Surface Plasmon Enhanced Optoelectronic Devices

by

Saba Siadat Mousavi

Thesis submitted to the University of Ottawa in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

Department of Electrical Engineering and Computer Science, Faculty of Engineering

University of Ottawa

Abstract

Surface Plasmon Polaritons (SPPs) are electromagnetic waves coupled to the free electrons at the surface of metals, which propagate along the interface of metal and dielectric at optical frequencies. SPPs have found many applications in communications, sensing, and photovoltaics, among others, due to their subwavelength confinement and high sensitivity. These qualities can significantly reduce device footprint, while enhancing device performance.

This thesis investigates three novel surface plasmon enhanced optoelectronic device concepts, namely two photodetectors and an electro-optic intensity modulator, at wavelengths in the photonic C-band. It is demonstrated, theoretically and numerically, that involving SPPs improves speed and sensitivity of these devices, while significantly reducing their dimensions, compared to conventional counterparts.

The first device proposed and investigated is a photodetector which employs arrays of nanodipoles, as a plasmonic metasurface, in order to localize light in subwavelength InGaAs detection regions, placed within the gaps of Au nanodipoles. As a result, the speed-responsivity trade-off, which is common in conventional photodetectors, is overcome. Numerically, responsivities of 100 mA/W and electrical bandwidths of up to 4 THz are predicted.

The second device is a photodetector which exploits tightly confined SPPs generated in a film of InGaAs, covered by arrays of Au nanomonopoles. By carefully designing these arrays of nanomonopoles, responsivities up to 200 mA/W were achieved for electrical bandwidths as high as 1 THz, at the wavelength of 1550 nm.

Finally, the fabrication of an electro-optic intensity modulator, incorporating grating couplers, is demonstrated and discussed. Modulation is based on enhanced perturbation of the effective refractive index of grating-coupled surface plasmon polaritons propagating along a metal–oxide–semiconductor structure on silicon. A front-side probing technique was employed, which enabled modulation in transmission, as well as reflection. Lithography techniques were optimized to produce high resolution devices.

Acknowledgments

Foremost, I would like to express my sincere gratitude to my advisor Prof. Pierre Berini for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me throughout my research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

My earnest thanks also go to Dr. Andreas Stohr for his valuable guidance and support, specially through the first portion of this work.

I must express my very profound gratitude to my dad “Hamid” and mom “Farideh”; for all their moral and emotional support, for their continuous encouragement throughout my years of study and through the process of research and writing this thesis, and for all the sacrifices they made to give the opportunity to grow in my desired field. This accomplishment would not have been possible without them.

I am also grateful to Anthony Olivieri and Ewa Lisicka-Skrzek, the lab managers at Berini’s Optics Lab and the NanoFab Lab at the Centre for Research in Photonics at the University of Ottawa for their unfailing help and support during my PhD research. Without their support the research could not have been successfully completed. I would also like to express my sincere gratitude to Howard Northfield for his help and motivation.

My special thanks goes to the person who believed in me when I lost hope: my Andisheh.

And finally, last but by no means least; to all my friends and colleagues in the Berini’s group for numerus hours of scientific and non-scientific discussions, and their invaluable friendship and support. It was great sharing the laboratory with all of you during this journey at the University of Ottawa.

Table of Contents

Table of Figures

Acronyms

AFM	Atomic Force Microscopy
ALD	Atomic Layer Deposition
BER	Bit Error Rate
BOE	Buffered Oxide Etch
e-beam	Electron beam
EPH	Electron-Hole Pair
FDTD	Finite Difference Time Domain
FEM	Finite Element Method
FIB	Focused Ion Beam
GSG	Ground-signal-ground
HB	Hard Bake
HGPWM	Hybrid Graphene-Plasmonic Waveguide Modulator
HHG	High Harmonic Generation
IPE	Internal Photoemission
IQE	Internal Quantum Efficiency
LN	Lithium niobate
LRDSPP	Long Range Dielectric-loaded Surface Plasmon Polariton
MCF	Multi-core Fiber
MOS	Metal-Oxide-Semiconductor

MSM	Metal-Semiconductor-Metal
MSM-PD	Metal-semiconductor-metal Photodetector
MZI	Mach-Zehnder Interferometer
MZM	Mach-Zehnder Modulator
NP	Nanoparticle
PEB	Post Exposure Bake
PML	Perfectly Matched Layers
PPM	Plasmonic Phase Modulator
PR	Photoresist
SEM	Scanning Electron Microscopy
SPP	Surface Plasmon Polariton
SPPAM	Surface Plasmon Polariton Absorption Modulator
TL	Transmission Line
TM	Transverse Magnetic

Symbols

A	Absorptance
A^*	Effective Richardson constant
BW	Bandwidth
C	Capacitance
c	Speed of light in free space
$\vec{D}$	Electrical displacement
d	Density of electrons
$\vec{E}$	Electric field
E_g	Bandgap energy of semiconductor
e	Charge of electron
$\vec{H}$	Magnetic field
h	Plank's constant
I_d	Dark current
I_p	Photocurrent
J	Current density
k	Wavenumber
k_B	Boltzmann's constant
m	Effective mass of electron
n	Refractive index
P	Poynting vector
P_{abs}	Absorbed power
Q'_{ss}	Equivalent fixed charge per unit area in oxide
q	Elemental charge
R	Reflectance
R_{esp}	Responsivity

r_{ij}	Pockel's coefficient
S_{abs}	Absorbed optical power
S_{inc}	Incident optical power
T	Transmittance
T_0	Absolute temperature
t	Time
x_{dT}	Maximum depletion width
α	Attenuation constant
β	Phase constant
γ	Characteristic collision frequency
ε_0	Permittivity of free space
ε_d	Permittivity of dielectric
ε_m	Permittivity of metal
ε_r	Complex relative permittivity
η_e	External quantum efficiency
η_i	Internal quantum efficiency
Λ	Grating period
υ	Frequency
ρ	Resistivity
τ	Time constant
Φ_B	Schottky barrier energy
φ	Electric potential distribution
ϕ_{ms}	Metal-semiconductor wavefunction difference
ω_P	Plasma frequency

CHAPTER 1. INTRODUCTION

This chapter presents an introduction to surface plasmon polaritons and surface plasmon-based optoelectronic devices, and their ever-increasing role in communication systems. In particular, this discussion focuses on plasmonic photodetectors and modulators, and describes the motivation for studying the topic of this thesis. The related body of literature is also reviewed. Finally, the scope and organization of this thesis is presented.

1.1. Surface plasmon polaritons

Self-sustained collective excitations of conduction electrons at the surface of metals are called plasma oscillations or plasmons. It is possible to couple external electromagnetic fields to these collective charge density oscillations at the interface of a metal and a dielectric, under certain circumstances. The result is an electromagnetic wave which propagates along the surface of the metal, and whose amplitude decays exponentially with increasing distance from the surface. These hybrid waves are known as surface plasmon polaritons (SPPs) [1]–[3].

The dispersive properties of metals (*e.g.* Au, Ag, Cu, and Al) at optical frequencies, particularly pertaining to the negative real part of permittivity, are such that visible and near-infrared light may couple to plasma oscillations at the surface. This coupling occurs only in the presence of *e.g.* surface roughness, gratings or prisms, which can manipulate the wavenumber of the incident radiation to match that of the surface electronic oscillations. Normally, the wavenumber of SPPs is larger than freely propagating light, which makes direct coupling impossible. SPP fields are transverse-magnetic (TM) polarized, which means their magnetic field lies in the plane of metal surface, perpendicular to the direction of propagation of SPPs, which is along the metal-dielectric interface. Figure 1.1-1 illustrate schematic views of the electromagnetic fields associated with an SPP, propagating along the planar interface of a semi-infinite dielectric medium and a semi-infinite metal. These electromagnetic waves are solutions of Maxwell's equations in each medium, subjected to the boundary conditions.

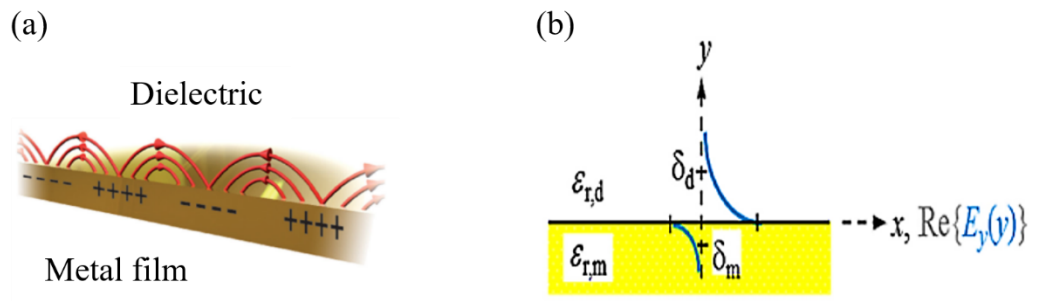

Figure 1.1-1 (a) Schematic of a propagating surface plasmon wave along a metal-dielectric interface (adapted from [4]) and (b) real component of the normal electric field of the surface plasmon polariton propagating along the interface of a semi-infinite metal and a semi-infinite dielectric. (Adapted from [5].)

For electromagnetic waves to propagate along a metal-dielectric interface, a component of the electric field $\vec{E}$ must exist normal to the interface. This condition limits the solutions of Maxwell's equations to those with a component of magnetic field $\vec{H}$ parallel to the interface. Also, according to the boundary conditions, electric and magnetic fields tangential to the interface must be continuous. On the other hand, due to the nature of SPPs, only solutions with fields decaying exponentially away from the interface are acceptable. Therefore, choosing the positive z-axis as the direction of propagation, we write

$$\vec{E} = \left[0, E_y(y), E_z(y)\right] e^{-jkz} e^{-j\omega t} \tag{1}$$

$$\vec{H} = \left[H_x(y), 0, 0\right] e^{-jkz} e^{-j\omega t} \tag{2}$$

where ω is the angular frequency, t is time, and $k = \beta - j\alpha$ is the SPP wavenumber, with β and α being the phase and attenuation constants, respectively. Further, the wavenumber of SPPs propagating along a metal-dielectric interface is obtained as [2]:

$$k = \frac{\omega}{c}\left(\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d}\right)^{1/2} \tag{3}$$

where ε_m and ε_d are the permittivity of the metal and dielectric, respectively. Based on the Drude model, the permittivity of the metal is given by [2]:

$$\varepsilon_m(\omega) = 1 - \frac{\omega_P^2}{\omega^2 + j\gamma\omega} \tag{4}$$

where γ is the characteristic collision frequency and $\omega_P^2 = \frac{de^2}{\varepsilon_0 m}$ is the plasma frequency of the free electron gas (confined to the metal), with d being the density of electrons, e the charge of electrons, and m the effective mass of electrons. ε_0 is the permittivity of free space.

1.2. Surface plasmon enhanced optoelectronic devices

As the name implies, optoelectronic circuits utilize a combination of electronics and optics for faster and more efficient signal processing, storage, and transmission of data. While optical transmission provides very wide bandwidths and low signal attenuation, the size mismatch between electronic and optical components becomes a limiting factor in many applications. For example, in processor chips, as the electronic transistors get increasingly smaller, the need for finding a way to focus light in subwavelength nanoscale structures becomes more evident. In the past couple of decades surface plasmon-based optoelectronic devices, such as waveguides and couplers, have proven to bridge this size mismatch between electronics and optics, by confining light in nanoscale areas [4], [6]–[9].

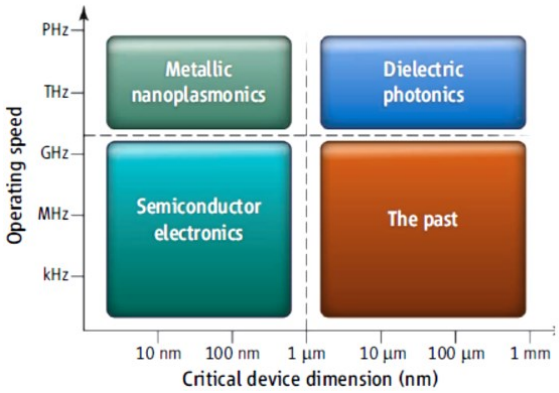

Figure 1.2-1 characterization of different technologies in terms of operation speed and device dimensions. The dashed lines indicate physical limitations of each technology. (Adapted from [8].)

By enhancing localized electric fields, SPPs increase the sensitivity of incident light to the optical properties of the medium. Therefore, it is possible to manipulate the properties of light by electrically manipulating the properties of the medium. SPP confinement to a metallic surface can be significant, with field decay lengths much smaller than the wavelength of light in air. This high confinement makes guiding light in subwavelength metallic structures possible. Thus, miniaturized optoelectronic components, such as waveguides, couplers, switches, modulators, and photodetectors can be constructed as different parts of the circuit, each approaching the size of state-of-the-art electronic devices.

Plasmonic waveguides include various geometries [1], however, they suffer from a fundamental trade-off between confinement and propagation loss. One solution to overcome this trade-off is hybrid waveguides, which consist of merging plasmonic and dielectric waveguides [10]. Alternatively, having a gain medium adjacent to the plasmonic waveguide can be considered, where the gain compensates for the propagation losses [11]. Plasmonic couplers, such as gratings and nanodot couplers, are required to connect plasmonic waveguide ports to conventional diffraction-limited photonic devices [12], [13].

Active control of SPPs, either electrically, optically, or via temperature, is the basis for plasmonic switches and modulators, where modulation depths of 80% and switching times in the order of picoseconds are expected [14]. Plasmonic devices also fulfill the need for achieving a nanometer-scale footprint, high-speed, and low power-consumption [7], [15].

Conventional photodetectors are constrained by the diffraction limit, in lateral dimensions, and by the finite absorption depth of semiconductors, in the vertical dimension. A photodetector's speed is generally determined by either its carrier transit time, which is proportional to the length of the detection region, or by its RC time constant, which is highly dependent on the device's effective capacitance (and area). Power consumption also scales with the area of the photodetector, and may be significantly reduced by down-sizing the device [16]. Plasmonic photodetectors have shown promising attributes over conventional semiconductor photodetectors, by employing plasmonic structures such as nanoantennas, nanogratings and waveguides [17]–[22].

1.3. Literature review

A large number of studies have been reported as researchers investigate and put to test new ideas within the context of SPP-enhanced photodetectors and modulators. This section reviews some of these studies.

1.3.1 Plasmonic Photodetectors

High speed photodetection is critical in integrated electronic-photonic systems. Surface plasmon photodetectors can provide fast and effective photodetection by using SPPs and their unique properties in the photodetection process [23]. Internal photoemission (IPE) and electron-hole pair (EHP) creation are the main detection mechanisms involved in surface plasmon photodetectors [24], [25]. When photons with energy $e = h\upsilon$ (h is Planck's constant, υ is the frequency) strike a metal-dielectric interface and are absorbed therein hot carriers are created in the metal and under the right conditions may be photo-emitted into the semiconductor region. An abrupt metal-

semiconductor interface (which is the structure of many surface plasmon photodetectors) forms a Schottky contact. Thus, as long as the photon energy $h\upsilon$ is greater than the Schottky barrier energy Φ_B ($\Phi_B < h\upsilon$) hot carriers transiting in the direction of the barrier may be emitted over the Schottky barrier into the semiconductor (IPE), where they are collected as photocurrent by applying a reverse bias voltage. This process is schematically shown in Figure 1.3-1. On the other hand, for EHP creation to occur, photon energies greater than the bandgap energy of the semiconductor E_g ($E_g < h\upsilon$) are required. In this case, the incident radiation is absorbed in the semiconductor which leads to the creation of electron-hole pairs therein. EHPs are then collected under a reverse bias voltage in the form of photocurrent. While creation of EHPs in the semiconductor is more efficient than IPE, the latter is particularly useful for detection at energies below the bandgap of the semiconductor ($\Phi_B < h\upsilon < E_g$) [24]. IPE and EHP creation can be enhanced by involving SPPs.

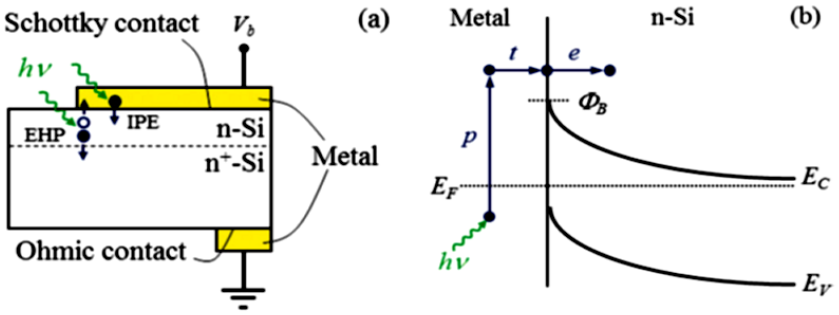

Figure 1.3-1 (a) IPE and EHP mechanisms illustrated in a representative MSM structure. Electrons are shown by filled circles and holes by unfilled ones. (b) Energy band diagram of a Schottky contact on n-Si and the three-step IPE process. (Adapted from [24].)

The most important specifications used to characterize photodetectors are internal and external quantum efficiencies, responsivity, speed, dark current, and sensitivity. The relative importance of each figure of merit depends on the application of the device.

Internal quantum efficiency η_i is a measure of the number of carriers that contribute to the photocurrent I_p *versus* the absorbed photons per second [26]:

$$\eta_i = \frac{I_p/q}{S_{abs}/h\upsilon} \tag{5}$$

where S_{abs} is the absorbed optical power, in the metal or semiconductor, leading to the photocurrent, and q is the elemental charge. Similarly, the external quantum efficiency is defined as the number of carriers that contribute to the photocurrent for the incident optical power S_{inc} [26]:

$$\eta_e = \frac{I_p / q}{S_{inc} / h\upsilon} \tag{6}$$

The responsivity R_{esp} is the ratio of the photocurrent to the incident optical power [26]:

$$R_{esp} = \frac{I_p}{S_{inc}} = \frac{\eta_e q}{h\upsilon} = \frac{A\eta_i q}{h\upsilon} \tag{7}$$

where A is the absorptance in the semiconductor. The speed of a photodetector typically refers to its 3-dB electrical bandwidth, determined by the maximum of its RC time constant or carrier transit time (for a lumped-element device):

$$BW = \frac{1}{2\pi(\tau_1 + \tau_2)} \tag{8}$$

where τ_1 and τ_2 are the RC time constant and the carrier transit time, respectively. The dark current is the main contributor to the noise performance of the photodetector [26]:

$$I_d = C_{area} A^* T_0^2 \exp(\frac{-q\Phi_B}{k_B T_0}) \tag{9}$$

where C_{area} is the Schottky contact area, A^* is the effective Richardson constant, T_0 is the absolute temperature, Φ_B is the Schottky barrier height, and k_B is Boltzmann's constant.

Plasmonic photodetectors combine various plasmonic nanostructures with photodetection structures to alter or improve these specifications.

One of the early studies on plasmonic photodetectors was done in 2008 by Tang and colleagues [18]. They proposed and experimentally demonstrated a subwavelength metal-semiconductor-metal (MSM) photodetector, based on open-sleeve dipole nanoantennas. The active region of the photodetector was made of germanium, placed in the gap of the dipole and extended under the two sleeves, which were employed here as electrodes. Dipole arms were 50 nm wide and 160 nm long. The target wavelength was 1310 nm to 1480 nm. The device was fabricated using a two-step focused-ion-beam (FIB) milling process. Photocurrent measurements showing enhancements by a factor of 20 were observed in the photocurrent due to the presence of the plasmonic nanoantenna.

A few years later, Tan and colleagues [27] numerically demonstrated a metal-semiconductor-metal (MSM) photodetector based on nanogratings and metal nanoparticles (NPs) embedded in a layer of amorphous silicon germanium on a GaAs substrate. As illustrated in

Figure 1.3-2(a), the photodetector structure consisted of two sections: the top section, which was the nanograting and a subwavelength slit, placed over a Ge or Si active layer, and the bottom section, which was formed by metal nanoparticles embedded in the Ge or Si layer, on a GaAs substrate. The incident light was coupled into SPPs propagating along the interface of the metal nanograting and the semiconductor active region (Ge or Si layer) below the nanograting. Although significant absorption enhancement was achieved, this enhancement was observed only within a few hundred nanometers below the subwavelength slit in the substrate. Incorporation of the bottom NPs expanded the light absorption area, hence the responsivity of the photodetector. Simulation results showed 28-fold and 3.5-fold enhancement in absorption compared to conventional non-plasmonic and plasmonic MSM-PD structures, respectively. Figure 1.3-2(b) plots the absorption enhancement factor as a function of wavelength for three cases of plasmonic MSM-PDs.

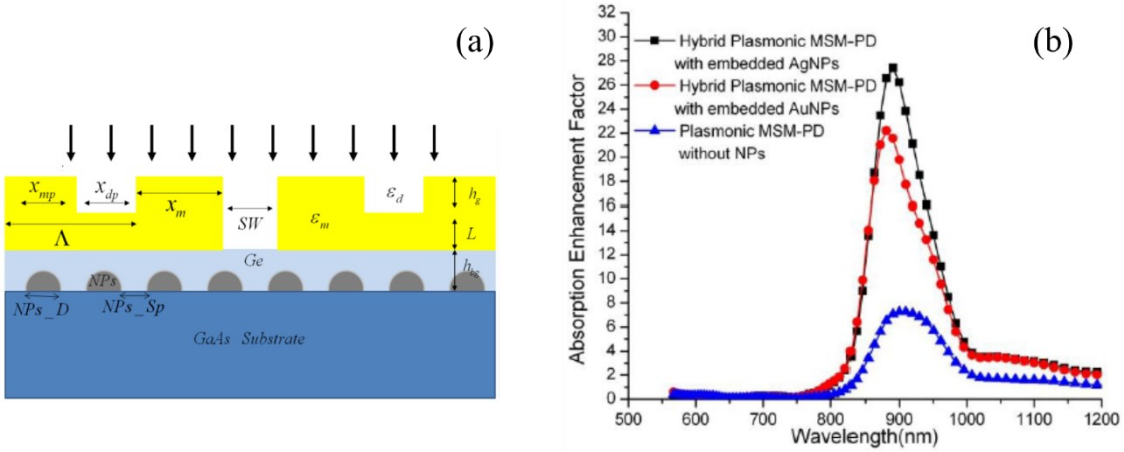

Figure 1.3-2 (a) Schematic view of the proposed hybrid plasmonic MSM-PD structure, employing a nanograting in conjunction with embedded metal NPs. (b) Absorption enhancement factor of the optimized hybrid plasmonic MSM-PD with embedded Au NPs, and the hybrid plasmonic MSM-PD with embedded Ag NPs and the conventional plasmonic MSM-PD. (Adapted from [27].)

Among other applications, optical to terahertz conversion through photoconduction has shown promising performance. Berry and colleagues [28] demonstrated the potential of plasmonic electrodes for enhancing the quantum efficiency and responsivity of conventional photoconductive emitter/detectors, both numerically and experimentally. For this purpose, plasmonic gratings were integrated into the contact electrodes of a conventional photoconductive emitter/detector. The contact electrodes were laid down on a low-temperature-grown GaAs substrate, and connected to a terahertz bowtie antenna. Photocarriers were generated upon incidence from a $\lambda_0 = 800$ nm laser pump, which in turn induced a photocurrent under an applied bias voltage. This photocurrent drove the bowtie antenna, generating terahertz radiation. The

gratings were designed to couple an incident optical beam to the surface plasmon waves excited along the interface of the metallic gratings and GaAs substrate. Confined SPPs significantly increased optical absorption and photocarrier generation in the vicinity of metal contacts, resulting in a 50 times increase in the emitted radiation power, and 30 times higher detection sensitivity comparing to a conventional photoconductive emitter/detector.

Luo and colleagues [29] demonstrated high-performance near-infrared photodetectors based on silicon nanowires. A film of graphene decorated by gold nanoparticles coated the array of Si nanowires. A responsivity of 1.5 A/W was estimated at zero bias voltage. The on/off switching ratio reached as high as 10^6. Numerical simulations showed that the nanoparticles deposited on graphene significantly enhanced light trapping and SPP excitation in Si nanowires, which improved the performance of the device. Figure 1.3-3 shows electric field energy density distributions in Si nanowires as a function of nanoparticle diameter.

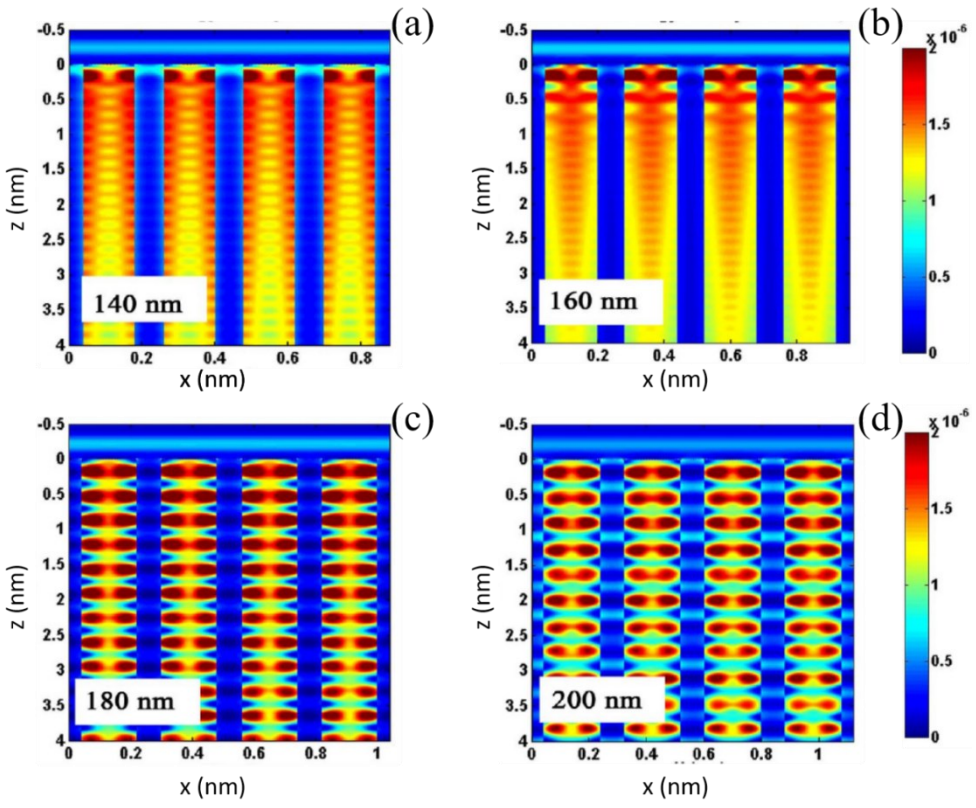

Figure 1.3-3 Cross sectional view of simulated electric field energy distribution in an array of graphene/SiNWs of diameter of (a) 140 nm, (b) 160 nm, (c) 180 nm, and (d) 200 nm. (Adapted from [29].)

A plasmonic photodetector based on an array of nano-scale active regions was proposed by Guo and colleagues [30]. A unit cell consisted of rectangular-shaped active regions, embedded in a metallic layer, surrounded by four circles at every corner. The circles were etched through

the metallic layer down to the dielectric layer in between the substrate and the top metal layer to couple the incident light to the SPPs propagating at the metal-dielectric interface and enhance light confinement, thus absorptance, in the active regions. InGaAs was chosen as the active material, and Ag as the metal. The wavelength of operation was 1550 nm. An absorptance of 74% and a responsivity of 0.74 A/W were numerically predicted for an optimized device. The 3 dB electrical bandwidth was estimated to reach 140 GHz for an array of 20 × 20 unit cells.

Around the same time, Panchenko and colleagues [31] proposed and experimentally demonstrated a plasmonic differential photodetector, which consisted of two back-to-back planar MSM Schottky photodetectors, as depicted in Figure 1.3-4(a). Three concentric spiral-shaped aluminum electrodes were utilized to form two planar intertwined spiral channels on Si, as the active regions of the photodetector. The spiral footprint of the device, 50 μm in diameter, conveniently overlapped the circular profile of conventional optical beams. Arrays of plasmonic nanoantennas were integrated to the active areas of photodetector to selectively distinguish orthogonal states of either linear or circularly polarized light. Two complementary signals were then applied between the two signal electrodes and the common electrode, which produced a differential photocurrent response. Figure 1.3-4 (b) shows an energy band diagram of the photodetector under bias. In this configuration the electrical signal changed sign as the polarization of the incident beam changed. The device was fabricated and experimentally tested over a wavelength range of 500 to 800 nm. Total responsivity of approximately 0.08 A/W was measured for each channel of the photodetector as a function of incident laser power under 3 V bias voltage at the wavelength of 625 nm.

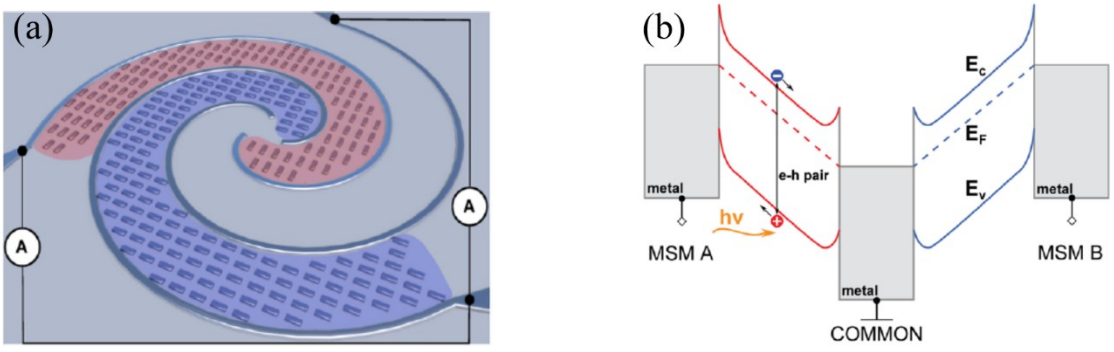

Figure 1.3-4 (a) Schematic representation of a spiral MSM photodetector with rectangular nanoantennas. (b) Energy band diagram under bias. (Adapted from [31].)

Rieger and colleagues [32] incorporated Yagi-Uda nanoantennas into a MSM photodetector. An array of 400 Yagi-Uda nanoantennas was lithographically fabricated on a Si substrate, where Schottky barriers were formed at the antenna-semiconductor interfaces. A bias

voltage was applied through two macroscopic Au electrodes located on either side of the nanoantenna array on the Si substrate. Upon illumination, SPPs were excited at the Schottky interface of the nanoantennas and the semiconductor substrate, which produced a photocurrent between the two electrodes. The measured photocurrent response showed resonances at the wavelengths 1110 nm and 1690 nm. Quantum efficiencies of 5.1% and 3.1% were estimated at 1110 nm and 1690 nm, respectively, representing a fourfold increase over a device lacking the nanoantenna array.

Recently, Gosciniak and colleagues [33] proposed a germanium plasmonic photodetector operating at the telecommunication wavelengths of 1310 nm and 1550 nm. The photodetector was based on the absorption of a long-range dielectric-loaded SPP (LRDSPP) mode, which propagated in a Ge waveguide located on a SiO_2 substrate. The Ge waveguide consisted of a Ge slab, a Ge ridge, and a metal stripe between the slab and the ridge, where mode field confinement is enhanced. Incident light was coupled to the Ge photodetector section through a joint Si waveguide. Two metal pads were placed on either side of the waveguide to apply a bias voltage and collect electron-hole pairs to form the photocurrent. The proposed device achieved bandwidths beyond 100 GHz and responsivities exceeding 1 A/W. The internal quantum efficiency (IQE) was 98% for 5 μm long waveguide operating at 1330 nm. At 1550, however, a 30 μm long waveguide was required to achieve IQE of 84%.

They also investigated [34] a plasmonic photodetector based on IPE by utilizing a metal stripe embedded into a semiconductor/insulator layer. In this configuration a Schottky barrier was formed at each metal-semiconductor interface. SPPs were excited at top and bottom metal-dielectric interfaces, where hot electrons with enough momentum could transition from the metal to the semiconductor, which significantly enhanced the internal quantum efficiency. A high coupling of 90% between the photonic waveguide and the photodetector also enhanced the external quantum efficiency. The authors calculated a responsivity of 0.5 A/W at telecommunication wavelength of 1550 nm and an electrical bandwidth exceeding 100 GHz.

Hybrid photodetectors have shown promise and attracted a lot of attention in recent years. Ding and colleagues [35] reported the design and experimental investigation of a graphene-based plasmonic photodetector on a silicon-on-insulator platform. The bandwidth of the device was shown to exceed 110 GHz at telecommunication wavelengths, while its responsivity reached 360 mA/W. A plasmonic slot waveguide, having a 120 nm gap, was utilized to confine the incident light to the subwavelength graphene regions and enhance the interaction of light with graphene. This narrow plasmonic gap also reduced the drift path for the photogenerated carriers. Contact

metals on either side of the slot waveguide were made of two different metallic stacks (Au/Pd and Au/Ti), resulting in different doping levels of graphene. While photovoltaic effects were the dominant cause of photocurrent generation, the potential difference caused by different doping levels gave rise to efficient separation of photogenerated carriers. Two alterations of the device were fabricated and tested. In the first one the length of the graphene-covered region of the photodetector was 2.7 μm, which resulted in 25 mA/W intrinsic responsivity. Increasing the length to 19 μm, in the second alteration, increased the responsivity to 360 mA/m.

Guo and colleagues [36] demonstrated silicon-graphene hybrid plasmonic photodetectors with reasonable bandwidth and responsivity at 1550 nm and beyond. The photodetector, as shown schematically in Figure 1.3-5, consisted of a silicon-on-insulator strip waveguide, as the input section, and a silicon-graphene hybrid plasmonic waveguide, as the active region. A mode converter based on a lateral taper structure connected the two sections. For the waveguide a 100 nm thick silicon core layer was used, which limited mode field confinement in the vertical direction, leading to enhanced light absorption in graphene. An ultrathin Al_2O_3 insulator layer and a graphene sheet were placed on the silicon core, capped by metal in the middle and on either side of the core. The middle metal part was used as the signal electrode and the two metals on the sides were used as ground electrodes. For the ground electrodes, the graphene sheet was sandwiched between two metal layers, which reduced the metal-graphene contact resistance. The middle metal part was used to confine light in graphene and enhance the responsivity of the photodetector, as well as its bandwidth, while taking advantage of the broad absorption band of graphene. The authors measured a responsivity in the range of 70 mA/W to 0.4 A/W, and a 3 dB bandwidth of 20 GHz to 40 GHz for operation at 2 and 1.55 μm, respectively.

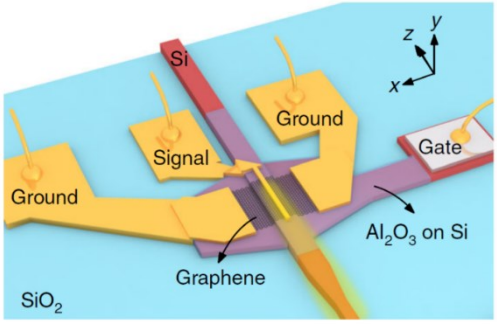

Figure 1.3-5 Schematic configuration of the silicon–graphene hybrid plasmonic waveguide with the signal electrode in the middle and the ground electrodes on both sides. A metal–graphene–metal sandwich structure results. (Adapted from [36].)

1.3.2 Plasmonic modulators

Electro-optic effects refer to changes in the refractive index n of a material, induced by the application of an external electric field, E. These changes modulate the optical modal properties when light propagates in the material. The frequency of the applied electric field must be smaller than the response time of the medium, so that the field appears static within the duration it takes for the optical properties of the medium to change. The refractive index in such medium can be described as $n = n(E)$. Expanding this function in a Taylor's series about $E = 0$, first and second order electro-optic effect coefficients are defined:

$$n(E) = n + a_1 E + \frac{1}{2} a_2 E^2 + \ldots \tag{10}$$

where a_1 and a_2 are the first and second order coefficients, respectively. Higher order terms in Eq. (10) are negligible for practical electric fields. The electro-optic effects are categorized as first or second order effects. The change in n due to the first term in E is called the Pockels effect, whereas the change due to the second term in E^2 is known as the Kerr effect. The Kerr effect is observed in all materials but is rather weak in commonly-used materials and not useful in most integrated optic applications. On the other hand, the Pockels effect is only observed in noncentrosymmetric crystals [37], [38]. Pockels and Kerr effects are the main mechanisms of optical phase modulation which effectively allows control of a carrier signal's phase via an externally applied voltage.

Electro-optic modulators generally can be divided into three classes: single-waveguide modulators, dual-channel waveguide modulators, and modulators employing reflection and diffraction [38].

Single-waveguide modulators can be implemented as a planar channel guide, and are sub-categorized as phase modulators, polarization modulators, intensity modulators, and electro-absorption modulators. In phase modulators the refractive index change in the waveguide, induced by the applied electric field (voltage), causes a change in the phase of the light waves traveling along the waveguide. The change in the refractive index due to an applied voltage V is described as

$$\Delta n_{EO} \propto n_{wg}^3 r_{ij} \frac{V}{2t_{wg}} \tag{11}$$

where Δn_{EO} is the change in refractive index of the medium, n_{wg} is the unperturbed refractive index of the waveguide core, r_{ij} is the Pockels coefficient that depends on crystal structure and

material, V is the applied voltage, and t_{wg} is the thickness of the waveguide. On the other hand, by definition,

$$\Delta n = \frac{\Delta\beta}{k_0} = \frac{\Delta\beta\lambda_0}{2\pi} \tag{12}$$

where β is the propagation constant, k_0 is the wavenumber in free space, and λ_0 is the free-space wavelength. Substituting (11) into (12) the phase change produced by the electric field, $\Delta\varphi_{EO}$, is obtained as

$$\Delta\varphi_{EO} = \Delta\beta L \propto \frac{\pi}{\lambda_0} n_{wg}^3 r_{ij} \frac{VL}{t_{wg}} \tag{13}$$

where L is the length of the modulator in the direction of propagation of light. One of the early works on electro-optic phase modulators was published by Kaminov *et al*. [39], where a 19 μm wide ridged channel waveguide made of $LiNbO_3$ was considered. Two evaporated metal stripes on either side of the waveguide introduced the required electric field into the waveguide. A voltage of 1.2 V produced a phase shift of 1 rad for a modulation power figure of only 20 μW/MHz/rad. In a later work the authors proposed a 5 μm wide Ti-diffused $LiNbO_3$ channel waveguide which reduced the required modulation power to 1.7 μW/MHz/rad.

The next sub-category of single-waveguide modulators is polarization modulators. Electrically-induced changes of refractive index in a noncentrocymmetric crystal lead to phase changes in the direction of the Pockels coefficient. Depending on the type and orientation of the nonlinear crystal, and on the direction of the applied electric field, the phase delay can depend on the polarization direction. A Pockels cell can thus be seen as a voltage-controlled waveplate, and it can be used for modulating the polarization state.

Intensity modulators are the most commonly used, because of their relatively convenient detection method. In an intensity modulator the changes in the refractive index of the waveguide, induced by application of an external electric field, is translated into changes in the intensity of the light beam propagating in the waveguide. Hall *et al*. [40] implemented intensity modulation in a planar waveguide structure by carefully tailoring the refractive index of the waveguide-substrate interface, as well as the dimensions of the waveguide, such that it guides only the lowest order mode in the absence of an external electric field. Applying a voltage of 130 V to the electrodes results in slight refractive index change, which brings the waveguide above cutoff; thus, the propagating mode is cutoff. Campbell *et al*. [41] demonstrated a channel-waveguide

intensity modulator in GaAs. A schematic view of the modulator structure is illustrated in Figure 1.3-6. A maximum modulation depth of 95%, and electrical bandwidth of 150 MHz were demonstrated with the driving power less than 300 μW/MHz.

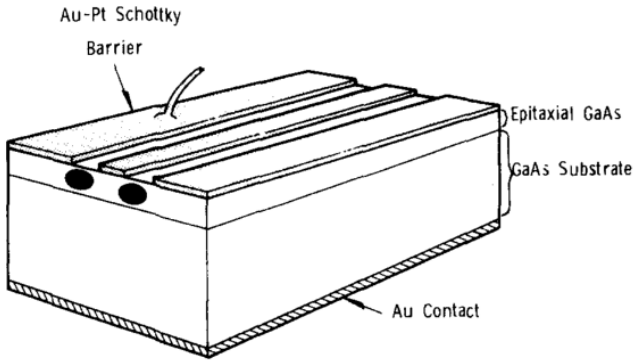

Figure 1.3-6 Schematic diagram of the electro-optic directional coupler switch. (Adapted from [41].)

Another well-established method of converting refractive index changes into intensity modulation is based on Mach-Zehnder interferometer concept [42](Figure 1.3-7). Their operation principle is that once the input signal is split between the two arms of the interferometer, by applying a differential external voltage across the arms, a differential refractive index change is induced in the two arm waveguides, which in turn results in a phase difference in the traveling waves. Once the two waves propagating along the two arms recombine, depending on their relative phase difference, constructive or destructive interference occurs. Thus, the output intensity is controlled by this applied voltage, though the relationship is not linear but rather sinusoidal. If the voltage is adjusted such that the two waves are out of phase by π the waves cancel each other, and the output intensity would, ideally, be zero. On the other hand, if the two waves are in phase, constructive interference at the output results in reconstruction of the input signal.

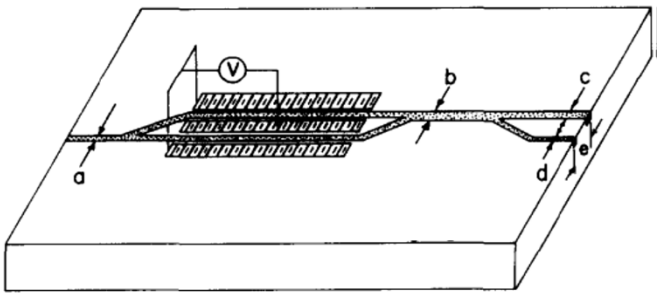

Figure 1.3-7 Waveguide switch/modulator configuration. (Adapted from [42].)

Modulation schemes reviewed thus far, are all based on the Pockels effect and changes in refractive index. Electro-absorption modulators, however, exploit the Franz-Keldysh effect,

which is a shift in the absorption edge of a semiconductor when a strong electric field is applied. In direct bandgap semiconductors with a steep absorption edge, very large changes in the absorption of wavelengths near the band edge are produced. Reinhart [43] proposed such modulator by considering a GaAlAs heterostructure waveguide, where the Al concentration was adjusted to obtain optimum performance at the wavelength of interest, $\lambda_0 = 900$ nm. For an applied voltage of -8 V a change in transmission of a factor of 100 was demonstrated. 90% modulation was achieved for a required power on the order of 0.1 mW/MHz.

Another class of electro-optic modulators is the dual-channel waveguide modulator. When two parallel optical waveguides are sufficiently close to each other light can be coupled from one waveguide to the other. If a modulating signal voltage is applied to the waveguides, according to the Pockels effect, a small change is produced in the refractive indices of the two waveguides, which results in a difference in their phase constants $\Delta\beta$. Power transmission from one waveguide to the other is a function of $\Delta\beta$. It can be shown that 100% coupling is enabled when $\Delta\beta = 0$, while zero coupling, or 100% modulation, occurs when $\Delta\beta = \pi\sqrt{3}/L_0$, L_0 is the coupling distance required for full transfer of power between the two waveguides when $\Delta\beta = 0$. Although the concept of using a dual-channel directional coupler as an electro-optic modulator was proposed as early as 1969 [44], the first successful realization of such device was reported in 1975 by Campbell *et al*. [45], and another, around the same time, by Papuchon *et al*.[46]. A schematic diagram of the coupler configuration in [46] is shown in Figure 1.3-8.

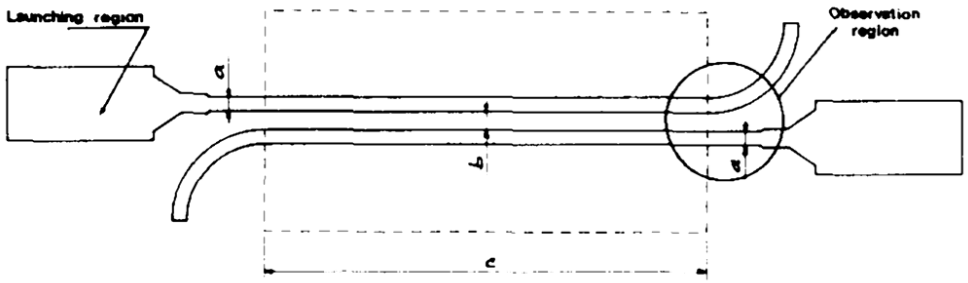

Figure 1.3-8 Coupler configuration- launching and observation regions are also shown. a = 2 μm, b = 3 μm, c = 3 mm. (Adapted from [46].)

Several modulators and switches have been demonstrated utilizing the concept of electro-optic control of either diffraction or reflection in optical waveguides. Diffraction modulators are generally based on the Bragg effect [38], [47]. A pair of comb-like interlocked electrodes are laid down on a planar waveguide of electro-optic material, as shown in Figure 1.3-9. Once a voltage is applied to the electrodes, the refractive index of the waveguide underneath changes such that a

Bragg grating pattern is formed in the waveguide. Light diffracts upon incidence on the Bragg grating pattern, causing a change in the direction of propagation of light. Similarly, the linear electro-optic effect may be utilized to reduce the index of refraction in a layer, such that total internal reflection of an optical beam happens upon application of voltage to the electro-optic region [48].

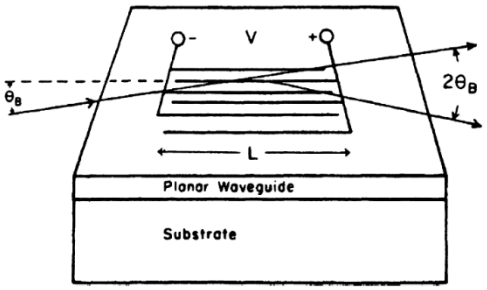

Figure 1.3-9 Schematic of a Bragg-effect electro-optic modulator. (Adapted from [38].)

One of the early studies on Si-based electro-optic modulators was done by Dionne and colleagues [49]. Their modulator consisted of a four-layer metal-oxide-Si-metal waveguide, where subwavelength slits, etched in the metal cladding layers, coupled light into and out of the modulator. In this geometry the input slit acts as the source in a MOS structure and is approximately 4 μm long, while the output slit acts as the drain, and the oxide layer as a channel connecting the source to the drain. The source-drain separation was varied from approximately 1 to 8 μm. The device was examined numerically and experimentally, which yielded switching energies of the order of femtojoule and potential modulation frequencies in the order of gigahertz, once illuminated by a Gaussian beam of λ= 1550 nm or λ= 685 nm. Modulation ratios greater than 10 dB were observed in simulations for drain-source separation of 2 μm at λ = 1550 nm.

Fabry-Perot structures have been of much interest as electro-optic modulators. Cai and colleagues [50] performed a theoretical study of electrically-activated plasmonic modulators, by numerically modeling three Fabry-Perot resonance structures. The first structure, illustrated in Figure 1.3-10(a), consisted of an infinitely long nanoslit in a silver slab on a transparent substrate. The nanoslit region was filled with an active dielectric material. They considered top illumination by a plane wave at λ_0 = 850 nm, polarized perpendicular to the length of the nanoslit. It was observed that for a given slit width, w, the number of resonances excited increased with the thickness of the slab. For a slit of w = 80 nm on a slab thick enough to support four resonances, time-averaged energy density in the nanoslit was eight times greater than that of the plane wave. The transmittance through the nanoslit was also three times larger than the incident light. The authors then added periodic corrugations in the silver slab on either side of the nanoslit, as

depicted in Figure 1.3-10(b), and optimized the dimensions of the gratings to give the maximum power transmittance from the slit resonator. In this case the time-averaged energy density inside the nanoslit was almost 160 times that of the incident wave, while the transmittance was enhanced 20 times compared to the previous case. The same structure was then proposed as a plasmonic modulator by assuming that the refractive index of the material inside the nanoslit is controlled electrically by applying an external voltage. The modulation depth of the device was modeled as a function of the changes in the real and imaginary parts of the refractive index within the slit. The authors observed that increasing the thickness of the silver slab led to better modulation performance for a given change in the refractive index, as it increases the interaction length between the SPPs and the active material in the slit. For example, a 3 dB modulation depth was calculated for a 830 nm thick slab with 2% change in the imaginary part of the refractive index, while the transmittance was determined to be 60 times larger than the input power. The authors also considered a waveguide-coupled plasmonic modulator device (Figure 1.3-10(c)). A metal-dielectric-metal waveguide was formed by considering a dielectric layer sandwiched between two silver slabs, where a resonant cavity was established by utilizing a pair of rectangular stubs separated by a length L in one of the silver slabs. By optimizing the dimensions of the resonant cavity, a 3 dB modulation depth was obtained by introducing either a 0.4% change in the imaginary part of the refractive index of the material within the cavity, or a 0.8% change in the real part of its refractive index. This strong modulation was a result of significant resonance enhancement of SPPs in the Fabry-Perot resonator formed inside the cavity. Given the small capacitance (~ fF/μm) due to compact cavity length (< 1 μm), and the estimated swing voltage of ~ 1V, a power consumption on the order of 1 fJ/bit was predicted in the device.

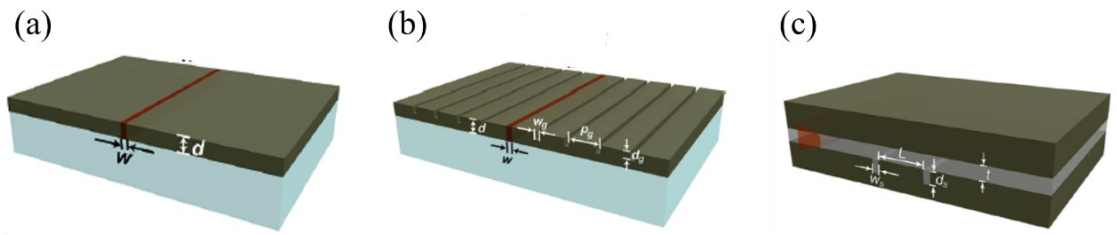

Figure 1.3-10 Schematic view of (a) a nanoslit resonator in a silver slab, (b) nanoslit resonator surrounded by periodic corrugations, (c) metal-dielectric-metal waveguide with a pair of stubs in one of the metal slabs (Adapted from [50].)

Melikyan and colleagues [51] proposed a surface plasmon polariton absorption modulator (SPPAM). The device consisted of an input coupler silicon nanowire which coupled incident light into an active plasmonic section, where modulation was done, and an output coupler silicon nanowire. The plasmonic section consisted of a pair of silver slabs which sandwiched an insulator/metal-oxide layer. Here, SiO_2 was chosen as the insulator and ITO as the metal oxide

layer. The silver slabs served as waveguides for SPPs, as well as electrodes for application of a voltage across the metal-oxide region. The free carrier density was changed in the ITO layer upon application of the external electric field (voltage), which resulted in a change of the complex permittivity, thus, the absorption of the medium. For example, applying an electric field of 10 V over 30 nm changed the free carrier density from $N_0 = 9.25 \times 10^{26}$ m^{-3} to 9.34×10^{26} m^{-3}, which changed the permittivity from ε = -1.667 + 0.824i to -1.721 + 0.832i. Thus, the absorption coefficient was decreased from $\alpha = 2.19$ μm^{-1} to 2.07 μm^{-1}, which, in turn, modulated the output power. The device operated at the wavelength of 1.55 μm, and was expected to reach an electrical bandwidth in the THz-region (being RC time constant limited). For an optimized 2 μm long modulator, a 1 dB extinction ratio and a total plasmonic loss in the range of 18 dB was predicted. The authors also fabricated a simplified prototype and performed a proof-of-principle experiment. The rather large distance between the two electrodes (61 μm) in their prototype, however, resulted in a weak electric field across the metal-oxide region, and therefore, weak modulation.

A few years later Melikyan and colleagues [52] proposed and experimentally demonstrated a compact, broadband, high-speed plasmonic phase modulator (PPM) that could bridge between optical and RF components in a communication link. Their PMM consisted of two metal tapers at the ends of a nanometer-scale metallic slot waveguide filled with a nonlinear organic material. While photonic-to-plasmonic mode conversion (and vice versa) was performed in the metal tapers, the phase modulation section was the slot waveguide between the two tapers. The Pockels effect was exploited by applying an electric field which changes the refractive index of the nonlinear polymer, thus modulating the phase of the SPPs. Due to enhanced electric fields in the nonlinear medium, and the highly confined nature of SPPs, shorter phase modulation lengths were possible. Also, the weakly dispersive nature of the nonlinear polymer, along with the optically broadband characteristics of metallic slot waveguides and metallic tapers yielded a modulator which performed well over the wavelength range of 1480 nm to 1600 nm. Additionally, the small *RC* time constant of the device, as well as the instantaneous Pockels effect, led to a flat frequency response up to at least 65 GHz and high-speed operation at 40 Gbit/s.

Haffner and colleagues [53] proposed and investigated an all-plasmonic Mach-Zehnder modulator (MZM). The device consisted of three sections. First, a photonic plasmonic interference section, which converted incident light from a silicon waveguide to SPPs and split it into the two arms of the MZM. The second section was phase shifter slot waveguides, which were formed between gold contact pads and a gold island, and were filled with DLD-164, a highly nonlinear material. The final section was another photonic-plasmonic interface which converted

the phase modulated SPPs into an amplitude modulated signal, guided to an output silicon waveguide. The total length of this MZM is 10 μm. the authors then tested the device's high-speed performance by feeding a two-level electrical signal to the device and detecting its output using a coherent receiver on the other side. High speed operation of 54 Gbit/s and 72 Gbit/s was observed with BERs of 7×10^{-5} and 3×10^{-3}, respectively. The electrical energy consumption was estimated to be ~25 fJ/bit for up to 54 Gbit/s operation, when driven by a ± 3 V external voltage. The electrical bandwidth of the device was governed by its RC time constant and turned out to be ~1.1 THz.

Hybrid plasmonic modulators have also attracted a lot of attention. Ansell and colleagues [54] fabricated and studied three types of hybrid graphene-plasmonic waveguide modulators (HGPWM) for telecom applications. They obtained modulation depth values greater than 0.03 dB/μm induced by gating voltages less than 10 V, for an active device area of 10 mm^2. Their simplest HGPWM modulator consisted of a metal-dielectric interface, where SSPs propagated adjacent to a graphene monolayer. Another configuration considered was a corrugated plasmonic waveguide, along with a graphene monolayer. The most efficient modulator the authors proposed was based on wedge SPPs propagating along the edge of planar section of the waveguide, which produced strong electrical field confinement along a graphene monolayer. The waveguide was 100 μm long. Incident light was coupled into plasmon modes by non-transparent gratings. In the waveguide section, voltage swings shifted the Fermi energy in graphene, which in turn changed graphene's conductivity, and resulted in modulation of SPPs propagating along the graphene monolayer. Output was then decoupled to light using another set of gratings. A schematic diagram of the experiment, as well as sketches of the three plasmonic modes are shown in Figure 1.3-11.

Olivieri and colleagues [55] proposed and tested a plasmonic intensity modulator based on a MOS structure. The modulator consisted of Au nanogratings on a plasmonic Au surface on a film of HfO_2, deposited on a doped Si substrate. A 2D numerical study was performed to optimize the nanograting design, as well as the thickness of the oxide layer. Signal and ohmic contacts were established, such that application of a bias voltage drove the MOS structure into accumulation/depletion, causing an electro-optic change in the material properties therein. Once illuminated from top, the optical input was partially coupled to the SPPs along the Au/Oxide/Si interface through the nanogratings, and partially reflected and eventually collected as a photocurrent. The electrical bandwidth of this device was determined to be about 500 MHz, with

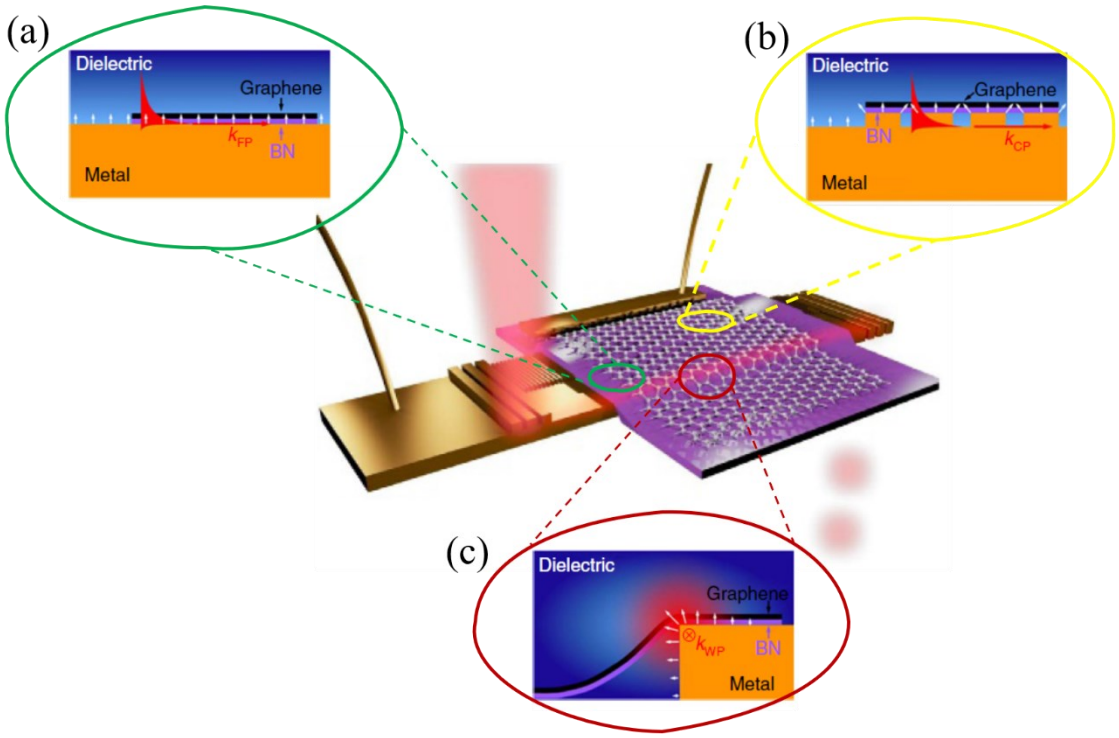

Figure 1.3-11 3D rendering of the experiment with the wedge plasmon mode central image). 3 types of plasmonic structures under investigation are sketched: (a) Flat, (b) corrugated, and (c) wedge. Different plasmon modes can be excited depending on the location of the incident beam on the device. (Adapted from [54].)

modulation depth of 3-6%. Details of the fabrication process of these modulators were reported in [56].

Ayata and colleagues [57] reported a high-speed, ultracompact, energy efficient modulator comprising a single gold layer on a glass substrate. This all plasmonic modulator was based on the Mach-Zehnder interferometer (MZI) architecture, where the interferometer arm slots were filled with a nonlinear optical material. A multi-core fiber (MCF) was used as the optical input/output means. A vertical two-step-etch grating coupler was employed to efficiently couple the incident light into the chip with high directionality. An integrated polarization rotator was designed to rotate the polarization of the incident light from *p* to *s* in order to obtain maximum plasmonic efficiency within the arm slots. The grating coupler also served as the electrical signal input, which modulated SPPs propagating along the arms of the MZI. The optical carrier was then coupled out into another core of the MCF via the output vertical grating coupler. The transmitted signal was then amplified and sampled by a coherent receiver. The authors observed high modulation speeds of up to 116 Gbit/s for a PAM4 signal, with a bit-error-rate of 3.8×10^{-3}, upon detection. The energy consumption of the modulator was estimated to be 30 fJ/bit for the same signal and driving voltage of ±2.8 V.

Wang and colleagues [58] proposed, fabricated, and tested monolithically integrated lithium niobate (LN – $LiNbO_3$) electro-optic modulators. For this purpose, LN thin films were nanostructured into waveguides, MZIs, and racetrack (ring) resonators, and subsequently, embedded in SiO_2, to further enhance electric field confinement. This allowed smaller distance between LN active regions and RF electrodes, which led to improved overlap between optical and electric fields. Optical waveguides were sandwiched between signal and ground electrodes with a gap of 3.5 μm. The authors measured 3-dB bandwidths of 30 GHz and 15 GHz, and extinction ratios of 3 dB and 8 dB for racetrack and MZI modulators, respectively. The on-chip insertion loss of the racetrack and MZI modulators was 1.5 and 2dB, respectively, for total waveguide lengths of 7 and 5 mm. Power consumptions of 0.24 and 1.6 pJ/bit were also measured for racetrack and MZI modulators, respectively.

Later on, Wang and colleagues [59] introduced another lithium niobate modulator, based on the MZI configuration, which operated at data rates up to 210 Gbit/s with a BER of 1.5×10^{-2}. The MZI arms were positioned in the gaps of a ground-signal-ground coplanar microwave strip line. Thus, by applying voltage to the micro-strip lines optical phase delay was induced in one arm of the MZI through Pockels effect. The operating voltage of this modulator was as low as 1.4 V for a 20 mm long device. On-chip integrated nature of the device also lowered optical loss to less than 0.5 dB. The authors measured an extinction ratio of 30 dB. The electrical energy dissipation for 8-level amplitude-shift-keying modulation was expected to be 14 fJ/bit.

Burla and colleagues [60] recently reported a plasmonic Mach-Zehnder phase modulator suitable for analog applications at telecom wavelengths. The device consisted of two slot waveguides of 20 μm long by 120 nm wide, filled with a nonlinear organic material, forming the two arms of the MZI. Application of a voltage to the MZI caused a change in the refractive index of the organic material, based on the Pockels effect, which in turn caused phase modulation in the SPPs propagating along the arms of the MZI. A 450 nm wide and 220 nm thick silicon stripe waveguide, connected to a 1 μm long linear taper were used to input light, and convert it to SPPs, respectively. An identical taper was used to decouple SPPs into the output light. The authors experimentally demonstrated a flat frequency response from 75 MHz up to 500 GHz, with an extinction ratio of ~25 dB.

Koch and colleagues [61] reported a plasmonic modulator array comprising four densely-arranged plasmonic phase modulators, as shown in Figure 1.3-12. Each modulator consisted of input and output grating couplers, for direct conversion of fiber to plasmonic modes and vice versa. The plasmonic phase modulator section was a slot waveguide of area 9 μm × 5.5 μm,

located between the input and output grating couplers, and filled with a nonlinear organic electro-optic material. The modulators were directly interfaced with an optical fiber array, which minimized conversion losses. Thus, the total footprint of this array was 90 µm × 5.5 µm, which provided total operation speed of 0.8 Tbit/s. Optical and electrical crosstalk between adjacent modulator devices were found to be negligible. The authors experimentally demonstrated operation at 100 GBd for each modulator in the array, at a single wavelength (space division multiplexing) and at multiple wavelengths (wavelength division multiplexing).

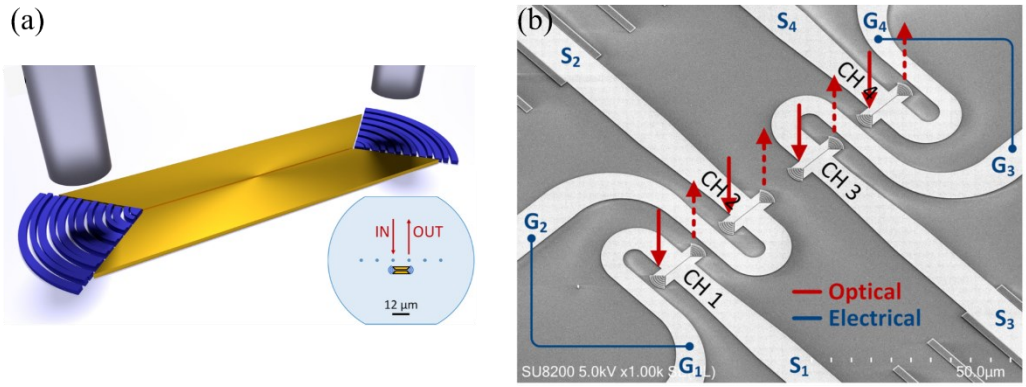

Figure 1.3-12 SEM image of the modulator array and arrangement of the parallel plasmonic modulators with the ground-signal contacts. Optical carriers (solid red) are coupled in by every second fiber core and the modulated light (dashed red) is coupled out through the remaining cores. The electrical signal (blue) is applied to the modulators by a multi-channel RF probe. The individual signal lines for channels 1–4 and the corresponding ground lines are marked by S_i and G_i, respectively.

Quite recently, Baeuerle and colleagues [62] proposed and experimentally demonstrated a plasmonic Mach-Zehnder modulator, which employed a dual-drive electrode system. A power multiplexer was used to produce the differential voltage required to drive the Mach-Zehnder modulator in a push-pull operation. Light was coupled via grating couplers to a pair silicon-photonics waveguides, which directed light into the plasmonic phase shifters embedded in the Mach-Zehnder arms, where the electro-optic modulation was performed. The organic electro-optic material HD-BB-OH/YLD124 was used here. The modulator's active length was 15 µm, and the slot width was 100 nm. A thermo-optic phase shifter tuned the phase relation of the two arms, hence, the modulator's operating point. The authors measured an electro-optic bandwidth of 108 GHz, which allowed for C-band transmission of 120 Gb/s signals through a single mode fiber. On the other hand, using differential drive voltage doubled the voltage swing in comparison with a conventional ground-signal-ground configuration, which led to a small drive voltage of 178 mV.

1.4. Thesis scope and outline

This thesis investigates the application of surface plasmon polaritons in optoelectronic devices, with foci on photodetectors and modulators. It has been demonstrated, theoretically and experimentally, that SPPs have the potential to significantly enhance the performance of photodetectors and modulators, by confining light into subwavelength regions, and enhancing localized electric fields. Novel designs are proposed, numerically modelled, or experimentally evaluated. This thesis comprises a collection of scientific articles, as well as some chapters in essay format. The remainder of this document is organized as follows:

Chapter 2 briefly discusses the numerical methods used, and examines their validity, by comparing results obtained to those produced using alternative numerical methods, as well as experimental results. The finite difference time domain (FDTD) method has been used for most of the numerical simulations in this thesis. Chapter 2 also covers details of simulations done for two projects in collaboration with Maryam Al-Shehab [63] and Giulio Vampa [64].

Chapter 3 explores two novel plasmonic photodetector designs. The photodetectors exploit nanoantennas to confine incident light to subwavelength detection regions. Numerical models demonstrate significant performance improvements over conventional photodetectors, in terms of speed and responsivity. The electrical performance of the photodetectors was also modelled.

Chapter 4 describes the fabrication of an improved version of the plasmonic reflection modulator previously proposed in [55] and [56]. The present device is capable of modulating in transmission, as well as in reflection. Additionally, the present device is expected to demonstrate a significant bandwidth improvement over the older version ([55], [56]), due to the reduced effective capacitance of the devices. Nanofabrication methods are discussed in detail, while limitations are pointed out.

Finally, in chapter 5 conclusions and suggestions for future work are presented.

CHAPTER 2. NANOANTENNA MODELLING

This chapter reviews the numerical model of nanoantennas used throughout this thesis, and in related publications. In addition to the methods, and results related to works in Appendix A [63] and Appendix B [64], the accuracy and reliability of our model, in comparison with other numerical methods, as well as experimental results, are considered.

2.1. Methods

Optical modeling of nanoantennas was achieved numerically using the commercial software Lumerical, which employs the finite difference time domain (FDTD) method. Maxwell's curl equations were numerically solved in three dimensions (assuming non-magnetic materials) to model the optical performance of the device under study:

$$\frac{\partial \vec{D}}{\partial t} = \nabla \times \vec{H} \tag{14}$$

$$\vec{D} = \varepsilon_0 \varepsilon_r \vec{E} \tag{15}$$

$$\frac{\partial \vec{H}}{\partial t} = -\frac{1}{\mu_0} \nabla \times \vec{E} \tag{16}$$

where $\vec{E}$ and $\vec{H}$ are the electric and magnetic fields, respectively, $\vec{D}$ is the electrical displacement, and ε_r is the complex relative permittivity.

Finite arrays of nanoantennas were modelled as infinite arrays, extended along the *x* and *y* axes [65]. This was achieved by considering one array element (unit cell) bounded on all lateral walls by periodic boundary conditions. Symmetry was also exploited by using symmetric and antisymmetric walls in order to reduce computational requirements. Perfectly matched layers (PMLs) were applied on the top and bottom walls to avoid reflections. To do so, the position of the PML boundaries was adjusted to ensure that the electric field distribution in the near-field region of the nanoantennas was independent of the location of the PMLs. The location of the lateral walls in a unit cell, or the pitch dimensions *p* and *q*, were adjusted to avoid coupling of adjacent nanoantennas, by, again, observing the electric field distributions. Symmetry was applied by utilizing proper boundary conditions whenever the geometry of the structure under investigation was suitable. This significantly reduced the required time and computational resources.

Meshing is an essential part of any numerical method, which ensures the accuracy of the results obtained. The smaller the mesh size, the higher the accuracy, given sufficient computational resources. A smaller mesh, however, means substantially longer simulation times, and more memory requirements. In order to achieve a balance between accuracy, time, and memory requirements, a medium-sized automatically-generated background mesh was selected, which was coarser in dielectric areas away from metallic structures, sources, or material boundaries, and finer where electric field discontinuities were expected, such as along material boundaries, or where the fields changed rapidly, such as near sharp metallic corners. In order to attain higher accuracy, manual meshing constraints were added to the automatically-generated background mesh in areas around nanoantennas. Figure 2.1-1 shows a top view of the simulation domain of a unit cell, where a nanomonopole is located at the centre. In Figure 2.1-1(a) the orange boundary area around the unit cell depicts PML boundary condition. Inside, the cross stripes show the area with finer mesh than everywhere else in the domain. The red circle encloses the nanomonopole (outlined by black dashed lines) at the centre, surrounded by an area of very fine fixed mesh (outlined by blue dashed lines) which extends 10 nm from perimeter of the nanomonopole. Enlarged views are shown in Figure 2.1-1(b) and (c). As will be discussed in Section 2.2, in most cases, a mesh of $1 \times 1 \times 1$ nm^3 is sufficiently small to obtain reliable results.

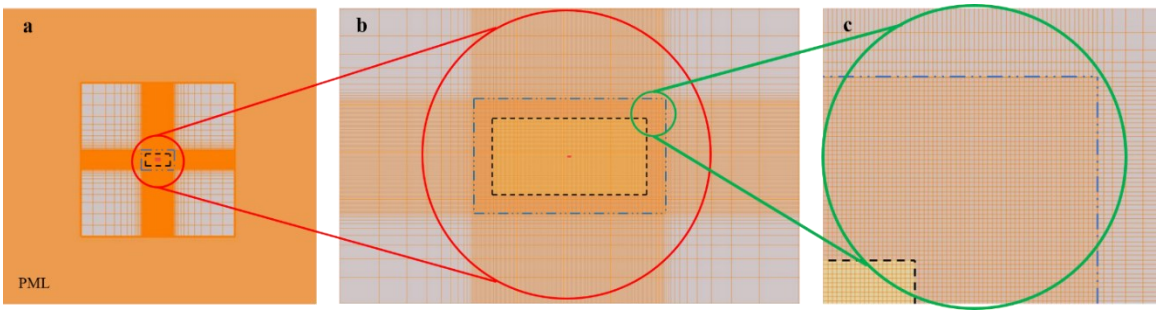

Figure 2.1-1 (a) Top view of a unit cell with a nanomonopole at its centre. The black dashed line outlines the nanomonopole, where the blue dashed line indicates the manually-meshed region. (b) Depicts an enlarged view of components in the red circle in (a). (c) Shows an enlarged view of the green circle in (b): top right corner of the nanomonopole.

Refractive index (n, k) data as a function of wavelength, for dispersive materials, were acquired from Palik's Handbook of Optical Constants for Solids [66] or from measurements and were incorporated using multi-coefficient material models. The arrays were illuminated perpendicularly by a plane wave through the substrate. The significantly smaller size of nanoantennas, in comparison with the beam size available experimentally, justifies using a plane wave in our numerical model, where a Gaussian laser beam illuminates the nanoantenna arrays during experimental tests. The plane wave was polarized along the length of the nanoantennas.

The dimensions and geometry of the nanoantennas were selected to maximize the absorptance at the wavelength of interest. For this purpose, the nanoantennas' length, l, width, w, thickness, t, and (applicable for dipole antennas) gap, g, were chosen such that the absorptance of an infinite array is maximized at the wavelength of interest [65], while considering the feasibility of fabrication as another governing criteria. For work in [67], [68] the thickness of the InGaAs layer, as the detection region, was also adjusted to maximize the absorptance at $\lambda_0 = 1550$ nm and yield maximum absorbed power in InGaAs in order to optimize photo-detection therein. The absorptance, A, was obtained as

$$A = 1 - T - R \tag{17}$$

where T and R are the transmittance and reflectance of an infinite array of nanoantennas, respectively. T and R were computed as a function of frequency at reference planes located in the far-field of the nano-antennas as:

$$T(f) = \int_S \mathrm{Re}\left\{P^m(f)\right\} \cdot dS \Big/ \int_S \mathrm{Re}\left\{P^s(f)\right\} \cdot dS \tag{18}$$

where P^m and P^s are Poynting vectors at reference planes where the transmittance (or reflectance) is computed, and at the location of source, respectively. f is the frequency, and S is the area of the reference plane, which covers the entire cross section of a unit cell. The reflectance, R, is calculated similarly, with data collected at a reference plane positioned at the location of the reflectance.

Field enhancement and electron-hole pair generation in the semiconductor region were then optimized by varying the dimensions of the InGaAs region, and calculating the power density absorbed therein. The integral of the absorbed power density over the InGaAs volume gives the absorptance in InGaAs. The absorbed power is given by

$$P_{abs} = -\frac{1}{2}\omega |E|^2 \, imag(\varepsilon) \tag{19}$$

where ω is the angular frequency, E is the electric field, and ε is the optical permittivity of InGaAs.

2.2. Accuracy of the Numerical Model

Resonant wavelength predictions obtained using our numerical model were compared to the results obtained using the finite element method (FEM) and the transmission line (TL) model, as well as experimental results [69], in order to verify the validity and accuracy of the employed numerical model.

An infinite array of nanomonopoles laid over a film of SiO_2 on a Si substrate was modeled, by periodically repeating a unit cell, as shown in Fig. 2.2.1(a), along the lateral dimensions. The graphene monolayer was not considered in this model since it is optically noninvasive. A plane wave polarized along the length of the nanomonopoles was placed in the far field region to illuminate the system from the top at normal incidence with a 1 V/m electric field magnitude. The dimensions of the nanomonopoles were critical in determining their respective resonance wavelength. The length of the nanomonopoles were varied from 60 to 120 nm. The width and thickness of the nanomonopoles, however, were fixed to 40 and 22 nm, respectively. Pitch dimensions were also fixed to 500 × 500 nm^2. The absorptance, A, was calculated following Eq. 4.1, where data for transmittance, T, and reflectance, R, were collected at the location of their respective reference planes. Figure 2.2.2 shows a comparison between results obtained from our model and those in [69]. Our results are in good agreement with experiments and these other numerical results.

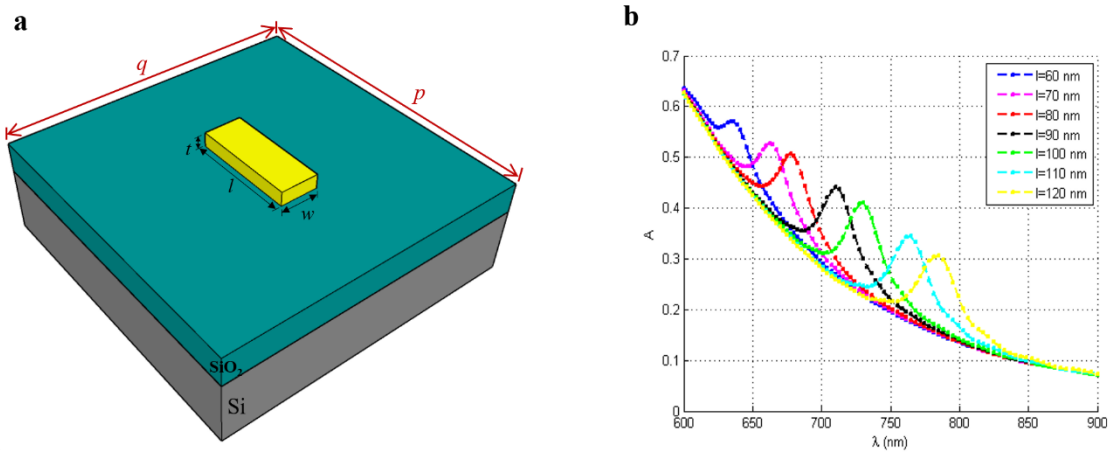

Figure 2.2-1. (a) Schematic of a unit cell of an infinite array of nanomonopoles. (b) Absorptance, A, as a function of wavelength, λ, for nanomonopoles of various lengths.

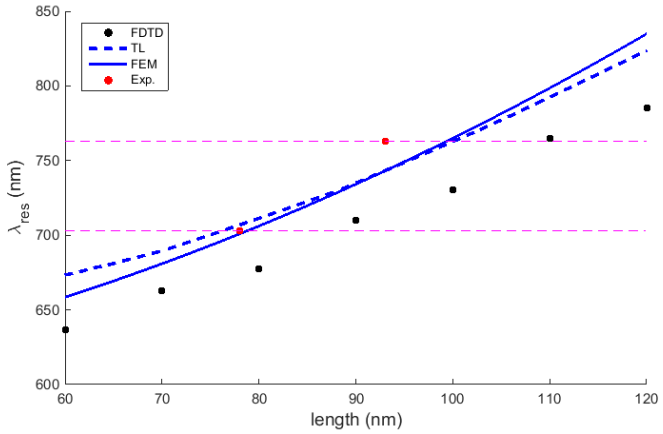

Figure 2.2-2. A comparison between results obtained from our model and those in [69]. Black dots were obtained using our numerical model. Blue dashed- and solid-lines are results of the transmission line model and finite element method, respectively. Red dots show experimental data.

A convergence study was also performed by decreasing size of the mesh and monitoring the position of the peak absorptance, A. Figure 2.2-3 give the absorptance, A, of an infinite array of nanomonopoles, where the mesh dimensions were successively halved from $2 \times 2 \times 2$ nm^3 to $0.25 \times 0.25 \times 0.25$ nm^3. As a result, the absorptance peak red-shifted monotonically, indicating convergence with decreasing mesh. These results also show that although there is a small red shift, the resonance peaks stay within a small enough vicinity. Hence, the resonant wavelength can be reliably determined given a prescribed expected error, as shown in Figure 2.2.2(b).

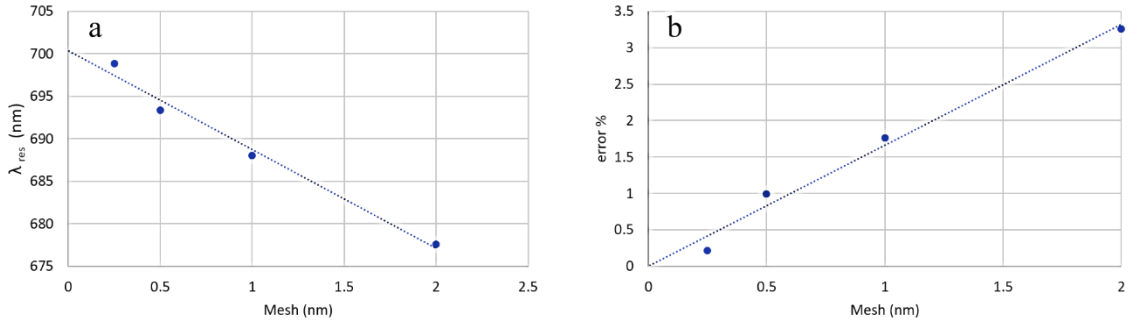

Figure 2.2-3 (a) resonant wavelength λ_{res} of an array of nanomonopoles as a function of mesh dimensions. Numerical results were extrapolated to zero using a linear fit. (b) The error % was calculated for each data point with respect to the extrapolated value at zero.

2.3. Nanoantennas for Enhanced Raman Scattering

The numerical model for the work published in [63] was implemented by following the method described in 2.1. However, due to difficulties in numerically modelling the graphene monolayer, and its expected minimal impact on the optical performance of the nanoantennas, this layer was disregarded in the numerical model. Also, due to fabrication limitations, the experimentally-tested structures were slightly different than those modelled numerically. These differences and their

impact on results are discussed in [63]. The sketch in Figure 2.2.1(a) depicts the modelled structure.

Nanomonopoles on graphene were modelled as infinite array of nanomonopoles. A plane-wave polarized along the length of the nanomonopoles illuminated the array from top at normal incidence, with a 1 V/m field magnitude. Mesh dimensions in the neighborhood of the nanoantennas were 0.5×0.5×0.5 nm^3, which were small enough to ensure convergence of electromagnetic waves, especially around sharp corners of nanomonopoles. The reflectance, *R*, was determined in a reference plane 2.5 µm above the Au-graphene interface, as shown in Figure 2.3-1. The reflectance acquired in the absence of nanomonopoles is also plotted as a reference. The resonance wavelengths red-shift by increasing the length, and blue-shift by increasing the width of the nanoantennas, as expected [65]. The resonance wavelengths obtained from this model were compared to the experimental results , as shown in Figure 2.3-1 (c) and (d) for various nanoantenna length and widths, respectively.

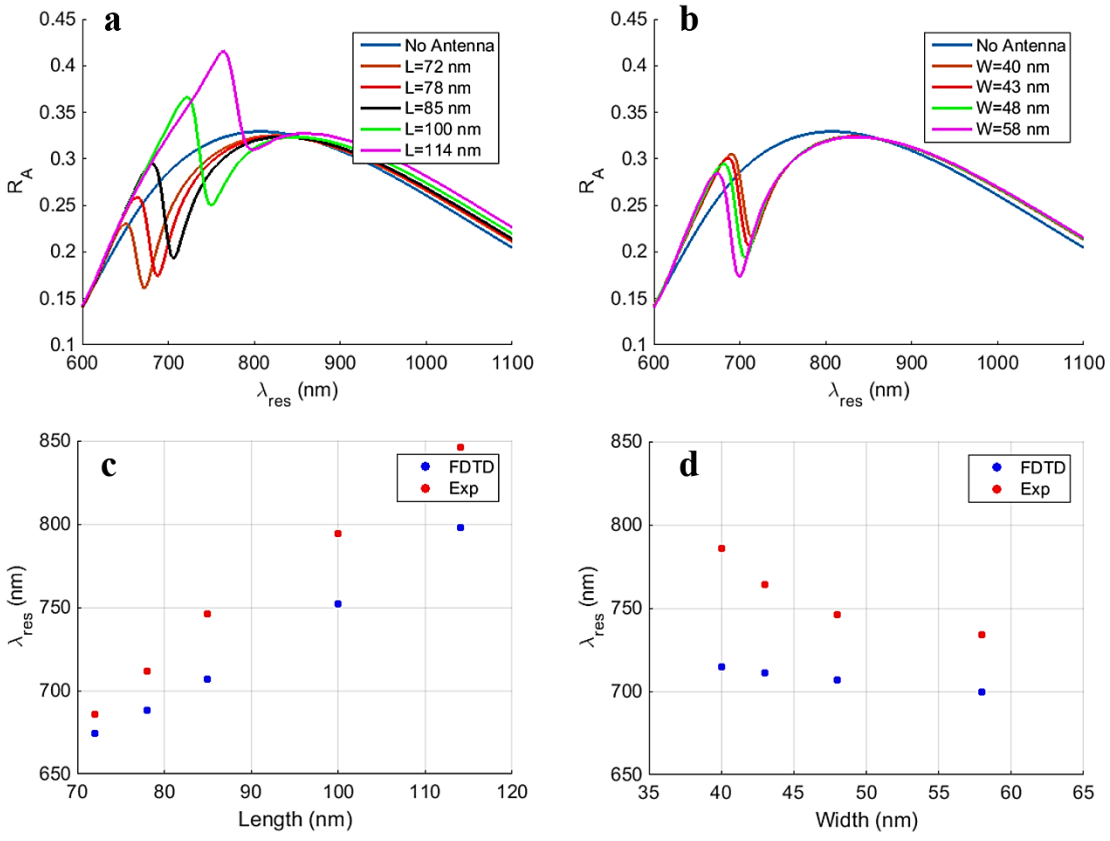

Figure 2.3-1 reflectance as a function of wavelength for various nanoantenna lengths (a), and widths (b). Comparison of results obtained from FDTD model with experimental results for different lengths (c) and widths (d).

2.4. Nanoantennas for High Harmonic Generation

The method described in 2.1 was applied to find the best nanomonopole design to resonate at λ_0 = 2100 nm, in support of work on plasmonic high-harmonic generation, while considering fabrication limitations. The absorptance, A, was determined, according to Eq. 1.4, for various nanomonopole lengths and widths, as shown in Figure 2.4-1. The thickness was fixed to t = 20 nm. Pitch dimensions were also varied to ensure the nanomonopoles are sufficiently distanced to avoid cross-coupling. The optimum design was found to be l = 205 nm, w = 80 nm, and t = 20 nm.

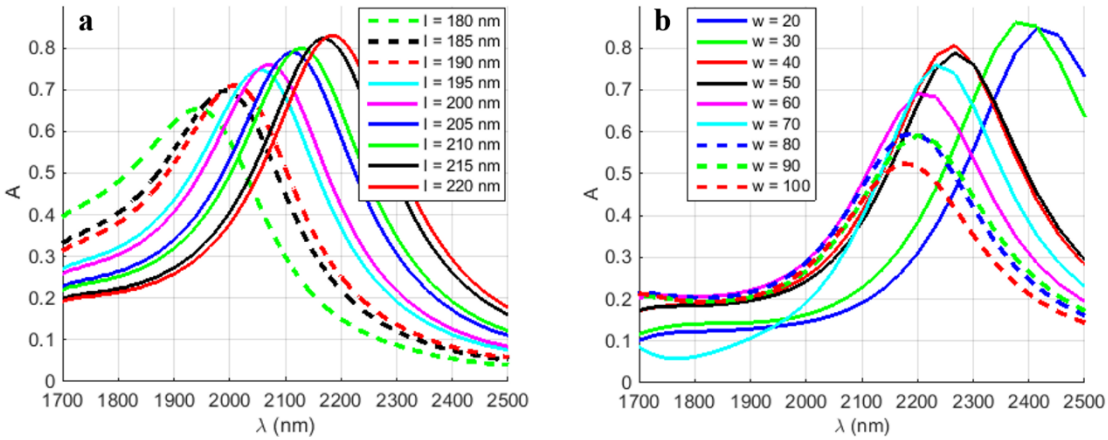

Figure 2.4-1 (a) Absorptance as a function of wavelength for various nanomonopole lengths, for w = 80 nm and t =20 nm. (b) The nanomonopole width was varied for l = 220 nm and t = 20 nm.

The distribution of the electric field intensity in Si, at different distances below the nanomonopoles, are shown in Figure 2.4-2. The incident beam was polarized along the length of the nanomonopoles. As expected, nanoantennas significantly enhanced and localized the electric fields in their proximity, which resulted in HHG in the Si [60].

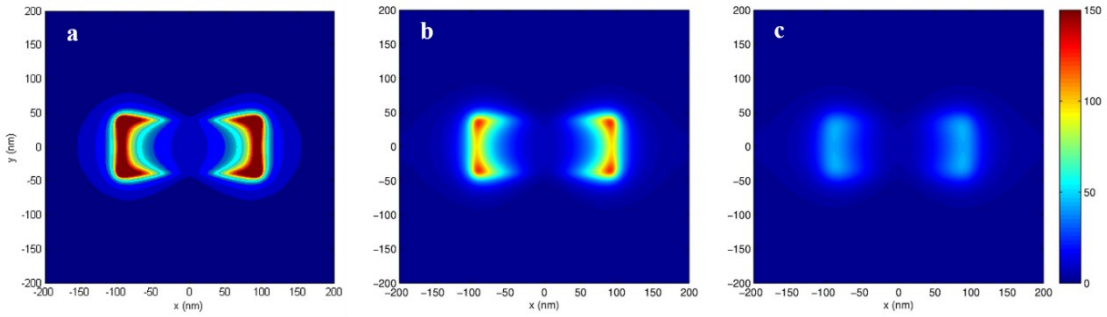

Figure 2.4-2 Electric field distribution along a cut (a) 5 nm, (b) 10 nm, and (c) 20 nm below the nanomonopole-Si interface.

CHAPTER 3. DESIGN OF PHOTODETECTORS

3.1. Summary

Plasmonic nanoantenna structures, combined with a metal-semiconductor-metal (MSM) structure, were investigated for the purposes of efficient and high-speed photodetection. The detection material was chosen to be InGaAs, where EHPs will be created upon excitation and will be separated through an applied reverse biased voltage, and collected in the form of photocurrent. Two slightly different plasmonic structures were selected for initial investigation, where feasibility of fabrication and experimental characterization were crucial factors. The first structure is an array of Au nanodipoles, filled with low-defect, low-doped InGaAs in their gaps. The photocurrent generated in this structure could be collected and measured through the dipole arms directly. The small size of the gaps leads to a very large electrical bandwidth (in the order of THz) for the device. Although this structure is advantageous in providing a very fast photodetector, fabrication of these arrays is challenging [67]. In order to tackle fabrication difficulties we considered a second structure consisting of an array of nanoantennas on a film of low-defect, low-doped InGaAs. The photocurrent, in this case, was be generated in a film of InGaAs, instead of nanoantenna gaps, and collected by applying a reverse biased voltage to the nanoantennas relative to the ground. The electric bandwidth is not as large as in the first case. However, the photodetector is predicted to be faster than commercially available devices, while yielding a reasonable responsivity [68].

3.2 Contribution

The results provided in this chapter were published as two articles in the journal of Applied Physics Letters and the Journal of Lightwave Technology. I implemented the numerical methods, generated and interpreted the results, and wrote the manuscript. Dr. Berini and Dr. Stöhr contributed to the interpretation of the results and revised the manuscript.

3.3 Article

The published articles follow verbatim.

APPLIED PHYSICS LETTERS **104**, 143112 (2014)

Plasmonic photodetector with terahertz electrical bandwidth

Saba Siadat Mousavi,[1] Andreas Stöhr,[2] and Pierre Berini[1,3,4]
[1]*School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada*
[2]*Institute of Optoelectronics, University of Duisburg-Essen, Duisburg, Germany*
[3]*Department of Physics, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada*
[4]*Centre for Research in Photonics, University of Ottawa, Ottawa, Ontario K1N 5N6, Canada*

(Received 24 February 2014; accepted 28 March 2014; published online 9 April 2014)

We propose and investigate a surface plasmon photodetector concept, based on the enhancement of electrical near-field in low-defect, low-doped $In_{0.53}Ga_{0.47}As$ detection volumes located in the gaps of an array of metal nanodipole antennas. We report enhancement in responsivity in the presence of nanodipoles and predict a maximum responsivity of ~100 mA/W at wavelengths near 1550 nm. The 3 dB electrical bandwidth of the device is estimated based on its RC rise time and the hole transit time through the detection volume for the cases of conventional and ballistic transport in InGaAs and is found to range from ~0.7 to 4 THz. Also, trends are observed relating the responsivity to the gap dimensions, revealing a trade-off between the field-enhancement in the gap and its volume, and leading to an optimum gap length producing the maximum responsivity.
 [http://dx.doi.org/10.1063/1.4870974]

Surface plasmon structures have been demonstrated to be useful in photodetection.[1,2] Among them, nanogratings and nanoantennas have attracted a lot of attention in recent years.[3–6] Schottky-based detectors[3,4] exploit enhanced absorptance in nanogratings and nanoantennas, which leads to higher rate of hot carrier generation at the Schottky barrier. The property of nanoantennas to enhance the electric near-field intensity at defined locations along the nanoantenna geometry, as well as in subwavelength semiconductor detection regions, has been investigated for fast and efficient photodetectors.[5,6] Using this property, one can significantly increase the absorptance in the semiconductor region, which leads to good photodetection, but in a very small volume, allowing the electrical bandwidth to be significantly extended. A recently published review discusses the topic of surface-plasmon photodetectors.[7]

In this Letter, we propose and investigate an array of nanodipoles with their gaps filled with InGaAs as a surface plasmon photodetector. Due to the enhanced electric field in the gap of the nanodipoles, enhanced photocurrent is produced due to increased production of electron-hole pairs in InGaAs. The nanoantennas' large effective aperture leads to a small detection volume, which means faster devices. The electrical bandwidth of this device is calculated to be in the range of a few THz, which is significantly higher than the bandwidth of, e.g., commercially available photodetectors.

Figure 1 depicts a schematic of the proposed structure. It consists of an array of Au nanodipoles on an InP substrate. The gaps of the nanodipoles are filled with low-defect, low-doped $In_{0.53}Ga_{0.47}As$ (lattice-matched to InP) as the detection medium. Photodetection occurs via the creation of electron-hole pairs therein, upon absorption of energy quanta at the wavelengths of interest (~1550 nm). The InGaAs forms Schottky contacts with the nanodipole arms, so each nanostructure behaves as metal-semiconductor-metal (MSM) photodetector. In order to collect the photocurrent produced in the InGaAs of each nanodipole, perpendicular Au connection lines of dimensions comparable to the nanodipoles are used and connected to a pad at each end. A DC bias applied to the pads establishes an electric field causing punch-through in the InGaAs regions, allowing the carriers to be separated and collected as the photocurrent.

An infinite array of nanodipoles is modeled optically where the pitch is fixed to 300×300 nm^2. The width and the thickness of nanodipoles are fixed ($w = 30$ nm, $t = 20$ nm), the gap (g) varies from 6 to 40 nm, and the length (l) is adjusted for each gap length such that the nanodipoles resonate near 1550 nm. An x-polarized plane wave source, with electric field magnitude of 1 V/m, illuminates the array from below. For this polarisation, the perpendicular interconnects are optically non-invasive. The finite difference time domain (FDTD) method was used to compute the electromagnetic response of the structure using measured optical parameters for the materials.[8,9]

The transmittance T was calculated as a function of frequency (wavelength) using

$$T(f) = \int_S \mathrm{Re}\{\boldsymbol{P}^{\boldsymbol{m}}(f)\} \cdot \boldsymbol{ds} \Big/ \int_S \mathrm{Re}\{\boldsymbol{P}^{\boldsymbol{s}}(f)\} \cdot \boldsymbol{ds}, \qquad (1)$$

where $\boldsymbol{P}^{m,s}$ is the Poynting vector at the reference planes where the transmittance is computed. f is the frequency and S is the surface of the reference plane. Equation (1) was also used to compute the reflectance R of the system by changing S to the appropriate reference plane. The absorptance is then determined as $A = 1 - T - R$.

T, R, and A are shown in Fig. 2(a) for example dimensions—this design resonates strongly near 1550 nm. Figure 2(b) shows the absorptance in the InGaAs volume only (A_{InGaAs}) for different gap lengths on the left axis, and the corresponding responsivities on the right axis. A_{InGaAs} is calculated as the integral of the absorbed power density within this volume

$$P_{abs} = -\frac{1}{2}\omega |E|^2 imag(\varepsilon), \qquad (2)$$

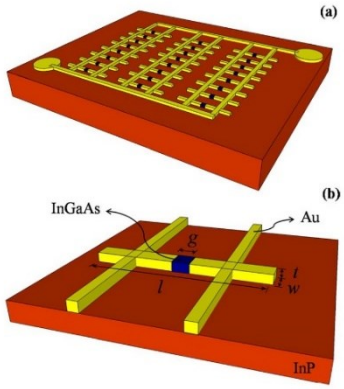

FIG. 1. Schematic of the system under study: (a) an array of Au nanodipoles, connected through interconnection lines, on an InP substrate. The gaps of the nanodipoles are filled with InGaAs (shown in blue); (b) a unit cell from the array in part (a), showing the dimensions of a nanoantenna. A plane wave source polarized along the length of the nanodipoles illuminates the array from below.

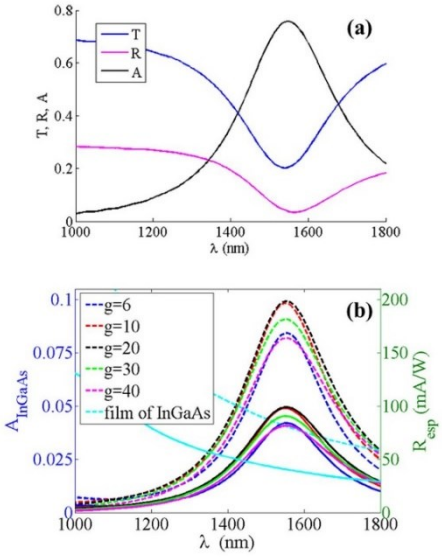

FIG. 2. (a) Transmittance T, reflectance R, and absorptance A of an array of nanodipoles of length $l = 214$ nm, $g = 20$ nm, width $w = 30$ nm, and thickness $t = 20$ nm. (b) Absorptance in the InGaAs volumes only (A_{InGaAs}) and detector responsivity R_{esp} on the left and right vertical axes, respectively. Different gap lengths were considered and in each case the length of the nanodipoles were adjusted such that they resonate at $\lambda = 1550$ nm ($g = 6$, $l = 166$; $g = 10$, $l = 182$; $g = 20$, $l = 214$; $g = 30$, $l = 236$; $g = 40$, $l = 254$; all dimensions in nm). The width and thickness are fixed at $w = 30$ nm and $t = 20$ nm. The case of an InGaAs film of the same thickness covering the entire area of a unit cell without nanodipoles is also computed for reference.

where ε is the optical permittivity of InGaAs. A_{InGaAs} is then used to obtain the responsivity R_{esp}, calculated as

$$R_{esp} = \eta_e \frac{q}{h\nu}, \tag{3}$$

where q is the charge of an electron, h is Plank's constant, ν is the optical frequency, and η_e is the external quantum efficiency given by

$$\eta_e = A_{InGaAs}\eta_i, \tag{4}$$

η_i is the internal quantum efficiency of InGaAs, which is taken to be 0.8.[10]

Comparing A and A_{InGaAs} in Figs. 2(a) and 2(b) (for $g = 20$ nm) shows significant difference between the two, which is due to the absorptance in the Au nanodipole arms. Absorption in Au could produce a photoresponse through the internal photoelectric effect,[7] but this detection mechanism is ignored relative to electron-hole creation in InGaAs.

A trend is observed in Fig. 2(b). As the gap increases, A_{InGaAs} increases until a maximum is reached, near $g = 10$–20 nm ($A_{InGaAs} \sim 0.1$ at 1550 nm). Further increase of the gap results in a drop of A_{InGaAs}. This trend exists because A_{InGaAs} is determined by the product of the electric field in InGaAs with its volume. As shown in Fig. 3, while a smaller gap produces more strongly enhanced electric fields, it also leads to a smaller volume leading to lower A_{InGaAs}; indeed, $A_{InGaAs} \to 0$ as $g \to 0$. On the other side of the maximum, as the gap increases, the field enhancement decreases leading to a lower A_{InGaAs}. Therefore, there exists an optimum value for g which maximises this product, in this case near $g = 10$–20 nm. The responsivity near this range reaches up to $\sim$100 mA/W.

It is worth noting that the presence of nanodipoles significantly enhances the absorptance in the InGaAs inside the gap. Two reference cases were considered to observe this enhancement. In the first case, the same geometry as shown in Fig. 1 was taken, but the nanodipoles were removed. The absorptance in the remaining volumes of InGaAs was found to be close to zero. In the second case, a continuous 20 nm thick film of InGaAs (same thickness as that used with the nanodipoles) on an InP substrate without nanodipoles was considered. In this case, the absorptance curve closely follows the material properties of InGaAs. As shown in Fig. 2(b) at 1550 nm, the level of absorptance in the film of InGaAs is roughly 0.04 which is significantly lower than our optimum value of $A_{InGaAs} \sim 0.1$, produced using nanodipoles with $g = 10$–20 nm. Although the film of InGaAs has a larger volume than the InGaAs inside the nanodipole gaps, the overall absorptance is much higher in the latter because the electric field is dramatically enhanced in the presence of nanodipoles ($\sim 30\times$ as shown in Fig. 3(b)).

The 3 dB electrical bandwidth is an important characteristic of photodetectors. We estimated this bandwidth using[11]

$$BW = \frac{1}{2\pi(\tau_1 + \tau_2)}, \tag{5}$$

where $\tau_1 = RC$ and τ_2 is the carrier transit time.

To obtain τ_1, we first estimate R as $\rho L_c/A_c$ by taking an average interconnect path length (perpendicular contacts in

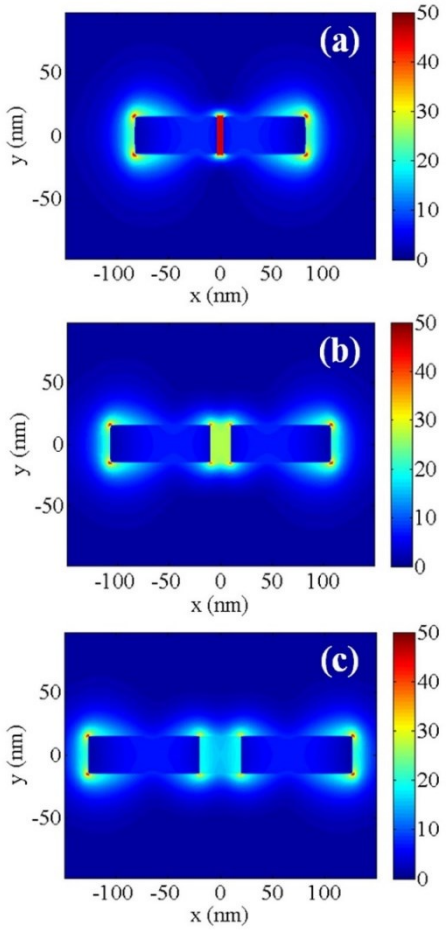

FIG. 3. Electric field distribution at $\lambda = 1550$ nm for nanodipoles of gap (a) $g = 6$ nm, (b) $g = 20$ nm, and (c) $g = 40$ nm. The lengths were adjusted to maintain resonance at $\lambda = 1550$ nm; (a) $l = 166$ nm, (b) $l = 214$ nm, and (c) $l = 254$ nm.

Fig. 1) for L_c and the cross-sectional area of the contact lines ($A_c = w \times t = 30 \times 20 = 600$ nm^2). A value of $R \sim 60\ \Omega$ is obtained for an average length of $L_c = 1650$ nm assuming a finite 6×6 array of 36 nanodipoles covering an area of $2 \times 2\ \mu$m^2—roughly the area of a tightly focused incident beam—and using the DC resistivity of Au ($\rho = 2.24 \times 10^{-8}\ \Omega$·m).The capacitance of a single nanodipole is then estimated as that of a parallel-plate capacitor defined by its gap filled with InGaAs, $C_d = \varepsilon A_c/g$ (neglecting fringing fields). Assuming the static permittivity of InGaAs ($\varepsilon \sim 14\varepsilon_0$ F/m), the same array of 36 nanodipoles ($A_c = 600$ nm^2) and the case $g = 10$ nm, yields a total capacitance of $C = 36C_d \sim 0.27$fF. Therefore, τ_1 works out to $\sim$16 fs.

In order to determine τ_2, we consider the transit of the slowest carriers (holes) for the cases of conventional and ballistic transport in InGaAs. For conventional transport, we take the maximum hole drift velocity in InGaAs[12] as $v_d \sim 5 \times 10^4$ m/s and estimate the transit time for a hole traversing the full length of the gap as $\tau_2 = g/v_d$, which works out to $\tau_2 \sim 200$ fs for $g = 10$ nm. Ballistic light hole transport has been observed in GaAs and a mean free path of 14 nm measured.[13] Given the comparatively short gap, ballistic hole transport could occur in the InGaAs resulting in a drift velocity about 10× larger, thus reducing τ_2 to $\sim$20 fs.[14]

The electrical bandwidth is then estimated by substituting the values for τ_1 and τ_2 into Eq. (5), yielding BW $\sim$ 0.7–4 THz, the latter occurring for ballistic transport. (The skin depth at 4 THz is $\sim$40 nm which is comparable to the interconnect dimensions—their 60 Ω resistance is therefore slightly underestimated but accurate enough for order-of-magnitude estimates.)

In summary, this study motivates and demonstrates the feasibility of using nanodipoles in conjunction with nanoscale semiconductor detection volumes to produce very fast photodetectors of broad optical bandwidth and reasonable responsivity. The properties of nanodipoles which enhance the electric field intensity inside the gap volume lead to significant increase of absorptance, which allows for a smaller detection volume, and thus a faster photodetector. The nanoantennas, particularly the field enhancement produced therein, alter the responsivity-speed trade-off encountered in conventional PIN photodetectors.

[1]A. Akbari, R. N. Tait, and P. Berini, Opt. Express **18**, 8505 (2010).
[2]I. Goykhman, B. Desiatov, J. Khurgin, J. Shappir, and U. Levy, Nano Lett. **11**, 2219 (2011).
[3]M. W. Knight, H. Sobhani, P. Nordlander, and N. J. Halas, Science **332**, 702 (2011).
[4]A. Karar, N. Das, C. L. Tan, K. Alameh, Y. T. Lee, and F. Karouta, Appl. Phys. Lett. **99**, 133112 (2011).
[5]T. Ishi, J. Fujikata, K. Makita, T. Baba, and K. Ohashi, Jpn. J. Appl. Phys., Part 2 **44**, L364 (2005).
[6]L. Tang, S. E. Kocabas, S. Latifi, A. K. Okyay, D. Ly-Gagnon, K. C. Saraswat, and D. A. B. Miller, Nat. Photonics **2**, 226 (2008).
[7]P. Berini, Laser Photonics Rev. **8**, 197–220 (2014).
[8]E. D. Palik, *Handbook of Optical Constants of Solid* (Academic Press, San Diego, 1997).
[9]See http://refractiveindex.info/?group=CRYSTALS&material=InGaAs for information about refractive index data of InGaAs.
[10]G. Keiser, *Optical Fiber Communications* (McGraw-Hill, 2000), p. 248.
[11]G. P. Agrawal, *Fibre-Optic Communication Systems* (John Wiley & Sons, New York, 2002).
[12]P. Hill, J. Schlafer, W. Powazinik, M. Urban, E. Eichen, and R. Olshansky, Appl. Phys. Lett. **50**, 1260 (1987).
[13]M. Heiblum, K. Seo, H. P. Meier, and T. W. Hickmott, Phys. Rev. Lett. **60**, 828 (1988).
[14]G. H. Dohler, F. Renner, O. Klar, M. Eckardt, A. Schwanhaußer, S. Malzer, D. Driscoll, M. Hanson, A. C. Gossard, G. Loata, T. Löffler, and H. Roskos, Semicond. Sci. Technol. **20**, S178 (2005).

Ultrafast Surface Plasmon III–V Photodetectors Based on Nanomonopoles

Saba Siadat Mousavi, Andreas Stöhr, *Senior Member, IEEE*, and Pierre Berini, *Fellow, IEEE*

***Abstract*—A nanoantenna-based plasmonic photodetector is proposed and investigated theoretically. An array of interconnected nanomonopoles on a film of InGaAs on n-doped InP is modeled. The infrared (∼1550 nm) responsivity of the structure can reach ∼200 mA/W for a 20-nm-thick InGaAs layer. The optimum 3-dB electrical bandwidth of the device is calculated to be ∼1 THz for the case of a 20-nm-thick InGaAs layer and a 4 × 4 nanomonopole array. The electrical bandwidth is shown to increase by decreasing the number of elements in the array. A simple linear model for the total resistance of the structure is proposed and validated. The total resistance increases by increasing the thickness of the InGaAs layer and by decreasing the number of array elements.**

***Index Terms*—Nanoantenna, photodetector, surface plamson, THz.**

I. Introduction

PLASMONIC photodetectors are of great interest, mainly because of their enhanced photo-response and high speed compared to conventional PIN photodetectors [1]. Enhanced and localized surface plasmon-polariton (SPP) fields can greatly increase the number of electron-hole pairs (EHP) created in the detection region of the photodetector, or improve the internal photoemission (IPE) process [2]–[12]. Nanoantennas and nanogratings have been employed in plasmonic photodetectors [13]–[18]. Such nanostructures, in conjunction with small detection regions, can produce extremely fast photodetectors [19]. Plasmonic photodetectors are used to generate THz sources based on photomixing [20].

In this work we propose and investigate a novel concept for a plasmonic photodetector based on nanoantennas on a very thin semiconductor detection layer–a design that is easier to fabricate than our previous proposal [19]. Due to the large effective aperture of the nanoantennas [21], input fields are strongly enhanced and confined to the nanoantenna geometry, which overlaps well with the detection region, resulting in a subwavelength-sized detector. The electrical bandwidth of the device reaches a few THz while maintaining a reasonable infrared responsivity at telecommunications wavelengths.

Manuscript received January 15, 2016; revised April 14, 2016 and May 9, 2016; accepted May 18, 2016. Date of publication May 31, 2016; date of current version September 25, 2016.

S. S. Mousavi is with the School of Electrical Engineering and Computer Science, and the Centre for Research in Photonics, University of Ottawa, Ottawa K1N 6N5, Canada (e-mail: ssiad031@uottawa.ca).

A. Stöhr is with the Center for Semiconductor Technology and Optoelectronics (ZHO), Department of Optoelectronics, University Duisburg-Essen, Duisburg, Germany (e-mail: andreas.stoehr@uni-due.de).

P. Berini is with the School of Electrical Engineering and Computer Science, the Department of Physics, and the Centre for Research in Photonics, University of Ottawa, Ottawa K1N 6N5, Canada (e-mail: berini@eecs.uottawa.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2016.2575002

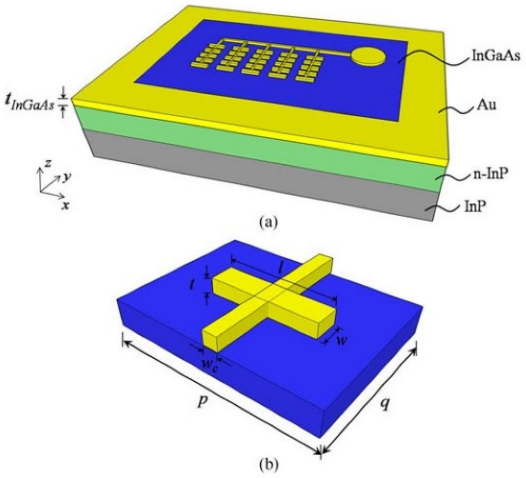

Fig. 1. (a) A 5 × 5 array of Au nanomonopoles on a thin film of unintentionally doped InGaAs on a layer of n-doped InP, deposited on an unintentionally doped InP substrate. The doping level of n-InP is assumed to be 2×10^{18} cm^{-3}, and provides the ground through an Ohmic contact to the top metal ring. Nanomonopoles are electrically connected to each other and to a device contact pad, forming a Schottky contact to the InGaAs layer. (b) Unit cell of the array in part (a), showing the dimensions of a nanomonopole; p and q are the dimensions of a unit cell.

II. Concept and Methods

An array of Au nanomonopoles is envisioned forming a Schottky contact on a film of low-defect, low-doped ($n \sim 10^{15}$ cm^{-3}) $In_{0.53}Ga_{0.47}$ As (lattice-matched to InP), located on a 0.5 μm thick n-doped layer of InP ($n \sim 2 \times 10^{18}$ cm^{-3}), which, in turn, is located on an unintentionally doped InP substrate ($n \sim 10^{15}$ cm^{-3}). The schematic is shown in Fig. 1. The array is assumed to be illuminated through the substrate by a tightly focused Gaussian beam (of diameter comparable to the array size) at an optical wavelength of 1550 nm, polarized along the length of the nanomonopoles. The excitation wavelength is far from the cut-off wavelength of the InGaAs layer (1680 nm), ensuring no bandedge effects. The nanomonopoles are interconnected by vertical, optically non-invasive Au connection lines, which gather to a contact pad located away from the array. Upon illumination, SPPs are excited on the Au nanomonopoles, with fields that overlap strongly with the InGaAs layer. These localized enhanced SPPs result in the creation of EHPs in the InGaAs film due to absorption, then collected as the photocurrent by applying a reverse bias voltage to the device contact pad relative to the ground contact ring forming an ohmic

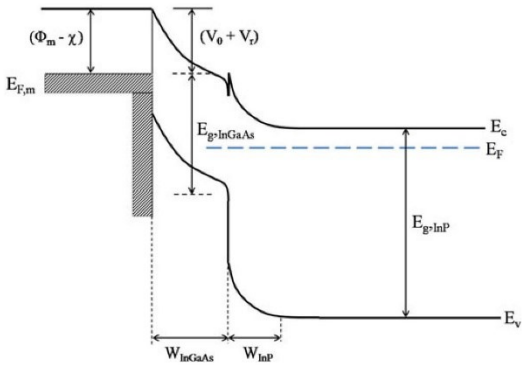

Fig. 2. Energy-band diagram of the Au/n-InGaAs/n-InP structure under a reverse bias voltage. A Schottky barrier is formed between Au (left) and n-InGaAs (middle layer). A heterojunction is formed at the n-InGaAs/n-InP junction. Band discontinuities and band bending at equilibrium are shown. E_c, E_v, E_F, and E_g are the conduction band, valance band, Fermi level, and band gap energies, respectively. V_0 is the built-in potential, V_r is the reverse bias voltage, ϕ_m is the metal work function, and χ is the electron affinity. W_{InGaAs}, and W_{InP} are the depletion widths of n-InGaAs (fully depleted) and n-InP, layers respectively.

contact to the n-doped InP layer. The bandgap energy of the InP regions is larger, precluding the creation of EHP's therein.

Fig. 2 shows the energy-band diagram of the device. A Schottky contact is formed at Au/n-InGaAs junction. The depletion width of an Au/n-InGaAs junction of the doping level assumed is $W_{\mathrm{InGaAs}} \sim 1.5$ μm under a reverse bias voltage of $V_r = 1$ V, which is significantly larger than the thickness of the n-InGaAs layer. Thus the n-InGaAs layer is always fully depleted.

A heterojunction is formed at the n-InGaAs/n-InP interface. To estimate the depletion width into the n-InP, W_{InP}, we neglect the n-InGaAs region, which is justified because the n-InGaAs layer is thin, lightly doped and fully depleted, thus producing little electric field drop across this layer. W_{InP} is therefore approximated as

$$W_{\mathrm{InP}} = \sqrt{\frac{2\varepsilon_s\,(V_0 - V_r)}{qN_d}} \approx 34 \text{ nm} \tag{1}$$

where ε_s is permittivity of semiconductor ($12.5\varepsilon_0$ for InP [22]), V_r is the reverse bias voltage (1 V), q is the electronic charge, N_d is the doping level of the semiconductor (2×10^{18} cm^{-3}), and $V_0 = \Phi_m - \Phi_s \sim 0.7$ V is the built in potential with Φ_m (5.1 V [23]) and Φ_s (4.405 V [22]) the work functions of the metal and semiconductor, respectively.

The finite-difference time-domain (FDTD) method was used to model numerically Maxwell's equations and obtain the optical performance of the system. Finite arrays were modelled as an infinite array of nanomonopoles illuminated by a plane wave through the substrate. This was achieved by considering one array element (unit cell) bounded on all lateral walls by periodic boundary conditions. Perfectly-matched layers were applied on the top and bottom walls to avoid reflections. The incident Gaussian beam diameter is assumed sufficiently larger than the dimensions of a unit cell that the former can be treated as a plane wave while modeling a single nanomonopole. The interconnection lines were not present in the optical model, since they are perpendicular to the direction of beam polarization, and therefore do not affect the optical performance of the system. The transmittance T and reflectance R were computed numerically as a function of frequency at reference planes located in the far-field of the nanomonopoles. The total absorptance of the detector A, defined as $A = 1 - T - R$, is then calculated to find the resonant wavelength of the designed structures. The power density absorbed in the InGaAs layer is calculated as

$$P_{\mathrm{abs}} = -\frac{1}{2}\,\omega|E|^2 \mathrm{imag}\,(\varepsilon) \tag{2}$$

where ω is the angular frequency, E is the electric field, and ε is the optical permittivity of InGaAs. For an input power normalized to 1 W, the volume integral of P_{abs} over the InGaAs volume corresponds directly to the absorptance in the InGaAs layer, A_{InGaAs}. The DC responsivity, R_{esp}, due to detection in the InGaAs layer is calculated as

$$R_{\mathrm{esp}} = \eta_e {}^{q}/_{h\nu} \tag{3}$$

where q is the charge of an electron, h is Planck's constant, ν is the optical frequency, and η_e is the external quantum efficiency, obtained as

$$\eta_e = A_{\mathrm{InGaAs}}\eta_i \tag{4}$$

with η_i, the internal quantum efficiency of InGaAs, taken as 0.8 [24]. It is worth noting here that although A_{InGaAs} is obtained using a plane wave source, as long as the Gaussian beam diameter is smaller than or equal to the array size R_{esp} remains unchanged.

In the electrical analysis, the structure was considered as a lumped element and its electrostatic properties were determined to estimate the electrical bandwidth. This is reasonable because of the small size of the array relative to the electrical wavelength of operation - the size of the largest array considered is 1.2 μm × 1.2 μm (6 × 6 element array) and the highest frequency of operation is expected to be about 2 THz for which the electrical wavelength is 150 μm, much larger than the array size. The electrostatic modeling was carried out using finite element analysis (FEA) to find the capacitance and resistance of the structure. A finite array was considered, where the pitch size and dimensions of the nanomonopoles were the same as in the optical FDTD analysis. The Au connection lines are present in this case, as well as the ground contact ring, since they affect the electrostatic properties of the system. The device contact pad, however, was not considered.

III. Results

A. Optical Analysis

The array parameters were chosen such that the nanomonopoles were uncoupled and would resonate at $\lambda = 1550$ nm [25]. The designs were also optimized for the maximum responsivity and electrical bandwidth, while considering technological limitations of fabrication. The parameters were chosen as follows: the pitch size is 200 nm × 200 nm, and the nanomonopoles are 108 nm long (l), 20 nm wide (w) and 20 nm thick (t). Fig. 3 shows A_{InGaAs} computed using (2) for different InGaAs thicknesses (left axis), and the responsivity associated with each case of A_{InGaAs} (right axis). The wavelength

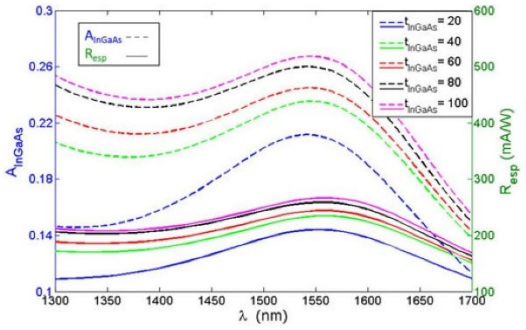

Fig. 3. (left) Absorptance in the InGaAs film, A_{InGaAs}, and (right) responsivity, R_{esp}, for an infinite array of nanomonopoles of dimensions $t = 20$ nm, $w = 20$ nm and $l = 108$ nm, located on a film of InGaAs. The thickness of the film of InGaAs varies from 20 to 100 nm in steps of 20 nm.

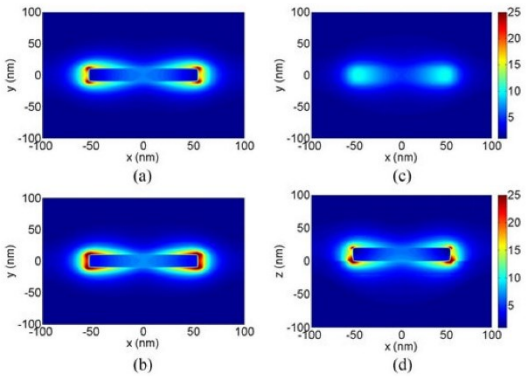

Fig. 4. Magnitude of the electric field [V/m] on (a) an *xy* plane through the middle of a nanomonopole, 10 nm above the InGaAs-Au interface, (b) an *xy* plane in InGaAs, 5 nm below the surface, (c) an *xy* plane in InGaAs, 10 nm below the surface, and (d) an *xz* plane through the longitudinal center of the antenna. The thickness of the InGaAs layer is 20 nm. The incident electric field is 1 V/m.

dependence observed is due to the wavelength response of the nanoantennas. We note that by increasing the thickness of the film of InGaAs from 20 to 100 nm, the peak R_{esp} increases from 211 to 267 mA/W. This trend is in agreement with expectations, because a thicker InGaAs film means a larger absorbing volume, which results in higher responsivity.

The electric field distribution at $\lambda = 1550$ nm is shown in Fig. 4. It is worth noting that although the fields are highly localized around the nanomonopoles, close to the Au-InGaAs interface, the fields maintain the same distribution and remain relatively undistorted up to 10 nm below the interface.

B. Electrical Analysis

Fig. 5 depicts the equivalent AC electrical circuit of the photodiode that is used to estimate the time constant of the photodetector. In this model, the junction resistance, R_j, is assumed to be very large, and the load resistance, R_L is 50 Ω.

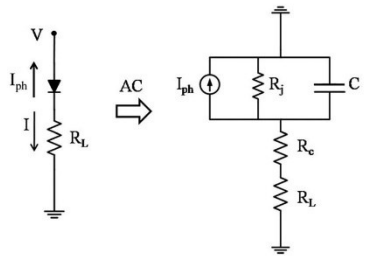

Fig. 5. AC equivalent circuit of the photodiode under reverse bias. I_{ph} is the photocurrent generated in the InGaAs layer. R_j and C are the junction resistance and capacitance. R_c is the contact resistance and R_L is the load resistance.

The junction capacitance, *C*, due to the depletion region, and contact resistance, R_c, are discussed in the following sections.

1) Capacitance: The capacitance of the detector under reverse bias is required to estimate its bandwidth. By applying a reverse bias voltage to the array and connection lines, and grounding the undepleted portion of the n-InP layer, the capacitance, *C*, of the structure is obtained as:

$$C = \frac{Q}{V_a} \tag{5}$$

Q is the charge on one capacitor terminal, say the interconnected array held at potential V_a, and obtained by integrating the volume charge density thereon, ρ_v, itself computed via

$$\nabla \cdot \vec{D} = \rho_v \tag{6}$$

$$\vec{E} = -\nabla\varphi. \tag{7}$$

The electric potential distribution, φ, is determined by solving Laplace's equation via FEA. φ is assumed zero on all external boundaries and takes on the potential V_a on the array (set to $V_a = 1$ V).

C is calculated as a function of the thickness of the InGaAs layer, for three arrays of 4 × 4, 5 × 5 and 6 × 6 elements, of the same dimensions as assumed for the optical analysis, with the addition of connection lines of width and thickness fixed to $w_c = t_c = t = 20$ nm. The depleted portion of the n-doped InP was taken into account (34 nm, (1)). The thickness of InGaAs varies from 5 to 50 nm in steps of 5 nm. As expected, increasing the thickness of InGaAs decreases *C*, following a $1/t_{\mathrm{InGaAs}}$ dependency, as shown in Fig. 6. On the other hand, increasing the number of elements in the array results in a larger *C*, as the capacitor area becomes larger.

The norm of the electrostatic electric field, $|E| = \sqrt{E_x^2 + E_y^2 + E_z^2}$, is shown in Fig. 7 for an array of 5 × 5 elements at different locations in the device. The distribution shows confined and localized static fields within the InGaAs layer at the position of the array elements with very little fringing or spreading, as expected for a thin InGaAs thickness.

2) Resistance: The resistance of the detector is also required in order to estimate its electrical bandwidth. The total resistance, R_t, of the device is computed numerically. An arbitrary forward bias voltage (e.g., 1 V) is applied to one of the end facets of

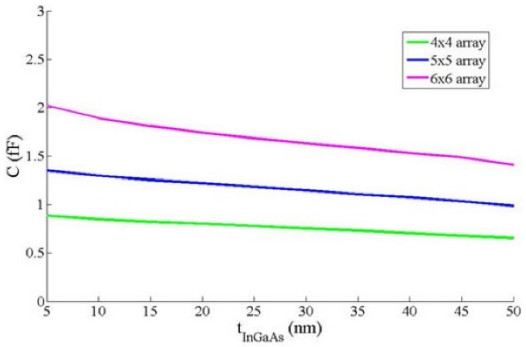

Fig. 6. Capacitance as a function of the thickness of the InGaAs layer for three arrays of 4 × 4, 5 × 5 and 6 × 6 elements. The element pitch size is 200 nm × 200 nm. Nanomonopole dimensions are $l = 108$ nm, $w = t = 20$ nm, and the connection lines are 20 nm wide by 20 nm thick.

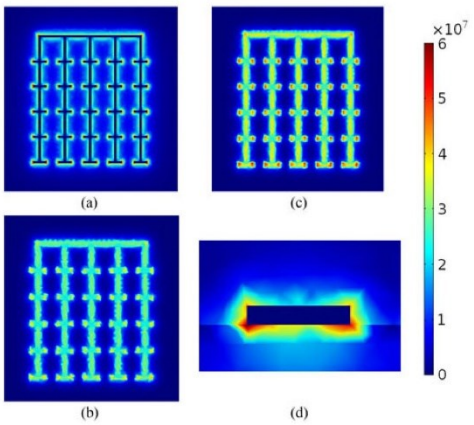

Fig. 7. Electric field distribution [V/m] on (a) an *xy* plane through the middle of the nanoantennas, 10 nm above the InGaAs-Au interface, (b) an *xy* plane in InGaAs, 5 nm below the InGaAs-Au interface, (c) an *xy* plane in InGaAs, 10 nm below the InGaAs-Au interface, and (d) a *xz* plane through the longitudinal center of a nanomonopole. A potential of 1 V was applied to the array of nanomonopoles relative to the grounded n-InP layer. The thickness of the InGaAs layer is 20 nm.

the connection line interconnecting the array columns to the contact pad (see Fig. 1(a)), relative to the ground ring. This models operation of the device under a forward bias (turn-on voltage neglected). Solving Laplace's equation for the electric potential distribution φ via the FEA, the current density J is then computed from

$$\vec{E} = -\nabla\varphi \tag{8}$$

$$\vec{J} = \sigma\vec{E} \tag{9}$$

Fig. 8 depicts the distribution of the current density magnitude for an array of 5 × 5 elements at various cross-sections in the device. Since the connection lines collect the current density from several nanomonopoles, the current density therein

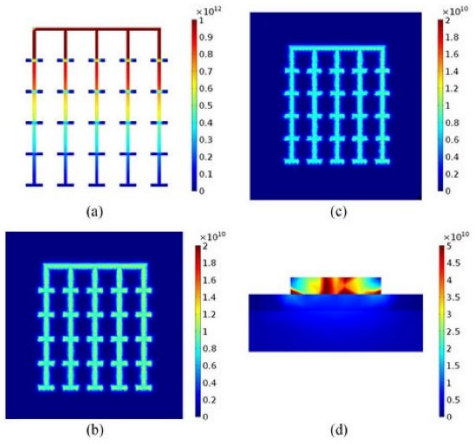

Fig. 8. Distribution of the current density magnitude [A/m^2] on (a) an *xy* plane through the middle of the nanoantennas, 10 nm above the InGaAs-Au interface, (b) an *xy* plane in InGaAs, 5 nm below the InGaAs-Au interface, (c) an *xy* plane in InGaAs, 10 nm below the InGaAs-Au interface, and (d) an *xz* plane through the longitudinal center of a nanomonopole. A potential of 1 V was applied to the right end facet of the horizontal connection line relative to the grounded n-InP layer.

is larger, as observed from Fig. 8(a). Fig. 8(b)–(d) shows confined current distributions within the InGaAs film with very little current spreading - this is reasonable given the small InGaAs thickness. Therefore, a simple model is proposed for the total resistance,

$$R_t = R_c + R_s \tag{10}$$

where R_c is the contact resistance of the array (i.e., of the nanoantenna network) and R_s is the resistance due to conduction through the film of InGaAs:

$$R_s = \rho\frac{t_{\mathrm{InGaAs}}}{A_c} \tag{11}$$

where ρ and A_c are the static resistivity of InGaAs (dopant dependant), and the conduction area, respectively.

In Fig. 9 R_t (computed numerically) is plotted as a function of ρ for $t_{\mathrm{InGaAs}} = 20$, 30, 40 and 50 nm, for three arrays of 4 × 4, 5 × 5 and 6 × 6 elements, showing a linear trend with ρ. By fitting our linear model (10) and (11) to the numerical computations, we extract the contact resistance of our designs R_c, yielding $R_c = 31, 28$ and $31\,\Omega$, for arrays of $4 \times 4, 5 \times 5$ and 6×6 elements, respectively. These contact resistances remain almost constant with the thickness of InGaAs for each array. They also remain constant irrespective of the bias (forward or reverse).

3) Electrical Bandwidth: The 3 dB electrical bandwidth of the photodetector is estimated as

$$BW = \frac{1}{2\pi\left(\tau_1 + \tau_2\right)} \tag{12}$$

where $\tau_1 = (R_L + R_c) \times C$ is the RC time constant and τ_2 is the carrier transit time, determined by the transit time of the electrons and holes for the cases of conventional and ballistic

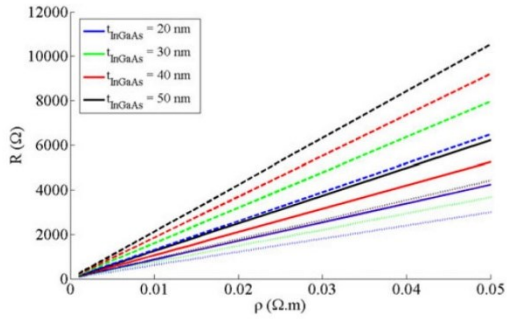

Fig. 9. (a) Total resistance of the structure (nanoantenna array and InGaAs layer), R_t, as a function of the resistivity, ρ, of the InGaAs. Dashed line: array of 4×4 elements; solid line: array of 5×5 elements; dotted line: array of 6×6 elements. All cases have pitch dimensions of $p = 200$ nm, $q = 200$ nm, and antenna dimensions of $t = 20$ nm, $w = 20$ nm and $l = 108$ nm. Four different InGaAs thicknesses are considered for each array. The contact lines are also 20 nm wide and 20 nm thick.

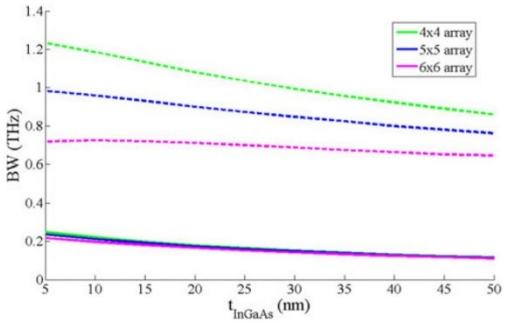

Fig. 10. Electrical bandwidth (BW) as a function of the thickness of the InGaAs film. Three arrays of 4×4, 5×5 and 6×6 elements are considered. The pitch dimensions are $p = 200$ nm, $q = 200$ nm, and the antenna dimensions are $l = 108$ nm, $w = t = 20$ nm. The contact lines are also 20 nm wide and 20 nm thick. Solid line: conventional transport; Dashed line: ballistic transport.

transport. Carrier transport is essentially vertical, because of the highly confined optical and bias fields (see Fig. 4 and 7). We determine the drift velocity of electrons and holes by finding the electric field in the depleted region ($t_{\mathrm{InGaAs}} + W_{\mathrm{InP}}$). Here, we assume the electric field to be uniform and continuous across the InGaAs/InP boundary, since the permittivity of both materials is very similar. Under a 1 V reverse bias, the electric field varies from ~260 kV/cm to ~120 kV/cm as t_{InGaAs} varies from 5 nm to 50 nm, for which we obtain the maximum drift velocity of holes in InGaAs as $v^h_{d,\,\mathrm{InGaAs}} = 6 \times 10^4$ m/s [26] and in InP as $v^h_{d,\mathrm{InP}} = 7 \times 10^4$ m/s [27]. The saturation drift velocity of electrons in InGaAs and InP are $v^e_{d,\mathrm{InGaAs}} = 1.5 \times 10^5$ m/s [28] and $v^e_{d,\mathrm{InP}} = 1.2 \times 10^5$ m/s [29]. The transit time for each carrier species is then determined by

$$\tau_2^{e,h} = x_d / v_d^{e,h} \tag{13}$$

where x_d is the maximum distance that each must travel. From Fig. 2, the maximum distance that the holes must travel corresponds to the thickness of InGaAs, because EHP's are photoexcited only in the InGaAs layer, and holes are collected in the Au region. On the other hand, the maximum distance that the electrons must travel is $t_{\mathrm{InGaAs}} + W_{\mathrm{InP}}$ in order to be collected in the undepleted portion of the n-InP layer. The carrier transit time is taken as $\tau_2 = \max(\tau_2^e,\ \tau_2^h)$ which gives $\tau_2 \approx 730$ fs for conventional transport in a 20 nm thick film of InGaAs and $\tau_2 \approx 73$ fs for ballistic transport (assuming $10 \times v_d$ [30]—a mean free path of 14 nm has been measured for ballistic light hole transport in GaAs [31]).

Fig. 10 shows the electrical bandwidth (BW) of the photodetector for three array sizes, assuming conventional (carrier saturation velocity) and ballistic transport. For t_{InGaAs} comparable to the ballistic mean free path, the *BW* is well-estimated assuming ballistic transport. However, for much larger thicknesses, assuming conventional transport is more appropriate. It is expected that as $t_{\mathrm{InGaAs}} \to 0$ the bandwidth drops because the capacitance and τ_1 diverge, and as t_{InGaAs} becomes large, the bandwidth drops because the carrier transit time τ_2 increases. The bandwidth is larger for smaller arrays, because τ_1 is smaller. Good thicknesses for the InGaAs film are in the range of 10 to 20 nm. In this range it is reasonable to assume mainly ballistic carrier transport, so the expected electrical BW would be ~1 THz, and the expected responsivity would be ~200 mA/W (Fig. 3).

4) Dark Current: The dark current is the main contributor to the noise performance of the device. The dark current I_d is estimated as [32]:

$$I_d = C_{\mathrm{area}} A^* T^2 \exp\left(\frac{-q\Phi_B}{kT}\right) \tag{14}$$

where C_{area} is the Schottky contact area, T is the absolute temperature, q is the electron charge, k is Boltzmann's constant, Φ_B is the Schottky barrier height (see Fig. 2) given by $\Phi_B = \Phi_M - \chi$ with Φ_M being the metal work function (5.1 V for Au [23]) and χ the electron affinity of the semiconductor (4.5 V for InGaAs [33]). A^* is the effective Richardson constant, calculated as:

$$A^* = \frac{4\pi m^* q k^2}{h^3} \tag{15}$$

with h being Plank's constant. Assuming $m^* = 0.045 m_0$ for $\mathrm{In_{0.53}Ga_{0.47}As}$ [34] gives $A^* = 5.41\ \mathrm{A\,cm^{-2}K^{-2}}$. Substituting A^* and $\Phi_B = 0.6$ eV in (14), the dark current is found to be $I_d = 68.9$ fA, where $C_{\mathrm{area}} = 1.7 \times 10^5$ nm^2 for a 5×5 array.

IV. Conclusion

In summary, we proposed a surface plasmon infrared (~1550 nm) photodetector based on nanomonopoles on III-V. The structure is investigated theoretically and modeled numerically. Nanomonopoles are used to enhance the absorptance of input radiation within a nanoscale InGaAs detection region, thus increasing the creation of EHPs, which leads to a good responsivity. Furthermore, due to their nanoscale dimensions, nanoantenna-based photodetectors offer a THz electrical bandwidth. The thickness of the InGaAs film and the array size are the main factors that determine the speed and responsivity of the photodetector, assuming fixed design of the array elements. Integrating plasmonic nanoantennas onto photodetectors alters the bandwidth-responsivity trade-off found in conventional photodetectors.

REFERENCES

[1] P. Berini, "Surface plasmon photodetectors and their applications," *Laser Photonics Rev.*, vol. 8, no. 2, pp. 197–220, 2014.

[2] P. Neutens, P. van Dorpe, I. de Vlaminck, L. Lagae, and G. Borghs, "Electrical detection of confined gap plasmons in metal–insulator–metal waveguides," *Nature Photonics*, vol. 3, pp. 283–286, Apr. 2009.

[3] T. Dufaux, J. Dorfmüller, R. Vogelgesang, M. Burghard, and K. Kern, "Surface plasmon coupling to nanoscale Schottky-type electrical detectors," *Appl. Phys. Lett.*, vol. 97, 2010, Art. no. 161110.

[4] F.-F. Ren, K.-W. Ang, J. Song, Q. Fang, M. Yu, G.-Q. Lo, and D. L. Kwong, "Surface plasmon enhanced responsivity in a waveguided germanium metal–semiconductor–metal photodetector," *Appl. Phys. Lett.*, vol. 97, 2010, Art. no. 091102.

[5] A. L. Falk *et al.*, "Near-field electrical detection of optical plasmons and single-plasmon sources," *Nature Phys.*, vol. 5, pp. 475–479, 2009.

[6] A. Akbari and P. Berini, "Schottky contact surface-plasmon detector integrated with an asymmetric metal stripe waveguide," *Appl. Phys. Lett.*, vol. 95, 2009, Art. no. 021104.

[7] S. Zhu, G. Q. Lo, and D. L. Kwong, "Theoretical investigation of silicide Schottky barrier detector integrated in horizontal metal-insulator-silicon-insulator-metal nanoplasmonic slot waveguide," *Opt. Express*, vol. 19, no. 17, pp. 15843–15854, 2011.

[8] Y.-M. Bahk *et al.*, "Plasmon enhanced terahertz emission from single layer Graphene," *ACS Nano*, vol. 8, no. 9, pp. 9089–9096, Aug. 2014.

[9] M. L. Brongersma, N. J. Halas, and P. Nordlander, "Plasmon-induced hot carrier science and technology," *Nature Nanotechnol.*, vol. 10, pp. 25–34, Jan. 2015.

[10] B. Desiatov, I. Goykhman, N. Mazurski, J. Shappir, J. B. Khurgin, and U. Levy, "Plasmonic enhanced silicon pyramids for internal photoemission Schottky detectors in the near-infrared regime," *Optica*, vol. 2, no. 4, pp. 335–338, 2015.

[11] M. Galanty, S. Yochelis, L. Stern, I. Dujovne, U. Levy, and Y. Paltiel, "Extinction enhancement from a self-assembled quantum dots monolayer using a simple thin films process," *J. Phys. Chem. C*, vol. 119, no. 44, pp. 24991–24995, Oct. 2015.

[12] S. Muehlbrandt *et al.*, "Plasmonic internal photoemission detectors with responsivities above 0.12 A/W," presented at the Lasers Electro-Optics Conf., San Jose, CA, USA, 2015.

[13] L. Tang *et al.*, "Nanometer-scale germanium photodetector enhanced by a near-infrared antenna," *Nature Photonics*, vol. 2, pp. 226–229, Mar. 2008.

[14] J. Hetterich, G. Bastian, N. A. Gippius, S. G. Tikhodeev, G. von Plessen, and U. Lemmer, "Optimized design of plasmonic MSM photodetector," *IEEE J. Quant. Electron.*, vol. 43, no. 10, pp. 855–859, Oct. 2007.

[15] S. Collin, F. Pardo, R. Teissier, and J.-L. Pelouard, "Efficient light absorption in metal–semiconductor–metal nanostructures," *Appl. Phys. Lett.*, vol. 85, pp. 194–196, 2004.

[16] D. Crouse, "Numerical modeling and electromagnetic resonant modes in complex grating structures and optoelectronic device applications," *IEEE Trans. Electron Devices*, vol. 52, no. 11, pp. 2365–2373, Nov. 2005.

[17] D. Okamoto, J. Fujikata, K. Nishi, and K. Ohashi, "Numerical study of near-infrared photodetectors with surface-plasmon antenna for optical communication," *Jpn. J. Appl. Phys.*, vol. 47, no. 4, pp. 2921–2923, 2008.

[18] Z. Fang, Z. Liu, Y. Wang, P. M. Ajayan, P. Nordlander, and N. J. Halas, "Graphene-antenna sandwich photodetector," *Nano Lett.*, vol. 12, no. 7, pp. 3808–3813, 2012.

[19] S. Siadat Mousavi, A. Stöhr, and P. Berini, "Plasmonic photodetector with terahertz electrical bandwidth," *Appl. Phys. Lett.*, vol. 104, 2014, Art. no. 143112.

[20] S. Yang and M Jarrahi, "Frequency-tunable continuous-wave terahertz sources based on GaAs plasmonic photomixers," *Appl. Phys. Lett.*, vol. 107, no. 13, 2015, Art. no. 131111.

[21] P. Bharadwaj, B. Deutsch, and L. Novotny, "Optical antennas," *Adv. Opt. Photonics*, vol. 1, pp. 438–483, 2009.

[22] [Online] Available: http://www.ioffe.ru/SVA/NSM/Semicond/InP/basic.html (for information about material parameters of InP).

[23] P. A. Tipler and R. A. Llewellyn, *Modern Physics*. San Francisco, CA, USA: Freeman, 2008.

[24] G. Keiser, *Optical Fiber Communications*. New York, NY, USA: McGraw-Hill, 2000, p. 248.

[25] S. Siadat Mousavi, P. Berini, and D. McNamara, "Periodic plasmonic nanoantennas in a piecewise homogeneous background," *Opt. Express*, vol. 20, no. 16, pp. 18044–18065, 2012.

[26] S. Adachi, *Physical Properties of III-Y Semiconductor Compounds*. Hoboken, NJ, USA: Wiley, 1992.

[27] K. Brennan and K. Hess, "Theory of high-field transport of holes in GaAs and InP," *Phys. Rev. B*, vol. 29, pp. 5581–5584, 1984.

[28] V. Balynas, A. Krotkus, A. Stalnionis, A. T. Gorelionok, N. M. Shmidt, and J. A. Tellefsen, "Time-resolved, hot-electron conductivity measurement using an electro-optic sampling technique," *Appl. Phys. A*, vol. 51, no. 4, pp. 357–360, 1990.

[29] T. J. Maloney and J. Prey, "Transient and steady-state electron transport properties of GaAs and InP," *J. Appl. Phys.*, vol. 48, no. 2, pp. 781–787, 1977.

[30] G. H. Dohler *et al.*, "THz-photomixer based on quasi-ballistic transport," *Semicond. Sci. Technol.*, vol. 20, no. 7, pp. 178–190, Jun. 2005.

[31] M. Heiblum, K. Seo, H. P. Meier, and T. W. Hickmott, "Observation of ballistic holes," *Phys. Rev. Lett.*, vol. 60, pp. 828–831, 1988.

[32] S. M. Sze and K. K. NG, *Physics of Semiconductor Devices*. Hoboken, NJ, USA: Wiley, 2007.

[33] [Online] Available: http://www.ioffe.ru/SVA/NSM/Semicond/GaInAs/bandstr.html (for information about material parameters of InGaAs).

[34] M. Jaffe and J. Singh, "Band structure and charge control studies of n- and p-type pseudomorphic modulation-doped field-effect transistors," *J. Appl. Phys.*, vol. 65, no. 1, pp. 329–338, 1989.

Saba Siadat Mousavi received the B.A.Sc. and M.A.Sc. degrees in electrical engineering from the University of Ottawa, Ottawa, ON, Canada, where she is currently working toward the Ph.D. degree in electrical engineering. Her research interests include plasmonic nanoantennas and their applications.

Andreas Stöhr (M'97) received the Dipl.Ing. and Dr. Ing. degrees in electrical engineering from Gerhard-Mercator-University, Duisburg, Germany, in 1991 and 1997, respectively. From 1987 until 1996, he was the CEO of MS Steuerungsanlagen GmbH in Germany. From 1996 until 2013, he was a Research Scientist at University Duisburg. During that period, in 1998 and 1999 he also joined the Communications Research Laboratory, Tokyo, Japan, where he worked on 60-GHz wireless systems employing radio over fiber techniques. He also worked with France Telecom Orange Labs, Lannion, France, in 2009, and with Corning in 2015. He is currently a Visiting Professor at the University of Ottawa, Ottawa, ON, Canada. Since 2011, he has been a Professor and the Head of the Department of Optoelectronics, Center for Semiconductor Technology and Optoelectronics (ZHO), University Duisburg-Essen (UDE).

His current research interests include III/V integrated microwave photonic device technology and RF photonic integration technologies for millimeter wave and THz communications, measurement systems as well as sensing applications. He has published more than 200 papers in refereed journals and conferences. He is a Senior Member in IEEE Photonics and MTT society, a Committee Member and the Chair of a number of international conferences and IEEE/OSA guest editor.

Pierre Berini (F'11) received the B.E.Sc. and B.Sc. degrees in electrical engineering and computer science from the University of Western Ontario, London, ON, Canada, and the M.Sc.A. and Ph.D. degrees in electrical engineering from Ecole Polytechnique de Montreal, Montreal, QC, Canada, respectively. He is a Professor of electrical engineering, a Professor of physics, the University Research Chair in Surface Plasmon Photonics, and the Director of the Centre for Research in Photonics, University of Ottawa. He was the Founder and Chief Technology Officer of a venture capital financed company and he collaborates on an ongoing basis with industry. He received an NSERC Steacie Fellowship, an NSERC Discovery Accelerator, a Premier of Ontario Research Excellence Award, the University of Ottawa Young Researcher of the Year Award, an URSI Young Scientist Award, a George S. Glinski Award for Excellence in Research, and a Canada Foundation for Innovation Researcher. He is a Fellow of the Canadian Academy of Engineering, and is a Member of the OSA and SPIE. He has published seven book chapters, approximately 300 scientific and 685 technical papers in peer-reviewed periodicals and conference proceedings, and is an inventor or coinventor of 20 patents. He was an Associate Editor of *Optics Express* and is currently an Associate Editor of *Nanophotonics*. He contributes on an ongoing basis to the organization of several international conferences in photonics. His research interests span many areas of optics and photonics, with surface plasmons and their applications being of particular interest.

CHA1PTER 4. NANOFABRICATION

4.1. Summary

This chapter presents details of the fabrication process developed for plasmonic intensity modulators, following [56] and [55]. The fabrication process flow, while mainly adapted from [56], was successfully modified to improve the device performance, as well as the repeatability and resolution of fabricated patterns. Devices were fabricated in five stages: plasmonic surfaces, ohmic contacts, nanogratings, interlayer dielectric, and contact metals. These stages will be discussed briefly in this chapter. Current-voltage and capacitance-voltage measurements were done to characterize the electrical properties of the devices. The electrical bandwidth of these modulators is expected to reach 100's of GHz.

4.2. Contributions

I fabricated the plasmonic surfaces in the first stage. The second stage comprised some photolithography steps, which were done by me, and a BOE wet etch process, which was done by Anthony Olivieri and Ewa Lisicka-Skrzek, since it required handling acids. The third stage was fabrication of nanogratings, using electron beam (e-beam) lithography, which was performed by Anthony Olivieri. I carried out the fourth stage of fabrication, which was the deposition of an interlayer dielectric polymer and clearing via holes through this layer. The fifth, and final, stage was the fabrication of electric contacts, which was done by me.

4.3. Article

The manuscript accepted for publication in the Journal of AIP Advances follows.

Fabrication of a high-speed plasmonic reflection/transmission modulator

Saba Siadat Mousavi,[1,2] Anthony Olivieri,[1,2] and Pierre Berini[1,2,3,a)]

[1]School of Electrical Engineering and Computer Science, University of Ottawa K1N 6N5, Canada
[2]Centre for Research in Photonics, University of Ottawa, Ottawa K1N 6N5, Canada
[3]Department of Physics, University of Ottawa, Ottawa K1N 6N5, Canada

[a)] Electronic mail:berini@eecs.uottawa.ca

The fabrication of a high-speed plasmonic reflection/transmission modulator for operation at λ_0 = 1550 nm is presented and described in detail. Front-side ground and signal contacts provide easy electrical probe access to the device, while allowing transmission of light through the substrate. Modulation is based on enhanced perturbation of the effective refractive index of grating-coupled surface plasmon polaritons propagating along a metal–oxide–semiconductor structure on silicon. Fabrication steps include deposition of a plasmonic metal patch, deposition of ohmic contacts, deposition of a Au nanograting coupler overlaid by e-beam lithography, and the application of an intermetal dielectric layer with metalized vias and metal electrical contacts. Current-voltage and capacitance-voltage characteristics verify the electrical integrity of the structure.

I. INTRODUCTION

Ever-increasing demands for high-speed data throughput in telecommunications systems motivate high-speed, broadband optical modulators in integrated electro-optic systems. Optical modulators based on Si photonics but exploiting 2D materials and/or

plasmonics are particularly promising due to their potential to overcome speed, bandwidth, and size trade-offs, and their integration possibilities with ancillary electronics [1]– [14], [29].

Surface plasmon modulators, in particular, can provide fast and effective modulation by using surface plasmon-polaritons (SPPs) and their unique properties in interacting with the modulation medium. The dispersive properties of metals (*e.g.* Au, Ag) at optical frequencies, particularly pertaining to the negative real part of permittivity, are such that visible and near-infrared light may excite SPPs at the surface. SPPs are transverse-magnetic (TM) polarized surface waves where the magnetic field is in the plane of the metal surface and is perpendicular to the direction of propagation which is along the metal-dielectric interface. The electric field has a weak longitudinal component but a strong transverse component perpendicular to the metal surface. This field component is enhanced relative to the exciting field, which leads to strong light-matter interaction – a motivating factor for the use of SPPs in electro-optic modulators. The wavenumber of an SPP is larger than that of freely propagating light, which makes direct coupling impossible. Coupling occurs in the presence of, *e.g.*, surface roughness, gratings or prisms, which can manipulate the wavenumber of the incident radiation to match that of the SPP.

Active control of SPPs, either electrically, optically, or via temperature, forms the basis for plasmonic switches and modulators. Notably, the metallic structures supporting SPPs can also be used as device driving electrodes (current, voltage) to ensure strong overlap between the SPP fields and the active region of the device. Modulation depths of 80% and switching times on the order of picoseconds are

expected [16]. Plasmonic devices also fulfill the need for achieving a nanometer-scale footprint, high-speed, and low power consumption [17], [18], [30], [31]. Depending on the property of light that is modulated, optical modulators are classified into different categories, such as amplitude modulators [6], phase modulators [5], [7], [9], [13], polarization modulators [13], and wavelength modulators [5]. The performance of a modulator is usually characterized by modulation speed, modulation depth, operating wavelength, power consumption, and insertion loss.

This paper describes the fabrication and realisation of a plasmonic modulator that can operate at high speed in reflection and/or transmission. The paper builds on previous work [14], [19], by introducing front-side coplanar electrodes compatible with microwave probes, thereby enabling high-speed operation and operation in transmission, and by introducing an inter-metal dielectric to reduce parasitic contact pad capacitance. Nanofabrication methods are discussed in detail along with fabrication results.

II. EXPERIMENTAL

A. Device structure and operating principle

A plasmonic reflection/transmission modulator, designed as a metal-oxide-semiconductor (MOS) capacitor, excited by a *p*-polarized continuous-wave optical beam, and electrically controlled by an external electrical drive signal, is shown schematically in Fig. 1. The device consists of a circular thin Au film (t = 20 nm) as a "plasmonic patch", on a thin (d = 5 nm) hafnium dioxide (HfO_2) layer, grown on a highly p-doped Si substrate via atomic layer deposition (ALD). The thin Au film, the HfO_2 layer, and the p-doped Si substrate form a metal-oxide-semiconductor (MOS)

structure. Au ridges forming a grating are overlaid on the thin Au film and couple the incident light to SPPs propagating along the bottom surface of the Au film, with SPP fields penetrating through the HfO_2 and into the Si substrate. The grating pitch varies from Λ = 420 to 480 nm. Its duty cycle is fixed to D = 50%, and the thickness of the ridges to H = 80 nm. The gratings are covered by a 700 nm thick layer of polymer (SU-8) as a transparent dielectric interlayer. In order to access the thin Au film electrically, a metal contact pad is deposited on the dielectric layer. A via extends from the metal contact pad through the SU-8 layer down to a small finger extending from the thin Au film. A pair of separate larger vias extend from the top contact metal layer down to ohmic pads in direct contact to the silicon surface, which act as the ground terminals of the device. The layout of the contact pads was selected to accommodate the standard pitch of microwave ground-signal-ground (GSG) probes. The dielectric interlayer ensures that the electrical performance of the MOS capacitor is not compromised by excess parasitic capacitance from the contact metal to the silicon substrate, and it passivates the device.

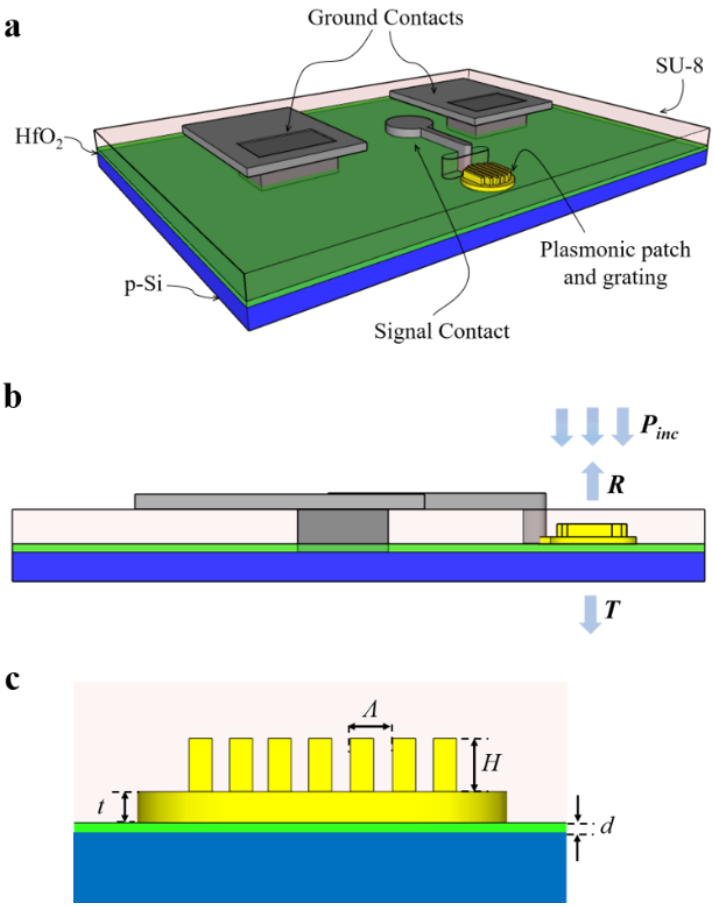

Fig. 1. Schematic diagram of the reflection/transmission modulator: (a) Isometric view, (b) cross-sectional view along the GSG contacts, (c) cross-sectional view through the grating.

A nanoscale grating structure (1st order) is used to couple the incident light to the SPPs propagating along the MOS interfaces. Tightly confined SPPs are sensitive to refractive index changes in the modulation region in the semiconductor along the oxide surface. Accumulation and depletion of carriers in this region, following the application of an external electrical signal to the MOS capacitor, changes the effective refractive index of the SPPs. This in turn alters the coupling efficiency of SPPs, and thus the modulation intensity of the reflected and transmitted light. In essence, the voltage

applied at the electrical contacts is used to modify the reflectance and transmittance of the device, which permits intensity modulation of the reflected and transmitted beams.

Numerical methods were used to model a cross-section of the device and determine the design parameters of the structure for operation at λ_0 = 1550 nm [20]. An optimized design maximises the sensitivity of the SPPs to refractive index perturbations in a thin region of semiconductor along the oxide, and thus the coupling efficiency of the incident beam to the SPPs. Computations via the transfer matrix method reveal that as the thickness of the oxide and metal decrease, the differential refractive index, Δn_{eff}, increases due to increasing overlap of the SPP fields with the perturbation. Taking into account fabrication limitations, the thickness of the metal was chosen as t = 20 nm and the thickness of oxide was chosen as d = 5 nm.

Using the momentum conservation equation for gratings

$$\Lambda = \frac{\lambda_0}{n_{eff}} \tag{20}$$

an initial value for the grating period was obtained as Λ = 420-480 nm, where broadside excitation along the surface normal and a first-order grating are assumed. The 2D finite-difference time-domain (FDTD) method was used to model the optical performance of the structure, yielding the reflectance, transmittance, absorptance and grating coupling coefficient [20].

B. Fabrication process flow

Figure 2 shows the flow of the major fabrication process steps that were developed and applied. In summary, single-side polished, prime grade, p-doped Si wafers were purchased. A thin film of HfO_2 was deposited on the wafers by atomic layer deposition

(Namlab GmbH, Dresden, Germany). Thin Au films (plasmonic patches) were deposited on HfO_2, followed by etching of HfO_2 in select locations to enable Pt-capped Al ohmic contacts to the p-Si substrate as ground terminals. Au nanogratings were then defined by e-beam lithography, evaporation and lift-off, centered on the plasmonic patches. A film of SU-8 2000.5 was deposited as the dielectric layer and via holes were created through the film of SU-8 by exposure and development. Finally, Cu/Cr contact pads were deposited on SU-8 and the vias were metallised to provide electrical access to signal and ground contacts on the device.

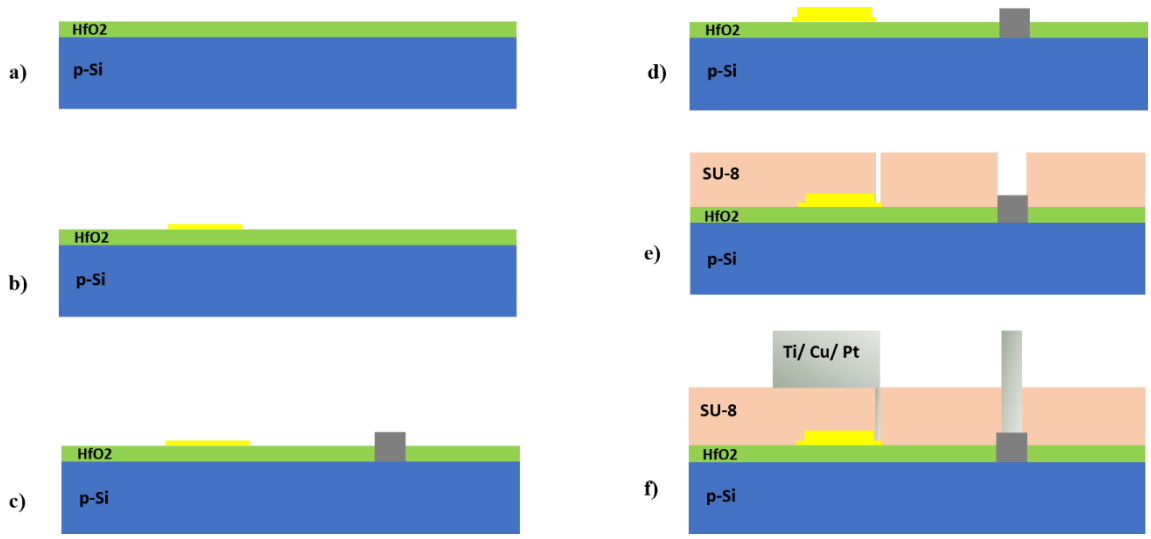

Fig. 2. Fabrication process flow of plasmonic intensity modulators: a) p-doped silicon wafer bearing a film of HfO_2 grown by ALD; b) deposition of thin Au film acting as the plasmonic patch; c) etching of windows in HfO_2 and deposition of Pt/Al ohmic contacts; d) e-beam exposure and lift-off of Au nanogratings; e) deposition of intermetal dielectric layer (SU-8) and via creation; f) deposition of via metal and contact pads.

1. Plasmonic patch

A bilayer lift-off process was carried out to define Au circular plasmonic patches of various diameters (3, 5, 11, 17, 22, 28, 56 µm) on HfO_2. For this purpose, a 2-inch p-

doped Si wafer (P/B, ρ = 0.001- 0.005 Ω.cm) was used, with a 5 ± 0.1 nm thick film of HfO_2 on its surface. The step-by-step process is summarized in Table 1 and Table 2.

Microchem LOR 1A lift-off resist was spun on the substrate surface (HfO_2) at 6000 RPM and baked at 190° C for 5 min. Megaposit SPR 955-CM photoresist (PR) was then spun on LOR 1A at 6000 RPM and baked at 90° C for 3 min. Afterwards, the bilayer resist stack was exposed to 84 mJ of UV light, using a photomask in an OAI Model 204IR mask aligner. Three I-line (λ_0 = 365 nm) bandpass filters were used to optimize pattern resolution. Using a stack of three filters allowed for more precise control over the exposure energy. The sample was then immersed in developer Microposit MF CD-26 for 60 s. A 20 nm thick layer of Au was deposited on a 0.3 nm thick layer of Ti used to promote adhesion. Both metals were deposited via thermal evaporation using an Angstrom Nexdep Evaporator. The unexposed parts of the bilayer resist stack, along with the Au film covering them, were lifted-off by immersing the sample in Remover PG (Microchem).

TABLE 1. Procedure for applying lift-off resist LOR 1A.

Step		Description
1	Load substrate and dispense LOR 1A	Ensure that wafer is centered and leveled on the spinner chuck. Dispense 5 mL of LOR 1A using a pipette.
2	Saturation	Let LOR 1A level out on the substrate for 30 s, while the spinner's chuck vacuum is disengaged.
3	Spin coating	15 s @ 200 rpm/s, 10 s @ 3000 rpm, 15 s @ 200 rpm/s, 30 s @ 6000 rpm, 6 s @ -1000 rpm/s
4	Relaxation	Disengage the vacuum and let the substrate relax for 30 s before moving it.

5	Soft bake	Place the substrate on a flat hotplate at 190 °C for 5 min to evaporate solvent in LOR 1A and harden the resist. Let the substrate cool for several minutes before the next step.

TABLE 2. Procedure for photolithography, metallization, and lift-off of the circular thin Au films used as plasmonic patches.

Step		Description
1	Load substrate and dispense SPR 955	Reload the substrate into the spinner and dispense 5 mL of photoresist using a pipette.
2	Saturation	Let SPR 955 level out on the substrate for 10 s, while the spinner's chuck vacuum is disengaged.
3	Spin coating	15 s @ 200 rpm/s, 10 s @ 3000 rpm, 15 s @ 200 rpm/s, 30 s @ 6000 rpm, 6 s @ -1000 rpm/s
4	Relaxation	Disengage the vacuum and let the substrate relax for 10 s before moving it.
5	Soft bake	Place the substrate on a flat hotplate at 100 °C for 3 min to evaporate the solvent and harden the PR. Let the substrate cool for several minutes before the next step
6	Exposure	Expose the PR with 84 mJ UV light. Three I-line filters were used to filter out UV wavelengths other than 365 nm.
7	Development	Fully immerse the sample in a fresh bath of CD-26 developer for 60 s. Rinse thoroughly in DI water, and blow dry with N_2.
8	Evaporation	0.3 nm Ti adhesion layer, e-beam evaporation at 0.1 Å/s; 20 nm Au, thermal evaporation at 0.5 Å/s, on Ti without breaking vacuum.
9	Lift-off	Immerse the sample into a bath of Remover PG for 20 min. Rinse thoroughly in Acetone, IPA and DI water. Blow dry with N_2.

2. Ohmic Contacts

To establish ohmic contacts, metal must be in direct contact with the semiconductor substrate. Therefore, a buffered oxide etch (BOE) process was practiced to open windows in HfO_2 to let the ohmic metal make direct contact with the p-Si substrate. A bilayer stack of SPR 955 on LOR 1A was utilized to etch the windows in HfO_2, then lift-off the evaporated ohmic metals in self-aligned processes (*i.e.*, the same lithography stack was used to carry out both processes). The PR was exposed, using the same parameters as in Table 2 and developed by immersion into MF CD-26 for 60 s. At this stage the sample surface was covered with the bilayer resist stack except for the openings where HfO_2 must be etched in order to realize the Ohmic contacts. The sample was then immersed in a 5% diluted hydrofluoric (HF) acid solution for 30 s, followed by a DI H_2O rinse and N_2 blow. The sample was then immediately loaded into the evaporator. Al was chosen as the ohmic contact metal due to its high conductivity, low ohmic contact resistance to p-type Si, and good adhesion to Si [21]. Thus, a 100 nm thick layer of Al, capped by a 20 nm thick layer of Pt, was e-beam evaporated, and lifted-off. Pt was introduced as a conductive protective layer of Al during subsequent process steps. See supplementary material at [URL will be inserted by AIP Publishing] for the process flow and a summary of steps for the realization of ohmic contacts.

3. Nanogratings

Nanogratings were deposited on the plasmonic surfaces in order to achieve coupling of the incident beam to SPPs. The dimensions of the nanogratings were selected to optimize the optical performance of the modulators (Λ = 420 - 480 nm, H = 80 nm, D = 50%) [20]. The size of the Au plasmonic patches affects the metal-oxide-semiconductor (MOS) capacitance of the structures and the electrical bandwidth of the

devices. Thus, it is important for the patch and grating dimensions to be as close as possible to the designed dimensions. Electron beam lithography, overlaid to the existing plasmonic patches, was used to produce gratings of high resolution. Au deposition was then done by thermal evaporation, followed by lift-off. The protocol for the fabrication of nanogratings is summarized in Table 3.

TABLE 3. Procedure for e-beam lithography, metallization and lift-off of the nanogratings.

	Step	Description
1	Load substrate and dispense PMMA 495	Dispense 5 mL of PMMA 495 6% using a pipette.
2	Saturation	Let PMMA 495 level out on the substrate for 1 min, while vacuum is disengaged.
3	Spin coating	5 s @ 200 rpm/s, 3 s @ 400 rpm/s, 90 s @ 2000 rpm, 5 s @ -400 rpm/s
4	Soft bake	Place the substrate on a flat hotplate at 190 °C for 1 hour to evaporate solvent and harden the resist. Let the substrate cool for several minutes before the next step.
5	Repeat steps 1- 4 to spin-coat PMMA 950	
6	E-beam exposure	e-beam acceleration voltage: 30 kV, dosage: 300 μC/cm^2
7	Development	Immerse the sample into a bath of MIBK developer at 20 °C for 2 min. Rinse in DI water and blow dry with N_2.
8	Descum	To remove residual PMMA O_2 RIE was performed for 1 min at 25 W.
9	Evaporation	80 nm Au thermally evaporated at 0.5 Å/s.
10	Lift-off	Immerse the sample in Acetone for 20 min. Rinse in IPA and DI water. Blow dry with N_2.

4. Dielectric interlayer

A dielectric interlayer is required between the top layer contact metal and the device level to ensure that the electrical performance of the MOS capacitor is not compromised by excess parasitic capacitance to the silicon substrate. It also provides some measure of protection to the grating surface. SU-8 2000.5 was selected for this purpose.

When exposed to UV light and subsequently baked, SU-8 becomes cross-linked and insoluble to liquid developers. SU-8 has very high optical transparency at wavelengths above λ_0 = 360 nm, which makes it optically non-invasive to the performance of our devices. SU-8 is best suited for permanent applications where it is imaged, developed, cured and retained. It comes in different viscosities. The least viscous type, SU-8 2000.5, was used for this application, since the desired film thickness, according to the numerical optimization [20], is ~700 nm, which can be achieved with SU-8 2000.5.

The first step was spin coating SU-8 2000.5 on the wafer bearing the plasmonic patches and gratings, as well as the ohmic pads. The second step was a soft bake (SB), where the sample was placed on a hotplate at 65 °C for 5 min. A temperature ramp-up was then used to reach 95 °C, where the sample was left for another 5 min. A ramp-down was then performed to bring the sample temperature to 50 °C, bringing the total time on the hotplate to 25 min. The soft bake is necessary to remove solvent from the SU-8. Temperatures, bake-times, and ramps are important to reduce stress formation and cracks in the film.

After a 10 min cool-down SU-8 was exposed to UV light, using a bright-field photomask to define via holes in the layer. An I-line filter was used to increase the pattern resolution. As a result, SU-8 cross-linking was initiated where exposed. In order

to avoid over-exposure around the edges, SU-8 was exposed with the minimum required energy. Over-exposure causes SU-8 to cross-link around the edges, where it is expected to be removed and thus perturbs dimensions from what is desired. Hard contact was used in the mask aligner in order to minimize the gap between the photomask and the sample.

Post exposure bake (PEB) is necessary to complete the cross-linking of the SU-8. The sample was placed on a perfectly flat hot plate set at 90 °C for 1 min. The temperature was then ramped up to 105 °C, and the sample baked for 1 min. The hot plate temperature was then ramped down to 75 °C, at which point the sample was removed from the hot plate.

SU-8 developer was used to develop the sample. The best results were achieved using vertical positioning of the wafer in the developer for 3 min, followed by two intervals of 10 s in an ultrasonic bath at 80 KHz and 30% power. Hard bake (HB) is required if further processes are to be done on SU-8, which is the case here. Thus, the sample was baked for 30 min on a hot plate at 120 °C. The sample was then flood exposed for 30 s, to improve the adhesion of the next layer metal to SU-8. See supplementary material at [URL will be inserted by AIP Publishing] for the fabrication flow and a summary of steps in processing the SU-8 layer.

It was found that defining via holes in SU-8 depends significantly on the size and geometry of the holes, as well as the thickness of the SU-8 layer. For a circular hole, if its diameter is small and the thickness of SU-8 layer is larger than the diameter, a bright spot ("Poisson spot" [22]) forms in the layer of SU-8 during exposure due to diffraction. Specifically, when a bright-field photomask bearing dark disks is

illuminated with a collimated exposure beam from the mask aligner, a bright spot appears behind the centre of the dark disks on the photomask within the SU-8 layer where the via holes are to be developed. The bright spot, located along the central axis of the dark disk, inside the SU-8, causes activation of the photoactive component therein and only partial development of the hole. The result was observed as pillars formed at the centre of the smallest diameter via holes.

Numerical modelling was carried out using Lumerical FDTD to verify this observation. A layer of SU-8 (700 nm thick) was assumed on a 1.5 μm thick Si substrate. A glass mask with 100 nm thick Cr masking disks of various diameters covered the SU-8 layer. The structure was illuminated from the top with a plane wave of wavelength 365 nm. The transmittance through a 100 nm film of Cr at the wavelength of interest was close to zero. Fig. 3 plots the distribution of the electric field magnitude over the cross-section of SU-8 layer and the masking disks. An interference pattern is visible throughout the SU-8 layer, due to the use of coherent illumination in the modelling – the actual illumination in the mask aligner is filtered but incoherent, so a well-defined interference pattern is not expected. The field diffracted below the mask disk clearly forms a Poisson's spot, observed to vary with disk diameter and becoming significant for a small diameter. The Poisson's spot will not be as well-defined using the mask aligner, again due to incoherent illumination, but will nonetheless be present.

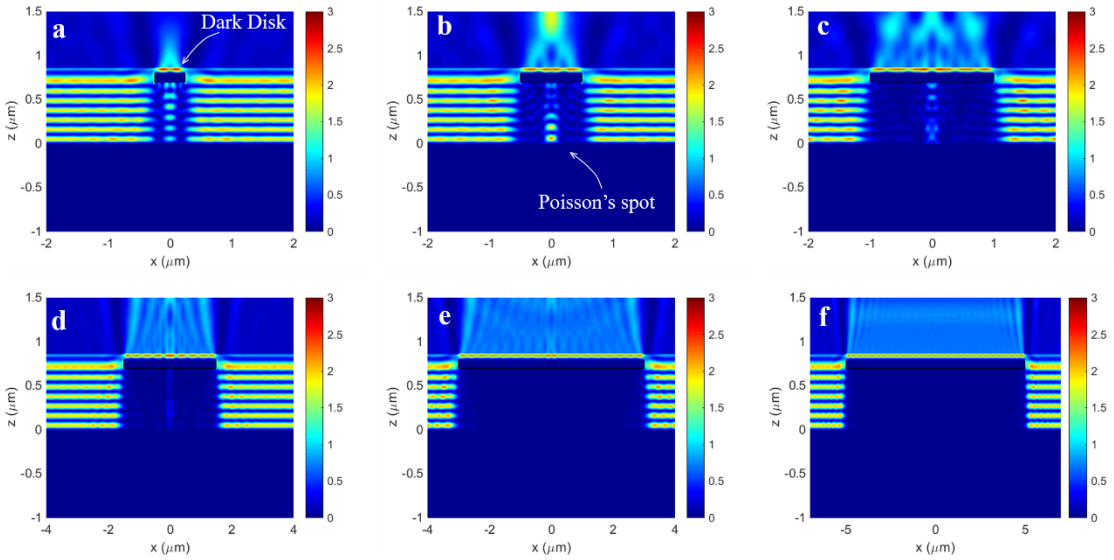

Fig. 3. Magnitude of the electric field distribution at λ_0 = 365 nm, over vertical cross-sections through a 700 nm SU-8 layer and mask disks of varying diameter q; (a) q = 0.5 µm, (b) q = 1 µm, (c) q = 2 µm, (d) q = 3 µm, (e) q = 6 µm, (f) 10×4 µm trench.

In another set of simulations, the same structure was studied, while the disk diameter was fixed to 2 µm, and the thickness of the SU-8 layer was varied from 0.4 to 3 µm. The Poisson's spot becomes more evident as the thickness of the SU-8 layer increases, as shown in Fig. 4. Fig. 4(d) plots the field distribution for the case where the Si layer is removed and replaced with an absorbing boundary condition (perfectly matched layer) in order to visualise the distribution without interference.

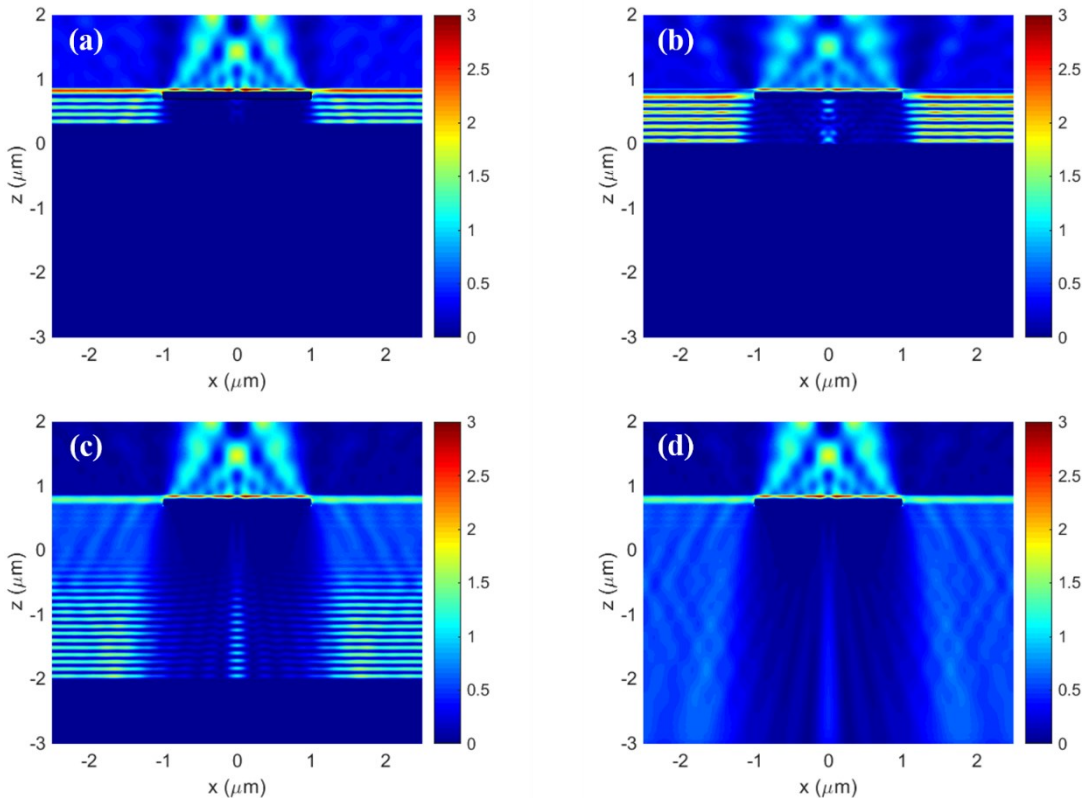

Fig. 4. Magnitude of the electric field distribution at λ_0 = 365 nm, over vertical cross-sections through a SU-8 layer of varying thickness *h* and mask disks 2 μm in diameter; (a) *h* = 0.4 μm, (b) *h* = 0.7 μm, (c) *h* = 2.7 μm, (d) *h* = 3 μm with the Si substrate replaced with an absorbing boundary condition.

It is evident from these computations (Figs. 3 and 4), that disk diameters less than 2 μm can produce a significant Poisson's spot in SU-8 layers that are about 700 μm or thicker. Thus, our via hole design was selected as a rectangular trench of dimensions 4 μm × 10 μm.

5. Contact layer

In order to facilitate electrical probing of devices and metallise the via holes, relatively thick, large contact pads are required. A bilayer lift-off process was used to fabricate

signal and ground contact pads. This process was, in principal, very similar to that used to create the Au plasmonic patches. However, LOR 1A was replaced by the much more viscous LOR 10B lift-off resist, to provide a thicker resist stack enabling lift-off of a thick contact metal layer. It is important to carefully align this layer in order to metallise all via holes and achieve contact to the modulator devices and their ohmic contacts. See supplementary material at [URL will be inserted by AIP Publishing] for the process flow and a summary of steps.

III. RESULTS AND DISCUSSION

A. Plasmonic patch and Ohmic contact

Following the steps in Sec.II.B, plasmonic modulator devices were successfully fabricated. Fig. 5(a) shows a microscope image of a plasmonic patch along with its nearby ohmic contact pads, metallized and lifted-off. Alignment-marks are also shown in this image. These are to facilitate alignment of a perpendicularly incident optical beam to the nanogratings during optical tests. Fig. 5(b) and (c) show an Atomic force microscopy (AFM) scan of an ohmic contact pad and its associated cross-sectional analysis, respectively. The Al ohmic pads were 93.4 nm thick, with a root mean square (rms) roughness of 1.5 nm, while the target thickness was 100 nm. However, this small difference may be neglected as it does not affect the performance of the device.

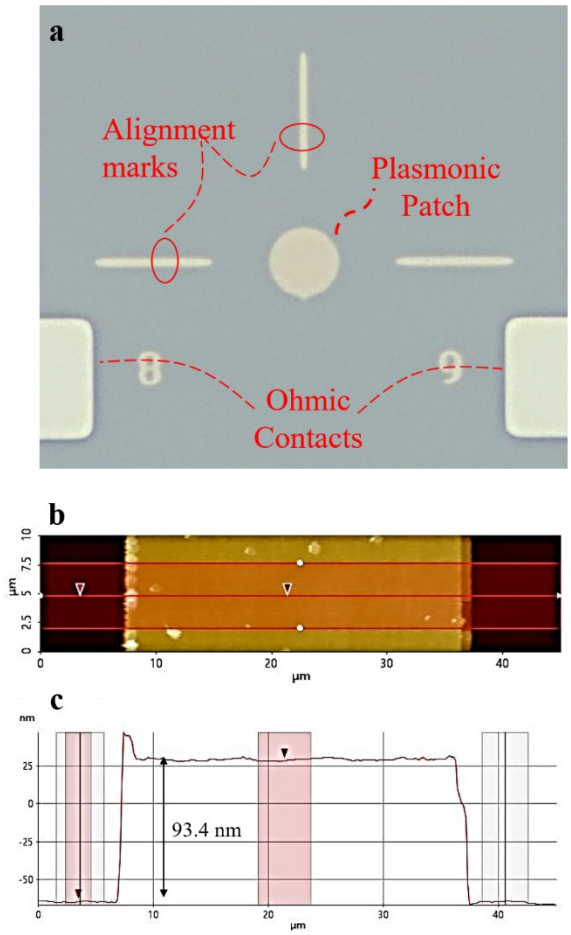

Fig. 5. (a) microscope image of a representative metallised plasmonic patch and its nearby ohmic contact pad; (b) AFM scan of a 30 µm wide Al ohmic contact on Si, and (b) its cross-sectional analysis.

B. Analysis of the Ohmic contacts

At this stage, current-voltage (I-V) measurements were performed to verify the ohmic contacts to the semiconductor. For this purpose, external probes were used to apply voltage to an arbitrary ohmic pad, while another nearby pad was grounded. The applied voltage was varied between -1 and 1 V, in steps of 0.1 V and the resulting current was measured for various signal to ground pad distances (the maximum current allowed was set to 100 mA). A typical I-V characteristic is shown in Fig. 7, where a linear relationship between current and voltage is observed, characteristic of an ohmic contact.

The resistance, taken as the inverse slope of this I-V characteristic, is 9.66 Ω, obtained for two adjacent ohmic pads, 225 μm apart, centre to centre. A probing test was performed to determine the resistance of the test setup by directly connecting the two test probes. The set-up resistance was measured to be 5.3 Ω, which was used as a calibration value. Thus, the total contact resistance is 4.36 Ω.

An equivalent circuit model of the contacts, shown in the inset of Fig.7, was used to represent measurements between an adjacent pair of contact pads. The total series resistance of this model, R_s, is:

$$R_s = 2(R_{Pt} + R_{Al} + R_{Al-Si}) + R_{Si} \quad (2)$$

The resistance of a single Al/p-Si ohmic contact is denoted $R_{Al\text{-}Si}$. All other resistances in this model (R_{Pt}, R_{Al}, R_{Si}) are computed using:

$$R = \rho \frac{L}{A_{res}} \quad (3)$$

where ρ is the resistivity of the material for which the resistance is computed, L is the length of the current path, and A_{res} is the surface area of the contact pads, the design of which produces an area of 4×10^3 μm^2. The bulk resistivity of Pt and Al was used, along with ρ = 0.005 Ω.cm for heavily p-doped Si. Setting R_s to the total measured value (R_s = 4.36 Ω) and isolating yields the contact resistance at the Al/p-Si interface as $R_{Al\text{-}Si}$ = 0.77 Ω.

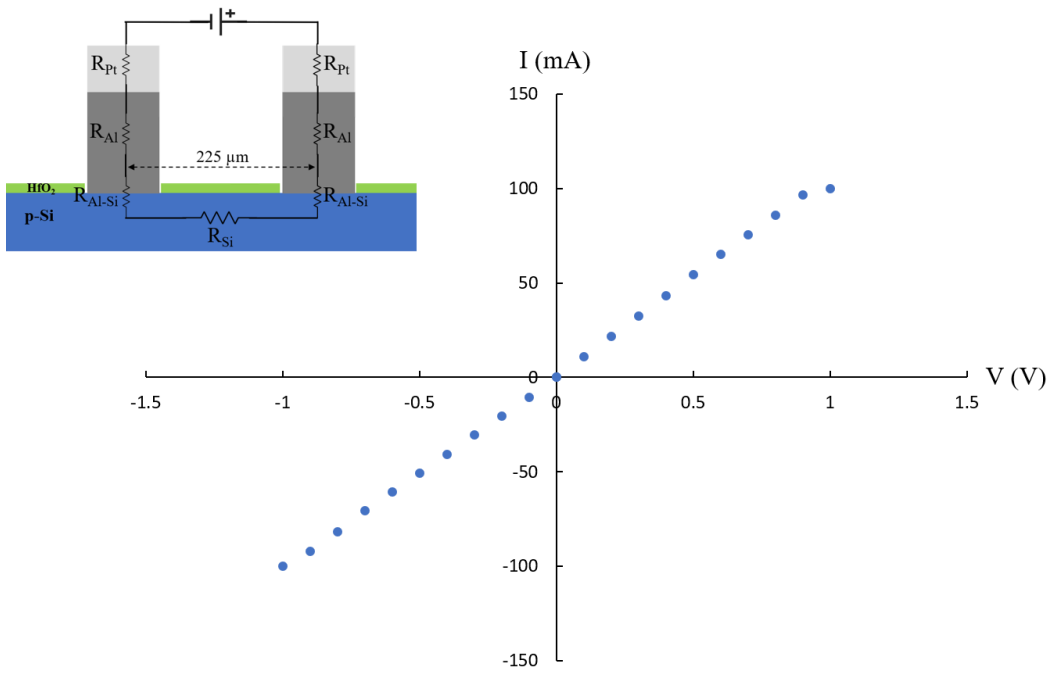

Fig. 7. Typical I-V measurements between a pair of ohmic pads. Inset shows an equivalent circuit model for ohmic contacts, where R_{Pt} is the resistance of the Pt layer, R_{Al} is the resistance of the Al layer, $R_{Al\text{-}Si}$ is the contact resistance of the Al-Si interface, and R_{Si} is the resistance of the semiconductor.

C. Analysis of the MOS capacitor

Capacitance-voltage (C-V) measurements were also obtained to characterize the electronic performance of the MOS capacitors formed as the Au-HfO_2-Si structure. Fig. 8 gives a typical high-frequency measurement for a plasmonic patch and a neighbouring ground contact, obtained using an HP 4284a LCR meter at a frequency of f = 1 MHz and applied AC voltage of v_s = 5 mV (maximum value). The accumulation capacitance C_{ox} is determined from Fig. 8 to be 40 pF. The diameter of the circular plasmonic patch probed was 56 µm. Therefore, given the thickness of the oxide layer as d ~ 5 nm

(verified independently by AFM), the relative permittivity of the oxide layer is found to be $\varepsilon_{ox} = 9.18$ by fitting to:

$$C_{ox} = \frac{\varepsilon_0 \varepsilon_{ox} A_c}{d} \tag{4}$$

where $A_c = 9852\ \mu m^2$ is the area of the capacitor.

C_{ox} consists of the capacitance over the film of hafnia C_{HfO2} in series with a capacitance over a layer of native SiO_2, C_{SiO2}, present on Si prior to ALD of HfO_2. Assuming the thickness of the SiO_2 layer to be $t_{SiO2} \sim 1$ nm, and its relative permittivity as $\varepsilon_{SiO2} = 3.9$, using Eq. 4 yields $C_{SiO2} = 85$ pF. From this value, $C_{HfO2} = 75.5$ pF was obtained via:

$$C_{HfO2} = \frac{C_{SiO2} C_{ox}}{C_{SiO2} - C_{ox}} \tag{5}$$

Subsequently, the relative permittivity of a 4 nm HfO_2 layer was determined as ε_{HfO2} =14, following Eq. 4. This is in the range of values reported in [14], [23]–[25] (ε_{HfO2} = 12-25). Thus, ε_{ox} is taken as an equivalent relative permittivity for the oxide stack.

The maximum depletion width x_{dT} was then calculated [26] to be 8.2×10^{-7} cm for our highly p-doped Si substrate (2.13×10^{19} cm^{-3}). The minimum capacitance per unit area C'_{min} which is defined at the threshold inversion point where the maximum depletion width is reached but the inversion charge density is essentially zero, can be obtained via [26]:

$$C'_{\min} = \frac{\varepsilon_0 \varepsilon_{ox}}{d + (\varepsilon_{ox} / \varepsilon_s) x_{dT}} \tag{6}$$

where ε_s = 11.7 is the relative permittivity of the semiconductor, Si. Thus, $C_{min} = C'_{min} \times A_c$ = 17.6 pF is obtained, which agrees with the minimum capacitance measured in Fig. 8.

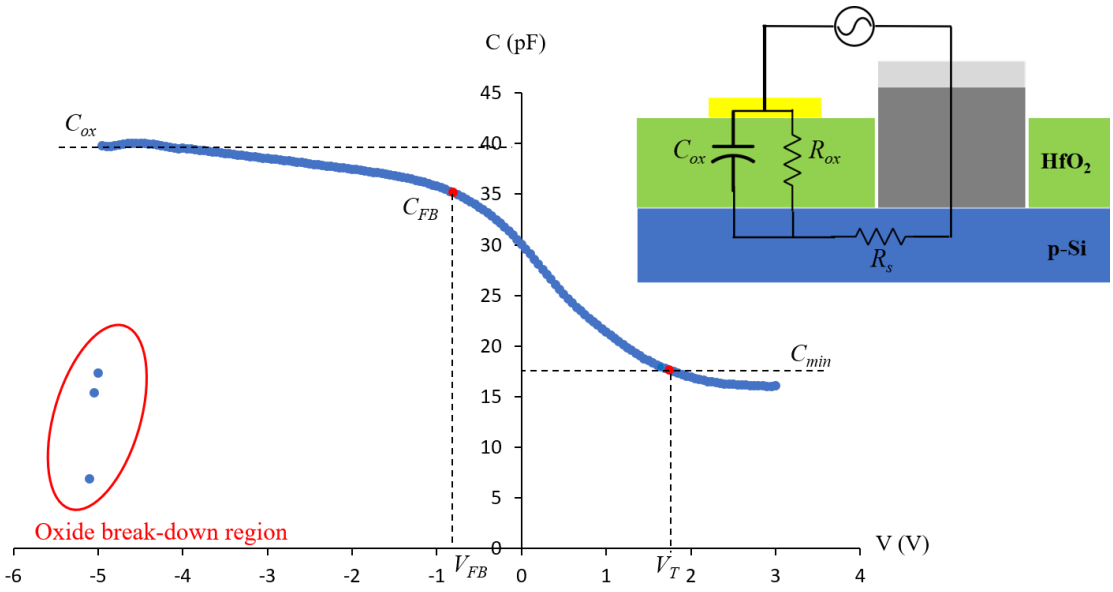

Fig. 8. C-V characteristic of a MOS capacitor formed between a Au plasmonic patch and a nearby ohmic contact of a modulator device. The inset shows an equivalent circuit model, where R_{ox} is the leakage resistance through the HfO_2 layer, and R_s is the series resistance through the semiconductor and the ohmic contact pad. A large value for R_{ox} and a small value for R_s renders these parasitic resistances negligible.

The point on the C-V curve which corresponds to the flat-band condition V_{FB} is of interest. The flat-band capacitance per unit area C'_{FB} is given by [26]

$$C'_{FB} = \frac{\varepsilon_0 \varepsilon_{ox}}{d + \left(\frac{\varepsilon_{ox}}{\varepsilon_s}\right)\sqrt{\left(\frac{kT}{e}\right)\left(\frac{\varepsilon_s}{eN_a}\right)}} \tag{7}$$

The flat-band capacitance $C_{FB} = C'_{FB} \times A_c = 35$ pF is obtained, which yields $V_{FB} = -0.8$ V, as shown on the C-V plot in Fig. 8.

The flat-band voltage was then used to determine the equivalent fixed charge per unit area Q'_{ss} in the HfO_2 layer. In an ideal oxide, with no fixed charges, the flat-band voltage is equal to the metal-semiconductor work function difference ϕ_{ms}. In a non-ideal situation, the presence of fixed charges in the oxide, introduced during fabrication, affects the flat-band voltage such that:

$$V_{FB} = \phi_{ms} - \frac{Q_{ss}}{C_{ox}} \tag{8}$$

Assuming $\phi_{ms} = -0.0016$ V results in the equivalent fixed charge $Q_{ss} = 3.14\times10^{-11}$ C. The equivalent fixed charge per unit area is then $Q'_{ss} = Q_{ss}/A_c = 1.27\times10^{-6}$ C/cm^2, which is comparable to the values reported in [27].

The C-V characteristic was driven in accumulation beyond break-down of the oxide, which occurred at a voltage of -5 V (Fig. 8). Given the thickness of the oxide layer as d = 5 nm, this yields a break-down electric field of ~10 MVcm^{-1}, which is close to the expected value for ALD HfO_2 [14], [25], [28].

The electrical bandwidth of this modulator device is given by:

$$BW = \frac{1}{2\pi(R_{sub} + R_g \,||\, R_L)C_{ox}} \tag{9}$$

where R_{sub} is the series resistance through the semiconductor and ohmic contact pads, R_g is the internal impedance of the generator driving the modulator, and R_L is the load resistance. R_g and R_L are normally assumed to be matched and set to a standard value such as 50 Ω. R_s is small compared with R_g and R_L and is neglected. The electrical bandwidth is then computed for different modulator designs. The smallest plasmonic

patch diameter that we have fabricated is 3 μm (large enough to accommodate a diffraction-limited incident optical beam), which yields the maximum bandwidth of BW = 55.4 GHz, for a 5 nm thick oxide layer.

D. Nanogratings

Nanogratings were exposed using e-beam lithography and then metallized as explained in Sec.II.B.3. An AFM scan of a grating on a 11 μm dimeter plasmonic patch is shown in Fig. 9(a). This scan yields a thickness for the grating ridges of H = 81 nm and for the plasmonic patch of t = 20 nm, which are very close to target (80 and 20 nm, respectively). The rms roughness of the plasmonic patch is in the range of 0.8 - 0.9 nm, and the rms roughness along the top of a grating ridge is 1.5 nm. Fig. 9(b) shows a scanning electron microscopy (SEM) image of a nanograting on a plasmonic patch. The ridge width and pitch of the grating were measured to be 270 nm and Λ = 460 nm, respectively.

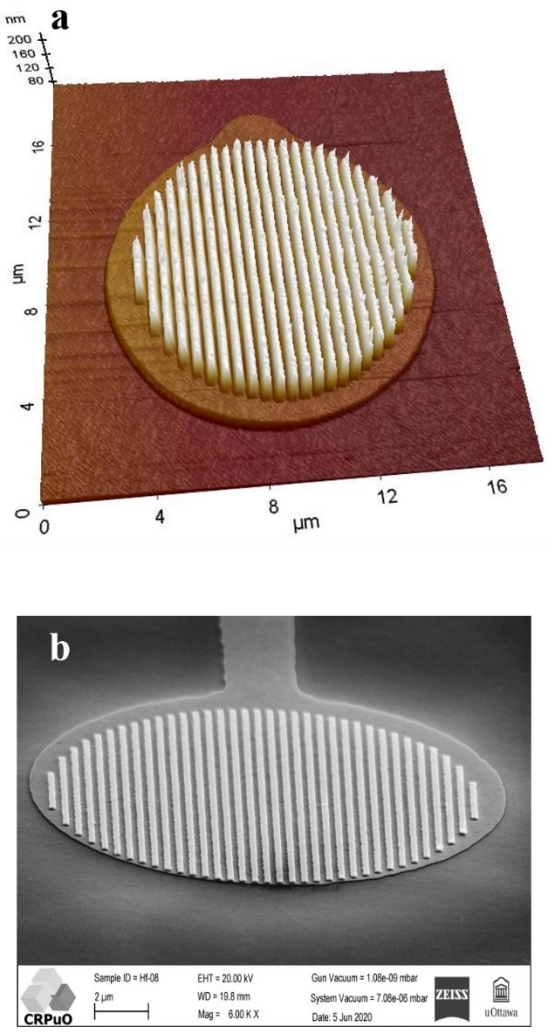

Fig. 9. (a) AFM scan of a grating on a plasmonic patch; (b) SEM image of Au nanogratings deposited on a plasmonic patch.

E. Dielectric interlayer and contact layer

The vias created following the steps in Sec.II.B.4 are shown in Fig. 10(a). In this image a via trench is overlapped with the contact tab of a plasmonic patch, while a bigger via window is opened over the ohmic contact. Fig. 10(b) shows an AFM scan of a via trench opening prior to metallization. The Au tab at the bottom of the via on top of the HfO_2 layer is visible on this AFM scan, as is the topographic contour of the

plasmonic patch. The depth of the via is measured to be 709 nm (Fig. 10(c)), which is close to the expected thickness of the SU-8 2000.5 layer (700 nm). This means the via trench was fully cleared down to the Au (plasmonic patch) and HfO_2 levels. Fig. 10(d) shows a zoomed-in view of the cross-sectional analysis in Fig.10(c), where the thickness of the tab of the plasmonic patch is measured and agrees well with the expected thickness. The rms surface roughness on the tap was 1.06 nm, which is very close to the rms surface roughness of the plasmonic patch (0.8 nm). This also verifies that the via trench was completely cleared.

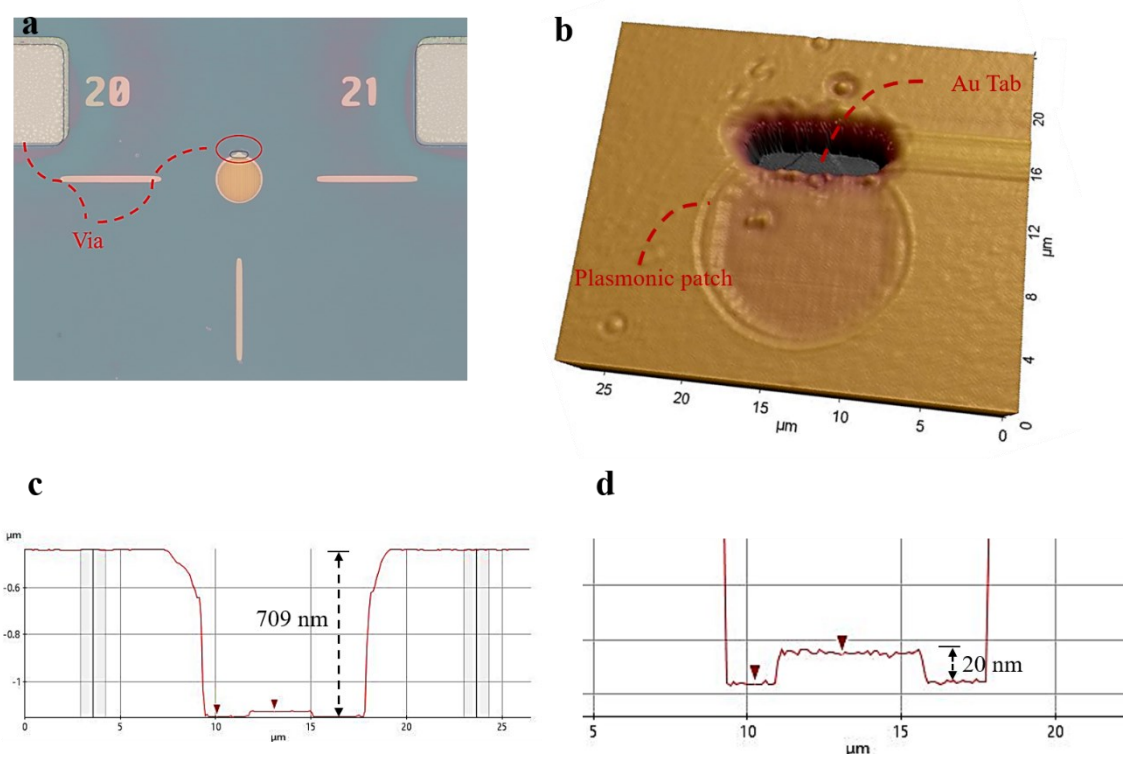

Fig. 10. (a) microscope image of an intermediate structure consisting of the plasmonic patch, the ohmic contact pad, the nanogratings, and the intermetal dielectric film with opened vias. (b) AFM scan of a via trench overlapped with a plasmonic patch. (c) cross-sectional analysis of a cut through the centre of the via. (d) a zoomed-in view of the cross-section in part (c). The tab of the plasmonic patch is clearly shown.

Finally, Fig. 11(a) shows a microscope image of the completed device, with signal and ground electric contacts. A SEM image of a completed device is also shown in Fig. 11(b). The physical continuity of the metal signal path down to the plasmonic patch is verified in Fig. 11(c).

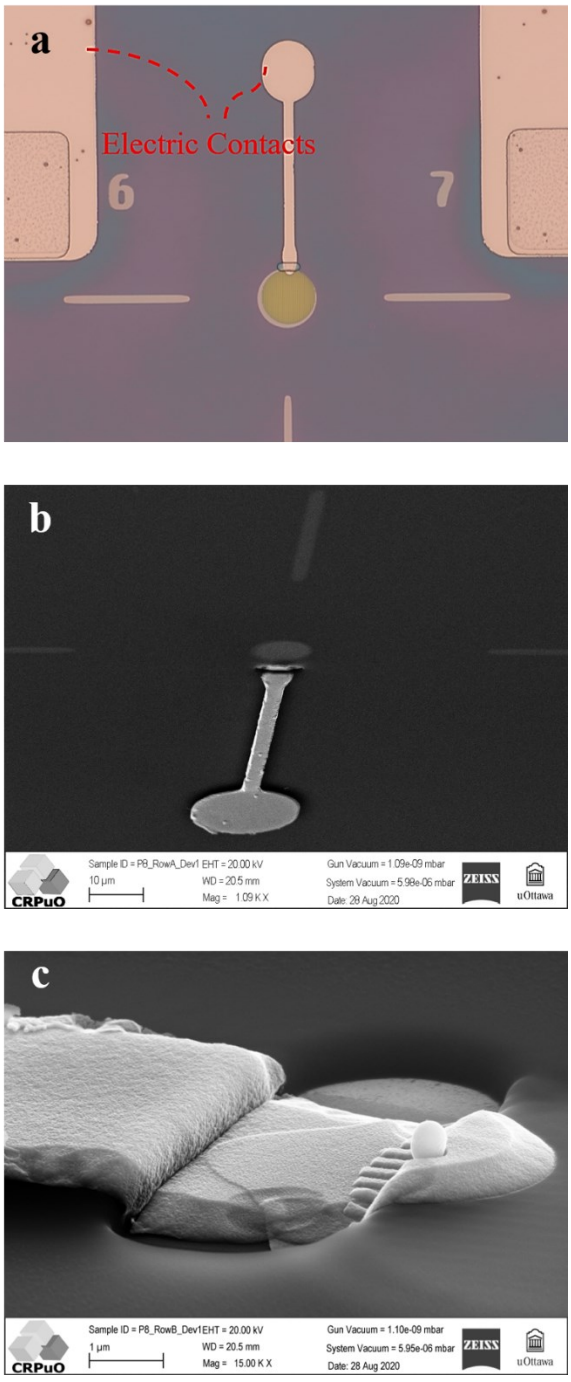

Fig. 11. (a) a complete device with ground and signal contacts. (b) SEM image of a complete device: signal contact pad and arm connected to a via trench. (c) Close-up SEM image of a metalized via, where the signal arm is connected to the underlying plasmonic patch. The plasmonic patch bearing a grating under the SU-8 layer can also be observed.

IV. SUMMARY AND CONCLUSIONS

Novel plasmonic reflection/transmission intensity modulators were successfully fabricated. A MOS capacitor was exploited as the active region of the modulator device, where, a thin film of HfO_2 was sandwiched between a heavily p-doped Si substrate and the Au plasmonic patches, deposited on the oxide layer. Nanogratings are employed to couple the incident beam into the SPPs propagating along the metal-oxide-semiconductor interfaces. Large front-side ground and signal pads were fabricated to allow easy probing of the devices, without obscuring the optical beams. A film of SU8 was utilized as an intermetal dielectric layer to minimize the parasitic capacitance due to the electrical contacts. Complete fabrication details were provided, and results given for intermediate structures.

Acknowledgements:

The authors are grateful to Ewa Lisicka-Skrzek and Howard Northfield for helpful discussions.

Data availability

The data that supports the findings of this study are available within the article [and its supplementary material].

[1] A. Chen and E. Murphy, *Broadband Optical Modulators: Science, Technology, and Applications*. CRC Press, 2011.
[2] K. Liu, C. R. Ye, S. Khan, and V. J. Sorger, "Review and perspective on ultrafast wavelength-size electro-optic modulators," *Laser & Photonics Reviews*, vol. 9, no. 2, pp. 172–194, Mar. 2015, doi: 10.1002/lpor.201400219.
[3] G. T. Reed, G. Mashanovich, F. Y. Gardes, and D. J. Thomson, "Silicon optical modulators," *Nature Photonics*, vol. 4, no. 8, pp. 518–526, Aug. 2010, doi: 10.1038/nphoton.2010.179.
[4] A. Liu *et al.*, "A high-speed silicon optical modulator based on a metal–oxide–semiconductor capacitor," *Nature*, vol. 427, no. 6975, pp. 615–618, Feb. 2004, doi: 10.1038/nature02310.
[5] A. Liu *et al.*, "High-speed optical modulation based on carrier depletion in a silicon waveguide," *Optics Express*, vol. 15, no. 2, p. 660, Jan. 2007, doi: 10.1364/OE.15.000660.
[6] Y. Ding *et al.*, "Effective Electro-Optical Modulation with High Extinction Ratio by a Graphene–Silicon Microring Resonator," *Nano Lett.*, vol. 15, no. 7, pp. 4393–4400, Jul. 2015, doi: 10.1021/acs.nanolett.5b00630.
[7] S. Ye, F. Yuan, X. Zou, M. K. Shah, R. Lu, and Y. Liu, "High-Speed Optical Phase Modulator Based on Graphene-Silicon Waveguide," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 23, no. 1, pp. 76–80, Jan. 2017, doi: 10.1109/JSTQE.2016.2545238.
[8] Z. Sun, A. Martinez, and F. Wang, "Optical modulators with 2D layered materials," *Nature Photonics*, vol. 10, no. 4, pp. 227–238, Apr. 2016, doi: 10.1038/nphoton.2016.15.
[9] A. Melikyan *et al.*, "High-speed plasmonic phase modulators," *Nature Photonics*, vol. 8, no. 3, pp. 229–233, Mar. 2014, doi: 10.1038/nphoton.2014.9.
[10] C. Haffner *et al.*, "All-plasmonic Mach–Zehnder modulator enabling optical high-speed communication at the microscale," *Nature Photonics*, vol. 9, no. 8, pp. 525–528, Aug. 2015, doi: 10.1038/nphoton.2015.127.
[11] M. Z. Alam *et al.*, "Plasmonic nanophotonic modulators," in *2017 IEEE Photonics Society Summer Topical Meeting Series (SUM)*, Jul. 2017, pp. 193–194, doi: 10.1109/PHOSST.2017.8012716.
[12] B. Huang, W. Lu, Z. Liu, and S. Gao, "Low-energy high-speed plasmonic enhanced modulator using graphene," *Optics Express*, vol. 26, no. 6, p. 7358, Mar. 2018, doi: 10.1364/OE.26.007358.
[13] M. Ayata *et al.*, "High-speed plasmonic modulator in a single metal layer," *Science*, vol. 358, no. 6363, pp. 630–632, Nov. 2017, doi: 10.1126/science.aan5953.
[14] A. Olivieri, C. Chen, S. Hassan, E. Lisicka-Skrzek, R. N. Tait, and P. Berini, "Plasmonic Nanostructured Metal–Oxide–Semiconductor Reflection Modulators," *Nano Letters*, vol. 15, no. 4, pp. 2304–2311, Apr. 2015, doi: 10.1021/nl504389f.
[15] W. Heni *et al.*, "Silicon–Organic and Plasmonic–Organic Hybrid Photonics," Jun. 26, 2017. https://pubs.acs.org/doi/abs/10.1021/acsphotonics.7b00224 (accessed Feb. 01, 2019).
[16] A. V. Krasavin and A. V. Zayats, "Photonic Signal Processing on Electronic Scales: Electro-Optical Field-Effect Nanoplasmonic Modulator," *Phys. Rev. Lett.*, vol. 109, no. 5, p. 053901, Jul. 2012, doi: 10.1103/PhysRevLett.109.053901.
[17] J. A. Schuller *et al.*, "Plasmonics for extreme light concentration and manipulation," *Nat. Mater.*, vol. 9, no. 3, pp. 193–204, Mar. 2010, doi: 10.1038/nmat2630.
[18] J. S. White *et al.*, "Extraordinary optical absorption through subwavelength slits," *Opt. Lett.*, vol. 34, no. 5, p. 686, Mar. 2009, doi: 10.1364/OL.34.000686.
[19] S. Hassan *et al.*, "Fabrication of a plasmonic modulator incorporating an overlaid grating coupler," *Nanotechnology*, vol. 25, no. 49, p. 495202, Dec. 2014, doi: 10.1088/0957-4484/25/49/495202.

[20] P. S. J. Berini and C. Chen, “Metal-insulator-semiconductor devices based on surface plasmon polaritons,” US9397241B2, Jul. 19, 2016.
[21] L. E. Terry and R. W. Wilson, “Metallization systems for silicon integrated circuits,” *Proceedings of the IEEE*, vol. 57, no. 9, pp. 1580–1586, Sep. 1969, doi: 10.1109/PROC.1969.7339.
[22] E. Hecht, *Optics*, 4th ed. Addison-Wesley, 2002.
[23] P. J. King, “Hafnium oxide-based dielectrics by atomic layer deposition,” p. 161.
[24] R. L. Puurunen *et al.*, “Hafnium oxide films by atomic layer deposition for high-κ gate dielectric applications: Analysis of the density of nanometer-thin films,” *Appl. Phys. Lett.*, vol. 86, no. 7, p. 073116, 2005, doi: 10.1063/1.1866219.
[25] G. Scarel *et al.*, “Trends of structural and electrical properties in atomic layer deposited HfO2 films,” *Mater. Sci. Eng. B*, vol. 109, no. 1, pp. 11–16, Jun. 2004, doi: 10.1016/j.mseb.2003.10.021.
[26] D. A. Neamen, *Semiconductor physics and devices: basic principles*, 4th ed. New York, NY: McGraw-Hill, 2012.
[27] X.-Y. Zhang *et al.*, “Simulation and Fabrication of HfO2 Thin Films Passivating Si from a Numerical Computer and Remote Plasma ALD,” *Applied Sciences*, vol. 7, no. 12, Art. no. 12, Dec. 2017, doi: 10.3390/app7121244.
[28] J. Molina *et al.*, “Physical and electrical characteristics of atomic-layer deposition-HfO_2 films deposited on Si substrates having different silanol Si-OH densities,” *J. Vac. Sci. Technol.*, vol. 31, no. 1, p. 01A132, Jan. 2013, doi: 10.1116/1.4769206.
[29] R. K. Vinnakota and D. A. Genov, “Terahertz Optoelectronics with Surface Plasmon Polariton Diode,” *Sci. Rep.*, vol. 4, no. 1, Art. no. 1, May 2014, doi: 10.1038/srep04899.
[30] K. Anglin *et al.*, “Voltage-controlled active mid-infrared plasmonic devices,” *J. Appl. Phys.*, vol. 109, no. 12, p. 123103, Jun. 2011, doi: 10.1063/1.3600230.
[31] J. A. Dionne *et al.*, “PlasMOStor: A Metal−Oxide−Si Field Effect Plasmonic Modulator,” *Nano Letters*, vol. 9, no. 2, pp. 897–902, Feb. 2009, doi: 10.1021/nl803868k.

Supplementary

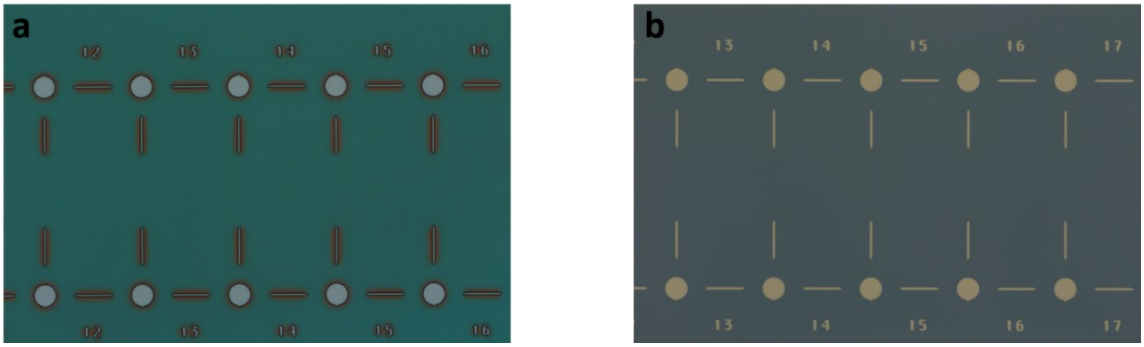

Fig. S-1. round plasmonic patches and alignment marks (a) after exposure and development, and (b) after Au deposition and lift-off.

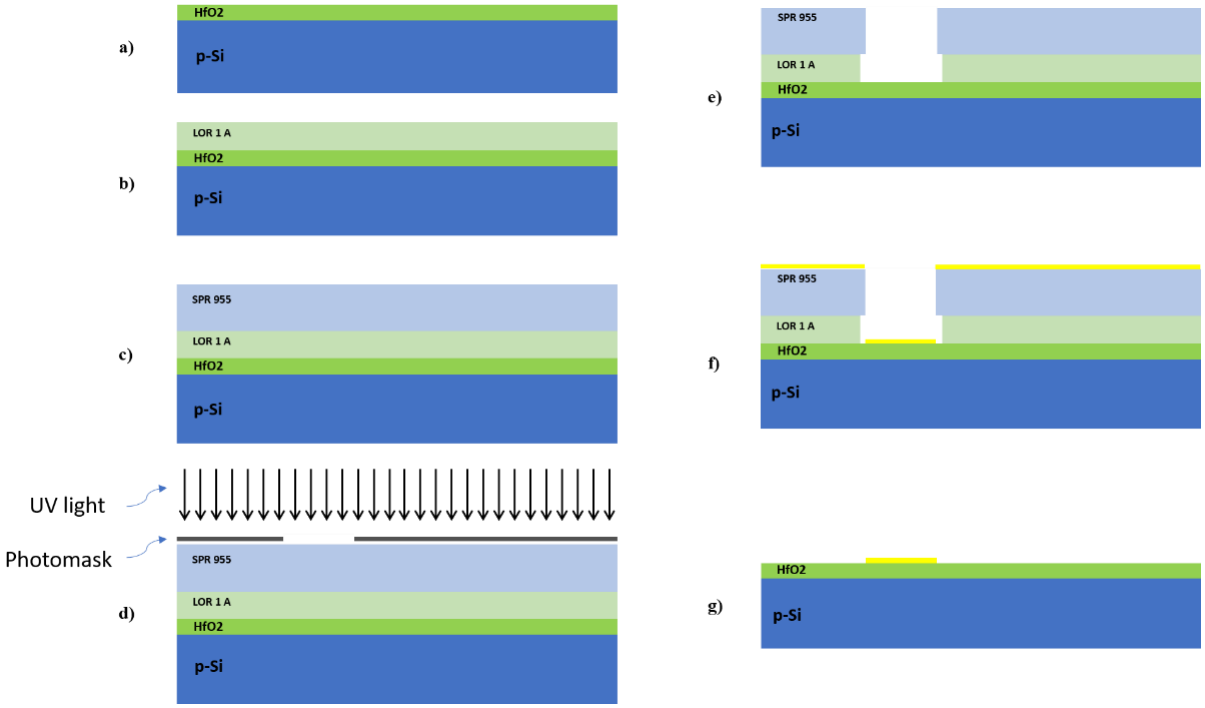

Fig. S-2. Fabrication process flow of plasmonic surfaces: a) selection of a p-doped silicon substrate with a film of HfO_2 grown on its surface, b) spin coating of LOR 1A, c) spin coating of SPR 955, d) exposure of PR using a photomask, e) development of exposed patterns, f) 20 nm Au evaporation, g) PR bilayer lift off.

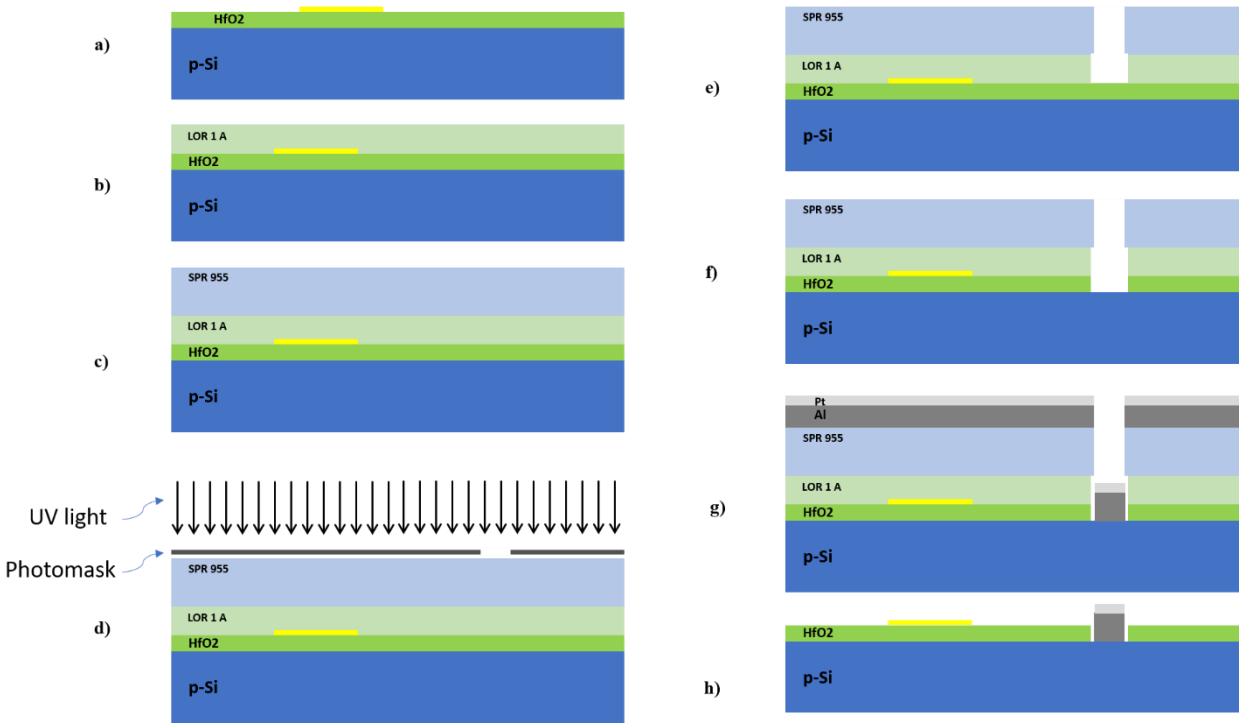

Fig. S-3. Fabrication process flow of ohmic contacts: a) substrate bearing HfO_2 and thin Au plasmonic patches, b) spin coating LOR 1A, c) spin coating SPR 955, d) UV exposure using photomask, e) development of PR, f) BOE wet etch of HfO_2, g) deposition of Pt/Al, h) lift-off.

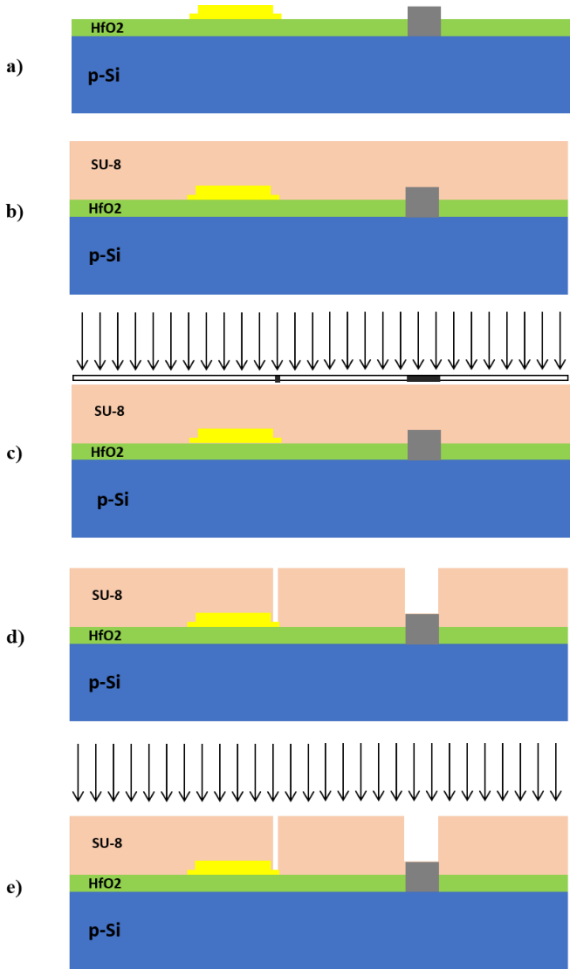

Fig. S-4. Fabrication process flow of interlayer dielectric and via holes: a) silicon substrate with Au patches, nanogratings, and ohmic contacts from the previous stages, b) spin coating of SU-8 2000.5, c) exposure, d) development, e) second exposure.

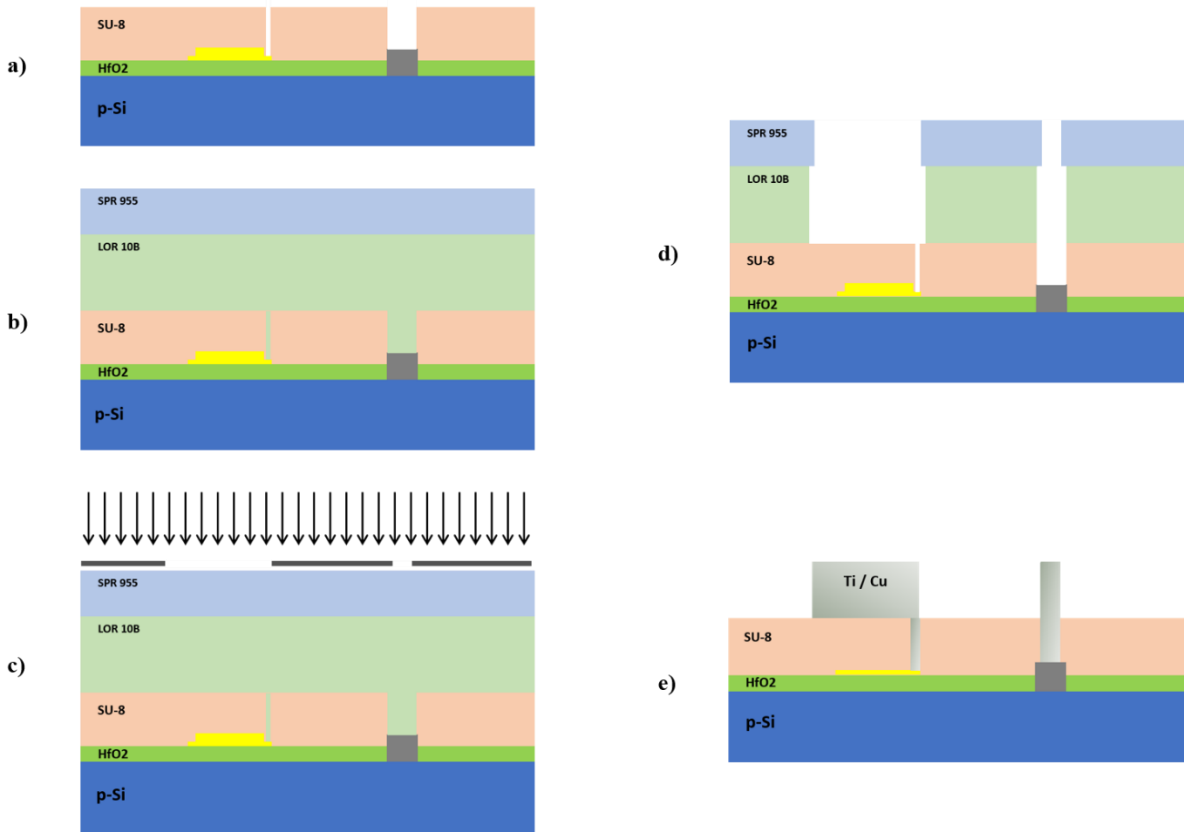

Fig. S-5. Fabrication process flow of metal contacts: a) substrate with plasmonic surfaces, ohmic contacts, nanogratings, and interlayer dielectric with via holes, b) spin coating of LOR 10B and SPR 955, c) exposure, d) development, e) metal deposition.

Table S-1. Procedure for ohmic contact photolithography and metallization.

Process	Step		Description
Spin coating of LOR 1A	1	Load substrate and dispense LOR 1A	Make sure wafer is centered and leveled on the spinner chuck. Dispense 5 mL of LOR 1A using a pipette.
	2	Soak	Let LOR 1A level out on the substrate for 30 s, while vacuum is disengaged.
	3	Spin coating	15 s @ 200 r/s, 10 s @ 3000 rpm, 15 s @ 200 r/s, 30 s @ 6000 rpm, 6 s @ -1000 r/s
	4	Relaxation	Disengage the vacuum and let the substrate relax for 30 s before moving it.
	5	Soft bake	Place the substrate on a flat hotplate at 190 °C for 5 min to evaporate solvent and harden the resist. Let the substrate cool down for several minutes before attempting the next step.
Spin coating of SPR 955	1	Dispense SPR 955	Reload the substrate into the spinner and dispense 5 mL of photoresist using a pipette.
	2	Soak	Let SPR 955 level out on the substrate for 10 s, while vacuum is disengaged.
	3	Spin coating	15 s @ 200 r/s, 10 s @ 3000 rpm, 15 s @ 200 r/s, 30 s @ 6000 rpm, 6 s @ -1000 r/s
	4	Relax	Disengage the vacuum and let the substrate relax for 10 s before moving it.
	5	Soft bake	Place the substrate on a flat hotplate at 100 °C for 3 min to evaporate solvent and harden the resist. Let the substrate cool down for several minutes.
Exposure	1	UV exposure	Exposed with 84 mJ of radiation. Three I-line filters were used to filter out UV- radiation wavelengths other than 365 nm.
Development	1	MF CD-26	Immersed the sample into a bath of CD-26 developer for 60 s. Rinse in DI water and blow dry with N_2.
Etching HfO_2	1	Solution preparation	Mix 5 units of HF with 100 units DI H_2O.
	2	Immerse	Immersed the sample into a bath of solution and hold for 40 s.
	3	DI H_2O rinse	Rinse in DI water and blow dry with N_2.
Metal deposition	1	100 nm Al and 20 nm Pt nm	100 nm Al e-beam evaporated at 5 Å/s. 20 nm Pt e-beam evaporated at 0.2 Å/s, over Al, without breaking vacuum.
	2	Lift-off	Immerse the sample in Remover PG for 20 min. Rinse in IPA and DI water. Blow dry with N_2.

Table S-2. Procedure for SU-8 interlayer dielectric with via holes.

Process	Step		Description
Spin coating of SU-8 2000.5	1	Load substrate and dispense, SU-8	Make sure wafer is centered and leveled on the spinner chuck. Dispense 5 mL of SU-8 using a pipette.
	2	Soak	Let SU-8 level out on the substrate, while vacuum is disengaged.
	3	Spin coating	10 s @100 r/s, 30 s @ 1000 rpm, 10 s @ -100 r/s
	4	Relaxation	Disengage the vacuum and let the substrate relax before moving it.
	5	Soft bake	Place the substrate on a flat hotplate at 65 °C for 5 min. Ramp up the temperature at 360 °C per hr to 95 °C and let it stand for 5 min. Ramp down the temperature at 360 °C per hr to 40 °C. Let the substrate cool down for 10 min. Total time 25 min.
Exposure	1	UV exposure	Exposed with 38 mJ radiation. One I-line filter was used to filter out UV- radiation wavelengths other than 365 nm.
	2	Post exposure bake	Place the sample on a flat hotplate at 80 °C for 1 min. Ramp up the temperature at 360 °C per hr to 95 °C and let it stand for 1 min. Ramp down the temperature at 360 °C per hr to 65 °C. Let the sample cool down for several minutes.
Development	1	SU-8 developer	The sample was fully immersed in a bath of SU-8 developer for 3 min, while held vertically. It was sonicated (80 KHz, 30% power) for two intervals of 10 s, during the first and the last minutes. It was then DI water rinsed, and N_2 blow dried.
	2	Hard bake	Place the sample on a flat hotplate at 120 °C for 30 min. Let the substrate cool down for several minutes before attempting the next step.
Exposure	1	UV exposure	Exposed with 180 mJ radiation. One I-line filter was used to filter out UV- radiation wavelengths other than 365 nm.
	2	Hard Bake	Place the sample on a flat hotplate at 120 °C for 30 min.

Table S-3. Procedure for contact metal layer photolithography, metallization, and lift-off.

Process	Step		Description
Spin coating of LOR 10B	1	Load substrate and dispense, LOR 10B	Ensure the wafer is centered and leveled on the spinner chuck. Pour enough LOR 10B from a vile to cover the surface of the sample.
	2	Soak	Let LOR 10B level out on the substrate, while vacuum is disengaged.
	3	Spin coating	3 s @ 100 r/s, 10 s @ 300 rpm, 3 s @ 400 r/s, 5 min @ 1500 rpm, 5 s @ -300 r/s
	4	Relaxation	Disengage the vacuum and let the substrate relax before moving it.
	5	Soft bake	Place the sample on a flat hotplate at 190 °C for 6 min. Let the sample cool down for several minutes.
	6	Cool	Leave sample on a leveled surface to cool down to room temperature.
Spin coating of SPR 955	1	Refer to steps 1-4, "Spin coating of SPR 955", in Table 2.	
Exposure	1	UV exposure	Exposed with 80 mJ radiation. Three I-line filters were used to filter out UV- radiation wavelengths other than 365 nm.
Development	1	Development	The sample was immersed in CD-26 for 35 s. It was then rinsed with DI H_2O, N_2 blow dried, and baked at 120 °C for 5 min followed by another 30 s in CD-26. It was then rinsed with DI H_2O, and N_2 blow dried, again.
Metal deposition	1	Evaporation	5 nm Ti adhesion layer, e-beam evaporation at 0.5 Å/s. 300 nm Cu, e-beam evaporation at 2 Å/s on Cr without breaking vacuum.
	2	Lift-off, Remover PG	The sample was immersed in Remover PG for 1 hr. It was then rinsed in Acetone, IPA, and DI H_2O, and N_2 blow dried.

CHAPTER 5. CONCLUSIONS

5.1. Summary and contributions

Chapter 2 describes the numerical models used throughout this thesis to design and optimize plasmonic structures and their performance in optoelectronic devices. For this purpose, a 3D model based on the finite-difference time-domain (FDTD) method was developed by employing a commercial software package (Lumerical FDTD). This model was validated and then used to design the nanoantenna-based photodetectors in Chapter 3, as well as the nanoantennas employed in the two collaborative projects, described in Appendices A and B.

In chapter 3, two novel *plasmonic photodetectors based on arrays of nanodipoles and nanomonopoles* were presented and modelled. In the first design [67] an array of gold nanodipoles was placed on an InP substrate. The gaps of the nanodipoles were filled with low-defect, low-doped $In_{0.53}Ga_{0.47}As$ as the detection volumes. A maximum responsivity of 100 mA/W was calculated for this device at wavelengths near 1550 nm. The electrical bandwidth of the photodetectors was found to reach as high as 4 THz, for an optimum gap length. In the second photodetector design [68] an array of interconnected gold nanomonopoles was considered and modeled. Nanomonopoles were assumed formed on a film of InGaAs, which was located on an n-doped InP substrate. In this case the responsivity of the photodetector device was estimated to reach ~200 mA/W at telecommunication wavelengths (~1550 nm). While the electrical bandwidth is shown to increase by decreasing the number of elements in the array, to an optimum bandwidth of ~1 THz which was predicted for the case of a 20-nm-thick InGaAs layer and a 4 × 4 nanomonopole array. A simple linear model for the total resistance of the structure was also proposed, validated, and employed to calculate the RC time constant of the device. The total resistance of the device was found to increase by increasing the thickness of the InGaAs layer and by decreasing the number of array elements.

Chapter 4 describes *a nanofabrication process developed to fabricate plasmonic intensity reflection/ transmission modulators.* The modulators were designed and modelled by Chengkun Chen [70]. A similar device was also fabricated and experimentally tested by Sa'ad Hassan and Anthony Olivieri in [55], [56]. Here, the modulator was improved by enabling modulation in transmission. Contrary to previous devices, here the modulators were made transparent at telecom wavelengths by employing a front-side probing technique, in which both the signal and ground electrodes are probed on the front-side of the device using a microwave coplanar waveguide

probe. A dielectric interlayer was also added in order to minimize the capacitance associated with the contact architecture, thereby increasing the electrical bandwidth of the device. Lithographic fabrication methods were also optimized to achieve higher resolution for the structures and to improve repeatability of fabrication.

5.2. Suggestions for future work

Fabrication and experimental assessment of the two photodetectors presented in chapter 3 is a natural next step. It is also important to measure the electrical bandwidth, as well as the noise characteristics of these devices. Considering their ultrafast bandwidth, they can also be envisioned as THz sources, which will open the doors to many new research projects.

Performing passive and active measurements on the modulators fabricated in chapter 4 was out of the scope of this thesis, yet is suggested as future work. These measurements could be performed in reflection, as well as in transmission, by placing the photoreceiver and its ancillary optics in the respective optical path. III-V semiconductor substrates may also be investigated in conjunction with Si for faster switching times. Ultimately, it would be desirable to combine modulators with other photonic components, such as source and photodetector, on a single chip.

Various designs of nanostructures including, but not limited to, different types of nanoantennas may be considered to replace nanomonopoles for enhanced Raman scattering and high-harmonic generation. It would also be interesting to replace Au with a more resilient plasmonic metal that survives high harmonic fields.

REFERENCES

[1] S. A. Maier and H. A. Atwater, "Plasmonics: Localization and guiding of electromagnetic energy in metal/dielectric structures," *J. Appl. Phys.*, vol. 98, no. 1, p. 011101, Jul. 2005, doi: 10.1063/1.1951057.

[2] S. A. Maier, *Plasmonics: Fundamentals and Applications*. Springer Science & Business Media, 2007.

[3] H. A. Atwater, "The Promise of PLASMONICS," *Sci. Am.*, vol. 296, no. 4, pp. 56–63, 2007.

[4] Z. Liang *et al.*, "Plasmonic Enhanced Optoelectronic Devices," *Plasmonics*, vol. 9, no. 4, pp. 859–866, Aug. 2014, doi: 10.1007/s11468-014-9682-7.

[5] P. Berini, "Figures of merit for surface plasmon waveguides," *Opt. Express*, vol. 14, no. 26, pp. 13030–13042, Dec. 2006, doi: 10.1364/OE.14.013030.

[6] E. Ozbay, "Plasmonics: Merging Photonics and Electronics at Nanoscale Dimensions," *Science*, vol. 311, no. 5758, pp. 189–193, Jan. 2006, doi: 10.1126/science.1114849.

[7] J. A. Schuller *et al.*, "Plasmonics for extreme light concentration and manipulation," *Nat. Mater.*, vol. 9, no. 3, pp. 193–204, Mar. 2010, doi: 10.1038/nmat2630.

[8] M. L. Brongersma and V. M. Shalaev, "The Case for Plasmonics," *Science*, vol. 328, no. 5977, pp. 440–441, Apr. 2010, doi: 10.1126/science.1186905.

[9] L. Jiang *et al.*, "Free-standing one-dimensional plasmonic nanostructures," *Nanoscale*, vol. 4, no. 1, pp. 66–75, 2012, doi: 10.1039/C1NR11445J.

[10] M. Hochberg *et al.*, "Integrated plasmon and dielectric waveguides," *Opt. Express*, vol. 12, no. 22, p. 5481, 2004, doi: 10.1364/OPEX.12.005481.

[11] M. P. Nezhad, K. Tetz, and Y. Fainman, "Gain assisted propagation of surface plasmon polaritons on planar metallic waveguides," *Opt. Express*, vol. 12, no. 17, p. 4072, 2004, doi: 10.1364/OPEX.12.004072.

[12] J. R. Krenn and J.-C. Weeber, "Surface plasmon polaritons in metal stripes and wires," *Philos. Trans. Royal Soc.*, vol. 362, no. 1817, pp. 739–756, Apr. 2004, doi: 10.1098/rsta.2003.1344.

[13] W. Nomura, M. Ohtsu, and T. Yatsui, "Nanodot coupler with a surface plasmon polariton condenser for optical far/near-field conversion," *Appl. Phys. Lett.*, vol. 86, no. 18, p. 181108, May 2005, doi: 10.1063/1.1920419.

[14] A. V. Krasavin and A. V. Zayats, "Photonic Signal Processing on Electronic Scales: Electro-Optical Field-Effect Nanoplasmonic Modulator," *Phys. Rev. Lett.*, vol. 109, no. 5, p. 053901, Jul. 2012, doi: 10.1103/PhysRevLett.109.053901.

[15] J. S. White *et al.*, "Extraordinary optical absorption through subwavelength slits," *Opt. Lett.*, vol. 34, no. 5, p. 686, Mar. 2009, doi: 10.1364/OL.34.000686.

[16] T. Barwicz *et al.*, "Silicon photonics for compact, energy-efficient interconnects [Invited]," *J. Opt. Netw.*, vol. 6, no. 1, p. 63, 2007, doi: 10.1364/JON.6.000063.

[17] T. Ishi, J. Fujikata, K. Makita, T. Baba, and K. Ohashi, "Si Nano-Photodiode with a Surface Plasmon Antenna," *Jpn. J. Appl. Phys.*, vol. 44, no. No. 12, pp. L364–L366, Mar. 2005, doi: 10.1143/JJAP.44.L364.

[18] L. Tang *et al.*, "Nanometre-scale germanium photodetector enhanced by a near-infrared dipole antenna," *Nat. Photon*, vol. 2, no. 4, pp. 226–229, Apr. 2008, doi: 10.1038/nphoton.2008.30.

[19] H. Ditlbacher *et al.*, "Organic diodes as monolithically integrated surface plasmon polariton detectors," *Appl. Phys. Lett.*, vol. 89, no. 16, p. 161101, Oct. 2006, doi: 10.1063/1.2362975.

[20] P. Neutens *et al.*, "Electrical detection of confined gap plasmons in metal–insulator–metal waveguides," *Nat. Photon*, vol. 3, no. 5, pp. 283–286, May 2009, doi: 10.1038/nphoton.2009.47.

[21] D.-S. Ly-Gagnon, S. E. Kocabas, and D. A. B. Miller, "Characteristic Impedance Model for Plasmonic Metal Slot Waveguides," *IEEE J. Sel. Top. Quantum Electron.*, vol. 14, no. 6, pp. 1473–1478, Nov. 2008, doi: 10.1109/JSTQE.2008.917534.
[22] Y. Liu *et al.*, "Plasmon resonance enhanced multicolour photodetection by graphene," *Nat Commun*, vol. 2, no. 1, p. 579, Sep. 2011, doi: 10.1038/ncomms1589.
[23] A. Dorodnyy *et al.*, "Plasmonic Photodetectors," *IEEE J. Sel. Top. Quantum Electron.*, vol. 24, no. 6, pp. 1–13, Nov. 2018, doi: 10.1109/JSTQE.2018.2840339.
[24] P. Berini, "Surface plasmon photodetectors and their applications," *Laser & Photonics Rev.*, vol. 8, no. 2, pp. 197–220, 2014, doi: 10.1002/lpor.201300019.
[25] M. Alavirad, L. Roy, and P. Berini, "Surface plasmon enhanced photodetectors based on internal photoemission," *JPE*, vol. 6, no. 4, p. 042511, Oct. 2016, doi: 10.1117/1.JPE.6.042511.
[26] S. M. Sze and K. K. NG, *Physics of Semiconductor Devices*. Hoboken, NJ, USA: Wiley, 2007.
[27] C. L. Tan *et al.*, "Optical absorption enhancement of hybrid-plasmonic-based metal-semiconductor-metal photodetector incorporating metal nanogratings and embedded metal nanoparticles," *Opt. Express*, vol. 21, no. 2, p. 1713, Jan. 2013, doi: 10.1364/OE.21.001713.
[28] C. W. Berry *et al.*, "Significant performance enhancement in photoconductive terahertz optoelectronics by incorporating plasmonic contact electrodes," *Nat. Commun.*, vol. 4, no. 1, Art. no. 1, Mar. 2013, doi: 10.1038/ncomms2638.
[29] L.-B. Luo *et al.*, "Light trapping and surface plasmon enhanced high-performance NIR photodetector," *Sci. Rep.*, vol. 4, no. 1, Art. no. 1, Jan. 2014, doi: 10.1038/srep03914.
[30] J. Guo *et al.*, "Design of plasmonic photodetector with high absorptance and nano-scale active regions," *Opt. Express*, vol. 24, no. 16, p. 18229, Aug. 2016, doi: 10.1364/OE.24.018229.
[31] E. Panchenko *et al.*, "Plasmonic Metasurface-Enabled Differential Photodetectors for Broadband Optical Polarization Characterization," *ACS Photonics*, vol. 3, no. 10, pp. 1833–1839, Oct. 2016, doi: 10.1021/acsphotonics.6b00342.
[32] W. Rieger *et al.*, "Yagi-Uda nanoantenna enhanced metal-semiconductor-metal photodetector," *Appl. Phys. Lett.*, vol. 113, no. 2, p. 023102, Jul. 2018, doi: 10.1063/1.5038339.
[33] J. Gosciniak and M. Rasras, "High-bandwidth and high-responsivity waveguide-integrated plasmonic germanium photodetector," *J. Opt. Soc. Am. B*, vol. 36, no. 9, p. 2481, Sep. 2019, doi: 10.1364/JOSAB.36.002481.
[34] J. Gosciniak *et al.*, "Plasmonic Schottky photodetector with metal stripe embedded into semiconductor and with a CMOS-compatible titanium nitride," *Sci Rep*, vol. 9, no. 1, p. 6048, Dec. 2019, doi: 10.1038/s41598-019-42663-3.
[35] Y. Ding *et al.*, "Ultra-compact integrated graphene plasmonic photodetector with bandwidth above 110 GHz," *Nanophotonics*, vol. 9, no. 2, pp. 317–325, Feb. 2020, doi: 10.1515/nanoph-2019-0167.
[36] J. Guo *et al.*, "High-performance silicon−graphene hybrid plasmonic waveguide photodetectors beyond 1.55 μm," *Light: Sci. Appl.*, vol. 9, no. 1, Dec. 2020, doi: 10.1038/s41377-020-0263-6.
[37] S. O. Kasap, *Optoelectronics & Photonics:Principles & Practices: International Edition*. Pearson Education Limited, 2013.
[38] R. G. Hunsperger, *Integrated Optics*. New York, NY: Springer New York, 2009.
[39] I. P. Kaminow, L. W. Stulz, and E. H. Turner, "Efficient strip-waveguide modulator," *Appl. Phys. Lett.*, vol. 27, no. 10, pp. 555–557, Nov. 1975, doi: 10.1063/1.88284.
[40] D. Hall, A. Yariv, and E. Garmire, "Observation of propagation cutoff and its control in thin optical waveguides," *Appl. Phys. Lett.*, vol. 17, no. 3, p. 4, 1970.

[41] J. C. Campbell *et al.*, “GaAs electro-optic directional-coupler switch,” *Appl. Phys. Lett.*, vol. 27, no. 4, pp. 202–205, Aug. 1975, doi: 10.1063/1.88428.
[42] W. E. Martin, “A new waveguide switch/modulator for integrated optics,” *Appl. Phys. Lett.*, vol. 26, no. 10, pp. 562–564, May 1975, doi: 10.1063/1.87992.
[43] F. K. Reinhart, “Electroabsorption in Al $_{y}$ Ga $_{1-y}$ As–Al $_{x}$ Ga $_{1-x}$ As double heterostructures,” *Appl. Phys. Lett.*, vol. 22, no. 8, pp. 372–374, Apr. 1973, doi: 10.1063/1.1654678.
[44] E. A. J. Marcatili, “Dielectric rectangular waveguide and directional coupler for integrated optics,” *The Bell System Technical Journal*, vol. 48, no. 7, pp. 2071–2102, Sep. 1969, doi: 10.1002/j.1538-7305.1969.tb01166.x.
[45] J. C. Campbell, F. A. Blum, and D. W. Shaw, “GaAs electro-optic channel-waveguide modulator,” p. 4, 1975.
[46] M. Papuchon *et al.*, “Electrically switched optical directional coupler: Cobra,” *Appl. Phys. Lett.*, vol. 27, no. 5, pp. 289–291, Sep. 1975, doi: 10.1063/1.88449.
[47] J. M. Hammer, “Digital electro-optic grating deflector and modulator,” *Appl. Phys. Lett.*, vol. 18, no. 4, pp. 147–149, Feb. 1971, doi: 10.1063/1.1653603.
[48] Chen Tsai, Bumman Kim, and F. El-Akkari, “Optical channel waveguide switch and coupler using total internal reflection,” *IEEE J. Quantum Electron.*, vol. 14, no. 7, pp. 513–517, Jul. 1978, doi: 10.1109/JQE.1978.1069829.
[49] J. A. Dionne *et al.*, “PlasMOStor: A Metal−Oxide−Si Field Effect Plasmonic Modulator,” *Nano Lett.*, vol. 9, no. 2, pp. 897–902, Feb. 2009, doi: 10.1021/nl803868k.
[50] W. Cai, J. S. White, and M. L. Brongersma, “Compact, High-Speed and Power-Efficient Electrooptic Plasmonic Modulators,” *Nano Lett.*, vol. 9, no. 12, pp. 4403–4411, Dec. 2009, doi: 10.1021/nl902701b.
[51] A. Melikyan *et al.*, “Surface plasmon polariton absorption modulator,” *Opt. Express*, vol. 19, no. 9, p. 8855, Apr. 2011, doi: 10.1364/OE.19.008855.
[52] A. Melikyan *et al.*, “High-speed plasmonic phase modulators,” *Nat. Photon.*, vol. 8, no. 3, pp. 229–233, Mar. 2014, doi: 10.1038/nphoton.2014.9.
[53] C. Haffner *et al.*, “All-plasmonic Mach–Zehnder modulator enabling optical high-speed communication at the microscale,” *Nat. Photon.*, vol. 9, no. 8, pp. 525–528, Aug. 2015, doi: 10.1038/nphoton.2015.127.
[54] D. Ansell *et al.*, “Hybrid graphene plasmonic waveguide modulators,” *Nat. Commun.*, vol. 6, no. 1, p. 8846, Dec. 2015, doi: 10.1038/ncomms9846.
[55] A. Olivieri *et al.*, “Plasmonic Nanostructured Metal–Oxide–Semiconductor Reflection Modulators,” *Nano Lett.*, vol. 15, no. 4, pp. 2304–2311, Apr. 2015, doi: 10.1021/nl504389f.
[56] S. Hassan *et al.*, “Fabrication of a plasmonic modulator incorporating an overlaid grating coupler,” *Nanotechnology*, vol. 25, no. 49, p. 495202, Dec. 2014, doi: 10.1088/0957-4484/25/49/495202.
[57] M. Ayata *et al.*, “High-speed plasmonic modulator in a single metal layer,” *Science*, vol. 358, no. 6363, pp. 630–632, Nov. 2017, doi: 10.1126/science.aan5953.
[58] C. Wang *et al.*, “Nanophotonic lithium niobate electro-optic modulators,” *Opt. Express*, vol. 26, no. 2, p. 1547, Jan. 2018, doi: 10.1364/OE.26.001547.
[59] C. Wang *et al.*, “Integrated lithium niobate electro-optic modulators operating at CMOS-compatible voltages,” *Nature*, vol. 562, no. 7725, pp. 101–104, Oct. 2018, doi: 10.1038/s41586-018-0551-y.
[60] M. Burla *et al.*, “500 GHz plasmonic Mach-Zehnder modulator enabling sub-THz microwave photonics,” *APL Photonics*, vol. 4, no. 5, p. 056106, May 2019, doi: 10.1063/1.5086868.
[61] U. Koch *et al.*, “Ultra-Compact Terabit Plasmonic Modulator Array,” *J. Lightwave Technol.*, vol. 37, no. 5, pp. 1484–1491, Mar. 2019, doi: 10.1109/JLT.2019.2899372.

[62] B. Baeuerle *et al.*, "120 GBd plasmonic Mach-Zehnder modulator with a novel differential electrode design operated at a peak-to-peak drive voltage of 178 mV," *Opt. Express*, vol. 27, no. 12, p. 16823, Jun. 2019, doi: 10.1364/OE.27.016823.
[63] M. AlShehab *et al.*, "Design and construction of a Raman microscope and characterization of plasmon-enhanced Raman scattering in graphene," *J. Opt. Soc. Am. B*, vol. 36, no. 8, p. F49, Aug. 2019, doi: 10.1364/JOSAB.36.000F49.
[64] G. Vampa *et al.*, "Plasmon-enhanced high-harmonic generation from silicon," *Nature Phys.*, vol. 13, no. 7, pp. 659–662, Jul. 2017, doi: 10.1038/nphys4087.
[65] S. S. Mousavi, P. Berini, and D. McNamara, "Periodic plasmonic nanoantennas in a piecewise homogeneous background," *Opt. Express*, vol. 20, no. 16, p. 18044, Jul. 2012, doi: 10.1364/OE.20.018044.
[66] E. D. Palik, *Handbook of Optical Constants of Solids*. Academic Press, 1998.
[67] S. S. Mousavi, A. Stöhr, and P. Berini, "Plasmonic photodetector with terahertz electrical bandwidth," *Appl. Phys. Lett.*, vol. 104, no. 14, p. 143112, Apr. 2014, doi: 10.1063/1.4870974.
[68] S. S. Mousavi, A. Stöhr, and P. Berini, "Ultrafast Surface Plasmon III–V Photodetectors Based on Nanomonopoles," *J. Lightwave Technol., JLT*, vol. 34, no. 20, pp. 4682–4687, Oct. 2016.
[69] B. G. Ghamsari *et al.*, "Enhanced Raman scattering in graphene by plasmonic resonant Stokes emission," *Nanophotonics*, vol. 3, no. 6, pp. 363–371, Dec. 2014, doi: 10.1515/nanoph-2014-0014.
[70] P. S. J. Berini and C. Chen, "Metal-insulator-semiconductor devices based on surface plasmon polaritons," US9397241B2, Jul. 19, 2016.

Appendix A

Research Article Vol. 36, No. 8 / August 2019 / Journal of the Optical Society of America B F49

Journal of the Optical Society of America B

OPTICAL PHYSICS

Design and construction of a Raman microscope and characterization of plasmon-enhanced Raman scattering in graphene

Maryam AlShehab,[1,2] Saba Siadat Mousavi,[1,2] Maude Amyot-Bourgeois,[2,3] Jaspreet Walia,[1,2] Anthony Olivieri,[2] Behnood Ghamsari,[1,2] and Pierre Berini[1,2,3,*]

[1]School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa K1N6N5, Canada
[2]Center for Research in Photonics, University of Ottawa, Ottawa K1N6N5, Canada
[3]Department of Physics, University of Ottawa, Ottawa K1N6N5, Canada
*Corresponding author: pberini@uottawa.ca

Received 28 February 2019; revised 13 April 2019; accepted 15 April 2019; posted 16 April 2019 (Doc. ID 361267); published 9 May 2019

Nanometallic structures efficiently convert light to surface plasmon-polaritons (SPPs) localized to ultra-small volumes. Such structures provide highly enhanced fields and are of interest in applications involving SPP-enhanced nonlinear optics. We report the design and construction of a spontaneous Raman microscope augmented with *in situ* reflectance measurement capabilities, and demonstrate its use for nonlinear plasmonics. The structures investigated consist of rectangular gold nanoantennas on graphene on a SiO_2/Si substrate. Specifically, SPP-enhanced Raman scattering from graphene is investigated using nanoantennas that are spectrally aligned with the Stokes wavelength of the graphene 2D peak. We use the microscope to demonstrate Raman scattering enhancement in graphene based on plasmonic resonant enhancement of the Stokes emission, where a maximum cross-sectional gain of ~500 per antenna was measured. We also measure the reflectance response of nanoantenna structures of different dimensions (length, width) to determine how the resonant wavelength shifts with dimensions and ensure spectral alignment with the Stokes wavelengths of interest.

https://doi.org/10.1364/JOSAB.36.000F49

1. INTRODUCTION

Over the last few decades, laser spectroscopy has become an important technique in materials science, nanophotonics, biophysics/biochemistry, and in the life sciences for disease detection and therapy control [1]. Laser spectroscopy allows the identification and structural characterization of materials and molecules by probing transitions in energy levels. The transition between molecular energy levels is a unique characteristic of a molecule or crystalline solid, and in some cases, amorphous materials as well. Measuring these transitions is generally useful for molecular identification, purity, and crystallinity. Surface-enhanced Raman spectroscopy (SERS) probes vibrational and rotational levels of molecules near roughened or nanostructured metallic surfaces, usually consisting of silver or gold, and provides enhanced signals even at sub-monolayer coverage [1]. The scattered field by a thin film or set of molecules contains reflection and/or Rayleigh scattered fields at the same wavelength as the incident beam plus the spontaneous Raman field at a slightly shorter or longer wavelength (anti-Stokes/Stokes). The biggest disadvantage of Raman spectroscopy is the weak signals generated by the Raman process; the magnitude of the cross section for the process is 10^{12}–10^{14} lower than a typical fluorescence cross section [2]. Furthermore, Raman scattering spectra and fluorescence spectra may overlap in many cases.

Weak Raman scattering has motivated research on the use of plasmonic substrates or nanostructures to produce enhanced local fields due to surface plasmon-polaritons (SPPs) that resonate thereon, aiming to enhance the Raman signal from nearby molecules or materials, in a process termed surface-enhanced Raman scattering (SERS) [1–3]. Much of the work on SERS has been directed toward improving the performance of SERS substrates by changing the geometry. For example, Scarabelli *et al.* presented the enhancement caused by the self-assembly of gold nano-triangles [3], and Funston *et al.* used nanorods in different arrangements, separations, and sizes, and deduced that the proximity of the nanorods played a significant role in the enhancement factors that were measured [4].

The integration of plasmonic nanostructures with graphene has received attention in the past few years due to the field enhancement in graphene that occurs when light interacts with the nanostructures. Khorasaninejad *et al.* showed that different shapes of nanostructures can yield different SERS spectra for graphene [5]. Specifically, they observed that the intensity of a Raman signal was dependent on the shape, diameter,

0740-3224/19/080F49-11 Journal © 2019 Optical Society of America

and spacing of the nanoantennas in an array. They found that the Raman intensity increased by a factor of 890 for the G-peak of graphene on crescents (compared to graphene on silicon dioxide), and by a factor of 38 for the 2D peak. Ghamsari *et al.* explored the use of SPPs to enhance graphene–light interaction, by aligning the nanoantennas spectrally with the Stokes wavelengths of the G and 2D peaks. They used gold nanorods as nanoantennas in the form of arrays on graphene on SiO_2 on Si [6], and measured array enhancement factors of 2–3 and single-antenna enhancement factors of ∼160. Aside from observing the graphene D, G, and 2D peaks at 1350 cm^{-1}, 1570 cm^{-1}, and 2700 cm^{-1}, respectively, the 520 cm^{-1} Si–Si bond peak and the higher-order two-phonon Si peak at 970 cm^{-1}, both originating from the substrate, were also observed. They also reported frequency pulling and broadening of the graphene peaks due to interaction with the nanoantennas [7]. Specifically, they showed that the position and linewidth of the G and 2D Raman peaks of graphene were altered by coupling to SPPs that resonated on nanostructures at the Stokes emission frequencies due to an increase in the optical density (OD) of states at the Stokes emission frequency. Li *et al.* used graphene as a spacer between a silver layer and silver nanospheres [8], and observed a graphene near-field enhancement factor of 1700 for the 2D peak. Similarly, Zhao *et al.* showed that sandwiching graphene between silver nanoparticles and graphene nanoparticles resulted in a very sensitive SERS detector, with a limit of detection of 10^{-13} M [9]. Another method of sandwiching graphene was shown by Liu *et al.* In this case, the substrate was bulk silver with a SiO_2 spacer topped by graphene and a silver disk [10]. Urich and colleagues deposited nano-islands of silver [11], and observed a 100-fold enhancement of the graphene Raman peaks due to localized SPPs.

The geometry dependence of the transmittance, reflectance and absorptance response of nanoantennas was studied by Mousavi *et al.* [12]. They found that increasing the length of monopole antennas shifts the transmittance, reflectance, and absorptance resonances to longer wavelengths. This shift is in correspondence with classical antenna theory where the resonance shifts to longer wavelengths with increasing antenna length. However, increasing the width of the antennas decreases the resonance wavelength.

In this paper, we describe the design and construction of a Raman microscope capable of measuring both the reflectance and Raman scattering responses *in situ* from gas, liquid, and solid-state samples. For maximum flexibility, the microscope combines a supercontinuum source with a fixed wavelength laser, enabling a wide variety of single wavelength, multiwavelength, and pump–probe measurements. In this study, the microscope was used to measure the wavelength response of plasmonic nanoantennas with varying rectangular dimensions, as well as the Raman scattering enhancement factor produced from a graphene layer underneath the nanoantenna array. A detailed description of the design and construction of the microscope is provided along with measurement results.

2. SETUP CONSTRUCTION

The confocal Raman microscope design, shown schematically in Fig. 1, focuses the pump laser beam to a very small, ideally diffraction-limited spot, on the sample. The scattered Raman field produced by the sample is then directed to a spectrometer to measure the Stokes shift. A second source is employed to obtain reflectance measurements from the same area of the sample. The spatial distribution of the Raman signal on the sample can be mapped by raster-scanning the sample using a computer-controlled piezo translation stage. Therefore, the Raman signal can be acquired for an array of points, thereby constructing a Raman image.

A. Excitation Arms

1. Pump Laser Arm (He–Ne)

The type of detection device used in the setup must be considered when choosing a laser for a Raman microscope. Our spectrometer functions well at wavelengths less than 900 nm and is a dispersive system with an intensified CCD camera having a cutoff wavelength of 1100 nm. Another point to consider is the laser linewidth, which can range from hundreds of cm^{-1} to much less than 1 cm^{-1}. For Raman spectroscopy, a laser linewidth of 1 cm^{-1} is sufficient. The laser used in a Raman microscope determines several key aspects such as the intensity of the Raman signal, which is inversely proportional to the fourth power of the wavelength (λ^{-4}) [13,14]. Thus, shorter wavelengths yield a higher intensity; however, short-wavelength photons can also excite fluorescence, which in turn can obscure the Raman bands. Therefore, red or near-infrared wavelengths are widely utilized in Raman spectroscopy, because a sufficiently high scattering cross section is accessed while suppressing background fluorescence [13,15]. Gas lasers normally provide stable and narrow emission lines.

This property is particularly desired for Raman spectroscopy, since the measured spectrum is the convolution of the actual Raman band with the laser line shape [14]. Gas lasers also have a very good beam quality and are convection cooled [14].

Thus, to meet the required spectral and spatial resolutions, stability, robustness, ease of use, and cost criteria, a 632.8 nm He–Ne laser producing 21 mW output power was chosen for our microscope (Thorlabs HNL210L). Furthermore, pumping at 632.8 nm results in Raman scattering in the 633 nm (0 cm^{-1}) to 815 nm (3532.84 cm^{-1}) wavelength range (for graphene), which falls within the high sensitivity region of Si CCDs [7,15]. To control the power of the laser input to the microscope, a neutral density filter was placed in front of the laser's output aperture. A Thorlab NDL-10s-4 filter was utilized for this purpose, allowing for attenuation in steps of 0.1, 0.2, 0.3, 0.4, 0.5, 0.8, 1, 2, 3, and 4 in OD, which yielded a transmittance ranging from 79.43% to 0.01%.

Lasers generally produce light output at wavelengths other than the main line of the laser, which can interfere with the Raman scattered signal. This is more commonly seen in gas lasers (Ar^+, Kr^+, He–Ne) and solid-state lasers because they emit atomic lines (plasma lines). Thus, lasers require a bandpass or a laser-line filter to reduce unwanted laser spectra. Although the extraneous emission is usually a few percent of the total power of the laser, the Raman scattered light is weak and even the smallest amount of non-lasing light from the laser may cause interference [14], as elastic scattering from such emission

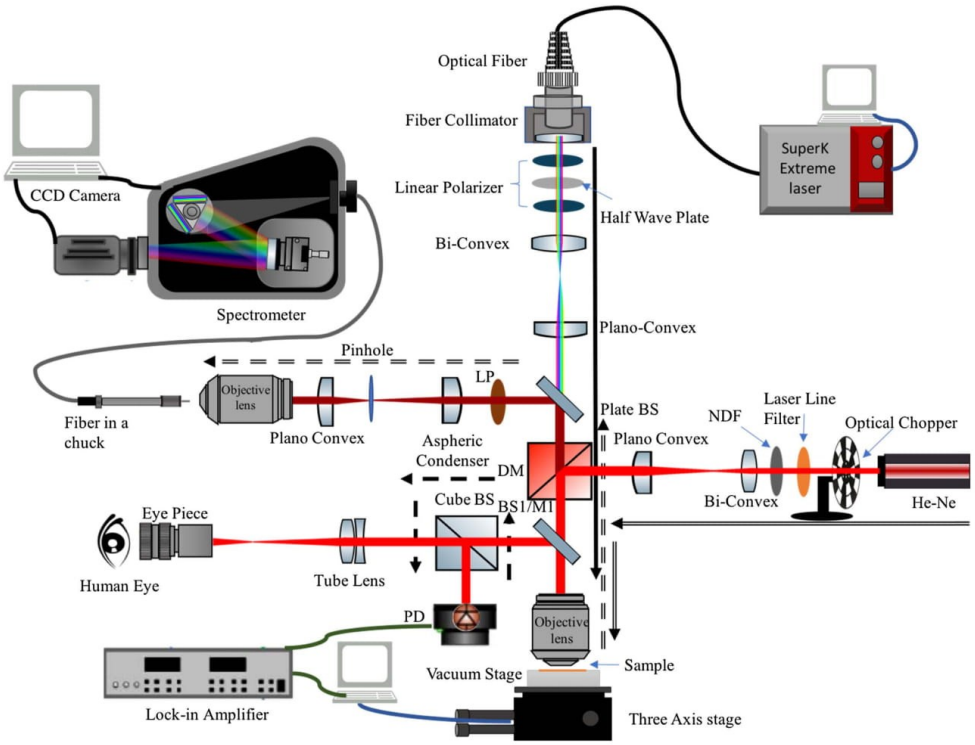

Fig. 1. Optical setup. Single and double solid arrows represent the incoming beam, and dashed double and single arrows are the reflected beam. NDF, neutral density filter; BS, beam splitter; M, mirror; DM, dichroic mirror; LP, low-pass filter; PD, photodetector [15].

may appear in Raman scattering spectra. The filter used here is a laser line filter (Thorlabs FL632.8-3) with a center wavelength at 632.8 ± 0.6 nm and a FWHM of 3 ± 0.6 nm.

The beam emerging from the laser passes through an optical chopper to modulate its intensity. The modulation enables the lock-in technique, which has been proven highly beneficial for measuring normalized reflectance measurements with a high signal-to-noise ratio. The lock-in amplifier used is the Scitec Instruments Ltd. 420 dual lock-in. The lock-in amplifier produces a very narrowband signal that automatically follows small changes in the modulating frequency [16]. It is important to mention that the chopper was moved to the supercontinuum arm to conduct reflectance measurements.

An objective lens achieves a diffraction-limited spot size only if a Gaussian beam entirely fills the input aperture. Under-filling a lens, in general, allows for stray reflections inside the lens that reduce the spatial resolution and peak intensity. More importantly, since the effective NA of an under-filled lens is less than its nominal value, the obtained focused spot will be larger than the diffraction-limit size. On the other hand, overfilling a lens will partially cut the laser beam. Therefore, the excitation profile on the lens' entrance pupil will no longer be perfectly Gaussian, resulting in diffraction as seen by fringing patterns in the output beam. Because the power distribution of diffracted beams is highly non-Gaussian and they possess high divergence angles, overfilling the objective aperture should be strictly avoided [17].

There is often a size mismatch between the laser beam output and the objective entrance pupil diameters. In our system, the former is 0.7 mm while the latter measures 1 cm. Therefore, the laser beam should be expanded while keeping it collimated to meet the desired coverage ratio of the objective entrance pupil. Also, all lasers have a divergence angle. The HNL210L has a divergence angle of 1.15 mrad, which can be corrected using a beam expander. A Keplerian beam expansion configuration was used to expand the beam to an aperture of 7 mm by using a bi-convex lens and a plano-convex lens of focal lengths 3 cm and 30 cm, respectively. Due to the mismatch between the refractive index of free space and the material of the lenses, ~4% of the light would be reflected at each surface of the lens; thus, anti-reflection coated (ARC) lenses were employed.

The objective lens is among the most important components that determine the functionality and performance of an optical microscope. The basic design of an objective lens depends on whether the image distance is finite or infinite,

and whether a compensating versus a fully corrected system is desired. The diffraction-limited laser spot, whose size depends on the lens' numerical aperture (NA) and the wavelength of light (λ), limits the resolution and quality of an image. For a specific laser wavelength, here $\lambda = 632.8$ nm, the diffraction-limited spot size is inversely proportional to the NA of the lens according to $d_{\text{spot}} = 1.22\lambda/\text{NA} = 0.771\ \mu\text{m}/(n \cdot \sin(\theta))$, with n being the refractive index of the imaging environment, usually air, and θ is the half angle of the light cone gathered by the lens. The quality and resolution of the image improves as the laser spot produced by the objective becomes smaller. Furthermore, the brightness of an image scales with the square of the NA. Thus, a large NA is highly desirable because it will provide a small diffraction-limited spot diameter, good resolution, and a bright image.

In order to avoid spherical aberration, a semi-plan or plan objective lens could be used [18]. Chromatic aberration is another important limiting factor. For Raman spectroscopy pumped at 632.8 nm, the lens needs to be able to correct into the near-infrared region in order to capture Stokes photons. Thus, an infinity-corrected plan fluorite objective lens seems an appropriate choice for Raman spectroscopy, providing a reasonable balance between quality and cost.

In contrast to finite objective lenses that map an object to a conjugate real image, an infinity-corrected objective lens maps the object to an "infinitely" distant point in space [19]. Thus, using an infinity-corrected lens focuses the input collimated beam and ensures that the output beam diverges very little [18]. The choice of an infinity-corrected plan fluorite lens narrows down the available options for NA and magnification. The chosen objective lens (Thorlabs RMS20X-PF) has a magnification of 20× and a NA of 0.5. This objective is corrected for chromatic aberrations at four wavelengths and spherical aberrations also at four wavelengths.

In Fig. 1, a few beam splitters (Thorlabs BSW26R) are needed to direct the light through the optical components to the sample and detection devices. To allow for Raman measurements in the system, a dichroic mirror (Thorlabs DMLP650R) is used between the beam expander and the objective lens to combine the second beam into the optical path. Clearly, the second source should have its own beam expander to again match the objective lens diameter.

2. Supercontinuum Arm

A secondary excitation source is used to deliver a linearly polarized and collimated beam to the sample through the objective lens to obtain reflectance measurements. The secondary source used here is a supercontinuum source with tunable bandpass filters delivering light through a linearly polarized fiber (NKT Photonics, SuperK Extreme). The supercontinuum produces white light, and tunable bandpass filters covering a range of wavelengths from 400 nm to 1800 nm are available. For the purposes of reflectance response measurements, a wavelength range of 600 nm to 1100 nm is required for the designed nanoantennas, since their resonance wavelength lie in that range.

There are two problems to solve in directing the supercontinuum beam to the objective lens and the sample. First, the beam should be linearly polarized in an orientation of interest for the nanostructures under investigation on the sample (e.g., nanoantennas), and second, the beam should cover approximately 90%–70% of the objective lens surface. To control the polarization, a half-wave plate (Thorlabs AHWP05M-980) was sandwiched between two polarizers (Thorlabs LPNIRE100-B). The first polarizer was used to determine and fix the linear polarization of the supercontinuum light, which emerges from a polarization-maintaining fiber followed by a beam collimator (Thorlabs F220APC-633 and F220APC-850) attached to the fiber output. The half-wave plate followed by the second polarizer allows the rotation of linear polarization to a desired orientation. Adding the mechanism of polarization rotation allows one to take advantage of the resonance properties of both the horizontal and vertical designs of the nanoantennas. This was seen when D'Andrea *et al.* engineered the plasmonic properties of nanoantennas to make them resonate in both the visible and infrared [20]. By rotating the excitation field polarization, they were able to take advantage of each resonance. In fact, for the purposes of this study the polarizations of the He–Ne and supercontinuum sources play an essential role in the measurements and should be aligned to the polarizations of the nanoantennas in order to see the desired results. Similar to the arm of the pump laser, the beam was expanded via a Keplerian beam expansion scheme, implemented with a bi-convex lens (Thorlabs LB1757-B) and a plano-convex lens (Edmund Optics 67-750) of focal lengths 3 cm and 10 cm, respectively. The beam in this arm follows the same path as that of the pump laser to the objective lens then to the sample under test.

B. Detection Arms

1. Photoreceiver Arm

A photoreceiver is used to measure the power reflected from the nanostructures to determine their reflectance response. In this case (linear scattering), the reflected power is strong, and thus easier to collect than Raman scattered signals. The photoreceiver selected for this purpose is a silicon-based PIN receiver suitable for light in the spectral range of 400 nm to 1100 nm, with an electrical bandwidth of 1.5 MHz (Thorlabs PDA100-A). This bandwidth is sufficient to monitor chopped light. Silicon has very good quantum efficiency, ranging from 80% to 90% over the wavelengths of 350 nm to 900 nm [21]. The photoreceiver has internal preamplifier stages with a switchable gain in steps of 7 × 10 dB (low-noise transimpedance amplifier followed by a voltage amplifier) and an aperture 9.8 mm in diameter, allowing full beams to be collected.

2. Spectroscopy Arm

To analyze the spectrum of the Raman signal, a spectrograph connected to an intensified CCD camera (Princeton Instruments, PI-MAX:1024X256) is used. The spectrograph (Princeton Instruments, IsoPlane SCT320) in the configuration used provides a spectral resolution of 0.08 nm with a 600 groove/mm grating. The elastic (Rayleigh) scattering resulting from direct illumination of the sample by the He–Ne laser or supercontinuum source is many orders of magnitude brighter than the Raman peaks. Consequently, a filter is used to eliminate as much of this spectral content as possible

from the Raman signal. The Raman bands are red shifted with respect to the pump laser, so a long-pass filter is used (FEL0650) that cuts off at 650 nm. To further clean the signal from unwanted beams, such as ghost beams and stray light, and provide a higher spatial resolution, a confocal scheme is employed. Thus, the filtered Raman beam is focused using an aspheric condenser (Thorlabs ACL2520-B) and passed through a small diameter pinhole (Thorlabs P5S, 5 μm) acting as a spatial filter. The pinhole must be slightly under-filled, e.g., by covering an area ∼4 μm in diameter, to avoid clamping the beam and causing diffraction. The beam is then collimated using a plano-convex lens (Thorlabs LA1422-B). The collimated beam must be directed to an objective lens (Newport 40× 0.65 NA) with an aperture diameter of 5 mm, but the beam is collimated to a diameter of $D_{\text{in}} = 7$ mm, so to cover 80% of the objective lens surface the beam is first contracted with an aspheric condenser (Thorlabs ACL2520-B) and a plano-convex lens (Thorlabs LA1422-B) of focal lengths 4 cm and 2 cm, respectively, to a diameter of ∼3.5 mm. Finally, the Raman signal entering the objective lens is focused onto the input facet of a multimode fiber of 62.5 μm core diameter. The output of the fiber is sent into the spectrograph, and the spectra are collected from the CCD camera. It was determined that an integration time of 10 s and an average of five spectra would produce a Raman scattering spectrum with a good SNR.

3. Visualization Arm

To complement the Raman microscope with visual access to the sample, which is required for alignment, a viewing microscope is constructed using the same objective lens by adding a tube lens and an eyepiece (see Fig. 1). The distance between the eyepiece and the tube lens is the sum of their focal length, following Keplerian beam expansion, where the two focal points meet in the middle. An achromat doublet lens (Thorlabs AC254-200-A) is suitable for correcting chromatic aberrations and has a focal length of 20 cm.

4. Software and Test Procedures

For ease of use, multiple LabVIEW codes were created and utilized. The piezo controller, lock-in amplifier, and supercontinuum laser were all controlled using LabVIEW, while the spectrometer was controlled using LightField (Princeton Instruments).

Reflectance responses are obtained using the super continuum source equipped with a tunable acousto-optic bandpass filter and the photoreceiver. The super continuum source is tuned to the desired wavelength using LabVIEW via the tunable filter. The input beam is traced through the system as the solid single line in Fig. 1. The reflected beam is represented by the dashed single lines leading to the photodetector. Measurements are typically carried out by scanning over an area of interest (e.g., 30×30 μm^2) that includes structures and the substrate, then finding the ratio of the reflectance from the structures to that from the substrate. Alternatively, an average of five or more measurement points can be taken on and off the structures to determine the ratio (this takes less time).

For spontaneous Raman measurements, the polarization of the pump He–Ne laser is first aligned to the structures of interest. The beam from the laser is guided through the optics to the objective lens and consequently onto the sample, as sketched by the solid double line in Fig. 1. The reflected Stokes light follows the path indicated by the dashed double lines, through the spectrometer arm and eventually into the spectrometer and camera for detection and analysis. The dichroic mirror and long-pass filter are used to prevent the He–Ne beam from entering the spectrometer and camera, and saturating it.

5. Microscope Characterization

Before acquiring reflectance responses and Raman spectra, it is essential to investigate the quality of the beams propagating through the microscope, and the quality of the reflectance responses. For these purposes, small mirrors were fabricated on a Si wafer as 100 nm thick 15×15 μm^2 square gold patches via e-beam lithography, metal evaporation, and lift-off. Figure 2(a) shows an aerial scan of such a structure. The image in Fig. 2(a) results from plotting the photoreceiver output while moving the stage over a 30×30 μm^2 area containing a mirror, then normalizing to the photoreceiver output obtained from the substrate only (normalized reflectance). The bottom two images [Figs. 2(b) and 2(c)] show line cuts along the horizontal and vertical axes in Fig. 2(a).

In Fig. 2(b), one can approximate the $1/e^2$ Gaussian beam intensity width by finding the derivative of the line cut [22]. This derivative is Gaussian fitted, as shown in Fig. 3, for the left and right edges of the line cut. From these plots, one can deduce the $1/e^2$ beam width by finding the maximum of the Gaussian fit, dividing by e^2, and taking the distance between the corresponding points as the beam width. Using both edges yields beam widths of 1.46 μm and 1.88 μm.

Figure 2(a) suggests that the beam width is different in x and y directions (as it would be for an elliptical beam). This is an artifact due to diffraction when the incident polarization is parallel to a sharp edge [23]. The scattering pattern produced by an incident hard polarization (parallel to an edge) is almost isotropic (radiating in all directions), which in turn blurs the edge. Thus, the knife edge technique is polarization sensitive, and one cannot deduce beam ellipticity by examining a single polarization. In this scan, the beam is horizontally polarized, resulting in blurred edges along vertical line cuts [Fig. 2(c)]. Thus, a circular beam may appear elliptical as a result of polarization-dependent scattering from the edges.

The normalized reflectance of the Au mirror was obtained as the ratio of the photoreceiver signal in the region of the mirror (yellow region) to the photoreceiver signal in the blue region (substrate). The ratio of these measured reflectances works out to $R_E = R_{\text{Au}}/R_{\text{sub}} = 2.5$. The complex refractive index of gold is $n_{\text{Au}} = 0.18377 - j3.41313$ [24], and that of silicon is $n_{\text{Si}} = 3.882 - j0.019$ [25] at 632.8 nm. Using the Fresnel reflection equation, the theoretical reflection coefficients in air from these materials are $r_{\text{Au}} = 0.82 - j0.53$ and $r_{\text{Si}} = 0.59 - j0.0016$, and the corresponding reflectances are $R_{\text{Au}} = |r_{\text{Au}}|^2 = 0.94$ and $R_{\text{Si}} = |r_{\text{Si}}|^2 = 0.35$, yielding a theoretical ratio for these reflectances of $R_T = R_{\text{Au}}/R_{\text{Si}} = 2.71$. This is in very good agreement with the experimental result of $R_E = 2.5$ (6.4% error).

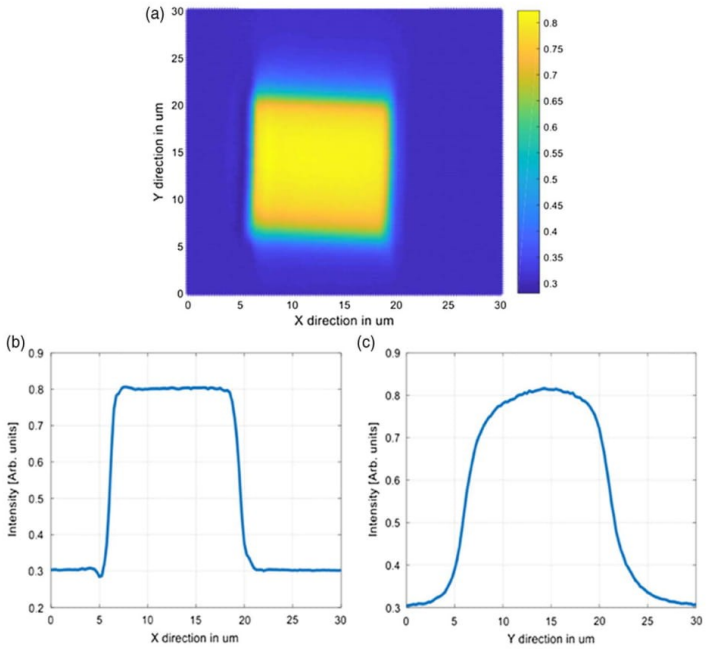

Fig. 2. (a) Aerial reflectance scan from a 30 × 30 μm^2 area containing a mirror consisting of a 15 × 15 μm^2 gold square (yellow region) normalized to the reflectance of the substrate (blue region); (b) horizontal line-cut graph at $y = 15$ μm; (c) vertical line-cut graph at $x = 15$ μm, which are used to measure the spot size of the laser.

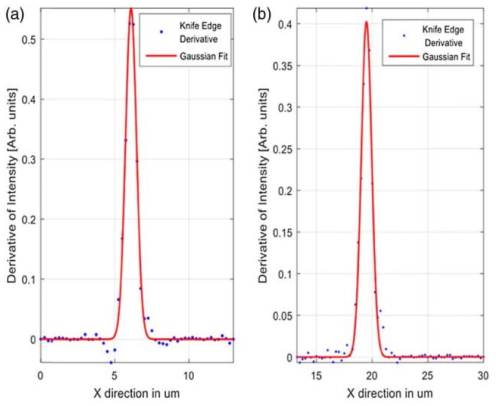

Fig. 3. Derivative of the line cut along the x axis to determine the spot size using the knife edge techniques in Fig. 2. (a) Left edge and (b) right edge.

3. EXPERIMENTAL

A. Structures

Arrays of gold monopole nanoantennas of different resonant frequencies were fabricated on graphene to explore SPP-enhanced Raman scattering using the microscope described in the previous section. Substrates consist of chemical vapor deposition (CVD)-grown graphene covering a 285 nm thick thermally grown SiO_2 layer on 460–500 nm thick single crystal silicon substrate (p-doped). Fabrication of the nanoantenna structures is achieved via e-beam lithography, Au evaporation, and lift-off on the CVD-grown graphene. The structures consist of multiple arrays of 22 nm thick rectangular gold nanorods of different lengths l and widths w. The nanoantennas are arranged over the surface following pitch p and q, as sketched in Fig. 4(a).

Each nanoantenna is part of an array covering an area of 20 × 20 μm^2, which contains 40 × 40 nanoantennas uniformly distributed in a pitch of $p = q = 500$ nm. The design lengths of the nanoantennas vary from 50 nm to 180 nm, and their widths from 30 nm to 45 nm, in steps of 10 nm and 5 nm, respectively, to produce arrays of nanoantennas that resonate at different wavelengths. The nanoantenna designs and resonant wavelengths were selected such that they would overlap spectrally with the Stokes wavelengths of graphene when a pump wavelength of 632.8 nm was used.

Figures 4(b) and 4(c) show scanning electron microscope (SEM) images of fabricated nanoantenna arrays. Comparing the actual length and width to the designed ones reveals a 6 nm to 8 nm increase in edge length dimensions. The SEM images are also essential in determining the quality of

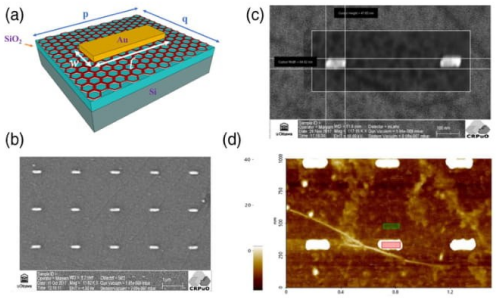

Fig. 4. (a) Sketch of the unit cell of the structure of interest. The sample consists of rectangular gold nanoantenna arrays on CVD graphene covering a 285 nm thick thermally grown SiO_2 layer on a 500 nm thick single crystal silicon substrate. (b) SEM images of nanoantenna arrays on graphene. Each nanoantenna is 85 nm by 48 nm, uniformly distributed in a pitch of $p = q = 500$ nm. (c) SEM image of a nanoantenna array. Each nanoantenna is 121 nm by 43 nm with the same pitch values. (d) AFM image of nanoantennas on graphene. The red region is a scanned area of interest on the top of a nanoantenna, and the green region is on graphene.

the nanoantenna arrays and whether they can be used in the experiments; thus, not all the arrays designed were actually used. Atomic force microscope (AFM) scans were used to measure the roughness of the nanoantennas as well as their thickness. As extracted from Fig. 4(d), the RMS roughness of the substrate is $R_q = 1.1$ nm (green area), whereas the roughness of a nanoantenna is $R_q = 3.011$ nm (red area). The thickness of the nanoantenna is 23 nm, which compares reasonably well with the designed value of 22 nm.

B. Reflectance Responses

The fabricated nanoantenna arrays on graphene were characterized in terms of their reflectance response using the microscope. A strong resonance is observed as shown in Fig. 5(a). The normalized measurements are the reflectance ratios R_A/R_{sub}, where R_A is the reflectance response of the nanoantennas, and R_{sub} is that of the substrate. One can deduce that for nanoantennas of fixed dimensions, the resonance wavelength shifts to longer wavelengths as the length of the nanoantennas increases. The shift is expected by correspondence to classical antennas, where resonance occurs when the antenna length is roughly half a wavelength. Increasing the length thus increases the wavelength at which the antenna is resonant. The effect of width is discussed later in the paper.

In Fig. 5(a), it is seen that arrays consisting of antennas longer than 114 nm (48 nm wide, 22 nm thick) have a very small reflectance ratio. This suggests that Raman scattering will not be enhanced significantly when using such arrays. Theoretical reflectance ratios were computed for the structures and are plotted in Fig. 5(b). We adopted the same nanoantenna thickness and width, allowing only the length to vary while neglecting the (single atom thick) graphene layer. One can see in Fig. 5(b) that the theoretical resonance response follows the same trend as the measured experimental response. There is a slight shift in the resonance wavelength that can be explained by the lack of graphene in the model, and possibly by different optical parameters for Au in the nanoantennas compared to the bulk values. Furthermore, the fabricated nanoantennas are not perfect; they have slightly rounded corners, as can be seen in Figs. 4(b) and 4(c), and slight variations in length.

The theoretical responses were computed via the finite difference time domain (FDTD) method (Lumerical). An infinite array was constructed by repeating the unit cell along x and y with pitch p and q, respectively [see Fig. 4(a)]. An x-polarized plane wave of electric field magnitude 1 V/m illuminates the array from the top at normal incidence. A cubic mesh of $0.5 \times 0.5 \times 0.5$ nm^3 was used in the region around the nanoantennas. Optical constants for Au, SiO_2 and Si were taken from [26]. The reflectance plane was located 2.5 μm above the surface (far field), and the reflectance was calculated as

$$R(f) = \int_s \mathrm{Re}\{P^m(f)\} \cdot \mathrm{d}s / \int_s \mathrm{Re}\{P^s(f)\} \cdot \mathrm{d}s, \quad (1)$$

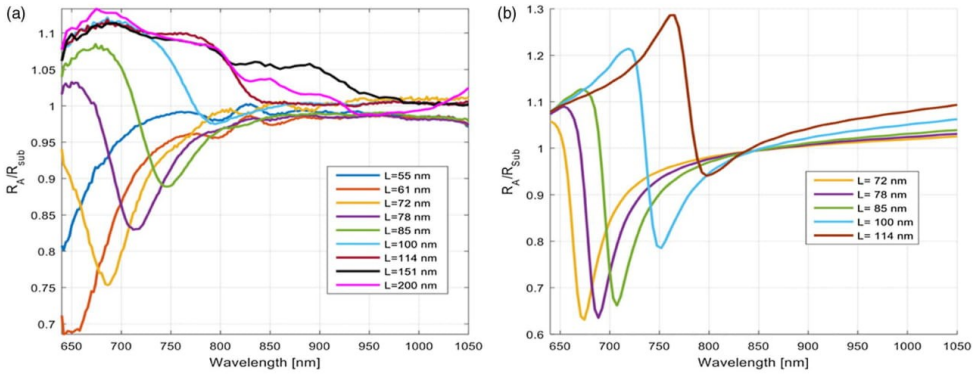

Fig. 5. Normalized reflectance response from arrays of nanoantennas of length ranging from 55 nm to 200 nm. The width and thickness of the nanoantennas are $w = 48$ nm and $t = 22$ nm, respectively. The ratio plotted on the y axis is R_A/R_{sub} as described in the text. (a) Experimental reflectance response of the nanoantennas and (b) theoretical reflectance responses calculated using FDTD simulations.

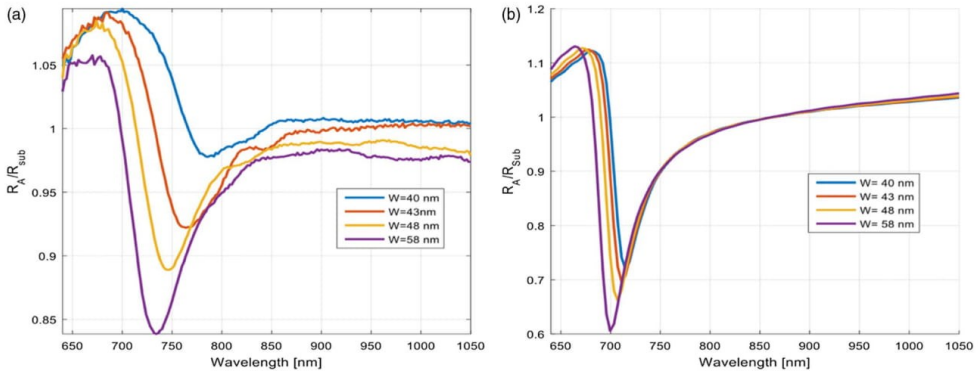

Fig. 6. Normalized reflectance response of arrays of nanoantennas of widths 40 nm, 43 nm, 48 nm, and 58 nm. The length and thickness of the nanoantennas are $l = 85$ nm and $t = 22$ nm, respectively. The ratio on the y axis is R_A/R_{sub} as described in the text. (a) Experimental reflectance response of the nanoantennas and (b) theoretical reflectance responses calculated using FDTD simulations.

where $P^{m,s}$ is the Poynting vector at the monitor and source location, f is the frequency, and S is the area of the reference plane [12].

The width of the nanoantennas is a second design parameter. The response of the nanoantennas to changes in width w, plotted in Fig. 6(a), reveals that increasing the width blue shifts the reflectance resonances. Furthermore, it is evident that the amount of shift decreases as $\Delta w/w$ decreases. This trend is a result of n_{eff} decreasing with an increase in width, following the relation for the resonance wavelength $\lambda_{\text{res}} = 2L_{\text{eff}}n_{\text{eff}}$ [27]. Additionally, modeling results were obtained as shown in Fig. 6(b), from which it is observed that theory and experiment follow the same trends. However, as mentioned earlier, the lack of graphene in the computations, the rounded antenna corners, and slight variations in dimensions within an array can alter the exact resonance position and cause inhomogeneous broadening.

C. Spontaneous Raman Measurements

Spontaneous Raman spectra for our nanoantennas on graphene were obtained using our microscope setup. The first part of this section compares graphene and Si Raman peaks. The second part determines the enhancement in the graphene Raman scattering due to metallic nanoantennas of different geometries.

While building the optical setup, the optics in the spectrometer arm were aligned to the He–Ne laser source. As a result, trying to find the Raman signal of graphene initially was challenging due to the mismatch of the focal points, reflectance angles, and the weakness of the signal. Raman scattering from Si is much stronger than from graphene, so the easiest way to optimize alignments (particularly coupling the fiber to the setup) is to start with the main Si peak appearing at 521 cm^{-1}.

Si–Si bonds are widely recognized as producing a sharp and intense Lorentzian peak at 521 cm^{-1}. Si also has a higher-order peak near 970 cm^{-1}, resulting from two-phonon scattering. To achieve the maximum Si signals, fiber–lens coupling is optimized, followed by optimizing the focal length to the sample. The same process is then repeated to maximize the intensity

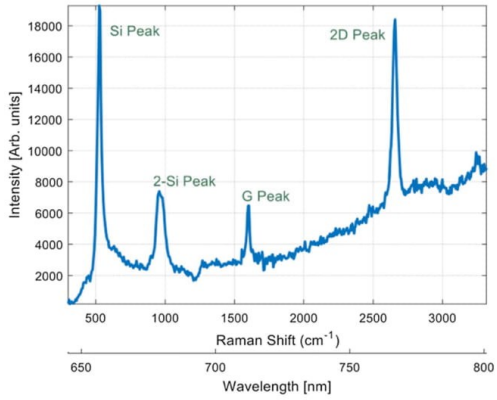

Fig. 7. Experimental Raman spectrum of graphene on SiO_2 on Si without plasmonic nanoantennas. Peaks associated with scattered signal from the underlying silicon wafer as well as graphene are readily resolved.

from graphene only (no nanoantennas)—a typical spectrum obtained using our setup is shown in Fig. 7. After optimization of the alignments, the graphene G and 2D peaks appear, as can be seen in Fig. 7, at shifts of 1594.9 cm^{-1} (703.83 nm) and 2650.9 cm^{-1} (760.35 nm), respectively, with the number in parentheses being the emission wavelength for an excitation wavelength of 632.8 nm. The two silicon peaks at 525.1 cm^{-1} and 970 cm^{-1} can also be seen in Fig. 7, labeled Si and 2-Si, respectively. Comparing this to spectra obtained using a commercial instrument (WiTec, used by Ghamsari *et al.* [6]) reveals very good agreement. During these measurements, our main attention was on the 2D peak. This is due to the fact that all carbon-based materials exhibit a peak at 1594.9 cm^{-1} (adventitious carbon and graphite, e.g.), but only graphene

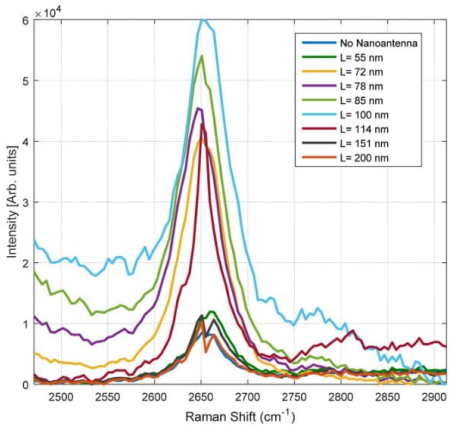

Fig. 8. Raman scattering intensity from the 2D peak of graphene covered by nanoantennas of different lengths (indicated by the color legend) but of fixed width ($w = 48$ nm). The Raman scattering intensity of a sample with no nanoantennas is also shown. Nanoantenna arrays of $l = 100$ nm and $w = 48$ nm produced the highest 2D Raman scattering intensity from graphene.

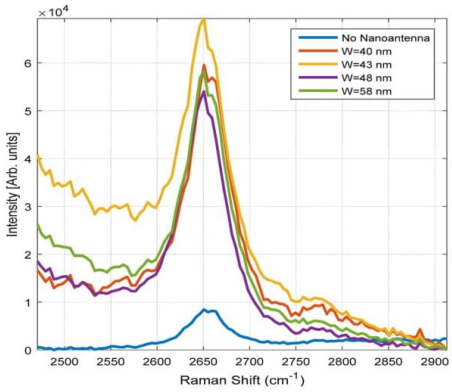

Fig. 9. Raman scattering intensity from the 2D peak of graphene covered by nanoantennas of different widths (indicated by the color legend) but of fixed length ($l = 85$ nm). It was observed that structures with $w = 43$ nm and $l = 85$ nm produced the highest Raman scattering intensity.

does at 2650.9 cm^{-1}. The D peak location and shape, i.e., a single peak at 2650.9 cm^{-1}, indicate that the graphene is in fact single-layer graphene as suggested by Ferrari *et al.* [28].

We now investigate the Raman enhancement produced by Au nanoantennas of various lengths and widths on the graphene signal. Specifically, we investigate the case where the nanoantennas resonate in conjunction with the Stokes emission (2D peak), as opposed to the more common approach of aligning the nanoantennas spectrally with the pump laser.

Eight arrays of nanoantennas of different lengths but fixed width of $w = 48$ nm were used to investigate the enhancement, as shown in Fig. 8. Comparing with Fig. 5, we note that as the resonance of the nanoantennas approaches 760 nm (wavelength of the 2D peak), a corresponding increase in 2D signal is observed. For nanoantennas ranging in length from 55 nm to 200 nm, the peak intensity ranges from 0.8 to 5.2, so the maximum array enhancement factor is ∼6.5 relative to graphene without nanoantennas.

Four arrays of nanoantennas of different widths but of fixed length were also investigated, as shown in Fig. 9. The length of the nanoantennas was selected as one that showed significant enhancement in Fig. 8, $l = 100$ nm. The largest array enhancement factor of ∼8.2 is produced by the array of nanoantennas of width $w = 43$ nm. This observation agrees with Fig. 6, as this array resonates closest to 760 nm. The increase in the background signal is due to the high reflectivity of the metallic nanostructures, which was subtracted in calculations.

It is clear that the nanoantennas enhance the graphene 2D peak when they are spectrally aligned with the latter. The nanoantennas of dimensions $l = 85$ nm and $w = 43$ nm produced the largest array enhancement factor of ∼8.2, as this array resonates near 760 nm (graphene 2D peak). In comparison, the array enhancement factors observed in Ref. [6] for similar structures were 2 to 3. The ability of our microscope to measure the reflectance response of nanostructures *in situ* is advantageous, as the measured responses can be used to select suitable nanostructures for subsequent nonlinear (Raman) experiments.

The enhancement factors deduced in Figs. 8 and 9 are for arrays of nanoantennas. In the present case, each nanoantenna is disposed in a pitch $p = q = 500$ nm. From the array enhancement factors, we seek to determine the enhancement per nanoantenna, or equivalently, the scattering gain S_G [6], which is defined as the ratio of the effective scattering cross section A_{eff} to geometric cross section A_g:

$$S_G = \frac{A_{\text{eff}}}{A_g} = \frac{\left(\frac{P_1}{P_0} - 1\right)A_0}{A_g}. \tag{2}$$

In the above, P_1 is the Raman power radiated by an arbitrary sub-area of the nanoantenna array (e.g., the pump beam area, as long as it is smaller than the array area), P_0 is the Raman power radiated from an un-patterned region of the same area, A_0 is the area of a unit cell, and A_g is the area of one nanoantenna. Thus, P_1/P_0 is the array enhancement factor, determined here directly in Figs. 8 and 9. Figure 10 plots the calculated scattering gains achieved for the 2D peak of graphene for all of the nanoantenna arrays considered in this work. The largest scattering gain achieved is ∼500, for the array comprising nanoantennas of length $l = 85$ nm and width $w = 43$ nm. This scattering gain is significantly higher than the one reported in Ref. [6], which is ∼160.

4. SUMMARY AND CONCLUSION

In conclusion, a novel microscope design was proposed and implemented for studies in surface nonlinear optics. The microscope incorporates a pump laser, a tunable supercontinuum

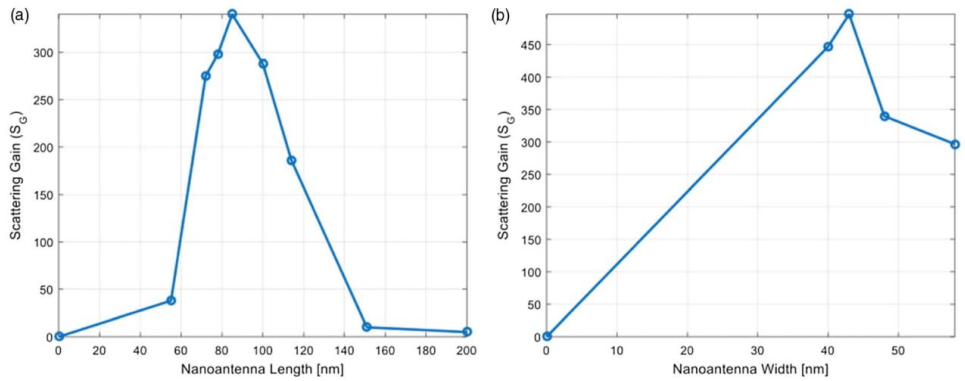

Fig. 10. Scattering gain, S_G, of a single nanoantenna as calculated using Eq. (2) for (a) nanoantennas of different lengths but of width $w = 48$ nm. The largest value of S_G was found to be for nanoantennas of length 85 nm. (b) Similarly, when the nanoantenna width is varied, the largest value of S_G occurs for nanoantennas of width $w = 43$ nm.

source, a viewing port, a photoreceiver arm, and a spectroscopy arm comprising a high-resolution spectrograph and intensified CCD camera. All arms of the microscope were shown to work effectively and efficiently. The constructed system is capable of measuring reflectance responses and capturing nonlinear scattered signals. The microscope was demonstrated by characterizing nanoantenna arrays, and capturing Raman spectra from Si and graphene. The G and 2D peaks for graphene were easily resolved at 1594.9 cm^{-1} and 2650.9 cm^{-1}, respectively.

The microscope was then used to investigate nanostructures consisting of arrays of rectangular Au nanoantennas of length l, width w, and thickness t, on graphene on SiO_2/Si. The antennas were arranged over the surface in pitches p and q. Fabrication was achieved using e-beam lithography, Au evaporation, and lift-off on CVD-grown graphene.

Reflectance responses were obtained with the microscope for nanoantennas of different geometries. It was observed that for a constant width, as the length increases, the optical resonance shifts towards a longer wavelength; alternatively, as the width increases and the length remains constant, the resonance moves toward shorter wavelengths (both trends in agreement with theoretical expectations). It was concluded that Au nanoantennas of dimensions $t = 22$ nm, $l = 85$ nm, and $w = 43$ nm resonate closest to 760 nm, which is significant because this is the Stokes wavelength of the 2D Raman peak of graphene for our pump wavelength of 632.8 nm.

Significant enhancement in Raman scattering intensities from the graphene (2D peak) were observed with the microscope as mediated by SPPs on the nanoantennas. The enhancement is due to the Stokes emission overlapping spectrally with Au nanoantennas. It was found that as the optical resonance of the nanoantennas approached 760 nm, the Raman scattering enhancement factor for the 2D signal increased markedly. The array enhancement factor relative to bare graphene was ~8.2 in the best case, corresponding to a scattering cross-section gain of ~500 per nanoantenna. These factors are significantly larger than what has been reported to date. Naturally this multiarm microscope can also be used in a pump–probe arrangement to examine the effect of exciting the nanoantennas at the Stokes wavelength while recording the Raman spectrum at another wavelength.

REFERENCES

1. M. Kerker, D.-S. Wang, and H. Chew, "Surface enhanced Raman scattering (SERS) by molecules adsorbed at spherical particles," Appl. Opt. **19**, 3373–3388 (1980).
2. A. Otto, I. Mrozek, H. Grabhorn, and W. Akemann, "Surface enhanced Raman scattering," J. Phys. Condens. Matter **4**, 1143–1212 (1992).
3. L. Scarabelli, M. Coronado-Puchau, J. J. Giner-Casares, J. Langer, and L. M. Liz-Marzán, "Monodisperse gold nanotriangles: size control, large-scale self-assembly, and performance in surface-enhanced Raman scattering," ACS Nano **8**, 5833–5842 (2014).
4. A. M. Funston, C. Novo, T. J. Davis, and P. Mulvaney, "Plasmon coupling of gold nanorods at short distances and in different geometries," Nano Lett. **9**, 1651–1658 (2009).
5. M. Khorasaninejad, S. M. Raeis-Zadeh, S. Jafarlou, M. J. Wesolowski, C. R. Daley, J. B. Flannery, J. Forrest, S. Safavi-Naeini, and S. S. Saini, "Highly enhanced Raman scattering of graphene using plasmonic nano-structure," Sci. Rep. **3**, 1–7 (2013).
6. B. G. Ghamsari, A. Olivieri, F. Variola, and P. Berini, "Enhanced Raman scattering in graphene by plasmonic resonant Stokes emission," Nanophotonics **3**, 363–371 (2014).
7. B. G. Ghamsari, A. Olivieri, F. Variola, and P. Berini, "Frequency pulling and line-shape broadening in graphene Raman spectra by resonant Stokes surface plasmon polaritons," Phys. Rev. B **91**, 1–5 (2015).
8. X. Li, W. C. H. Choy, X. Ren, D. Zhang, and H. Lu, "Highly intensified surface enhanced Raman scattering by using monolayer graphene as the nanospacer of metal film-metal nanoparticle coupling system," Adv. Funct. Mater. **24**, 3114–3122 (2014).
9. Y. Zhao, W. Zeng, Z. Tao, P. Xiong, Y. Qu, and Y. Zhu, "Highly sensitive surface-enhanced Raman scattering based on multi-dimensional plasmonic coupling in Au-graphene-Ag hybrids," Chem. Commun. **51**, 866–869 (2015).
10. B. Liu, C. Tang, J. Chen, Q. Wang, M. Pei, and H. Tang, "Dual-band light absorption enhancement of monolayer graphene from surface plasmon polaritons and magnetic dipole resonances in metamaterials," Opt. Express **25**, 12061–12068 (2017).
11. A. Urich, A. Pospischil, M. M. Furchi, D. Dietze, K. Unterrainer, and T. Mueller, "Silver nanoisland enhanced Raman interaction in graphene," Appl. Phys. Lett. **101**, 153113 (2012).

12. S. S. Mousavi, P. Berini, and D. McNamara, "Periodic plasmonic nanoantennas in a piecewise homogeneous background," Opt. Express **20**, 18044–18065 (2012).
13. P. L. Stiles, J. A. Dieringer, N. C. Shah, and R. P. Van Duyne, "Surface-enhanced Raman spectroscopy," Annu. Rev. Anal. Chem. **1**, 601–626 (2008).
14. R. L. McCreery, *Raman Spectroscopy for Chemical Analysis* (Wiley, 2000).
15. M. H. Alshehab, *Design and Construction of a Raman Microscope for Nano-Plasmonic Structures* (University of Ottawa, 2018).
16. J. H. Scofield, "Frequency-domain description of a lock-in amplifier," Am. J. Phys. **62**, 129–133 (1994).
17. M. M. Yang, W. J. Coleman, C. M. Silva, and M. R. Dilworth, "High resolution imaging microscope," Biotechnol. Alia **4**, 1–20 (1998).
18. H. E. Keller, *Objective Lenses for Confocal Microscopy* (2006).
19. D. J. Brady, "Geometric imaging," in *Optical Imaging and Spectroscopy* (Wiley, 2009), pp. 11–50.
20. C. D'Andrea, J. Bochterle, A. Toma, C. Huck, F. Neubrech, E. Messina, B. Fazio, O. M. Marago, E. Di Fabrizio, M. Lamy de La Chapelle, P. G. Gucciardi, and A. Pucci, "Optical nanoantennas for multiband spectroscopy," ACS Nano **7**, 3522–3531 (2013).
21. R. A. Yotter and D. M. Wilson, "A review of photodetectors for sensing light-emitting reporters in biological systems," IEEE Sens. J. **3**, 288–303 (2003).
22. M. A. de Araújo, R. Silva, E. de Lima, D. P. Pereira, and P. C. de Oliveira, "Measurement of Gaussian laser beam radius using the knife-edge technique: improvement on data analysis," Appl. Opt. **48**, 393–396 (2009).
23. C. A. Balanis, "Geometrical theory of diffraction: edge diffraction," in *Advanced Engineering Electromagnetic* (Wiley, 1989), pp. 765–777.
24. P. B. Johnson and R. W. Christy, "Optical constants of the noble metals," Phys. Rev. B **6**, 4370–4379 (1972).
25. D. E. Aspnes and A. A. Studna, "Dielectric functions and optical parameters of Si, Ge, GaP, GaAs, GaSb, InP, InAs, and InSb from 1.5 to 6.0 eV," Phys. Rev. B **27**, 985–1009 (1983).
26. E. D. Palik, *Handbook of Optical Constants of Solids* (Academic, 1985).
27. M. Moskovits, "Surface-enhanced spectroscopy," Rev. Mod. Phys. **57**, 783–826 (1985).
28. A. C. Ferrari, J. C. Meyer, V. Scardaci, C. Casiraghi, M. Lazzeri, F. Mauri, S. Piscanec, D. Jiang, K. S. Novoselov, S. Roth, and A. K. Geim, "Raman spectrum of graphene and graphene layers," Phys. Rev. Lett. **97**, 1–4 (2006).

Appendix B

nature physics

LETTERS

PUBLISHED ONLINE: 3 APRIL 2017 | DOI: 10.1038/NPHYS4087

Plasmon-enhanced high-harmonic generation from silicon

G. Vampa[1,2*], B. G. Ghamsari[1,3], S. Siadat Mousavi[1], T. J. Hammond[1], A. Olivieri[1], E. Lisicka-Skrek[1], A. Yu Naumov[3], D. M. Villeneuve[3], A. Staudte[3], P. Berini[1] and P. B. Corkum[1,3*]

Plasmonic antennas can enhance the intensity of a nanojoule laser pulse by localizing the electric field in their proximity[1]. It has been proposed that the field can become strong enough to convert the fundamental laser frequency into high-order harmonics through an extremely nonlinear interaction with gas atoms that occupy the nanoscopic volume surrounding the antennas[2–4]. However, the small number of gas atoms that can occupy this volume limits the generation of high harmonics[5–7]. Here we use an array of monopole nano-antennas to demonstrate plasmon-assisted high-harmonic generation directly from the supporting crystalline silicon substrate. The high density of the substrate compared with a gas allows macroscopic buildup of harmonic emission. Despite the sparse coverage of antennas on the surface, harmonic emission is ten times brighter than without antennas. Imaging the high-harmonic radiation will allow nanometre and attosecond measurement of the plasmonic field[8] thereby enabling more sensitive plasmon sensors[9] while opening a new path to extreme-ultraviolet-frequency combs[10].

Traditional high-harmonic generation occurs in rare-gas atoms[11,12]. Therefore, early experiments with plasmonic antennas[2] or funnels[3] assumed that high harmonics were generated through local field enhancement in the gas surrounding the nanostructures. However, considerable enhancement is also achieved below and around the antennas, in the material that supports them, as demonstrated in Supplementary Fig. 1.

In recent years, high harmonics have also been generated without nanostructures directly from the bulk of a ZnO crystal[13], and subsequently from various other materials[14–16]—in some cases with peak intensities as low as 10^{11}–10^{12} W cm^{-2}. These intensities are two to three orders of magnitude lower than those required in the gas phase. Not only is the threshold intensity to drive this nonlinear optical process lower than in gases, but the density of solids is ∼1,000 times greater. Hence, generating plasmon-assisted high harmonics within a crystalline substrate should overcome the limit of previous experiments (that is, the need to increase the number of emitters), and improve the longevity of the nanostructures to high-power irradiation (the longevity depends on their structural quality). These advantages have been exploited in a recent experiment conducted on Au-coated sapphire nano-cones[17], where the surface plasmon is adiabatically excited—rather than resonantly excited as in this letter.

Figure 1a shows a sketch of the experimental setup (see also Supplementary Information). We focus infrared femtosecond laser pulses with a central wavelength of 2.1 μm onto an array of Au monopolar nano-antennas fabricated on a thin film of single-crystal Si (500 nm thick) grown on an Al_2O_3 single-crystal substrate (0.5 mm thick). Figure 1b shows a high-resolution scanning electron micrograph of the structures (design and fabrication details are reported in the Supplementary Information). Two designs are implemented, with the antennas parallel to either the [110] or the [100] orientation of the Si crystal (therefore, rotated by 45°).

Figure 2a shows the high-harmonic spectrum produced by antennas in the configuration parallel to the [110] Si lattice. The emission extends from the 5th to the 9th harmonic for an incident intensity (in vacuum) of 3×10^{10} W cm^{-2}. The 11th harmonic, at 190 nm, lies beyond the detectable spectral range of our experimental setup. When the laser polarization is parallel to the antennas (red line), the high-harmonic emission is 5 to 10 times stronger than for polarization orthogonal to the antennas (green line), a configuration that does not lead to field enhancement (see Supplementary Fig. 1). In fact, in the latter case, the high-harmonic power spectrum is almost identical to that emitted from an area without any nano-antennas (black line). Parallel and perpendicular polarizations correspond to equivalent crystal directions; therefore, harmonic emission from bulk silicon is unchanged. Despite the small area covered by the antennas (8%), and the small volume over which the field is enhanced, emission from the array is stronger than the emission from the bulk, that is, off the array (black line). We estimate that the high-harmonic emission is confined to regions of $\sim 20 \times 20\,\mathrm{nm}^2$ on either side of each antenna, and that the high-harmonic emission density in these regions is increased by $\sim 3 \times 10^3$–3×10^4 times with respect to the bulk (see Supplementary Information).

In Fig. 2b we show the intensity dependence of the harmonic radiation yield. Emission from antennas whose axis is parallel to the [100] direction (coloured diamonds in Fig. 2b) is weaker than for those aligned parallel to the [110] direction (coloured circles in Fig. 2b), as similarly measured for the bulk[16] (compare grey circles with grey diamonds in Fig. 2b). Harmonic emission from Au would be irrespective of the orientation of the substrate. Further proof that the emission occurs from Si rather than Au is reported in Supplementary Fig. 3, which shows an image of the 5th harmonic generated from Si (bright areas) and not from micrometre-sized thin Au electrodes (black areas).

For antennas parallel to the [110] direction, all harmonics from the bulk scale approximately as $I^{3.5}$, where I is the incident laser intensity, whereas perturbative nonlinear optics predicts I^n, where n is the harmonic order[18]. A slightly slower scaling, also non-perturbative, has been measured from bulk ZnO (ref. 13). All harmonic orders emitted from the array (coloured lines) show the same intensity dependence, scaling approximately with I^3. This suggests that they are also generated non-perturbatively. The different scaling can arise from saturation of ionization above a threshold intensity (as found for ZnO; ref. 13): because of enhancement, emission from

[1]University of Ottawa, Ottawa, Ontario K1N 6N5, Canada. [2]PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. [3]National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada. *e-mail: gvampa@stanford.edu; pcorkum@uottawa.ca

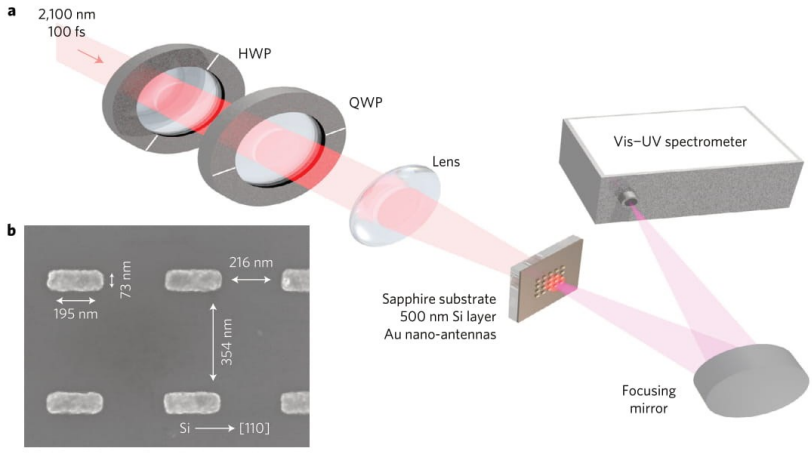

Figure 1 | Experimental setup. a, Laser pulses with a duration of 100 fs and with a centre wavelength of 2.1 μm are focused with a CaF_2 lens on an array of Au monopolar nano-antennas. The antennas are fabricated on a 500-nm-thick single-crystal Si film grown over a 500 μm sapphire substrate (M-plane cut). A combination of a half-wave plate (HWP) and a quarter-wave plate (QWP) compensates for the birefringence of the sapphire substrate to achieve horizontal linear polarization inside the Si film. **b**, High-resolution scanning electron micrograph of an Au nano-antenna array. The antenna major axis is aligned along the [110] direction of the Si crystal, which yields the strongest high-harmonic emission from bulk Si. The length, width and height of the monopoles determine the resonant wavelength to 2.1 μm, the fundamental laser wavelength (see Supplementary Information for details about the design).

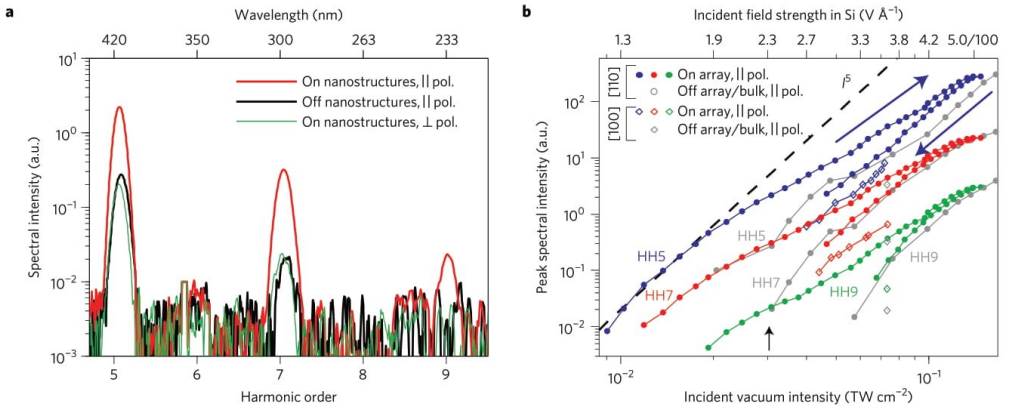

Figure 2 | Non-perturbative high-harmonic spectrum. a, High-harmonic spectrum extending up to the 9th harmonic (233 nm), when antennas aligned along the [110] direction are illuminated with a polarization parallel (red line) and perpendicular (green line) to the major axis of the antennas. The black line is the emission from bulk Si, when the sample is illuminated on the unpatterned material beside the array. The detection limit of our spectrometer extends to 200 nm, which is not enough to measure the 11th harmonic (at 190 nm). The vacuum intensity of the infrared driver is 3×10^{10} W cm^{-2}. **b**, The peak harmonic signal from an array of antennas with their axis parallel to the [110] direction of the Si crystal (coloured circles) and from the bulk (grey circles), for polarization parallel to the axis of the antennas, is plotted as a function of the intensity of the infrared laser. The On/Off contrast decreases with increasing laser intensity (blue arrow pointing up-right), and remains close to unity for decreasing laser intensity (blue arrow pointing down-left). The scaling of bulk harmonics is reversible (not shown). When the axis of the antennas is parallel to the [100] direction instead, harmonic emission is weaker (open diamonds with coloured edges). Bulk emission from the [100] direction (open diamonds with grey edges) is also similarly weaker than from [110], suggesting that harmonics are emitted from Si rather than from Au. All harmonics scale non-perturbatively. The dashed black line, which scales with I^5, is a guide to the eye.

the array corresponds to a higher intensity inside Si, which results in slower scaling. Alternatively, it may also arise from progressive damage to the antennas, as explained below.

The contrast between 'on' and 'off the array' steadily decreases with increasing laser intensity, approaching unity at the highest intensity, and then remains close to one for decreasing laser intensity. We interpret this irreversible behaviour as being due to progressive damage, but it is unclear whether it occurs to the Si or to the antennas (the antennas do not seem to be obviously altered, see Supplementary Fig. 6). Harmonics from unpatterned Si behave reversibly. In higher-bandgap materials, the intensity required for high-harmonic generation increases, and extensive damage to the antennas is expected[19,20]. However, generating harmonics from the substrate, rather than from gas atoms placed near the antennas,

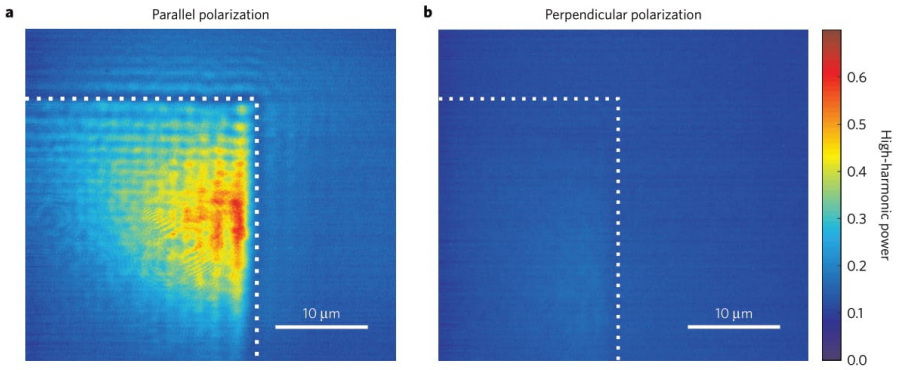

Figure 3 | Imaging high-harmonic emission. Image with the 5th harmonic of a corner of an illuminated array. **a**, The harmonic is emitted only for the portion of the beam that overlaps with the array and for polarization parallel to the antennas' major axis—a proof of field enhancement. **b**, The emission disappears for polarization perpendicular to the antennas' major axis. At this intensity, no harmonic emission is observed from the bulk. The contour of the array is marked by the white dotted line.

offers opportunities to overcome the damage limit by encapsulating the antennas in a transparent dielectric medium, such as SiO_2 or MgF_2. Alternatively, the antennas can be buried in the Si film or fabricated with materials more resilient than Au, such as TiN (ref. 21).

The diffraction of the high-harmonic beam carries information about the spatial distribution of nanoscopic emission regions. The simulations predict an inhomogeneous field enhancement, which is strong at the ends of the antennas and all along the edges (see Supplementary Fig. 1). Although these tiny features cannot be imaged with the infrared beam (unless electrons[22] or near-field probes[23] are used), they should be distinguishable with a short-wavelength high-harmonic beam, whose diffraction limit of $\lambda_{HH}/2$ approaches 100 nm (λ_{HH} is the wavelength of a high harmonic order). Figure 3 shows a corner of the array imaged with a magnification of 160 taken using a lens with NA = 0.75 and the 5th harmonic ($\lambda = 420$ nm), for polarization of the driving field parallel (Fig. 3a) and perpendicular (Fig. 3b) to the axis of the antennas. The emission is clearly brighter over the array for parallel polarization—another indication of enhancement. The blurring of the sharp edge of the array can be used to estimate a resolution of $\sim$1.6 μm (4.6 times larger than the expected diffraction limit, defined by the Rayleigh criterion $r = 0.61\lambda/\mathrm{NA}$). This resolution is not sufficient to resolve individual nano-antennas. A larger-numerical-aperture lens should allow individual regions of field enhancement to be imaged. Lens-less imaging can also be used to exceed the diffraction limit[24,25].

Our results have a number of important implications. First, because high-harmonic emission lasts only a fraction of the optical cycle, imaging the high harmonics can map the sub-cycle dynamics of the collective electronic motion, which influence the temporal[8] and spatial properties[26] of the local plasmonic field. For example, if the antennas are resonant with a 'probing field' that excites plasma oscillations, then the resulting local field will perturb the high-harmonic generation process driven by a non-resonant driver. The perturbation will alter the spatial[27] or spectral[28,29] properties of the harmonic beam. By measuring the beam, we will be able to track the spatial evolution of the local field with attosecond temporal resolution.

Second, the interference between near field and bulk harmonics will contain a wealth of information about the fields in the antennas. For example, as the laser wavelength is scanned across the plasmonic resonance, the amplitude and phase of the near field will be modified relative to the bulk—which is largely unaffected by the frequency sweep. Each harmonic created in the array will inherit the phase shift of the plasmonic field, multiplied by the harmonic order. Thus, this homodyne detection scheme offers a highly sensitive method for studying plasmon responses that differs from previously proposed methods[26], where an electron microscope images electrons that are simultaneously photoionized by attosecond pulses and accelerated in the local field of an infrared pulse. Coherent detection of high harmonics will enable a new generation of plasmon sensors of unprecedented sensitivity, in which small phase shifts of the resonance herald tiny variations in the environment surrounding the antennas.

Third, an inhomogeneous field will accelerate electron–hole pairs differently throughout the volume and even between successive laser half-cycles, since the electrons and holes follow oppositely directed paths. If the variation in the field is sufficient across the path length of the electron and the hole, the inversion symmetry is broken, allowing even harmonics[30] to be produced. In fact, high harmonics can be thought of as a quantum sensor for weak fields or for field inhomogeneity. We do not observe emission of even harmonics. As discussed in the Supplementary Information, the opposite field gradients at the two ends of each antenna average the even harmonic signal in the far field to zero, and the asymmetry on either side of the antennas is predicted to be too small to yield appreciable even harmonic emission. The spatial average would be removed by imaging single nano-antennas with a sufficiently high NA lens, or by designing asymmetric antennas. In addition, the effect of the anisotropic field will be enhanced with a longer wavelength driver. We predict that at 3.7 μm the even harmonic signal from the dominant short-trajectory electron–hole pairs will reach $\sim$3% of the nearest odd harmonic.

To conclude, we have observed non-perturbative harmonic generation from crystalline bulk Si assisted by plasmonic field enhancement from an array of monopole nano-antennas. Using solids overcomes the difficulty of generating macroscopic emission with low-density gas—a scheme that was tried in previous experiments[2–4]. Harmonic emission is sensitive to the orientation of the incident linear laser polarization with respect to the major axis of the antennas. This effect can be exploited to engineer the polarization of the harmonic beam in the near field. For example, circularly polarized harmonics could be obtained by illuminating with a circularly polarized driver two overlapping arrays of antennas, with their major axis perpendicular to one another. Each antenna array will couple to the component of the electric field that aligns with the antenna's major axis, therefore with a quarter-cycle delay between the two arrays. This delay is inherited by the harmonics (multiplied by their order). Overlapping the emission from the two arrays will result in circularly polarized odd-order

harmonics. Complicated geometries, such as metasurfaces[31], will allow even more precise control of many properties of high-harmonic beams, such as amplitude and orbital angular momentum.

Looking forward, light can be confined to volumes smaller than a few cubic nanometres[32]. In this regime, the atomic arrangement of the atoms of the antenna becomes relevant. Such tight confinement provides an opportunity to control strong-field excitation at the level of a single unit cell. This will be a valuable tool to address several conceptual issues in high-harmonic generation from solids, such as the role played by the relative diffusion of the electron and the hole, the plausibility of collisions between neighbouring electrons and holes, and the effect of boundaries and impurities on propagating electron–hole pairs. Combining solid-state technology with strong-field physics, high-harmonic generation can be engineered at will.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of this paper.

Received 3 October 2016; accepted 8 March 2017; published online 3 April 2017

References

1. Stockman, M. I. Nanoplasmonics: the physics behind the applications. *Phys. Today* **64,** 39–44 (February, 2011).
2. Kim, S. *et al*. High-harmonic generation by resonant plasmon field enhancement. *Nature* **453,** 757–760 (2008).
3. Park, I.-Y. *et al*. Plasmonic generation of ultrashort extreme-ultraviolet light pulses. *Nat. Photon.* **5,** 677–681 (2011).
4. Park, I.-Y. *et al*. Generation of EUV radiation by plasmonic field enhancement using nano-structured bowties and funnel-waveguides. *Ann. Phys.* (*Leipz.*) **525,** 87–96 (2013).
5. Sivis, M. *et al*. Nanostructure-enhanced atomic line emission. *Nature* **485,** E1–E3 (2012).
6. Sivis, M. *et al*. Extreme-ultraviolet light generation in plasmonic nanostructures. *Nat. Phys.* **9,** 304–309 (2013).
7. Raschke, M. B. High-harmonic generation with plasmonics: feasible or unphysical? *Ann. Phys.* (*Leipz.*) **525,** A40–A42 (2013).
8. Feist, J., Homer Reid, M. T. & Kling, M. F. Nanoplasmonic near-field synthesis. *Phys. Rev. A* **87,** 033816 (2013).
9. Stewart, M. E. *et al*. Nanostructured plasmonic sensors. *Chem. Rev.* **108,** 494–521 (2008).
10. Cingöz, A. *et al*. Direct frequency comb spectroscopy in the extreme ultraviolet. *Nature* **482,** 68–71 (2012).
11. Li, X. F. *et al*. Multiple-harmonic generation in rare gases at high laser intensity. *Phys. Rev. A* **39,** 5751–5761 (1989).
12. Pfeifer, T., Spielmann, C. & Gerber, G. Femtosecond x-ray science. *Rep. Prog. Phys.* **69,** 443–505 (2006).
13. Ghimire, S. *et al*. Observation of high-order harmonic generation in a bulk crystal. *Nat. Phys.* **7,** 138–141 (2011).
14. Schubert, O. *et al*. Sub-cycle control of terahertz high-harmonic generation by dynamical Bloch oscillations. *Nat. Photon.* **8,** 119–123 (2014).
15. Luu, T. T. *et al*. Extreme ultraviolet high-harmonic spectroscopy of solids. *Nature* **521,** 498–502 (2015).
16. Vampa, G. *et al*. Generation of high harmonics from silicon. Preprint at http://arXiv.org/abs/1605.06345 (2016).
17. Han, S. *et al*. High-harmonic generation by field enhanced femtosecond pulses in metal-sapphire nanostructure. *Nat. Commun* **7,** 13105 (2016).
18. Boyd, R. W. *Nonlinear Optics* (Academic, 2003).
19. Pfullmann, N. *et al*. Bow-tie nano-antenna assisted generation of extreme ultraviolet radiation. *New J. Phys.* **15,** 093027 (2013).
20. Pfullmann, N. *et al*. Nano-antenna-assisted harmonic generation. *Appl. Phys. B* **113,** 75–79 (2013).
21. Guler, U., Shalaev, V. M. & Boltasseva, A. Nanoparticle plasmonics: going practical with transition metal nitrides. *Mater. Today* **18,** 227–237 (April, 2015).
22. Barwick, B., Flannigan, D. J. & Zewail, A. H. Photon-induced near-field electron microscopy. *Nature* **462,** 902–906 (2009).
23. Kawata, S., Inouye, Y. & Verma, P. Plasmonics for near-field nano-imaging and superlensing. *Nat. Photon.* **3,** 388–394 (2009).
24. Zhang, B. *et al*. High contrast 3D imaging of surfaces near the wavelength limit using tabletop EUV ptychography. *Ultramicroscopy* **158,** 98–104 (2015).
25. Chapman, H. N. *et al*. Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nat. Phys.* **2,** 839–843 (2006).
26. Förg, B. *et al*. Attosecond nanoscale near-field sampling. *Nat. Commun.* **7,** 11717 (2016).
27. Kim, K. T. *et al*. Petahertz optical oscilloscope. *Nat. Photon.* **7,** 958–962 (2013).
28. Dudovich, N. *et al*. Measuring and controlling the birth of attosecond XUV pulses. *Nat. Phys.* **2,** 781–786 (2006).
29. Vampa, G. *et al*. Linking high harmonics from gases and solids. *Nature* **522,** 462–464 (2015).
30. Ciappina, M. F. *et al*. High-order-harmonic generation from inhomogeneous fields. *Phys. Rev. A* **85,** 033828 (2012).
31. Yu, N. & Capasso, F. Flat optics with designer metasurfaces. *Nat. Mater.* **13,** 139–150 (2014).
32. Barbry, M. *et al*. Atomistic near-field nanoplasmonics: reaching atomic-scale resolution in nanooptics. *Nano Lett.* **15,** 3410–3419 (2015).

Acknowledgements

We thank D. Crane and B. Avery for technical support. G.V. thanks M. Sivis for insightful discussions. This material is based on work supported by the Air Force Office of Scientific Research under award number FA9550-16-1-0109 and the AFOSR MURI grant number FA9550-15-1-0037. The authors also acknowledge financial support from the NRC, NSERC and CFI/ORF.

Author contributions

G.V. and P.B.C. conceived the experiment; G.V. and T.J.H. performed the high-harmonic measurements; B.G.G. and S.S.M. designed the nano-antennas; B.G.G., A.O. and E.L.-S. fabricated the antennas; A.S. and A.Y.N. maintained the laser source; D.M.V., P.B.C. and P.B. supervised the experiment; all authors contributed to the manuscript.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to G.V. or P.B.C.

Competing financial interests

The authors declare no competing financial interests.

Methods

Fabrication of the antennas. Several square monopole arrays were fabricated by electron-beam lithography, metal evaporation and lift-off on the (001) surface of a 500-nm-thick single-crystal silicon film (http://www.mtixtl.com/sossiliconsapphire-2.aspx). The silicon was grown on an R-plane 500-nm-thick sapphire substrate. The Si/sapphire substrate was first annealed at a temperature of 200 °C in ambient conditions for two hours. Immediately after cool down, two layers of polymethyl methacrylate (PMMA), each with a thickness of 40 nm, were spun on the substrate to produce a re-entrant electron-beam resist bi-layer. PMMA 495 A2 and 950 A2 were used for the first and second layer, respectively. Both layers were baked at 180 °C for one hour and cooled to room temperature before any further processing. The nano-antennas were patterned by electron-beam lithography at 30 keV followed by a one-hour bake at 95 °C. The samples were developed in MIBK/IPA 1:3 at 20 °C. A 2-Å-thick chromium adhesion layer was deposited directly on the substrate followed by evaporation of 200 Å of gold, both using electron-beam evaporation. The metal lift-off took place in an acetone bath at 40 °C, which was sonicated at 30 kHz for approximately one minute. Figure 1 shows a scanning electron micrograph of a nano-antenna array.

Generation and detection of high harmonics. An optical parametric amplifier (Light Conversion OPA TOPAS-Prime) is pumped with a titanium sapphire femtosecond regenerative amplifier (Coherent Legend Elite Cryo) and delivers infrared laser pulses of 100 fs duration with a central wavelength of 2.1 μm at 10 kHz repetition rate. The beam is spatially filtered with a diamond pinhole of 150 μm diameter and refocused into the silicon sample with a magnification of 2/3. The beam waist on the sample is comparable to the size of a nano-antenna array (100 μm). The average power of the laser system is significantly attenuated from $P = 150$ mW to the required values of <1 mW (<100 nJ per pulse) by a combination of a spatial filter and a small aperture placed before the filter. Control over the power is achieved by slightly varying the aperture size. The waist on the sample is unaffected since it is a fixed fraction of the pinhole size.

Harmonics with photon energies above the direct bandgap of Si (at 3.4 eV) are strongly absorbed. Because the region of field enhancement extends only a few nanometres below the Si surface, to detect these harmonics the Si film must face the detector. The projection of the sapphire c axis on the surface aligns to the (100) direction of the Si crystal, and therefore shows birefringence for the desired output polarization parallel to the (110) direction. We used a combination of a half-wave and a quarter-wave plate to ensure that the polarization exiting the sapphire substrate and entering the Si film is linear and aligned along the (110) direction.

The harmonics are detected by focusing them through the slit of a visible–ultraviolet spectrometer from Ocean Optics (model USB2000+).

Data availability. The data that support the plots within this paper and other findings of this study are available from the corresponding author on reasonable request.