operate at a central wavelength of 633 nm allows efficient energy-harvesting.

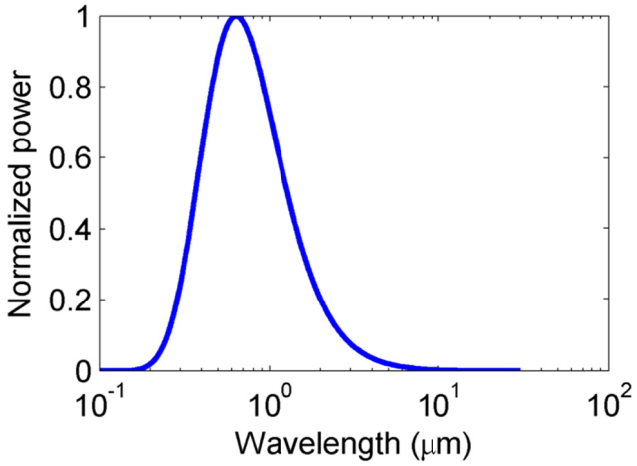


Figure IV-7 Normalized distribution of power in a black body spectrum as a function of wavelength.

As calculated in Section IV.C, the wavelength at which a rectenna operates efficiently is determined by the RC time constant of the diode. In Figure IV-8, I calculate the fraction of solar energy carried by wavelengths greater than the cutoff to find out the maximum wavelength up to which solar thermo-photovoltaic rectennas are beneficial. For example, if the rectenna harvests wavelengths greater than 4 μm, only 1% of the total solar energy is harvested.

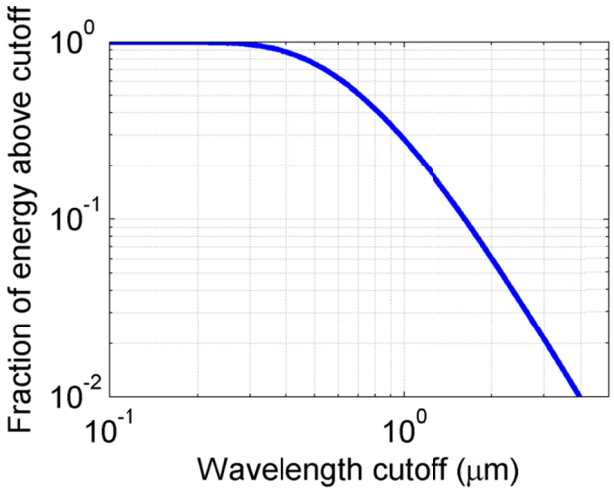


Figure IV-8 Fraction of solar energy as a function of minimum wavelength that can be harvested efficiently.

From Figure IV-8, it is evident that 60% of the solar energy exists at wavelengths less than 0.8 μm. For an infrared solar cell, operating above 0.8 μm, 40% of the total solar energy is available for harvesting.

F. Ultimate efficiency of rectenna solar cells

The maximum efficiency with which the energy radiated by a black body source can be converted to electrical energy is given by the Landsberg efficiency (Landsberg, 1979). In this thermodynamic model for efficiency the entropy of the photons generated at a high temperature is discarded at a low temperature resulting in a corresponding loss of energy. This gives rise to the temperature dependence for the efficiency as shown in Figure IV-9.

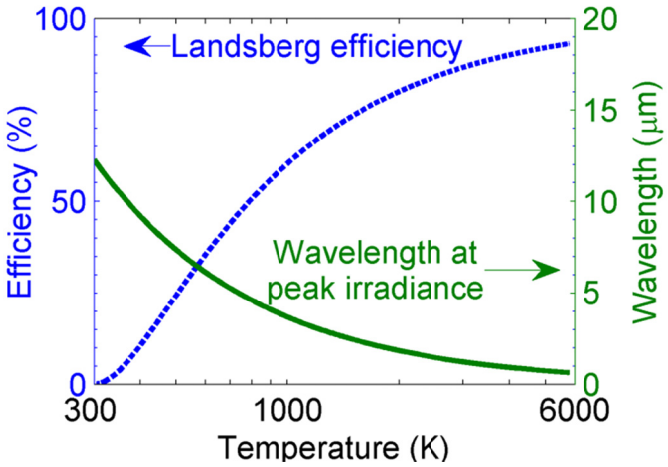


Figure IV-9 Thermodynamically-limited maximum efficiency of a solar cell as a function of temperature of the black body source. The solar cell is assumed to be at 300 K. Also shown is the peak wavelength of the black body spectrum as a function of temperature.

The Landsberg efficiency is based on a thermodynamic model for black body source and is not applicable to single frequency rentennas. As explained below, erroneous attempts have been made to develop a maximum efficiency model specific to rectennas (Corkish, 2002). In the intermediate absorber model (de Vos, 1992), the radiation is assumed to be converted into heat by the antenna and then Carnot efficiency is applied to the new heat source. In another model the antenna is modeled by a resistive noise source at the temperature of the sun (Sokolov, 1998). The first approximation for the antenna is incorrect as the antenna only channels the radiation from free space to the diode. The second approximation is correct only at low frequencies where the noise spectrum of a resistor is equal to the black body spectrum (Dicke, 1946).

Even in the absence of a theory that gives the maximum efficiency of a rectenna, the Landsberg efficiency is a valid upper limit. In context of the sun, this efficiency is 93%. However, in the context of harvesting waste heat, this efficiency is low and depends on the temperature of the heat source. This fact is overlooked by

researchers who propose to harvest the energy radiated out by the earth at night (Midrio, 2010). Energy can be harvested by a solar cell only at a temperature lower than that of the source. Again, the efficiency depends on the temperature ratio of the cell and the source as given by Landsberg.

CHAPTER V

SEMICLASSICAL THEORY OF RECTENNAS

In the circuit analysis presented in the previous chapter, the resistance and responsivity of the diode were used to calculate the efficiency of the rectenna. Classically, differential (or small signal) resistance is the inverse of the first derivative of current with respect to voltage and the responsivity is proportional to the ratio of the second to the first derivative. These calculations are based on the DC $I(V)$ characteristics of the diode. However, only at a relatively low frequency can a tunnel diode be considered as a classical rectifier (Eliasson, 2001). When the energy of the incident photons ($E_{ph} = \hbar\omega$) exceeds the voltage scale times e of the nonlinearity of the diode, a semiclassical analysis for the photon-assisted tunneling is required (Tien, 1963). At high frequencies, the quanta of energy of the incident photons is manifested in the illuminated $I(V)$ characteristics of the nonlinear device (Dayem, 1962).

The qualification of high frequency is relative to the voltage scale on which the nonlinearity in the diode appears. For the superconducting diodes investigated by Dayem, the voltage scale on which the diode nonlinearity is apparent corresponds to the the superconducting bandgap of the metal films ~ 0.1 meV. Photons in this energy range fall under microwave frequencies. For optical frequencies, the relevant energy scale is greater than 100 meV, which is also the voltage scale for the nonlinearity in low-barrier MIM diodes. This scale corresponds to the metal-insulator barrier height (Tucker, 1978).

The modification of the electronic properties of a device due to the interaction of the photon with the electron is termed as photon-assisted transport, which can be

analyzed using several approaches (Platero, 2004). Here, I follow the derivation given by Tien and Gordon (Tien, 1963) for a superconducting junction and generalized for tunnel devices by Tucker (Tucker, 1978). I examine the properties of an MIM diode in light of the semiclassical analysis. I also establish a correspondence between the classical and the first-order approximation of the semiclassical rectification theories. In Chapter VIII and IX, I present a generalized approach for examining the photon-assisted transport properties of diodes, applicable to non-tunneling devices as well.

The analyses presented in this chapter assume perfect impedance matching between the antenna and the diode. Also, the bandwidth limitation due to the circuit RC time constant is ignored. In the previous chapter, I ascertained that a small-signal rectification theory is sufficient for light intensities of interest. This calculation assumed a nominal antenna impedance of 100 Ω to which the diode should be matched. However, on having a higher resistance diode and therefore antenna, for a constant incident power the amplitude of the AC signal may no longer be smaller than the voltage scale over which the diode nonlinearity becomes apparent. In the semiclassical theory presented here, I assume that $eV_\omega << \hbar\omega$ so that a 1st order approximation can be applied to the results derived. A summary of models discussed in this chapter are listed below.

- First-order semiclassical model
    - Resistance and responsivity (section A)
    - Illuminated diode $I(V)$ characteristics (section B)
- Semiclassical load-line analysis (solar cell efficiency calculation) using a piece-wise linear diode
    - Derivation and results for first-order model
    - Discussion regarding higher-order model

- Correspondence and comparison of circuit model and semiclassical rectification theory
    - Efficiency for a classical small-signal circuit model
    - Efficiency for a classical large-signal circuit model

Both the first-order and higher-order semiclassical models are obtained from the theory presented in the next section.

A. Semiclassical theory

Consider an MIM tunnel diode that is biased at a DC voltage of V_D and excited by an AC signal of amplitude V_ω at a frequency ω. The overall voltage across the diode is

$$V_{diode} = V_D + V_\omega \cos(\omega t)$$

Eq. V-1

In a classical theory, the effect of the AC signal is modeled by modulating the Fermi level on one side of the tunnel junction while holding the other side at a fixed potential, as shown in Figure V-1. Effectively, the low frequency AC signal results in an excursion along the DC $I(V)$ curve around a bias point given by V_D.

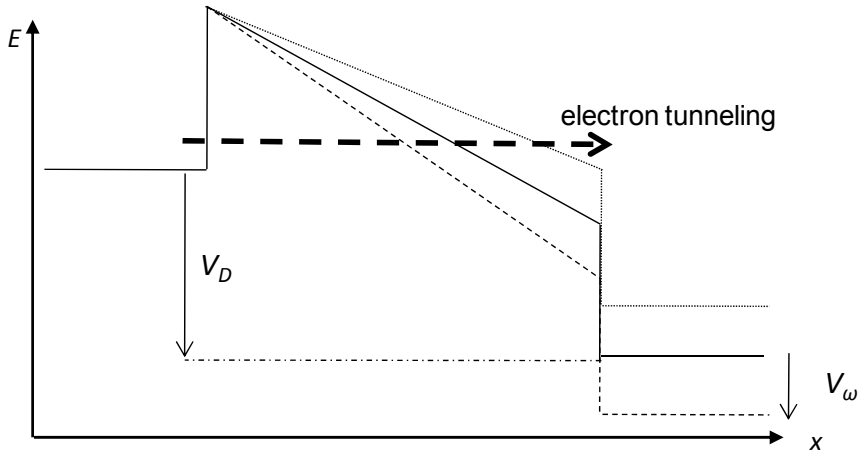


Figure V-1 Conduction band profile of an MIM diode modulated by an AC voltage. Classical model to account for the applied AC signal that modulates the Fermi level on one side of the tunnel junction.

For a high frequency signal, the effect of V_ω is accounted through a time dependent term in the Hamiltonian H for the contact (Tien, 1963), written as

$$H = H_0 + eV_\omega \cos(\omega t)$$

Eq. V-2

where, H_0 is the unperturbed Hamiltonian in the contact for which the corresponding wavefunction is of the form

$$\psi(x,y,z,t) = f(x,y,z)e^{-iEt/\hbar}$$

Eq. V-3

The harmonic perturbation in Eq. V-2 leads to an additional phase term whose effect can be included in the wavefunction as

$$\psi(x,y,z,t) = f(x,y,z)e^{-iEt/\hbar} \exp\left[-(i/\hbar)\int_t dt' eV_\omega \cos(\omega t')\right]$$

Eq. V-4

Integrating over time and using the Jacobi-Anger expansion, the wavefunction can be written as

$$\psi(x,y,z,t) = f(x,y,z)\sum_{n=-\infty}^{+\infty} J_n\left(\frac{eV_\omega}{\hbar\omega}\right)e^{-i(E+n\hbar\omega)t/\hbar}$$

Eq. V-5

where J_n is the Bessel function of order n. The modified wavefunction indicates that an electron in the metal, previously located at energy E, can now be found at a multitude of energies separated by the photon energy ($E_{ph} = \hbar\omega$) as shown in Figure V-2. The amplitude of the electron being found at energy $E+nE_{ph}$ is given by the Bessel function of order n, where n corresponds to the number of photons absorbed or emitted by the electron in a multi-photon process. The time dependent wavefunction is normalized since the infinite sum of the square of Bessel terms is unity. The electron density is proportional to the modulus squared of the wavefunction and therefore to the square of the Bessel function.

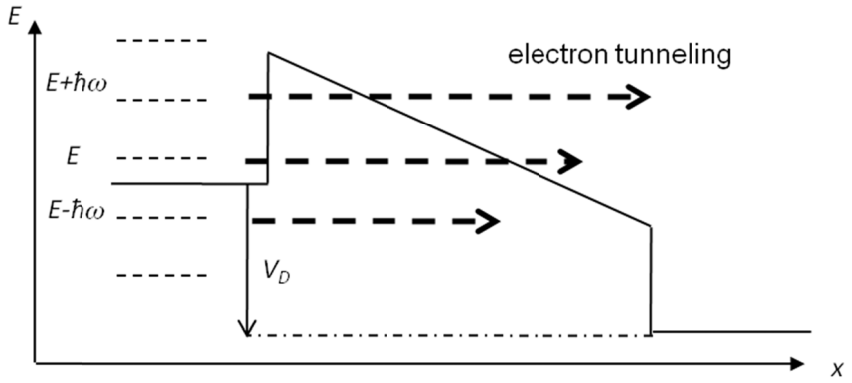


Figure V-2 Conduction band profile of an MIM diode with photon assisted transport. An electron at energy E absorbs or emits photons. The transmission probability of the electron increases at higher energies.

A heuristic explanation for the effect of the wavefunction modulation on the tunnel current is based on the fact that all such single-electron states will undergo the same modulation (Tucker, 1985, p.1067). This implies that in addition to the DC voltage, there is an additional voltage nE_{ph}/e that is applied across the diode with a

weighting factor $J_n^2(\alpha)$ where $\alpha = eV_\omega/E_{ph}$. The DC current under illumination is then given by

$$I_{illum}(V_D, V_\omega) = \sum_{n=-\infty}^{\infty} J_n^2(\alpha) I_{dark}\left(V_D + n\frac{\hbar\omega}{e}\right)$$

Eq. V-6

where $I_{dark}(V)$ is the tunnel current in the un-illuminated diode.

A similar derivation exists for the current in a diode illuminated by several frequencies (Tucker, 1985, p.1077). Here I restrict the analysis to a single frequency.

Apart from the DC component of the tunnel current given by Eq. V-6, there is a time dependent current which consists of the harmonics of ω. From a time dependent formulation for the illuminated tunnel current (Tucker, 1979), the first harmonic is given by

$$I_\omega = \sum_{n=-\infty}^{\infty} J_n(\alpha)\left[J_{n+1}(\alpha) + J_{n-1}(\alpha)\right] I_{dark}\left(V_D + n\frac{\hbar\omega}{e}\right)$$

Eq. V-7

From Eq. V-6 and Eq. V-7 one can calculate the resistance and responsivity under illumination using the equations

$$R_D^{SC} = \frac{V_\omega}{I_\omega}; \beta_i^{SC} = \frac{\Delta I_{illum}}{\frac{1}{2} V_\omega I_\omega}$$

Eq. V-8

where the superscript *SC* denotes the semiclassical formulation for calculating the photon-assisted transport. R_D^{SC} is the AC resistance for a signal at frequency ω and ΔI_{illum} is the incremental DC current due to the illumination. The $\Delta I_{illum} = I_{illum}(V_\omega) - I_{illum}(0)$ and the approximation used for Bessel functions are $J_0(x) \approx 1 - x^2/4$, and $J_{\pm n}(x) \approx (\pm x/2)^n/(n!)$.

From here on up to the end of section V.B, a small-signal AC voltage is assumed such that Bessel function terms only up to first order in n are needed. From Eq. V-7, to first order in n(=-1, 0, 1), R_D^{SC} is given by (Tucker, 1979)

$$R_D^{SC} = \frac{2E_{ph}/e}{I_{dark}(V_D + E_{ph}/e) - I_{dark}(V_D - E_{ph}/e)} \xrightarrow{classical} \frac{1}{I'}$$

Eq. V-9

which, in the classical limit ($E_{ph} \to 0$) leads to the differential resistance.

The semiclassical responsivity is similarly found out from the first order approximation of Eq. V-6 and Eq. V-7, and is given by (Tucker, 1979)

$$\beta_i^{SC} = \frac{e}{E_{ph}} \left[\frac{I_{dark}(V_D + E_{ph}/e) - 2I_{dark}(V_D) + I_{dark}(V_D - E_{ph}/e)}{I_{dark}(V_D + E_{ph}/e) - I_{dark}(V_D - E_{ph}/e)} \right] \xrightarrow{classical} \frac{1}{2} \frac{I''}{I'}$$

Eq. V-10

Again, Eq. V-10 in the limit of small photon energies leads to the classical formula for responsivity given by 1/2 the ratio of second derivative of current to the first derivative. I now use an experimentally measured DC $I(V)$ for a Nb/Nb_2O_5(3 nm)/Ta_2O_5(1.75 nm)/NbN double-insulator diode and calculate the semiclassical quantities given by Eq. V-9 and Eq. V-10. As shown in Figure V-3, the semiclassical resistance is the reciprocal of the slope of the secant between the currents at $V_D \pm \hbar\omega/e$, instead of the usual tangent at V_D for the classical case.

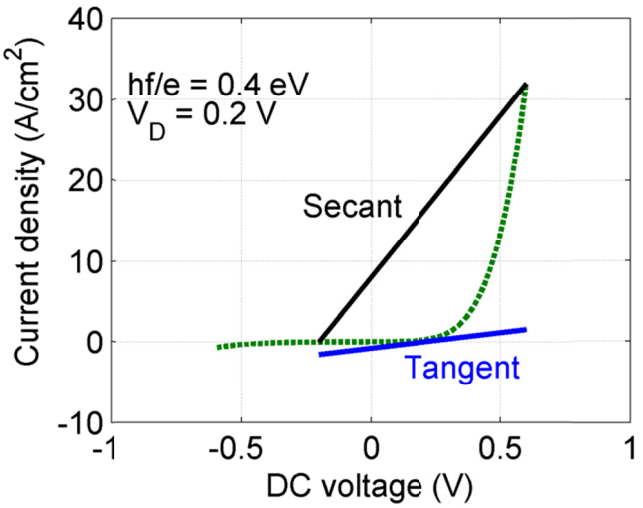


Figure V-3 The semiclassical resistance is obtained from the inverse of the slope of secant between $I_{dark}(V_D \pm \hbar\omega/e)$. An experimentally measured I_{dark} for a Nb/Nb_2O_5(3 nm)/Ta_2O_5(1.75 nm)/NbN double insulator diode is used.

The semiclassical responsivity reflects the change in the slope of the secant, rather than the curvature at the bias point for the classical case. In Figure V-4, I plot the semiclassical resistance and responsivity at zero bias vs. the photon energy (E_{ph}) using the $I_{dark}(V)$ shown in Figure V-3. As the photon energy increases, the resistance of the diode decreases as the slope of the secant is larger than the slope of the tangent.

The AC resistance given by Eq. V-9 is the diode impedance seen by the antenna. The drop in the AC resistance at high frequencies improves the impedance match between the antenna and the diode. As shown in Chapter IV, the semiclassical diode resistance leads to a higher coupling efficiency.

The change in the slope of the secant decreases as it becomes more vertical, therefore the responsivity decreases with increasing E_{ph}. However, at high E_{ph} the responsivity approaches the limit of $e/\hbar\omega$, which is the maximum achievable responsivity corresponding to a quantum efficiency of 1. Thus, even a diode with a poor quantum efficiency at low E_{ph} becomes more (quantum) efficient and adequate

at high E_{ph}. This occurs because the semiclassical responsivity is a strong function of the diode asymmetry instead of the nonlinearity.

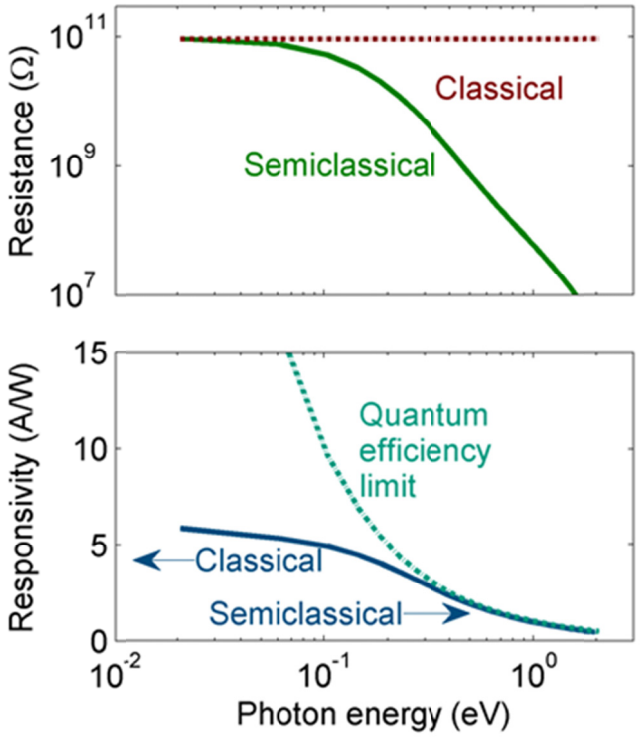


Figure V-4 Resistance and responsivity vs. photon energy calculated classically and semiclassically using the finite difference forms given by Eq. V-9 and Eq. V-10 for the diode of Figure V-3, with V_D=0 V. At high photon energy, the semiclassical resistance is significantly lower than the classical value and the semiclassical responsivity approaches that for unity quantum efficiency.

The Tien-Gordon formulation used in above derivation assumes that the AC excitation is applied on one of the contacts. This leads to the problem of gauge invariance whereby the response of the diode is not strictly dependent on the difference in voltage applied across the junction (Pedersen, 1998). A modification to the Tien-Gordon approach that essentially leads to the same results was proposed by Büttner (Büttner, 1968). Another assumption implicit in the Tien-Gordon approach is that the reservoir charge distribution extends up to the junction region

and that the applied field does not influence the charge distribution in the junction (Pedersen, 1998). For tunnel diodes, these approximations are valid as there is a high concentration of electrons in the metals and a low concentration in the insulating tunnel barrier.

B. Semiclassical rectifier-properties

In response to an alternating voltage on top of a DC bias, an assumed zero-frequency or DC current generated in the diode is given by the I_{illum} vs. voltage relation of Eq. V-6. In Figure V-5, I show the procedure for calculating the illuminated $I(V)$ for a diode with step characteristics (Eliasson, 2001, p.142), (Eliasson, 2005). In the next section, I will argue that the step diode is not realistic and the efficiency calculated from it is misleading. Throughout this section, I assume that the strength of the AC signal (V_ω) is constant. This implies that for different bias points along the illuminated $I(V)$ curve the power delivered by the source varies. If V_ω is applied using a function generator, this would indeed be the case, but for a rectenna illuminated with a constant intensity signal, the V_ω varies with the bias voltage. This case is discussed in Section IV.C.

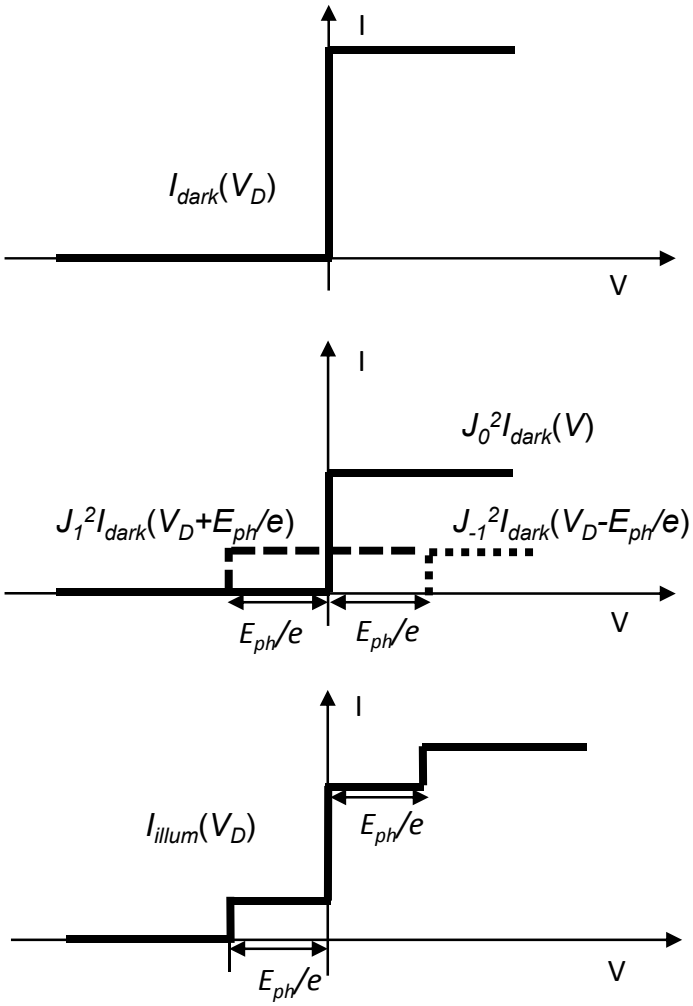


Figure V-5 Obtaining I_{illum} vs. voltage curve starting with a step I_{dark}, and applying a first order approximation of Eq. V-6.

In the case of a detector, the DC voltage is fixed and the additional current generated due to the AC signal is given by $I_{illum}(V) - I_{dark}(V)$. The I_{illum} for the ideal case gives an open-circuit voltage of E_{ph}/e and a short circuit current I_{SC} that depends on the strength of the AC signal V_ω via the argument of the Bessel function. Eliasson (Eliasson, 2001, p.143) shows that in the ideal diode, the DC power given by the product of $I_{SC}*E_{ph}/e$ implies a conversion efficiency of 100%. As explained in the next section, the maximum conversion efficiency of a realistic diode achievable under the first order semiclassical region of operation is 50%.

For a rectenna solar cell, the I_{illum} vs. voltage relation determines the quadrant of the $I(V)$ curve where it provides power. Unlike conventional semiconductor junction solar cells, which operate in the fourth quadrant, rectennas operate as solar cells in the second. Using the $I_{illum}(V)$ characteristics, the operating point of the solar cell can be calculated from a load-line analysis as discussed in the next section.

In Figure V-6, I calculate I_{illum} using the experimentally measured MIIM diode I_{dark} shown in Figure V-3. This diode has a high forward-to-reverse current ratio, which is required for obtaining a significant short-circuit current or open-circuit voltage.

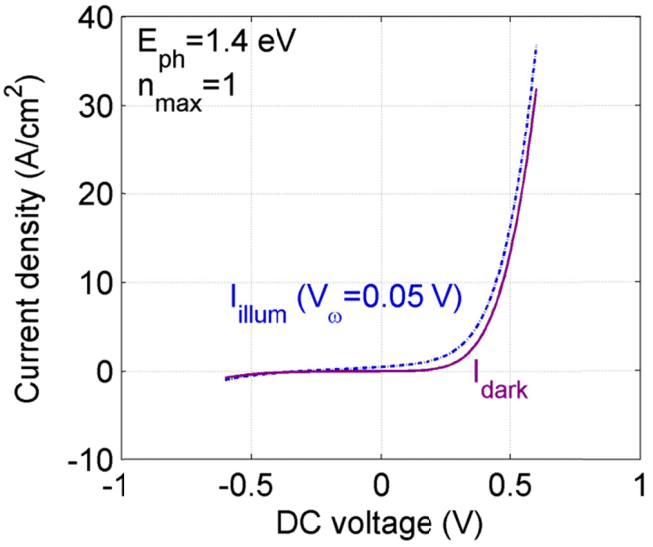


Figure V-6 $I(V)$ characteristics for an illuminated diode computed using the dark $I(V)$ of the MIIM diode shown in Figure V-3. The I_{illum} is calculated from I_{dark} for an AC signal amplitude of 0.05 V using a first-order approximation of Eq. V-6.

The effect of the illumination is more noticeable on a log scale. In Figure V-7, I plot the I_{illum} vs. voltage curves for four different values of V_ω, with V_ω=0 corresponding to the I_{dark}. As V_ω increases, corresponding to an increasing number

of photons incident on a diode, the zero-crossing of the illuminated $I(V)$ shifts leftwards. This shows that as the power incident on the diode increases, a higher operating voltage and thus a greater efficiency can be achieved by the rectenna. This conclusion is consistent with the requirement for a large collection area described in Chapter IV.

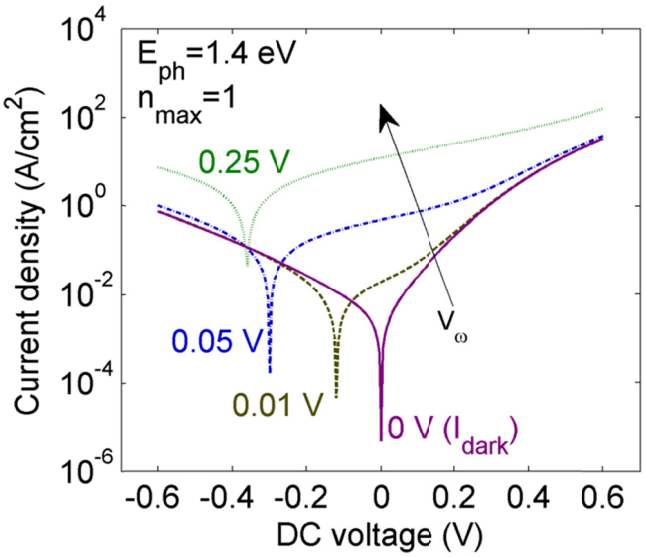


Figure V-7 Calculated $|I(V)|$ characteristics on a log-scale under various levels of illumination for the MIIM diode. The open-circuit voltage and short-circuit current increasing with increased illumination (V_ω). The ratio of eV_ω/E_{ph} is kept small, allowing a first order approximation of Eq. V-6 to be used.

In Chapter IV, the strength of the AC source (V_ω) was calculated as the equivalent of the overall solar spectrum. This approach cannot be applied here as the frequency dependence of the source also needs to be accounted for. Eliasson (Eliasson, 2001, p.132) has calculated an electric field strength (and voltage) by equating the energy in a capacitive field to that of a harmonic oscillator. This approach appears to be erroneous as the equation used for the energy in the field is only applicable to a static field. For an AC field, energy is not 'stored' across a capacitor. Further problems with Eliasson's method are the arbitrary choice of a 10

nm window around each discrete step in photon energy, and counting the number of photons incident in a time period of oscillation (Eliasson, 2001, p.136). These assumptions lead to the number of photons being fractional, which he considered unnatural and arbitrarily set it to 1 for all frequencies. This causes a significant change in the spectral distribution of energy.

To remove these errors, I propose the following changes in the calculation of V_ω. First, as the perturbation is applied as a classical field or voltage, it is unnecessary to find the number of photons corresponding to the field. This is reflected in the description of the semiclassical model given by Eq. V-2, where the electromagnetic field is included classically ($V_\omega\cos(\omega t)$) but its effect on the electrons is modeled quantum mechanically (H). In the case of a detector, where the input power is dissipated only as AC power in the diode, the voltage can be found out directly from the field strength or the power at a frequency as

$$V_\omega = \sqrt{2P_\omega R_D^{SC}}$$

Eq. V-11

where P_ω is the incident power at the frequency of interest.

The second change is required in selecting the interval over which the spectrum should be integrated to find the power around a particular frequency. For this, I propose that the window of integration be such that the frequencies within it exhibit a certain temporal coherence, which is denoted by a coherence time τ_c (Donges, 1998). This time-scale over which the radiation around a particular frequency remains coherent should be greater than the average time over which at least one electron flows across the barrier. This time is proportional to the inverse of the current (I) and is given by $\tau_{1e} = e/I$.

Donges considers the coherence of the overall black body spectrum, which results in τ_c = 2 fs. Correspondingly, the current in the illuminated diode current

should be greater than 80 μA. If only a part of the spectrum is considered, the coherence time will be larger. For a minimum of one electron to flow during this time will result in a smaller current requirement. However, a weaker illumination will lead to lesser current generation. A numerical solution is required to determine the appropriate bandwidth around a central frequency, which satisfies the condition of the illuminated current equal to e/τ_c and determine the P_ω and V_ω.

C. Load-line analysis

1. First-order semiclassical model

The step I_{dark} used in Figure V-5 is a misleading representation of a diode. Near $V_D=0^+$ there is essentially zero resistance to the flow of current. As the voltage increases, there is a corresponding increase in resistance that keeps the current constant. Such an increase in resistance may actually be due to the series resistance of the diode. However, this trend for resistance is opposite of that for a real diode in which the resistance is higher at zero voltage and decreases with increasing voltage. A more realistic model for the ideal diode is one with no current for $V<0$ and a linear $I(V)$ for $V>0$. I modify this $I(V)$ as having a higher resistance (R_N) for $V<0$ and lower resistance (R_P) for $V>0$. The forward bias resistance is kept fixed in the ensuing calculations at 100 Ω. I also define a current asymmetry ratio $A = R_N/R_P$. The I_{dark} for such a diode is shown in Figure V-8.

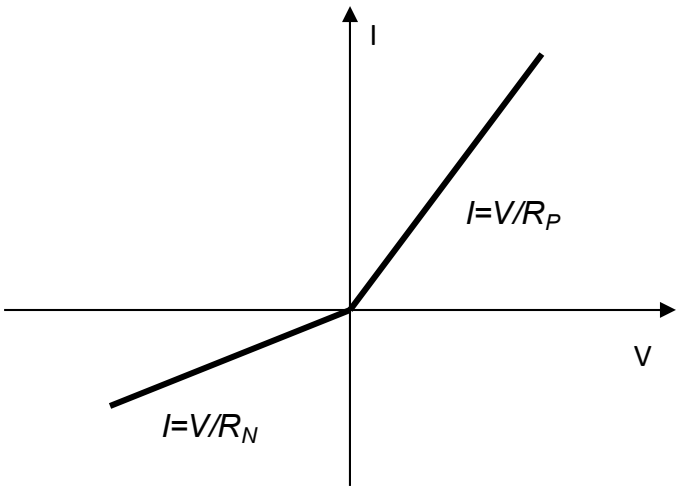


Figure V-8 I_{dark} for a piecewise linear diode with a higher resistance in reverse than in forward bias.

If a low-frequency small-signal voltage is applied to the above diode, it will rectify only around V_D=0 V. However, at high frequencies, the semiclassical analysis implies that the MIM $I(V)$ curve is sampled in steps of E_{ph}/e even if the amplitude of the AC signal is small. Therefore, there is non-zero responsivity even for $V_D \neq 0$ V.

Following Figure V-6, a diagrammatic calculation of the illuminated $I(V)$ is shown in Figure V-9, with a fixed V_ω. The I_{illum} corresponds to the current that can be extracted out of a diode and the operating point depends on the load resistance.

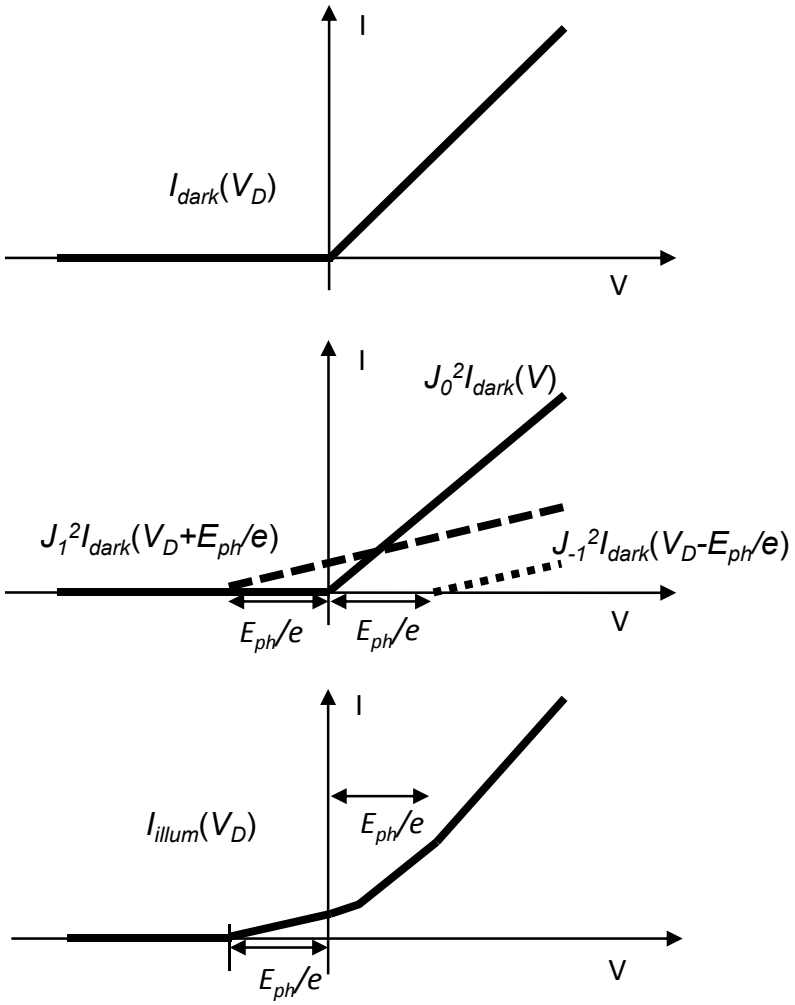


Figure V-9 I_{illum} vs. voltage for a piecewise linear $I(V)$ with a high forward-to-reverse current ratio.

As mentioned earlier, a rectenna is likely to develop a V_ω that varies with V_D in order to keep the input power constant. Under this condition, I now derive the modified load-line curve and find out the operating point corresponding to the maximum obtainable power from a solar cell based on this diode. The rectified power given by $P_{rect}=V_D*I_{illum}$ can be expressed in terms of I_{dark} using Eq. V-6 as

$$P_{rect} = V_D\left[\left(1-\frac{\alpha^2}{4}\right)^2 I_{dark}\left(V_D\right)+\frac{\alpha^2}{4}\left(I_{dark}\left(V_D+E_{ph}/e\right)+I_{dark}\left(V_D-E_{ph}/e\right)\right)\right]$$

Eq. V-12

where I_{illum} can be separated into I_{dark} and the remaining terms that depend on the strength of the signal making up I^{ω}_{DC}. Therefore I^{ω}_{DC} gets divided into the load current and the dark current in the diode, as shown in Figure IV-6. The total power received by the rectenna is equal to the sum of AC power dissipated in the diode and the rectified DC power, which is equal to $V_D * I^{\omega}_{DC}$. When the DC resistance of the diode at the bias point is much larger than the load resistance, the DC power is most efficiently delivered to the load. For power generation, the diode has to operate in the second quadrant implying that $-E_{ph}/e < V_D < 0$. The total power equal to the input power is given by

$$P_{in} = P_{AC} + P_{DC} = \frac{1}{2} V_{\omega} I_{\omega} + \left| V_D I^{\omega}_{DC} \right| = \frac{1}{2} V_{\omega} I_{\omega} + \left| V_D \beta_i^{SC} \right| \frac{1}{2} V_{\omega} I_{\omega}$$
$$= \frac{V_{\omega}^2}{2 R_D^{SC}} \left(1 + \left| V_D \beta_i^{SC} \right| \right)$$

Eq. V-13

The DC voltage dependent V_{ω} can be determined from the above equation.

Under the assumption that little or no power is lost to the diode dark current at reverse bias, the efficiency of the rectenna solar cell at a bias point is given by

$$\eta_{load} = \frac{\left| V_D \beta_i^{SC} \right|}{\left(1 + \left| V_D \beta_i^{SC} \right| \right)}$$

Eq. V-14

According to the I_{illum} in Figure V-9, the largest absolute operating voltage of the rectenna for power generation in the second quadrant is $\hbar\omega/e$, while the maximum responsivity of the diode from Eq. V-10 is $e/\hbar\omega$. Therefore the maximum value of η_{load} is 1/2 implying that under the first-order semiclassical regime, the maximum efficiency of operation for the rectenna is 50%. The remaining 50% or more of the power is dissipated in the diode as AC power.

I now apply the above analysis to a piecewise linear and an exponential dark $I(V)$ curve. The forward bias resistance of the piecewise linear $I(V)$ is 100 Ω and the forward to reverse current ratio is 10^8. The reverse saturation current of the exponential characteristics is 10^{-10} A and the coefficient in the exponential is 100 V^{-1}. An input power of 10^{-8} W is assumed, at a photon energy of 1.5 eV. The power corresponds to the solar intensity of 100 W/cm^2 received over an area of 10 μm^2.

The basic assumption in the above analysis is that the first order theory is applicable requiring $\alpha = eV_\omega/\hbar\omega$ to be much smaller than 1. In Figure V-10, I plot the illuminated $I(V)$, the variation of α, and the load efficiency as a function of the DC bias voltage.

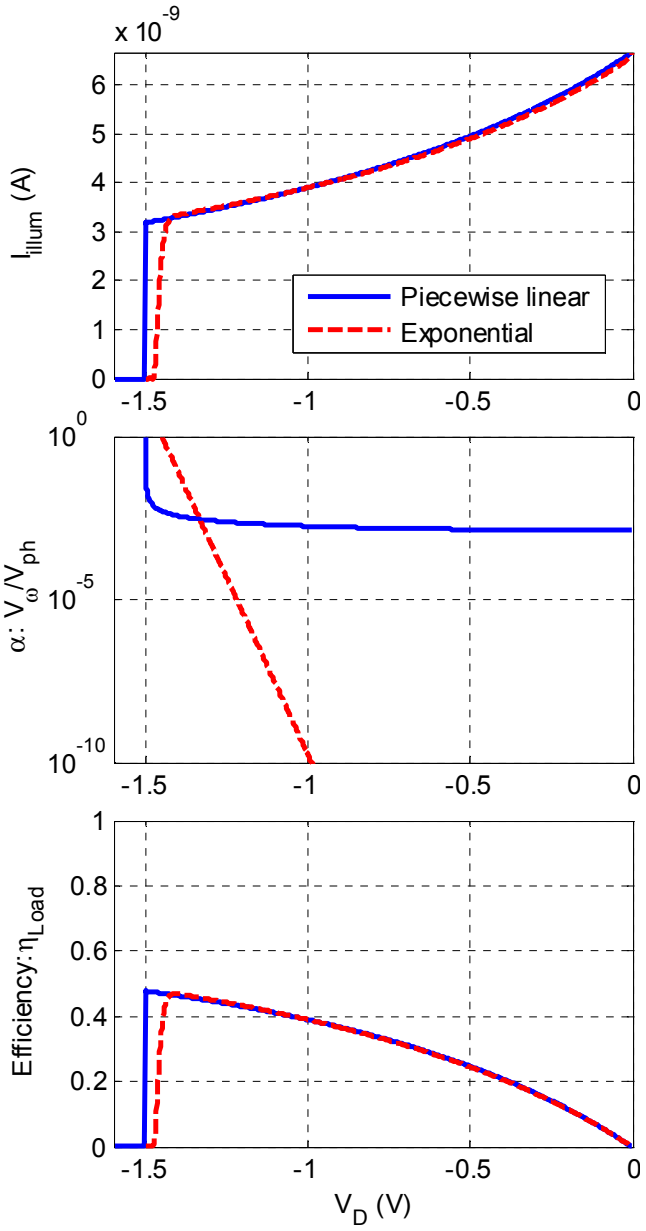


Figure V-10 Illuminated characteristics for a piecewise linear, and an exponential dark $I(V)$ curve.

As expected, at $V_D = V_{OC} = -\hbar\omega/e$, I_{illum} is zero, also, the product of $V_{OC}I_{SC}$ is equal to the input power of 10^{-8} W. The assumption of $\alpha<<1$ is satisfied for majority of the illuminated characteristics and is violated only near $V_D=V_{OC}$. The effect of a varying V_ω is evident in the illuminated $I(V)$ curve leading to the trapezoidal shape that results in a peak efficiency close to 50%. This is to be contrasted with Figure

V-9, where the triangular illuminated $I(V)$, under the assumption of constant V_ω, incorrectly suggests a peak efficiency of only 25%.

As V_ω increases at large negative V_D, the efficiency rises. This suggests that enhancement of the field by confining a fixed power in a smaller diode region leads to higher efficiency (Dagenais, 2010). However, this requires high impedance antennas to match to the high impedance of extremely confined diode structures and the low RC requirement discussed in Chapter IV.

2. Higher-order semiclassical model

When $\alpha = eV_\omega/\hbar\omega$ is on the order of one or larger, the first-order approximation is no longer applicable. Under this condition, the semiclassical formulae for resistance, responsivity, illuminated I(V) and power need to be derived with higher-order Bessel terms included in Eq. V-6 and Eq. V-7. Another consideration at large AC voltage is higher-order harmonics which haven't been considered in the first-order analysis. These factors may modify the illuminated characteristics and efficiency of the rectenna.

D. Correspondence between circuit model and semiclassical rectification theory

In the limit of photon energy smaller than the voltage scale of nonlinearity in the diode, the semiclassical formulas for resistance and responsivity reduce to their classical analogues. In a similar manner, there needs to be a correspondence between the classical and semiclassical theories for rectification. Here, I analyze this correspondence and also explain why a small-signal classical rectifier and even the first-order semiclassical one are only able to operate at 50% efficiency.

Both the classical small-signal and the semiclassical first-order models, under the condition of large reverse-bias DC resistance lead to the efficiency formula of $\eta = |\beta V_D|/(1+|\beta V_D|)$. Maximum quantum efficiency limits the value of $|\beta V_D|=1$ leading to $\eta = 1/2$. With half the power delivered to the load at DC, the other half is dissipated as AC power in the diode. At any $\eta < 1/2$, the AC power dissipation is correspondingly increased.

In the classical rectifier, the AC signal from the antenna feeds into an AC diode resistance R_D, causing dissipation in the diode. Under the first-order semiclassical operating regime, a similar AC resistance is applicable. This resistance is influenced by the large photon energy, suggesting that the dissipation occurs due to photon-assisted tunneling.

I have emphasized the fact that the diode is always dissipating in order to point out the difference between the classical small-signal and large-signal (linear) rectifier. In the large-signal circuit model, the diode turns ON only for a short duration of the cycle as shown in Figure V-11.

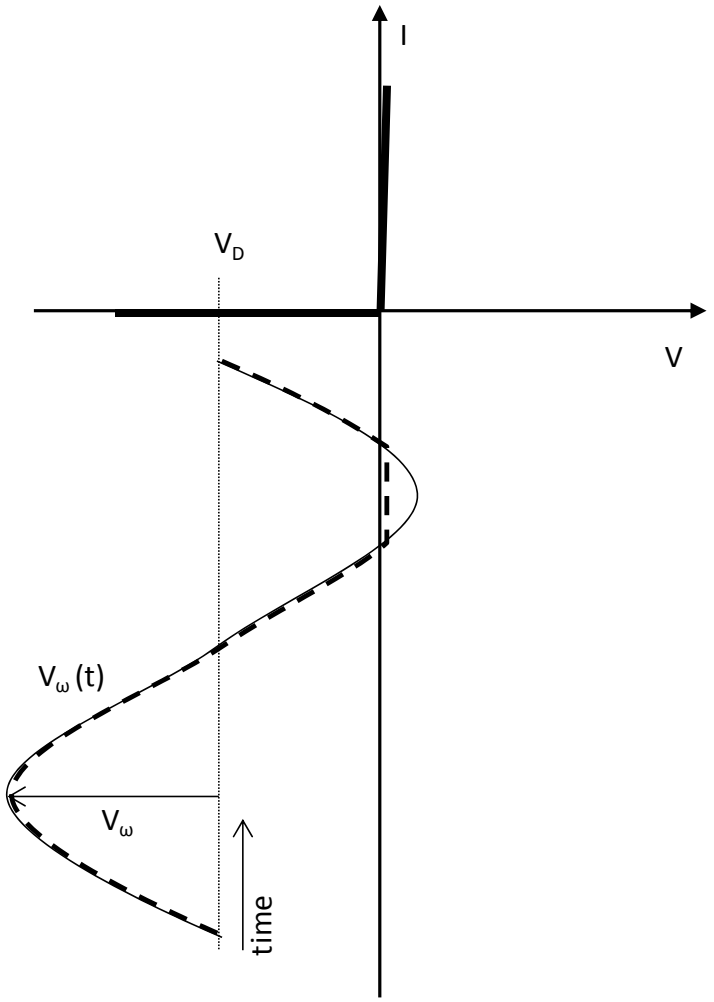


Figure V-11 Representation of a classical large-signal model superimposed on a diode $I(V)$ with a sharp turn on for $V>0$ and zero current for $V<0$.

In this case, the best case self-bias DC voltage that can develop across the diode is $V_D=-V_\omega$. In doing so, the rectenna essentially becomes a level shifter (Sedra, 2004, p.194). If the level shift is such that the diode never turns on, there is no DC current generated. For supplying power to a load, the diode must turn on for a part of the AC voltage cycle. This happens for $V_\omega(t)+V_D > 0$ allowing the rectenna to draw charge at a very low resistance and therefore low loss. In the remaining cycle, the diode is turn off, and the charge that was pulled up is supplied to the load. The AC-to-DC conversion efficiency now depends on the power that is dissipated in the

diode when it turns ON. Depending on the diode asymmetry, conversion efficiency close to 100% is achievable.

There is no evident correspondence between the large-signal circuit model described above and the higher-order semiclassical model considered in the previous section. Such a correspondence should exist to make the semiclassical theory consistent with the circuit model in the limit of small photon energies. This is crucial for understanding whether efficiencies higher than 50% are achievable in rectennas receiving large photon energies at large AC amplitudes.

CHAPTER VI

MIM TRAVELING-WAVE DETECTOR

In this chapter, I evaluate a technique to improve the performance of antenna-coupled diode rectifiers working in the infrared. Efficient operation of conventional, lumped-element MIM rectifiers as described in Chapter IV is limited to the low terahertz due to a large RC time constant. By using the femtosecond fast metal/insulator/metal (MIM) diodes in a traveling-wave (TW) configuration, a distributed rectifier with improved bandwidth and efficiency is possible (Estes, 2006). I develop a method for calculating the responsivity of the antenna-coupled TW detector. Three TW devices, made from different materials, are simulated to obtain their impedance and responsivity at 1.5, 3, 5, and 10 μm wavelengths. I compare the responsivity of a TW device with the lumped-element device. To gauge the performance of the TW detector with respect to non-rectenna based detectors from literature, I calculate its normalized detectivity for comparison with several IR detectors. I also identify ways for improving the detectivity of the TW device.

A. Theory of operation

As introduced in Chapter I, a rectenna coverts the incident electromagnetic wave to an AC signal using a micro-antenna , which then gets rectified to a DC signal in the diode. The difference between rectennas using a lumped-element diode and the TW diode is twofold. First, they differ in the manner in which the signal is transferred from the antenna to the diode. Second, as the names suggest, the traveling-wave rectifies a wave while the lumped-element has the same signal appearing across the whole diode.

In Chapter IV, I have already discussed the requirements of impedance matching and RC time constant for a lumped-element model. Here I describe the concept of a TW device and explain how it can achieve the above mentioned requirements.

The MIM traveling-wave detector was proposed as a method for improving the efficiency of antenna coupled MIM rectifiers operating in the infrared (Estes, 2006). A 3D view of this device is shown in Figure VI-1(a). It consists of an antenna connected to two metals M1 and M2, with a thin insulator between them. The metal/insulator/metal sandwich forms an extended MIM tunnel diode, which has the characteristics of a plasmonic waveguide (Zia, 2004). The radiation received by the antenna excites a surface-plasmon in the TW section, which propagates in the z-direction. The plasmon develops a voltage between the two metal electrodes causing tunneling of electrons in the y-direction. The resulting current is rectified due to an asymmetry in the $I(V)$ curve about the bias point, as characterized by the semiclassical diode responsivity discussed in Chapter V. Due to losses, the high frequency components of the rectified signal do not travel over a large distance. Thus the output at the contacts is a net DC current.

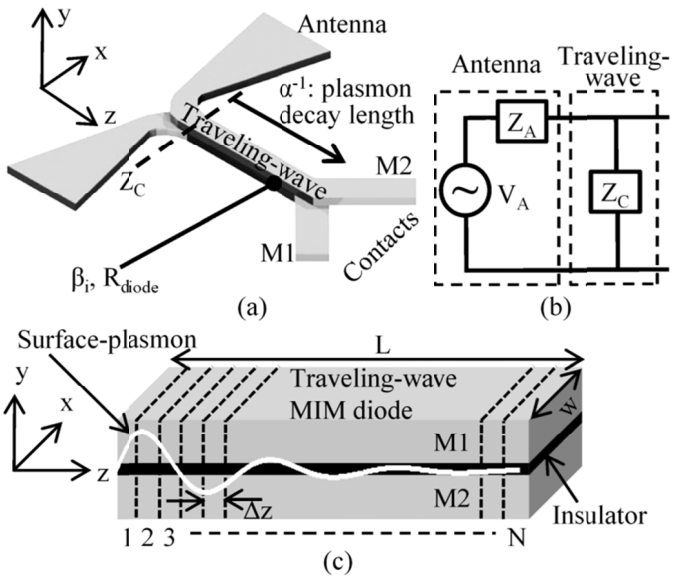


Figure VI-1 (a) An isometric view of the antenna-coupled traveling-wave detector. The antenna arms converge into a parallel-plate waveguide with a thin (2 nm) insulator between the metals M1 and M2. On the other end, these metals form the leads to the contact pads (not shown) or the load. Parameters required for calculating the performance of the traveling-wave include the characteristic impedance of the waveguide (Z_C); the plasmon decay length (α^{-1}); and the tunnel diode resistance (R_D) and responsivity (β_i). (b) Small signal circuit representation of the detector. Impedance of the traveling-wave diode can be readily matched to the antenna ($Z_A = Z_C^*$). (c) A 3D view of the traveling-wave MIM diode. The structure has a length L and a width w. To calculate the responsivity, I divide the entire length of the traveling-wave into N equal sections of width Δz.

The technique of rectifying a surface-plasmon wave extends the bandwidth and provides a higher efficiency as compared to the lumped-element detector. In the TW detector, the antenna can be matched to the characteristic impedance of the waveguide as shown in the equivalent circuit in Figure VI-1(b). This leads to an improved power transfer from the antenna to the diode. Also, the distributed rectifier in the traveling-wave does not have an RC bandwidth limitation. In the next section, I explain the techniques required to model the performance of the traveling-wave detector.

A variation of the traveling-wave has been implemented (Hobbs, 2007) by coupling the antenna-coupled traveling-wave to a silicon waveguide.

B. TW concept and modeling

Similar to the lumped-element detector responsivity calculated in chapter IV, here I give a formulation for the responsivity of the traveling-wave detector. To model the responsivity of the distributed rectifier, the MIM waveguide is divided into N parts of width Δz as shown in Figure VI-1(c). The antenna excites a surface-plasmon wave in the MIM waveguide. Assuming that a power of 1 W is incident on the waveguide cross-section (x-y), this wave develops a certain voltage distribution ($V_{1W}(z)$) between the two metal plates. The 1 W of power is chosen for normalization. In an actual device, the incident power would be several orders-of-magnitude smaller. Resistive losses in the metal cause the plasmon-wave amplitude to decay exponentially along-z. The resulting voltage distribution is given by

$$V_{1W}(z) = V_0 e^{-\alpha z}$$

Eq. VI-1

where α is the decay constant or the inverse of the decay length of the plasmon, and V_0 is the voltage at $z = 0$. Here I have ignored the decay of the plasmon wave due to rectification. As calculated later, the rectified energy is a small fraction of the total. To simplify the derivation, I assume a constant voltage distribution along the x-axis. The net AC power across the diode is given by the following sum

$$P_{AC} = \sum_{1}^{N} \left(\frac{{V_{1W}}^2(z)}{2R_{D/A}(w\Delta z)} \right)$$

Eq. VI-2

where $R_{D/A}$ is the resistance per unit area of the diode, and w is the width along the x-axis. The DC current that results from P_{AC} is given by

$$i_{dc} = \beta_i P_{AC}$$

Eq. VI-3

Combining Eq. VI-2 and Eq. VI-3, and applying the limit of $N \to \infty$ ($\Delta z \to 0$) gives

$$i_{dc} = \frac{\beta_i V_0^2 w}{2R_{D/A}} \left(\frac{1 - e^{-2\alpha L}}{2\alpha} \right)$$

Eq. VI-4

where L is the length of the waveguide. Since the incident power was normalized to 1 W, the DC current in Eq. VI-4 also gives the responsivity. Choosing the length of the waveguide to be larger than the decay length of the surface-plasmon $\alpha L > 1$, the responsivity is approximated as

$$\Re_{TW} = \frac{\beta_i V_0^2}{4R_{D/A}} \frac{w}{\alpha}$$

Eq. VI-5

Eq. VI-5 provides an estimate for the efficiency of the traveling-wave detector and helps to identify ways for improving the device, which are discussed later.

I now describe how to obtain the parameters required for computing the responsivity using Eq. VI-5. The semiclassical diode-properties (β_i and $R_{D/A}$) have already been discussed in chapter V. Next, I describe the procedure for calculating the V_0, α, and the characteristic impedance (Z_C) of the TW device. These quantities are then used to estimate the responsivity and noise performance of the TW detector.

To find the parameters for the traveling-wave section, I first calculate the transverse magnetic (TM) modes of an MIM waveguide. To simplify the analysis, it is assumed that the waveguide is infinite in the x-direction, ensuring that only TM modes exist. These modes can have an even- or an odd-symmetric field distribution (Raether, 1988). The incoming radiation induces an odd-symmetric current

variation on the antenna arms, which matches with the odd TM mode of the MIM waveguide. By designing the TW diode to have characteristic impedance equal to the antenna impedance, the impedance matching and the mode matching allow efficient coupling between the antenna and the rectifier.

I calculate the complex propagation constant ($\gamma=\alpha+ik$), of the infinite waveguide by using a matrix method (Davis, 2009) to obtain the dispersion relation for the TM mode. With the estimate for γ and fixing the width of the waveguide to 100 nm I solve for the corresponding hybrid mode (TE + TM) using a commercial finite element solver (COMSOL, 2004). This provides an accurate propagation constant (γ) from which the plasmon wavelength and the distance over which it propagates can be calculated.

The solver also provides the field distribution in the 2D cross-section of the waveguide. In Figure VI-2, I plot these results for a Ni-NiO-Ni (MIM) TW structure simulated at 100 THz. As seen in Figure VI-2(a), the wave is confined mainly to the 2-nm thick insulator region.

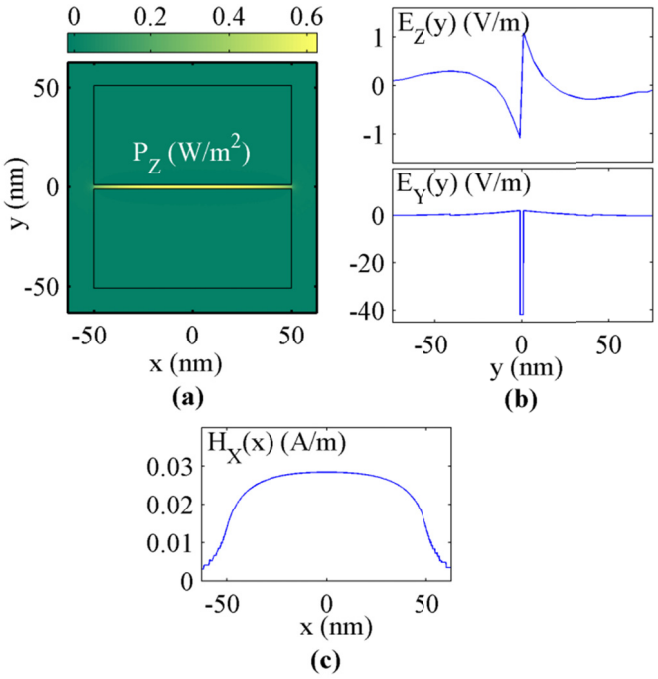


Figure VI-2 Finite element analysis of a Ni-NiO-Ni traveling-wave diode for 100 THz. (a) Cross-sectional distribution of power flowing in the *z*-direction. The black rectangles mark the metal regions of the MIM diode. The power is confined mainly to the 2-nm thick insulator. (b) The *y*- and *z*-directed electric fields as a function of the vertical position for x=0. (c) The x-directed magnetic field as a function of the horizontal position for y=0. The $E_Y(y)$ and the $H_X(x)$ are used to calculate the characteristic impedance of the traveling-wave, which is 45 Ω for this analysis.

From the field distribution, the characteristic impedance (Huang, 2009) of the TW structure is obtained as

$$Z_C = \frac{V}{I} = \frac{\int_{M1}^{M2} \vec{E}.dy}{\oint \vec{H}.d\vec{r}} = \frac{\int_{M1}^{M2} E_Y dy}{\int_{-\infty, y=0}^{\infty} H_X dx}$$

Eq. VI-6

Here the voltage is given by the integral of the electric field, which is shown in Figure VI-2(b). The current is calculated using Ampère's law by integrating the magnetic field in a semi-infinite loop around either of the conductors. As seen in

Figure VI-2(c), the magnetic field falls rapidly outside the waveguide. Thus at an infinite distance, it is negligible and the current can be approximated by an integral along x from -∞ to +∞. For the validity of the above formula, the metal needs to be a perfect electric conductor or have its magnitude of dielectric constant much larger than the insulator $|\varepsilon_{metal}| >> \varepsilon_{insulator}$. At infrared wavelengths, the latter is the case for the materials that are considered here.

Next, I calculate the V_0, which is proportional to the voltage given by the numerator of Eq. VI-6. Since V_0 corresponds to an incident power of 1 W, I apply a normalization factor given by the integration of the z-directed power density in the TW cross-section. This leads to

$$V_0 = \frac{\int_{M1}^{M2} E_y dy}{\int_{area} P_Z dA}$$

Eq. VI-7

Using the above results, the performance of the traveling-wave detector is calculated in the next section.

C. Performance calculation

As explained earlier, the TW diode is expected to facilitate a good impedance match between the antenna and the diode, and overcome the RC bandwidth limitation, allowing efficient operation in the IR. I quantify these attributes by calculating the characteristic impedance and the responsivity of the TW detector at several IR wavelengths of interest. I choose to analyze these properties for three MIM material combinations that provide a small resistance ($R_{D/A}$) for the tunnel barrier. As given by Eq. VI-5, a small resistance is required to obtain a high responsivity for the TW detector. The MIM diodes compared are Ni/NiO/Ni (0.2 eV)

(Hobbs, 2007), $Nb/Nb_2O_5/Nb$ (0.1 eV) (Bain, 1985), and $Ta/Ta_2O_5/Ta$ (0.4 eV). The bracketed quantity denotes the barrier height of each diode. Symmetric tunnel-barriers have been chosen to keep the analysis simple.

1. Characteristic impedance and responsivity

The characteristic impedance (Z_C) of the traveling-wave diode is plotted as a function of wavelength (1.5 to 10 μm) in Figure VI-3. The Z_C is well within the range of typical antenna impedances. For a precise match with the antenna, the impedance of the TW structure can be fine tuned by varying its width along-x. The impedance match can be achieved over a large wavelength range, which is useful for broadband detection. The small variation in impedance with frequency is attributed to variation in the experimental values for the metal dielectric constants (ε_{metal}) (Palik, 1985), (Palik, 1991). The impedances for the Nb and Ta devices are comparable as their ε_{metal} are similar.

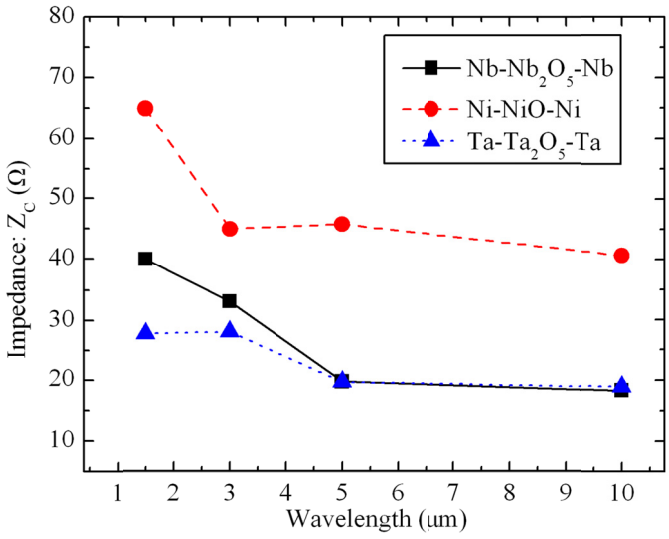


Figure VI-3 Calculated characteristic impedance (Z_C) of the traveling-wave diode vs. free-space wavelength. Three symmetric MIM structures have been analyzed. The values for Z_C are in the range of typical antenna impedances. Tuning of the impedance is possible by varying the width of the traveling-wave diode.

Assuming no losses in the antenna, the responsivity of the TW detector is calculated at a DC bias of 0.1 V. The bias ensures an operating point at which the diode responsivity (β_i) is non-zero. I assume that a perfect impedance match can be made between the traveling-wave diode and the antenna. The detector responsivity calculated from Eq. VI-5 is shown in Figure VI-4. The responsivity is higher for the low-barrier MIM diodes with the highest for the Nb diode. The peak in the curves for Nb and Ni, and its absence in Ta, can be explained from the trends in resistance ($R_{D/A}$) and responsivity (β_i). In the semiclassical model, these quantities vary with frequency. Since the Ta-Ta_2O_5-Ta is a higher tunnel-barrier than the Nb and Ni MIM diodes, it has a different variation of $R_{D/A}$ and β_i due to its larger turn-on voltage.

The maxima in the Nb curve at a wavelength of 3 μm gives a quantum efficiency of 3.6%.

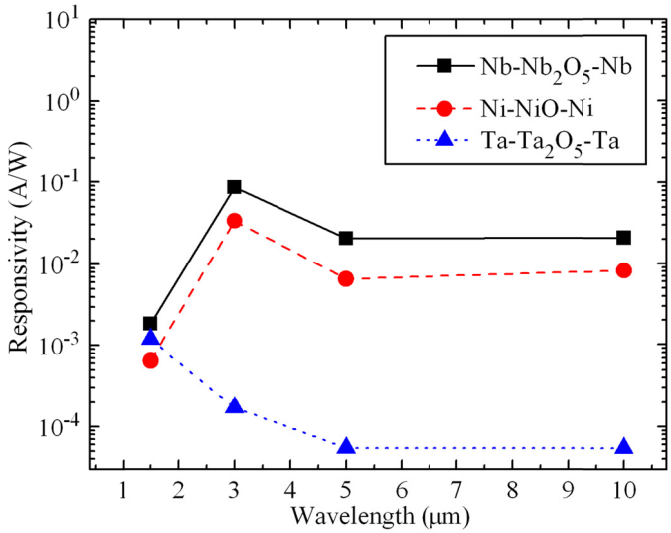


Figure VI-4 Calculated responsivity of the traveling-wave diode vs. free-space wavelength. The responsivity is higher for Ni and Nb since they have a lower-barrier oxide than Ta. The peak in responsivity of the Nb device corresponds to a quantum efficiency of 3.6%.

2. Comparison with the lumped-element detector

I compare the responsivity of the TW detector with a lumped-element detector that has an antenna impedance of 100 Ω and a diode area of 100x100 nm^2. Both the detectors are made from the Nb-Nb_2O_5-Nb MIM diode. The resulting comparison is shown in Figure VI-5. For the lumped-element device, the calculation is based on $\eta = \eta_c \beta_i$ where η_c accounts for the RC time constant and the impedance mismatch. The performance of the lumped-element detector degrades with increase in frequency. As explained in Chapter IV, only a minor improvement in the lumped-element device can be achieved by choosing a smaller area for the diode.

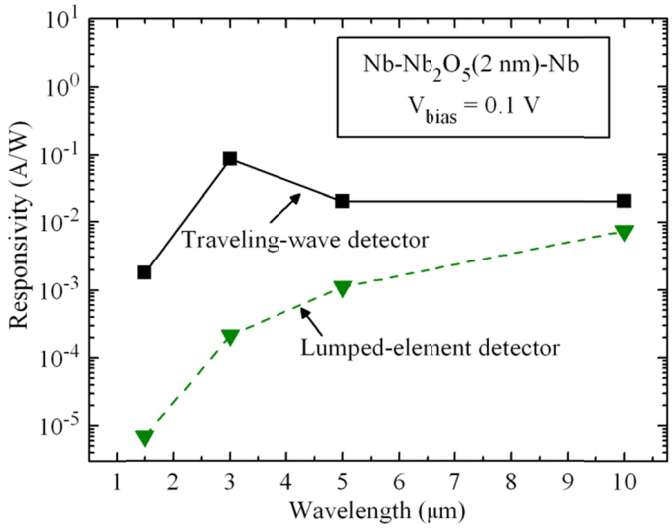


Figure VI-5 Calculated responsivity comparison of the lumped-element and the traveling-wave detectors. The traveling-wave detector shows significantly better performance that does not degrade at smaller wavelengths. The responsivity of the lumped-element detector decreases as the diode admittance increases with decreasing wavelength.

Thus, the responsivity of the TW detector is significantly improved and much less frequency dependent than the lumped-element detector. Both the impedance and the responsivity allow the detection bandwidth to extend over several microns

in the MWIR and the LWIR ranges. The spectral response of the TW detector can be tuned with an appropriate design for the antenna (Fumeaux, 1998).

D. Comparison with IR detectors

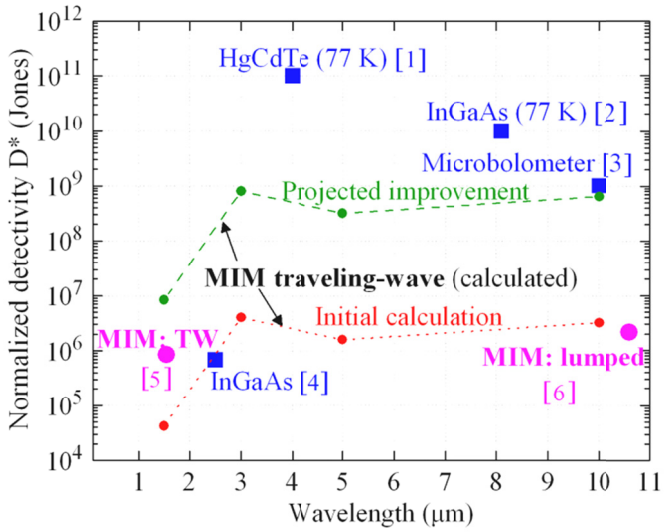
In order to compare the TW detector with other detector technologies used for infrared imaging, its normalized detectivity (D^*) (Rogalski, 2003) needs to be calculated. The D^* is a measure of the noise-performance of the detector and is defined as

$$D^* = (A_d \Delta f)^{1/2} \frac{\Re_{TW}}{I_n}$$

Eq. VI-8

where A_d is the detector area over which the radiation is received, Δf is the bandwidth of the readout circuitry, and I_n is the noise current in the diode.

The I_n is taken to be a combination of Johnson and shot noise (Spieler, 2005). In Figure VI-6, I compare the calculated D^* of the traveling-wave detector with some of the existing technologies. The initial calculation corresponds to the $Nb/Nb_2O_5/Nb$ traveling-wave detector responsivity shown in Figure VI-4. The D^* for the experimental TW detector (Phiar, 2002) is close to the predicted performance. The curve for the projected improvement in performance is based on several factors explained later in this section. If implemented, the improved noise performance will place the traveling-wave detector at par with thermal detectors (Rogalski, 2003).



References

[1] (Rogalski, 2000)
[2] (Gunapala, 2008)
[3] (Rogalski, 2000)
[4] (Jayaweera, 2008)
[5] (Phiar, 2002)
[6] (Tiwari, 2009)

Figure VI-6 Calculated detectivity comparison of the traveling-wave (TW) detector with semiconductor, thermal and MIM lumped-element detectors. The initial calculation of the TW device response is based on the Nb-Nb_2O_5-Nb device. With improvements, the performance of the TW detector can be at-par with thermal detectors.

With its projected performance capability on par with thermal detectors, the traveling-wave detector can be used in a range of applications including active IR imaging. A significant advantage of the TW detector over the semiconductor based ones is its ease of fabrication. With a small number of steps involving lithography and deposition of amorphous thin-films, the detector can be easily made into a focal-plane-array on top of existing CMOS circuits. The antenna preserves the phase of the incoming signal (Middlebrook, 2008), therefore the TW detector can also be used in IR phased-arrays. Unlike thermal detectors (Rogalski, 2000, p.68), the TW detector rectifies the incoming signal using tunneling and can support readout bandwidths in the terahertz. The large bandwidth is beneficial in communication applications where the detector can be used for sensing the envelope of amplitude-modulated signals (Rockwell, 2007).

1. Scope for improvement

Three criteria for improving the performance of the traveling-wave detector can be identified from Eq. VI-8. Since the detector is not limited by background radiation noise, i.e., it is not a background limited infrared photodetector (BLIP) (Rogalski, 2000, p.16), the receiving area can be increased without affecting the noise performance. Also, the responsivity of the TW diode, given in Eq. VI-5, can be improved. This can be done by improving the design of the traveling-wave structure and/or by using a diode with low resistance and high responsivity. These diode characteristics can be achieved by appropriately designing a multi-insulator diode as discussed in Chapter III. Finally, reducing the current noise I_n can also improve the D^*. This can be done by eliminating the shot noise, which is proportional to the DC current of the diode by operating at zero bias. For a rectenna to work at zero bias, diodes with asymmetric $I(V)$ characteristics having a large zero-bias responsivity are required.

CHAPTER VII

FABRICATION & CHARACTERIZATION OF GEOMETRIC DIODES

The concept of a geometric diode was introduced in Chapter I and the efficiency of rectennas based on it was analyzed in Chapter IV. Conductors with extremely small critical dimensions on the scale of the mean-free path length for charge carriers are required for geometric rectification. Successful fabrication at such dimensions demands extremely high resolution lithography. Low cost is a key criterion for the use of geometric diodes in optical-frequency detectors and photovoltaic rectifiers. Due to the planar structure and essentially a single lithography step process, a technique like nano-imprint lithography (Schift, 2008) can be used for low-cost-production. A glancing-angle deposition technique (Moddel, 2009) can also be used as a non-lithographic approach for making asymmetric conductors.

For the initial development of geometric diodes, I have focused on using electron-beam (e-beam) lithography. Here I give the fabrication details for the e-beam process carried out at the University of California, Santa Barbara. Antenna-coupled geometric diodes made from both metal and graphene thin-films were fabricated.

The $I(V)$ characteristics of the fabricated diodes were measured at DC. Rectifying behavior was also measured by illuminating the antenna-coupled diodes with infrared radiation.

A. Metal geometric diodes

The neck size in a geometric diode is required to be on the size of the mean free path, which, in a bulk-metal like gold is 42 nm (Ashcroft, 1976, pp.10,38). From a previous attempt at making aluminum geometric diodes (Zhu, 2010), it was realized that the extremely low resistance of a metal film is difficult to measure. Therefore I chose to make the gold geometric diodes with a metal-thickness of 25 nm along with a 5 nm chromium or titanium adhesion layer. The diode was designed to have a neck size of 20 nm as shown in Figure VII-1, and was connected at the feedpoint of a bow-tie antenna. The antenna was designed to receive 10.6 μm radiation (González, 2005) in an edge-fed configuration (Weiss, 2003).

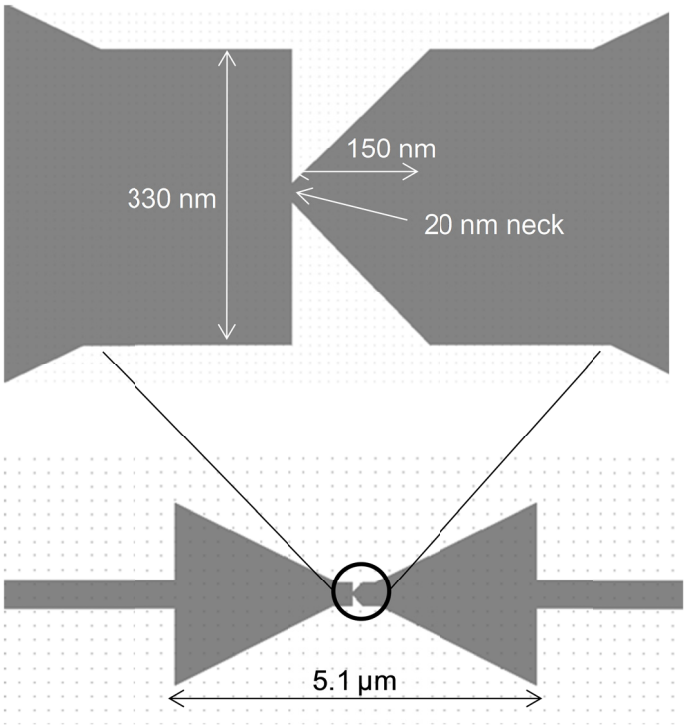


Figure VII-1 Antenna-coupled metal geometric diode with a 20 nm neck.

An oxidized silicon wafer with 750 nm oxide was used as the substrate on which Cr/Au (5/45 nm) metal-electrodes were placed using photolithography and lift-off. Alignment marks for e-beam lithography were also patterned during this step. The

next step was to pattern the shape of the geometric diode and the antenna. This was carried out using ZEP (styrene methyl acrylate based positive e-beam resist), which allows high resolution lithography. The lithography was done by Bill Mitchell in the Nanotech lab. at UCSB (Nanotech, 2011). A dose test was conducted to determine the appropriate exposure-dose for ZEP on the substrate used. After patterning the antenna region at a dose of 400 μC/cm^2, the optimal dose for patterning the diode was found to be 200 μC/cm^2. While writing the antenna region, the diode area gets partially exposed due to the proximity. Therefore, for patterning the diode, a smaller dose is required to compensate for this unintended exposure. The resulting patterns are shown in Figure VII-2 and Figure VII-3.

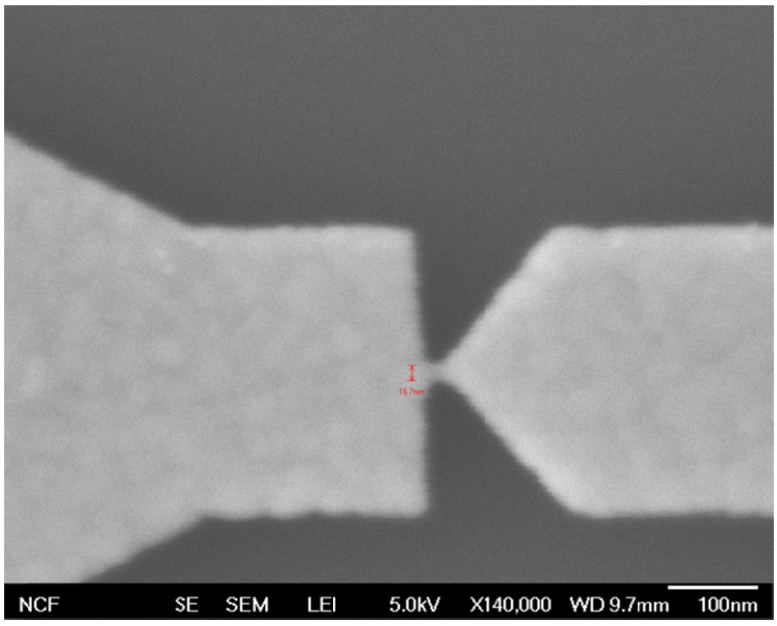


Figure VII-2 Scanning electron microscope (SEM) image of a gold geometric diode with a neck size of approximately 17 nm.

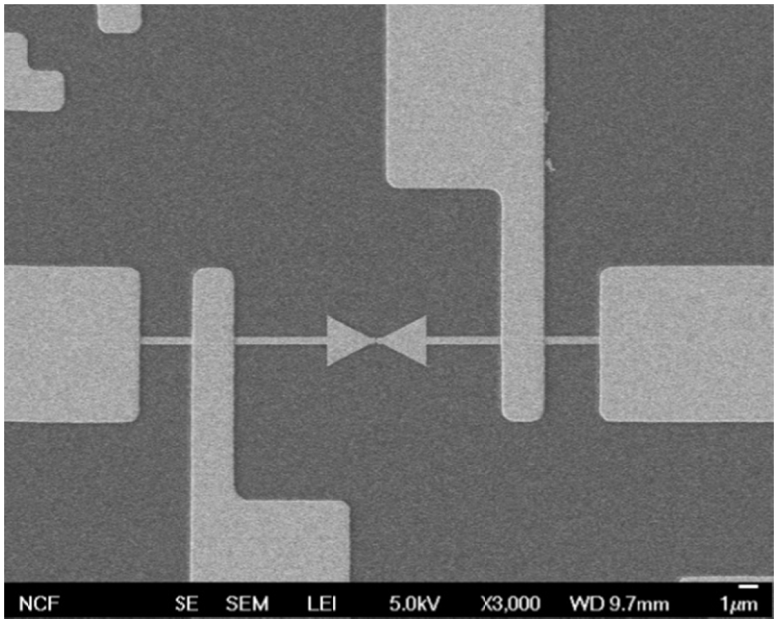


Figure VII-3 SEM image of the antenna and the four-point probe contacts for to the geometric diode shown in Figure VII-2.

After depositing the Cr/Au (5/25 nm) using e-beam evaporation, the ZEP resist was lifted-off using Shipley 1165 resist-stripper heated to 80 °C. These steps were also carried out by Bill Mitchell at UCSB.

The DC $I(V)$ curves for the metal geometric diodes were observed to be linear. A possible cause for the linear characteristics is the series resistance between the device and the probes, as seen in Figure VII-3. The devices were also tested for response to infrared illumination using the setup described in section C. No rectification was observed.

The failure of metal based geometric diodes can be attributed to the grain boundaries in the thin-film that limit the mean-free-path. As seen in the SEM image in Figure VII-4, the grain boundaries occur on a scale smaller than the bulk mean-free-path and comparable to the thickness of the film and the neck-size. The conduction is therefore dominated by electron reflections at the grain boundaries (Durkan, 2007, p.107) rather than at the edges of the conductor.

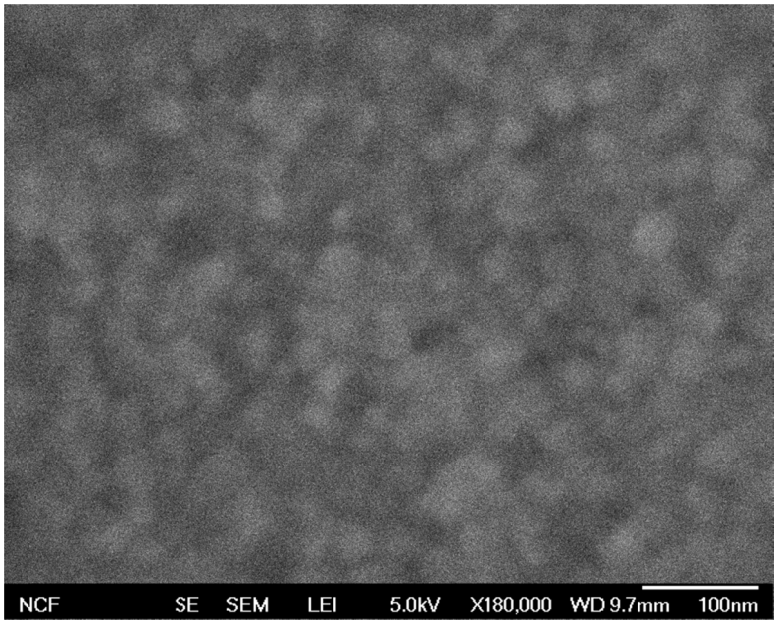


Figure VII-4 Grains in the evaporated Cr/Au thin-film.

B. Graphene geometric diodes

To overcome the mean-free path limitation of metal devices, graphene was chosen as the material for the geometric diode. As described in Chapter I, graphene is a two-dimensional crystalline sheet of carbon atoms with a large mean-free-path for charge carriers. The 2D electron gas in graphene is conducive for observing geometric effects as, unlike in the thin metallic films (Durkan, 2007, p.101), the thickness does not limit the mean-free path length. For the initial development, I chose to make both the diode and the antenna using graphene. Later efforts have modified the process to have a metal antenna connected to a graphene geometric diode.

The first step in the fabrication of graphene geometric diodes was the exfoliation of graphene flakes onto an oxidized silicon wafer with a 90 nm thick oxide. The 90 (or 300 nm) SiO_2 enhances the contrast of the graphene, making it easier to locate and to estimate its thickness using an optical microscope (Roddaro, 2007). Graphene films with 1-10 layers were used. As the graphene flakes had

random locations and orientations, they were separated by cleaving the wafer and processed individually. While cleaving the wafer, precautions were taken to prevent silicon dust from falling on the graphene. This was achieved by shielding the graphene using a glass slide that also serves as a line to scribe along.

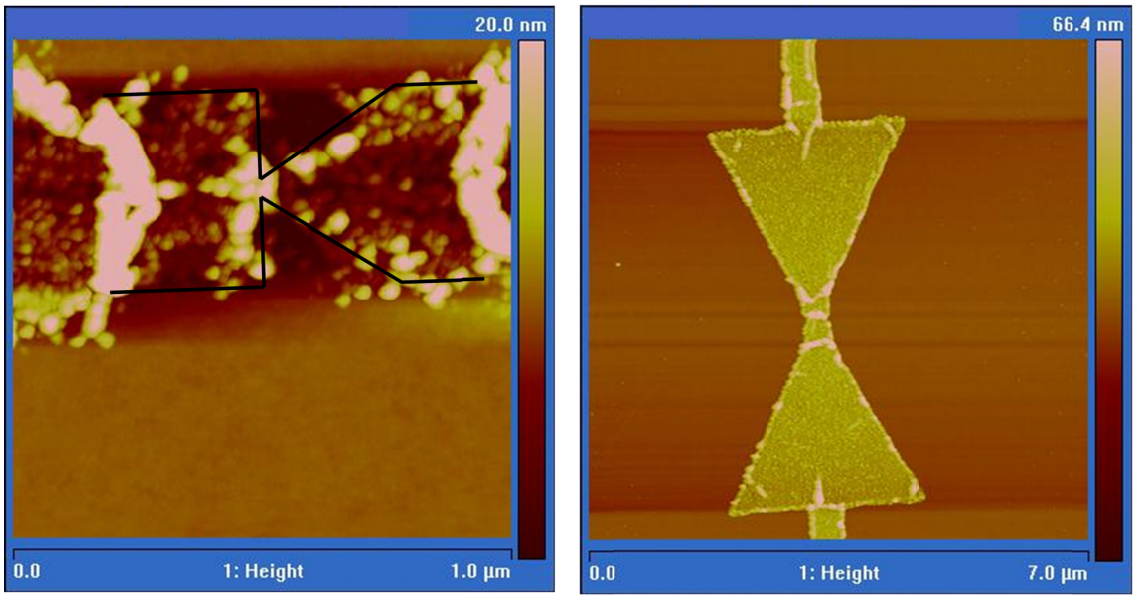
Cr/Au (5/45 nm) contacts for four-point probe measurement were placed using lift-off. The contacts also helped to keep the graphene clamped to the substrate during the subsequent processing steps. Negative resist NR9-1000PY was used for patterning the contacts. The resist was spun at 3000 rpm for 40 s and prebaked for 1 min at 150 °C. Alignment between the contact patterns on the mask and the graphene on the wafer was guided with scratch marks that were made while locating the graphene. The grayscale images displayed on the screen of the MJB4 mask-aligner made it easier to locate and align the graphene. Once aligned, the NR9 resist required a 3 s exposure on the MJB4. The resist was post-baked at 100 °C for 1 min followed by a 7 s dip (agitated) in the resist developer RD6. The usual step after developing a resist is to remove residual-resist layers using oxygen plasma. As graphene is sensitive to oxygen plasma this step was not performed. To compensate for the lack of a plasma-clean, I overdeveloped the resist by 3 s, without a noticeable change on the shape of the patterns. This was followed by a 20 s dip in DI water and N_2 blow-dry. The metals for the contacts were evaporated using a thermal evaporator. The lift-off was done in acetone (without ultrasonic) followed by an IPA rinse and N_2 blow-dry.

To pattern the graphene, e-beam lithography had to be done with a negative e-beam resist (hardens where exposed) so that the unprotected areas of graphene could be removed by etching in oxygen plasma. Negative e-beam resist maN-2403, which has a resolution of about 50 nm, was used with HMDS as an adhesion layer. A 300 nm thick resist layer was obtained by coating the wafer at 3000 rpm for 30 s. The exposure dose for e-beam lithography done at UCSB was between 1000-1400

$\mu C/cm^2$. The resist was developed for 45 secs in AZ 300MIF developer followed by a 60 s rinse in DI water. The exposed graphene was etched using oxygen plasma (100 W, 30 mTorr, 5 s) in the (Technics PEII-A) reactive ion-etch (RIE) at UCSB. Finally the resist was stripped by soaking the wafer in Shipley 1165 at 80 °C for 15 min.

Several precautions had to be observed during the fabrication of the graphene devices. After imaging the patterns generated on the resist with an SEM, stripping the resist became difficult. It was concluded that the SEM beam was depositing a thin carbonaceous coating on the imaged area (e-beam induced decomposition of background CO, CO_2, or hydrocarbons) that prevented the stripper from accessing the resist (Mitchell, 2010). Furthermore, ultrasonic could not be used to speed-up the lift-off or the resist removal as the contacts and graphene were prone to developing cracks.

An atomic force microscope (AFM) image of a graphene device is shown in Figure VII-5 (a). The diode is connected to a graphene antenna as shown in Figure VII-5 (b).



(a) (b)

Figure VII-5 (a) AFM image of a graphene geometric diode with a neck-size of 50 nm and shoulder of 400 nm. This diode was patterned at a dose of 1200 µC/cm^2. The bumps are due to residual e-beam resist. (b) AFM image of a graphene antenna coupled to the geometric diode shown in (a) rotated by 90°. The antenna was patterned at a dose of 1000 µC/cm^2. Again, the residual resist is visible along the edges.

The antenna-coupled graphene devices were not tested for DC $I(V)$ due to the series resistance of the leads and also because the graphene devices were prone to burn-out under the parameter analyzer (HP4145B). Graphene diodes without antennas fabricated using a similar process have subsequently been tested using a source-meter (Keithley2602/2612). As shown in Figure VII-6, these measurements were performed with a four-point probe setup to avoid nonlinearity at the metal-graphene contact (Perello, 2010), (Ran, 2009).

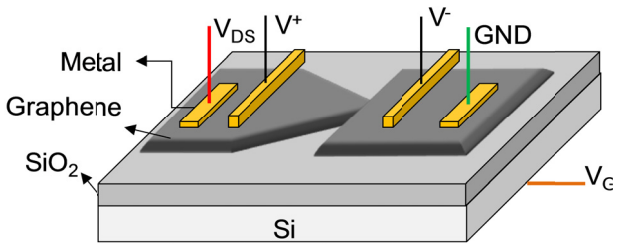


Figure VII-6 Four point probe configuration for measuring the diode I(V) without including nonlinearity at contacts.

In the above figure, the Si substrate is connected to a gate bias that modifies the Fermi level relative to the bandstructure of graphene. At zero gate voltage, the graphene is p-type due to its interaction with the SiO_2 substrate (Kang, 2008). The majority carriers in graphene can be made n-type by applying a positive gate voltage greater than the voltage corresponding to the Dirac point (Novoselov, 2005), which occurs near V_G = 25 V for the sample reported here.

The DC characteristics of a graphene geometric diode at gate voltages of 20 V and 40 V are shown in Figure VII-7. The asymmetry in current is consistent with the rectifying characteristics for majority p-type carriers for V_G = 20 V and n-type carriers for V_G = 40 V. At V_G = 20 V, the current is higher at positive than at negative V_{DS} indicating that the preferred direction of current flow is along the funnel, which is also the preferred direction for p-type carrier transport. The resistance is higher for the opposite voltage direction that blocks the flow. The asymmetry is also seen in the zero-bias responsivity. The opposite characteristics are observed for V_G = 40 V as the majority carriers are n-type. The preferred direction for charge flow is still along the funnel however n-type carriers require the opposite bias polarity to flow in that direction.

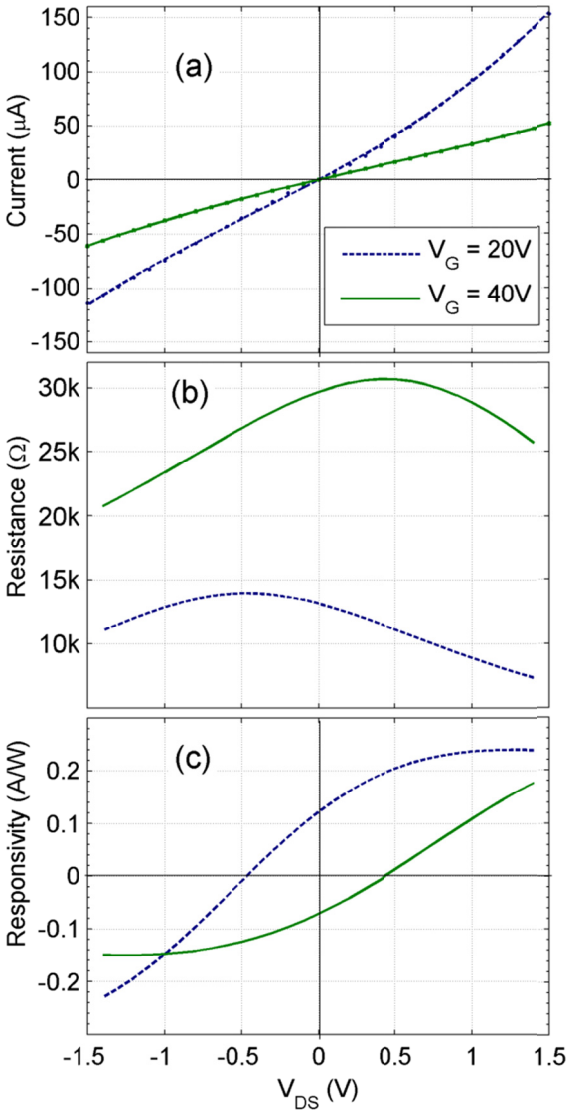


Figure VII-7 (a) Current, (b) resistance, and (c) responsivity vs. voltage for a graphene geometric diode made using exfoliated graphene.

To avoid artifacts from numerical processing of discrete points, a 4th order polynomial-fit is applied to the $I(V)$ curve to obtain the resistance (I') and responsivity (I''). The asymmetry is evident in the resistance, which is higher for the direction of bias that blocks the flow of holes. The asymmetry in the $I(V)$ curve also leads to a non-zero responsivity at zero-bias, thereby allowing geometric-diode-based rectennas to operate without an external bias.

C. Infrared characterization of antenna-coupled diodes

To verify the rectifying behavior of a geometric diode at optical frequency, I measured the response of antenna-coupled geometric diodes under illumination from a CO_2 laser (λ=10.6 μm). Performing this measurement at far-infrared instead of visible or near-infrared is preferred as the resistive losses, both in metals (Fox, 2001, p.146) and graphene (Jablan, 2009), are lower.

A Synrad CO_2 laser (Synrad, 2010), model number 48-1 SWJ, is used in the experiment. The laser is powered by a 30 V power supply and requires cooling. Using a function generator, a tickle (5 V) pulse with an on-time of 1 μs at a frequency of 5 kHz is applied to keep the gas ionized without lasing. The laser needs to be warmed-up in this state for more than 15 min to obtain a stable output. The high-time of the pulse is increased to enable lasing. The output of the laser is pulsed according to the applied signal. Therefore, the output power of the laser is controlled by varying the duty-cycle.

The response to illumination is measured by the DC short-circuit current (I_{SC}) or open-circuit voltage (V_{OC}) generated by rectification. Due to the small nonlinearity of the diode, V_{OC} and I_{SC} are expected to be extremely low. This necessitates the use of a lock-in detection technique with the setup shown in Figure VII-8.

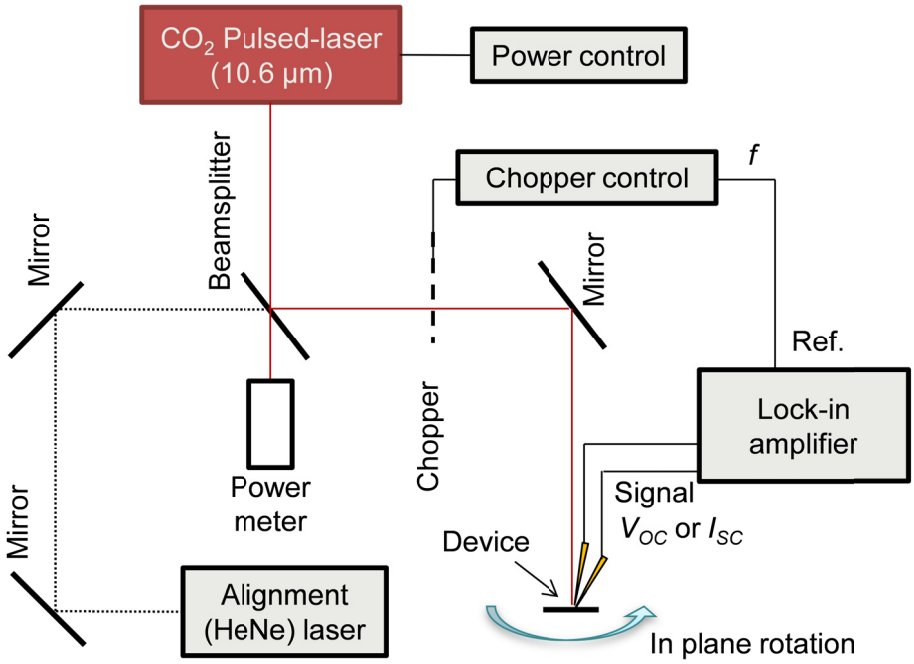


Figure VII-8 Setup for measuring response of antenna-coupled geometric diodes to 10.6 µm illumination. The wafer is rotated in the horizontal plane to change the angle between the polarizations of the radiation and the antenna.

Infrared radiation from the CO_2 laser is split into two beams, one of which is used for measuring the power and the other is directed onto the wafer. A HeNe laser is used to align the infrared beam onto device, which gets blocked during measurement. A heat sink is used to block the IR laser prior to and after the measurement. The wafer is rotated to change the angle (θ) between polarization of the radiation and the antenna. The radiation emitted by the laser is vertically polarized.

The infrared beam is chopped at a frequency (f). The rectified signal in the device is also modulated at the same frequency. From the chopper controller, a frequency reference is fed into the reference port of the lock-in amplifier which then measures the amplitude of the signal from the device at that frequency. The lock-in input can be set to detect either current or voltage.

The variation of I_{SC} with θ for a graphene geometric diode connected to a graphene antenna is shown in Figure VII-9. The maximum response occurs at $\theta = 0^{o}$ when the antenna is aligned parallel to the electric-field. An order of magnitude calculation for the efficiency of rectification is calculated by the responsivity given as the ratio of the current generated to the approximate power incident on the antenna. The laser beam has a diameter of 4 mm at the source and expands at a rate of 4 mm for every 1 m of propagation. In the absence of an exact knowledge of the beam center and its alignment with respect to the antenna, I approximate the laser beam as uniform over an 8 mm diameter. For a total power of 1 W, chopped at 50% duty cycle, the intensity on the wafer is 0.5 W/cm^2. With an antenna receiving area of 10 μm^2 the power received is 50 nW. At θ=0^{o}, this leads to a responsivity of 3 pA/50 nW ~ 6x10^{-5} A/W. The low efficiency can be attributed to a number of loss mechanisms including antenna efficiency, impedance mismatch, and poor quantum-efficiency of the diode.

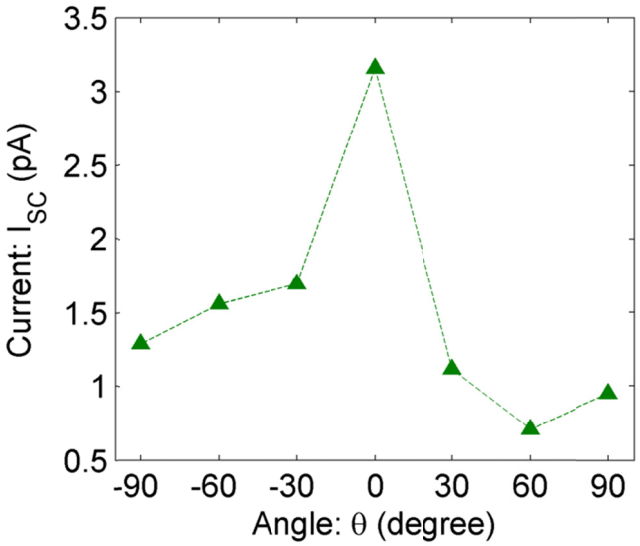


Figure VII-9 Short circuit current I_{SC} vs. angle between field and antenna polarization. Maximum response occurs when the field is parallel to the antenna.

As the complete device is made out of graphene, the sheet resistance is constant throughout the antenna and the diode. The rectified current generated in the geometric diode flows through the antenna and the leads before being detected, which weakens the current signal measured by the lock-in detector. This is avoided by measuring the open-circuit voltage. In Figure VII-10, I plot the V_{OC} vs. θ for varying power of the laser.

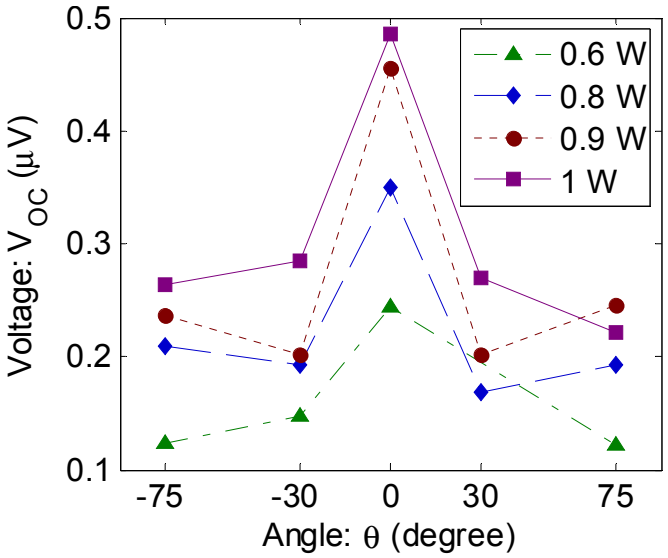


Figure VII-10 Open-circuit voltage vs. angle between the polarization of antenna and radiation for varying power of the laser source. As with I_{SC}, peak response occurs at θ=0°.

For θ=0° and 1 W power on the laser, the diode + series resistance is given by the ratio V_{OC}/I_{SC} (=R_{D+S}). The measurements shown in Figure VII-9 and Figure VII-10 are on different devices, therefore the calculation of resistance gives at best an order of magnitude estimate. The R_{D+S} = 0.5 μV/3 pA = 167 kΩ also includes the resistance at the contacts.

The polarization (θ) sensitive infrared response of the antenna-coupled geometric diode indicates that the signal occurs due to the presence of the antenna and thereby a device a connected to the antenna. Ideally, the polarization sensitive

response should be smallest at ±90°. As seen in Figure VII-9, this may be true for the negative angles but is not the case for positive angles. Also, in Figure VII-10, this is not the case for all power levels. As the response is extremely small, it is conceivable that this is due to measurement error. It could also be due to variation in the position of the device with respect to the IR beam in going from one rotation to the other. A possible way to eliminate the latter is to mount the device on a TO-8 can, making sure that the position of the rectenna coincides with the rotation axis of the can.

The response observed is not bolometric (Richards, 1994) as both V_{OC} and I_{SC} were measured in the absence of any external bias. Thermo-electric effect arising at the metal-graphene interface due to unequal heating of different regions can lead to a V_{OC} and I_{SC}. However, it does not explain the θ dependence.

CHAPTER VIII

QUANTUM SIMULATION OF GEOMETRIC DIODE

In Chapter I, I have described the operating principle of the geometric diode using the Drude model and also using a wavefunction approach. Characteristics of geometric diodes simulated using the Drude model (Zhu, 2011) are consistent with the expected direction of asymmetry in the $I(V)$ curve. However, these simulations are based on a classical model and account for the shape of the conductor through the specular reflection of charges from the boundaries. Moreover, a constant field distribution across the patterned thin-film is assumed, without accounting for the effect of charge or shape on the potential. Material properties only enter through the mean-free-path length and Fermi velocity. Finally, it is not possible to model the effect of high-frequency excitation through this approach.

Several formulations exist for modeling the transport properties in nanoscale systems (Ventra, 2008). A particularly interesting approach for modeling ballistic transport based on the transfer-matrix method was applied to an irregular graphene waveguide (Li, 2009). However, this approach assumed the wavefunction to be coherent across the device. Modeling the geometric diode accurately requires inclusion of phase-breaking mechanisms and determining the charge and potential self-consistently. Furthermore, simulating the response of a geometric diode to a high-frequency signal requires modeling the photon-assisted transport arising from the interaction of the charge-carriers with quanta of energy ($=\hbar\omega$) of the AC signal. A technique that allows such effects to be incorporated at the appropriate level of complexity is the non-equilibrium Green's function (NEGF) method. Starting with the Hamiltonian that describes the material and the structure of the geometric

diode, the NEGF approach can be used to model geometry-dependent transport, effect of contacts, and interaction of electrons with photons and phonons.

I model the characteristics of an asymmetrically-shaped graphene-film using the NEGF method. Development of the simulator is an ongoing project and in its current version, it can model ballistic transport through graphene while calculating the charge and potential self-consistently. The simulator can be further improved to include scattering and, as explained in Chapter IX, it can be used in conjunction with a semi-analytical perturbation approach for modeling photon-assisted transport.

Though all the results in this chapter are specific to the 2D hexagonal lattice of graphene, the method can be extended to any lattice. A numerical solution is obtained by converting the physical problem defined by the NEGF formalism into a set of matrix equations that are solved to obtain the $I(V)$ curve for the geometric diode.

A. NEGF formalism

The Green's function is used to solve inhomogeneous differential equations in a manner similar to the use of Fourier series for solving partial differential equations (Weisstein, 2011). In quantum physics, the Green's function provides information regarding experimental observables and allows the construction of a systematic perturbation theory (Haug, 2008, p.41). An added level of complexity is introduced if the system under consideration is open, i.e. it interacts with the surroundings. For the geometric diode, such an interaction occurs due to the flow of current to and from the contacts and can be simulated using the NEGF approach (Datta, 2002).

Consider a central device region with contacts at two ends that act as charge reservoirs held at fixed potentials as shown in Figure VIII-1. The transport

properties, shape, charge, and field in the device region are modeled by the Hamiltonian (H) and the potential (U). In the absence of contacts, the solution to (E - [H+U])Ψ=0 gives discrete eigenenergies for the device region.

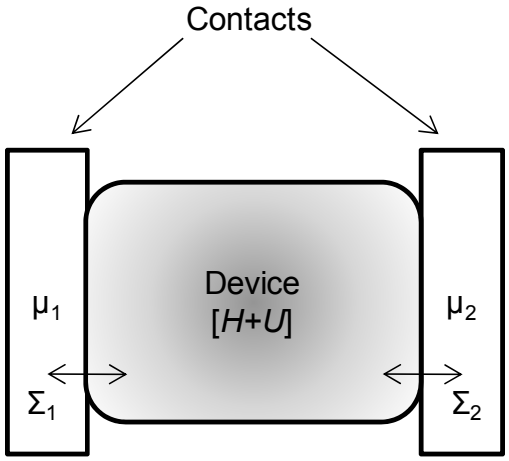


Figure VIII-1 A central device region with two contacts that act as charge reservoirs that are held at fixed potentials (μ). The device is represented by a Hamiltonian H and potential U.

The effect of the contacts on the central device region is three-fold (Datta, 2005, p.183). First, the contacts determine the potential boundary-conditions at the two ends of the device. Second, they act as charge reservoirs from which carriers are injected (or extracted) with an escape-rate (γ) describing the quality of the connection between the device and the reservoirs. Third, on connecting a set of discrete states (device) with a continuum of states (contact) there is a 'spillover' of states from the contact to the device. The last two effects are a consequence of the *time*energy* uncertainty relation between the lifetime of an interacting state and its spread in energy (Datta, 2005, p.11).

Even though the reservoirs are significantly larger in size compared to the device, their effect on the device is modeled by a self-energy matrix (Σ) related to the escape rate, that has the same size as H (Datta, 2000). A detailed explanation of this is given in the next section. Unlike H, Σ is a function of energy (E).

Scattering mechanisms can be modeled using NEGF through the appropriate choice of a self-energy matrix (Σ_S) corresponding to a virtual scattering-contact (Golizadeh-Mojarad, 2007). Unlike the reservoir contacts, the scattering contact is not held at a fixed potential, and Σ_S is determined differently than Σ. As it affects and gets affected by the charge distribution in the device region, Σ_S has to be determined self-consistently with the charge. In the results presented in the later sections, I have not accounted for scattering. However, it can be added to the basic framework of the simulator that I have developed.

The formulation of the simulator along with the explanation of the relevant quantities is given below. If the reader finds it difficult to follow the matrix-version of these equations, understanding the concepts for a single energy-level may be helpful (Datta, 2008).

For an isolated system, the eigenenergies and the corresponding set of wave functions $\{\Psi\}$ can be found out from the Schrödinger equation

$$\left[EI-\left(H+U\right)\right]\{\psi\}=0$$

Eq. VIII-1

where I is the identity matrix. In an open system, Eq. VIII-1 is modified by a broadening matrix Γ which is the matrix version of the escape-rate γ.

$$\left[EI-\left(H+U\right)+\frac{i\Gamma_1}{2}+\frac{i\Gamma_2}{2}\right]\{\psi\}=0$$

Eq. VIII-2

By adding Γ, the wavefunction in the device gets modified such that it decays with time. The self-energy matrices are related to the broadening as

$$\Sigma=-\frac{i\Gamma}{2}$$

Eq. VIII-3

More generally, Σ can have a real part as well, which represents the effect of the contact on the device Hamiltonian. The outflow or the decay of electrons from the device is compensated by the reservoirs that act as a source. Therefore a source-term $\{S\}$ needs to be added to Eq. VIII-2 as

$$\left[EI-(H+U)-\Sigma_1-\Sigma_2\right]\{\psi\}=\{S\}$$

Eq. VIII-4

The strength of the source, representative of the number of electrons coming into the device, is expressed as

$$\left[\Sigma^{in}\right]=\{S\}\{S\}^{+}=[\Gamma]f$$

Eq. VIII-5

Heuristically, Σ^{in} depends on the Fermi distribution (f) that determines the occupation of energy-levels in the reservoir, and the coupling between the device and the contact given by Γ.

A solution to Eq. VIII-4 is obtained as

$$\{\psi\}=[G^r]\{S\}$$

Eq. VIII-6

where G^r is the retarded Green's function given by

$$G^r=\left[(E+i\eta)I-(H+U)-\Sigma_1-\Sigma_2\right]^{-1}$$

Eq. VIII-7

in which η is an infinitesimal positive number added to make the Green's function zero for $t<0$ (retarded). As seen in Eq. VIII-5, the physical quantities in the system (e.g. the strength of the source) are given by the square of the wavefunction-like terms. This is analogous to measuring intensity instead of field. Similarly, G^r by

itself has no physical significance. However, two essential quantities that can be derived from it are the electron density and the density of states.

G^n represents the correlation of the wavefunction, and its diagonal elements give the electron density for the corresponding location at a particular energy.

$$G^n = \{\psi\}\{\psi\}^+ = G^r\{S\}\{S\}^+ G^{r+} = G^r \Sigma^{in} G^{r+}$$

Eq. VIII-8

Here, Σ_{in} includes the contribution from both contacts

$$\Sigma^{in} = \Sigma_1^{in} + \Sigma_2^{in} = \Gamma_1 f_1 + \Gamma_2 f_2$$

Eq. VIII-9

The density of states is given by the diagonal elements of the spectral function (A) defined as

$$A = G^r \Gamma_2 G^{r+} + G^r \Gamma_1 G^{r+}$$
$$or$$
$$A = i[G^r - G^{r+}]$$

Eq. VIII-10

The expression for A is derived from the sum of G^n and a function similar to it with f replaced by (1-f) in Eq. VIII-9, such that both filled and empty states are counted.

Finally, the current in the device from both the contacts is calculated for a particular energy as

$$I(E) = I_1 + I_2 = \frac{e}{h} Trace\left(\left[\Sigma_1^{in} A - \Gamma_1 G^n\right] + \left[\Sigma_2^{in} A - \Gamma_2 G^n\right]\right)$$

Eq. VIII-11

The $\Sigma^{in}A$ term represents the current flowing from the reservoir to the device and depends on the strength of the source and the density of states in the device.

The ΓG^n term represents the current flowing out of the device, which is a function of the escape-rate and the electron density in the device. The total current is the integral of Eq. VIII-11 over energy. For coherent transport, an alternative expression for current is written in terms of a transmission distribution $T(E)$ as (Datta, 2000)

$$I(E) = \frac{e}{h} T(E)(f_1 - f_2); T(E) = Trace\left(\Gamma_2 G \Gamma_1 G^+\right)$$

Eq. VIII-12

I will now describe the procedure for obtaining H and Σ.

B. Tight-binding model (H)

The graphene lattice consists of hexagonally arranged carbon atoms. Depending on the angle of the edges with respect to the lattice, two edge profiles can be obtained. Based on the type of edge, the graphene is referred to as zigzag and armchair. Although I use the example of an armchair graphene, zigzag graphene can be simulated by appropriately modifying the H (Datta, 2009).

Using a nearest-neighbor tight-binding model (TBM) (Reich, 2002), the Hamiltonian (H) of the device region is given by a matrix whose diagonal elements give the Bloch energy (ε_{2p}) at each atomic location and the off-diagonal elements (γ_1) give the interaction energy with the first nearest-neighboring atoms. Interactions with second- and third-nearest neighboring atoms can also be taken into account. However, the change in characteristics is small (Wu, 2011). For graphene, $\varepsilon_{2p} \approx 0$ eV, $\gamma_1 = -2.7$ eV. In Section VIII-F, I will introduce the effect of a gate voltage on the Fermi level in graphene through a term E_{offset}, which is accounted for as $\varepsilon_{2p} = E_{offset}$.

The H-matrix is of the size NxN where N is the number of atoms in the device. The hexagonal structure of the graphene is translated into matrix form by appropriate assignment of the interaction energies. In Figure VIII-2, an armchair

graphene structure is divided into sub-matrices for ease of formulating the *H*-matrix.

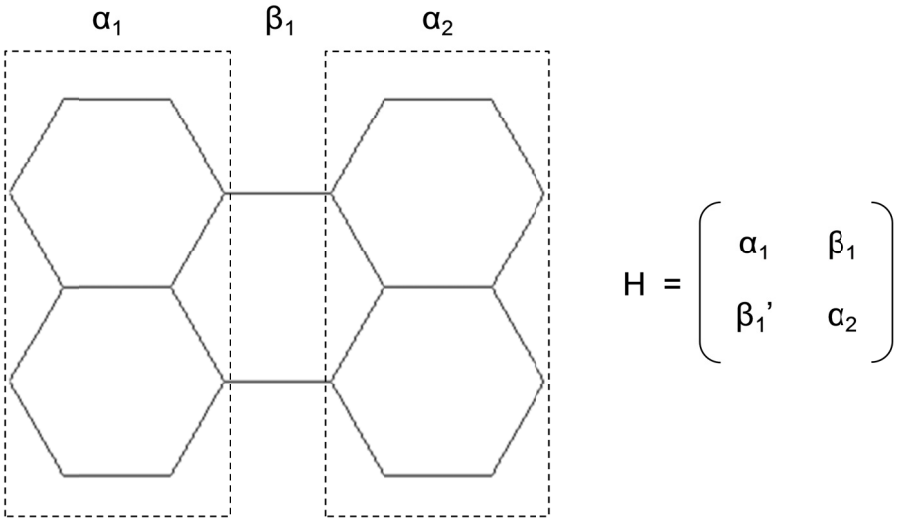


Figure VIII-2 Determining the tight-binding Hamiltonian (*H*) for armchair graphene.

The on-site energy for each carbon atom is contained within the α-matrices along with the nearest-neighbor interaction between the atoms enclosed in the same rectangle. The interaction between two α-matrices is provided in the β-matrix. The TBM Hamiltonian consists of the α-matrices along the diagonal and the β-matrices (and their transpose) as the upper (and lower) off-diagonal terms. The α-matrices have a dimension of nxn, where n is equal to the number of atoms enclosed by the rectangle. The β-matrices have a dimension equal to nxm, where n(m) is the number of atoms enclosed by the rectangle on the left (right).

The simulator is programmed to accept an input for the neck and shoulder sizes, along with the angle of the slanting edges. It then generates a rectangular sheet of armchair graphene and cuts out a right-triangle from the top and bottom to obtain the asymmetric hourglass geometry shown in Figure VIII-3. The width of the uncut-regions on the left and the right can be varied.

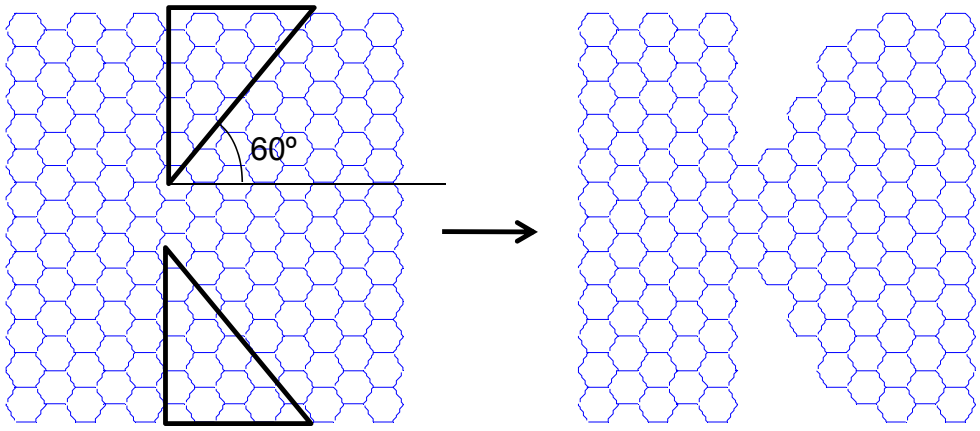


Figure VIII-3 Programming a geometric-diode-shaped graphene sheet starting from a rectangular piece. The patterned graphene has 5 nm shoulders, 1 nm neck and an angle of 60°. The neck size is greater than the programmed 1 nm as the incomplete hexagons formed on the top and bottom of the neck having more than 5 atoms around the center are assumed complete.

For the graphene geometric diode in Figure VIII-3, the assignment of α and β are shown in Figure VIII-4.

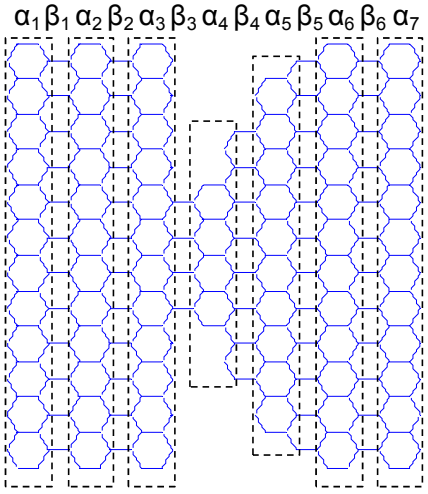


Figure VIII-4 Assignment of α and β matrices for the patterned graphene sheet of Figure VIII-3.

C. Contact self-energy (Σ)

In an experimental device, the electrical contact to graphene would be made using a metal. A tight-binding description of the graphene-metal interface is feasible (Blanter, 2007). However, such an interface may not have linear $I(V)$ characteristics (Ran, 2009). In order to ensure that the contacts do not contribute to the nonlinearity or asymmetry in the simulated $I(V)$ curve, I choose to model the contacts as semi-infinite graphene leads stretched out to the left and the right instead of metal leads.

A semi-infinite contact with a periodic structure can be modeled using a surface Green's function (g_R) (Datta, 2005, p.190). The g_R is a subset of the overall Green's function in the reservoir and accounts for the atoms in the reservoir closest to the device region. This simplification is based on the periodic lattice of the contact and accounts for the atoms in the first period. Due to the limited computational capability, I restrict the width of the contact to be equal to the shoulder-size of the diode. If α_1 and α_7 in Figure VIII-4 extend indefinitely to the left and the right respectively, the contacts are represented by α (= α_1 = α_7) and β (=β_1 = β_6) matrices as shown in Figure VIII-5. Condensing the vertical direction into matrix form allows a quasi-1D representation of the contacts.

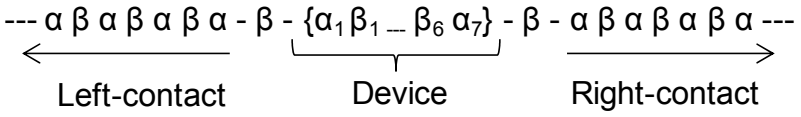


Figure VIII-5 Representation of semi-infinite quasi-1D reservoirs connected to a device.

To obtain g_R, the following equation needs to be solved for g iteratively (Datta (a), 2009)

$$g_{n+1} = \left[(E + i\eta) I - \alpha - \beta g_n \beta^+ \right]^{-1}$$

Eq. VIII-13

The self-energy matrix Σ is calculated from g_R as

$$\Sigma = \beta g_R \beta^+$$

Eq. VIII-14

For a description of the iterative procedures for solving Eq. VIII-13, see the work of Sancho and Rubio (Sancho, 1984) and references therein. To avoid an iterative solution that can be computationally intensive, a closed-form solution was proposed (Umerski, 1997). However, both the iterative and the closed-form methods are unstable due to the ill-conditioned β-matrices for graphene. The difficulty arises in calculating the inverse of matrices with determinant equal to zero.

A two-step semi-analytical approach for finding the self-energy, even with ill-conditioned matrices, is given by Rungger and Sanvito (Rungger, 2008). In the first step, the singular (non-invertible) matrix is rewritten in a form that separates the singularities. This is called singular value decomposition (SVD) (Weisstein (a), 2011). Next, the singularities are removed by reducing the size of the matrix. The algebra for finding the self-energy is performed on a reduced set of matrices followed by the inverse process of SVD and singularity-elimination to get back to the original form. I have implemented this method in the simulator.

D. Self-consistent NEGF-Poisson solver

In the previous two sections, I have outlined the procedure for solving the quantum physics of the geometric diode. Another factor that needs to be accounted for in the overall simulation is the potential distribution across the device. The potential is calculated by solving the Poisson's equation with appropriate boundary conditions. Since the potential affects the solution of the NEGF equations, which in turn modifies the Poisson's equation via the charge distribution in the device, the two equations need to be solved self-consistently (Datta, 2005, p.163).

I use the partial differential equation (PDE) solver in MATLAB (MATLAB, 2010) to solve the Poisson's equation on a 2D region defined by the boundary of the patterned graphene region as shown in Figure VIII-6. Dirichlet boundary conditions are applied on the left and the right ends as the voltage there is held fixed by the reservoirs, while the top and the bottom boundaries are floating with Neumann boundary conditions.

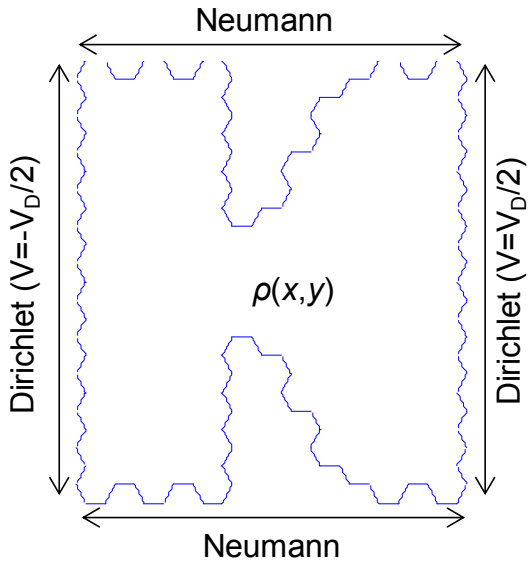


Figure VIII-6 Geometry for solving Poisson's equation with appropriate boundary conditions and charge distribution.

The solver is implemented by first calculating the potential distribution U_0 without any charge (ρ_0=0). This is used as a starting value in the NEGF equations and the spatial distribution of the charge is obtained as

$$\rho = \frac{1}{2\pi Vol.} \int dE \; G^n(E)$$

Eq. VIII-15

where *Vol.* is the volume associated with each atomic location. This is approximated as $a^2 * t_G$ where a (=2.46 Å) is the lattice constant and t_G (=3.4 Å for monolayer) is the thickness of the graphene.

Using the charge ρ_1 calculated from Eq. VIII-15 as a starting point for the Poisson's equation, an updated potential U_1 is obtained. Convergence of the self-consistent procedure implies that the difference between the updated (U_1) and the previous (U_0) potential should be less than a critical value (*RelTol*).

$$change = \frac{\sum_x \sum_y |U_1 - U_0|}{\sum_x \sum_y |U_1 + U_0|} < RelTol$$

Eq. VIII-16

Until Eq. VIII-16 is satisfied, the self-consistent NEGF-Poisson loop shown in Figure VIII-7 is repeated with a new estimate for the potential (U_{new}) used for solving the NEGF equations.

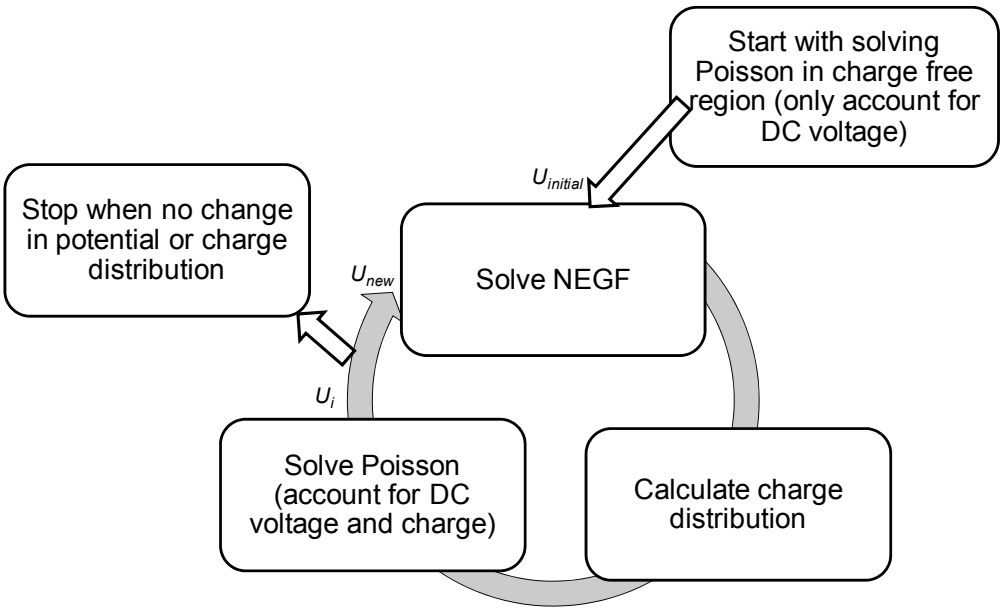


Figure VIII-7 Self-consistent NEGF-Poisson solver.

If $U_{new} = U_i$ is used, the convergence is slow. A faster convergence is obtained by predicting U_{new} based on several previous U's. The prediction technique that I have used is called Anderson mixing (Eyert, 1996) and is based on minimizing the residual vector. The residual vector is the difference in potential used in the NEGF

equations and the potential obtained from solving the Poisson equation. Anderson mixing calculates weights for the linear combination of previous U's such that the residue is minimized. Here I choose to keep no more than 5 U's in the memory.

E. A simulation example

I will now compute and explain the various quantities that are obtained from the simulation conducted on the geometry shown in Figure VIII-3. Applying a bias of V_D = 0.28 V across the graphene results in the potential distribution shown in Figure VIII-8. This is the starting point of the iterative loop in Figure VIII-7, obtained by solving the Poisson's equation in a charge-free region.

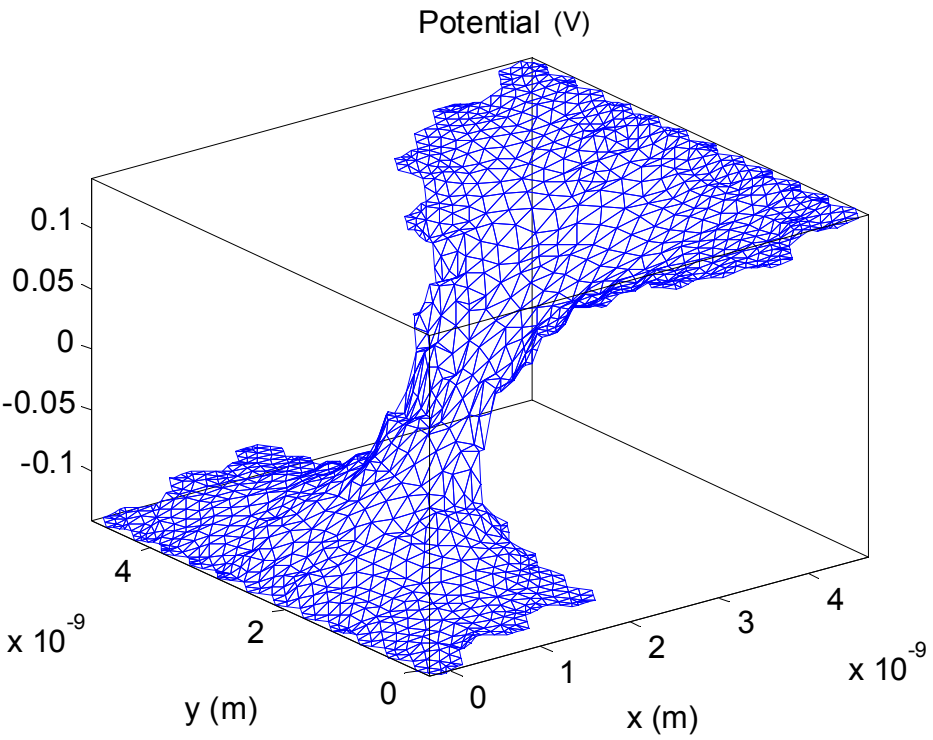


Figure VIII-8 Potential distribution in a charge-free graphene at V_D = 0.28 V.

On the termination of the self-consistent iterative procedure described in the previous section (with a *RelTol* of 10^{-4}), the potential distribution takes the form shown in Figure VIII-9.

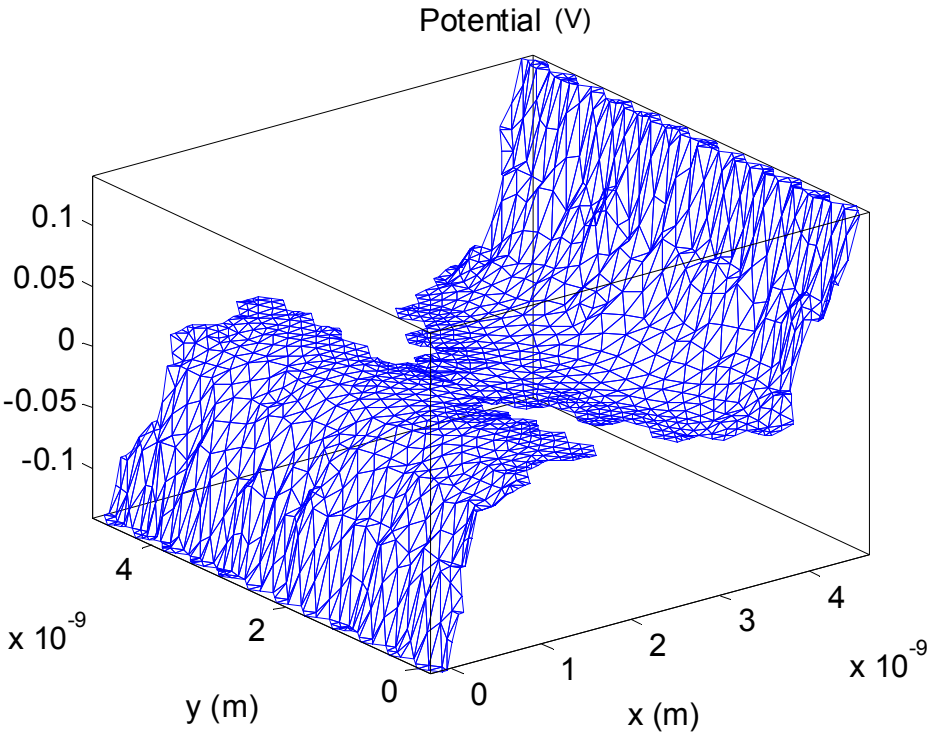


Figure VIII-9 Self-consistent potential distribution after accounting for the charge in the graphene.

Almost all the voltage drop in the potential distribution of Figure VIII-9 occurs equally at the two contacts. This is due to the absence of scattering in the device region (Datta, 2000). The voltage drop also results in a higher charge density near the contacts with the electrons accumulated at the positive- and holes at the negative-voltage as shown in Figure VIII-10.

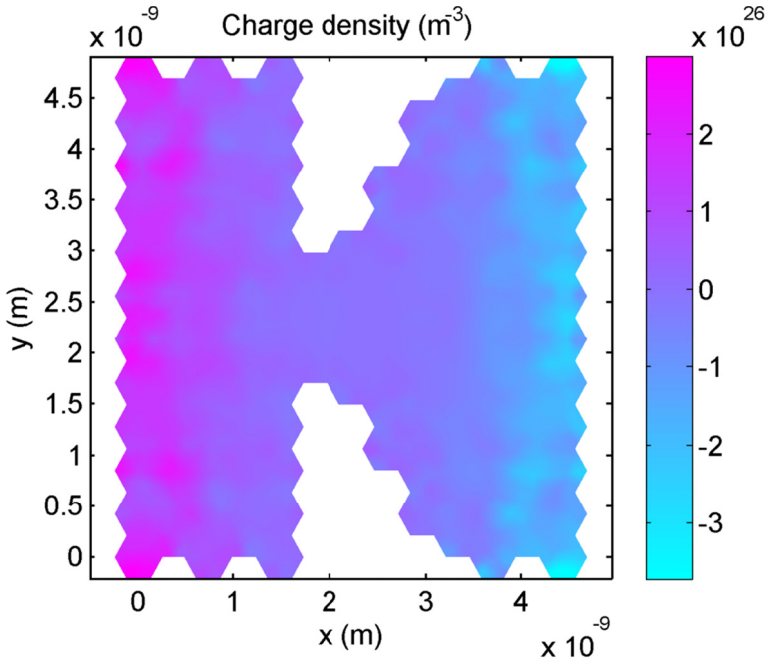


Figure VIII-10 Self-consistent charge distribution corresponding to the potential shown in Figure VIII-9.

I calculate the current using Eq. VIII-12, which requires finding the transmission as a function of energy as shown in Figure VIII-11.

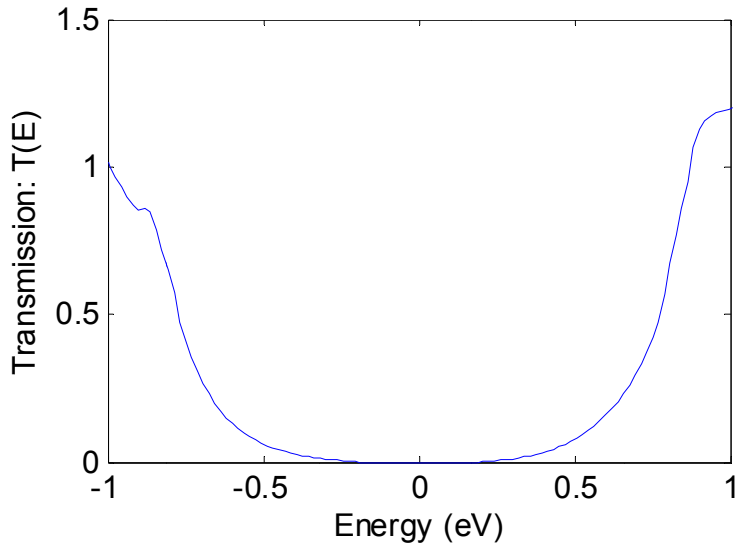


Figure VIII-11 Transmission vs. energy obtained from the self-consistent solution.

At negative energies, the transmission is due to holes and at positive energies it is due to electrons. The confinement due to the neck introduces a bandgap (0.37 eV in this case) that prevents the transmission at energies close to zero. This bandgap is also observed in the plot of density of states vs. energy shown in Figure VIII-12.

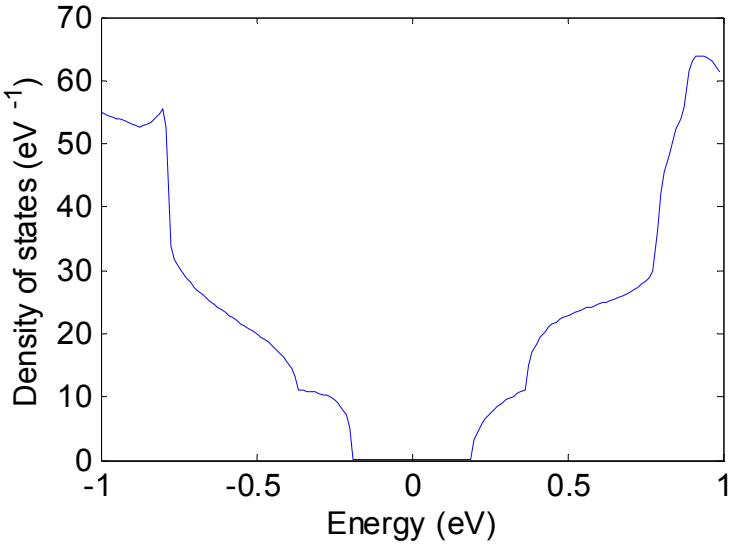


Figure VIII-12 Density of states vs. energy. The bandgap and step change with increasing energy are evident.

The density of states has a step change with energy due to the abrupt addition of conduction channels with increasing energy. The rounding of the steps occurs due to the interaction with the contacts that leads to broadening of the channels in the device.

In Figure VIII-13, I plot the energy dependent current calculated from the transmission $T(E)$ and Fermi distribution according to Eq. VIII-12. Even though the $T(E)$ increases at higher $|E|$, the difference in Fermi distributions on the left- and right-contact only allows conduction in a narrow range of energies near the bandgap.

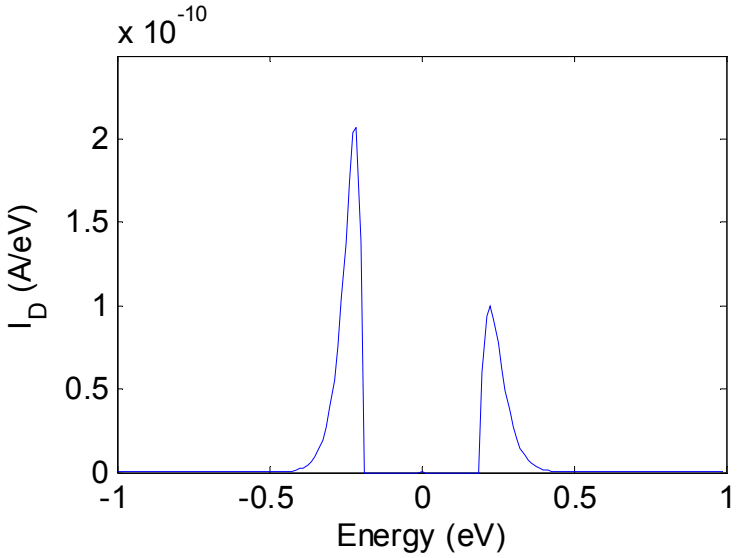


Figure VIII-13 Current vs. energy calculated from $T(E)$ and Fermi distribution.

The integral of the above curve gives the total current at the DC bias of 0.28 V.

F. Simulated geometric diode $I(V)$ characteristics

Electrons and holes move in opposite directions under a drift field. However, the geometric effect funnels both polarities of charges in the same direction. For an equal contribution from electrons and holes to the total current, the geometric effect gets canceled and reversing the voltage does not lead to a change in the magnitude of current. Hence, there is no asymmetry in the $I(V)$ curve and thereby no geometric rectification under such a condition.

As seen in Figure VIII-13, electrons and holes have unequal contributions to the current. This unequal, net p-type conduction is attributed to the sharper rise in density of states for holes as compared to electrons near the bandgap as seen in Figure VIII-12. Even though it is not visible in Figure VIII-11, this difference also exists for $T(E)$. The type of carriers that carry the larger current is determined by the direction in which the asymmetry favors transport in conjunction with the

applied bias, and also by the relative location of the Fermi-level with respect to the conduction and valence bands of graphene.

The effect of asymmetry and bias on the contribution of n- and p-type carriers to the current is shown in Figure VIII-14, where I plot the current vs. energy for a positive and a negative biased diode.

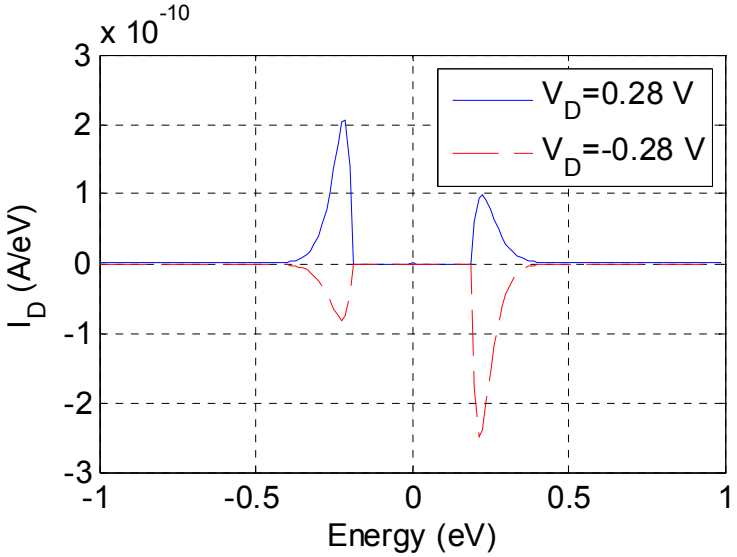


Figure VIII-14 Current vs. energy for opposite voltages across the diode.

As seen before, for the positive bias, p-type carriers have a larger contribution as they flow in the direction favored by the geometry. However, for the same reason, at negative bias n-type carriers dominate. Overall, the current under negative bias is larger as seen in Figure VIII-15.

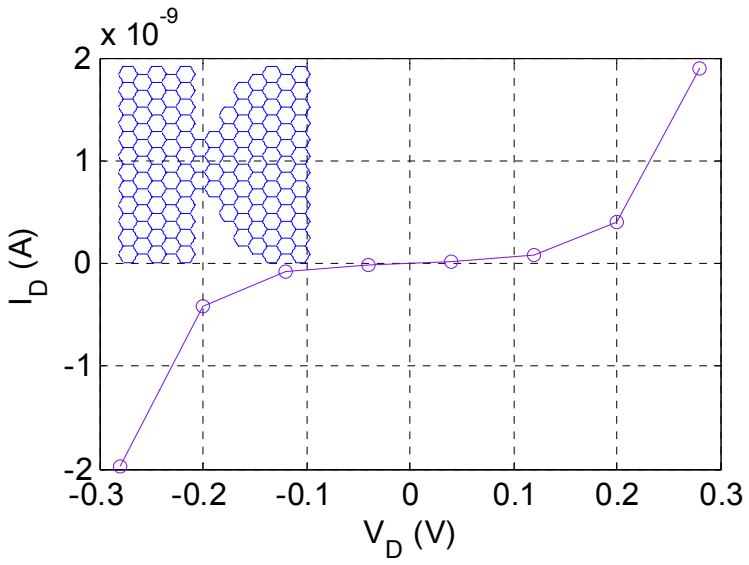


Figure VIII-15 Asymmetric $I(V)$ characteristics of the geometric diode obtained from simulation.

A possible technique for improving the asymmetry is by making the graphene strongly n- or p-type by changing the location of the Fermi level w.r.t. the bandstructure. This is achieved by doping the graphene or by applying a gate voltage. In the simulation, this effect is modeled by changing the Bloch energy (ε_{2p}) at each graphene atom by an amount E_{offset} as was introduced in section VII.B. This change is equivalent to changing the position of the Fermi level with respect to the bandstructure.

To demonstrate the effect of E_{offset}, in Figure VIII-16, I plot the current vs. energy curves for two bias voltages with different offset energies chosen to have dominant hole (E_{offset}, = 0.05 eV) or electron (E_{offset} = -0.05 eV) transport for the same geometric diode shape as before.

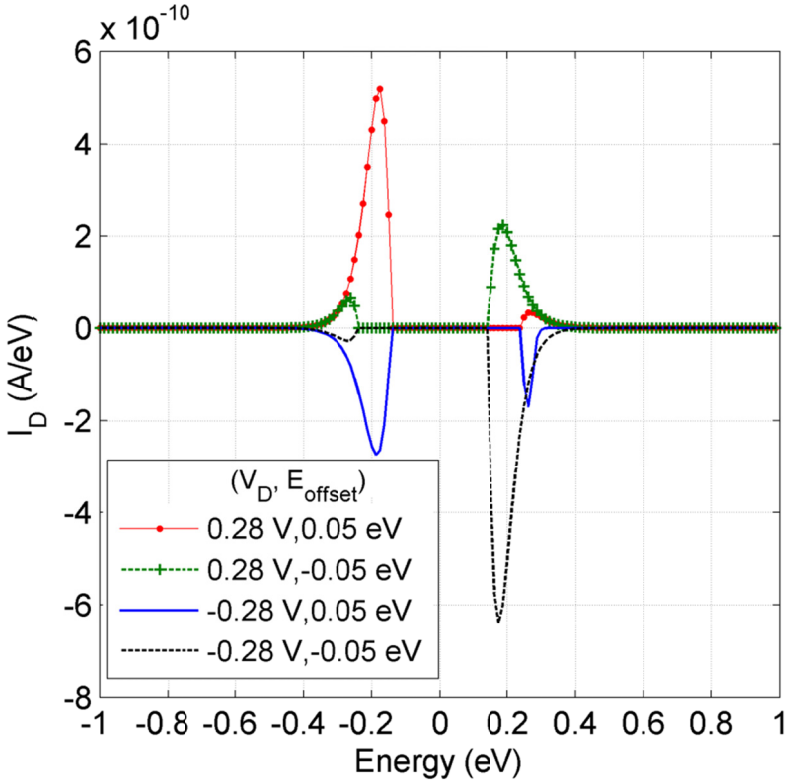


Figure VIII-16 Comparison of current vs. energy at positive- and negative-bias for majority electron (E_{offset} = -0.05 eV) and hole (E_{offset} = 0.05 eV) carriers.

When holes are the majority carriers, the current for a positive voltage is larger than at a negative voltage implying that a greater current flows along the direction of the funnel in the geometrically asymmetric conductor. The opposite current-asymmetry is observed when electrons are the majority carriers as a larger electron-flow occurs along the direction of the funnel for the opposite bias polarity.

Change in the preferred direction of current flow while transitioning from an n- to p-type graphene is shown in Figure VIII-17. For E_{offset} > 0.02 eV, the graphene is p-type and a higher current flows in the reverse bias. For E_{offset} < 0.02 eV, the graphene is n-type and a higher current flows at positive bias. The saturation in asymmetry for negative offset values suggests that an optimal E_{offset} may be found.

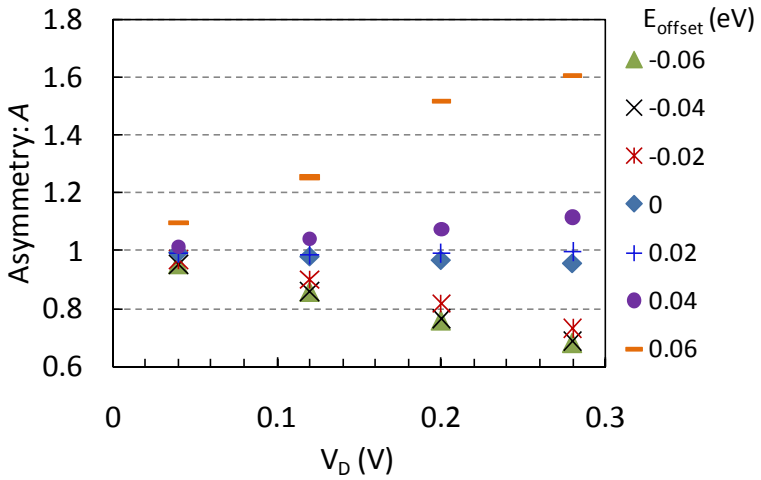


Figure VIII-17 Forward-to-reverse current asymmetry ($A=|I_D(+V_D)/I_D(-V_D)|$) vs. diode voltage with changing E_{offset}.

Transmission from contact 1 to 2 or vice-versa at a particular voltage is equal. The asymmetry in the $I(V)$ curve occurs due to unequal transmission at different voltages. The effect of the potential on transmission increases with the magnitude of the applied voltage. Therefore, the asymmetry in Figure VIII-17 is prominent at higher voltages.

As a check, the asymmetric shape used for the above simulations was flipped along the vertical and simulated. The $I(V)$ curve obtained was the same as the original one mirrored at the origin, indicating no geometry related simulation error.

To further ensure that the asymmetry in $I(V)$ occurs due to the geometry, I simulated two symmetric structures. The first one is a sheet of armchair graphene with 5 nm edge lengths. The $I(V)$ curve for this conductor is shown in Figure VIII-18. The magnitude of current flowing under forward and reverse bias is equal.

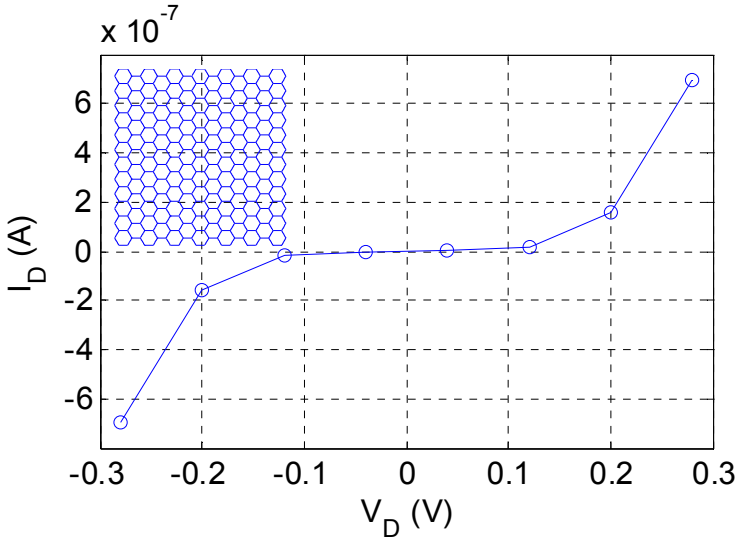


Figure VIII-18 Symmetric *I*(*V*) curve for a symmetric sheet of graphene.

The second shape is a symmetric hourglass shaped conductor that emulates the confinement that occurs in a geometric diode as shown in Figure VIII-19. Again, no asymmetry is observed in the *I*(*V*) curve for this conductor.

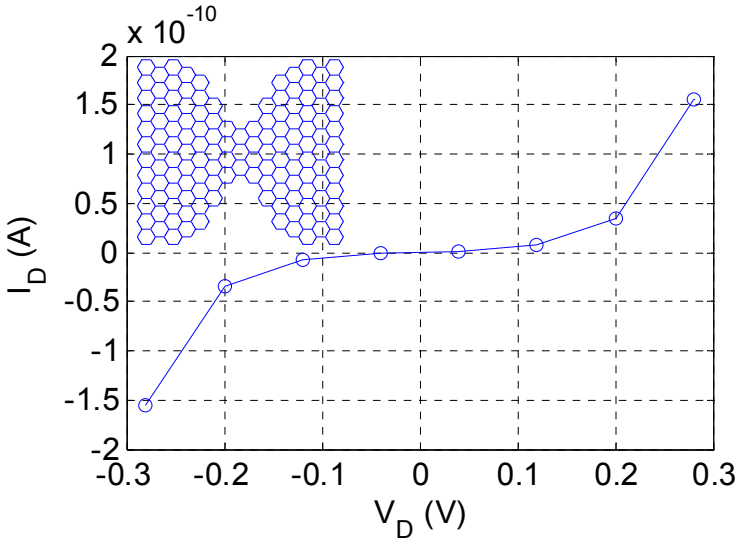


Figure VIII-19 Symmetric *I*(*V*) curve for a symmetric conductor with confinement.

G. Comparison of simulated and experimental characteristics

The simulations shown above are on a length scale two orders of magnitude smaller than the experimental devices. The smaller dimensions lead to conduction sub-bands that are farther apart in energy than in an actual device. Therefore, transport properties, e.g., transmission, in a smaller device may differ greatly from a larger device. However, as seen in Figure VIII-20, the effect of geometry on the diode asymmetry is consistent for the quantum simulation, Drude model (Zhu, 2011), and an experimental diode. To plot them together, the curves are scaled by a factor given in the legend. The smaller asymmetry in the experimental device is partly due to series resistance. Greater asymmetry is observed in measured $I(V)$ curves at a larger voltage scale.

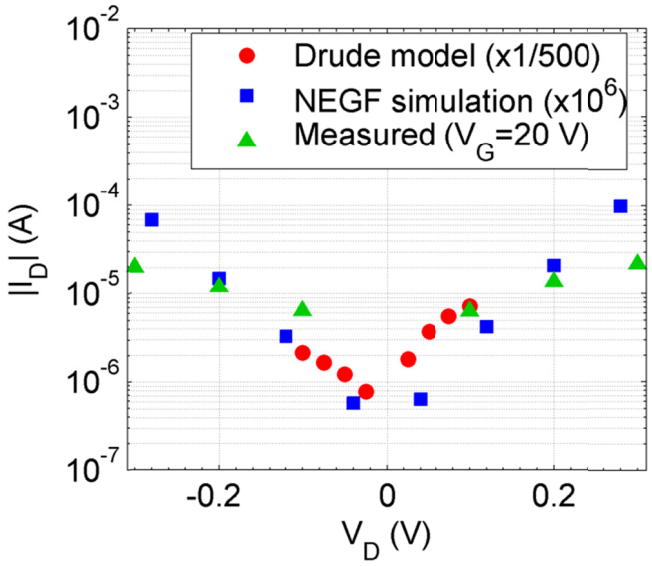


Figure VIII-20 Comparison of $I(V)$ curves obtained from Drude model (Zhu, 2011), self-consistent NEGF + Poisson solver, and measurement.

The NEGF simulation in Figure VIII-20 corresponds to a diode with a neck size of 1 nm and a shoulder width of 7 nm with an angle of 60°. The Drude simulation is for a device with a neck size of 50 nm, shoulder size of 2000 nm, angle of 45°, and

mean-free path length of 200 nm. The measured $I(V)$ curve is for the device reported in Figure VII-7.

CHAPTER IX

SEMICLASSICAL THEORY OF OPTICAL FREQUENCY RECTIFICATION IN MESOSCOPIC DIODES

From the semiclassical analysis for the optical response of an MIM diode, presented in Chapter V, the characteristics under illumination are obtained from the DC $I(V)$ curve of the diode. A theory similar to the Tien and Gordon approach (Tien, 1963), but applicable even to a non-tunneling based device, is required for the optical response of geometric diodes.

Platero and Aguado (Platero, 2004) have reviewed several techniques that can be used to study photon-assisted transport in semiconductor nanostructures. Other than Tien and Gordon's approach, all the analyses culminate in a form requiring numerical computation of the transport mechanism, and do not provide insights into the optical behavior through a simple extension of the DC characteristics.

The simplicity of the Tien and Gordon formulation comes with the drawback that the theory is not gauge-invariant. Another limitation of this method is that it does not account for charge and current conservation. To ensure conservation, an AC transport theory such as the one that solves the non-equilibrium Green's function (NEGF) and Poisson equations self-consistently (Kienle, 2010) is required. In a geometric diode, the dependence of the transport properties on the geometry adds to the complexity.

Here I derive a formula analogous to the Tien and Gordon approach but also applicable to an illuminated mesoscopic junction. It is based on the NEGF, which was introduced in Chapter VII. Several approximations are made to enable an

analytical relation between the DC and illuminated characteristics. Even though these approximations limit the applicability of the result, it is nevertheless helpful in understanding the illuminated characteristics of mesoscopic junctions.

A. Mesoscopic junction under illumination

The starting point for the derivation is the NEGF theory for an illuminated junction given by Datta and Anantram (Datta, 1992). In this section I reproduce and explain some of their results.

Consider a mesoscopic junction connected to two contacts as shown in Figure IX-1. It is assumed that charge transport from one contact to the other occurs phase-coherently, and the coherence is broken only by scattering in the contacts. Here, the contacts refer to charge reservoirs, much larger than the device region, held at a fixed potential. Electrons gain or lose energy through scattering in the contacts. The interaction of charge carriers with photons occurs in the device region through a time-varying potential $V(\boldsymbol{r},t)$. This interaction, even though inelastic (changes the energy of charge particles), does not cause phase incoherence.

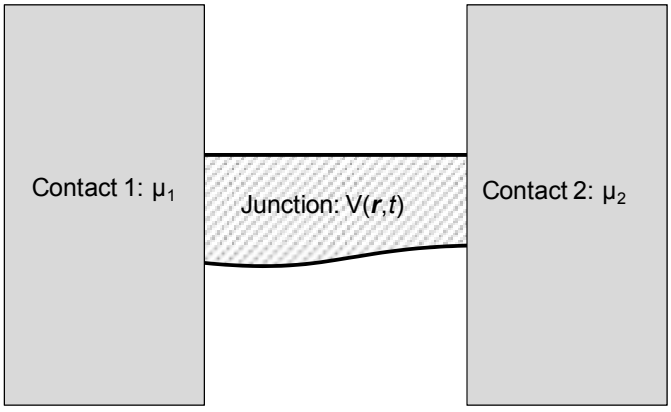


Figure IX-1 A two terminal device with an arbitrary shape, subjected to an AC potential $V(\boldsymbol{r},t)$.

The energy-domain version of Schrödinger's equation for the device in the absence of illumination is (Datta, 1992)

$$\left[E+\frac{\hbar^2\nabla^2}{2m}-V_S(\vec{r})+\frac{i\hbar}{2\tau_\varphi(\vec{r},E)}\right]G_0^R(\vec{r},E;\vec{r}\,',E')=\delta(\vec{r}-\vec{r}\,')\delta(E-E')$$
$$\Rightarrow H_0(\vec{r},E)G_0^R(\vec{r},E;\vec{r}\,',E')=\delta(\vec{r}-\vec{r}\,')\delta(E-E')$$

Eq. IX-1

where G_0^R is the retarded Green's function that represents the impulse response of Schrödinger's equation. The subscript '0' refers to the Green's function for the un-illuminated case. The wavefunction at any energy E can be obtained from G_0^R. V_s is the static potential in the device. The τ_φ is the scattering (phase-breaking) time in the contacts.

The current in the device is obtained as

$$I=e\hbar\int dE\int dE'\int d\vec{r}\int d\vec{r}\,'\left[t_{21}(E,E')f_1(E')-t_{12}(E',E)f_2(E)\right]$$

Eq. IX-2

where f_1 and f_2 are the Fermi-Dirac distributions in contact 1 and 2 respectively. The $t_{21}(E,E')$ is the transmission from an input energy mode E' in contact 1 to an output energy mode E in contact 2, and is given by

$$t_{21}(E,E')=\frac{\hbar}{\tau_{avg}}\iint\limits_{\substack{\vec{r}\in contact1\\ \vec{r}\,'\in contact2}}d\vec{r}d\vec{r}\,'\frac{\left|G^R(\vec{r},E;\vec{r}\,',E')\right|^2}{4\pi^2\tau_\varphi(\vec{r},E)\tau_\varphi(\vec{r}\,',E')}$$

Eq. IX-3

The τ_{avg} is the time over which the current or the transmission is averaged to find the DC component. The above equations are analogous to Eq. VIII-12 with $1/\tau_\varphi$ representing the broadening (Γ) due to the contacts and $|G^R|^2$ representing GG^+. The un-illuminated case is obtained by replacing G^R by G_0^R in Eq. IX-3.

Eq. IX-2 is different from the usual form for the transport equation that takes into account the exclusion principle by counting the filled states on one contact and the empty states on the other. The applicability of this equation for phase-coherent transport is explained by Landauer (Landauer, 1989) and the equation is explicitly derived by Datta (Datta, 1992).

Under illumination, the Schrödinger equation is modified as (Datta, 1992), (Ferry, 2009, p.624)

$$H_0(\vec{r},E)G^R(\vec{r},E;\vec{r}',E') = \delta(\vec{r}-\vec{r}')\delta(E-E') + \sum_{\omega} V(\vec{r},\hbar\omega)G^R(\vec{r},E-\hbar\omega;\vec{r}',E')$$

Eq. IX-4

where V(***r***,*t*) is represented by its Fourier transform components Σ*V*(***r***, *ħω*). This perturbation term added to the RHS is the strength of the broadening due to interaction with the field (Datta, 2005, p.259). This effect is included via a self-energy similar to the one used for the contacts (Kienle, 2010). Under a first order Born approximation, the solution to the equation of motion given by the modified Schrödinger equation in Eq. IX-4, is

$$\begin{aligned} G^R(\vec{r},E;\vec{r}',E') &= G_0^R(\vec{r},\vec{r}';E)\delta(E-E') \\ &+ \sum_{\omega}\int d\vec{r}''\, G_0^R(\vec{r},\vec{r}'';E)V(\vec{r}'',\hbar\omega)G_0^R(\vec{r}'',\vec{r}';E')\delta(E-\hbar\omega-E') \end{aligned}$$

Eq. IX-5

In the presence of illumination, the current is obtained by substituting G^R in Eq. IX-3 by the above expression. The G^R can be computed numerically using the technique described in Chapter VII. However in the next section, I simplify the above equation such that the illuminated characteristics can be predicted by an analytical extension of the DC *I*(*V*).

B. Projecting illuminated characteristics from DC *I*(*V*)

I propose two simplifications to the expression for G^R given in Eq. IX-5. The first is a uniform strength of interaction with the field over the device area ($V(\boldsymbol{r}, \hbar\omega) = V(\hbar\omega)$). This is achieved by coupling an AC scalar potential through a gate electrode (Pedersen, 1998) or by applying the dipole approximation for a vector potential gauge (Eliasson, 2001, p.127). The dipole approximation requires that the wavelength of the EM field be much larger than the size of the device. This condition is easily satisfied for an MIM diode, and for small geometric diodes. A further complication in geometric diodes is the field non-uniformity due to the shape of the conductor. For this, a field strength averaged over the geometry would serve as an initial correction.

The second approximation relates to the transport properties of electrons at energies separated by $\hbar\omega$. Here, I claim that the transport properties defined by G^R do not differ significantly for two energy levels spaced apart by $\hbar\omega$. The G^R for the two energies are similar if the photon energy is small compared to the energy of electrons. As the majority of conduction occurs due to electrons at the Fermi surface, the relevant energy for comparison is the Fermi energy ($\hbar\omega << E_f$) measured with respect to the band edge. This assumption is similar to the nearly elastic scattering case considered by Datta (Datta, 2005, p.274). The ramifications of this approximation are discussed at the end of the chapter.

Under these assumptions, Eq. IX-5 is simplified to

$$G^R(\vec{r},E;\vec{r}\,',E') = G_0^R(\vec{r},\vec{r}\,';E)\,\delta(E-E') + \sum_{\omega} V(\hbar\omega)\int d\vec{r}\,''\,G_0^R(\vec{r},\vec{r}\,';E)\,\delta(E-\hbar\omega-E')$$

Eq. IX-6

where the Green's function in the integral is simplified under the assumption $E \approx E'$ as

$$G_0^R(\vec{r},\vec{r}';E)=G_0^R(\vec{r},\vec{r}'';E)G_0^R(\vec{r}'',\vec{r}';E')$$

Eq. IX-7

The spatial integral on the RHS of Eq. IX-6 leads to the volume of the devices region (*vol.*) as the ***r***" dependence in the integrand has been removed. The Green's function under illumination then becomes

$$G^R(\vec{r},E;\vec{r}',E')=G_0^R(\vec{r},\vec{r}';E)\delta(E-E') + vol.\sum_{\omega}V(\hbar\omega)G_0^R(\vec{r},\vec{r}';E)\delta(E-\hbar\omega-E')$$

Eq. IX-8

Substituting the above expression in Eq. IX-3

$$\begin{aligned}
&t_{21}(E,E')\propto \\
&\left|G_0^R(\vec{r},\vec{r}';E)\delta(E-E')+vol.\sum_{\omega}V(\hbar\omega)G_0^R(\vec{r},\vec{r}';E)\delta(E-\hbar\omega-E')\right|^2 \\
&=\left|G_0^R(\vec{r},\vec{r}';E)\right|^2\delta^2(E-E') \\
&+\left[G_0^R(\vec{r},\vec{r}';E)\delta(E-E')\left(vol.\sum_{\omega}V(\hbar\omega)G_0^R(\vec{r},\vec{r}';E)\delta(E-\hbar\omega-E')\right)^{cc}+cc\right] \\
&+\left|vol.\sum_{\omega}V(\hbar\omega)G_0^R(\vec{r},\vec{r}';E)\delta(E-\hbar\omega-E')\right|^2
\end{aligned}$$

Eq. IX-9

where *cc* denotes complex conjugate. The term inside the square bracket has a product of two delta-functions $\delta(E\text{-}E")*\delta(E\text{-}\hbar\omega\text{-}E")$, which is always zero. Also, in the third term, the square of the summation over delta functions is equal to the summation of the squares as all cross terms with different ω are always zero due to the product of delta-functions $\delta(E\text{-}\hbar\omega_1\text{-}E")*\ \delta(E\text{-}\hbar\omega_2\text{-}E")$. Therefore Eq. IX-9 reduces to

$$\begin{aligned}
&t_{21}(E,E')\propto \\
&\left|G_0^R(\vec{r},\vec{r}';E)\right|^2\delta^2(E-E')+vol.\sum_{\omega}\left|V(\hbar\omega)\right|^2\left|G_0^R(\vec{r},\vec{r}';E)\right|^2\delta(E-\hbar\omega-E')
\end{aligned}$$

Eq. IX-10

The second term in the above equation represents a first-order, low-photon-energy correction to the Green's function for the un-illuminated case. This leads to the expression for the illuminated current given by

$$I_{illum} = \frac{e\hbar^2}{T}\int dE \int dE' \int_{\vec{r}_2} d\vec{r} \int_{\vec{r}_1} d\vec{r}\,'$$
$$\times \left[\begin{array}{l} \dfrac{\left|G_0^R(\vec{r},\vec{r}\,';E)\right|^2 \delta(E-E')}{4\pi^2 \tau_\varphi(\vec{r},E)\tau_\varphi(\vec{r}\,',E')} f_1(E') - \dfrac{\left|G_0^R(\vec{r}\,',\vec{r};E')\right|^2 \delta(E'-E)}{4\pi^2 \tau_\varphi(\vec{r},E)\tau_\varphi(\vec{r}\,',E')} f_2(E) + \\ Vol^2 \sum_\omega \left|V(\hbar\omega)\right|^2 \left\{ \begin{array}{l} \dfrac{\left|G_0^R(\vec{r},\vec{r}\,';E)\right|^2 \delta(E-\hbar\omega-E')}{4\pi^2 \tau_\varphi(\vec{r},E)\tau_\varphi(\vec{r}\,',E')} f_1(E') - \\ \dfrac{\left|G_0^R(\vec{r}\,',\vec{r};E')\right|^2 \delta(E'-\hbar\omega-E)}{4\pi^2 \tau_\varphi(\vec{r},E)\tau_\varphi(\vec{r}\,',E')} f_2(E) \end{array} \right. \end{array} \right]$$

Eq. IX-11

I now perform the integral with respect to E' in Eq. IX-11 using the sifting property of the delta function. Under the condition $\hbar\omega << E_f$, I also approximate that τ_φ does not vary significantly from E to $E+\hbar\omega$. The validity of this is in-line with the approximation made for G^R in that the interaction of an electron with the reservoirs is similar at two closely spaced energies. With this simplification, the illuminated current is given by

$$I_{illum} = \frac{e\hbar^2}{T}\int dE \int_{\vec{r}_2} d\vec{r} \int_{\vec{r}_1} d\vec{r}\,'$$
$$\times \left[\begin{array}{l} \dfrac{\left|G_0^R(\vec{r},\vec{r}\,';E)\right|^2}{4\pi^2\tau_\varphi(\vec{r},E)\tau_\varphi(\vec{r}\,',E)} f_1(E) - \dfrac{\left|G_0^R(\vec{r}\,',\vec{r};E)\right|^2}{4\pi^2\tau_\varphi(\vec{r},E)\tau_\varphi(\vec{r}\,',E)} f_2(E) + \\ Vol^2 \sum_\omega \left|V(\hbar\omega)\right|^2 \left\{ \begin{array}{l} \dfrac{\left|G_0^R(\vec{r},\vec{r}\,';E)\right|^2}{4\pi^2\tau_\varphi(\vec{r},E)\tau_\varphi(\vec{r}\,',E)} f_1(E-\hbar\omega) - \\ \dfrac{\left|G_0^R(\vec{r}\,',\vec{r};E)\right|^2}{4\pi^2\tau_\varphi(\vec{r},E)\tau_\varphi(\vec{r}\,',E)} f_2(E) \end{array} \right\} \end{array} \right]$$

Eq. IX-12

which can be simplified to

$$I_{illum} = \frac{e\hbar^2}{T}\int dE \int_{\vec{r}_2} d\vec{r} \int_{\vec{r}_1} d\vec{r}\,'$$
$$\times \left[\begin{array}{l} \dfrac{\left|G_0^R(\vec{r},\vec{r}\,';E)\right|^2}{4\pi^2\tau_\varphi(\vec{r},E)\tau_\varphi(\vec{r}\,',E)} \left\{ f_1(E) - f_2(E) \right\} + \\ Vol^2 \sum_\omega \left|V(\hbar\omega)\right|^2 \dfrac{\left|G_0^R(\vec{r},\vec{r}\,';E)\right|^2}{4\pi^2\tau_\varphi(\vec{r},E)\tau_\varphi(\vec{r}\,',E)} \left\{ f_1(E-\hbar\omega) - f_2(E) \right\} \end{array} \right]$$

Eq. IX-13

The second term in the above equation that represents the additional current due to illumination can be written in terms of the un-illuminated (dark) current given by the first term. This is done by combining the $\hbar\omega$ with the DC voltage applied between the two contacts (V_D). Assuming that contact 2 is the ground, Eq. IX-13 can be written in terms of the Fermi distribution $f(E)=[1+\exp((E-E_f)/kT)]^{-1}$ as

$$I_{illum} = \frac{e\hbar^2}{T}\int dE \int_{\vec{r}_2} d\vec{r} \int_{\vec{r}_1} d\vec{r}\,' \times \left[\begin{array}{l} \dfrac{\left|G_0^R(\vec{r},\vec{r}\,';E)\right|^2}{4\pi^2\tau_\varphi(\vec{r},E)\tau_\varphi(\vec{r}\,',E)}\left\{f(E-eV_D)-f(E)\right\}+ \\ Vol^2\sum_\omega \left|V(\hbar\omega)\right|^2 \dfrac{\left|G_0^R(\vec{r},\vec{r}\,';E)\right|^2}{4\pi^2\tau_\varphi(\vec{r},E)\tau_\varphi(\vec{r}\,',E)}\left\{f(E-(eV_D+\hbar\omega))-f(E)\right\} \end{array} \right]$$

Eq. IX-14

This can be interpreted as

$$I_{illum}(V_D) = I_{dark}(V_D) + Vol^2\sum_\omega \left|V(\hbar\omega)\right|^2 I_{dark}\left(V_D + \frac{\hbar\omega}{e}\right)$$

Eq. IX-15

C. Discussion

Eq. IX-15 allows an analytical evaluation of the illuminated characteristics from a dark $I(V)$ curve, similar to the first order version of Eq. V-6 depicted graphically using a piece-wise linear dark $I(V)$ curve in Figure V-9. Although the equation for an MIM diode was derived for a single-frequency illumination, its multi-spectral extension (neglecting the mixing terms) also consists of an integral on the contribution from different frequencies (Tucker, 1985).

In analogy with the high-frequency operating mechanism described for an MIM diode in Figure V-2, Eq. IX-15 indicates that the interaction of the electrons in the device region is equivalent to the modulation of electron energies in the contact. This simplified picture for the interaction emerges because the electronic transport properties are assumed to be constant in a narrow range of energies given by the additional photon energy ($\hbar\omega$) acquired by the electrons in the device region. The derivation presented in the previous section essentially transferred the electron interaction with the photon from the device region to the contacts.

The assumption of $\hbar\omega << E_f$ is the most significant consideration of the derivation. Earlier, I stated this condition without analyzing its physical significance. Transport properties, e.g. tunneling probability in an MIM diode, can be a strong function of energy as seen in Figure II-2(a). However, in the case of MIM diodes the photon-electron interaction does occur only in the contacts due to the absence of electrons in the insulator. This is the basis for the semiclassical theory described in Chapter V. In a mesoscopic junction like the geometric diode, the transmission is a weaker function of energy than in MIM diodes. Therefore, in narrow energy ranges, the assumption of constant transport behavior has greater validity. Ultimately, the effect of the added photon energy is relative to the existing electron-energy. If this energy is large, such that the transmission is highly likely, the change in transmission with additional photon energy will be small. A measure of interest for the existing electron energy is the Fermi energy and hence the condition $\hbar\omega << E_f$.

A final point of discussion concerns the material for the thin-film used in the device region. Eq. IX-15 is applicable under the assumption that the photon energy (E_{ph}) is small compared to the Fermi energy (E_f) . This depends on the value of the E_f. For metals, E_f is on the order of a few electron-volts so that the result holds even at far-to-mid-infrared. However, for graphene, the E_f is closer to zero, but can be varied by applying a gate-voltage or by doping.

CHAPTER X

CONCLUSIONS & FUTURE WORK

I have analyzed the roadblocks in achieving practically useful efficiencies in optical rectennas with MIM and geometric diodes as potential rectifiers. Efficient optical rectennas require diodes that are inherently fast, highly nonlinear and/or asymmetric and have the product of resistance and capacitance (RC) smaller than the inverse of frequency to be rectified.

Below, I have grouped my conclusions into four sections based upon metal-insulator diodes, circuit model and design of rectennas, semiclassical analysis of MIM diodes, and geometric diodes. This is followed by suggestions for future work.

A. Metal-insulator diodes

The transport in MIM diodes is femtosecond-fast but the single-insulator low-resistance diodes are not sufficiently nonlinear. Well engineered multi-insulator diodes can have improved current-voltage $I(V)$ characteristics satisfying both these requirements. Two mechanisms for improved MIIM diodes, resonant tunneling and abrupt change in tunneling through a step barrier, are examined through simulation. Either of these mechanisms can be made to dominate by the appropriate choice of insulators and variation of barrier thicknesses.

Single-insulator diodes were fabricated using sputter deposition and their characteristics were consistent with the simulated $I(V)$ curves. Successful demonstration of simulated characteristics of metal-insulator diodes is dependent on the thermodynamic stability of the material interfaces. An initial attempt to

quantify the stability was carried out using a demonstrational web-software and also via book values for thermodynamic quantities used in a Gibb's free energy analysis.

B. Rectenna circuit and design

I have analyzed the requirements imposed by the operating frequency on the circuit parameters of the rectenna. Independent of the antenna impedance, diodes with low resistance and capacitance are required for the RC time constant of the rectenna to be smaller than the reciprocal of the operating frequency. This ensures adequate bandwidth and efficient coupling between the antenna and the diode.

Existing MIM diodes fail to meet these requirements for visible-to-near-infrared operation. Infrared detectors with reduced efficiency due to the high RC of MIM diodes may still be practical. However, the impact of RC on the power conversion efficiency of solar cells is less forgiving.

I have considered two approaches for improving the efficiency of rectennas. The more radical approach is to use the planar geometric diode, which unlike MIM diodes does not have a parallel-plate structure and thereby has a smaller RC time constant. The less radical approach uses MIM diodes in a traveling-wave structure to obtain a distributed rectifier that removes the RC limitation on bandwidth.

At infrared wavelengths, the traveling-wave detector was shown to have improved responsivity and bandwidth. I derived a formula for calculating the responsivity of the traveling-wave detector, which improves with increased diode responsivity and reduced resistance per unit area. A multi-insulator diode can be designed to meet these requirements. The responsivity of the traveling-wave detector is comparable to the lumped-element detector at 10 μm, but is much higher at shorter wavelengths with scope for further improvement. Both the impedance

and the responsivity of the traveling-wave detector allow the detection bandwidth to extend over several microns in the MWIR and the LWIR ranges. Its efficiency is limited by the resistive losses in the waveguide-metals.

C. Semiclassical analysis of MIM diodes

I have modeled the $I(V)$ characteristics of MIM diodes illuminated at optical frequencies using a semiclassical approach that accounts for the photon energy of the radiation. Instead of classical small-signal rectification, in which a continuous span of the DC $I(V)$ curve is sampled during rectification, at optical frequencies the radiation samples the DC $I(V)$ curve at discrete voltage steps separated by the photon energy (divided by the electronic charge). As a result, the diode resistance and responsivity differ from their classical values. At optical frequencies, a diode even with a moderate forward-to-reverse current asymmetry exhibits high quantum efficiency.

The power efficiency in an optical rectifier with a non-ideal diode improves with a stronger AC signal. For solar radiation, the maximum illumination area that can be tapped by one rectenna element corresponds to a coherence area of 283 μm^2. This limits the strength of the AC voltage at the diode. The AC voltage can be increased by making the diode smaller, effectively, increasing the field strength for the same input power. With the AC signal signal strength larger than the photon energy divided by electron charge, the first-order assumption breaks down and higher order terms need to be considered.

The maximum rectification efficiency allowed by the small-signal classical and the first-order semiclassical models is 50%. On the other hand, while the small-signal circuit model gives 50% efficient rectennas, the large-signal classical circuit model allows efficiencies close to 100%.

D. Geometric diodes

Asymmetrically patterned thin films were investigated with the objective of obtaining low capacitance and low resistance diodes for optical rectennas. As theorized, asymmetric constrictions on the length-scale of the mean free path length of charge carriers were found to have asymmetric $I(V)$ characteristics. Geometric diodes fabricated using graphene yielded characteristics that are consistent with Drude-model and quantum simulations. Finer lithography is expected to further improve the device responsivity and provide reduced capacitance.

Quantum simulations based on the non-equilibrium Green's function and the Poisson equation were carried out. These simulations use a tight-binding Hamiltonian and solve for the charge and potential self-consistently. Although these simulations are performed device geometries that are much smaller than fabricated structures, the trends in simulated characteristics are consistent with the experimental curves.

A generalized semiclassical analysis was carried out for predicting the optical response of geometric diodes. Several assumptions were made to simplify the intermediate equations to a meaningful expression. For photon energies low compared to the Fermi energy measured from the band edge, the response turns out to be similar to that for MIM diodes.

E. Future work

The double insulator tunnel diodes that I have simulated are only demonstrative and do not represent optimized diodes. Exploring the optimization space consisting of the barrier heights, insulator thicknesses, and dielectric constants may lead to diodes with higher nonlinearity and lower resistance. An unexplored possibility for achieving a lower RC is to modify the multi-insulator

tunnel diode with a low-barrier low-dielectric-constant insulator. Doing so may decrease the capacitance without affecting the resistance significantly. From a survey of different dielectrics, there appears to be an inverse relationship between the barrier height and dielectric constant. Therefore, implementing this approach requires a deeper study to find the appropriate combination of metals and insulators.

The tunnel diode simulations were carried out using a transfer-matrix method to calculate the probability of electron tunneling. A possible improvement in simulating the MIIM diode is to consider the occupation of the resonant well. A better estimate for the effective mass, measured or extracted from curve-fitting, is also desirable.

The assumption of single frequency illumination was implicit in the small-signal circuit analysis as well as the semiclassical analysis for the MIM diode. In applications such as infrared detectors and solar cells, the radiation is expected to be broadband. For the rectenna, not only does this imply that the antenna should be broadband but also requires the diode to rectify each frequency optimally. If and how this is possible requires consideration both from the perspective of circuit analysis and photon-assisted transport.

The correspondence between the large-signal circuit and semiclassical models needs to be analyzed in order to understand whether optical rectennas can operate at efficiencies higher than 50%. This may require including the higher-order terms in the semiclassical analysis.

For the geometric diode, so far, only sub-10 nm geometries have been simulated quantum mechanically. This is largely due to the computationally intensive tight-binding model that is used to define the device Hamiltonian. A simpler approach, e.g., an effective mass model, may allow larger geometries to be simulated quantum mechanically at reduced computational cost. Further, only charge and potential

have been accounted self-consistently in the simulations. Scattering phenomenon including interaction with photons and phonons can be added onto the basic framework of the simulator.

In the theory developed for the semiclassical analysis of geometric diodes, a significant approximation was the unchanged transport properties of an electron before and after the absorption of an energy quantum equal to the photon energy. I have provided qualitative implications of this assumption. It would be worthwhile to quantify the error incurred and also to analyze the applicability of the derived expression outside the constraints of the assumption.

REFERENCES

Abdel-Rahman, M.R., Gonzalez, F.J., Zummo, G., Middleton, C.F. & Boreman, G.D., 2004. Antenna-coupled MOM diodes for dual-band detection in MMW and LWIR., *Proc. SPIE*, 5410, 238.

ADOSE, 2008. *Reliable Application Specific Detector of Road Users with Vehicle On-board Sensors*. European Commision.

Agarwal, G.S., Gbur, G. & Wolf, E., 2004. Coherence properties of sunlight. *Optics Letters*, 29(5), pp.459-61.

Ashcroft, N.W. & Mermin, N.D., 1976. *Solid state physics*. Orlando: Harcourt College Publishers.

Bailey, R.L., 1972. A proposed new concept for a solar-energy converter. *Journal of Engineering for Power*, (73).

Bain, R.J.P. & Donaldson, G.B., 1985. Sputtered all-niobium Josephson tunnel junctions with barrier oxide protected by gold. *J. of Phys. C: Solid State Phys.*, 18(12), pp.2539-48.

Bale, C.W., 2011. *FACT-EQUILIB-Web*. [Online] Available at: http://www.crct.polymtl.ca/equiwebmenu.php.

Berger, C., Song, Z., Li, X., Wu, X., Brown, N. et al., 2006. Electronic Confinement and Coherence in Patterned Epitaxial Graphene. *Science*, 312(5777), pp.1191-96.

Berland, B., 2003. *Photovoltaic Technologies Beyond the Horizon: Optical Rectenna Solar Cell*. Final report. NREL Report No. SR-520-33263.

Blanter, Y.M. & Martin, I., 2007. Transport through normal-metal--graphene contacts. *Phys. Rev. B*, 76(15), p.155433.

Born, M. & Wolf, E., 1999. *Principles of Optics*. 7th ed. Cambridge University Press.

Brown, W.C., 1976. Optimization of the Efficiency and Other Properties of the Rectenna Element., 1976.

Brown, W.C., 1984. The History of Power Transmission by Radio Waves. *Microwave Theory and Techniques, IEEE Transactions on*, 32(9), pp.1230-42.

Brown, E.R., 2004. A system-level analysis of Schottky diodes for incoherent THz imaging arrays. *Solid-State Electronics*, 48, pp.2051-53.

Büttner, H. & Gerlach, E., 1968. Microwave assisted tunneling between superconductors. *Phys. Lett.*, 27A(4), pp.226-27.

Camp, M. & Lecchini, S.M.A., 1965. The work function of polycrystalline tungsten foil. *Proc. of the Physical Society*, 85(4), p.815.

Chapline, M.G. & Wang, S.X., 2007. Analytical formula for the tunneling current versus voltage for multilayer barrier structures. *Journal of Applied Physics*, 101(8), p.083706.

COMSOL, 2004. FEMLAB Electromagnetics Module User's Guide, FEMLAB 3.1.

Corkish, R., Green, M.A. & Puzzer, T., 2002. Solar energy collection by antennas. *Solar Energy*, 73(6), pp.395-401.

Dagenais, M., Choi, K., Yesilkoy, F., Chryssis, A.N. & Peckerar, M.C., 2010. Solar Spectrum Rectification Using Nano-Antennas and Tunneling Diodes. *Proc. of SPIE*, 7605, pp.76050E-1.

Datta (a), S., 2009. ECE 659 Lecture 19: Self Energy. http://nanohub.org/resources/6430.

Datta, S., 1992. Steady-state transport in mesoscopic systems illuminated by alternating fields. *Phys. Rev. B*, 45(23), pp.13761-64.

Datta, S., 2000. Nanoscale device modeling: the Green's function method. *Superlattices and Microstructures*, 28(4), pp.253-78.

Datta, S., 2002. The Non-Equilibrium Green's Function (NEGF) Formalism: An Elementary Introduction. In *IEDM Digest.*, 2002.

Datta, S., 2005. *Quantum Transport: Atom to Transistor*. Cambridge.

Datta, S., 2008. Lecture 2A: Quantum Transport. http://nanohub.org/resources/5263.

Datta, S., 2009. Lecture 5: NEGF Simulation of Graphene Nanodevices. *http://nanohub.org/resources/7422.*

Datta, S. & Anantram, M.P., 1992. Steady-state transport in mesoscopic systems illuminated by alternating fields. *Phys. Rev. B*, 45(23), pp.13761-64.

Davis, T.J., 2009. Surface plasmon modes in multi-layer thin-films. *Optics Communications*, 282(1), pp.135-40.

Dayem, A.H. & Martin, R.J., 1962. Quantum Interaction of Microwave Radiation with Tunneling Between Superconductors. *Phys. Rev. Lett.*, 8(6), pp.246-48.

de Vos, A., 1992. *Endoreversible thermodynamics of solar energy conversion*. Oxford: Oxford Science.

Dicke, R.H., Beringer, R., Kyhl, R.L. & Vane, A.B., 1946. Atmospheric Absorption Measurements with a Microwave Radiometer. *Phys. Rev.* , 70, pp.340-48.

Donges, A., 1998. The coherence length of black-body radiation. *Eur. J. Phys.* , 19, pp.245-49.

Durkan, C., 2007. *Current at the nanoscale: an introduction to nanoelectronics*. 1st ed. London: Imperial College Press.

Eliasson, B.J., 2001. *Metal-Insulator-Metal Diodes For Solar Energy Conversion*. PhD Thesis. Boulder: University of Colorado at Boulder.

Eliasson, B. & Moddel, G., 2005. Detectors at the Intersection of Photons and Electromagnetic Fields or, Where Einstein Meets Maxwell. In *Spring Meeting of the Materials Research Society*. San Francisco, March 28-April 1, 2005.

Estes, M.J. & Moddel, G., 2005. Terahertz interconnect system and applications. *US Patent 6,967,347*.

Estes, M.J. & Moddel, G., 2006. Surface plasmon devices. *U S Patent No. 7,010,183*.

Evenson, K.M., Wells, J.S., Matarrese, L.M. & Elwell, L.B., 1970. Absolute Frequency Measurements of the 28- and 78- μm cw Water Vapor LASER Lines. *Applied Physics Letters*, 16(4), pp.159-62.

Eyert, V., 1996. A Comparative Study of Methods for Convergence Acceleration of Iterative Vector Sequences. *J. of Computational Physics*, 124, pp.271-85.

Fal'ko, V., 1989. Nonlinear Properties of Mesoscopic Junctions under High-Frequency Field Irradiation. *Europhysics Letters*, 8(8), pp.785-89.

Ferry, D.K., Goodnick, S.M. & Bird, J., 2009. *Transport in Nanostructures*. 2nd ed. Cambridge: Cambridge University Press.

Fox, M., 2001. *Optical Properties of Solids*. 1st ed. Oxford University Press.

Fumeaux, C., Herrmann, W., Kneubühl, F.K. & Rothuizen, H., 1998. Nanometer thin-film Ni-NiO-Ni diodes for detection and mixing of 30 THz radiation. *Infrared Physics & Technology*, 39(3), pp.123-83.

Golizadeh-Mojarad, R. & Datta, S., 2007. Nonequilibrium Green's function based models for dephasing in quantum transport. *Phys. Rev. B*, 75, p.081301.

González, F.J. & Boreman, G.D., 2005. Comparison of dipole, bowtie, spiral and log-periodic IR antennas. *Infrared Physics & Technology*, 46(5), pp.418-28.

Green, M.A., 2001. Third generation photovoltaics: Ultra-high conversion efficiency at low cost. *Progress in Photovoltaics: Research and Applications*, 9(2), pp.123-35.

Grover, S., Dmitriyeva, O., Estes, M.J. & Moddel, G., 2010. Traveling-Wave Metal/Insulator/Metal Diodes for Improved Infrared Bandwidth and Efficiency of Antenna-Coupled Rectifiers. *Nanotechnology, IEEE Transactions on*, 9(6), pp.716-22.

Gunapala, S., Bandara, S., Hill, C., Ting, D., Liu, J. et al., 2008. Quantum Well and Quantum Dot Based Detector Arrays for Infrared Imaging. *Mater. Res. Soc. Symp. Proc.*, 1076.

Guo, Y., Gu, B.-L., Yu, J.-Z., Zeng, Z. & Kawazoe, Y., 1998. Resonant tunneling in step-barrier structures under an applied electric field. *Journal of Applied Physics*, 84(2), pp.918-24.

Hagerty, J.A., Helmbrecht, F.B., McCalpin, W.H., Zane, R. & Popovic, Z.B., 2004. Recycling ambient microwave energy with broad-band rectenna arrays. *Microwave Theory and Techniques, IEEE Transactions on*, 52(3), pp.1014-24.

Hartstein, A. & Weinberg, Z.A., 1978. On the nature of the image force in quantum mechanics with application to photon assisted tunnelling and photoemission. *Journal of Physics C: Solid State Physics*, 11(11), p.L469.

Haug, H.J.W. & Jauho, A., 2008. *Quantum Kinetics in Transport and Optics of Semiconductors*. 2nd ed. Berlin: Springer-Verlag.

Hegyi, B., Csurgay, A. & Porod, W., 2007. Investigation of the nonlinearity properties of the DC I-V characteristics of metal-insulator-metal (MIM) tunnel diodes with double-layer insulators. *Journal of Computational Electronics*, 6, pp.159-62. 10.1007/s10825-006-0083-9.

Hobbs, P.C., Laibowitz, R.B., Libsch, F.R., LaBianca, N.C. & Chiniwalla, P.P., 2007. Efficient waveguide-integrated tunnel junction detectors at 1.6 μm. *Opt. Express*, 15(25), pp.16376-89.

Huang, J.-S., Feichtner, T., Biagioni, P. & Hecht, B., 2009. Impedance Matching and Emission Properties of Nanoantennas in an Optical Nanocircuit. *Nano Letters*, 9(5), p.1897–1902.

Hübers, H.-W., Schwaab, G.W. & Röser, H.P., 1994. Video detection and mixing performance of GaAs Schottky-barrier diodes at 30 THz and comparison with metal-insulator-metal diodes. *Journal of Applied Physics*, 75(8), pp.4243-48.

Jablan, M., Buljan, H. & Soljačić, M., 2009. Plasmonics in graphene at infrared frequencies. *Phys. Rev. B.*, 80, p.245435.

Jayaweera, P.V., Matsik, S.G., Perera, A.G., Liu, H.C., Buchanan, M. et al., 2008. Uncooled infrared detectors for 3–5 mu m and beyond. *Applied Physics Letters*, 93(2), p.021105.

Jonsson, B. & Eng, S.T., 1990. Solving the Schrodinger Equatoin in Arbitrary Quantum-Well Potential Profiles Using the Transfer Matrix Method. *IEEE J. of Quantum Elec.*, 26(11), pp.2025-35.

Kale, B.M., 1985. Electron tunneling devices in optics. *Optical Engineering*, 24(2), pp.267-74.

Kang, Y., Kang, J. & Chang, K.J., 2008. Electronic structure of graphene and doping effect on SiO2. *Phys. Rev. B*, 78, p.115404.

Kazemi, H., Shinohara, K., Nagy, G., Ha, W., Lail, B. et al., 2007. First THz and IR characterization of nanometer-scaled antenna-coupled InGaAs/InP Schottky-diode detectors for room temperature infrared imaging. *Infrared Technology and Applications XXXIII*, 6542(1), p.65421.

Kienle, D., Vaidyanathan, M. & Léonard, F., 2010. Self-consistent ac quantum transport using nonequilibrium Green functions. *Phys. Rev. B*, 81, p.115455.

Kittel, C., 1996. *Introduction to Solid State Physics*. 7th ed. John Wiley & Sons.

Korotkov, A. & Likharev, K., 1999. Resonant Fowler-Nordheim Tunneling through Layered Tunnel Barriers and its Possible Applications. In *Techn. Dig. IEDM.*, 1999.

Kotter, D.K., Novack, S.D., Slafer, W.D. & Pinhero, P., 2008. Solar Nantenna Electromagnetic Collectors. *ASME Conference Proceedings*, 2008(43208), pp.409-15.

Kroemer, H., 1994. *Quantum Mechanics*. New Jersey: Prentice-Hall.

Landauer, R., 1989. Johnson-Nyquist noise derived from quantum mechanical transmission. *Physica D*, 38, pp.226-29.

Landsberg, P.T. & Tonge, G., 1979. Thermodynamics of the conversion of diluted radiation. *J. Phys. A: Math. Gen.*, 12(4), pp.551-62.

Lent, C.S. & Kirkner, D.J., 1990. The quantum transmitting boundary method. *Journal of Applied Physics*, 67(10), pp.6353-59.

Liu, J. & Giordano, N., 1990. Nonlinear response of a mesoscopic system. *Physica B*, 165&166, pp.279-80.

Li, H., Wang, L., Lan, Z. & Zheng, Y., 2009. Generalized transfer matrix theory of electronic transport through a graphene waveguide. *Phys. Rev. B*, 79(15), p.155429.

Luryi, S., Xu, J. & Zaslavsky, A., 2010. *Future Trends in Microelectronics: From Nanophotonics to Sensors to Energy*. John Wiley \& Sons.

Marchenko, O., Kazantsev, S. & Windholz, L., 2007. *Demonstrational Optics Part 2, Coherent and Statistical Optics*. Springer.

Marks, A.M., 1984. Device for conversion of light power to electric power. *US Patent No. 4445050*.

MATLAB, 2010. *MATLAB R2010a, The Mathworks Inc.* Natick, MA.

Matsumoto, Y., Hanajiri, T., Toyabe, T. & Sugano, T., 1996. Single Electron Device with Asymmetric Tunnel Barriers. *Jpn. J. Appl. Phys.*, 35, pp.1126-31.

Middlebrook, C.T., Krenz, P.M., Lail, B.A. & Boreman, G.D., 2008. Infrared phased-array antenna. *Microwave and Optical Technology Letters*, 50(3), pp.719-23.

Midrio, M., Boscolo, S., Locatelli, A., Modotto, D., Angelis, C.D. et al., 2010. Flared Monopole Antennas for 10 µm Energy Harvesting. In *Proc. 40th European Microwave Conf.* Paris, France, 2010.

Mitchell, B., 2010. *Private communication*, University of California, Santa Barbara.

Moddel, G., 2001. Fractional bandwidth normalization for optical spectra with application to the solar blackbody spectrum. *Applied Optics*, 40(3), pp.413-16.

Moddel, G., 2009. Geometric Diode, Applications and Method. *US Patent Application 20110017284*.

Moddel, G. & Eliasson, B., 2004. High speed electron tunneling device and applications. *U.S. Patent No. 6,756,649*.

Nagae, M., 1972. Response Time of Metal-Insulator-Metal Tunnel Junctions. *Jpn. J. Appl. Phys.*, 11(11), pp.1611-21.

Nanotech, 2011. http://www.nnin.org/nnin_ucsb.html.

Nimtz, G. & Marquardt, P., 1988. On the corpuscle-wave transition of electrons in mesoscopic systems. *J. Phys. A: Math. Gen.*, 21, pp.L85-88.

Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D., Katsnelson, M.I. et al., 2005. Two-dimensional gas of massless Dirac fermions in graphene. *Nature*, 438, pp.197-200.

Palik, E.D., 1985. *Handbook of Optical Constants of Solids*. Orlando: Academic Press, Inc.

Palik, E.D., 1991. *Handbook of Optical Constants of Solids II*. San Diego: Academic Press, Inc.

Pedersen, M.H. & Büttiker, M., 1998. Scattering theory of photon-assisted electron transport. *Phys. Rev. B*, 58(19), pp.12993-3006.

Perello, D.J., Lim, S.C., Chae, S.J..L.I., Kim, M.J., Lee, Y.H. et al., 2010. Current anisotropy of carbon nanotube diodes: Voltage and work function dependence. *Appl. Phys. Lett.*, 96, p.263107.

Periasamy, P., Bergeson, J.D., Parilla, P.A., Ginley, D.S. & O'Hayre, R.P., 2010. Metal-insulator-metal point-contact diodes as a rectifier for rectenna., 2010.

Phiar, 2002. *Infrared Photon Field Detectors*. University of Colorado at Boulder and Phiar Corporation.

Phiar, C., 2007. Private communication.

Platero, G. & Aguado, R., 2004. Photon-assisted transport in semiconductor nanostructures. *Physics Reports*, 395(1-2), pp.1-157.

Probst, O.M., 2002. Tunneling through arbitrary potential barriers and the apparent barrier height. *American Journal of Physics*, 70(11), pp.1110-16.

Puri, A. & Schaich, W.L., 1983. Comparison of image-potential theories. *Phys. Rev. B*, 28(4), pp.1781-84.

Raether, H., 1988. *Surface Plasmons, vol. 111 of Springer-Verlag Tracts in Modern Physics*. New York: Springer-Verlag.

Ran, Q., Gao, M., Guan, X., Wang, Y. & Yu, Z., 2009. First-principles investigation on bonding formation and electronic structure of metal-graphene contacts. *Appl. Phys. Lett.*, 94(10), p.103511.

Reich, S., Maultzsch, J., Thomsen, C. & Ordejón, P., 2002. Tight-binding description of graphene. *Phys. Rev. B*, 66, p.035412.

Reinhart, K.F. & al., e., 2009. Low-cost Approach for Far-Infrared Sensor Arrays for Hot-Spot Detection in Automotive Night Vision Systems. In *Advanced Microsystems for Automotive Applications*. Springer Berlin Heidelberg.

Riccius, H.D., 1978. Improved metal-insulator-metal point-contact diodes for harmonic generation and mixing. *Appl. Phys. A*, 17(1), pp.49-52.

Riccius, H.D. & Siemsen, K.J., 1984. Point-contact diodes. *Applied Physics A: Materials Science & Processing*, 35, pp.67-74. 10.1007/BF00620632.

Richards, P.L., 1994. Bolometers for infrared and millimeter waves. *J. Appl. Phys.*, 76(1), pp.1-24.

Richards, P.L., 1997. Bolometers for infrared and millimeter waves. *J. Appl. Phys.*, 76(1), pp.1-24.

Rockwell, S., Lim, D., Bosco, B.A., Baker, J.H., Eliasson, B. et al., 2007. Characterization and Modeling of Metal/Double-Insulator/Metal Diodes for Millimeter Wave Wireless Receiver Applications. In *Radio Frequency Integrated Circuits (RFIC) Symposium, IEEE.*, 2007.

Roddaro, S., Pingue, P., Piazza, V., Pellegrini, V. & Beltram, F., 2007. The Optical Visibility of Graphene: Interference Colors of Ultrathin Graphite on SiO2. *Nano Lett.*, 7(9), p.2707–2710.

Rogalski, A., 2000. *Infrared Detectors*. Gordon and Breach Science Publishers.

Rogalski, A., 2003. Infrared detectors: status and trends. *Progress in Quantum Electronics*, 27(2-3), pp.59-210.

Rungger, I. & Sanvito, S., 2008. Algorithm for the construction of self-energies for electronic transport calculations based on singularity elimination and singular value decomposition. *Phys. Rev. B*, 78, p.035407.

Sanchez, A., C. F. Davis, J., Liu, K.C. & Javan, A., 1978. The MOM tunneling diode: Theoretical estimate of its performance at microwave and infrared frequencies. *Journal of Applied Physics*, 49(10), pp.5270-77.

Sancho, M.P.L., Sancho, J.M.L. & Rubio, J., 1984. Quick iterative scheme for the calculation of transfer matrices: application to Mo(100). *J. Phys. F: Met. Phys.*, 14, pp.1205-15.

Sarehraz, M., 2005. *Novel rectenna for collection of infrared and visible radiation*. PhD Thesis. University of South Florida.

Sarehraz, M., Buckle, K., Weller, T., Stefanakos, E., Bhansali, S. et al., 2005. Rectenna developments for solar energy collection. In *31st IEEE Photovoltaic Specialists Conference.*, 2005.

Schift, H., 2008. Nanoimprint lithography: An old story in modern times? A review. *Journal of Vac. Sc. & Tech. B: Microelectronics and Nanometer Structures*, 26(2), pp.458-80.

Sedra, A.S. & Smith, K.C., 1997. *Microelectronic circuits*. 4th ed. New york: Oxford University Press.

Sedra, A.S. & Smith, K.C., 2004. *Microelectronic Circuits*. 5th ed. New York: Oxford University Press.

Shinohara, N. & Matsumoto, H., 1998. Experimental study of large rectenna array for microwave energy transmission. *Microwave Theory and Techniques, IEEE Transactions on*, 46(3), pp.261-68.

Silberberg, M.S., 2007. Principles of general chemistry. New York: McGraw-Hill. p.666.

Simmons, J.G., 1963. Generalized Formula for the Electric Tunnel Effect between Similar Electrodes Separated by a Thin Insulating Film. *Journal of Applied Physics*, 34(6), pp.1793-803.

Simmons, J.G., 1971. Conduction in thin dielectric films. *Journal of Physics D: Applied Physics*, 4(5), p.613.

Singh, P., Kaneria, S., Anugonda, V.S., Chen, H.M., Wang, X.Q. et al., 2006. Prototype silicon micropower supply for sensors. *Sensors Journal, IEEE*, 6(1), pp.211-22.

Sokolov, I.M., 1998. On the energetics of a nonlinear system rectifying thermal fluctuations. *Europhys. Lett.*, 44(3), pp.278-83.

Solymar, L. & Walsh, D., 2010. *Electrical Properties of Materials*. 8th ed. Oxford University Press.

Song, A.M., Lorke, A., Kriele, A., Kotthaus, J.P., Wegscheider, W. et al., 1998. Nonlinear Electron Transport in an Asymmetric Microjunction: A Ballistic Rectifier. *Phys. Rev. Lett.*, 80(17), pp.3831-34.

Spieler, H., 2005. *Semicondutor Detector Systems*. New York: Oxford University Press Inc.

Stratton, R., 1962. Volt-current characteristics for tunneling through insulating films. *Journal of Physics and Chemistry of Solids*, 23(9), pp.1177-90.

Šunjić, M. & Marušić, L., 1991. Dynamical effects in electron tunneling: Self-consistent semiclassical image potentials. *Phys. Rev. B*, 44(16), pp.9092-95.

Synrad, 2010. *Series 48, L version lasers: Operator's manual*. [Online] Available at: http://www.synrad.com/Manuals/48series_manual.pdf.

Sze, S.M. & Ng, K.K., 2006. *Physics of Semiconductor Devices*. Wiley-Interscience.

Thouless, D.J., 1980. The effect of inelastic electron scattering on the conductivity of very thin wires. *Solid State Comm.*, 34, pp.683-85.

Tien, P.K. & Gordon, J.P., 1963. Multiphoton Process Observed in the Interaction of Microwave Fields with the Tunneling between Superconductor Films. *Phys. Rev.*, 129(2), pp.647-51.

Tiwari, B., Bean, J.A., Szakmany, G., Bernstein, G.H., Fay, P. et al., 2009. Controlled etching and regrowth of tunnel oxide for antenna-coupled metal-oxide-metal diodes. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, 27(5), pp.2153-60.

Torrey, H.C. & Whitmer, C.A., 1964. *Crystal Rectifiers*. Lexington: Boston Technical Publishers, Inc.

Tucker, J.R., 1979. Quantum limited detection in tunnel junction mixers. *IEEE J. of Quantum Electronics*, QE-15(11), pp.1234-58.

Tucker, J.R. & Feldman, M.J., 1985. Quantum detection at millimeter wavelengths. *Reviews of Modern Physics*, 57(4), pp.1055-113.

Tucker, J.R. & Millea, M.F., 1978. Photon detection in nonlinear tunneling devices. *Appl. Phys. Lett.*, 33(7), pp.611-13.

Umerski, A., 1997. Closed-form solutions to surface Green's functions. *Phys. Rev. B*, 55(8), pp.5266-75.

Ventra, M.D., 2008. *Electrical Transport in Nanoscale Systems*. 1st ed. Cambridge: Cambridge University Press.

Weiss, M.D., Eliasson, B.J. & Moddel, G., 2003. Device integrated antenna for use in resonant and non-resonant modes and method. *US Patent No. 6664562*.

Weisstein (a), E.W., 2011. Singular Value Decomposition. *From MathWorld-A Wolfram Web Resource*. http://mathworld.wolfram.com/SingularValueDecomposition.html.

Weisstein, E.W., 2011. Green's Function. *From MathWorld-A Wolfram Web Resource*. http://mathworld.wolfram.com/GreensFunction.html.

Wilke, I., Oppliger, Y., Herrmann, W. & Kneubühl, F.K., 1994. Nanometer thin-film Ni-NiO-Ni diodes for 30 THz radiation. *Applied Physics A: Materials Science & processing*, 58(4), pp.329-41.

Wu, Y. & Childs, P.A., 2011. Conductance of Graphene Nanoribbon Junctions and the Tight Binding Model. *Nanoscale Research Letters*, 6(62), pp.1-5.

Yon, J.J., Mottin, E., Biancardini, L., Letellier, L. & Tissot, J.L., 2003. Infrared Microbolometer Sensors and Their Application in Automotive Safety. In Valldorf, J. & Gessner, W. *Advanced Microsystems for Automotive Applications 2003*. Springer Berlin Heidelberg. pp.137-57.

Yoo, T. & Chang, K., 1992. Theoretical and Experimental Development of 10 and 35 GHz Rectennas. *IEEE Tran. on Microwave Theory and Techniques*, 40(6), pp.1259-66.

Zhu, Z., 2010. Geometric diodes fabrication using focused ion beam etching of Al-film. *Private communication*.

Zhu, Z., Grover, S., Krueger, K. & Moddel, G., 2011. Optical rectenna solar cells using graphene geometric diodes. *accepted in the 37th IEEE PVSC*.

Zia, R., Selker, M.D., Catrysse, P.B. & Brongersma, M.L., 2004. Geometries and materials for subwavelength surface plasmon modes. *J. Opt. Soc. Am. A*, 21(12), pp.2442-46.

APPENDIX - A

METAL/INSULATOR/METAL FIELD EFFECT TRANSISTOR

A. Introduction

As described in Chapter I, the metal/insulator/metal (MIM) diode consists of a nanometer-thin insulator sandwiched between two metal electrodes. Current flow between the two metal electrodes occurs via the tunneling of electrons across the insulator. The current depends exponentially on the distance of tunneling leading to the nonlinear dependence of current on diode voltage. Another method for changing the distance of tunneling, proposed by Fujimaru (Fujimaru, 1996), is via a gate voltage (V_G) that modifies the band profile of the MIM diode as shown in Figure A-.

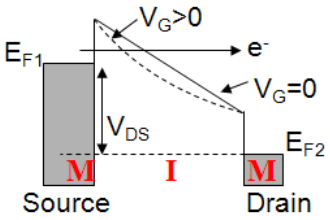


Figure A-1 Energy band diagram of the tunnel diode influenced by a gate voltage.

A positive gate bias causes the insulator band profile to bend downwards, which decreases the effective distance for Fowler Nordheim (FN) tunneling of electrons. A change in the band potential by fractions of a volt can result in a large change in the tunnel current. A structural schematic for such a device, referred to as the MIM field effect transistor (MIMFET), is shown in Figure A-2.

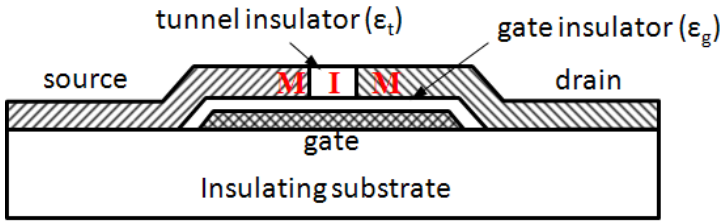


Figure A-2 Structural schematic for a metal-insulator-metal field effect transistor

In this appendix, I give the first demonstration of consistent simulated and experimental characteristics of a MIMFET. I also verify the operation of the MIMFET upto 100 kHz. Finally, I estimate its performance capability related to the maximum frequency of operation and the transistor gain achievable.

B. Background

Fujimaru et al. (Fujimaru, 1996) suggested the modification of a Schottky barrier tunnel transistor (Jang, 2003) using an MIM diode. The basic requirement in such a transistor is to have a high-barrier gate insulator for negligible gate leakage, and a low-barrier tunnel insulator chosen to have an adequate drive current and a low thermionic-leakage. Effective coupling of the gate voltage into the tunnel insulator (Shaker, 2003) requires that the ratio of dielectric constants $\varepsilon_g/\varepsilon_t$ be high and the thickness of the gate oxide be small.

Previous attempts at fabricating the MIMFET were based on two approaches: (i) anodic oxidation of a metallic strip (titanium) using an AFM tip in the presence of moisture to produce the tunnel junction (Snow, 1998), and (ii) sidewall oxidation of a metal strip followed by a second metal deposition (Fukushima, 1999). The AFM method is limited in scalability by the size of the tip, yielding a 40 nm wide $Ti/TiO_x/Ti$ tunnel junction for which results were reported only at 77 K. The second fabrication method uses the same Ti-based barrier with a 30 nm wide tunnel

insulator. Even for relatively low current densities, both these devices required high V_{DS}.

Desirable performance of MIMFETs has not been achieved in any of the previously reported results. In some of these works (Snow, 1998), (Fujimaru, 1999), field effect measurements are only reported for low temperatures. Extremely low current and a large gate-voltage that causes only a small change in tunnel current also limit the applicability of these devices. Moreover, simulations (Snow, 1998) predict large source-drain currents, which are far from the measured values. Two possible explanations for the inconsistency between the simulated and experimental results are, (i) the simulations are on a structure largely different from the fabricated device, (ii) tunneling is not the dominant conduction mechanism due to the large (> 10 nm) thickness of the tunnel insulator.

In lieu of these observations, I aim for a sub-10 nm thickness of the tunnel barrier expecting tunneling to be the dominant conduction mechanism. I also use Al_2O_3 as the gate oxide since it has a higher dielectric constant than SiO_2, which leads to better coupling between the gate and the tunnel dielectric.

C. Device design and fabrication

As shown in Figure A-2, the targeted MIMFET geometry has the source, drain, and gate electrodes separated by the tunnel and gate insulators on a sub-10 nm scale. Previous fabrication methods are limited in the choice and design of the tunnel-insulator. Also, in both those methods, surface conduction paths can possibly connect the source and the drain. I describe two alternative techniques for fabricating the MIMFETs. These methods differ in the orientation of the MIM tunnel diode and the gate electrode. I classify them as having a horizontal or a vertical gate. As discussed later, the horizontal-gate MIMFET avoids surface

conduction paths. Before proceeding, I acknowledge the support of Phiar Corporation in the development of this process.

1. Horizontal gate MIMFET

This approach is similar to the sidewall oxidation method (Fujimaru, 1999), except that I deposit the tunnel insulator(s) instead of oxidizing the sidewall. This technique of creating a nanoscale-thin, vertically-oriented tunnel-insulator is depicted in Figure A-. The tunnel insulator is conformally deposited over a step and anisotropically etched to create a thin sidewall without the use of lithography.

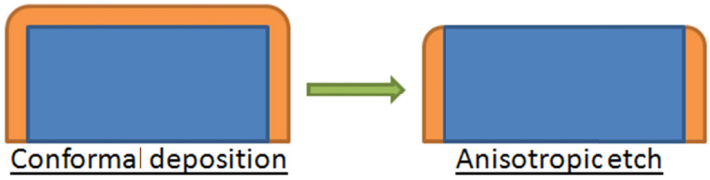


Figure A-3 Sidewall deposition technique for creating a nanoscale tunnel-insulator by conformal deposition of the oxide.

A cross-section schematic of the fabricated device is shown in Figure A-. As shown in the figure, two variations of tunnel insulators were used, which are discussed in Section C.3.

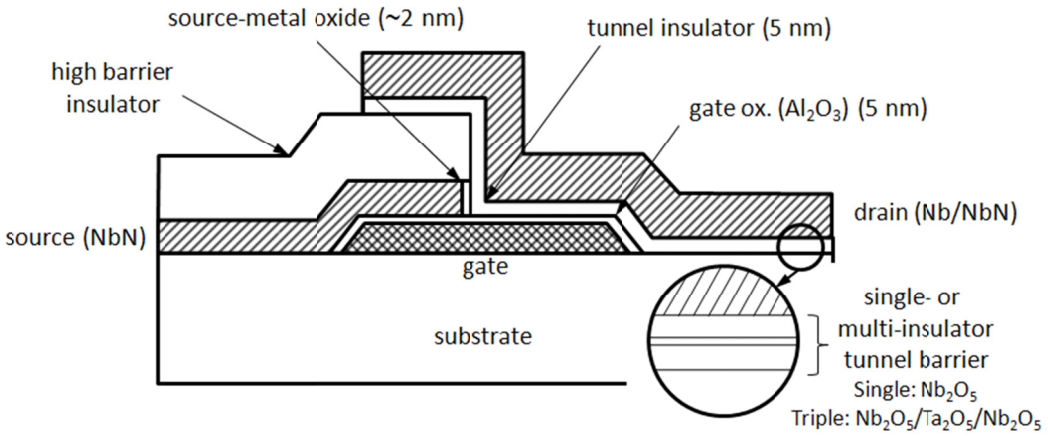


Figure A-4 Cross-section schematic for the MIMFET.

The device is fabricated on an oxidized silicon wafer using the following sequence of steps:

i. Pattern and deposit contact pads (Cr/Au) using metal evaporation and lift-off.
ii. Pattern and deposit gate metal electrode (Al) using evaporation and lift-off.
iii. Oxidize the gate metal to produce the gate oxide (Al_2O_3) using plasma oxidation done in the sputter deposition chamber.
iv. Conformally deposit source metal (NbN) by sputtering Nb in presence of nitrogen.
v. Conformally deposit SiO_2 using PECVD or sputtering. The SiO_2 serves as a spacer layer.
vi. Pattern and deposit a Cr mask using lift-off. This serves as an etch mask for the source.
vii. Etch the SiO_2 using reactive ion etching (RIE) in CHF_3+Ar plasma with endpoint detection.
viii. Etch the source metal using RIE using a mixture of CF_4 and O_2.
ix. Sputter deposit tunnel-insulator(s) and drain metal.
x. Pattern and etch drain metal and tunnel-insulator(s) in CF_4 and O_2 plasma, to obtain a narrow strip of materials overlapping with the patterned source metal.
xi. Pattern and deposit metal (Cr/Au) to contact to the drain electrode.

A TEM cross-section of the fabricated device, along with a schematic cross-section and the top-view microscope image is shown in Figure A-5. The fabricated device resembles the intended structure, except for the undercut and the sloped sidewall of the SiO_2 region. The undercut can be reduced by decreasing the height of the NbN source region. The sloped SiO_2 sidewall does not affect the performance of the MIMFET.

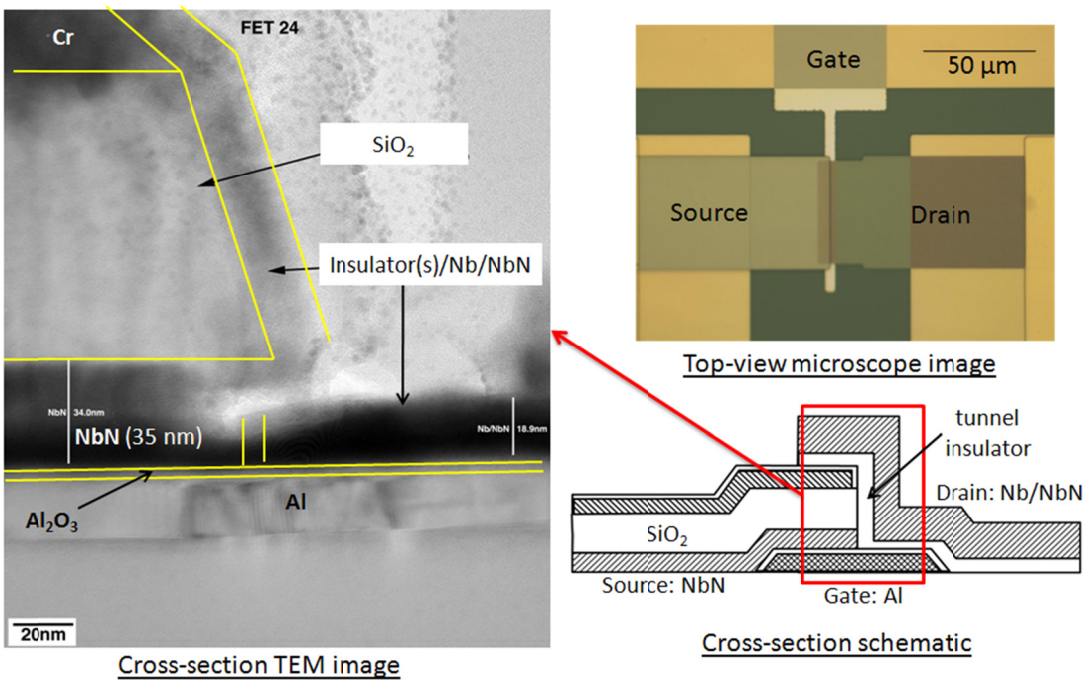


Figure A-5 Cross-section TEM image of the MIMFET. The various regions of the cross-section schematic are marked with yellow lines.

In the horizontal MIMFET fabricated above, the tunnel insulator layer separates the source and the drain electrodes, preventing a direct connection between them. Thus, a defect or a surface conducting layer cannot be formed between the source and the drain, ensuring that tunneling is the dominant conduction mechanism. However, this also implies an unwanted (tunnel) insulator below the drain electrode that increases the separation between the gate and the tunnel diode, which decreases the effectiveness of the gate. Also, as the MIM tunnel junction is not formed in a single step, the device is prone to interface impurities between the source metal and the tunnel insulator. This can be avoided by appropriate cleaning or preconditioning before the deposition of the tunnel insulator.

Any technique for fabricating the MIMFET, in which the tunnel-insulator is deposited instead of grown, gives the freedom to design the tunnel barrier. This is

the case with both the horizontal-gate MIMFET discussed above and the vertical-gate MIMFET described next.

2. Vertical gate MIMFET

In the vertical gate MIMFET, a gate electrode is placed on the sidewall of a narrow vertical stack of MIM as shown in Figure A-6. The gate electrode is electrically insulated from the stack by a gate insulator.

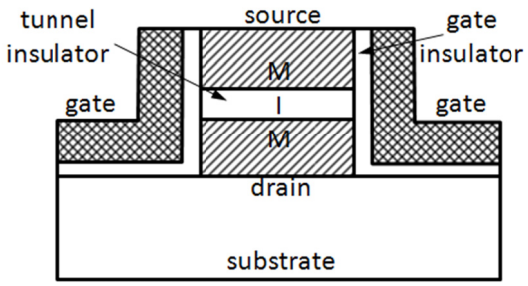


Figure A-6 Cross section schematic of the vertical gate MIMFET.

In the vertical MIMFET, the MIM tunnel barrier is deposited without breaking vacuum, which makes it less likely to have defects along the interfaces across which the tunneling of electrons occurs. The primary constraint for this structure is the width of the stack, which needs to be in the vicinity of 10 nm for effective coupling of the gate field into the tunnel insulator. However, in the fabricated device, this dimension is larger by an order of magnitude due to limitations in lithography.

The fabrication of this structure was done by Limin Cao using the diode process developed at Phiar Corporation. A CVD SiO_2 was deposited as the gate insulator followed by the gate metal deposition using evaporation.

Characterization of the vertical MIMFET is given in the next section. As described there, no modulation of the tunnel current was achieved using the vertical MIMFET. The primary reason for this is the undercut in the MIM stack, which occurs while defining the vertical edges using the RIE. In Figure A-7, I show the

TEM micrograph of an MIM diode, where the trapezoidal structure is a metal-mask used for etching. The CF_4 based RIE etch leads to a large undercut and the subsequent deposition of insulator produces air voids. On realizing this problem, work on this structure was discontinued.

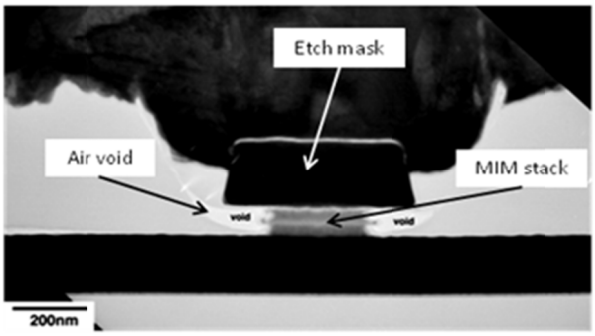


Figure A-7 TEM cross-section of a diode stack showing the undercut below the metal mask and the resulting air voids from the deposition of an insulator. This image is for an MIM diode and the placement of a gate electrode was not attempted. *Courtesy: Phiar Corporation.*

3. Tunnel insulator design

Better transistor characteristics can be obtained by designing the tunnel insulator(s) with appropriate dielectric constant and barrier height profile. To show the effect of dielectric constant, I compare the electrostatic potentials in a single- and a triple-insulator barrier for different gate voltages. The dimensions and the dielectric constant profiles for both these barriers are given in Figure A-8. The dielectric constants do not correspond to real materials, but are chosen to exemplify the advantage of a low-high-low dielectric profile in the triple-insulator diode.

As shown in the figure, both the tunnel-diodes are held fixed at a bias of 0.5 V, and the potential is calculated when V_G is 0 and 2 V. At V_G = 0 V, both the diodes have almost the same potential profile. However at V_G = 2 V, the change in potential of the triple-insulator is larger than the single-insulator. This happens

due to the dielectric profile of the triple-insulator where the low-ε insulators sandwich a high-ε insulator.

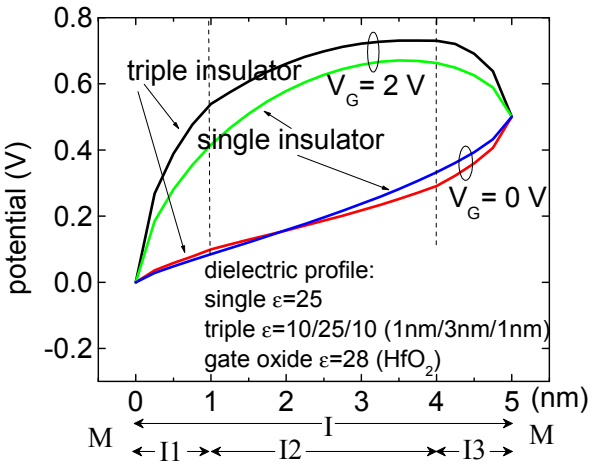


Figure A-8 Absolute potential in single- and triple-insulator for comparing the change in potential on applying a gate voltage.

With this arrangement, continuity of the electric displacement vector implies that most of the potential drop from the drain to the source occurs in the outer insulators. This causes a larger change in potential. The tunnel-current has a nonlinear dependence on the change in potential in the tunnel-insulator. Thus for the same change in gate voltage, a stronger gate effect is obtained by using a low-high-low dielectric profile.

Another design consideration is the barrier height of the insulators in the tunnel barrier. Here I again consider a single- and a triple-barrier as shown in Figure A-9.

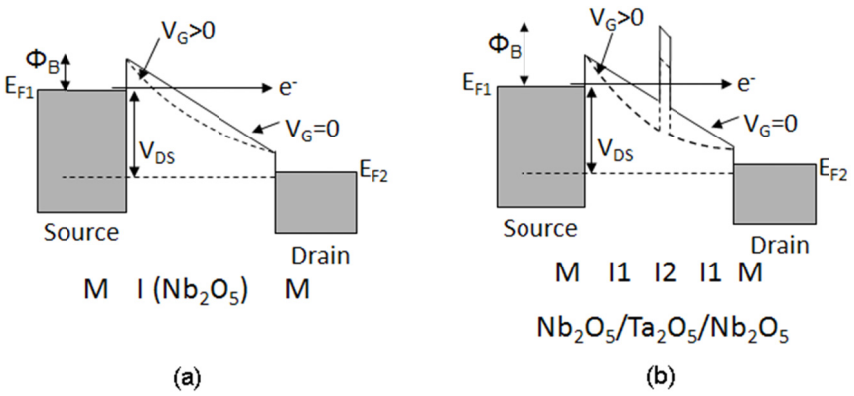


Figure A-9 Barrier profile for (a) single- and (b) triple-insulator tunnel diodes.

The triple-barrier has several advantages. First, the triangular quantum-well formed to the left of the centrally located high-barrier goes in and out of resonance as the gate voltage is varied. The presence or absence of a resonant level in the well, controlled by the gate voltage, implies a sharper turn-on of the transistor. Also, as shown in Figure A-8, the change in potential due to the gate is highest close to the center of the tunnel insulator. Therefore, the effect of the gate-voltage on the tunnel current increases due to the presence of the high barrier that operates as the current controlling element.

The high barrier also helps prevent thermionic current between the source and the drain. Thermionic current is caused by electrons, with sufficient thermal energy, flowing over the top of the tunnel barrier (Simmons, 1964). This excess current is not desirable as thermionic emission is a slow mechanism and is only regulated by the maximum barrier height. Having a narrow, high-barrier at the center allows the surrounding low-barrier insulator (I1) to be made even lower without affecting the thermionic current. This increases the current driving capability of the transistor.

In next section, I give the performance characteristics of the triple-insulator device. An experimental comparison of MIMFETs having single- and triple-

insulators was attempted. However, as explained later, the process variability did not allow conclusive results.

D. Electrical characterization

A two step DC characterization of the MIMFET is carried out to ascertain if an MIM diode exists and to see if the gate has any effect on the diode current. The measurements are made using the HP4145B parameter analyzer. The devices that showed desirable behavior at DC were also tested with an AC gate voltage to ensure that the gate-effect was due to the modulation of the tunnel current and not due to slow moving interface charges. For the horizontal MIMFET, the results reported are for the triple-insulator device with the largest gate effect. All measurements are done at room temperature.

1. Characteristics and failure mechanism for vertical gate MIMFET

While describing the fabrication of the vertical MIMFET, I have alluded to the structural impossibility of a gate effect due to the undercut from the RIE. Despite this, a change in the drain-source current occurred on varying the gate voltage as shown in Figure A-20.

As seen in the figure, the I_{DS} vs. V_G curves are sensitive to the direction in which the gate-voltage is scanned. Moreover, the current increases significantly above a threshold gate voltage, which is not expected from a MIMFET.

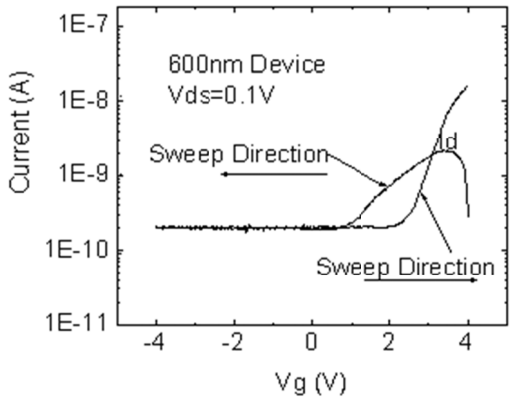


Figure A-20 Drain-source current vs. gate-voltage for the vertical-gate MIMFET. Although there is considerable change in current with V_G, it is not due to the modulation of the tunnel current. *Courtesy: Limin Cao*

To develop a better understanding for the directional dependence, I tested these devices for variation of I_{DS} using an AC gate voltage. I examined the I_{DS} vs. V_G on the x-y mode of the oscilloscope, starting with an extremely small frequency < 0.1Hz. For this frequency range, the characteristics resembled those seen in Figure A-20, with the forward and the reverse sweep not coinciding with each other at high V_G. However, as the frequency was cranked up from 0.1 to 1 Hz, the maximum value of I_{DS} dropped significantly until the x-y mode showed just a horizontal line. A possible explanation for this is a surface conduction layer along the vertical edge of the MIM stack that starts conducting above a certain gate voltage. This causes a change in the conductance between the source and the drain, leading to an increase in current only for a slowly increasing gate-voltage.

2. DC characteristics of horizontal-gate MIMFET

In Figure A-31, I show the variation of I_{DS} with V_{DS} for step change in V_G from 0 to 2 V. The curve for $V_G = 0$ V is identical to the diode I(V) curve for the MIM tunnel diode. Even for the $V_G \neq 0$ curves the characteristics resemble those of a diode with the resistance decreasing as the gate-voltage is increased.

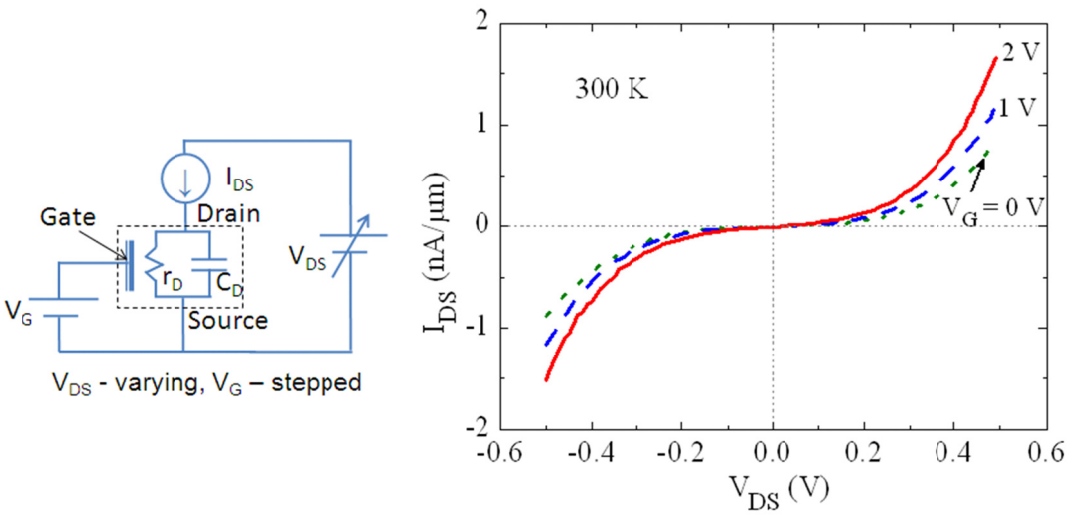


Figure A-31 Drain-source current vs. voltage curves for fixed gate-voltages.

Next, I varied the V_G while keeping the V_{DS} fixed. The result is shown in Figure A-42. The gate current is at least three orders of magnitude lower than the drain current, therefore the change in the drain current is primarily due to the effect of V_G on I_{DS}. At negative gate voltages, the change in current appears to saturate as the tunnel insulator bends convex up, resulting in a smaller fractional increase in tunnel distance as the gate voltage become more negative. On the other hand, the total tunnel distance at positive gate voltages decreases as V_G increases, thereby the percentage change in the tunnel distance increases. This leads to a sharper change in tunnel current for V_G>0.

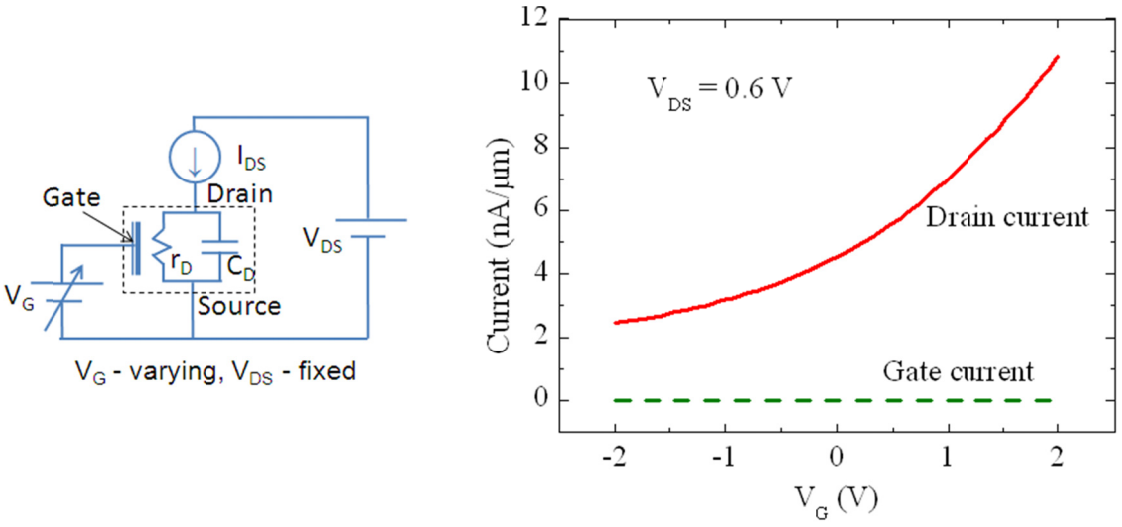


Figure A-42 I_{DS} vs. V_G measured at a constant V_{DS}. The gate current is orders of magnitude smaller than the drain current, implying low gate leakage.

Even though the overall change in current with gate-voltage is small, the horizontal MIMFET demonstrates the gate-effect in a device with a smaller width for the tunnel-insulator and at a lower V_{DS} than previously reported. As compared to the previous devices, the tunnel diode in the MIMFET has a higher current density resulting in an improved drive current at a lower V_{DS}. In Figure A-53, I compare the characteristics shown in Figure A-42 with the device reported by Fukushima et.al. (Fukushima, 1999).

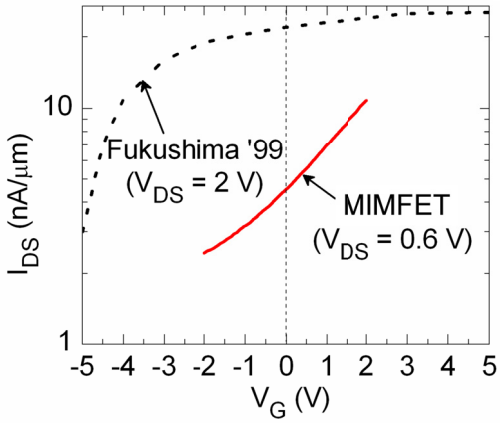


Figure A-53 Comparison of transistor characteristics of the MIMFET reported here and Fukushima's device.

The transconductance, defined as $g_m = \partial I_D / \partial V_G$, for the MIMFET is larger than Fukushima's near V_G = 0 V. The saturation in transconductance seen in Fukushima's device is inconsistent with the characteristics observed for the MIMFET. Later, I show that the behavior of the MIMFET is closer to what is expected from modeling.

3. AC characteristics of horizontal-gate MIMFET

To confirm the gate effect observed at DC, I test the transistor with a large-signal AC gate voltage. On applying an AC voltage signal on the gate a corresponding change in I_{DS} is observed. Since the response time for slow moving charges in the insulators and surface conduction states is expected to be large, their effect should disappear at a high enough frequency. Due to the small magnitude of the tunnel current and the large capacitive current that might overshadow it at high frequencies, I use a lock-in amplifier setup as shown in Figure A-64.

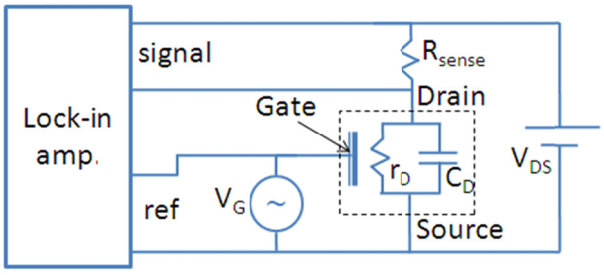


Figure A-64 (a) AC measurement setup for the MIMFET.

By measuring the current flowing in R_{sense} for V_{DS} = 0 and 0.8 V, I extract out the modulation in the tunnel current caused by the gate-voltage. In Figure A-75, this extraction is represented using current vectors for the two V_{DS} at low and high frequency. At V_{DS} = 0 V, the r_D vector is zero, and only the current through C_{GD} is sensed. At a finite V_{DS}, the total current represented by R_{sense} is sensed. The current

corresponding to the modulation of the tunnel insulator is extracted by the vector difference of C_{GD} and R_{sense}.

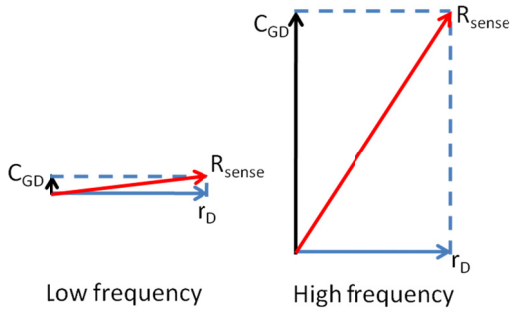


Figure A-75 Extracting the resistive component from the overall current sensed by the lock-in amplifier.

As shown in Figure A-86, this signal is constant with frequency. The gate-effect in the MIMFET is confirmed to exist at non-zero frequencies and remains constant up to the range of the lock-in amplifier used. The roll-off near 100 kHz is due to the gain-bandwidth limitation of the lock-in.

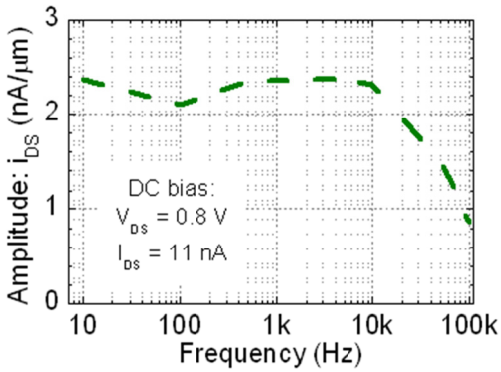


Figure A-86 Amplitude of AC drain-source current (i_{DS}) vs. frequency.

E. Device modeling

1. Simulator methodology and initial results

I model the $I(V)$ curve of the MIMFET to verify that the experimental device characteristics are consistent with expected behavior. The simulator is also used to project the performance of improved MIMFET structures. The simulation is divided into two parts. A 2D electrostatic potential calculation is required to estimate the effect of the gate voltage on the potential in the tunnel insulator. This is implemented using the PDE solver in MATLAB. Next, the quantum tunneling simulator developed in Chapter II is used to find the current flowing in the tunnel diode whose band profile is modified by the potential change due to the gate voltage.

In Figure A-97(a), I show a simplified MIMFET structure that is used for simulation.

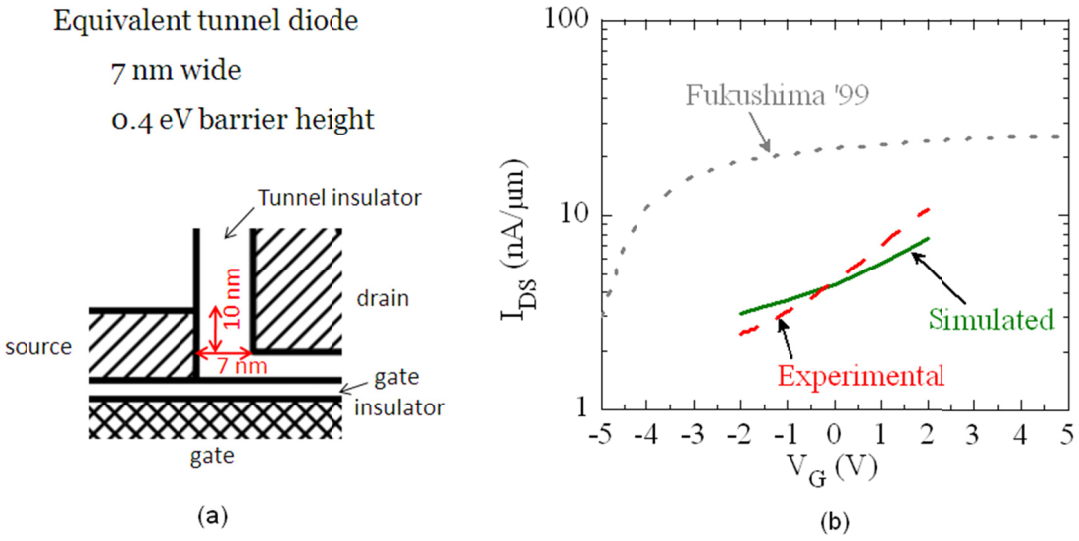


Figure A-97 (a) Simplified MIMFET structure used for simulation. (b) Comparison of simulated and experimental I_{DS} vs. V_G.

This structure is based on an MIM diode that is 7 nm wide and has a barrier height of 0.4 eV. Here the 7 nm width is expected from the processing. The 0.4 eV barrier height is chosen to match the current in the simulated diode with that in the

experimental device at V_G = 0 V. As shown in Figure A-97(b), the shape of the simulated curve is similar to the experimental one. I also show one of the previously reported (Fukushima, 1999) result for comparison.

This is the first demonstration of consistency between simulated and experimental characteristics of the MIMFET. Our simulated characteristics for the MIMFET are quite different from those simulated earlier (Shaker, 2003). Contrary to the previous report, the MIMFET does not show any saturation of I_{DS} at large V_G.

2. Simulation of improved structures

With experimental constraints in mind, I investigate possible improvements in the basic structure of the MIMFET. In Figure A-108, I compare three structures described below:

i. The first one (original) is the same as Figure A-97(a) and is repeated here for reference.
ii. The second structure is a modified version of the original one, obtained by removing the extra 5 nm insulator layer between the gate and the drain. This brings the tunnel junction closer to the gate electrode. It requires an extra RIE step to be performed before the deposition of the drain metal.
iii. In the third device, the dielectric constant of the gate insulator is increased from 10 to 28. This is possible by using a high-K dielectric like HfO_2.

In the I_{DS}-V_G curve shown in Figure A-108, the structural improvements lead to a substantial increase in the transconductance of the device (from 1 nS to 0.63 μS). Also, a drive current of 1μA/μm is achieved at a low V_{DS}.

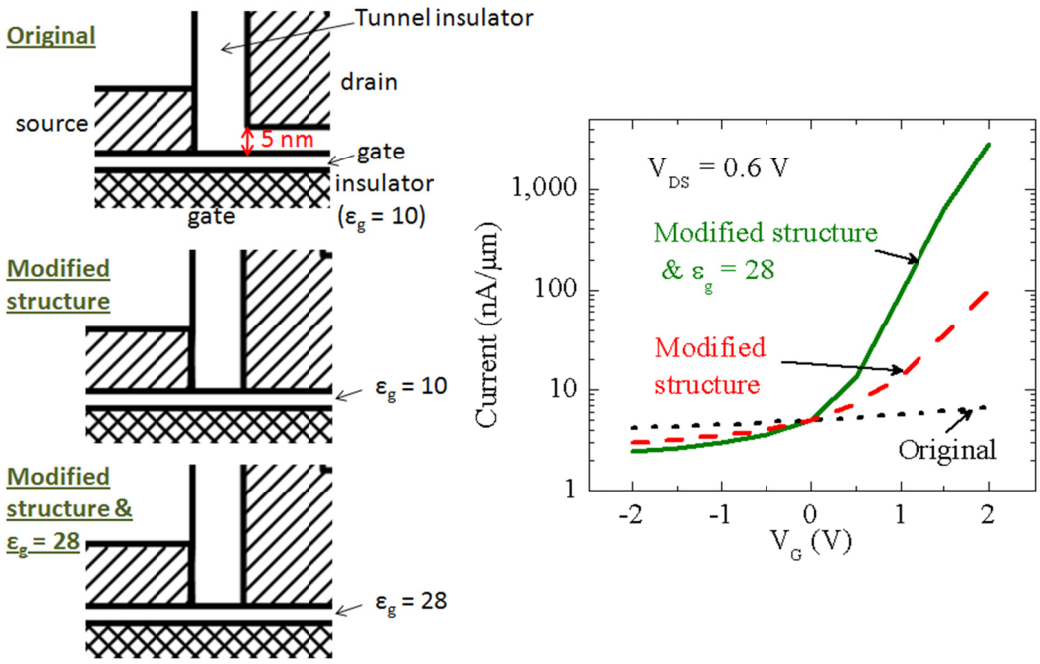


Figure A-108 Simulation of improved MIMFET structures. The I(V) curves indicate that a significant improvement in transconductance and drive-current is possible, with changes in the existing structure. Starting with the original device simulated in the previous sub-section, I modify it by removing the excess insulator between the gate and the drain. In the second improvement, the dielectric constant of the gate insulator is increased.

In the next section, I use the simulated characteristics to predict the maximum frequency of operation for the MIMFET.

3. Frequency of operation

The reason for pursuing a transistor compatible with metal-insulator electronics is twofold. First, to satisfy the need for a transistor that can be fabricated using the same technology that is employed for making other metal-insulator devices. Second, the MIMFET is a contender for a high frequency thin-film transistor (Reuss, 2005). Here I project the maximum frequency of operation that can be achieved by the MIMFET. I model the AC characteristics of the MIMFET using the circuit model shown in Figure A-19.

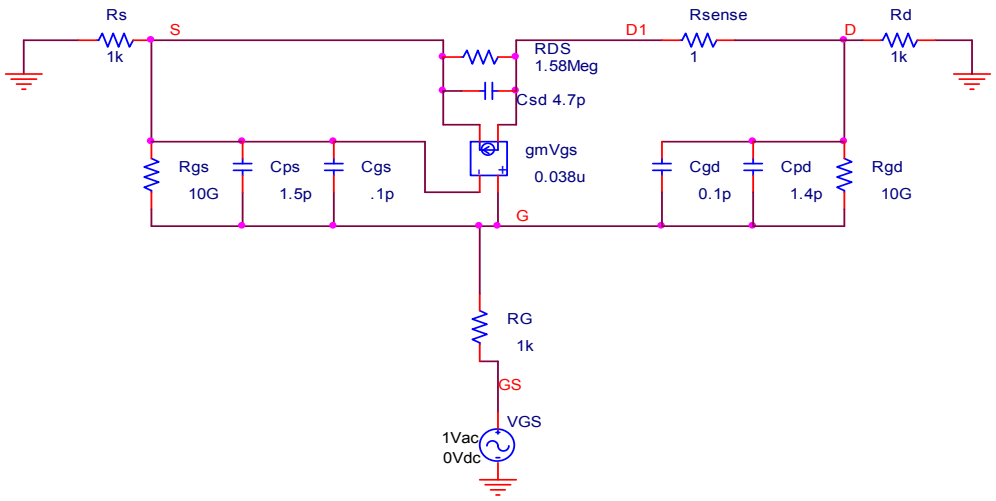


Figure A-19 Circuit model used in the frequency response simulation.

The diode is modeled by a gate controlled current source and a resistance in parallel. The parameters for these components are calculated from the measured characteristics. By simulating the frequency response of this circuit in SPICE, I determine the unity current-gain frequency (f_c). At f_c the modulated current in the drain-source diode becomes equal to the capacitive current flowing through the gate electrode. In Figure A-110, I show the results from three SPICE simulations with the parameters described below:

i. The first curve, with the lowest falloff frequency corresponds to the experimental result shown in Figure A-53. Apart from the transconductance (g_m = 1 nS/μm), a gate-source capacitance (C_{GS}) value of 33 fF/μm is measured.

ii. The second curve corresponds to the simulated device shown in Figure A-108 that has a modified structure and a high-K gate oxide. This device has an improved transconductance of 0.63 μS/μm. The C_{GS} in this simulation is reduced to 2.2 fF/μm, assuming that an extremely small gate width of 15 nm can be patterned using nanolithography.

iii. In the third simulation, I assume a transconductance of 1 mS/μm along with the reduced capacitance. The increased transconductance is achievable by designing a resonant tunnel diode as described in section C.3 or by applying a larger drain-source voltage. However, at large V_{DS}, the breakdown of the tunnel insulator may be problematic.

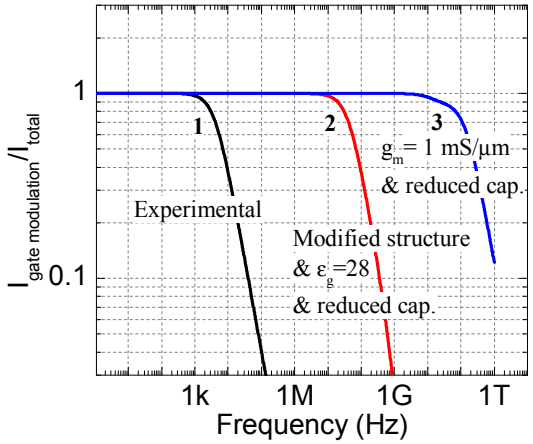


Figure A-110 Frequency response of gate modulated drain-source current. The ordinate equal to 0.5 gives the unity current-gain frequency estimate around 10 kHz, 100 MHz, and 100 GHz for the three curves.

Even though our experimental device is limited in its frequency of operation, there is substantial room for improvement in performance as see in Figure A-110. A modified structure along with an increased dielectric constant and a reduced capacitance lead to a unity current-gain frequency approaching 100 MHz. With further improvements in the MIM diode, a 100 GHz transistor is possible.

F. Conclusions and suggestions for future work

I have made the first demonstration of a metal/insulator/metal field effect transistor that has characteristics consistent with simulations. The horizontal-gate MIMFET has diode-like characteristics with the differential resistance decreasing

as the gate voltage is increased. The transconductance of this device, at zero gate voltage, is higher than previously reported. The device has also been demonstrated to work when an AC signal is applied on the gate. This verifies that the gate is controlling the tunnel barrier, and confirms the absence of slow moving charges in the insulators or along surfaces.

The fabrication method for the horizontal MIMFET is improved compared to the previous fabrication techniques as it allows significantly smaller tunnel-insulator widths. This allows the MIM diode to have high-current density at low voltages. The nanoscale junction in the MIM diode is made without using nanolithography. The key to this process is the deposition of the tunnel-insulator on the sidewall of the metal, where the width of the insulator is governed by the thickness of deposited material instead of lithography.

A drawback of the sidewall deposition method is the non-uniformity in electrical characteristics across the wafer. Due to the limited size of the sputtering source, the thickness and the angle of deposition vary at different locations on the wafer. Electrical properties of the MIMFET are extremely sensitive to variations in thickness of the tunnel insulator, which may occur due to the insulator deposition step or the reactive ion etching. The variation makes it extremely difficult to compare devices on different wafers. It also led to a null result for a number of fabrication experiments, including the comparison of the triple-insulator MIMFET with the single-insulator device.

The key requirement here is to develop a device that is less sensitive to unintended variations in fabrication, but allows the design to be reliably tweaked. Unlike MOSFETs, MIMFET fabrication is not based on a self-aligned process, which makes it difficult to align an extremely small gate with the sub-10 nm tunnel diode. Thus, a self-aligned approach is the key to have a narrow gate, which will

make the parasitic capacitance low. Finally, low barrier-height tunnel diodes need to be investigated for increasing the drive current of the MIMFET.

Apart from verifying experimental results, the device simulations have served as a tool to evaluate improved designs for the MIMFET. Results for some of these projected improvements indicate that minor changes in structure can lead to a substantial improvement in performance. Based on these changes, a MIMFET with a transconductance of 0.63 μS is achievable.

In light of the performance estimates, the MIMFET is a potential candidate for a thin-film transistor for analog signal processing. For digital applications, power dissipation may be a limiting factor as the I_{on}/I_{off} ratio is only around 10^4 as seen in Figure A-108. Another consideration is to have a thin-film technology that can operate at high frequency. The cutoff for the experimental device is low. However, in light of the suggested improvements, cutoff frequency in gigahertz is feasible.

References

Fujimaru, K. & Matsumura, H., 1996. Theoretical Consideration of a New Nanometer Transistor Using Metal/Insulator Tunnel-Junction. *Jpn. J. Appl. Phys.*, 35, pp.2090-94.

Fujimaru, K., Sasajima, R. & Matsumura, H., 1999. Nanoscale metal transistor control of Fowler-Nordheim tunneling currents through 16 nm insulating channel. *J. Appl. Phys.*, 85(9), pp.6912-16.

Fukushima, K., Ryouta, S., Kouji, F. & Hideki, M., 1999. A Novel Nanoscale Metal Transistor Fabricated by Conventional Photolithography. *Jpn. J. Appl. Phys.*, 38, pp.7233-36.

Jang, M., Kang, K., Lee, S. & Park, K., 2003. Simulation of Schottky barrier tunnel transistor using simple boundary condition. *Appl. Phys. Lett.*, 82(16), pp.2718-20.

Reuss, R.H., Chalamala, B.R., Moussessian, A., Kane, M.G., Kumar, A. et al., 2005. Macroelectronics: Perspectives on Technology and Applications. *Proc. IEEE*, 93(7), pp.1239-56.

Shaker, A. & Zekry, A.H., 2003. Theoretical Investigation of Single- and Dual- Gate Metal Insulator Tunnel Transistors. In *ICM 2003.*, 2003.

Simmons, J.G., 1964. Potential Barriers and Emission-Limited Current Flow Between Closely Spaced Parallel Metal Electrodes. *J. Appl. Phys.*, 35(8), pp.2472-81.

Snow, E.S., Campbell, P.M., Rendell, R.W., Buot, F.A., Park, D. et al., 1998. A metal/oxide tunneling transistor. *Semicond. Sci. Technol.*, 13, pp.A75-78.

Snow, E.S., Campbell, P.M., Rendell, R.W., Buot, F.A., Park, D. et al., 1998. A metal/oxide tunneling transistor. *Appl. Phys. Lett.*, 72(23), pp.3071-73.