Technologies for Room-Temperature Mid-Infrared Photodetection using Graphene

by

Jordan A. Goldstein

B.S., Massachusetts Institute of Technology (2014)
M.Eng., Massachusetts Institute of Technology (2016)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021

© Massachusetts Institute of Technology 2021. All rights reserved.
Technologies for Room-Temperature Mid-Infrared Photodetection using Graphene

by

Jordan A. Goldstein

Submitted to the Department of Electrical Engineering and Computer Science on August 09, 2021, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Abstract

Mid-infrared light is used for thermal imaging, relying on thermal radiation, and chemical analysis, based on vibrational absorption spectra. Ironically, these applications require detector materials and architectures with resilience to thermal noise and infrared-transparent optical materials with minimal vibrational absorption, restricting the mid-infrared material toolbox. 2D materials, which promise to combine high crystallinity with inexpensive and low-temperature processing paradigms, may alleviate some of the material compatibility issues that complicate the design of advanced mid-infrared systems beyond photodetectors and imagers. Graphene is a particularly promising 2D material whose photoresponse has been shown to range from visible to terahertz wavelengths and enjoys fairly mature synthesis and processing technology. Thus, in this thesis, I demonstrate two different mid-infrared systems with novel features enabled by graphene as the optically active material. First, I introduce a multispectral imager concept based on metasurfaces composed of differently-sized, graphene-loaded slot antennas. Here, the tight juxtaposition of sub-wavelength antennas allows broadband transfer and wavelength-sorting of incident mid-IR light into graphene patches with a theoretical efficiency of up to \( \sim 58\% \). I develop a compact circuit model which accurately predicts the absorption spectra of these slot antennas, and demonstrate an electroplating process for fabricating such metasurfaces. This research paves the way towards CMOS-integrable mid-infrared spectral imagers. Second, I demonstrate a chalcogenide glass-on-CaF\(_2\) platform accommodating waveguide-integrated split-gate photothermoelectric graphene photodetectors. These devices achieve waveguide-integrated photodetection at a record-long wavelength of 5.2 \( \mu \)m with a Johnson noise-limited noise-equivalent power of 1.1 nW/Hz\(^{1/2}\). They also feature fast response, with no fall-off in photoresponse up to \( f = 1 \) MHz and a predicted 3-dB bandwidth of \( f_{3\text{dB}} > 1 \) GHz. The demonstrated platform can be readily extended to longer wavelengths and opens the door to distributed gas sensing and portable dual-comb spectroscopy applications. Taken together, these results demonstrate the ability of graphene to enable novel mid-infrared microsystems with unique features and capabilities.
Thesis Supervisor: Dirk R. Englund
Title: Associate Professor of Electrical Engineering and Computer Science
Acknowledgments

First I would like to thank my collaborators and peers—Dr. Marek Hempel, who was kind enough to show me how to do CVD graphene transfer during the early stages of my Ph.D., and who was always diligent in providing us with CVD graphene, as well as Ang-Yu Lu, who continued to grow graphene for us after Marek graduated, and who helped us run pyrolyze the absorber of our bolometer samples, and their advisor, Prof. Jing Kong. Prof. Kong was also kind enough to let us use their AFM tool, which Ang-Yu maintains. I would also like to thank Dr. Hongtao Lin, now professor at Zhejiang University, who helped deposit and gave me advice for working with chalcogenide glass towards the beginning of the chalcogenide glass waveguide-integrated detector project, and Dr. Skylar Deckoff-Jones for continued to help with chalcogenide glass deposition after Hongtao left, as well as their advisor Prof. Juejun Hu. Prof. Hu was also generous enough to allow me to characterize my devices in his lab using his quantum cascade laser and the associated setup, and Hongtao and Skylar were patient enough to get me acquainted with the setup; Prof. Hu’s group has been exceedingly inviting overall and I cannot thank them enough. In addition to Ang-Yu and Prof. Jing Kong, I would also like to thank Dr. Yuxuan “Cosmi” Lin and Elaine McVay as well as their advisor Prof. Tomás Palacios with whom we applied for and received funding together; without their participation in the grant application and maintenance process it’s not clear where the money for my research would have come from. Finally, I would like to thank Drs. Ren-Jye Shiue, Dmitri Efetov and Cheng Peng, who were helpful early in my Ph.D. for their advice and help on the physics and fabrication of graphene.

Second, I would like to thank my research advisor Prof. Dirk Englund, for whom I have also worked since undergrad, for contributing valuable ideas and suggestions over the course of my research, giving me the freedom and funding to pursue my interests, and for helping me keep within sight the bigger picture during my Ph.D. I have on several occasions gotten stuck in a rut with my research and lost sight of the goal while trying to optimize minutiae, and Prof. Englund has never failed to help
me recognize this and put me back on track. Prof. Englund has also always been
diligent in finding and pursuing funding opportunities, without which my research
would not have been possible.

I would like to acknowledge the staff that have maintained the shared resources we
used for fabrication as well as some simulations of our devices. Among the MIT.Nano
(formerly MTL Fabrication), Nanosystems Laboratory, Harvard Center for Nanoscale
Systems, and MTL Computation staff, I would in particular like to thank Mark Mon-
dol, Jim Daley, Dave Terry, and Kurt Broderick for maintaining fabrication equip-
ment that I found particularly critical for my research, namely the Elionix e-beam
lithography tool, the Plasmatherm RIE and Helium Ion Microscope tools, the e-beam
evaporation tool, and the ALD tool, respectively. Mark and Kurt also provided very
useful practical fabrication advice on multiple occasions.

I would like to thank Prof. Rajeev Ram, my academic advisor since undergrad. It
was Prof. Ram’s optoelectronics class that led me choose to pursue this field, and he
always helped provide me with a sense of direction, not only in terms of my research
but also my career. During stressful times, I always left our meetings feeling refreshed
and energized. He has been as helpful of an academic advisor as one could ask for. I
would also like to thank my thesis committee, Profs. Qing Hu and Tomás Palacios,
for agreeing to participate on my committee, and for their helpful feedback about my
research.

Last, I would like to thank the people in my life who have been there for me
over this difficult and emotionally taxing last several years. I would like to thank my
friends at MIT, both in the Quantum Photonics Group and otherwise, for sharing
good times and for lending an ear or a metaphorical shoulder when necessary. I
would like to thank my family back at home for their love and support, for giving
me the resources to be where I am today, and for periodically coming to visit and
spend time in Boston. Finally, I would like to thank my wife, Qin Zhang, who has
been unbelievable supportive during this journey, has offered so much love, through
the highest highs and the lowest lows, who has always been there to listen to me,
provide encouragement when I felt down, and helped me so much during my job
hunt. Especially during this last year, where COVID has cast a shadow over our lives, I don’t know how I would have made it without you.
Contents

1 Introduction

1.1 Key performance metrics of thermal imagers and detectors 26
   1.1.1 “Sensitivity” and how to quantify it 26
   1.1.2 Speed 29
   1.1.3 Spectral range and resolution 29

1.2 State-of-the-art room-temperature detectors and FPAs 30
   1.2.1 Bolometers and other thermal FPAs 30
   1.2.2 HOT photon and other high-speed detectors 34

1.3 Motivations for improving uncooled detector and imager technology 35
   1.3.1 Long-range imaging 36
   1.3.2 Optical gas imaging 37
   1.3.3 Robust imaging in defense with spectral resolution 37
   1.3.4 Robust free-space optical communication 38
   1.3.5 Robust simultaneous localization and mapping 39

1.4 Motivations for waveguide-integrated mid-IR photodetection 39
   1.4.1 Point gas sensors 40
   1.4.2 Liquid-phase chemical sensors 41
   1.4.3 Robust coherent free-space optical communications 41

1.5 State of integrated mid-IR detector research 42

2 Fundamentals and Applications of Graphene 45

2.1 Graphene introduction and photoresponse origin 45
   2.1.1 Bandstructure and optical properties 45
2.1.2 Carrier cooling and photothermoelectric effect .................................. 49
2.2 Comparison with other mid-IR detector technologies ............................. 53
  2.2.1 Sensitivity and speed for communications and DCS ....................... 53
  2.2.2 Cost and adaptability ............................................................... 55

3 Multispectral Imaging Metasurfaces based on Graphene-Loaded Slot Antennas

  3.1 Introduction .................................................................................. 60
  3.2 Single-antenna results ................................................................. 63
    3.2.1 Single antenna circuit model ................................................ 63
    3.2.2 Circuit model comparison with FDTD results ....................... 65
  3.3 Metasurface results ...................................................................... 66
    3.3.1 Baseline design ................................................................. 66
    3.3.2 Effect of pitch variation ...................................................... 69
    3.3.3 Discussion of angular acceptance ...................................... 72
    3.3.4 Different numbers of spectral channels ................................. 75
    3.3.5 Metasurface from slitted antennas .................................... 79
  3.4 Sensitivity performance estimation ................................................... 79
  3.5 Methods ....................................................................................... 81
    3.5.1 Simulation details ............................................................. 81
    3.5.2 Impedance model details ...................................................... 83
    3.5.3 Detectivity estimation ......................................................... 83
  3.6 Fabrication ................................................................................... 85
    3.6.1 Flip-chip fabrication process ............................................... 85
    3.6.2 Cross-polarized scattering measurements ............................. 87

4 Waveguide-Integrated Mid-Infrared Photodetection using Graphene on a Scalable Chalcogenide Glass Platform

  4.1 Introduction ................................................................................. 92
  4.2 Results ....................................................................................... 95
    4.2.1 Device design ................................................................. 95
4.2.2 Raw measurement results ........................................... 95
4.2.3 Comparison with device model ..................................... 96
4.2.4 Frequency response .................................................. 98
4.2.5 Noise measurements .................................................. 99
4.2.6 Estimated gas sensing performance .............................. 100
4.3 Discussion ................................................................. 100
4.4 Methods ................................................................. 102
  4.4.1 Photodetector fabrication ......................................... 102
  4.4.2 Optical setup, alignment, power and loss calibration ......... 104
  4.4.3 Device modelling .................................................... 106
  4.4.4 Model parameter extraction and modelled resistance and transmittance ................................................... 109
  4.4.5 Gate capacitance measurement ................................... 112
  4.4.6 Measurement of TLM structure for gate resistance extraction 113
  4.4.7 Calculation of gas sensitivity ..................................... 114
4.5 Additional information ................................................ 116
  4.5.1 Modulators and hysteresis ....................................... 116
  4.5.2 Optical absorption of graphene for longer wavelengths ..... 117

5 Discussion and outlook ................................................. 119
  5.1 Graphene photodetection in general ................................ 119
    5.1.1 Plasmonic enhancements to device performance ............ 119
    5.1.2 Bilayer graphene as an active material ...................... 121
    5.1.3 Improvement of graphene processing technology ........... 121
  5.2 Multispectral imaging metasurfaces .............................. 122
    5.2.1 What could we learn from an experimental demonstration? 122
    5.2.2 Prospects for CMOS integration ............................... 123
    5.2.3 Metasurface images based on materials besides graphene ... 124
    5.2.4 Imaging metasurfaces in the terahertz ....................... 125
  5.3 Waveguide-integrated mid-IR detectors and chip-scale gas sensing 126
5.3.1 Extended wavelength operation ..................................... 126
5.3.2 Device enhancements .................................................. 127
5.3.3 Demonstration of gas sensing ....................................... 128
5.3.4 Scalability ............................................................... 128
List of Figures

1-1 Spectrally-averaged detectivity $D_0^*$ versus time constant $\tau$ for several individual noise sources and for all noise sources combined in VOx using parameters from Niklaus et al[1]. I also plot with dashed lines the limiting detectivities associated with only the radiative thermal fluctuation of bolometer pixels at ambient and VO$_x$ operating temperatures. The vertical black line represents the longest possible $\tau$ for the assumed value of $c_A$ in the case of no conductive thermal exchange between the pixel and its environment (I.E., only radiative thermal exchange). Points $(\tau, D_0^*)$ extracted from a set of exemplary published room-temperature thermal detectors, VO$_x$-based and otherwise, illustrate the validity of our claimed performance limit[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. I do not address readout (ROIC) noise here, which in practice cannot be neglected and likely accounts for the difference between our noise model and the best reported detectivities in practice[1]. 33

2-1 Graphene’s hexagonal crystal structure. By AlexanderAlUS and licensed under CC-BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0/deed.en). No changes made. ................................. 46

2-2 Graphene’s bandstructure at charge neutrality. Its Fermi level coincides with the intersection between its conduction and valence bands, which meet at a point in a conical manner. By Ponor and licensed under CC-BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0/deed.en). No changes made. ................................. 47
2-3 a) Real and b) imaginary parts of graphene’s optical conductivity versus frequency at various Fermi levels for $T = 300\,\text{K}$ and $\tau = 100\,\text{fs}$. c) Imaginary/real ratio of the conductivity. The further from zero, the more reactive the graphene is; positive values correspond to capacitive behavior, and negative values correspond to inductive behavior, so-called “kinetic inductance”.

2-4 Depiction of a split-gate PTE graphene photodetector. Metal back-gates are used to tune the Fermi level on each side of the graphene channel resulting in a $p_n$-junction and thus two different Seebeck coefficients, $S_1$ and $S_2$, on the two sides of the device. Concentrated illumination along the center of the device heats up electrons in the channel relative to the lattice, with $\Delta T_{\text{el}} = 0$ boundary conditions at the source and drain. The asymmetric Seebeck coefficient profile combined with the elevated electron temperature in the channel gives rise to a thermal electromotive force between the source and drain as described by Eqn. 2.3.

3-1 a) Illustration of a broadband absorbing slot antenna metasurface consisting of six differently tuned slot antennas tiled with sub-wavelength periodicity. The graphene patches are color-coded by antenna length, and the diagram is drawn to proportion based on the device dimensions used to produce Fig. 3-6a. b) Depiction of a single graphene-coupled slot antenna-based photodetector based on the photothermoelectric effect. c) Depiction of the gating scheme of the proposed photodetector. Asymmetric placement of the channel with respect to the slot results in a p-n junction in the center of the channel.
3-2 a) Graphene-loaded slot antenna with physical features corresponding to the components in circuit b) labelled. b) Circuit schematic of a graphene-coupled slot antenna. $V_A$ represents incoming light, $Z_A$ is the radiation impedance of the slot aperture, $Z_{gr}$ is the impedance of the graphene sheet and $Z_{wg}$ is that of the slot, effectively a short-circuit waveguide stub.

3-3 Wavenumber dependence of the impedances in the circuit in Figure 2a. a) Linear scaling. b) Log-abs scaling. The antenna featured here has dimensions $d = 2.0 \mu m$, $h = 5.5 \mu m$, and $w = 0.4 \mu m$.

3-4 a) Comparison of simulated and modeled $\eta_{gr}$ for Antenna Design A. Modeled antennas are 400 nm wide. Dashed lines represent the fullwave FDTD absorption results, while the solid lines represent the impedance model results. b) $\eta_{gr}$ versus wavenumber for various values of the load sheet conductance for Antenna Design A, where $\sigma_{gr}$ is the optical conductivity of intrinsic graphene. Dashed lines represent the fullwave FDTD absorption results, while the solid lines represent the model results. The antenna featured here has dimensions $d = 2.0 \mu m$, $h = 5.5 \mu m$, and $w = 0.4 \mu m$.

3-5 a) Cross-section of Antenna Design B, as well as dimension definitions. b) FDTD-simulated $\eta_{gr}$ spectra for various such antenna dimensions. In all cases, $w = 0.6 \mu m$, $d_i = 60 \text{ nm}$ and $w_i = 150 \text{ nm}$.

3-6 a) Graphene and total absorption efficiencies of a six-antenna meta-surface based on Antenna Design A with a 7\text{ \mu m} by 12.667\text{ \mu m} unit cell. The colored curves represent the contributions of each antenna to the overall graphene absorption of the metasurface. The curve colors here match the antenna colors in Fig. 3-1a. Shorter antennas yield higher resonance wavenumbers. b) Same as a), but from Antenna Design B. Here the unit cell, which contains four different antenna sizes, is reduced to 3.333\text{ \mu m} by 12\text{ \mu m}.
3-7 Total absorption efficiency (including both graphene and metal contributions) and individual antenna contributions for both Antenna Design A- and B-based metasurfaces corresponding to Fig. 3-6.

3-8 a) Total graphene absorption efficiency versus wavenumber for metasurfaces with varied unit cell pitch. The legend indicates the width of the unit cell as illustrated in Fig. 3-1a. b) Mean absorption efficiency for each of the curves in a), averaged between 1050 cm\(^{-1}\) and 1600 cm\(^{-1}\). c) Unit cell dimensions used to obtain the results in a) and b) and the corresponding maximum wavenumbers for which normally incident light experiences no higher order diffraction. The unit cell x and y pitch were varied together to attain a roughly uniform antenna-antenna proximity.

3-9 Graphene absorption efficiency and individual antenna contributions for Antenna Design A-based metasurfaces of varied pitch. Full unit cell dimensions for each case are given in Fig. 3-8.
3-10 Different representations of the on-resonance far-field radiation pattern of a 6.5 µm-long slot antenna, which is qualitatively very similar to those of antennas of other lengths. a) 3D-representation of the far-field directivity pattern. \( D(\theta, \phi) \) goes to zero for light incident along the long axis of the aperture. b) Directivity vs. elevation angle \( \theta \) along the azimuth of the antenna’s long axis, \( \phi = 90^\circ \). c) Directivity in directional cosine space. Note also that \( k_x = k_0 u \) and \( k_y = k_0 v \) where \( k_x \) and \( k_y \) are the \( x \)- and \( y \)-components of the incident wavevector. d) Depiction in \( k_x k_y \)-space of the light cone edges (colored rings) and the specular reflection-only incident light directions (colored patches) for three different wavenumbers for the 7 µm × 12.667 µm 6-antenna unit cell design in Fig. 3-1a. The black dots represent the reciprocal lattice points of the metasurface. For each wavelength, the specular reflection-only region is the set of remaining \( \{k_x, k_y\} \) after subtracting from the light cone all copies of the light cone transposed by all nonzero reciprocal lattice vectors.

3-11 Graphene absorption efficiency and individual antenna contributions for a perfectly conducting metasurface with various off-angle excitation directions. Using Perfect Electrical Conductor (PEC) as the metasurface material allows me to use a coarser mesh of 40 nm, as the electric field does not penetrate the conductor (the “skin depth” which requires careful modelling for realistic metal). I also adjust the metasurface dimensions to align all features to the mesh: The unit cell is 7.2 × 12.64 µm, and the antenna lengths are 3.4, 3.8, 4.4, 5.04, 5.8, and 7.36 µm. The topmost plot shows the absorption spectra for normally incident light. The second row depicts the case of light incident perpendicular to the antennas’ long axes (along the line of maximum directivity), and the third row depicts light incident perpendicular to the antennas’ short axes.
3-12 Graphene absorption efficiency and individual antenna contributions for Antenna Design A-based metasurfaces with varied numbers $N$ of antennas of different lengths. The lower-right table shows the arrangements of the antennas within the unit cell, with 1, 2...$N$ corresponding to the shortest, second shortest, etc. up to longest antenna. The table also lists the lengths of the antennas in each case, chosen roughly to maintain a constant degree of frequency-space overlap between adjacent spectral channels as simulated on an individual antenna basis. Antennas are placed on a $1.167 \mu m \times 6.333 \mu m$ grid as in the reference metasurface design plotted in Fig. 3-6a except for the last spectrum, which is placed on a $0.875 \mu m \times 6.333 \mu m$ grid to match the overall unit cell size and hence $\lambda_c$ of the reference design.

3-13 Graphene absorption efficiency and individual antenna contributions for Antenna Design B-based metasurfaces of varied pitch. The unit cell dimensions are $2.667 \mu m \times 11.5 \mu m$, $3.333 \mu m \times 12 \mu m$, $4.333 \mu m \times 12.667 \mu m$, $3.0 \mu m \times 13.333 \mu m$, $7.333 \mu m \times 14 \mu m$.

3-14 Graphene absorption efficiency and individual antenna contributions for Antenna Design B-based metasurfaces with varied numbers $N$ of antennas of different lengths. The lower-right table shows the arrangements of the antennas within the unit cell, with 1, 2...$N$ corresponding to the shortest, second shortest, etc. up to longest antenna. The table also lists the lengths of the antennas in each case, chosen roughly to maintain a constant degree of frequency-space overlap between adjacent spectral channels as simulated on an individual antenna basis. Antennas are placed on a $0.833 \mu m \times 6 \mu m$ grid as in Fig. 3-6b.

3-15 Illustration of how Antenna Design B could be implemented in the wiring layers of a CMOS chip.
3-16 Graphene absorption efficiency and individual antenna contributions for a 9 µm by 13.333 µm metasurface with varied minimum FDTD mesh size as discussed in the Methods section of the main paper. The inset plots the mesh dependence of the mean graphene absorption efficiency in the 1050 cm\(^{-1}\) to 1600 cm\(^{-1}\) range.

3-17 a) Graphene absorption efficiency and individual antenna contributions for a 7 µm by 12.667 µm metasurface with the same geometric parameters used to generate Fig. 3-6a, but with the graphene doped to 0.05 eV which decreases the graphene sheet conductivity to a degree, especially for longer wavelengths. Here, the mean efficiency averaged between 1050 cm\(^{-1}\) and 1600 cm\(^{-1}\) is 33%. This doping level is chosen to maximize the Seebeck coefficient for the graphene model used to approximate the device detectivity as discussed in the main text. b) Same as a), but for the 4-antenna Antenna Design B metasurface design used to generate Fig. 3-6b as discussed in the main text.

3-18 Process flow for gold slot antennas.

3-19 a) Scanning electron micrograph of SU-8 fins, the mold for the gold slots. b) Scanning electron micrograph of the completed gold metasurface with a sheet of graphene transferred on top.

3-20 Comparison between measured and simulated cross-polarized scattering for rectangular arrays of individual antennas with \(w = 240\) nm, \(d = 1.6\) µm, and \(h\) varied from 1.75 µm to 4.0 µm in increments of 250 nm. Each array had a pitch of \(2h \times 2h\), accounting for the reduced scattering for longer antennas. Inset: Earlier slot antenna fabrication process gave rise to gold nanoparticles precluding graphene transfer, shown here along the edges of a slot.
4-1 a) Illustration of the device cross-section perpendicular to the waveguide axis. The optical mode supported by the GSSe waveguide evanescently couples to and is absorbed by the graphene channel, which is gated by two graphene back-gates to induce a p-n junction. b) Optical image of the device depicting source, drain and gate contact pads. c) Depiction of the optical guided mode at $\lambda = 5.2 \mu m$.

4-2 a) Measured zero-bias photovoltage produced by the device as a function of the two gate voltages. b) Total device resistance as a function of the two gate voltages. c) Lock-in signal reflecting power measured by an InAsSb photodetector at the focal point of the output facet collection lens, used to monitor transmission of the device as a function of the gate voltages. d, e, f) Plots of line sections indicated with dashed lines in figures a, b, and c, respectively.

4-3 a, b) Contour plots of the a) measured and b) modelled responsivity maps of the device, evaluated with $\tau_{DC} = 3.5 \text{fs}$, $\tau_{IR} = 40 \text{fs}$, $\sigma_n = 2 \times 10^{12} \text{cm}^{-2}$, $\tau_{eph} = 50 \text{ps}$, and $\alpha_e = 2.5 \text{mm}^{-1}$. c) Electron temperature increase $\Delta T_{el}$ and absorbed optical power per area $\dot{Q}$ profiles in the graphene channel per guided optical power at gate voltages of $\{-2.35 \text{V}, 0.35 \text{V}\}$, chosen to maximize the modelled photoreponse, and other parameters as above.

4-4 a) Comparison of the frequency response of the photodetector with that of the laser current modulation itself. The consistency between the two indicates that the photodetector frequency response exceeds 1 MHz. Inset: Simulated GHz-range photodetector frequency response, with and without considering the impact of the electron-phonon cooling time $\tau_{eph}$. b) Measured noise spectral density versus resistance and corresponding Johnson noise spectral density of Device B, without illumination, for the 49 pairs of gate voltages $\{V_{g1}, V_{g2}\}$ where each $V_{gn}$ is varied from $-6 \text{V}$ to $6 \text{V}$ in steps of $2 \text{V}$. Measurement was performed at $T = 293 \text{K}$. 


4-5  a) Depiction of the in- and out-coupling beam path for optical and optoelectronic measurements. Lengths are not to scale; the chip and waveguide are magnified for clarity. b) Output photodiode signal versus waveguide length for four waveguide kickback structures. The decaying exponential fit reveals the waveguide loss and the in-coupled power, subject to a conversion efficiency accounting for the collection efficiency of the collection optics and the voltage responsivity of the photodiode. Inset: Layout of the kickback structures. The longest of the five kickbacks was not included in the fit as the out-coupled signal was indistinguishable from stray light.

4-6  a, b) Contour plots of the measured resistance and transmittance maps as a function of gate voltages. c, d) Contour plots of the modelled resistance and transmittance maps as a function of gate voltages using the graphene quality and waveguide loss parameters listed in Table 4.1

4-7  Transmission line method (TLM) device used for gate capacitance and contact resistance measurements.

4-8  Capacitance versus measurement frequency for five devices.

4-9  Y-intercepts of resistance versus channel length as a function of gate voltage offset from the resistance peak for the five TLM devices.

4-10 a) Illustration of the modulator cross-section perpendicular to the waveguide axis. b) Transmission versus applied voltage for one of my modulators using a zigzag voltage sweep to reveal hysteresis. Sweep direction is color-coded and indicated with arrows.

4-11 Contour plot of real part of graphene’s infrared optical conductivity as a function wavelength and mean carrier concentration, assuming a carrier concentration spread of $\sigma_n = 2.0 \times 10^{12} \text{cm}^{-2}$ and a Drude scattering time of $\tau_{IR} = 40 \text{fs}$.
List of Tables

1.1 Comparison of LWIR photon detectors and other technologies for fast photodetection in the LWIR range. .......................... 34

3.1 Estimated sensitivity figures for graphene imaging array. ............. 80

4.1 Device parameters and approximate values .......................... 98

4.2 Comparison of the detector with inferred room-temperature performance metrics for two HgCdTe photodiodes optimized for two different wavelengths (from [14]) and a VO\textsubscript{x} bolometer (from [2]) available off the shelf. For the photodiodes, the NEP is extrapolated from the specified detectivity for a detector scaled down to match the size of a diffraction-limited spot with NA = 0.3, which is the acceptance NA of these detectors. For the bolometer, I give the NEP of a single 17\,\mu m \times 17\,\mu m bolometer pixel as calculated from the specified noise-equivalent temperature difference as described in Rogalski[15]. ....... 101
Chapter 1

Introduction

Thermal imagers (shortened to “imagers” here unless otherwise specified), referring to the class of imagers sensitive in the 3–8µm mid-wave infrared (MWIR) and/or 8–15µm long-wave infrared (LWIR) atmospheric transparency bands, enjoy extensive application in areas such as security and defense[15, 16], remote gas sensing[17, 18, 19], medical diagnostics[20], HVAC and electrical diagnostics and maintenance[21, 22], and research, among others. Mid-IR (MIR) photodetectors (“detectors”), which I use here to describe single-pixel photodetectors operating within the above wavelength ranges, are used extensively in infrared spectroscopy[23], gas sensing[24, 25] and environmental monitoring[26, 27], and are being investigated for free-space optical communication (FSO) applications due to potentially lower atmospheric attenuation at these wavelengths [28, 29]. Thermal imagers and detectors fall neatly into two categories: Photon detectors, in which optically excited charge carriers in a semiconductor are directly converted into a photovoltaic or photoconductive signal, and thermal detectors, in which radiative heating of a sensitive pixel is read out via some temperature-dependent property, such as a resistance in the case of bolometers[15]. Photon detectors are usually cooled to cryogenic temperatures to suppress thermal charge carrier generation that would otherwise be a severe source of noise; this, and the cost of epitaxial materials (typically a III-V semiconductors or HgCdTe), contribute to their high cost. Thermal detectors, on the other hand, are usually operated at ambient temperature, and are relatively inexpensive, but they are generally slower
and less sensitive than cooled photon detectors.

1.1 Key performance metrics of thermal imagers and detectors

Several performance metrics are relevant when discussing thermal imagers and detectors, and the relations between these metrics are often subtle and complex. Here, I will summarize the most important metrics to the extent necessary to motivate my research.

1.1.1 “Sensitivity” and how to quantify it

In discussing the sensitivity of MIR detectors and imagers it is important to keep in mind that matter at ambient temperatures constantly emits blackbody radiation with significant intensity in the MIR (for this reason also often referred to as “thermal IR”) wavelength range[30]. Therefore in the following discussion, the “signal” one refers to in calculating noise equivalent power, signal to noise ratio and related metrics must be understood to mean the “desired signal” one wishes to measure, not including any excess radiative exchange between the sensor and internal camera or detector components.

Noise-equivalent power (NEP)

The noise-equivalent power of a detector is defined as the power spectral density (“spectrum” referring not to wavelength spectrum, but rather to fluctuations in optical power at comparatively long time scales) of incident light that yields an output signal power spectral density equal to the noise spectral density of the detector[15]. With dimensions of W/Hz\(^{1/2}\), it may be thought of as the minimum input optical power that can be distinguished from noise. NEP is calculated as \(\text{NEP} \equiv \sqrt{\mathcal{S}_v/R_v} \) or \(\sqrt{\mathcal{S}_i/R_i}\), where \(\mathcal{S}_v\) or \(\mathcal{S}_i\) is the output noise spectral density of the detector with the relevant dimensions for the given detector (V\(^2\)/Hz or A\(^2\)/Hz, respectively), and \(R_v\) or \(R_i\) is
the responsivity of the detector with dimensions of V/W or A/W, respectively.

**Detectivity (D*)**

The majority of imager and detector technologies, including bolometers and photoconductive and photovoltaic photon detectors, feature an $A_d^{1/2}$ dependence of NEP where $A_d$ is the detector area, and it is thus convenient to define the detectivity $D^* \equiv \sqrt{A_d}/\text{NEP}$, yielding a sensitivity metric that is scale-invariant and correlates positively with “sensitivity”. $D^*$ is invariably quoted in units of cm Hz$^{1/2}$/W, also referred to as “Jones” after R. C. Jones who proposed the concept in the 1960s[15]. It is the metric most commonly used when comparing different detector technologies.

I shall point out an ambiguity in the definition of $D^*$ that will become relevant later on. As defined above, $D^*$ is a frequency-dependent metric that falls off above the cutoff frequency of the detector as well as at low frequencies for detectors that exhibit $1/f$ noise[31]. Some authors do not account for the frequency dependence of the responsivity $\mathcal{R}$ and instead define $D_0^* \equiv \sqrt{A_d} \mathcal{R}(0)/\sqrt{\mathcal{S}(f)}$ for consistency with the definition of noise equivalent difference in temperature (NEDT, defined below)[32].

**Noise-equivalent difference in temperature (NEDT)**

As the canonical goal of thermal imaging is to image the temperatures of a scene, it is natural to use the temperature distinguishing ability of a thermal imaging system as a key performance metric. This is quantified as the NEDT, which Rogalski[15] defines as $\text{NEDT} \equiv V_{\text{rms}}/(dV/dT)$, where $V_{\text{rms}}$ is the RMS noise voltage of the focal plane array (FPA) after sampling at the frame rate and $dV/dT$ is the change in pixel voltage with respect to changes in scene temperature at low measurement frequencies (below the cutoff frequency of the detector). Some authors distinguish between spatial and temporal NEDT depending on whether $V_{\text{rms}}$ is defined as spatial noise across an array of pixels or temporal noise for a single pixel[33]. For digital imagers, NEDT may be defined not in terms of voltages but instead in terms of post-ADC digital units, or least significant bits (“LSBs”), allowing digitization noise to be accounted for. In practice, NEDT is empirically determined for an imaging system based on
established protocols rather than calculated in terms of more fundamental system and
detector parameters [34]. This allows manufacturers to quote values that best reflect
temperature distinguishing performance for typical use cases and avoids ambiguities
in NEDT’s technical definition, such as how 1/f noise (which in theory produces a
logarithmically infinite RMS noise voltage) is handled.

The limitations of today’s readout integrated circuits (ROIC), silicon integrated
circuits that read out a video stream from an array of detector pixels, are crucial
in understanding the thermal imager tradespace. Readout of a thermal imager pixel
invariably involves accumulating electrical charge representing the optical power incident
on the pixel (in the case of photon detectors, the number of electrons accumulated
in an exposure is precisely the number of incident photons multiplied by the quantum efficiency of the detector) in a capacitor over the course of a given integration
time \( t_{int} \)[35]. In most applications involving cooled photon detectors, the number of
photons incident on a pixel within a single frame period would far exceed the charge
handling capacity of this capacitor, and thus the integration time is artificially short-
ened so as to not overfill it. It is then typically found that the NEDT of the imager is
limited not by the detectivity of the detector itself, but by shot noise associated with
the ROIC charge handling capacity, which is independent of the imaging optics[36].
In contrast, for imagers based on thermal detectors such as bolometers, the bolome-
ter pixel itself typically produces considerable noise that dominates the NEDT of the
imager. Here, the NEDT also depends on the f-number of the imaging optics, as a
larger focusing lens is capable of focusing more light from the scene onto each pixel
and producing a greater photoresponse[15]. For this reason, bolometer manufactures
quote NEDT at a given f-number, typically \( f/1 \).

The intended takeaway from this discussion is that neither detectivity nor NEDT is
a one-size-fits-all metric for comparing the sensitivity of different detector and imager
technologies, especially when comparing thermal detectors with photon detectors.
Because of the complex interplay of different noise sources and design constraints,
any discussion of the relative sensitivity of different technologies cannot be abstracted
away from the application at hand.
1.1.2 Speed

The speed of a detector can be expressed in terms of the -3 dB cutoff of its frequency response, \( f_{3\,dB} \), or the response time, \( \tau = 1/(2\pi f_{3\,dB}) \). For bolometers, \( \tau \) is typically on the order of 10 ms, limited by the thermal time constant of the suspended pixels[37]. The speed of photon detectors, in contrast, is limited by electronic time scales such as carrier diffusion and recombination times as well as RC times constants. Typical orders of magnitude for \( \tau \) are 0.1–10 ns for HgCdTe photodiodes and 1–100 ps for quantum well infrared photodetectors (QWIPs)[38, 39].

1.1.3 Spectral range and resolution

Spectral range refers to the range of wavelengths to which a detector is sensitive. Since the photoresponse of thermal detectors is based on heating from absorbed light, very broad spectral ranges are possible, typically limited by the choice of focusing optic. In contrast, for photon detectors based on bulk semiconductors, the spectral range is limited by the bandgap of the semiconductor bandgap, with photoresponse falling off rapidly for photon energies lower than the bandgap. For reduced-dimensionality photon detectors such as QWIPs, the photoresponse is limited to wavelengths in the vicinity of the relevant intersubband transition. It is important to point out that narrower-bandgap photon detectors and higher operating temperatures are both associated with lower detectivities, as narrower bandgap materials feature greater thermal carrier generation rates due to Auger, radiative and Shockley-Read-Hall processes[15]. Therefore, one typically selects a photon detector with a bandgap just narrow enough to sense the longest wavelengths of interest. For instance, a series of commercially available HgCdTe photodiodes designed for room-temperature operation with cutoff wavelengths of 5.0 µm, 8.0 µm, and 10.6 µm feature nominal detectivities of \( 9 \times 10^9 \), \( 3 \times 10^8 \), and \( 1 \times 10^8 \), respectively[14]. Since, on the other hand, the detectivity of thermal detectors has essentially no relation with spectral range, the relative advantage of cooled photon detectors over thermal detectors is greater for operation at shorter wavelengths.
Spectral resolution refers to the ability of a thermal imaging system to provide information about the wavelength of detected light. It is a desirable feature in military settings for improved target detection and identification[40] and for missile early warning systems to reduce false alarms using spectral information from the missile plume[41]. Civilian applications may include remote gas and atmospheric analysis, potentially allowing simultaneous detection and identification of multiple different gases. Such functionality is not intrinsic to any detector technology, but is instead achieved at either the “meta-device” level by combining semiconductor materials with different bandgaps on the focal plane array[40, 41, 36], or at the system level using dispersive optics, filters or interferometers[42]; these latter approaches tend to feature much finer spectral resolution, but are often quite bulky due the required optics and/or the larger aperture sizes needed to compensate for light lost in absorptive spectral filters, which limits the scope of applications. For this reason, I will from now on limit the discussion to devices achieving spectral resolution at the FPA level.

1.2 State-of-the-art room-temperature detectors and FPAs

Having introduced the key performance parameters of thermal imagers and detectors, I will now offer a more detailed discussion of the state of the art of detectors and imagers capable of room-temperature operation. Due to the great disparity in their performance parameters (especially speed), I will address thermal detectors and photon detectors separately.

1.2.1 Bolometers and other thermal FPAs

Today’s uncooled thermal FPA market is dominated by bolometers based on VO$_x$ thin films, a material that undergoes a crystalline phase transition between 50° and 70° C[15], and the lowest-NEDT commercially available (and, to my knowledge, experimentally demonstrated) uncooled thermal imagers available have used VO$_x$ tech-
nology. VO$_x$ benefits from a high thermal coefficient of resistance associated with its phase transition as well as low cost of manufacture due to its compatibility with thin-film deposition techniques (see [43] and references associated with Fig. 1-1 below). As with all thermoresistive bolometer technologies, however, it suffers from 1/$f$ noise, which happens to be the dominant noise source[37]. To visualize the fundamental limitations of VO$_x$ bolometers and compare with a few other thermal FPA technologies and demonstration s, I expand upon the noise performance model developed by Zerov et al[32], in which expressions for the 1/$f$, Johnson, and thermal fluctuation noise spectral densities are presented. (Thermal fluctuation noise refers to noise due to random fluctuations of the bolometer pixel temperature associated with radiative and conductive heat transfer pathways between the pixel and its environment[44].) Algebraic manipulation of the expressions in Zerov et al. allows the limiting detectivity associated with each noise source to be written entirely in terms of parameters independent of pixel area as follows in the limit of 100% pixel on-time as follows:

\[
D^*_{0,1/f}(f) = \frac{\eta \alpha \tau}{c_A} \sqrt{\frac{h_{VO_x}}{K}} f^{1/2},
\]

\[
D^*_{0,Johnson}(f) = \eta \alpha \left( \frac{\tau}{4k_Bc_A} \left( 1 - \frac{T_0}{T_{VO_x}} \right) \right)^{1/2},
\]

\[
D^*_{0,therm}(f) = \frac{\eta}{T_{VO_x}} \left( \frac{\tau}{4k_Bc_A} \right)^{1/2},
\]

with variables defined as follows:

- \(\eta\): Thermal infrared absorption efficiency of pixel
- \(\alpha\): Thermal coefficient of resistance of VO$_x$ material at operating temperature
- \(\tau\): Thermal time constant of pixel, \(\tau = C/G\)
- \(C\): Total heat capacity of pixel
- \(G\): Total thermal conductance between pixel and environment
- \(c_A\): Heat capacity per area of pixel, \(c_A = C/A_d\)
- \(h_{VO_x}\): Thickness of VO$_x$ layer within pixel
- \(K\): Volumetric 1/$f$ noise parameter of VO$_x$ material
- \(k_B\): Boltzmann’s constant
- \(T_0\): Temperature of pixel’s ambient environment
$T_{\text{VO}_x}$:  Pixel operating temperature

These expressions are frequency-dependent. I then define a frequency-averaged detectivity as follows:

$$
\bar{D}_0^* = \sqrt{f_2 - f_1} \left( \int_{f_1}^{f_2} \sum_{\text{noise sources}} [D_0^*(f)]^{-2} df \right)^{-1/2},
$$

(1.4)

where $f_1 = 0.147t^{-1}_\text{meas}$, $t_{\text{meas}} = 4.27\text{s}$, and $f_2 = \tau^{-1}/2\pi$. This definition is chosen to yield a detectivity reflecting the correct signal-to-noise (SNR) ratio for a signal with uniformly distributed frequency spectrum. The lower frequency limit is chosen to accurately reproduce the total $1/f$ noise power obtained from a 128-frame acquisition at a frame rate of 30 Hz, consistent with commercial NEDT measurement protocols[34], and the upper frequency limit is chosen to reflect the useful frequency range of the device limited by its thermal time constant. Fig. 1-1 plots the various noise source-limited $\bar{D}_0^*$ and that of the combined noise sources as a function of $\tau$ using VO$_x$ material parameters ($\alpha$, $K$, $T_{\text{VO}_x}$) and pixel layer structure design parameters ($c_A$, $h_{\text{VO}_x}$) from Niklaus et al.[1]; I assume $T_0 = 290\text{K}$ and $\eta = 1$. I also include data points representing approximate values of $(\tau, \bar{D}_0^*)$ extracted from various publications and product datasheets for exemplary devices; most of these report NEDT at an f-stop of $f/1$ without giving detectivity values, so I approximate $\bar{D}_0^*$ using the relation from Rogalski[15],

$$
\text{NEDT} = \frac{4\sqrt{A}\Delta f}{N^2A_1^{1/2}D_0^*[\Delta P/\Delta T]_{\lambda_1-\lambda_2}},
$$

(1.5)

where $N$ is the f-number, $[\Delta P/\Delta T]_{\lambda_1-\lambda_2}$ is the temperature contrast of radiation within the 8 µm–14 µm range at $T_0$ and the noise bandwidth $\Delta f$ is taken to be half the frame rate for the given publication; that is, the Nyquist sampling rate. Comparison between the theoretically predicted detectivity limit (the salmon-colored curve in Fig. 1-1) and the datapoints for reported devices show that existing VO$_x$ devices are already quite close to the performance limit set by $1/f$ noise, controlled by the noise parameter $K$. The difference is likely made up by readout (ROIC) noise, which can
Figure 1-1: Spectrally-averaged detectivity $\overline{D}_0^*$ versus time constant $\tau$ for several individual noise sources and for all noise sources combined in VOx using parameters from Niklaus et al[1]. I also plot with dashed lines the limiting detectivities associated with only the radiative thermal fluctuation of bolometer pixels at ambient and VO$_x$ operating temperatures. The vertical black line represents the longest possible $\tau$ for the assumed value of $c_A$ in the case of no conductive thermal exchange between the pixel and its environment (I.E., only radiative thermal exchange). Points ($\tau, \overline{D}_0^*$) extracted from a set of exemplary published room-temperature thermal detectors, VO$_x$-based and otherwise, illustrate the validity of our claimed performance limit[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. I do not address readout (ROIC) noise here, which in practice cannot be neglected and likely accounts for the difference between our noise model and the best reported detectivities in practice[1].
be quite significant for these systems\cite{1}, non-unity $\eta$, and uncertainty in the model input parameters. With no clear path to reducing $K$\cite{37} and NEDT values ceasing to decrease over the last two decades (see citations in Fig. 1-1), I find it not unreasonable to suggest that the sensitivity of VO$_x$ detectors has reached its upper limit. Instead, manufacturers appear to have shifted focus towards reducing the pixel pitch without sacrificing performance.

### 1.2.2 HOT photon and other high-speed detectors

Photon detectors exhibiting good noise performance at non-cryogenic temperatures—High-operating-temperature (HOT) photon detectors, a subset of which includes room temperature photon detectors—and other high-speed room-temperature mid-IR detector technologies are an exciting and active area of research, motivated by applications such as free-space optical communication\cite{28, 29} and dual-comb gas spectroscopy (DCS)\cite{51} which require higher speeds than thermal detectors can offer. In Table 1.1, I summarize the key performance metrics of some exemplary devices in this category that exhibit an LWIR photoresponse. HgCdTe\cite{45} and InAsSb\cite{46}

<table>
<thead>
<tr>
<th>Technology</th>
<th>$D^*$ [Jones] @ $\lambda$ [$\mu$m]</th>
<th>Speed</th>
<th>$\lambda$ Range [$\mu$m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>HgCdTe photovoltaic, optically immersed\cite{45}</td>
<td>$1.0 \times 10^8 @ 10.6 \mu$m</td>
<td>$\tau \leq 1.5$ ns</td>
<td>2.2–10.6</td>
</tr>
<tr>
<td></td>
<td>$1.6 \times 10^8 @ 9 \mu$m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HgCdTe photoconductor, optically immersed\cite{45}</td>
<td>$8 \times 10^7 @ 10.6 \mu$m</td>
<td>$\tau \leq 3$ ns</td>
<td>1–11.3</td>
</tr>
<tr>
<td></td>
<td>$1.0 \times 10^8 @ 9 \mu$m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>InAsSb photovoltaic, not immersed\cite{46}</td>
<td>$2.5 \times 10^7 @ 10.6 \mu$m</td>
<td>$\tau \leq 3$ ns</td>
<td>2–11.2</td>
</tr>
<tr>
<td></td>
<td>$4.5 \times 10^7 @ 9 \mu$m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antenna-coupled quantum well\cite{47}</td>
<td>$2.89 \times 10^7 @ 8.15 \mu$m</td>
<td>Usable up to 4 GHz</td>
<td>7.7–9.5</td>
</tr>
<tr>
<td>Graphene-plasmonic photothermoelectric\cite{48}</td>
<td>$1.1 \times 10^9 @ 8.15 \mu$m</td>
<td>$f_{3dB} = 8$ MHz</td>
<td>7–14</td>
</tr>
<tr>
<td>Graphene-plasmonic photoconductor\cite{49}</td>
<td>$3.8 \times 10^5 @ 12.2 \mu$m</td>
<td>$\tau_{rise} = 0.6$ ns \textit{(simulated)}</td>
<td>(unreported)</td>
</tr>
<tr>
<td>Graphene rectenna\cite{50}</td>
<td>$2.6 \times 10^6 @ 10.6 \mu$m</td>
<td>(unreported)</td>
<td>(unreported)</td>
</tr>
</tbody>
</table>

Table 1.1: Comparison of LWIR photon detectors and other technologies for fast photodetection in the LWIR range.
photovoltaic and photoconductive detectors achieving room-temperature detectivities around $10^8$ Jones in the important 9–11 µm range are commercially available, and detectivities as high as $1.1 \times 10^9$ Jones and speeds up to 4 GHz in this temperature and wavelength range have been reported for detectors based on graphene plasmons[48] and antenna-coupled quantum well patches[47], respectively. The detectivities of the semiconductor photon detectors may be increased further by cooling the sensors to temperatures readily achievable with thermoelectric coolers (TECs)[45, 46, 47]; for instance, the detectivity of a TEC/detector co-packaged version of the above HgCdTe photovoltaic detector is enhanced by a factor of 20 over the uncooled version for a modest power dissipation of 0.28 W. Of course, there are still circumstances in which it makes sense to insist on room temperature operation. It is undesirable, for instance, to have to frequently replace batteries or stay tethered to a power source for portable or miniature gas sensors and analyzers that are deployed in the field or around an industrial site.

1.3 Motivations for improving uncooled detector and imager technology

Despite the incredible range of applications enabled by modern thermal imager and detector technology, there is no lack of motivation (or government funding) to improve upon the state of the art of uncooled devices. Here, I motivate efforts to improve upon the three performance metrics discussed in Section 1.1: Sensitivity, speed, and spectral resolution. Before moving on to specific applications, though, I offer some comments about sensitivity. In essence, increased sensitivity means being able to extract the same amount of information from fewer photons. The generality of this intuitive concept reflects the wide range of applications for which high sensitivity (or high detectivity, or low NEP, as makes sense for the application at hand) is crucial, and indeed it underpins all other performance metrics. It would not make sense to say, for instance, that speed is more important than sensitivity for free
space optical communications applications. When one asks for a faster receiver in a
data link to increase the symbol rate, less light reaches the receiver within a given
symbol period, and the SNR to infer a data symbol is reduced by the square root
of the symbol rate increase. If the original receiver was already operating within a
tight SNR budget, the faster receiver would also have to have an improved noise-
equivalent power alongside higher speed to make use of the increase in bandwidth
without increasing transmitter power. Therefore the sensitivity of the receiver, in its
own way, limits the data rate just as the bandwidth does—both are important. A
similar story may be told for spectral resolution where, for a given received spectral
power (or spectral radiance or spectral irradiance, as the case may be), finer spectral
resolution means less photon flux per spectral channel, and thus greater sensitivity
is needed to maintain the same SNR per channel. With this in mind, I first discuss
applications that purely require increased sensitivity, followed by applications for
which speed and then spectral resolution are also important.

1.3.1 Long-range imaging

Cooled detectors are currently the standard choice for long-range imaging in defense
and security applications[52, 53]. High-angular resolution imaging requires focusing
lenses with long focal lengths. In light of Eqn. 1.5, a low f-number must be used
to achieve an acceptable NEDT in the regime of detectivity-limited NEDT (recall
from Sect. 1.1.1 the NEDT of cooled systems is often limited by ROIC charge han-
dling capacity, so this applies more for uncooled imagers). Therefore, for uncooled
imagers, the finer the angular resolution requirement, the larger the lens must be to
achieve acceptable NEDT. Beyond a certain resolution, it simply makes more sense
to pay for a cooled imager rather than an expensive and bulky infrared optic. Of
course, cooled imagers themselves suffer from a several-$10,000 barrier to entry in
addition to the power requirement of keeping an integrated Dewar assembly (IDA) at
cryogenic temperatures[15]. Higher-detectivity uncooled imagers would thus enable a
new generation of inexpensive, portable long range-imagers, extending the range for
detection, recognition and identification (DRI) in defense and security settings[54].
1.3.2 Optical gas imaging

Optical gas imaging is used for the detection of gas plumes and leaks in a variety of industrial and infrastructure settings, such as petrochemical and chemical plants, power plants, and electrical utility stations[55], as gases tend to feature characteristic absorption peaks in the mid-IR wavelength range. For typical applications where the goal is to detect leaks of a known gas species or family of species with similar absorption spectra (for instance, hydrocarbons[23]), it is not necessary to have wavelength resolution at the FPA level. Instead, it suffices to use a band-pass filter to eliminate light from outside the spectral region of interest, reducing background clutter and, in the limit of background-noise-limited imaging, photon shot noise. Passive remote gas sensing relies on a temperature difference between a gas plume and the background to achieve contrast: A relatively hot gas plume will emit more thermal radiation than it absorbs, and a relatively cool gas plume will absorb more than it emits. The sensitivity of a gas imaging camera may be quantified as the minimum line integral of gas concentration over a ray terminating at the camera, with units of concentration-length product. In ideal conditions where the gas plume is not imaged in front of a noisy background (for instance, vegetation), and with all else the same, the minimum detectable gas concentration-length is proportional to the NEDT of the camera. More sensitive thermal imagers would thus be able to detect smaller and lower-concentration gas plumes, especially in the LWIR where the absorption spectra of the strongest greenhouse gases lie[56, 57, 55]. Implementing spectral resolution at the FPA level would provide additional functionality by allowing discrimination and analysis of different gas species without having to physically exchange optical filters, which may be useful for environmental monitoring applications[41].

1.3.3 Robust imaging in defense with spectral resolution

As previously mentioned in Sect. 1.1.3, the spectral information provided by imagers with multiple spectral bands is useful in military settings, allowing improved contrast, DRI at reduced ranges and reduced false positive rates for missile launch
detection. These applications have prompted manufacturers to produce dual-color HgCdTe focal plane arrays able to discriminate between incident radiation falling within one of two spectral bands, which allows absolute (uncalibrated) temperature measurements[40, 41, 36]. Multispectral functionality is also helpful in degraded visual environments such as fog and dust, as different wavelengths are scattered and attenuated differently by different aerosols; simultaneous imaging in multiple spectral bands can thus provide a degree a resilience against poor visual conditions, and algorithms have been developed to produce crisp images in such conditions by combining data from multiple spectral bands. For instance, LWIR wavelengths have been reported by various authors to experience reduced attenuation through fog and clouds[58, 59], whereas shorter wavelengths offer fundamentally better angular resolution, and these can be combined to produce the best possible image[60]. While multispectral imaging systems mounted on large aircraft and vehicles are likely to opt for the improved performance provided by cooled imagers at a higher cost, inexpensive, low-power and lightweight uncooled imagers featuring spectral resolution offer the possibility of bringing these benefits to miniature unmanned aerial vehicles (UAVs) and handheld or helmet-mounted visibility enhancement systems.

1.3.4 Robust free-space optical communication

Free-space optical communication (FSOC) has been proposed as an alternative to radio frequencies for data links between airborne, terrestrial, nautical, and space-based platforms[61]. FSOC is especially appealing in the military domain as a secure, high-bandwidth data link solution between aircraft, persistent threat detection systems and other network nodes, as collimated beams of light are more difficult to intercept and eavesdrop on than RF radiation[62]. The aforementioned weather resilience of LWIR light, besides imaging applications, also motivates its application in FSOC with robustness against poor weather conditions. Studies comparing MWIR and LWIR optical links with SWIR links have found the former to exhibit reduced power fluctuations in fog[28] and radically reduced attenuation in fog, haze and rain[29]. Exploiting the full potential of robust FSOC in settings with constrained power, mass
and/or cost budgets, such as miniature UAVs, will require fast and sensitive mid-IR detectors operating at or near room temperature.

1.3.5 Robust simultaneous localization and mapping

Simultaneous localization and mapping (SLAM) refers to the process by which an autonomous vehicle infers its surroundings and position using data collected from various sensors and imagers. It has been shown that visual orientation cues obtain from cameras on-board miniature UAVs can supplement data obtained from inertial measurement units (IMUs) to perform SLAM with less noise than from IMUs alone[63], and motion estimation at night has been demonstrated by combining sparse depth data from LiDAR with thermal camera data[64]. SLAM under dark or degraded visual conditions is thus a promising application for thermal imagers. For miniature UAVs, fast imaging would be important for accurate and responsive SLAM systems, as the cutoff frequency of the camera must be faster than the maximum angular velocity divided by angular resolution of the camera to prevent motion blur; this value can easily enter the kilohertz range for small, agile UAVs while maneuvering[63]. I therefore anticipate the fast, lightweight and low-power thermal imagers will enable agile operation of miniature UAVs in dark or poor visual conditions.

1.4 Motivations for waveguide-integrated mid-IR photodetection

Until now, all of the photodetectors that I have discussed have been free space-coupled devices. Recently, there has been interest in optical waveguide-integrated mid-IR detectors as a feature of chip-integrated mid-IR microsystems. Chip-integration of mid-IR-based systems offers the promise of inexpensive and scalable manufacturing due to reduced packaging costs compared to free-space systems, and integrating the photodetectors themselves on chip instead of coupling off-chip into a packaged detector allows the cost benefit to be best realized. Additionally, waveguide-integration
of photodetectors allows in principle for the sensitivity of such systems to be maxi-
mized, as the NEP of most detector technologies is proportional to the square root
of the sensitive area as discussed in Sect. 1.1.1, and it is easier to align and fab-
ricate a micron-scale detector onto a single-mode waveguide than it is to align and
coopackage a discrete detector of the same size. Below, I provide some examples of
chip-integrated mid-IR microsystems that would benefit from waveguide-integrated
detectors.

1.4.1 Point gas sensors

Point gas sensing has applications in greenhouse gas emission monitoring, industrial
toxic gas monitoring, pipeline gas leak detection, ecological monitoring, and domestic
air quality monitoring. In the last few decades, multiple authors have proposed and
demonstrated chip-integrated optical gas sensing systems based on gases’ absorption
of a guided mid-IR probe beam. Here, the gas and probe light interact in the evanes-
cent fields of an integrated optical waveguide [65, 66, 67, 68, 69] or ring resonator[70].
Su et al. took the concept further by integrating the photodetector on the same chip
as the waveguide to sense methane [71]. Such systems have the potential to be smaller,
lighter and less expensive than state of the art gas sensors that require discrete optics
and packaging and assembly costs. However, development of this technology will re-
quire surmounting several difficulties. First, blackbody sources are not bright enough
for integrated gas sensors based on single-mode waveguides, as thermodynamics places
a low limit on the spectral power that can be thermally radiated into a single optical
mode[30], offsetting any benefit from reducing the size of the detector. It is thus nec-
ecessary to use high-brightness infrared light sources such as quantum cascade lasers
(QCLs), quantum cascade superluminescent light emitters[72, 73] or QCL frequency
combs[74, 75] to achieve acceptable gas sensitivity performance. Chip-integration of
such light sources is a subject of ongoing research[76, 77]. Second, it is necessary to
engineer low-loss optical waveguides to allow long gas-light interaction lengths and
compete with the sensitivity of technologies based on multi-pass gas cells[78, 65].
Finally, of course, it is necessary to develop highly sensitive waveguide-integrated
mid-IR detectors operating at or near room temperature to realize the cost, size, and power consumption benefits of chip-integrated optical gas sensors.

### 1.4.2 Liquid-phase chemical sensors

Related to waveguide-integrated point gas sensing is the concept of liquid-phase chemical sensing using fluid-immersed or microfluidic integrated microsystems, which may enable portable and inexpensive tools for medical diagnostics, water quality measurement, food safety assessment, and forensics. Several demonstrations of integrated fluid-phase chemical sensing using off-chip sources and detectors have been published using various waveguide platforms and chemical analytes[79, 80, 81, 82]. The main caveat of these technologies is that liquid water strongly absorbs all mid-infrared wavelengths[83], therefore it is necessary to use a non-aqueous sensing medium. One solution to this problem is to use microfluidic liquid-liquid extraction techniques allowing in-situ exchange of analytes from an aqueous medium into a water-immiscible and IR-transparent solvent[84].

### 1.4.3 Robust coherent free-space optical communications

In Sect. 1.3.4, I discussed the benefits of mid-IR light for FSOC with resilience against poor weather conditions. Incorporating coherent detection into such a communication link offers two distinct SNR benefits. First, coherent detection provides optical gain, allowing the NEP of a receiver to be decreased below that of the detector itself[85]. Second, coherent detection allows coherent combination of free-space optical beams with poor spatial phase coherence. In field-deployed FSOC systems, atmospheric turbulence causes time-varying decoherence of the optical beam’s phase front, preventing the beam from being focused down to a diffraction-limited spot with a static optic, and thus reducing the achievable SNR. If the overall receiver size is larger than the scale of variations of the phase front (the lateral coherence length), typically on the order of centimeters within Earth’s atmosphere, ideal performance can be recovered by splitting the single focusing lens into several sub-lenses of size less than the lateral
coherence length and coherently recombining the captured light to be detected with the smallest possible detector area and thus minimum NEP\cite{61}. Implementing such a system in a compact fashion would inevitably require integrated subsystems featuring local oscillators, phase modulators and photodetectors for tracking the light phase and performing the eventual optical to electrical signal conversion. Therefore, fast and sensitive waveguide-integrated mid-IR detectors are also critical for implementing cutting-edge coherent FSOC systems.

1.5 State of integrated mid-IR detector research

A great variety of mid-IR waveguide platforms has been demonstrated to date involving various arrangements of mid-IR transparent materials: Silicon-on-insulator (SOI), silicon-on-sapphire, germanium-on-silicon, suspended germanium and silicon, waveguides based on silicon nitride, chalcogenide glasses, and more, all with different advantages and disadvantages\cite{86}. A subset of these technologies, such as suspended Ge and chalcogenide glass waveguides, have wide transparency windows extending well into the LWIR. To date, waveguide-integrated photodetection has been demonstrated at wavelengths up to about 4μm; these longest-wavelength demonstrations in the 3.6–4μm range have all been achieved using 2D materials black phosphorus and graphene on SOI platforms\cite{87, 88, 89}. Indeed, 2D materials are a convenient choice for waveguide-integrated mid-IR photodetectors for two reasons. First, due to their extreme thinness, introduction of a 2D material onto an optical waveguide typically only introduces a perturbative loss to the optical mode without changing the mode profile or effective index significantly, which makes efficient coupling from a bare waveguide into a photodetector trivial\cite{90}. Second, 2D material transfer techniques can be applied to essentially any substrate. This latter advantage of 2D materials will only become more pronounced at longer wavelengths where fewer waveguide platforms are available; SiO\textsubscript{2} cladding becomes too lossy beyond ≈ 4μm, followed by sapphire beyond ≈ 5.5μm, and silicon waveguides become lossy beyond ≈ 8μm\cite{86}. At the same time, however, wavelengths beyond 4μm fall below the bandgap of even the most
narrow-gapped 2D semiconductors such as black phosphorus and tellurene\cite{91} that have enabled some of the longest-wavelength waveguide-integrated photodetectors so far. This leaves only the semimetallic graphene to carry the flag of 2D material advantage for waveguide-integrated photodetection beyond the MWIR, where increasingly unconventional waveguide platforms make integration with standard LWIR-sensitive materials and systems difficult. Encouraged by recent demonstrations showing exceptional sensitivity and speed in the LWIR wavelength range\cite{48, 92, 93}, I see this not as a disadvantage but as an impetus to explore graphene photodetectors for the waveguide-integrated mid-IR applications described above. I thus move on to introduce the basics of graphene and graphene-based photodetectors in the next chapter.
Chapter 2

Fundamentals and Applications of Graphene

Graphene is a 2D material of interest for a plethora of applications, including tissue engineering, drug delivery and gene delivery[94], strain and electrochemical sensors[95], transparent electrodes for organic and thin-film optoelectronics[96], energy storage and conversion, water decontamination, deionization and heavy metal sequestration[97], terahertz modulation and photodetection[98], infrared modulation and, of course, infrared photodetection[99]. In this chapter, I summarize graphene’s basic electrical, optical and optoelectronic properties as background for my research, followed by a more targeted and application-oriented discussion of graphene’s benefits and drawbacks compared to other technologies.

2.1 Graphene introduction and photoresponse origin

2.1.1 Bandstructure and optical properties

Graphene’s unique properties that make it useful for optoelectronic applications ultimately derive from its semimetallic bandstructure, which itself derives from graphene’s so-called honeycomb crystal structure, depicted in Fig. 2-1. Tight-binding analysis of graphene’s half-filled \( \pi \)-band leads to the bandstructure depicted in Fig. 2-2, in
which the conduction and valence bands of charge-neutral graphene intersect at two $k$-points, the $K$ and $K'$ high-symmetry points representing the two distinct corners of graphene’s hexagonal Brillouin zone. Fourier transforming and Taylor expanding the graphene bandstructure Hamiltonian around small energy offsets from the charge-neutral Fermi level yields equations of motion resembling two copies of the 2D massless Dirac equation for ultra-relativistic electrons, except with the Fermi velocity $v_F \approx 1 \times 10^6 \text{m/s}$ in place of the speed of light $c$. The ability for clean graphene samples to achieve impressive carrier mobilities of above $50,000 \text{cm}^2\text{V}^{-1}\text{s}^{-1}$ at room temperature even after undergoing CVD synthesis and transfer[100] originates not only from the high $v_F$ but also from Dirac equation dynamics, in particular the ability of Dirac charge carriers to tunnel through potential barriers, providing graphene’s mobility a degree of resilience against disorder-induced potential wells[101]. Is is this high carrier mobility that gives graphene promise as an optoelectronically active material in applications requiring high bandwidth, such as detectors and modulators for communications applications.

Graphene’s conductivity at optical frequencies may be derived from the Kubo formulas and expressed as follows[102]:

\[
\begin{align*}
\text{Graphene’s conductivity at optical frequencies may be derived from the Kubo formulas and expressed as follows:}
\end{align*}
\]
Figure 2-2: Graphene’s bandstructure at charge neutrality. Its Fermi level coincides with the intersection between its conduction and valence bands, which meet at a point in a conical manner. By Ponor and licensed under CC-BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0/deed.en). No changes made.

\[
\sigma(\omega, E_F, \tau, T) = \frac{je^2(\omega - j\tau^{-1})}{\pi\hbar^2} \times \left[ \frac{1}{(\omega - j\tau^{-1})^2} \int_0^\infty \varepsilon \left( \frac{\partial f_d(\varepsilon)}{\partial \varepsilon} - \frac{\partial f_d(-\varepsilon)}{\partial \varepsilon} \right) d\varepsilon \right.
\]

\[
\left. - \int_0^\infty \frac{f_d(-\varepsilon) - f_d(\varepsilon)}{(\omega - j\tau^{-1})^2 - 4(\varepsilon/\hbar)^2} d\varepsilon \right] \tag{2.1}
\]

where \(\omega\) is the optical frequency, \(E_F\) is the Fermi level, \(\tau\) is the Drude scattering time, \(T\) is temperature, \(e\) is the elementary charge, \(f_d(\varepsilon) = (\exp((\varepsilon - E_F)/k_BT) + 1)^{-1}\) is the Fermi-Dirac distribution, \(k_B\) is Boltzmann’s constant, and fields vary as \(e^{j\omega t}\). Note that \(\sigma\) is invariant upon negation of \(E_F\). In Fig. 2-3, I plot the resulting curves as a function of \(\omega\) for various \(E_F\) and for \(T = 300\) K and for \(\tau = 100\) fs, which is a fairly typical Drude scattering time for CVD-transferred graphene[103]. Two different regimes of absorption (that is, high \(\text{Re}[\sigma]\)) can be distinguished. At high frequencies/short wavelengths where interband transitions dominate, the conductivity approaches a constant value of 60.6 \(\mu\)S, corresponding to 2.3% absorption of light.
Figure 2-3: a) Real and b) imaginary parts of graphene’s optical conductivity versus frequency at various Fermi levels for $T = 300$ K and $\tau = 100$ fs. c) Imaginary/real ratio of the conductivity. The further from zero, the more reactive the graphene is; positive values correspond to capacitive behavior, and negative values correspond to inductive behavior, so-called “kinetic inductance”.

---

**Diagram Description:**

- **Graph a:** Real Conductivity vs. Wavelength [μm] for various Fermi levels.
  - Y-axis: Real Conductivity [μS/μm]
  - X-axis: Wavelength [μm]
  - Different colors represent different Fermi levels: 0 eV, 0.1 eV, 0.2 eV, 0.3 eV, 0.4 eV, 0.5 eV, 0.6 eV.

- **Graph b:** Imaginary Conductivity vs. Wavelength [μm] for various Fermi levels.
  - Y-axis: Imag. Conductivity [μS/μm]
  - X-axis: Wavelength [μm]
  - Same Fermi levels as in graph a.

- **Graph c:** Loss ratio (Imag./Real Cond.) vs. Optical frequency [THz] for various Fermi levels.
  - Y-axis: Loss ratio (Imag./Real Cond.)
  - X-axis: Optical frequency [THz]
  - Same Fermi levels as in graphs a and b.
normally incident on suspended graphene. As the frequency falls below \(2E_F/\hbar\), the conductivity rapidly decreases due to Pauli blocking as conduction states at the upper end of the optical transition fill with electrons (or for negative \(E_F\), as valence states at the bottom end of the transition get depleted). At low frequencies, \(\text{Re}[\sigma]\) again increases as intraband (i.e., Drude) absorption becomes significant, approximately following a \(\text{Re}[\sigma] \propto E_F \lambda^2 \tau^{-1}\) relation in the mid- and far-IR\(^{102}\). Graphene devices operating in the MWIR and LWIR thus roughly straddle these two regimes, with the wavelength above which intraband effects dominate depending on \(T, E_F\) and \(\tau\).

2.1.2 Carrier cooling and photothermoelectric effect

The photothermoelectric (PTE) photoresponse mechanism in graphene, the main response mechanism for unbiased devices, is based on optically-induced thermal gradients in graphene's electronic system that get converted into a voltage by the Seebeck effect \(^{104, 105}\). This is somewhat similar to a thermopile infrared detector, whose photovoltage also derives from thermal gradients read out by a junction between two materials with disparate Seebeck coefficients\(^{44}\). The key difference between the two is that in the thermopile, it may be assumed that the electron gas and phonon bath in the detector are in thermal equilibrium. In graphene, however, this is not the case, and thus it was deemed that this photoresponse mechanism deserved its own separate name. When light is absorbed in a patch of graphene, either via interband or intraband absorption, the electrons in the targeted patch thermalize into a Fermi-Dirac distribution with an elevated temperature relative to the phonon bath. This process occurs on a sub-picosecond timescale, mediated by electron-electron interaction mechanisms such as Auger processes. If the exciting photon energy greatly exceeds the optical phonon energy of \(\approx 200\) meV, optical phonons may be emitted in the process, reducing the efficiency of optical electronic heating\(^{106}\). Following thermalization, this excess electronic heat undergoes thermal diffusion as well as direct electron-lattice cooling via acoustic phonons. This latter process is relatively slow due the two-order-of-magnitude discrepancy between the Fermi velocity and sound velocity of graphene, and ranges from 1 psi to hundreds of picoseconds\(^{107}\). This relatively long electron-
lattice cooling timescale, combined with graphene's low electronic density of states, gives rise to a considerable discrepancy between the electronic and lattice temperatures even under modest illumination levels. For graphene targeted with pulsed light sources, electronic temperatures of up to 3000 K have been observed[108].

The effects of electron-lattice cooling and thermal diffusion in graphene are captured in the thermal transport equation for graphene:

\[-\nabla \cdot (\kappa \nabla (\Delta T_{el})) + \tau_{eph}^{-1} C_{el} \Delta T_{el} = \eta \dot{Q} - j \cdot \nabla \Pi.\]  

(2.2)

Here \(\kappa\) represents the electronic planar thermal conductivity of the graphene; \(\Delta T_{el}\) is the difference between the electronic temperature \(T_{el}\) and the lattice temperature \(T_0\), \(\tau_{eph}\) represents the electron-lattice cooling time constant, and \(C_{el}\) represents the electronic heat capacity of graphene. \(\eta\) is the efficiency with which optical energy absorbed by the graphene is deposited into the electronic system on a sub-picosecond timescale, roughly unity for MWIR and longer-wavelength light[106]. \(\dot{Q}\) represents the power per area absorbed in the graphene. \(j\) is the electrical current density, and \(\Pi\) refers to the Peltier coefficient. Charge transport is described by the equation[109]:

\[j = -\sigma (\nabla V + S \nabla (\Delta T_{el})),\]  

(2.3)

where \(\sigma\) here is the DC conductivity of graphene, related to \(\kappa\) by the Weidemann-Franz law[109]; \(V\) is voltage, and \(S\) is the Seebeck coefficient. Simultaneous solution of these equations yields the responsivity of an arbitrary unbiased graphene photodetector to optical excitation described by \(\dot{Q}\). When \(\Delta T_{el} \ll T_0\), the temperature dependence of the coefficients may be ignored, yielding a linear system of equations.

Song et al. presented an analytical solution to Eqns. 2.2 and 2.3 in the simple case of a graphene channel of length \(L\) and width \(W\), where the graphene on each half of the channel is gated to a different Fermi level and light is incident along the center of the channel at the junction between the two differently-gated section of graphene,
Figure 2-4: Depiction of a split-gate PTE graphene photodetector. Metal back-gates are used to tune the Fermi level on each side of the graphene channel resulting in a $p_n$-junction and thus two different Seebeck coefficients, $S_1$ and $S_2$, on the two sides of the device. Concentrated illumination along the center of the device heats up electrons in the channel relative to the lattice, with $\Delta T_{el} = 0$ boundary conditions at the source and drain. The asymmetric Seebeck coefficient profile combined with the elevated electron temperature in the channel gives rise to a thermal electromotive force between the source and drain as described by Eqn. 2.3.

as shown in Fig. 2-4. They find an open-circuit voltage of:

$$V = \frac{\eta(S_2 - S_1)(P_{abs}/W)}{\frac{\kappa_1}{\ell_1} \coth \frac{L}{2\ell_1} + \frac{\kappa_2}{\ell_2} \coth \frac{L}{2\ell_2}}$$

(2.4)

where $P_{abs}$ is the total power absorbed over the channel width, $\ell_i = \sqrt{\kappa_i \tau_{eph}/C_{el}}$ is the cooling length—the coefficient of exponential $\Delta T_{el}$ versus position dependence in the channel due to electron-lattice cooling—and the subscript index refers to the two sides of the device. In the simplified case where $E_{F1} = -E_{F2}$ and thus $S_1 = -S_2$, $\kappa_1 = \kappa_2$ and $\ell_1 = \ell_2$, the Johnson noise-limited NEP of the device is thus given by:

$$\text{NEP}_{conc} = \sqrt{4k_B T_0 \frac{L}{\sigma W} \frac{\kappa \coth(L/2\ell)}{\eta \ell |S|(1/W)}} = \frac{\pi^2 k_B^{5/2}}{3} \frac{\sigma^{1/2} T_0^{3/2}}{e^2} \frac{2 \coth(L/2\ell)}{\eta |S|} \frac{1}{\ell} \sqrt{LW},$$

(2.5)

by applying the Weidemann-Franz relation. It can be found by solving Eqns. 2.2 and 2.3 that if the illumination is not concentrated along the centerline of the device but
rather distributed uniformly over the graphene channel, the NEP then becomes:

\[
\text{NEP}_{\text{dist}} = \frac{\pi^2 k_B^{5/2}}{3 e^2} \frac{\sigma^{1/2} T_0^{3/2}}{\eta S} \frac{L}{2 \ell^2} \left(1 + \coth^2 \frac{L}{4 \ell}\right) \sqrt{LW}.
\]  

(2.6)

NEP\text{conc} is minimized when \( L \approx 2.18 \ell \) and NEP\text{dist} is minimized when \( L \approx 1.77 \ell \). In each case, the NEP reduces to:

\[
\text{NEP}_{\text{conc, min}} = 3.71 \frac{\pi^2 k_B^{5/2}}{3 e^2} \frac{\sigma^{1/2} T_0^{3/2}}{\eta S} \sqrt{\frac{W}{\ell}}
\]

\[
\text{NEP}_{\text{dist, min}} = 7.99 \frac{\pi^2 k_B^{5/2}}{3 e^2} \frac{\sigma^{1/2} T_0^{3/2}}{\eta S} \sqrt{\frac{W}{\ell}}
\]

(2.7)

For excellent-quality CVD-grown graphene at room temperature, \( S \approx 100 \mu V/K \) and \( \sigma \approx 0.5 \text{mS} \) have been reported\[110\], where I have extracted \( S \) from a reported resistance versus gate voltage curve using the Mott formula\[109\]. Assuming \( \eta = 1 \), one arrives at:

\[
\text{NEP}_{\text{conc, min}} \approx 0.4 \sqrt{\frac{W}{\ell}} \text{pW Hz}^{-1/2}
\]

\[
\text{NEP}_{\text{dist, min}} \approx 0.8 \sqrt{\frac{W}{\ell}} \text{pW Hz}^{-1/2}
\]

(2.8)

Using for example \( \ell = 1 \mu m \) as has been found for high-quality graphene near room-temperature and assuming \( W = \ell \), one can then make comparisons to other detector technologies. For instance, a highly sensitive free-space-coupled germanium photodiode with a detectivity of \( 2.6 \times 10^{11} \text{Jones} \)[111] with an area of \( 1 \mu m \times 1 \mu m \) would have an NEP of \( 0.38 \text{fW Hz}^{-1/2} \), which bluntly illustrates graphene’s disadvantage for highly sensitive detection of near-IR radiation. For high-speed communications applications, the situation for graphene is much more favorable, as the NEP of fiber communications systems is typically limited by the transimpedance amplifier (TIA) and not the detector itself\[112\]; for a state-of-the-art silicon photonics platform-integration Ge PIN photodiode\[113\] amplified by a modern high-speed TIA designed for 53 GBAud operation\[114\], one can achieve an amplified NEP of about \( 10 \text{pW Hz}^{-1/2} \), suggesting that graphene may be a competitive technology for communications applications.
course, due to graphene’s notoriously weak absorption, actually absorbing a beam of light in a $1\,\mu\text{m} \times 1\,\mu\text{m}$ or smaller patch requires resonant or plasmonic field enhancement, not to mention limitations simply due to the wavelength of light, but recent demonstrations based on metal plasmonic structures have gotten quite close\cite{115}.

2.2 Comparison with other mid-IR detector technologies

It goes without saying that in the mid-IR, the applications, tradespaces, and device- and system-level design considerations are completely different than in the near-IR, where interest in graphene is largely predicated upon its promise to unchain high-bandwidth integrated detectors from epitaxy on monocrystalline substrates. Here I argue in favor of graphene as a means to achieve high speed and decent sensitivity as necessary for some of the applications described in Chapter 1, while pointing out other practical benefits of graphene technology that give it an additional edge above competing technologies. (A discussion of graphene as a means to enable spectral resolution will be left for Chapter 3.)

2.2.1 Sensitivity and speed for communications and DCS

Bandwidth is a driving factor in communications applications. Plans for 5G wireless telecommunications networks offering data rates of up to 20 GBit/s\cite{116} highlight the thirst for bandwidth in today’s highly interconnected world. Virtual and augmented reality media are some of the major driving applications for 5G network deployment\cite{117}, and they are also likely to find use in the military domain (for instance, for remote control of unmanned vehicles) where FSOC is most appealing. I therefore anticipate the importance of bandwidth to extend to FSOC applications.

To date, the fastest non-graphene room-temperature LWIR detection has been demonstrated with quantum well infrared photodetectors\cite{39}, shown to operate up to 4 GHz and with a detectivity of about $3 \times 10^7$ Jones at room temperature for
Their detectivity increases to \(8 \times 10^7\) Jones at temperatures within the reach of thermoelectric coolers. In contrast, waveguide-integrated graphene photodetectors operating at telecom wavelengths have been shown to enable 100 Gbit/s communication links [115] and, separately, bandwidths of up to 67 GHz under zero bias[118]. Considering that the latter device uses a PTE and thus thermal response mechanism, there is no reason to doubt that similar bandwidths would be achievable at longer wavelengths. On the other hand, demonstrated sensitivity in the mid-IR has varied substantially. Castilla et al. demonstrated an unbiased PTE-based antenna-coupled encapsulated graphene photodetector around 3\(\mu\)m in size with an NEP of 82 pW Hz\(^{-1/2}\) operating between \(\lambda = 6–7\) \(\mu\)m[119]. Relying on a different response mechanism, the bulk photovoltaic effect, Wei et al. demonstrated an unbiased metal plasmon-enhanced graphene photodetector about 6\(\mu\)m in size with an NEP of 0.12 pW Hz\(^{-1/2}\) operating at \(\lambda = 4\) \(\mu\)m. Roughly, these devices suggest detectivities in the mid-10\(^6\) Jones range. On the other hand, a few recent demonstrations of biased graphene photodetectors utilizing plasmonic phenomena have reported detectivities up to around \(1 \times 10^9\) Jones[48, 120] in the LWIR. Although the physical origin of these incredible detectivities is not well understood, their observation strongly motivates further investigation of graphene-based high-speed high-sensitivity detectors in the mid-IR as well as platforms to incorporate these detectors into incoherent or coherent FSOC systems.

The performance considerations for detectors for dual comb spectroscopy systems differ from those of FSOC. The bandwidth required for DCS is based on the fastest beat note between any two comb teeth, which can be no greater than half the repetition rate. Thus, bandwidths beyond the 1–10 GHz are not likely to ever be necessary, limited by the highest repetition rates of chip-integrated lasers. Many actual demonstrations of DCS have involved maximum beat frequencies in the 100’s of MHz, which can be comfortably detected with cooled HgCdTe detectors. In principle, DCS can be performed with much lower maximum beat frequencies, but this is limited by the relatively stability of the two comb sources, and in practice it is most desirable for the frequency difference between adjacent beat notes to greatly
exceed the linewidth of any given beat note over the time scale of the acquisition. Compact and inexpensive DCS implementations, considered here as a motivation for fast and sensitive room-temperature mid-IR detectors, will most likely require long integration times to compensate for increased sensor NEP and decreased gas- or chemical-light interaction lengths. I therefore anticipate it will be most advantageous at the system level to have fast photodetection to reduce the stability burden for the comb sources, especially given the limited space for laser stabilization infrastructure in compact implementations[51]. A heuristic goal, then, may be to maximize detectivity while meeting the $\approx$100’s of MHz detector bandwidth used in many previous DCS demonstrations—well out of reach of bolometer bandwidths. Room-temperature HgCdTe may just barely meet this bandwidth requirement as shown in Table 1.1, but the potential for up to order of magnitude increases in detectivity again motivates investigation of graphene detectors for compact DCS applications.

### 2.2.2 Cost and adaptability

One of the great promises of graphene technology has always been its adaptability and cost: A high-quality, nearly monocrystalline material that can be conveniently transferred onto any substrate, eliminating the need for costly crystalline substrates and slow epitaxial material growth for optoelectronic applications. While claims of graphene’s superior ease of fabrication may have elicited skepticism in the past, products using CVD graphene have already been commercialized. For instance, Emberion Oy produces a visible–shortwave IR (SWIR) camera based on a colloidal quantum dot (CQD)-photogated graphene channel technology[121, 122]. Their technology offers sensitivities of about half that of InGaAs, but they claim a substantially reduced price point. Another company, Cardea Bio, has commercialized biological chemical sensors based on functionalized graphene field-effect transistors fabricated in a 150 mm Si wafer foundry [123]. They report a cost for graphene of $0.019 \text{ cm}^2$, far less than the costs of processing or silicon substrates [124]. Multiple teams have reported scalable CVD growth of graphene via roll-to-roll processing[125], and very recently LG Electronics has unveiled a system for roll-to-roll graphene growth on 400 mm-wide copper
films at a feed rate of 17 mm/s and a capacity of $2.8 \times 10^4 \text{m}^2/\text{year}$ for commercial production[126]. These are just a few examples illustrating the momentum behind the commercialization of graphene and other 2D material technology, and with other commercial applications under development it stands to reason that the supporting infrastructure for 2D material technologies will only grow in the coming years.

For comparison, the manufacturing costs of infrared optoelectronics remain high after decades of development. For instance, molecular beam epitaxy systems (one of the main technologies for epitaxial growth of infrared materials) come at a cost of over $1M for a system capable of simultaneously processing eight 6" wafers[127]. The growth of HgCdTe on Si for the focal plane array has been reported to take roughly half a day, so assuming 24/7 operation one would expect a capacity of about $10^2 \text{m}^2/\text{year}$[128]. The huge disparity between the production capacity of bulk semiconductor and graphene technologies highlights the potential for graphene devices to achieve disruptively low costs of production.

“Adaptability” in the context of graphene refers to its ability to be integrated into any device architecture, with no need for an epitaxial substrate to achieve perfect crystallinity. Indeed, clean and crackless graphene transfer has even been demonstrated on bumpy substrates such as pennies[129], and graphene anodes have been shown to enable flexible OLEDs with high luminous efficacy on PET substrates[130]. Although pennies and flexible substrates are unlikely to be used for infrared detection and sensing applications, graphene’s adaptability is nevertheless valuable in this domain due to the varied and unusual nature of infrared substrates and materials. Many infrared-transparent materials, such as chalcogenide glasses, are not amenable to high-temperature processing[131], which makes graphene a suitable material for heterointegration as its fabrication does not require high temperatures. As another example, graphene could feasibly be integrated onto mid-IR waveguide platforms based on suspended waveguides[86] by first transferring a thicker layer of hBN for mechanical support over a gap or perforations, followed by graphene. Although these are only two examples, they illustrate the adaptability advantage of graphene compared to materials requiring epitaxy or high temperature growth or annealing.
processes.

The strongest potential competitor to graphene in the integrated mid-IR microsystem space would likely be transfer-printed photodetector coupons composed of III-V materials or HgCdTe. III-V transfer printing was originally pursued for laser integration on silicon photonics platforms[132]. Such an approach could in principle solve both the cost and adaptability problems of III-V materials, as an integrated gas sensor or similar system would only require one very small photodetector coupon that could be conveniently placed on any waveguide platform. Recently, transfer printing of antimonide-based type-II superlattice (T2SL) mid-IR detectors has been demonstrated with no degradation of performance compared to before transfer, but operating at liquid nitrogen temperatures[133]. Similarly, epitaxial liftoff of thin-film monocrystalline CdTe solar cells has been demonstrated by wet etching a MgTe sacrificial layer[134]; this could feasibly be adapted to implement HgCdTe stamps for heterointegration. Although I am not aware of heterointegration of such devices on waveguides, I believe that it, too, merits investigation.

In summary, graphene photodetector technology offers an array of benefits in terms of speed, sensitivity, cost and adaptability that make it a worthwhile material to investigate for a variety of next-generation mid-infrared applications. In the following two chapters, I describe my work towards two different graphene-enabled advancements in mid-IR detection technology: Spectrally-resolved mid-IR imaging using slot antenna-coupled graphene patches, and graphene-enabled waveguide-integrated photodetection at $\lambda = 5.2\,\mu m$. 
Chapter 3

Multispectral Imaging Metasurfaces based on Graphene-Loaded Slot Antennas

Spectral imagers, the classic example being the color camera, are ubiquitous in everyday life. However, most such imagers rely on filter arrays that absorb light outside each spectral channel, yielding $\sim 1/N$ efficiency for an $N$-channel imager. This is especially undesirable in thermal infrared (IR) wavelengths, where sensor detectivities are low. I propose an efficient and compact thermal infrared spectral imager comprising a metasurface composed of sub-wavelength-spaced, differently-tuned slot antennas coupled to photosensitive elements. Here, I demonstrate this idea using graphene, which features a photoresponse up to thermal IR wavelengths. The combined antenna resonances yield broadband absorption in the graphene exceeding the $1/N$ efficiency limit. I establish a circuit model for the antennas’ optical properties and demonstrate consistency with full-wave simulations. I also theoretically demonstrate $\sim 58\%$ free space-to-graphene photodetector coupling efficiency, averaged over the $1050 \text{ cm}^{-1}$ to $1700 \text{ cm}^{-1}$ wavenumber range, for a four-spectral-channel gold metasurface with a $0.883 \mu\text{m}$ by $6.0 \mu\text{m}$ antenna pitch. This research paves the way towards compact CMOS-integrable thermal IR spectral imagers.
3.1 Introduction

We take spectral imaging for granted in daily life. Our eyes are spectral imagers, providing information about the composition of what we see. The infrared electromagnetic band covers many chemical absorption resonances and thus also reveals compositional information. In particular, infrared spectral imaging is applied in areas such as gas emission monitoring[19, 18, 135, 136, 17, 137, 138], ecological monitoring[139, 140, 141], food quality control[142, 143, 144], waste sorting[145], biological research[146] and oceanography[147].

Spectral imaging aims to measure a “data cube” representing light intensity over two spatial dimensions $x$ and $y$ and one spectral dimension $\lambda$ with $N$ channels. Scanning spectral imagers sequentially measure different portions of the data cube over multiple exposures to form the full data cube. A common example is the pushbroom scanner which measures $x \times \lambda$ data cube slices while scanning $y$ and is thus typically associated with satellites and conveyor belts where either the camera or subject is gradually moving in one direction[148, 19, 143, 145]. The spectral axis may also be scanned such as in tunable filter-based imagers [135, 149] or Fourier transform interferometer-based imagers which are bulky and have moving parts[18]. The former class of imagers is, roughly speaking, limited to a $1/N$ light utilization efficiency in the sense that at any given moment only light falling within one of the $N$ spectral channels reaches the focal plane, while the remaining light is discarded. In contrast to scanning imagers, snapshot spectral imagers (SSIs) capture a data cube with a single exposure[42]. This may be achieved using a color filter array similar to that of a color camera, thus again limiting light utilization to $1/N[150, 151, 42]$. Another category of SSIs uses dichroic or dispersive optics to break up incoming light by wavelength before arriving on a focal plane array (FPA). There are many variations of this approach[152, 153, 154, 155, 146], but they all require increasing the etendue of the incoming light beam by a factor of $N$, leading to an unfavorable trade-off between total FPA area, input acceptance angle and spectral resolution[42].

Compared to these technologies, the category of imagers based on inherently mul-
Figure 3-1: a) Illustration of a broadband absorbing slot antenna metasurface consisting of six differently tuned slot antennas tiled with sub-wavelength periodicity. The graphene patches are color-coded by antenna length, and the diagram is drawn to proportion based on the device dimensions used to produce Fig. 3-6a. b) Depiction of a single graphene-coupled slot antenna-based photodetector based on the photothermoelectric effect. c) Depiction of the gating scheme of the proposed photodetector. Asymmetric placement of the channel with respect to the slot results in a p-n junction in the center of the channel.

tispectral pixels is less explored. One such example is the Foveon RGB sensor, which extracts three different electrical signals from different depths in the optically active silicon region, as shorter wavelengths are absorbed closer to the sensor surface[156]. Another approach uses nanoantennas, optical resonators of sub-wavelength dimensions which nevertheless feature absorption cross sections of order $\lambda^2$ if the antenna is conjugate impedance matched with its load; or, equivalently, if the antenna is critically coupled to the vacuum. This concept has been illustrated using various antenna geometries, such as plasmonic nanoribbons[157, 158] and silicon nanorods[159]. Here, I propose a thermal IR multispectral imager where $N$ differently sized metallic slot antennas with infrared-sensitive loads targeting $N$ spectral channels are tiled to form a metasurface featuring efficient free space-to-load optical energy transfer. Fig. 3-1a
shows such a metasurface for $N = 6$. I model graphene as the photosensitive load because its broadband absorption in the mid-IR\cite{160} and processing flexibility\cite{161, 162} make it suitable for this platform. Not only do the antennas sort incident light by spectral channel, but they also enhance the absorption of the graphene load, bridging the gap between the impedance of free space and graphene, which, when undoped, has an optical sheet resistance no lower than 16.1 kΩ\cite{160}. Fig. 3-1b shows in detail an example of a single such antenna-coupled graphene photodetector. This detector is designed for a strong photothermoelectric response, in which absorbed light heats up the electron gas in the graphene, resulting in an electromotive force due to the Seebeck effect. The graphene channel is assumed to be isolated from the metasurface by a several-nanometer layer of dielectric, thin enough to not impact the optical properties of the system. Fig. 3-1c shows how the asymmetric position of the graphene channel with respect to the slot allows half of the graphene channel to be gated by metal underneath, yielding the asymmetric graphene Fermi level profile needed for a nonzero net photoresponse\cite{104, 105}. Note that while perfect absorption in this wavelength range has been demonstrated in heavily doped graphene accompanied with metal nanostructures\cite{163}, I limit my consideration to very lightly doped graphene as the peak Seebeck coefficient occurs at very low doping levels allowing for a strong overall photoresponse\cite{104}, and because leveraging high doping to increase device performance would come at a considerable cost in terms of device complexity due to the need to maintain sharp carrier density junctions. I also emphasize that although broadband nanostructure-enhanced light absorption in undoped graphene has been demonstrated previously\cite{158, 164, 165, 166}, I am the first to my knowledge to specifically target broadband, wavelength-resolved photodetection in graphene and to present detectivity predictions.

For optical absorption, slot antennas offer a few advantages over planar designs such as dipole or bowtie antennas. They have unidirectional radiation patterns, and thus an array of them can perfectly absorb an incident beam, whereas planar antennas require a quarter-wave back-reflector to do so \cite{167, 168}. The wavelength dependence of the back-reflector phase complicates design of broadband absorbing
metasurfaces and exacerbates undesirable antenna-antenna coupling. Additionally, since planar antennas must be supported by transparent dielectric, they cannot be embedded in a CMOS process as the inter-layer dielectric strongly absorbs thermal infrared radiation[169, 170], whereas for slot antennas the dielectric on top and in the perforations may be etched away without sacrificing the mechanical integrity of the antenna.

3.2 Single-antenna results

3.2.1 Single antenna circuit model

Figure 3-2: a) Graphene-loaded slot antenna with physical features corresponding to the components in circuit b) labelled. b) Circuit schematic of a graphene-coupled slot antenna. $V_A$ represents incoming light, $Z_A$ is the radiation impedance of the slot aperture, $Z_{gr}$ is the impedance of the graphene sheet and $Z_{wg}$ is that of the slot, effectively a short-circuit waveguide stub.

Fig. 3-2a depicts the cross-section of what I shall refer to as Antenna Design A—a simple rectangular slot with a metallic surface. I model this as an aperture antenna fed by a rectangular waveguide terminated in a short circuit a distance $d$ behind the aperture. I represent the graphene sheet as a shunt impedance connected in parallel with the waveguide stub. Fig. 3-2b illustrates this circuit with a Thévenin equivalent radiation impedance $Z_A$ and source $V_A$ representing the aperture antenna[171]. The
rectangular waveguide stub presents an impedance

\[ Z_{wg} = Z_0 \ \frac{e^{jn_eff k_0 d} + r e^{-jn_eff k_0 d}}{e^{jn_eff k_0 d} - r e^{-jn_eff k_0 d}}, \]  

where \( Z_0 \) and \( n_{eff} \) are the characteristic impedance and effective index of the TE\(_{10}\) mode of the slot and \( k_0 \) is the vacuum wavenumber. \( r \) is the Fresnel reflection coefficient between vacuum and metal for an \( s \)-polarized plane wave at an incident angle of \( \text{arccos}(n_{eff}) \), which describes the TE\(_{10}\) mode.

The graphene presents a mostly resistive impedance of

\[ Z_{gr} = \frac{\pi^2 w}{8h \sigma_{gr}}, \]  

using power/current impedance normalization\[172\]. \( w \) and \( h \) are defined in Fig. 3-1b and \( \sigma_{gr} \) is the sheet conductance of the graphene, modeled here as intrinsic. I calculate \( Z_A \) using finite element simulations for various \( h \) and \( w \); I provide more details on these calculations in the Methods section. Fig. 3-3 provides an example of the frequency dependence of the impedances in the circuit.

![Figure 3-3](image_url)  
Figure 3-3: Wavenumber dependence of the impedances in the circuit in Figure 2a. a) Linear scaling. b) Log-abs scaling. The antenna featured here has dimensions \( d = 2.0 \mu m, h = 5.5 \mu m, \) and \( w = 0.4 \mu m \).

Define \( \eta_{gr} \) as the fraction of available power from Thévenin source \( V_A, Z_A \) that is
dissipated in $Z_{gr}$. Solving the above circuit, I arrive at

$$\eta_{gr} = 4 \left| \frac{Z_{gr} \parallel Z_{wg}}{Z_{gr}} \right|^2 \frac{\text{Re}(Z_A) \text{Re}(Z_{gr})}{|Z_A + (Z_{gr} \parallel Z_{wg})|^2},$$

(3.3)

where $\parallel$ represents reciprocal addition. Define $A_{gr} = \frac{P_{inc}}{I_{inc}}$ as the partial absorption cross section of light of intensity $I_{inc}$ coupled into the graphene and $P_{gr}$ as the power absorbed in the graphene. For a lossless, conjugate matched antenna, antenna theory predicts $A_{gr,\text{max}} = \frac{D(\theta, \phi)}{4\pi} \lambda^2$ where $D(\theta, \phi)$ is the antenna’s directivity at the given incident angle and polarization[171], which I calculate using finite element simulations. $\theta$ here is the polar angle and $\phi$ is the azimuthal angle. I omit the polarization angle in my notation, implicitly setting it to maximize $D$. Additional ohmic losses and impedance mismatch then reduce the actual absorption cross-section into graphene by a factor $\eta_{gr}$, i.e.

$$A_{gr} = \eta_{gr} \frac{D(\theta, \phi)}{4\pi} \lambda^2.$$

(3.4)

I obtain $A_{gr}$ from FDTD simulations of plane waves incident on the graphene-loaded antennas, which I then use to calculate $\eta_{gr}$ via Equation 3.4.

### 3.2.2 Circuit model comparison with FDTD results

Fig. 3-4a compares $\eta_{gr}$ between the model described by Equation 3.3 and FDTD results for antennas of various dimensions, both evaluated using an optical model for evaporated gold[173]. The data show that the model is accurate to within 10% of the $\eta_{gr}$ peak amplitude and 2% of the resonance wavenumber. I attribute these discrepancies to aspects not captured by the quasi-analytical model, such as my assumption of a perfectly conducting outer antenna face and finite meshing. Despite these shortcomings, the present model allows me to predict slot antenna absorption properties to tolerances comparable to the uncertainty due to variations in metal quality.

To further validate this model, I artificially scale the sheet conductance of the graphene load by factors ranging from 0.33 to 5 and compare the resulting $\eta_{gr}$ between the model and FDTD for a single such antenna. The results, shown in Fig. 3-4b,
show that the model accurately predicts the sublinear scaling of $\eta_{gr}$ with respect to the load conductivity, with the $\eta_{gr}$ peak amplitude reaching 0.8 for a load conductivity of $5\sigma_{gr}$.

With realistic metal, the efficiency of these devices is ultimately limited by ohmic losses. However, more advanced antenna designs can be used to improve $\eta_{gr}$. Fig. 3-5a introduces Antenna Design B, in which the slot inlet is narrower than the internal cavity width. This design concentrates the electric field around the graphene, which reduces $Z_{gr}$ without increasing the TE$_{10}$ mode loss. This allows the peak $\eta_{gr}$ to increase to up to 0.63 as shown in Fig. 3-5b.

### 3.3 Metasurface results

#### 3.3.1 Baseline design

Having modeled the individual components, I now discuss broadband absorbing metasurfaces incorporating differently tuned antennas tiled in a periodic array. I use a
three-step process to design such metasurfaces. I first constrain $d$ and $w$ to be the same for all antennas in the metasurface, and choose their values to yield high peak $\eta_{gr}$ for antennas resonant in the targeted wavelength range. Secondly, I choose the values of $h$ for the antennas, following the heuristic that at the wavelength where one antenna’s $\eta_{gr}$ falls to half its peak amplitude, the next antenna’s $\eta_{gr}$ should have risen to half its peak amplitude. Finally, I choose the arrangement and pitch of the antennas to be as closely packed as possible while satisfying qualitative fabricability constraints. I also avoid juxtaposing antennas of adjacent wavelength channels to minimize antenna-antenna crosstalk.

The antenna pitch, more accurately described by the Bravais lattice vectors of the array, is a critical parameter in determining the potential absorption efficiency of the array. Light incident from a given direction can only be scattered by the two-dimensional diffraction orders of the lattice. The array can only perfectly absorb an incoming light beam if no nonzero diffraction orders fall within the light cone, barring the event that all higher diffraction orders overlap with nodes in the individual unit cell radiation pattern. By “light cone”, I refer here to the region in the Fourier transform space of the $xy$–plane for which radiation can occur, namely $k_x^2 + k_y^2 < k_0^2$. For a square lattice, if the lattice pitch $a < \lambda_{\text{min}}/2$, no higher diffraction orders are within the light cone for any incident angle. In practice, the limited numerical
aperture of imaging systems relaxes this constraint.

Figure 3-6: a) Graphene and total absorption efficiencies of a six-antenna metasurface based on Antenna Design A with a 7 µm by 12.667 µm unit cell. The colored curves represent the contributions of each antenna to the overall graphene absorption of the metasurface. The curve colors here match the antenna colors in Fig. 3-1a. Shorter antennas yield higher resonance wavenumbers. b) Same as a), but from Antenna Design B. Here the unit cell, which contains four different antenna sizes, is reduced to 3.333 µm by 12 µm.

Targeting the 6 – 10 µm wavelength range and using Antenna Design A, I follow the above methodology and arrive at an array of six antennas with \(d = 2.25 \text{µm}, w = 400 \text{nm}, \text{and} h = 3.41, 3.81, 4.41, 5.04, 5.80, 7.36 \text{µm},\) uniformly spaced and tiled as shown in Fig. 3-1a with a 7 µm by 12.667 µm unit cell. Fig. 3-6a shows the absorption efficiency of normally incident x-polarized light into the graphene load of each antenna as well as their sum, simulated with FDTD. This graphene absorption efficiency is defined as the proportion of incident optical power absorbed in the graphene patches on the metasurface. I also plot the total absorption of the metasurface, which is the proportion of incident optical power absorbed overall (including in the gold); its antenna-by-antenna breakdown is plotted in Fig. 3-7a. With an average efficiency of 36% across the 1050 cm\(^{-1}\) to 1600 cm\(^{-1}\) band, this structure improves upon the \(1/N\) limit of filter array-based imagers by more than a factor of two, albeit with the caveat that y-polarized incident light gets reflected with negligible absorption as it does not couple to the slot mode. Note that the unit cell highlighted in Fig. 3-1a is not the primitive unit cell of the lattice, although it was used as the FDTD simulation region.
due to software constraints.

Figure 3-7: Total absorption efficiency (including both graphene and metal contributions) and individual antenna contributions for both Antenna Design A- and B-based metasurfaces corresponding to Fig. 3-6.

3.3.2 Effect of pitch variation

To further understand the physics of these metasurfaces, I vary the pitch of the unit cell while keeping all other parameters constant. The resulting total absorption efficiency spectra for six different cases are shown in Fig. 3-8a, and their mean efficiencies averaged between 1050 cm$^{-1}$ and 1600 cm$^{-1}$ are plotted in Fig. 3-8b. I show the individual absorption contributions from each antenna for each metasurface pitch in Fig. 3-9, and I list the unit cell widths and heights in Fig. 3-8c. The data show that the absorption efficiencies are roughly constant, and comparable with the peak efficiencies obtained in Fig. 3-4a, up to the 7 µm–wide unit cell. For larger unit cells, the mean efficiency decreases with increasing unit cell size. This can be understood by analyzing the diffraction characteristics of the various metasurfaces. Sparser metasurfaces have tighter reciprocal lattices and thus more diffraction orders are available within the light cone. For normally incident illumination as I am using here, the minimum wavelength for which no higher-order diffraction occurs (i.e., completely specular reflection) is given by $\lambda_c = \max\left(\left(\frac{a_1^{-2} + a_2^{-2}}{2}\right)^{-1/2}, \frac{a_1}{2}, \frac{a_2}{2}\right)$, where $a_1$ and $a_2$ are the unit cell dimensions. The corresponding wavenumbers for the various unit
Figure 3-8: a) Total graphene absorption efficiency versus wavenumber for metasurfaces with varied unit cell pitch. The legend indicates the width of the unit cell as illustrated in Fig. 3-1a. b) Mean absorption efficiency for each of the curves in a), averaged between 1050 cm\(^{-1}\) and 1600 cm\(^{-1}\). c) Unit cell dimensions used to obtain the results in a) and b) and the corresponding maximum wavenumbers for which normally incident light experiences no higher order diffraction. The unit cell x and y pitch were varied together to attain a roughly uniform antenna-antenna proximity.
Figure 3-9: Graphene absorption efficiency and individual antenna contributions for Antenna Design A-based metasurfaces of varied pitch. Full unit cell dimensions for each case are given in Fig. 3-8.
cells used here are listed in Fig. 3-8c. For the 9 µm and 11 µm width unit cells, the threshold wavenumber falls in the middle of the range of interest, resulting in reduced efficiency as diffracted light cannot participate in destructive interference with specularly scattered light. This is especially apparent for the 9 µm unit cell, which exhibits a clear spectral transition between high and low efficiency at the diffraction threshold. Note that choosing an excessively tight antenna spacing is also detrimental, not only due to fabrication difficulty, but also because there also must be room between adjacent antennas to place the back-gated part of the graphene channel as well as the contacts as shown in Fig. 3-1. For the design modeled in Fig. 3-6a, the antenna-antenna spacing is 1.167 µm, leaving sufficient room for an 800 nm graphene channel (to cover the 400 nm antenna slot, and then an additional 400 nm to one side to complete the p-n junction).

3.3.3 Discussion of angular acceptance

For a spectrally sensitive metasurface to be practical, not only must it maintain a high absorption efficiency for a reasonable range of incoming light directions, but also the absorption spectra of the individual antennas must not shift or distort too strongly as the incoming light angle varies. The directional dependence of the metasurface arises from two factors: The directivity profile $D(\theta, \phi)$ of the individual antennas, and array effects resulting from interference and antenna-antenna coupling. Although full angle-dependent simulation results for these gold metasurfaces are outside the scope of this paper due to the extreme computational overhead of off-angle periodic structure simulations[174], I can still provide insight by elaborating on the aforementioned factors, and I also perform off-angle simulations of a simplified metasurface constructed of Perfect Electrical Conductor (PEC) which permits a much larger mesh size. Figures 3-10a, b and c display the directivity profile of a 6.5 µm × 400 nm antenna on resonance. This profile is similar to those of the other antenna lengths used in the metasurface. Intuitively, the directivity decreases as the incident angle approaches the long axis of the antenna, and I thus expect a similar trend in the directional dependence of the array. In Fig. 3-10d, I plot for three wavenumbers the sets of incident
Figure 3-10: Different representations of the on-resonance far-field radiation pattern of a 6.5 µm-long slot antenna, which is qualitatively very similar to those of antennas of other lengths. a) 3D-representation of the far-field directivity pattern. $D(\theta, \phi)$ goes to zero for light incident along the long axis of the aperture. b) Directivity vs. elevation angle $\theta$ along the azimuth of the antenna’s long axis, $\phi = 90^\circ$. c) Directivity in directional cosine space. Note also that $k_x = k_0 u$ and $k_y = k_0 v$ where $k_x$ and $k_y$ are the $x$- and $y$-components of the incident wavevector. d) Depiction in $k_x k_y$–space of the light cone edges (colored rings) and the specular reflection-only incident light directions (colored patches) for three different wavenumbers for the 7 µm × 12.667 µm 6-antenna unit cell design in Fig. 3-1a. The black dots represent the reciprocal lattice points of the metasurface. For each wavelength, the specular reflection-only region is the set of remaining $\{k_x, k_y\}$ after subtracting from the light cone all copies of the light cone transposed by all nonzero reciprocal lattice vectors.
Figure 3-11: Graphene absorption efficiency and individual antenna contributions for a perfectly conducting metasurface with various off-angle excitation directions. Using Perfect Electrical Conductor (PEC) as the metasurface material allows me to use a coarser mesh of 40 nm, as the electric field does not penetrate the conductor (the “skin depth” which requires careful modelling for realistic metal). I also adjust the metasurface dimensions to align all features to the mesh: The unit cell is $7.2 \times 12.64 \mu m$, and the antenna lengths are $3.4$, $3.8$, $4.4$, $5.04$, $5.8$, and $7.36 \mu m$. The topmost plot shows the absorption spectra for normally incident light. The second row depicts the case of light incident perpendicular to the antennas’ long axes (along the line of maximum directivity), and the third row depicts light incident perpendicular to the antennas’ short axes.
light directions, represented as components $k_x$, $k_y$ of the incident wavevector $\mathbf{k}$, for which only specular reflection from the metasurface occurs. For $\lambda^{-1} = 1050$ cm$^{-1}$, all light incident within 45° of normal is only specularly reflected. This range decreases with increasing wavenumber until normally incident light is pinched off at 1579 cm$^{-1}$.

As in Fig. 3-8, I expect efficiency to suffer when the specular reflection-only condition is not met. Fig. 3-11 shows the results of off-angle simulations of the simplified PEC metasurface. The data show that for light incident off-angle but perpendicular to the antennas’ long axes, antenna resonances falling outside the specular reflection-only region are subject to decreased absorption efficiency as well as blue-shifting. For light incident off-angle and perpendicular to the antennas’ short axes, I obtain similar results, except that the peaks red-shift instead of blue-shift, and I observe an overall reduction of the absorption at steep incident angles due to the reduced directivity.

3.3.4 Different numbers of spectral channels

I also investigate metasurfaces comprising numbers of spectral channels besides $N = 6$. Fig. 3-12 shows the geometric details and simulation results for metasurfaces with $N = 3$, 4, 5, and 8 as well as the default value of 6. I achieve good results with uniformly high absorption efficiency for $N = 5$ and 6. For lower values of $N$, the wide frequency spacing between the individual resonances yields deep troughs in the overall absorption efficiency curve, whereas for higher values of $N$, excessive overlap between the antenna resonances causes the overall metasurface efficiency to suffer. Moving to $N = 8$ but maintaining the same antenna-antenna pitch as the reference design plotted in Fig. 3-6a, $\lambda_c$ falls within the wavelength range of interest which impacts the spectral response, so I also include spectra for an 8-antenna metasurface with tighter-spaced antennas to better isolate the effect of having overlapping antenna resonances.
Figure 3-12: Graphene absorption efficiency and individual antenna contributions for Antenna Design A-based metasurfaces with varied numbers $N$ of antennas of different lengths. The lower-right table shows the arrangements of the antennas within the unit cell, with 1, 2...$N$ corresponding to the shortest, second shortest, etc. up to longest antenna. The table also lists the lengths of the antennas in each case, chosen roughly to maintain a constant degree of frequency-space overlap between adjacent spectral channels as simulated on an individual antenna basis. Antennas are placed on a $1.167 \, \mu m \times 6.333 \, \mu m$ grid as in the reference metasurface design plotted in Fig. 3-6a except for the last spectrum, which is placed on a $0.875 \, \mu m \times 6.333 \, \mu m$ grid to match the overall unit cell size and hence $\lambda_c$ of the reference design.
Figure 3-13: Graphene absorption efficiency and individual antenna contributions for Antenna Design B-based metasurfaces of varied pitch. The unit cell dimensions are 2.667 µm × 11.5 µm, 3.333 µm × 12 µm, 4.333 µm × 12.667 µm, 3.0 µm × 13.333 µm, 7.333 µm × 14 µm.
Figure 3-14: Graphene absorption efficiency and individual antenna contributions for Antenna Design B-based metasurfaces with varied numbers $N$ of antennas of different lengths. The lower-right table shows the arrangements of the antennas within the unit cell, with $1, 2...N$ corresponding to the shortest, second shortest, etc. up to longest antenna. The table also lists the lengths of the antennas in each case, chosen roughly to maintain a constant degree of frequency-space overlap between adjacent spectral channels as simulated on an individual antenna basis. Antennas are placed on a $0.833 \mu \text{m} \times 6 \mu \text{m}$ grid as in Fig. 3-6b.
3.3.5 Metasurface from slitted antennas

Antenna Design B features a broader absorption peak than Antenna Design A; therefore, in assembling such antennas into a metasurface, fewer different antenna sizes can be placed in the desired wavelength range without causing excessive resonance overlap. However, since the antenna slots are thinner, the graphene channel on each antenna is correspondingly shorter, so the antennas can be spaced closer together, affording improved angular acceptance range. Following the same design principles as above, I arrive at a metasurface design with graphene and total absorption curves shown in Fig. 3-6b. This design features $w = 0.6 \, \mu m$, $d_i = 60 \, nm$ and $w_i = 150 \, nm$ for all antennas, and $h = 3.00, 3.76, 5.09, 8.03 \, \mu m$, arranged as described in Fig. 3-14. It has an mean absorption efficiency of 58% averaged in the spectral range from 1050 cm$^{-1}$ to 1700 cm$^{-1}$ (note the wider wavelength range). As with the previous metasurface design, I investigate the antenna-by-antenna contributions to the total absorption of the of the metasurface as well as the variation in the absorption spectrum as I vary the antenna pitch and spectral channel count; the results are shown in Figures 3-7b, 3-13 and 3-12 respectively and display qualitatively similar characteristics to the metasurfaces based on Antenna Design A.

![Figure 3-15: Illustration of how Antenna Design B could be implemented in the wiring layers of a CMOS chip.](image)

3.4 Sensitivity performance estimation

I can combine data describing the resistivity of gated graphene with the metasurface simulation results to estimate the room temperature detectivity of the spectral
Table 3.1: Estimated sensitivity figures for graphene imaging array.

<table>
<thead>
<tr>
<th>Design</th>
<th>$R_c$ [Ω·μm]</th>
<th>$\Re$ [A/W]</th>
<th>NEP [pW Hz$^{-\frac{1}{2}}$]</th>
<th>$D^*$ [Jones]</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0.70</td>
<td>11.</td>
<td>$5.8 \times 10^7$</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.36</td>
<td>14.</td>
<td>$4.6 \times 10^7$</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>4.0</td>
<td>3.3</td>
<td>$1.4 \times 10^8$</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.1</td>
<td>5.5</td>
<td>$8.2 \times 10^7$</td>
</tr>
</tbody>
</table>

imagers based on antenna designs A and B described above, corresponding to the efficiency spectra in Fig. 3-6. Using the methods described in Section 2.1.2 and the electrical properties of polycrystalline, non-annealed graphene achieved by de Fazio et al.[110], I arrive at the values in Table 3.1 with and without accounting for graphene contact resistance for normally incident $x$-polarized light. In arriving at these figures, to reflect the reality that achieving a pn-junction in the graphene requires some degree of doping, I re-execute the metasurface simulations with the graphene doped to $E_F = 0.05$ eV; I discuss this and other details related to the detectivity calculations in the Methods section. For comparison, more conventional bolometer-based thermal IR FPAs operating at room temperature have been reported to achieve detectivities in the $1-2 \times 10^9$ Jones range, although such devices do not feature spectral resolution and are limited to millisecond-range response times[175, 176].

Standard CMOS processes provide a potential avenue for the practical implementation of these antenna structures. Copper, which is universally used as wiring in the back end of CMOS devices [177], has been shown to exhibit similar or even slightly superior optical properties to gold in the mid-IR wavelength range when care is taken in its deposition. Given the immense effort which has been directed towards minimizing the resistivity of CMOS wiring, I am optimistic that this material would exhibit optical properties consistent with those achieved in such “alternative metal” plasmonic research [178]. To comply with typical CMOS design rules in which wiring consists of alternating “metal” (horizontal wire) and “via” layers, it would be necessary to perforate the slot walls, as shown in Fig. 3-15. If the perforations, considered as rectangular waveguides, are sub-cutoff for the resonant wavelength and sufficiently deep, they do not leak light. Additionally, it would be necessary to etch away the
inter-layer dielectric to prevent mid-IR light absorption. In academic environments, on the other hand, one may consider fabricating such antennas by electroplating gold around a mold of the slot cavity.

I would like to emphasize the general applicability of the antenna metasurface concept to other wavelength ranges, photosensitive elements and antenna designs. I propose that besides graphene and other 2D materials, III-V or HgCdTe semiconductor photosensitive elements could also be incorporated by a transfer printing heterointegration process[132]. The slot antenna-based metasurface imager approach could also scale to terahertz, where ohmic losses are much less than in the mid-IR[179] and the antennas could be fabricated directly in a circuit board-like platform.

In conclusion, I introduced a six-spectral-channel graphene-coupled slot antenna metasurface with 36% efficiency functioning as a spectral imager in the thermal IR, as well as a model for estimating the optical properties of the individual antennas therein. I also introduced a more complex four-spectral-channel metasurface based on slots with thinned inlets which allows me to increase the efficiency to 58% averaged between $1050\,cm^{-1}$ and $1700\,cm^{-1}$. This device is appropriate for integration in the wiring layers of a CMOS process with suitable post-processing to remove inter-layer dielectric within the cavity and transfer graphene. Further research on this concept may focus on experimental demonstration of the absorption enhancement functionality, or on optimizing the device design to meet certain engineering goals.

3.5 Methods

3.5.1 Simulation details

Unless otherwise specified, I use the evaporated gold optical model described in Palik et al. throughout the paper [173]. I model graphene as an infinitely thin conductive sheet using the optical conductivity model described in Hanson [102]. As input parameters to the model, I use a temperature of 300 K, intrinsic graphene (zero Fermi level), and a scattering rate $\Gamma = 0.514\,meV$. 

81
Figure 3-16: Graphene absorption efficiency and individual antenna contributions for a 9\(\mu\)m by 13.333\(\mu\)m metasurface with varied minimum FDTD mesh size as discussed in the Methods section of the main paper. The inset plots the mesh dependence of the mean graphene absorption efficiency in the 1050 cm\(^{-1}\) to 1600 cm\(^{-1}\) range.

I use the Lumerical FDTD package for all FDTD simulations. For simulations of individual antennas, I use \(x\)-, \(y\)- and \(z\)-meshes of 8 nm in the vicinities of the slot aperture and slot bottom, as well as a \(z\)-mesh of 40 nm within the slot. As such, the finest meshes coincide with metallic surfaces and corners, allowing me to capture the nonzero skin depth of the metal, whereas the more gradual \(z\)-dependence of the fields inside the slot permits a coarser mesh. I find that a minimum mesh size of 8 nm yields converged results for these simulations. I use PML boundary conditions on all sides except for the \(-z\) side, where I use a metallic boundary condition as the light does not penetrate beyond the slots anyway. I also apply symmetry conditions across the \(xz\) and \(yz\) planes. For the metasurface simulations, I use the same meshing scheme, but with a fine mesh of 15 nm and a \(z\) mesh of 100 nm within the slot due to computational resource availability limits. To illustrate the error associated with this choice of mesh, I plot mesh-dependent absorption efficiency curves for a 9\(\mu\)m by 13.333\(\mu\)m unit cell, \(N = 6\) metasurface in Fig. 3-16, with the mean efficiency
averaged between 1050 cm\(^{-1}\) and 1600 cm\(^{-1}\) plotted in the inset. The results validate my qualitative conclusions and put an approximately 3% relative error bound on the spectrally averaged efficiencies of the metasurfaces, although the coarse mesh does somewhat distort the actual spectra. For the metasurface simulations, I change the \(x\)- and \(y\)- boundary conditions to Bloch boundary conditions to reflect the periodic nature of the metasurface, and I apply symmetry across only the \(xz\) plane.

### 3.5.2 Impedance model details

I calculate \(Z_0\) and \(n_{\text{eff}}\) for the impedance model using the Lumerical MODE waveguide mode solver with an 8 nm mesh. I use the graphene model from Falkovsky \[160\], which gives almost identical results to the Hanson model used by Lumerical for the parameters I use. I use Ansys HFSS finite element software to calculate \(Z_A\). These simulations excite the aperture from within by its \(\text{TE}_{10}\) mode yielding the \(S_{11}\) scattering coefficient of the internally reflected light, from which the software calculates \(Z_A\). In the finite element simulations, I model the aperture and slot as perfect electrical conductors, as I expect the real part of the antenna impedance to be dominated by radiative loss (rather than ohmic loss) and the imaginary part by energy storage in the near field of the aperture (rather than plasmonically within the metal). From these same simulations I also extract the antenna directivity \(D(\theta, \phi)\).

### 3.5.3 Detectivity estimation

Detectivity estimations follow the graphene photothermoelectric response calculations introduced in Section 2.1.2, in which the electronic temperature profile in the graphene under illumination is described by Eqn. 2.2. I choose an antenna with \(w = 400\) nm and \(h = 5.5\) \(\mu\)m in these calculations, extracting the spatial dependence of \(\dot{Q}\) from Ansys HFSS finite element simulations. For the graphene’s conductivity \(\sigma\) as a function of Fermi level, I use the data measured by de Fazio et al. for unannealed, polycrystalline graphene \[110\]; this is then used to calculate \(\kappa\) via the Wiedemann-Franz law and
Π as well as the Seebeck coefficient $S$ via the Mott formula. The value of $\gamma C_{el}$ is estimated by assuming a electronic thermal cooling length of $\sqrt{\kappa/\gamma C_{el}} = 1 \mu$m, an empirical value[180]. As shown in Fig. 3-1b, the graphene is assumed to be terminated at the slot edge on one side, and is taken to extend 400 nm past the slot edge on the other side, where its Fermi level is gated through the metal to the n-type Seebeck coefficient peak. The graphene Fermi level in the suspended region is simply taken to be the zero-gate-voltage Fermi level from de Fazio et al. as it cannot be controlled. For simplicity I assume a sharp jump in the values of $\sigma$, $\kappa$, $\Pi$ and $S$ between the n- and p-doped sides of the device, neglecting fringing fields from the gate. The graphene channel is assumed to be short-circuited with graphene-metal contact resistances $R_c$ of either 0 or 1000 $\Omega \mu$m per contact, the latter being consistent with 1-dimensional contacts to graphene near the Dirac point[181]. The average $\Delta T_{el}$ at the graphene p-n junction and the Seebeck coefficients on either side determine the thermal electromotive force via $E = -S \nabla T[109]$, which in turn determines $j$ via the total device resistance. Thus, Equation 2.2 including the Peltier term may be solved directly, as $j$ is a linear functional of $\Delta T_{el}$. Solving for $\Delta T_{el}$ over two spatial dimensions, I find linear thermal decay profiles along the $+x$ and $-x$ directions away from the $\Delta T_{el}$ peak which indicates that the device is in the short-channel regime where carrier cooling is dominated by the $\Delta T_{el} = 0$ boundary conditions, and $\sqrt{\kappa/\gamma C_{el}}$ is large enough for the electron-phonon interaction term to be inconsequential.

Having obtained the short-circuit responsivity $R$ under zero bias as such, I calculate the noise-equivalent power of the device assuming Johnson noise at 300 K as the dominant noise source, a reasonable assumption for an unbiased device[182]. The detectivity for the array is calculated incorporating the antenna pitch, noting that there are two antennas per unit cell. To account for the decreased optical absorption efficiency of a metasurface loaded with graphene doped to the Seebeck coefficient peaks of roughly $\pm 0.05$ eV, I redo the simulation used to generate Fig. 3-6a with the graphene doped as such. I plot the resulting absorption spectra in Fig. 3-17, which shows a mean absorption efficiency of 33% averaged between 1050 cm$^{-1}$
Figure 3-17: a) Graphene absorption efficiency and individual antenna contributions for a 7µm by 12.667µm metasurface with the same geometric parameters used to generate Fig. 3-6a, but with the graphene doped to 0.05 eV which decreases the graphene sheet conductivity to a degree, especially for longer wavelengths. Here, the mean efficiency averaged between 1050 cm$^{-1}$ and 1600 cm$^{-1}$ is 33%. This doping level is chosen to maximize the Seebeck coefficient for the graphene model used to approximate the device detectivity as discussed in the main text. b) Same as a), but for the 4-antenna Antenna Design B metasurface design used to generate Fig. 3-6b as discussed in the main text.

and 1600 cm$^{-1}$ for the Antenna Design A metasurface and 54% averaged between 1050 cm$^{-1}$ and 1700 cm$^{-1}$ for the Antenna Design B metasurface. I finally calculate the external values of responsivity, NEP, and detectivity by scaling the internal values by the average single-antenna peak efficiency.

3.6 Fabrication

Over the course of this theory project, I also investigated various approaches to fabricate the proposed antenna metasurfaces structures. Here I describe the most successful process and present the fabrication results as well as preliminary infrared scattering measurements.

3.6.1 Flip-chip fabrication process

Fig. 3-18 shows the process flow for the slot antenna process. I first coat a thermal oxide (285 nm on silicon) chip with SU-8 2002 photoresist to a thickness of 2.2 µm,
followed by a 60s at 70°C, 60s at 100°C, and finally 60s at 70°C hot plate bake schedule. I expose the SU-8 metasurface pattern at a nominal dose of 20 µC/cm² with proximity effect correction used to compensate for pattern density variations. The SU-8 layer is then post-baked for 60s at 70°C, 300s at 90°C, and 60s at 70°C, followed by 60s of development in PGMEA, 60s IPA rinse and blowdry to produce an array of high-aspect ratio SU-8 fins, show in Fig. 3-19a. I then use oxygen plasma to remove residual organic residue from the chip without removing too much material from the fins, after which deposit 1.5 nm Ti followed by 20 nm Au conformally onto the fins using e-beam evaporation with a planetary sample holder. I then followed by electroplating in a sulfite gold plating solution (Transene TSG-250 solution with a 0.5% addition of Transene TSG Hardener/Brightener) at nominal mean current of 2.2 mA/cm² with a 50% duty cycle to completely cover the fins with thick gold and provide mechanical integrity and chemical resilience. I then use Norland Optical Adhesive NOA-61 to flip and glue the gold slot chip upside-down onto a quartz carrier chip as shown in Fig. 3-18, followed by exposing the NOA-61 to an ultraviolet light source to harden the glue and an overnight bake at 60°C to produce a full cure. I
then use a diamond grid polishing sheet and motorized rotary lapping tool to remove most of the original Si substrate, followed by sonication in water, a 3-minute dip in room temperature Nano-Strip (a stabilized piranha solution formulation) to clean away organic residues, and a 3-minute dip in 7:1 buffered oxide etch (BOE) to remove any native oxide from the silicon, followed by thorough rinsing in deionized water. The chip is then very quickly transferred from the beaker of water to a XeF$_2$ etcher to minimize native oxide regrowth and etched until none of the silicon substrate remains. Finally, the 285 nm silica layer is removed with a 5 minute BOE etch, and the original SU-8 mold is removed by placing a few drops of piranha solution (3:1 sulfuric acid to hydrogen peroxide) onto the chip, heated to 100°C on a hot plate, and allowing the SU-8 to etch out for 10 minutes followed by a deionized water rinse and dry. The resulting structure is shown in Fig. 3-19b with a sheet of graphene transferred on top; the graphene transfer process is the same as that described in Section 4.4.1 except with critical point drying used after removing the PMMA support film to gently suspend the graphene over the gold slots.

### 3.6.2 Cross-polarized scattering measurements

I performed FTIR cross-polarized scattering measurements on arrays of slot antennas, in this case fabricated with an earlier process variation involving direct gold plating...
Figure 3-20: Comparison between measured and simulated cross-polarized scattering for rectangular arrays of individual antennas with $w = 240$ nm, $d = 1.6 \mu$m, and $h$ varied from $1.75 \mu$m to $4.0 \mu$m in increments of $250$ nm. Each array had a pitch of $2h \times 2h$, accounting for the reduced scattering for longer antennas. Inset: Earlier slot antenna fabrication process gave rise to gold nanoparticles precluding graphene transfer, shown here along the edges of a slot.

Through SU-8 fins produced via the same process as described in Section 3.6 on a Ti/Au-coated silicon substrate. In this case, the electroplating was terminated after depositing approximately $1.6 \mu$m of material, followed by removal of the SU-8 fin in piranha solution, leaving behind a gold slot suitable for infrared reflection measurements but insufficiently smooth for subsequent integration of 2D materials as shown in the inset to Fig. 3-20. Each array consisted of a $500 \mu$m $\times$ $500 \mu$m square array of slot antennas of varying $h$, $w = 240$ nm, and $d = 1.6 \mu$m. Each array had a pitch of $2h$ along both directions. To measure the resonant frequency and quality factor of the antennas, I illuminated each array in an FTIR microscope at a $45^\circ$ angle from the slots, and measured the reflection filtered through another polarizer placed $90^\circ$ rotated from the incident light polarization. This allows elimination of light specularly reflected from the metal surface, such that only light which interacts with the electromagnetic mode supported by the slot antennas can be scattered from the incident polarization to the measured polarization. The resulting spectra for $h$ varied from $1.75 \mu$m to $4.0 \mu$m in increments of $250$ nm are represented as solid lines in Fig.
3-20, with shorter $h$ corresponding to higher-wavenumber resonances. To validate these measurements, I also performed FDTD simulations designed to emulate the measurement conditions, calculating the absolute efficiency with which incident light (normally incident in the case of the FDTD simulations, which differs from the wide angle of incidence range occurring in FTIR measurements) scatters between the two polarizations diagonal to a slot. The results are shown in dashed lines in Fig. 3-20. The FTIR measurements, which were not power-calibrated, have been vertically scaled to match that of the FDTD simulations. The comparison shows excellent agreement between the measured and simulated resonance frequencies, although the quality factors of the as-fabricated antennas are found to be less than as predicted, which may be due to the particulate contamination around the edges of the slots as well as the quality of the electroplated metal. This validates our theoretical conception of the antenna slots as supporting electromagnetic resonances.
Chapter 4

Waveguide-Integrated Mid-Infrared Photodetection using Graphene on a Scalable Chalcogenide Glass Platform

The development of compact and fieldable mid-IR spectroscopy devices represents a critical challenge for distributed sensing with applications from gas leak detection to environmental monitoring[86]. Recent work has focused on mid-IR photonic integrated circuit (PIC) sensing platforms[65, 67, 66] and waveguide-integrated mid-IR light sources[77] and detectors based on semiconductors such as PbTe[71], black phosphorus[87] and tellurene[91]. However, material bandgaps and reliance on SiO$_2$ substrates limit operation to wavelengths $\lambda \lesssim 4\mu$m. Here I overcome these challenges with a chalcogenide glass-on-CaF$_2$ PIC architecture incorporating split-gate photothermoelectric graphene photodetectors. My design extends operation to $\lambda = 5.2\mu$m with a Johnson noise-limited noise-equivalent power of 1.1 nW/Hz$^{1/2}$ with no fall-off in photoresponse up to $f = 1$ MHz and a predicted 3-dB bandwidth of $f_{3dB} > 1$ GHz. This mid-IR PIC platform readily extends to longer wavelengths and opens the door to applications from distributed gas sensing and portable dual comb spectroscopy (DCS)[51] to weather-resilient free space optical communications[29].
4.1 Introduction

Mid-IR absorption spectroscopy is a critical tool for chemical sensing and analysis, especially for inert gases that evade detection by chemical reaction-based sensors such as metal oxide semiconductor and electrochemical sensors. The halogenated chemistry that renders many such gases inert also gives them potent global warming potential due to carbon-halogen stretching modes resonant in the thermal IR[23, 183], augmenting the importance of rapid detection of leaks of these gases. Besides inert gas sensing, optical gas sensors also find use in applications requiring differentiation between chemically similar analytes that might not be distinguishable via other sensing techniques due to the superior selectivity afforded by their spectroscopic sensing mechanism[184]. Current infrared gas sensors use co-packaged discrete optics, the size and cost of which precludes broad, networked deployment around chemical plants, electrical utility sites, and other locations liable to greenhouse gas leaks, hampering sensitive and responsive leak detection. For instance, a commercially available handheld IR gas sensor may consist of a thermal radiation source, a gas flow tube, a spectral filter and an infrared sensor to measure changes in light absorption at a specific absorption peak[24]. A spectroscopic gas analyzer, capable of distinguishing tens or hundreds of different gases, would typically use a Fourier Transform Infrared (FTIR) operating mechanism, incorporating a mechanically tunable interferometer and occupying roughly the volume of a desktop computer[78, 185].

To address the challenges of optical gas sensor size and fieldability, MIR photonic integrated circuit platforms have been investigated to reduce optical gas sensors to the size of a chip. Recent work demonstrated integrated optical methane[66] and volatile organic compound[67] sensing, but required coupling to off-chip sources and detectors. However, integrating the detector on-chip is more compact and can improve sensitivity by reducing the volume of active material able to generate thermal noise. Su et al. achieved integration of a PbTe photoconductor and demonstrated methane sensing at a wavelength of $\lambda = 3.31 \mu m$[71], but their platform is limited to $\lambda \lesssim 4 \mu m$ by absorption in the SiO$_2$ substrate[86] and by PbTe’s absorption cutoff[44]. Waveguide-
Figure 4-1: a) Illustration of the device cross-section perpendicular to the waveguide axis. The optical mode supported by the GSSe waveguide evanescently couples to and is absorbed by the graphene channel, which is gated by two graphene back-gates to induce a $p$-$n$ junction. b) Optical image of the device depicting source, drain and gate contact pads. c) Depiction of the optical guided mode at $\lambda = 5.2 \mu m$.

Integrated detectors based on narrow-gap 2D materials black phosphorus[87] and tellurene[91] have also been demonstrated, but they too are bandgap-limited to $\lambda \lesssim 4 \mu m$.

Here we exceed the wavelength limit of previous demonstrations using graphene-based detectors on an extended-transparency waveguide platform. While graphene integrated detectors have shown promise at telecom wavelengths[115], the material’s advantages improve at longer wavelengths due to the thermal nature of the photothermoelectric (PTE) response mechanism[104, 105] and due to the impact of optical plasmon scattering at short wavelengths[106]. Integrated photodetection with graphene was demonstrated at $\lambda = 3.8 \mu m$, but on a SiO$_2$ platform[89]. To access longer wavelength operation and achieve good sensitivity at zero bias, I introduce a Ge$_{28}$Sb$_{12}$Se$_{60}$ (GSSe)-on-CaF$_2$ waveguide platform supporting gated PTE-based graphene photodetectors.
Figure 4-2: a) Measured zero-bias photovoltage produced by the device as a function of the two gate voltages. b) Total device resistance as a function of the two gate voltages. c) Lock-in signal reflecting power measured by an InAsSb photodetector at the focal point of the output facet collection lens, used to monitor transmission of the device as a function of the gate voltages. d, e, f) Plots of line sections indicated with dashed lines in figures a, b, and c, respectively.
4.2 Results

4.2.1 Device design

Figs. 4-1a and 4-1b illustrate the platform and photodetector design. The device consists of a 750 nm-tall by 3.2 μm-wide single-mode GSSe waveguide fabricated on top of a 5.4 μm wide by 300 μm long, CVD-grown graphene channel, flanked on either side by source and drain contacts that run along the full length of the channel\(^1\) and are connected to electrical probe pads. Beneath the graphene channel is a pair of CVD graphene back-gates, separated by a 400 nm and used to electrostatically induce a p-n junction along the center of the channel. I use HfO\(_2\) as the gate dielectric and as an airtight capping layer to prevent oxidative deterioration and contamination of the graphene channel. The device is fabricated on a CaF\(_2\) substrate, which is transparent up to \(\lambda = 8 \text{ μm}\) and has an refractive index of \(n = 1.4\), making it a suitable cladding material. Fig. 4-1c depicts the resulting waveguide mode at \(\lambda = 5.2 \text{ μm}\). Light enters and exits the chip via a pair of cleaved waveguide facets on opposing sides of the chip, which are laterally offset by 5 mm to reduce the amount of stray input light passing through the aperture stop of our collection optics.

4.2.2 Raw measurement results

I use lock-in measurement to characterize the detectors, focusing light from a \(\lambda = 5.2 \text{ μm}\) QCL source into the chip’s input facet via a molded aspheric lens. Light exiting the chip is focused by a high-NA germanium aspheric collection lens onto an InAsSb photodetector and amplified for transmission measurement. Fig. 4-5a in the Methods section depicts this beam-path in more detail. Operating under zero bias, I measure the voltage signal produced by the device directly using a separate lock-in amplifier. I also measure the device resistance between source and drain as a function of the gate voltages with no input light.

\(^1\)I choose definitions of channel “length” and “width” that are opposite those used when discussing FETs to maintain consistency with the notion of wavelength length.
signal versus both gate voltages for one such photodetector (“Device A”). Here, I chop the $\lambda = 5.2 \, \mu\text{m}$ QCL source at 3.78 kHz with a guided “on” power of 11 $\mu$W at the detector input. Prior to each data point collection, both gate voltages are reset to $-7 \, \text{V}$ for 80 ms to reset the gate dielectric hysteresis (see Section 4.5.1), then set to the desired gate voltages and allowed to dwell for 200 ms for the lock-in signal to stabilize. The lock-in filter is set to a 30 ms time constant with a 12 dB/octave falloff. The detector photovoltage for in Fig. 4-2a is measured directly by the lock-in amplifier with no additional amplification. For the resistance map in Fig. 4-2b, I use a lock-in amplifier to bias the device with a 1 VRMS sine wave at 3.78 kHz through a 100 k$\Omega$ resistor to act as a current source and measure the voltage across the device with the lock-in. From the photovoltage and resistance maps, alongside the power and loss calibrations described in Section 4.4.2, I infer gate voltage pairs to realize maximum voltage responsivity, maximum current responsivity, and minimum NEP with respect to Johnson noise, indicated with green markers in Fig. 4-2. For these, I arrive at $1.5 \, \text{V/W}$, $10. \, \text{mA/W}$, and $1.1 \, \text{nW/Hz}^{1/2}$, respectively. The observed photovoltage gate map indicates a PTE response mechanism, evidenced by the six-fold sign change pattern around the graphene channel’s charge neutral point[104]. I also observe that the device exhibits a relatively strong photoresponse in the far upper-left and far lower-right corners of the gate map, overlapping with regions of high transmitted power. I attribute this to intraband absorption in the graphene, which is weaker than interband absorption for charge-neutral graphene in this wavelength but still strong enough to produce a significant photoresponse.

4.2.3 Comparison with device model

To confirm my understanding of device operation and elucidate the prospects for performance improvement, I apply the formalism introduced in Section 2.1.2 to calculate the electronic temperature distribution and Seebeck photovoltage in the graphene channel under illumination. Figs. 4-3a and 4-3b compare my measured and modelled voltage responsivities using calculations described in Section 4.4.3. Due to the material quality variation of as-transferred CVD graphene samples, the performance of
Figure 4-3: a, b) Contour plots of the a) measured and b) modelled responsivity maps of the device, evaluated with $\tau_{\text{DC}} = 3.5 \, \text{fs}$, $\tau_{\text{IR}} = 40 \, \text{fs}$, $\sigma_n = 2 \times 10^{12} \, \text{cm}^{-2}$, $\tau_{\text{eph}} = 50 \, \text{ps}$, and $\alpha_e = 2.5 \, \text{mm}^{-1}$. c) Electron temperature increase $\Delta T_{\text{el}}$ and absorbed optical power per area $\dot{\mathcal{Q}}$ profiles in the graphene channel per guided optical power at gate voltages of $\{-2.35 \, \text{V}, 0.35 \, \text{V}\}$, chosen to maximize the modelled photoresponse, and other parameters as above.

The device depends on several fitting parameters, whose definitions and approximate values (derived from the measured data) I provide in Table 4.1. I describe my fitting process in Section 4.4.4. Critically, all features of the modelled responsivity map in Fig. 4-3b up to an overall scale factor from $\tau_{\text{eph}}$ are established \textit{a priori} from fitting parameters extracted from the device transmittance and resistance maps, with only $\tau_{\text{eph}}$ obtained by matching the scales of the measured and modelled responsivities. The resemblance between Figs. 4-3a and 4-3b thus reflects the validity of my PTE model and is not due to over-fitting. In Fig. 4-3c, I plot the solution to Eqn. 4.5, $\Delta T_{\text{el}}(x)$, as well as the source term $\dot{\mathcal{Q}}(x)$. The model predicts that $9 \, \mu\text{W}$ of guided
Table 4.1: Device parameters and approximate values

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_{DC}$</td>
<td>Drude scattering time at DC</td>
<td>$\approx 3.5 \text{ fs}$</td>
</tr>
<tr>
<td>$\tau_{IR}$</td>
<td>Drude scattering time at IR</td>
<td>30–50 fs</td>
</tr>
<tr>
<td>$\sigma_n$</td>
<td>Standard deviation of native carrier concentration due to spatial inhomogeneity</td>
<td>$1.5–2.5 \times 10^{12} \text{ cm}^{-2}$</td>
</tr>
<tr>
<td>$E_{Fc}$</td>
<td>Native Fermi level of graphene channel</td>
<td>$\approx 0.17 \text{ eV}$</td>
</tr>
<tr>
<td>$E_{Fg}$</td>
<td>Native Fermi level of graphene gates</td>
<td>$\approx 0.48 \text{ eV}$</td>
</tr>
<tr>
<td>$\tau_{eph}$</td>
<td>Electron-phonon cooling time</td>
<td>$\approx 50 \text{ ps}$</td>
</tr>
<tr>
<td>$\alpha_e$</td>
<td>Excess light attenuation within device</td>
<td>2–3 mm$^{-1}$</td>
</tr>
</tbody>
</table>

Figure 4-4: a) Comparison of the frequency response of the photodetector with that of the laser current modulation itself. The consistency between the two indicates that the photodetector frequency response exceeds 1 MHz. Inset: Simulated GHz-range photodetector frequency response, with and without considering the impact of the electron-phonon cooling time $\tau_{eph}$. b) Measured noise spectral density versus resistance and corresponding Johnson noise spectral density of Device B, without illumination, for the 49 pairs of gate voltages $\{V_{g1}, V_{g2}\}$ where each $V_{gn}$ is varied from $-6 \text{ V}$ to $6 \text{ V}$ in steps of $2 \text{ V}$. Measurement was performed at $T = 293 \text{ K}$.

4.2.4 Frequency response

To measure the device frequency response, I apply a sinusoidal signal to the current modulation input of our QCL source, which allows us to modulate the illumination intensity at frequencies up to 1 MHz. I then amplify the photovoltage by 40 dB and record the signal at the modulation frequency using an RF lock-in amplifier. To
account for the modulation response of the laser, I also record the photovoltage of a fast InAsSb photodiode illuminated by the same modulated QCL source light coupled through a single-mode waveguide on the chip with no devices on it. In all cases, I use a dwell time of 1.5 s, and the lock-in filter is set to 100 ms with a 12 dB/octave falloff. The comparison shown in Fig. 4-4a indicates that the device is faster than the laser’s modulation bandwidth. I thus use a COMSOL model to find the actual RC contribution to the device’s frequency response, plotted in the inset of Fig. 4-4a. I also plot the product of the RC-limited frequency response and the $\tau_{\text{eph}}$-limited frequency response with an assumed first-order $(1 + (2\pi\tau_{\text{eph}}f)^2)^{-0.5}$ dependence, which is valid as the electron-phonon cooling length $\ell = \sqrt{\kappa\tau_{\text{eph}}/C_{\text{el}}} \approx 230$ nm is narrower than the device channel[105]. I thus predict a 3-dB cutoff frequency of $f_{-3\text{dB}} \approx 1.3$ GHz, dominated by the capacitance between the graphene back-gates. One may improve upon this bandwidth reducing the overlap between the graphene contacts for the channel and gates, or by artificially increasing the contact resistance to the gates; for instance, by removing lengths of graphene at the gate-metal contact. In this case, the device bandwidth would increase to $f_{-3\text{dB}} = 1/(2\pi\tau_{\text{eph}}) \approx 3.2$ GHz.

4.2.5 Noise measurements

To investigate my device’s noise performance, I modulate the QCL current at 30 kHz, amplify the photovoltage by 60 dB with a low-noise preamplifier and inspect using a signal analyzer. I observe in Device A no broadening of the 30 kHz photoresponse peak at offset frequencies as low as 0.1 Hz, indicating long-term responsivity stability, and I observe no illumination-dependence of the noise floor. I then measure the un-illuminated noise spectral density and resistance versus both gate voltages. To do so, I feed the 60 dB-amplified device noise to an FFT signal analyzer and record the averaged power spectral density between 22 and 32 kHz (where I observe no electromagnetic interference-related peaks) while stepping through all pairs of gate voltages from $-6$ V to 6 V in steps of 2 V. Fig. 4-4b shows the resulting data for a Device B of identical design to Device A, organized by resistance and compared to the expected Johnson noise spectral density. I observe excellent consistency between the
measured and predicted noise, with a 2–4 dB discrepancy consistent with the specified noise figure of the preamplifier (specified as 1–3 dB over the measured range of input impedances) combined with ambient electromagnetic noise, corroborating my earlier claim of Johnson-noise-limited NEP.

4.2.6 Estimated gas sensing performance

To demonstrate the device’s utility, I analyze its predicted gas-sensing performance, summarized from Section 4.4.7. The minimum detectable gas concentration for a given waveguide platform and photodetector is given by:

\[ p_{\text{gas, min}} = e^{\frac{\alpha_{\text{base}} \cdot \text{NEP}}{a n_g \Gamma E I_0}}, \]  

where \( I_0 \) is the source power, \( \alpha_{\text{base}} \) is the attenuation coefficient of the gas-light interaction waveguide in the absence of gas, \( a \) is the specific attenuation coefficient of the gas, \( n_g \) is the guided mode group index, \( \Gamma E \) is the confinement factor of electric field energy within the gaseous medium, and \( e = \exp(1) \). For detection of nitric oxide (NO), with an absorption peak at \( \lambda = 5.24 \mu m \) and a specific attenuation of approximately \( a \approx 70 \text{ m}^{-1}\text{atm}^{-1} \) at low concentrations[186], I arrive at \( p_{\text{gas, min}} = 74 \mu \text{atm}/\sqrt{\text{Hz}} \) for a 1 mW illumination source. Assuming a measurement bandwidth of 0.1 Hz over which I have measured the photoresponse to be stable, I find \( p_{\text{gas, min}} = 23 \text{ ppm} \), roughly equal to the National Institute of Occupational Safety and Health (NIOSH) recommended exposure limit (REL) of 25 ppm[187]. Removing the slightly lossy HfO\(_2\) dielectric underneath the gas-light interaction waveguide could decrease \( p_{\text{gas, min}} \) considerably, as waveguide losses down to 0.7 dB/cm have been demonstrated at the same wavelength using a similar chalcogenide glass and liftoff process[188].

4.3 Discussion

Although my demonstration is limited to \( \lambda = 5.2 \mu m \) by light source availability, the optical conductivity of my as-transferred graphene inferred from the fitting parame-
HgCdTe PD, $\lambda_{\text{opt}} = 5.0\, \mu m$

HgCdTe PD, $\lambda_{\text{opt}} = 10.6\, \mu m$

VO$_x$ bolometer

This work

<table>
<thead>
<tr>
<th>NEP</th>
<th>HgCdTe PD $\lambda_{\text{opt}} = 5.0, \mu m$</th>
<th>HgCdTe PD $\lambda_{\text{opt}} = 10.6, \mu m$</th>
<th>VO$_x$ bolometer</th>
<th>This work</th>
</tr>
</thead>
<tbody>
<tr>
<td>[pW/$\sqrt{\text{Hz}}$]</td>
<td>1, $\lambda = 5.2, \mu m$</td>
<td>10, $\lambda = 5.2, \mu m$</td>
<td>0.9</td>
<td>1100</td>
</tr>
<tr>
<td>$f_{-3\text{dB}}$ [MHz]</td>
<td>1.3</td>
<td>106</td>
<td>0.00001 (pred.)</td>
<td></td>
</tr>
<tr>
<td>Vacuum required?</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Waveguide-integrated?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 4.2: Comparison of the detector with inferred room-temperature performance metrics for two HgCdTe photodiodes optimized for two different wavelengths (from [14]) and a VO$_x$ bolometer (from [2]) available off the shelf. For the photodiodes, the NEP is extrapolated from the specified detectivity for a detector scaled down to match the size of a diffraction-limited spot with $\text{NA} = 0.3$, which is the acceptance NA of these detectors. For the bolometer, I give the NEP of a single $17\, \mu m \times 17\, \mu m$ bolometer pixel as calculated from the specified noise-equivalent temperature difference as described in Rogalski[15].

ters in Table 4.1 remains relatively constant and even increases at longer wavelengths due to intraband absorption as shown in Fig. 4-11. I thus expect this platform to scale to $\lambda = 10\, \mu m$ and beyond, perhaps requiring a BaF$_2$ substrate for extended transparency, with little reduction in performance owing to the PTE effect’s thermal nature. In Table 4.2 I compare my device’s performance with various off-the-shelf detectors. Although its NEP is not yet on par with commercial options, its predicted bandwidth may be useful for DCS-based integrated gas analyzers[51]. Additionally, the vacuum requirement of VO$_x$ bolometers may complicate co-packaging and introduce coupling losses, and the high cost of HgCdTe may preclude use in broadly deployed sensor networks. It is important to point out that development of sensitive waveguide-integrated mid-IR photodetectors is not the only challenge that much be tackled to develop integrated mid-IR gas sensing and spectroscopy systems. As is readily apparent in Eqn. 4.1, not only must NEP be low; so must $\alpha_{\text{base}} n_g^{-1} \Gamma_E^{-1}$, and $I_0$ must be high. The former pertains to the gas-light interaction waveguide, which must have low loss and feature a high gas/field overlap. The latter pertains to the light source, which must be bright. A thermal, Planckian emitter will not work for the
light source, as the optical power that may be thermally emitted into a single optical mode is fundamentally limited for a given source temperature[30]. Therefore, further research is required on all three parts of the system to enable compact and sensitive gas-sensing systems—especially spectroscopic systems which inherently suffer from reduced signal-to-noise ratios due to the multiplicity of spectral channels[51]—and the simple multiplicative relation of Eqn. 4.1 implies that the source, waveguide and detector are all equally important.

In conclusion, I have demonstrated a PTE-based graphene photodetector, integrated in a scalable chalcogenide glass waveguide platform with an NEP of $1.1 \text{nW/Hz}^{1/2}$ and a bandwidth exceeding $f_{-3\text{dB}} = 1 \text{MHz}$. I have modeled the bandwidth to approach $1.3 \text{GHz}$ and I predict similar performance at longer wavelengths for scaled-up devices enabled by the transparency of GSSe beyond $\lambda = 10 \mu\text{m}[189]$. Finally, I have shown that my device and waveguide platform would enable NO detection at concentrations comparable to its REL. Substantial improvements are likely by shrinking the optical mode to the cooling length $\ell$ using metal-insulator-metal plasmonic waveguides[115] or dielectric slot waveguides, which would improve the utilization of absorbed light[105]. The PIC platform further promises to support a full toolkit of mid-IR active devices that may even leverage the same graphene material platform for devices such as modulators[190] or high-intensity hot-electron-based light sources[191]. This research represents the first foray into waveguide-integrated detectors operating beyond $\lambda = 4 \mu\text{m}$, paving the way towards 2D-material-enabled integrated mid-IR microsystems for gas sensing, spectroscopy[51] and free-space optical communications[29].

4.4 Methods

4.4.1 Photodetector fabrication

A continuous monolayer graphene film was grown on Cu foil (99.8%, Alfa Aesar, annealed, uncoated, item no. 46365) cut to a size of $15 \times 2 \text{ cm}^2$ in a 1-inch-diameter
quartz tube furnace under atmospheric pressure. The furnace was heated to 1060°C over 30 minutes under 500 sccm of Ar flow; afterwards, 15 sccm of H₂ and 10 sccm of dilute CH₄ (1% in Ar) were introduced as reducing gas and carbon source, respectively, and flowed for 4 hours to ensure the continuity of the graphene film. Finally, the furnace was allowed to cool to 100°C without modifying the gas flow before the CVD graphene was removed from the chamber. The devices were fabricated on a 1" diameter by 1.0 mm thick (111)-cut CaF₂ substrate (MTI Corporation, item CFc25D10C2). I first coated the substrate with a PMMA bilayer for liftoff (495 PMMA A6 followed by 950 PMMA A2), which features a slightly re-entrant sidewall profile after developing. I then performed e-beam lithography using an Elionix FLS-125 125 keV electron beam lithography system to pattern alignment marks on the substrate, followed by room-temperature development in 3:1 isopropanol:methyl isobutyl ketone, e-beam evaporation of 5 nm Ti/100 nm Au (Temescal VES2550), and liftoff. To transfer the first layer of graphene, I first coated one side of the CVD graphene-on-Cu sheet with PMMA and removed the graphene from the other side using 90 seconds of oxygen RIE (16 sccm He and 8 sccm O₂ at a pressure of 10 mTorr and an RF power of 100 W, “oxygen RIE process”). I then etched away the Cu using a FeCl₃-based etchant, followed by 2 DI water rinses, a 30-minute clean in 5:1 DI water:HCl 37% in water to reduce metal ion contamination, and two more DI water rinses. After letting the graphene film sit overnight in the final evaporating dish of water, I scooped it out with the CaF₂ substrate, blew N₂ on the film to eliminate most of the trapped water, and then baked the sample at 80°C for 30 minutes followed by 160°C for 2 hours. I then removed the PMMA from the graphene using acetone at room temperature, rinsed it in isopropanol and blew it dry, and baked the sample in N₂ for 1 hour to improve adhesion. To pattern the graphene back-gates, I spun on a layer of 950 PMMA A6, patterned the gates in the Elionix and developed as described previously, etched away the exposed graphene using “oxygen RIE process”, and removed the PMMA as described previously. I then repeated the Ti/Au liftoff process described above to define the contacts to the graphene gates, after which I evaporated 1.5 nm Al (Temescal VES2550) as an ALD seed layer, allowed the thin
Al layer to oxide in air, and deposited 300 cycles $\approx 30$ nm of HfO$_2$ ALD at 200°C (Cambridge Nanotech Savannah 200). To define the graphene channel and channel contacts, I performed another graphene transfer as described above and repeated the subsequent graphene patterning and contact metallization steps, followed by another Al seed layer and 150 cycles of HfO$_2$ ALD to protect the graphene channel. Finally, to pattern the GSSe waveguides, I coated the chip with 495 PMMA A11, used the Elionix to define the waveguides, developed as described previously and evaporated 750 nm of Ge$_{28}$Sb$_{12}$Se$_{60}$, followed by a quick liftoff in boiling acetone, IPA rinse and N$_2$ blow-dry, and cleaving of the chip to expose waveguide facets.

4.4.2 Optical setup, alignment, power and loss calibration

Fig. 4-5a depicts the optical beam path during all optical and optoelectronic measurements. Chopped, collimated illumination at a wavelength of $\lambda = 5.2 \mu$m is coupled into a cleaved waveguide facet at the edge of the chip using a molded aspheric focusing lens to achieve a diffraction-limited spot. The waveguides containing the photodetectors and modulators as well as the kickback waveguides for loss measurement are designed to exit the chip at a 5 mm offset with respect to the input to reduce the amount of stray light picked up by the collection optics. The collection optics consists of a 1" high-NA germanium asphere which images the output facet onto the center of an iris, behind which I place an InAsSb photodiode that monitors the out-coupled power. To align the setup, I flip the Ge collection lens, iris and photodiode out of the output beam path and use a long focal length CaF$_2$ lens and liquid nitrogen-cooled InAsSb camera to image the output facet of the chip while adjusting the chip position to achieve coupling first through a straight multimode waveguide (not shown) followed the desired S-shaped device or kickback waveguide. I then replace the long focal length lens with the high-NA germanium collection lens, flip the iris back into the collection beam path in the “open” position, and place the camera behind the iris. By adjusting the position of the collection lens to focus the out-coupled light at the center of the iris while gradually reducing the iris aperture size and monitoring the focal point on the camera, I am able to localize the focus of the collection lens at
Figure 4-5: a) Depiction of the in- and out-coupling beam path for optical and optoelectronic measurements. Lengths are not to scale; the chip and waveguide are magnified for clarity. b) Output photodiode signal versus waveguide length for four waveguide kickback structures. The decaying exponential fit reveals the waveguide loss and the in-coupled power, subject to a conversion efficiency accounting for the collection efficiency of the collection optics and the voltage responsivity of the photodiode. Inset: Layout of the kickback structures. The longest of the five kickbacks was not included in the fit as the out-coupled signal was indistinguishable from stray light.
the center of the iris. Finally, I replace the camera with the InAsSb photodiode and adjust the in-coupling and out-coupling optics to maximize the signal measured by the photodiode.

To determine the loss of my waveguides and predict the optical power immediately incident upon the graphene photodetector during characterization, I have fabricated a set of five waveguide “kickback” structures of varying length, the layout of which is shown in the inset of Fig. 4-5b. A large radius of 75 µm is used for the waveguide turns to reduce the associated light leakage to negligible levels. I individually align and optimize the coupling for each of these waveguides and record the lock-in signal of the collection photodiode; these data are plotted versus the total kickback waveguide length in Fig. 4-5b. I could not collect a data point for the longest kickback because I found the out-coupled light to be indistinguishable from stray illumination. Fitting the data to a decaying exponential reveals a waveguide loss of 1.1 dB/mm and an in-coupled power corresponding to a lock-in signal of 46.2 mV. To obtain the actual in-coupled optical power, I must divide this value by the product of the voltage responsivity of the photodiode and the efficiency of the collection optics. For the former, I obtain a value of $1.76 \times 10^3$ V/W from the manufacturer-provided calibration data. For the latter, I perform a full-wave electromagnetic simulation of the waveguide facet and extract the far-field profile of out-coupled light. Integrating the simulated optical power falling within the entrance pupil of the collection optics, I obtain a collection efficiency of 0.45. Combining the above three figures, I arrive at 58 mW coupled into the waveguide facet at optimal alignment. Accounting for waveguide loss, I obtain a power of 11 mW immediately incident upon the measured photodetector, from which I can thus calculate the responsivity figures reported in Section 4.2.3.

### 4.4.3 Device modelling

I use Eqn. 2.1 to model graphene’s conductivity at DC and infrared frequencies, albeit with different values of the Drude scattering time $\tau$ for the different frequency ranges. As I will show below, graphene’s low frequency conductivity $\sigma_{\text{DC}}$ and infrared conductivity $\sigma_{\text{IR}}$ affect various intermediate model parameters; $\sigma_{\text{DC}}$ and $\sigma_{\text{IR}}$
themselves depend strongly on $E_F$, which features spatial variation due to the back-gates. For the graphene channel, I assume a constant $N_c = N_{0,c} + e^{-1}C_g V_g$ in the region above each gate, where $N_c$ is the carrier concentration in the channel (positive for positive $E_F$, negative for negative $E_F$), $N_{0,c}$ is the native carrier concentration at zero gate voltage, $C_g$ is the capacitance per area of the gate dielectric, and $V_g$ is the voltage applied to the gate in question. (Using a set of test devices, I measure $C_g = 34 \text{ fF/µm}^2$ on my chip, corresponding to a back-gate dielectric constant of $K \approx 12$; this is described in more depth in Section 4.4.5.) In the part of the graphene channel above the gap between the two gates, I assume a linear slope between $N_c, 1$ and $N_c, 2$. For the gates, $N_g = N_{0,g} - e^{-1}C_g V_g$, with $N_g$ and $N_{0,g}$ defined similarly to $N_c$ and $N_{0,c}$. In general, the graphene’s Fermi level and carrier concentration are related by $E_F = \hbar v_{gr} \sqrt{\pi |N| \text{sign}(N)}$, where $v_{gr}$ is graphene’s Fermi velocity. To incorporate the blurring of the graphene’s Fermi level-dependent properties due to spatial carrier concentration variations, I convolve the Kubo formula with a Gaussian as follows:

$$\sigma_{DC}(N) = \frac{1}{\sigma_n \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(n-N)^2}{2\sigma_n^2}} \sigma(0, E_F(N), \tau_{DC}, T_0) \, dn \quad (4.2)$$

and similarly for $\sigma_{IR}(N)$ using $\omega = 2\pi c/\lambda$ instead of 0 and $\tau_{IR}$ instead of $\tau_{DC}$. Finally, $R = \sigma_{DC}^{-1}$, $\kappa = \pi^2 k_B^2 T_0 \sigma_{DC}/3e^2$ via the Wiedemann-Franz law, and $S = -d(\log \sigma_{DC})/dE_F$ [109]. $C_{el}$ is obtained by convolving the heat capacity of pristine graphene with a Gaussian of standard deviation $\sigma_N$ as in Eqn. 4.2, where the pristine heat capacity is given by [109, 101]:

$$C_{el}(N)|_{\sigma_n=0} = \int_{-\infty}^{\infty} \varepsilon \frac{2|\varepsilon|}{\pi \hbar^2 v_{gr}^2} \frac{\partial f_d(\varepsilon - E_F(N))}{\partial T} \, d\varepsilon. \quad (4.3)$$

I use a waveguide eigenmode solver to find the mode profile of my waveguide design at $\lambda = 5.2 \text{ µm}$, using refractive indices of 1.4, 2.6, and 1.88 for the CaF$_2$, GSSe, and HfO$_2$, respectively. The resulting mode profile enters into the expression for $\dot{Q}$ as follows [90]:

$$\frac{\dot{Q}}{P} = \frac{(|E_x(x, y_c)|^2 + |E_y(x, y_c)|^2) \sigma_{IR,c}(x)}{\iint_{g^2} \text{Re}(E \times H^*) \cdot \hat{z} \, dx \, dy}. \quad (4.4)$$
Here $y_c$ is the $y$-coordinate of the graphene channel, and $y_g$ would be the $y$-coordinate of the graphene gates. One may then write $\alpha_c = P^{-1} \int_{-W/2}^{W/2} \dot{Q}(x) \, dx$. Similar expressions hold for $\alpha_g$ in terms of $\sigma_{IR,g}(x)$, noting of course that $\sigma_{IR,g}(x) = 0$ for $x$ within the gap between the gates where there is no graphene. Finally, $\rho_\Omega = \int_{-W/2}^{W/2} R(x) \, dx$.

Having thus obtained expressions for $\kappa(x)$, $C_{el}(x)$, $\dot{Q}(x)$, $S(x)$, $\Pi(x)$, $\alpha_c$, $\alpha_g$ and $\rho_\Omega$ as a function of the gate voltages as well as $\tau_{DC}$, $\tau_{IR}$, $\sigma_n$, $E_F$, $E_F$, $\tau_{eph}$, $\alpha_e$, and $\rho_c$, I then solve for the increase in electronic temperature per guided power $\Delta T_{el}(x)/P = (T_{el}(x) - T_0)/P$ using the equation and effective one-dimensional version of Eqn. 2.2:

$$-\frac{d}{dx} \left( \kappa \frac{d \Delta T_{el}}{dx} \right) + \tau_{eph}^{-1} C_{el} \Delta T_{el} = \eta \dot{Q} - J_x \frac{d \Pi}{dx},$$

(4.5)

Here, I am approximating the electric field to run exclusively in the $x$-direction, valid for sufficiently gradual light absorption. I assume $\eta = 1$ as has been previously reported in pump-probe experiments at this wavelength range[106]. Integrating Eqn. 2.3 between the source and drain with $J_x = 0$, one obtains the thermal electromotive force (EMF) arising from the Seebeck effect:

$$\mathcal{E}_x = -\int_{-W/2}^{W/2} S \frac{d \Delta T_{el}}{dx} \, dx,$$

(4.6)

where $W = 5.4 \mu m$ is the channel width. In Eqns. 4.5 and 4.6, $\kappa$, $C_{el}$, $S$, and $\Pi = S T_{el} \approx S T_0$ (for small $\Delta T_{el}$) are all dependent on the local Fermi level $E_F$ of the graphene, and thus have a gate-tunable $x$-dependence, which I account for in my calculations. Combining the equations, the $\eta \dot{Q}$ source term in Eqn. 4.5 gives rise to a proportional photo-induced EMF, whereas the Peltier term $J_x \frac{d \Pi}{dx}$ gives rise to a current-dependent EMF which appears as a resistance in series with the Ohmic and contact resistances of the channel. One can thus write:

$$V = \mathcal{R}_v \alpha_c P(z) - (\rho_\Omega + \rho_\Pi + \rho_c) J_x(z)$$

(4.7)

where $V$ is the voltage across the contacts, $\mathcal{R}_v$ is the photovoltage per absorbed power
per length of a cross-sectional slice of the device (i.e., dimensions of $V/(W/m)$), $\alpha_c$ is the component of the waveguide power attenuation coefficient arising from absorption in the graphene channel, $P(z)$ is the guided power at a position along the waveguide, and $\rho_\Omega$, $\rho_\Pi$, $\rho_c$ are the Ohmic, Peltier, and contact line resistivities (dimensions of $\Omega \cdot m$), respectively. Averaging over $z$ along the length of the waveguide one obtains:

$$V = \frac{R_c}{L \alpha_{tot}} \alpha_c \left(1 - e^{-\alpha_{tot}L}\right) P_{in} - (R_\Omega + R_\Pi + R_c) I,$$

(4.8)

where $I$ is the current produced by the photodetector, thus describing a Thévenin equivalent source. Here, $\alpha_{tot} = \alpha_c + \alpha_g + \alpha_e$ is the total guided power attenuation coefficient within the detector, including contributions not only from the graphene channel but also from the graphene gates ($\alpha_g$) as well as a gate-independent excess loss $\alpha_e$ due to scattering and any other light absorption that may occur along the channel. Thus the total device resistance is equal to $R = R_\Omega + R_\Pi + R_c$, and the voltage responsivity is given by:

$$R_v = \frac{R_c}{L \alpha_{tot}} \alpha_c \left(1 - e^{-\alpha_{tot}L}\right),$$

(4.9)

which I plot versus both gate voltages in Fig. 4-3b for the best-fit device parameters given in Table 4.1 obtained as described in Section 4.4.4. All calculations are carried out in Mathematica.

### 4.4.4 Model parameter extraction and modelled resistance and transmittance

Fig. 4-6 compares the measured and modelled resistance and transmittance gate maps obtained using the mean graphene quality and waveguide loss parameters listed in Table 4.1; namely, $\tau_{DC} = 3.5 \text{fs}$, $\tau_{IR} = 40 \text{fs}$, $\sigma_n = 2 \times 10^{12} \text{cm}^{-2}$, and $\alpha_e = 2.5 \text{mm}^{-1}$. I determined the values of these parameters to achieve the best fit simultaneously between both pairs of maps. Generally speaking, $\tau_{DC}$ is inversely proportional to $R_\Omega$, $\tau_{IR}$ affects the scale and modulation contrast of the transmittance $\mathcal{T} = e^{-\alpha_{tot}L}$.
and $\sigma_N$ affects the sharpness (width at half-max) and contrast of both. First, I infer $E_{F_c}$ simply from from the gate voltage of the charge neutral point (peak in the case of $R$ and dip in the case of $T$). Since the sharpnesses of both $R$ and $T$ are largely determined by $\sigma_N$ for relatively high $\sigma_N$ as is the case for these devices, I then determine $\sigma_N$ to best match both maps. In particular, I find that the resistance map is best fit by $\sigma_n = 1.5 \times 10^{12} \text{ cm}^{-2}$ and the transmittance map by $\sigma_n = 2.5 \times 10^{12} \text{ cm}^{-2}$, from which I obtain the error margins quoted for $\sigma_n$ in Table 4.1; $\sigma_n = 2 \times 10^{12} \text{ cm}^{-2}$ represents a compromise between these two values. I then determine $\tau_{DC}$ to roughly match the scale of $R$, but I allow the modelled resistance to be on the order of $10 \Omega$ less the measured resistance to take into account the possibility of a contact resistance.
in this range as justified in Section 4.4.6. The uncertainty in the actual contact resistance, as well as the imperfect fit, both contribute to uncertainty in the actual value of \( \tau_{\text{DC}} \). I determine \( \tau_{\text{IR}} \) and \( E_{\text{Fg}} \) to best match the measured transmittance map, and the values quoted in Table 4.1 represent a best attempt to simultaneously reflect multiple features of this map; namely, the modulation contrast between the center and corners of the map, the contrast between the upper right and other corners, and the falloff at the low-voltage edges of the map (where intraband absorption of the graphene gates is strongest) and the high-voltage edges of the map (where interband absorption of the graphene gates is strongest). The quoted error margins of \( \tau_{\text{IR}} \) reflect the range over which these different features of the transmittance map are best rendered in my model, and the error margins of \( \alpha_e \) reflect the range required to match the scale of the measured transmittance map over the error range of \( \tau_{\text{IR}} \). I finally use these six parameters to predict the relative gate dependence of the voltage responsivity. In this way, only the overall scale factor of the responsivity is subject to a fitting parameter (namely, \( \tau_{\text{eph}} \)); the contour of the responsivity map is, in this device’s performance regime, purely predicted from parameters extracted from the resistance and transmittance maps. Therefore, the resemblance between the measured and modelled responsivity maps is not the result of tweaking parameters to achieve
4.4.5 Gate capacitance measurement

I used Transmission Line Method (TLM) devices to measure both the gate capacitance and contact resistance of my devices. Shown in Fig. 4-7, these devices consist of a large, contacted graphene back gate, gating a set of graphene FETs of increasing channel lengths, ranging from 5 µm to 65 µm. To extract the capacitance per area of the hafnia dielectric, I measure the capacitance between the lower graphene gate and the upper graphene/gold structure, and divide by the overlap area of these two structures. To measure the capacitance, I use a lock-in amplifier to apply a 0.5 VRMS sinusoid of variable frequency to the rightmost two metal contacts in Fig. 4-7, and in the current return path I place a 10 kΩ shunt resistor, the voltage across which I monitor with the lock-in amplifier. At each frequency and for each TLM that I measure, I perform a measurement with both pads contacted and with only one pad contacted to compensate for any stray capacitance in the setup. I then subtract the measured capacitances in the “connected” and “disconnected” cases to obtain the actual device capacitance as a function of frequency, which I plot in Fig.
4-8. Excluding the orange curve as an outlier, I measure a capacitance of $C \approx 30 \text{ pF}$ corresponding to a capacitance per area of $C_g = 3.4 \text{ fF/µm}^2$ and a dielectric constant of $K \approx 12$.

4.4.6 Measurement of TLM structure for gate resistance extraction

In additional to capacitance measurement, I also use the TLM structures for their intended purpose of evaluating the resistance of the graphene-metal contacts. For each of the six channels in each of the five TLM devices, I measure the channel resistance as a function of gate voltage using an upward voltage sweep each time to compensate for the hysteresis discussed in Section 4.5.1. Since the gate voltage of the charge neutral point may shift slightly between measurements due to trapped charges in the gate, I then shift the measured resistance curves so that their peaks overlap. Finally, for each measured voltage offset from the Dirac peak and for each TLM, I fit the data of resistance versus channel length to a line and plot the resulting y-intercept as a function of voltage offset, manually eliminating any gate voltage sweeps showing malformed (for instance, flattened or bimodal) resistance peaks. The resulting y-intercept curves are shown in Fig. 4-9. Unfortunately, I find highly inconsistent intercept resistances between the five TLM devices, with the intercept even going negative in several cases. Therefore, I am unable to draw a quantitative conclusion regarding the contact resistance of the devices. I can at least, however, estimate the total contact resistance (summing over both contacts) for the TLM devices to be generally in the $\approx 1 \times 10^2 \text{ Ω}$ range; therefore, since the TLM channels are 40 µm wide, I would expect the total contact resistance of the actual photodetectors to be in the $\approx 1 \times 10^1 \text{ Ω}$ range, which is an order of magnitude lower than my measured resistances; therefore I conclude that it can be safely ignored in my modelling as other sources of error (such as the imperfect fit between the measured and modelled resistance and transmittance maps) are much more likely to dominate the uncertainty in my analysis.
Figure 4-9: Y-intercepts of resistance versus channel length as a function of gate voltage offset from the resistance peak for the five TLM devices.

4.4.7 Calculation of gas sensitivity

My gas sensitivity analysis follows that of Siebert at el.[65], which considers a light source at the gas absorption peak illuminating a long gas-light interaction waveguide, partially cladded with the ambient air in which the gas’s presence is suspected, and terminated in a noisy photodetector. However, the authors do not give a rigorous analysis of the attenuation coefficient of guided light due to the absorbing gas, which I supply here. The power attenuation coefficient in a waveguide with a perturbative source of absorption is given by [90]:

\[
\alpha_{\text{wg}} = \omega_0 \frac{\int\int_{\mathbb{R}^2} 2\kappa |\mathbf{E}|^2 \, dx \, dy}{\int\int_{\mathbb{R}^2} \text{Re}((\mathbf{E} \times \mathbf{H}^*) \cdot \mathbf{\hat{z}}) \, dx \, dy},
\]

(4.10)

where the complex refractive index \( \tilde{n} \equiv n - j\kappa \). For a partial pressure of target gas \( p_{\text{gas}} \), the free space attenuation coefficient is:

\[
\alpha_{\text{free-space}} = a p_{\text{gas}} = 2\kappa \omega_0 / c,
\]

(4.11)
therefore $\kappa = a_p \text{gas}/(2\omega_0)$, and thus

$$\alpha_{\text{gas}} = a_p \text{gas} c \frac{\iint_{\mathbb{R}^2} n_0 |E|^2 \, dx \, dy}{\iint_{\mathbb{R}^2} \text{Re}(E \times H^*) \cdot \hat{z} \, dx \, dy}, \quad (4.12)$$

where $n_0 \approx 1$ is the refractive index of the gaseous medium, as $\kappa$ is only nonzero in the gaseous region of the waveguide cross-section. Since the waveguide mode is absorbed only very weakly, I now reference the expressions for group velocity $v_g$ and propagation constant $\beta$, respectively, given in Snyder and Love for non-absorbing waveguides[90]:

$$v_g = \frac{c^2 \beta}{\omega_0} \frac{\iint_{\mathbb{R}^2} \text{Re}(E \times H^*) \cdot \hat{z} \, dx \, dy}{\iint_{\mathbb{R}^2} n^2 \text{Re}(E \times H^*) \, dx \, dy}, \quad (4.13)$$

and

$$\beta = \frac{\omega_0}{c^2} \frac{\iint_{\mathbb{R}^2} n^2 \text{Re}(E \times H^*) \, dx \, dy}{\iint_{\mathbb{R}^2} n^2 |E|^2 \, dx \, dy}. \quad (4.14)$$

Combining these two equations, the group index $n_g$ is thus:

$$n_g = \frac{c}{v_g} = c \frac{\iint_{\mathbb{R}^2} n^2 |E|^2 \, dx \, dy}{\iint_{\mathbb{R}^2} \text{Re}(E \times H^*) \cdot \hat{z} \, dx \, dy}. \quad (4.15)$$

By comparison with Eqn. 4.12, one finally arrives at:

$$\alpha_{\text{gas}} = a_p \text{gas} n_g \frac{\iint_{\mathbb{R}^2} n_0 |E|^2 \, dx \, dy}{\iint_{\mathbb{R}^2} n^2 |E|^2 \, dx \, dy} = a n_g n_0^{-1} \Gamma_E \rho_{\text{gas}}, \quad (4.16)$$

where $\Gamma_E$ is the electric field confinement factor:

$$\Gamma_E \equiv \frac{\iint_{\mathbb{R}^2} n_0^2 |E|^2 \, dx \, dy}{\iint_{\mathbb{R}^2} n^2 |E|^2 \, dx \, dy}. \quad (4.17)$$

I emphasize that this is not, in general, equal to the “traditional” confinement factor $\Gamma$ defined based on the proportion of modal Poynting vector in a gain/loss region, as has been previously noted[192].

The remainder of my gas sensitivity analysis follows that of Siebert et al.[65], in which an ideal gas-light interaction wavelength is found as a function of the gas concentration where maximum sensitivity to deviations is desired. For gas detection
4.5 Additional information

4.5.1 Modulators and hysteresis

In addition to photodetector devices, I also fabricated gated graphene modulator devices as illustrated in Fig. 4-10a. These are similar to the detector devices described in Section 4.2.1, except with only a single graphene sheet and a single contact on each graphene layer. Indeed, the detector devices can be made to behave similarly by applying the same voltage to each of the back-gates, but I nevertheless report the modulator devices here for completion as well as to illustrate the observed hysteresis effect. Fig. 4-10b shows the transmission response versus applied voltage for one of the modulator devices. Applying a zigzag voltage sweep reveals an undesirable hysteresis pattern showing distinct curves for rising and falling voltages sweeps, which I label red and blue respectively. The same effect is observed in the photodetector devices as well. Hysteresis has been widely reported for as-deposited HfO$_2$ gates, and
Figure 4-11: Contour plot of real part of graphene’s infrared optical conductivity as a function wavelength and mean carrier concentration, assuming a carrier concentration spread of $\sigma_n = 2.0 \times 10^{12}$ cm$^{-2}$ and a Drude scattering time of $\tau_{\text{IR}} = 40$ fs.

is attributed to trapped charge carriers within the dielectric\cite{193, 194}. To prevent the effects of hysteresis from appearing in the gate sweeps presented in Fig. 4-2, I “reset” both gate voltages to 6.5 V prior to collecting each data point.

### 4.5.2 Optical absorption of graphene for longer wavelengths

In Fig. 4-11, I apply Eqn. 4.2 to predict how the real part of the as-transferred graphene’s optical conductivity (and thus optical absorption) would change at longer wavelengths using the carrier concentration spread and Drude scattering time material parameters extracted from the device model. I find that the optical absorption remains roughly the same if not even higher at longer wavelengths, which is due to increased intraband absorption; therefore I anticipate similar device performance can be achieved at these wavelengths.
Chapter 5

Discussion and outlook

5.1 Graphene photodetection in general

Before discussing specific opportunities to build upon my research, I will first examine the general prospects for improvement of graphene photodetectors, highlighting the applicability of such developments to my research.

5.1.1 Plasmonic enhancements to device performance

In the context of graphene optoelectronics, “plasmonic enhancements” may mean one of two things: It can refer to concentration of electromagnetic fields via metallic structures[115], or it can refer to enhancements in the photoresponse of graphene devices arising from plasmonic resonances in the graphene itself, typically engineered by periodic removal or doping of the graphene to enable localization of what would otherwise be propagating graphene plasmon modes[48, 195, 196]. As I allude to in Section 4.3, metal plasmonics will allow the size of the optical mode to be reduced to the graphene cooling length $\ell$, improving the utilization efficiency of absorbed light. In fact, this is an oversimplification of the situation, and the reality is even more optimistic. The full range of benefits of metal-plasmonic mode concentration is apparent in Eqn. 2.6 describing the NEP of a uniformly-illuminated PTE-based detector, which approximates the case of a realistic dielectric or plasmonic waveguide-
integrated detector where the source and drain contacts are placed as close as possible to the waveguide mode to minimize $L$, as typically $L \gg \ell$. The responsivity and thus NEP benefit is reflected in the $(L/2\ell^2)(1+\coth^2(L/4\ell))$ term, which describes how the temperature at the $pn$-junction varies with $L$: Most radiation absorbed more than about a distance $\ell$ from the junction dissipates due to electron-phonon scattering before it can diffuse to the junction and contribute to photoresponse, and this shows up as a factor of $L$ in the NEP equation, as the proportion of radiation absorbed out of reach of the junction grows in proportion to $L$. There is also, however, the $\sqrt{LW} = \sqrt{A}$ term, which reflects the reality that more graphene means more noise. Applying metal plasmonics not only permits lower $L$, but in the case of waveguide-integrated detectors incorporating metal-plasmonic guided modes, also lowers $W$ due the enhanced attenuation coefficient due to graphene absorption of light. This can be seen via Eqn. 4.4; if a waveguide mode is uniformly shrunk by a factor of $s$ while maintaining a given $P$, the fields must be amplified by a factor of $s$; therefore $\dot{Q}$ increases as $s^2$ and $\alpha_c$ as $s$, allowing reduction of $W$ by a factor of $s$. Therefore, in theory, the overall reduction in NEP using plasmonic waveguide modes could be as much as $s^2$, potentially up to two orders of magnitude. While I take advantage of the benefits of metal plasmonics for the metasurfaces I simulate in Chapter 3, demonstration of metal-plasmonic waveguide-integrated graphene detectors in the mid-IR is a prime area of exploration, building upon similar demonstrations at near-IR wavelengths[115, 197].

Besides metal plasmonics, graphene plasmonic effects have emerged as another means to improve the sensitivity of graphene detectors. Plasmon propagation has been observed in unpatterned graphene[198], and graphene plasmons can be coupled to radiating modes using periodic perturbations such as removal[163] or doping[199] of the graphene sheet. Two different mechanisms enable the potential benefits of graphene plasmonics. First, plasmonic resonances in periodically patterned graphene are associated with a dramatic increase in optical conductivity on resonance[163], which can improve absorption efficiency in free-space incident devices. For waveguide-integrated devices, this enhanced conductivity could be leveraged to concentrate
light absorption at a Seebeck coefficient step in a PTE device (for instance, using a $p$/plasmonic $n+/n$ device design) and to enhance absorption and reduce the necessary waveguide length, both enabling higher NEP. Second, graphene plasmonic effects have been found to enable very high responsivity mid-infrared photodetection based on a plasmon-to-hot carrier decay pathway[48]. This response mechanism does not yet seem to be fully understood, but the performance of reported devices encourages further investigation.

5.1.2 Bilayer graphene as an active material

Bilayer graphene has been shown to exhibit high Seebeck coefficients[200] and slow carrier-phonon scattering times[201], both key components for a strong PTE response, and indeed exceptional sensitivity has been demonstrated in bilayer graphene PTE detectors for terahertz radiation[202]. Compared to monolayer graphene, bilayer graphene has a potential advantage due to the appearance of a narrow bandgap under an applied electric field[203]. Within the regime of unbiased PTE operation, a bandgap may reduce achievable NEPs by reducing conductivity, referring to Eqn. 2.7. More importantly, however, the decrease in conductivity associated with a bandgap could allow biasing of bilayer graphene photodetectors without introducing excessive electron shot noise, permitting photovoltaic device operation in which electric field drift is applied to direct hot carriers to a $pn$-junction for readout. The noise performance of such devices may also benefit from the slow carrier-phonon scattering in bilayer graphene, which in principle would reduce thermal carrier generation rates and thus reduce noise. To my knowledge, this behavior has not been conclusively demonstrated, but the theoretical benefits motivate future research.

5.1.3 Improvement of graphene processing technology

Since the beginning of this research, graphene processing technology has progressed considerably. Deterministic and scalable CVD growth and placement of graphene single crystals has been demonstrated, yielding as-transferred graphene with exception-
ally high mobility and Seebeck coefficients up to $|S| = 140 \mu V/K$ even after coating with a polymer dielectric[118]. Such high Seebeck coefficients yield proportionally high responsivities as is evident from Eqn. 2.3. Improvements in graphene quality may also enhance the aforementioned light utilization of PTE devices by increasing $\tau_{eph}$ and $\kappa$ and thus $\ell$, which is beneficial in circumstances where reducing the channel length to $L \approx 2\ell$ is not feasible. While incorporating such novel graphene processing technology into mid-IR devices could enable considerable improvements in material quality and thus performance, maintaining transparency in the mid-IR remains an area for further investigation, as the graphene cladding materials used in such demonstrations are not mid-IR-transparent. Chalcogenide glasses are a promising cladding and dielectric option that have been shown to preserve the Raman spectra and electrical properties of graphene[190], and merit further investigation. Improved ALD processes using seeding approaches that reduce damage to the graphene may also be investigated; this has been researched extensively but not conclusively resolved[204]. Future work towards damage-free ALD cladding of graphene for mid-IR applications may focus on oxides with established mid-IR transparency, such as bismuth oxide or yttrium oxide[205].

5.2 Multispectral imaging metasurfaces

5.2.1 What could we learn from an experimental demonstration?

The lack of a clear experimental demonstration of broadband, multispectral absorption in graphene integrated on slot antenna metasurfaces limits the range of conclusions that can be drawn from the research described in Chapter 3. What I do purport to demonstrate is twofold. First, I claim to have established that slot antennas are an efficient way to couple normally incident radiation into optically active materials with poor absorption, and that this behavior is well-understood and can be described analytically. Second, I offer the observation that such antennas can be
assembled into a multispectral metasurface with high absorption because they have small footprints and make use of the downward direction for resonant energy storage. However, multiple questions remain. Although one might be able to suggest multiple ways to fabricate the proposed structures—for instance, the approach put forward in Section 3.6, or a CMOS BEOL with some dielectric removed—demonstrating robust, reliable suspension of a gatable graphene film is a challenging but necessary endeavour to prove the practicality of such an approach. (Gating of the graphene film will be necessary for an infrared absorption demonstration to bring the graphene close to its Dirac peak, where Pauli blocking is minimized.) While I was successful in transferring a single layer of graphene onto a slot metasurface with minimal tearing using graphene wet transfer followed by critical point drying, I found that this surface could not be wetted effectively by the liquid electrolyte I attempted to use to gate it. Using a multilayer sheet of hexagonal boron nitride (hBN) to support the monolayer graphene film slightly improved the wettability, but this solution is not ideal due to the influence of hBN’s strong optical phonon resonances on the metasurface optical properties. A more pragmatic approach may be to prepare a dielectric-graphene-dielectric-graphene-dielectric thin film stack to allow counter-doping of the graphene, with a dielectric relatively transparent in the infrared. The practical feasibility of the proposed device is predicated on such a development.

Besides questions related to fabricability, this research does not address uncertainty regarding the performance of as-fabricated metals. Different growth or deposition methods can incur dramatic differences in the optical properties of plasmonic metals [178], and it is largely for this reason that our theoretical results based on reported optical properties of thin-film gold[173] have limited predictive ability for experimental performance. Therefore, an experimental demonstration will be necessary to provide actual performance benchmarks for any given fabrication approach.

5.2.2 Prospects for CMOS integration

As discussed in Chapter 3, CMOS integration of the proposed slot antenna metasurfaces and graphene devices would provide a context to explore questions of graphene
integration and metal performance while also offering a realistic pathway towards fabrication of actual optoelectronic devices. While the fabrication process in Section 3.6 is not conducive to large-scale device fabrication due to its flip-chip nature and contiguous gold ground plane, a CMOS process would accommodate device integration as electrical connections to contact the devices could be routed down to the transistor layer in the gaps between adjacent antennas. CMOS integration of graphene devices is not a new concept, having been demonstrated in the context of a visible to short-wave IR imager using a commercial ROIC[16]. A most feasible path forward for demonstrating a slot antenna metasurface-based imager would likely require taping out a custom ROIC with slots either exposed to the top layer or easily accessible with minimal post-processing. Since the pitch between adjacent antennas is much smaller than the minimum ROIC pixel size (limited by the area needed for the circuitry for each pixel), it would be necessary to electrically connect multiple antennas from the same spectral channel for readout such that each ROIC pixel represents the total signal for a given spectral channel for several metasurface unit cells.

5.2.3 Metasurface images based on materials besides graphene

Imagers based on slot antenna metasurfaces need not be limited to use with graphene or other 2D materials. More traditional infrared active materials, such as HgCdTe or III-V materials, are also potential subjects for slot antenna metasurface integration. Indeed, a periodic patch antenna metasurface integrated with a QWIP active layer already been shown to enable room-temperature $\lambda = 9\mu m$ photodetection with a detectivity of $3 \times 10^7$ Jones, as the antennas concentrate the optical to electronic conversion into small volumes of active material, thus reducing thermal generation noise[47]. This is conceptually identical to the functionality offered by a slot antenna metasurface, and hints at the applicability of my research to alternative material platforms. The feasibility of slot antenna metasurface integration with traditional IR materials is further bolstered by the availability of epitaxial liftoff and membrane transfer processes available for various infrared material platforms, as previously mentioned in Section 3[206, 134]. Integration of such epitaxial HgCdTe or III-V mem-
branes with the proposed slot antenna metasurfaces thus offers another route towards multyspectral IR detection at room or TEC-accessible temperatures.

5.2.4 Imaging metasurfaces in the terahertz

The multispectral antenna metasurface concept is not limited to the mid-infrared. In fact, performance may be enhanced at terahertz wavelengths, which have been exploited for spectroscopic atmospheric and astronomic gas analysis as well as medical imaging[207]. The Drude model of free electrons in metal predicts that the imaginary part of the permittivity of a metal is given by:

\[
\text{Im}(\varepsilon) = \frac{\omega_p^2 \tau}{\omega(1 + \omega^2 \tau^2)}.,
\]

(5.1)

where \(\tau\) is the Drude scattering time and \(\omega_p\) is the plasma frequency[109]. The AC sheet resistance of a metal surface with perturbative loss is given by \(R_S = Z_0/\sqrt{2\text{Im}(\varepsilon)}\), where \(Z_0\) is the vacuum impedance, and the attenuation coefficient of a metal waveguide is proportional to \(R_S\)[208]. Therefore, the power lost per wavelength of propagation in a metal waveguide, and thus the power lost per cycle, will be proportion to \(R_S/\omega\). Hence for a metal waveguide resonator,

\[
Q \propto \frac{\omega}{R_S} = \frac{1}{Z_0} \sqrt{\frac{2\omega \omega_p^2 \tau}{1 + \omega^2 \tau^2}}.
\]

(5.2)

This is maximized at \(\omega = 1/\tau\), which falls in the 4–8 THz range for noble metals. Insofar as our the metasurfaces discussed in Chapter 3 operate around 40 THz, one would thus expect similar or better resonator performance down to roughly 1 THz. An impressive variety of terahertz detectors have been demonstrated, including thermal detectors such as bolometers and thermocouples, detectors based on plasmon waves in FET channels, Schottky diode detectors, quantum well detectors, nonlinear detectors, gated photoconductors[209] and even graphene-based PTE detectors[210]. Integrating such detectors with slot antenna metasurfaces scaled up to resonate at terahertz wavelengths could enable spectrally resolving focal plane arrays for terahertz
imaging. Since the antennas would have to be physically scaled up to accommodate terahertz wavelengths, a CMOS BEOL-based process would no longer be suitable for fabrication of the metasurfaces. Instead, a fabrication process similar to those used to produce multilayer circuit boards may be more practical.

5.3 Waveguide-integrated mid-IR detectors and chip-scale gas sensing

The range of research that could build upon the presented waveguide-integrated photodetectors is potentially quite broad. Here I discuss the most relevant steps forward.

5.3.1 Extended wavelength operation

A key limitation of my demonstration is that the photodetector characterization was limited to $\lambda = 5.2\,\mu m$, in this case due to light source availability. In principle, achieving operation at longer wavelengths would be as simple as scaling up the waveguides and devices to accommodate the larger optical modes. The graphene itself should also still absorb light as shown in Fig. 4-11, transitioning from an interband absorption-dominated to intraband absorption-dominated regime as $\lambda$ increases. In reality, however, some of the waveguide materials I use will begin to absorb at longer wavelengths, and finding appropriate substitutes may bring about engineering challenges. First, while CaF$_2$, the substrate material, is well-documented to be transparent up to about 8–10 $\mu m$[211], the onset of absorption of the HfO$_2$ is less clear and would need to be measured experimentally. For operation at wavelengths beyond 8–10 $\mu m$, CaF$_2$ may be replaced with BaF$_2$, which is transparent to 11–13 $\mu m$ and features similar optical, mechanical and crystallographic properties but is considerably more soluble in water. This may cause difficulties during the graphene transfer step, where the substrate is used to scoop a sheet of graphene out of water. To address this problem, one may consider first coating the BaF$_2$ substrate with a layer of a suitable ALD material to prevent dissolution, although finding an ALD material with sufficiently low ab-
sorption even for very thin layers may be challenging in and of itself. Alternatively, the graphene film may be transferred to an alternative solvent prior to placement on the substrate, or an different graphene transfer process can be used altogether. The waveguide material itself, GSSe, exhibits absorption at wavelengths beyond 11 µm, but may still be usable up to 14 µm if long gas-light interaction waveguides are not required[189]. The onset of absorption may be extended by a micron by instead using As₂Se₃ glass[212], which features otherwise similar optical properties. Finally, one may consider using chalcogenide glasses as the dielectric and cladding layers instead of ALD-deposited materials at longer wavelengths[91]. Operation beyond 14 µm is unlikely to be necessary for gas sensing applications due to the limited range of gas absorption peaks[23].

5.3.2 Device enhancements

I discussed the benefits of and some options for achieving enhanced Seebeck coefficients and improved light utilization for PTE devices in general in Sections 5.1.1 and 5.1.3, and these suggestions apply to my mid-IR waveguide on graphene platform. Beyond those already discussed, there are some additional options for improving light utilization that are unique to guided platforms with high refractive index contrast. For instance, a dielectric slot waveguide design may be used to concentrate the optical mode near the graphene pn-junction[213], taking advantage of the high refractive index of GSSe glass. More advanced dielectric waveguide cross-sections permit even stronger mode concentration than slot waveguides at the cost of increased difficulty of fabrication[214]. Finally, one could consider implementing a hybrid slot-plasmonic mode in which the source and drain contacts are placed close enough to a dielectric slot mode to perturb the mode shape. While metal-plasmonic waveguide modes offer the benefit of bringing the source and drain contacts closer together and thus minimizing excess resistance in the channel, slot or other light-concentrating dielectric waveguides do not suffer from ohmic losses as metal waveguides do. Further investigation would be required to select the best option for the present platform.
5.3.3 Demonstration of gas sensing

A clear next step for this research is to use the developed platform to perform gas sensing. Such a demonstration could for instance follow the methodology adopted by Su et al.[71], in which both the gas interaction waveguide and photodetector are encapsulated in a PDMS microfluidic chamber attached to the substrate via plasma-bonding. In this approach, the input and output waveguides as well as metal lines for electrical contact are extended beyond the walls of the PDMS chamber for access. This thin region of overlap between the optical mode and a PDMS wall would not incur enough absorption loss to preclude an experimental demonstration to wavelengths of at least 7 µm[215]. For demonstrations at longer wavelengths, it may instead be necessary to contain the entire chip in a gas flow chamber with IR-transparent windows and electrical feedthroughs, or identify an alternative material to clad over the access waveguide prior to incorporation of the PDMS chamber.

5.3.4 Scalability

Most aspects of the presented platform are readily scalable. However, the substrate and the CVD graphene are areas of uncertainty. The single crystal CaF$_2$ substrates used for this research are small, expensive, and difficult to handle. For scalable manufacturing, it would be preferable to instead begin with silicon wafers, followed by deposition of a thick layer of a low-index, IR-transparent material. However, the absorption of infrared thin films is considerably stronger than their monocrystalline counterparts, with an extinction coefficient of $k = 3.1 \times 10^{-4}$ at $\lambda = 5.3$ µm having been reported for evaporated CaF$_2$ films[216] and $k \geq 0.01$ at wavelengths beyond 10 µm for a variety of infrared-transmitting oxide and fluoride thin films[205]. While the observed absorption at $\lambda = 5.3$ µm is likely acceptable for a range of applications, incurring a loss of about 4 dB/cm for the present waveguide design and potentially lower for optimized designs, the absorption of infrared thin films at longer wavelengths is much more problematic and ultimately may require switching to a different platform outright, such as suspended Ge waveguides, to achieve scalability.
The scalability of CVD graphene transfer is a perpetual concern in graphene optoelectronics research. However, several promising techniques for high-quality large-area transfer have been developed. Kim et al. reported a particularly promising semi-dry transfer approach involving lamination of a reusable pressure-sensitive adhesive film onto a CVD graphene-on-Cu film, followed by copper etching and dry stamping of the graphene onto the target substrate[217]. The authors demonstrated transfer of large areas of graphene with high mobility onto various substrates, including silicon wafers and PET plastic films. Yet other reported techniques for large area graphene transfer include semi-dry transfer using PDMS stamps and fully dry stamp transfer[218]. Mass manufacturing of gas sensors and other optoelectronic devices based on graphene will thus require investigating these various options and selecting a transfer technique that yields material of sufficient quality at the desired manufacturing scale.


[118] Vaidotas Miseikis, Simone Marconi, Marco A Giambra, Alberto Montanaro, Leonardo Martini, Filippo Fabbri, Sergio Pezzini, Giulia Piccinini, Stiven Forti,


[190] Hongtao Lin, Yi Song, Yizhong Huang, Derek Kita, Skylar Deckoff-Jones, Kaiqi Wang, Lan Li, Junying Li, Hanyu Zheng, Zhengqian Luo, Haozhe Wang,
Spencer Novak, Anupama Yadav, Chung-Che Huang, Ren-Jye Shiuie, Dirk En- 
glund, Tian Gu, Daniel Hewak, Kathleen Richardson, Jing Kong, and Juejun 
805, 2017.

[191] Laura Kim, Seyoon Kim, Pankaj K Jha, Victor W Brar, and Harry A Atwater. 
Mid-infrared radiative emission from bright hot plasmons in graphene. *Nature 

[192] T. D. Visser, H. Blok, B. Demeulenaere, and D. Lenstra. Confinement fac-
tors and gain in optical amplifiers. *IEEE Journal of Quantum Electronics*, 

Thermal stability and electrical characteristics of ultrathin hafnium oxide gate 
dielectric reoxidized with rapid thermal annealing. *Applied Physics Letters*, 


2012.


[197] Jakob E. Muench, Alfonso Ruocco, Marco A. Giambra, Vaidotas Miseikis, 
Dengke Zhang, Junjia Wang, Hannah F. Y. Watson, Gyeong C. Park, Sha-
hab Akhavan, Vito Sorianello, Michele Midrio, Andrea Tomadin, Camilla Co-
letti, Marco Romagnoli, Andrea C. Ferrari, and Ilya Goykhman. Waveguide-
integrated, plasmonic enhanced graphene photodetectors. *Nano Letters*, 

[198] Jianing Chen, Michela Badioli, Pablo Alonso-González, Sukosin Thongrat-
tanasiri, Florian Huth, Johann Osmond, Marko Spasenović, Alba Centeno, 
Amaia Pesquera, Philippe Godignon, et al. Optical nano-imaging of gate-

[199] Lin Xiong, Carlos Forsythe, Minwoo Jung, AS McLeod, SS Sunku, YM Shao, 
GX Ni, AJ Sternbach, S Liu, JH Edgar, et al. Photonic crystal for graphene 

[200] Seung-Geol Nam, Dong-Keun Ki, and Hu-Jong Lee. Thermoelectric transport 
2010.


