Towards next generation of optoelectronics: from quantum plasmonics and 2D materials to advanced optimization techniques of nanophotonic devices

Thesis by
Yury Tokpanov

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy in Applied Physics

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2020
Defended May 21st 2020

Yury Tokpanov
ORCID: 0000-0001-5123-7428

ACKNOWLEDGEMENTS

Making it through my PhD adventure without support from my family, my friends, and Caltech community would have been impossible.

I would like to thank my beloved wife Anya for supporting me along this journey, for being closer to me every day, and for sharing her experience in optics and material science. I am grateful to my mother Svetlana and all my family for their support, advice, and understanding.

I would like to thank my adviser, Harry Atwater, for his support and for giving me the opportunity to work on challenging and diverse projects, from which I learned a lot.

I am grateful to all my multi-disciplinary collaborators. For my quantum plasmonics project, I would like to thank James Fakonas for providing initial guidance, Benjamin Vest for his insight, help with data analysis and paper, and for good time overall, Yousif Kelaita for initial design of experimental setup, Artur Davoyan and Ruzan Sokhoyan for in-depth discussions, Ragip Pala, Dagny Fleischman, Zach Aitken, Krishnan Thyagarajan, Yulia Tolstova, and Sunita Darbe for equipment training, and Julia Greer for providing access to the focused ion beam in her laboratory. For my 2D materials project, I would like to thank Cora Went and Joeson Wong for their deep insights and eagerness to have long discussions with me, Marco Bernardi, Maurizia Palummo, and I-Te Lu for help regarding ab initio theory and practice and some aspects of the project. For my optimization project, I would like to thank nanophotonic team Dagny Fleischman, Prachi Thureja, and Kate Fountaine for their insights, discussions, and datasets, and the machine learning team Yisong Yue, Yuxin Chen, and Jialin Song for algorithmic insights and intricacies explanations, and all of members of this collaboration for fruitful brainstorming sessions, during which many interesting ideas arose.

Many administrators in the Atwater lab, APhMS department, graduate office, and ISP helped with navigating the formal aspects of PhD life and paperwork and provided timely emotional support. I am gratetul to Christy Jenstad, Jennifer Blankenship, Laura Flower Kim, Kam Flower, Daniel Yoder, and Natalie Gilmore.

I would like to thank Andrei Faraon, Kerry Vahala, and Yisong Yue for being on my PhD committee, I appreciate you time and effort. I would like to thank Oskar Painter, who served on my candidacy committee.

ABSTRACT

In this thesis, we explore different novel concepts and materials for the next-generation of nanophotonic and optoelectronic devices that could be used both in classical and quantum settings.

First, we study quantum coherence properties of surface plasmon polaritons (SPPs) in the regime of extreme dispersion. Most experiments to date, that tested quantum coherence properties of SPPs, used essentially weakly-confined plasmons, which experience limited light-matter hybridization, thus restricting the potential for decoherence. Our setup is based on a hole-array chip supporting SPPs near the surface plasma frequency, where plasmonic dispersion and confinement is much stronger than in previous experiments, making the plasmons much more susceptible for decoherence processes. We generated polarization-entangled pairs of photons and transmitted one of the photons through this plasmonic hole array. Our results show that the quality of photon entanglement after the highly-dispersive plasmonic channel is unperturbed. Our findings provide a lower bound of 100 femtoseconds for the pure dephasing time of dispersive plasmons in our materials, and show that even in a highly dispersive regime, surface plasmons preserve quantum mechanical correlations, making possible harnessing the power of extreme light confinement for integrated quantum photonics.

Second, we systematically study different passivation schemes of sulfur vacancies in 2D MoS_2 using first-principles calculations based on density functional theory. We aim at building a microscopic understanding of passivation mechanisms of treatment with TFSI superacid – a popular approach of to improve optical properties. Since superacids have a strong ability to donate protons, we consider hydrogenation and protonation of sulfur vacancies as a possible passivation scheme. Our calculations show that effects of protonation and hydrogenation on properties of 2D MoS_2 are very similar. Moreover, we find that four hydrogen atoms can fully "heal" sulfur vacancies in this material. Our results are an important step towards controllable defects design in 2D transition metal dichalcogenides.

And third, we study applications of advanced methods of optimization and machine learning to the design of different nanophotonic devices. We explore feasibility of using novel multi-fidelity Gaussian processes optimization technique to optimize plasmonic mirror filters for hyperspectral imaging. We compare our results

with other common optimization approaches. Then we apply deep-learning inspired techniques to optimize control voltages of individual pixels of active metasurfaces to achieve dynamic beamsteering. We obtain interesting results that pave the way for future experiments both in nanophotonics and machine learning fields.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Yury S. Tokpanov, James S. Fakonas, Benjamin Vest, and Harry A. Atwater. "Quantum Coherence Preservation in Extremely Dispersive Plasmonic Media". In: *Physical Review Applied* 12.4 (2019). Publisher: American Physical Society, p. 044037. DOI: 10.1103/PhysRevApplied.12.044037.
Yu.T. participated in the conception of the project, designed and fabricated samples, designed and assembled optical experimental setup, collected and prepared the data, and participated in the writing of the manuscript.

[2] Jialin Song, Yury S. Tokpanov, Yuxin Chen, Dagny Fleischman, Kate T. Fountaine, Harry A. Atwater, and Yisong Yue. "Optimizing Photonic Nanostructures via Multi-fidelity Gaussian Processes". In: *NeurIPS 2018 Workshop on Machine Learning for Molecules and Materials*. 2018. URL: http://arxiv.org/abs/1811.07707.
Yu.T. participated in the conception of the project, established collaboration between research groups, assisted in collecting data, prepared the data, and participated in the writing of the manuscript.

TABLE OF CONTENTS

Acknowledgements . . . iii
Abstract . . . iv
Published Content and Contributions . . . vi
Table of Contents . . . vi
List of Illustrations . . . viii
Chapter I: Introduction . . . 1
1.1 Motivation . . . 1
1.2 Scope of this thesis . . . 6
Chapter II: Preservation of quantum entanglement by highly-dispersive surface plasmons . . . 7
2.1 Introduction . . . 7
2.2 Generation of entangled pairs of particles . . . 9
2.3 Elliptical hole array in the linear dispersion regime . . . 14
2.4 Hole array with nonlinear dispersion . . . 16
2.5 Results and discussion . . . 23
Chapter III: First-principles study of passivation of sulfur vacancies in MoS_2 . . . 27
3.1 Introduction . . . 27
3.2 Computational methods . . . 28
3.3 Comparison of properties of pristine MoS_2 and MoS_2 with sulfur vacancy . . . 35
3.4 Hydrogenation of sulfur vacancies in 2D MoS_2 . . . 39
3.5 Passivation by hydrohalic acids (simulating protonation) . . . 45
3.6 Passivation by carbon and oxygen . . . 48
3.7 Conclusion and outlook . . . 54
Chapter IV: Optimization of nanophotonic devices . . . 55
4.1 Introduction . . . 55
4.2 Optimization of plasmonic mirror filters . . . 55
4.3 Numerical approaches to black-box optimization . . . 57
4.4 Multi-fidelity Bayesian optimization . . . 60
4.5 Experimental setup and results . . . 63
4.6 Optimization of universal metasurface . . . 67
4.7 Solving inverse problem using deep-learning . . . 70
4.8 Optimizing directivity with perceptron-like network . . . 74
4.9 Optimization of positions of metasurface pixels . . . 82
4.10 Summary and outlook . . . 87
Bibliography . . . 89

LIST OF ILLUSTRATIONS

Number *Page*

1.1 This figure is adapted from [2] (a) Optical micrograph of 26-mode, 88-MZI PNP by [3]. PCBs are visible above and below the chip. (b) Artistic rendering of a U(4) PNP by [4]. (c) Germanium-doped glass six-mode, 15-MZI PNP by [5]. (d) Four-mode, six-MZI PNP by [6] implemented in the SOI platform. 3

1.2 An illustration by [21] of liquid-crystal metasurface (LCM) beam steering depicts its operation. The incident light is TM-polarized. Shown here, the LCM steers to two different output angles (a and b), determined by the spatial frequency of the phase-modulation pattern applied to the array of tunable resonators on the LCM (c and d). Higher spatial frequencies steer the output beam closer to the incident beam. 5

2.1 Experimental setup for the measurement of polarization-entanglement preservation. Pump photons at 406 nm are sent toward a pair of BiBO crystals and generate pairs of polarization entangled photons that propagate along two separate paths. Along the upper path, we can insert a metallic hole array, and measure the transmission of the entangled light that has been coupled to plasmons. 10

2.2 Calibration of the setup, and entanglement between produced pairs of photons. Number of coincidence counts as a function of polarizers angles without plasmonic sample (solid lines are fits to cosine). The visibility of the different cosine fits is nearly equal to one, indicating quasi-perfect entanglement between the photons of our SPDC source. 13

2.3 Design of plasmonic elliptical hole array. Cross-sectional schematic and dispersion relation of elliptical hole arrays for SPPs supported at the gold/glass interface. At the wavelength of the down-converted photons (812 nm, as shown by the red solid line), the dispersion is "photon-like", i.e., linear, and very close to the light line. 15

2.4 Study of entanglement preservation with the elliptical hole array. Elliptical hole array used to study the influence of hole geometry on the preservation of photon entanglement: (A) SEM image, where the orientation of the minor and major axis can be seen oriented at ±45°; (B) Transmission factor of the device for different polarizations of incident light. The holes have no rotation symmetry anymore, so that the transmission maximum varies between 12.5% for horizontally polarized light and 7% for vertically polarized light. (C)&(D) Normalized number of coincidence counts as a function of polarizer angles in the presence of an elliptical plasmonic hole array (solid line represents fit to the full model) in two configurations : (C) With entangled photons and for different fixed directions of the polarizer beta; (D) With classical light and similar choice for beta. The plots with $\beta = 45°$ and $\beta = 135°$ show a similar decrease in visibility in both configurations, indicating that this results from a purely classical effect. 16

2.5 Plasmonic hole array design in our experiment: cross-sectional schematic (top) and dispersion relation of circular hole arrays for SPPs supported in the silicon/gold/silicon structure, which exhibits strongly nonlinear dispersion at 812 nm (bottom). The red line shows the energy of the plasmons excited in our experiment. 18

2.6 Investigation of the evolution of the transmission resonances for various structures. The transmission spectra at normal incidence of nine different hole array structures, with periods ranging from 600 nm to 1000 nm have been experimentally measured. The transmission amplitude is displayed as a color map, and as a function of both the energy (vertical axis) and of the periodicity of the hole array (horizontal axis). Experimental data have been interpolated between the nine sets of data points. The different white branches displayed on the colormap correspond to transmission resonances. The branch corresponding to the plasmon resonance exploited in the main experiment is labeled as P0 and the operating point of the experiment (structure periodicity of 850 nm, plasmon resonance at 812 nm) is marked with a black X. Another branch of plasmon resonance at lower energies is identified as P1. These branches were fitted using relation 2.4.1 for different sets of parameters (solid color lines). For both P0 and P1, agreement between experimental data and model is good and allows us to determine the plasmon wavevectors at resonance. 20

2.7 Comparison of the analytical dispersion relation with the transmission measurements. The colormap of Fig. 2.6 is represented this time as a function of the wavevector, using the conversion between periodicity of the structures and resonant wavevectors given by relation (2.4.1) with $n = 6$ and $d = 300$ nm. The two branches P0 and P1 are identified on the plot, the operating point of the main experiment is marked with a black X. P0 is well fitted by a segment of the dispersion relation that corresponds to a non-linear highly-dispersive regime for the plasmons (blue solid line), and the position of the operating point is in agreement with the initial design of the experiment. 22

2.8 Unfolding the transmission spectra to measure indirectly the dispersion relation. We compare the analytical dispersion relation (black solid line) of the structure with the position of the local transmission maxima measured experimentally with structures of various periodicities (vertical cross sections of Fig. 2.6). For each measured transmission maximum, two values of associated wavevector are derived, by using (2.4.1) with two sets of parameters. If the transmission maximum belongs to the branch P0, the two wavevectors values are calculated using $n = 6$, $d = 300$ nm (dashed blue line) or $n = 5$, $d = 430$ nm (dashed red line). If the transmission maximum belongs to the branch P1, the two wavevectors values are calculated using $n = 4$, $d = 300$ nm (dashed yellow line), or $n = 3$, $d = 430$ nm (dashed purple line). This shows that experimental transmission spectra are in great agreement with the expected dispersion relation. . 23

2.9 Hole array for plasmons in higly-dispersive regime. (A) SEM image of the sample. The period of the 2D array is 850 nm. The different material species experience different milling rates that slightly affect the shape of the holes' perimeter. (B) Transmission spectrum of the hole array. With the holes being circular, there is no polarization dependence. The broadening of the transmission feature around 810 nm can be attributed to the imperfect shape of the holes. Note that the plasmons experience a significant absorption. (C) Number of coincidence counts as a function of polarizer angles in the presence of the hole array in a highly-dispersive regime (solid lines are fits to cosine). Whatever the choice of β is, and even when placed at 45°or 135°, the visibility of quantum interference remains almost equal to one, indicating near perfect preservation of entanglement between particles. 24

3.1 Crystalline structure of a MoS_2 monolayer, blue balls are molybdenum atoms, yellow circles are sulfur atoms. Illustration adapted from https://www.ossila.com/pages/molybdenum-disulfide-mos2. 27

3.2 Influence of the size of supercell on bandstructure of 2D MoS_2. . . . 33

3.3 Influence of the strain on bandstructure of primitive 2D MoS_2 (computed on primitive cell). 34

3.4 Influence of spin-orbit coupling on dielectric function of 2D MoS_2. . 34

3.5 4x4 supercells of pristine (top) and defective (bottom) MoS_2 (slightly rotated top view). 35
3.6 Comparison of band structures of pristine (top) and defective (bottom) MoS_2 (4x4 supercell). 37
3.7 Comparison of densities of states of pristine (top) and defective (bottom) MoS_2 (4x4 supercell). 38
3.8 Comparison of imaginary parts of dielectric functions of pristine and defective MoS_2. 38
3.9 4x4 supercells of 2D MoS_2 in which sulfur vacancies were filled with different number of hydrogen atoms. 40
3.10 Band structure of MoS_2 in which sulfur vacancy is filled with one hydrogen atom (4x4 supercell). 41
3.11 Band structure of MoS_2 in which sulfur vacancy is filled with two hydrogen atoms (4x4 supercell). 41
3.12 Band structure of MoS_2 in which sulfur vacancy is filled with three hydrogen atoms (4x4 supercell). 41
3.13 Band structure of MoS_2 in which sulfur vacancy is filled with four hydrogen atoms (4x4 supercell). 42
3.14 Projected densities of states of 2D MoS_2 in which sulfur vacancies were filled with different number of hydrogen atoms. 42
3.15 Imaginary parts of dielectric functions of 2D MoS_2 in which sulfur vacancies (SV) were filled with different number of hydrogen atoms. . 43
3.16 Comparison of imaginary parts of dielectric functions of pristine MoS_2 and defective 2D MoS_2 in which sulfur vacancies (SV) were filled with four hydrogen atoms. 44
3.17 Cross-sectional view of 4x4 supercells of 2D MoS_2 in which sulfur vacancy is filled with different hydrohalic acids. 46
3.18 Bandstructure of 2D MoS_2 in which sulfur vacancy is filled with different hydrohalic acids. 47
3.19 Projected densities of states of 2D MoS_2 in which sulfur vacancies were filled with different hydrohalic acids in comparison with case of passivation by one hydrogen atom. 48
3.20 Imaginary parts of dielectric functions of 2D MoS_2 in which sulfur vacancies were filled with different hydrohalic acids in comparison with case of passivation by one hydrogen atom. 48

3.21 4x4 supercell of 2D MoS_2 in which sulfur vacancies were filled with two carbon atoms. 50
3.22 Bandstructure of 2D MoS_2 in which sulfur vacancy is filled with a different number of carbon atoms. 50
3.23 Projected densities of states of 2D MoS_2 in which sulfur vacancies were filled with a different number of carbon atoms. 51
3.24 Imaginary part of dielectric function of 2D MoS_2 in which sulfur vacancies were filled with different two carbon atoms. 51
3.25 4x4 supercell of 2D MoS_2 in which sulfur vacancies were filled with two oxygen atoms. 52
3.26 Bandstructure of 2D MoS_2 in which sulfur vacancy is filled with a different number of oxygen atoms. 52
3.27 Projected densities of states of 2D MoS_2 in which sulfur vacancies were filled with a different number of oxygen atoms. 53
3.28 Imaginary parts of dielectric functions of 2D MoS_2 in which sulfur vacancies were filled with a different number of oxygen atoms. 53
4.1 Benefit from multi-fidelity Bayesian optimization. The left panel shows normal single fidelity Bayesian optimization where locations near a query point (crosses) have low uncertainty. When there is a lower fidelity cheaper approximation in the right panel, by querying a large number of points of the lower fidelity function, the uncertainty in the target fidelity can also be reduced significantly. 59
4.2 Influence of mesh size on the results of FDTD simulations. 65
4.3 Multi-fidelity based on conformal mesh size. Every method is run 20 times, and we plot the mean plus/minus one standard error in the figures. 67
4.4 Multi-fidelity based on conformal simulation time. Every method is run 20 times, and we plot the mean plus/minus one standard error in the figures. 67
4.5 Reflection coefficient amplitude-phase relation. 70
4.6 Solving inverse design problem directly using neural networks. 71
4.7 Results of neural network approximating inverse calculation for linear phase profiles applied to ideal device. 72
4.8 Solving inverse design problem directly using autoencoder-like network. 73
4.9 Results of autoencoder-like neural network. 74
4.10 Solving inverse design problem directly using perceptron-like network. 75

4.11 Optimized directivity of ideal device (orange curve) in comparison with directivities computed using conventional linear phase profile (blue curve). 77
4.12 Optimized directivity of semi-realistic device (orange curve) in comparison with directivities computed using conventional linear phase profile (blue curve). 77
4.13 Optimized directivity of realistic device (orange curve) in comparison with directivities computed using conventional linear phase profile (blue curve). 78
4.14 Array factor (top) and phases (bottom) optimized for steering into 30°using real device. 79
4.15 Array factor (top) and phases (bottom) optimized for steering into75°using real device. 80
4.16 Array factor (top) and phases (bottom) optimized for steering into two angles using real device: -20°and 50°. 81
4.17 Comparison of periodic (blue dots) and optimized (orange) positions for array with average spacing between pixels equal to 2 μm. 83
4.18 Comparison of directivities between: blue - ideal device with optimized phases and positions; green - realistic device with optimized phases and positions; orange - realistic device with optimized phases but random positions. 83
4.19 Array factor (top) and phases (bottom) optimized for steering into 50°for device with average spacing of 2 μm. 84
4.20 Comparison of directivities for cases of optimized and periodic positions for ideal device with average spacing of 10 μm. 85
4.21 Comparison of directiviites for cases of optimized and periodic positions for realistic device with average spacing of 10 μm. 85
4.22 Comparison of directivities for cases of optimized and periodic positions for ideal device with average spacing of 100 μm. 86
4.23 Comparison of directivities for cases of optimized and periodic positions for realistic device with average spacing of 100 μm. 86
4.24 Possible schema for co-design of metasurface parameters. 87

Chapter 1

INTRODUCTION

1.1 Motivation

Many areas of technological progress significantly depend on advances in solid state physics, electronics, and optics. Sustaining the pace of innovation requires novel approaches and breakthroughs in these areas.

One of the most famous example, is "Moore's law", which is an observation about exponential increase with time of the density of transistors in state-of-the-art integrated circuits. It is not a physical law, but an empirical observation, which turned out to be quite accurate for several decades. It was used by semiconductor industry for long-term planning and setting up milestones in research and development. In the past decade, however, improvements in industrial transistor technology have slowed down significantly due to critical dimension approaching the physical limits of devices. Current circuit design is still mostly based on classical physics, even though to accurately simulate properties of individual elements, one has to rely on quantum mechanics. However, increasing density of transistors naturally leads to appearance of "undesired" quantum effects, such as tunneling of electrons between circuit elements, which increases error rate.

Another significant related problem is heat dissipation. Actually, in current computing devices, that allow "loss" of information during computation (e.g., clearing part of memory), heat generation is inevitable, since physical information is not actually lost, but instead is dumped to the environment, increasing its entropy. However, we are far away from the such a limit, as most of the generated heat is due to inefficiencies of elements of the circuit.

Improving power efficiency is of very significant importance. At the moment of writing of this thesis, US Department of Energy is partnering with Hewlett-Packard and AMD to build the El Capitan supercomputer to be installed at Lawrence Livermore National Laboratory. It should have performance of 2 exaFLOPS ($2 \cdot 10^{18}$ floating point operations per second). Power consumption is projected to be a record 40MW, several times higher than current fastest machines. Supercomputers are extremely important for scientific and technological advances, however, due to multiple reasons (including those outlined above), their development has stagnated

in the last decade.

One can argue that due to the slowing down of Moore's law, most recent advancements are being done through so-called architecture improvements: e.g., more computing cores on a single chip, optimization of hardware implementation of machine commands, etc. Making physical improvements requires novel approaches in element base.

One of possible alternatives is so-called optical (or photonic) computing. It has been a holy grail of technology since 80's [1]. Many optical equivalents to traditional electronic elements were demonstrated. Proposed optical components have superior bandwidth and potentially orders of magnitude better power efficiency.

With advances in fabrication of integrated circuits from different materials, a new area of optics has emerged - nanophotonics. It is a scientific and engineering field that studies behaviour and manipulation of light in integrated circuits at the nanoscale. Many nanophotonic devices can be organically combined with electronic or optoelectronic techniques, making them a very promising candidate for a novel elements base of next-generation information processing.

Many research groups and industry leaders have been actively working on implementation of programmable nanophotonic processors (PNPs) (see Fig. 1.1). Even though a general purpose programmable version of such a device is still in the future, application-specific nanophotonic integrated circuits have been already demonstrated experimentally. One the of the most prominent examples of such chips is a general purpose matrix multiplication nanophotonic accelerator, which generated a lot of interest in academia, government applied research, and industry. Such a device has broad range of applications in both in classical and quantum computing [2].

For example, a few years years ago, Google introduced Tensor Processing Units (TPUs) [7] for accelerating their in-house deep learning algorithms. To make a bet on significantly more power-efficient technology, Google Ventures recently invested heavily in a new startup called Lightmatter, which is based on the idea of applying PNPs for hardware acceleration of neural networks [8]. Though, the idea to use optical computing for implementing neural networks was suggested in the 80s by Caltech scientists [9].

However, the main disadvantage of nanophotonic devices is still relative bulkiness due to size of wavelengths of visible and infrared light, which are being utilized

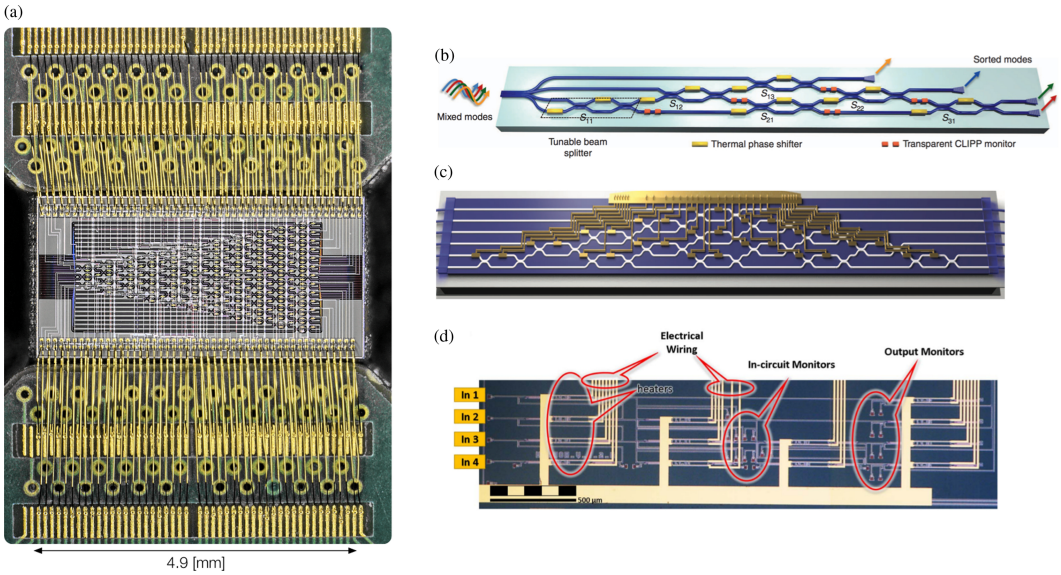


Figure 1.1: This figure is adapted from [2] (a) Optical micrograph of 26-mode, 88-MZI PNP by [3]. PCBs are visible above and below the chip. (b) Artistic rendering of a U(4) PNP by [4]. (c) Germanium-doped glass six-mode, 15-MZI PNP by [5]. (d) Four-mode, six-MZI PNP by [6] implemented in the SOI platform.

in such devices. Plasmonics, one of the subfields of nanophotonics, is trying to address this issue. Plasmonic devices utilize a special kind of electromagnetic waves that can exist on the boundary between metal and dielectric – so-called surface plasma waves, that are guided along the interface [10]. Quanta of such waves are called surface plasmon polaritons (SPPs). SPPs are hybrid excitations of charge motion in metals (plasmons) and electromagnetic waves in dielectrics (polaritons). Due to such hybridization, these surface waves have significantly smaller wavelength than incident photons, enabling significantly tighter spatial confinement and higher local field intensity.

Both of these effects allow smaller-sized plasmonic devices in comparison to their photonic counterparts. Unfortunately, plasmonic structures usually experience much higher losses, limiting potential applications. However, there are several important cases where plasmonics can be especially helpful. Due to increased light-matter interaction and decreased modal volume, plasmonics is excellent for sensing [11, 12, 13], sub-diffraction limit imaging [14, 15], and improving properties of quantum emitters [16, 17].

In addition to being a promising platform for computing, photonics can help with data transfer. The current revolution in communications has been partially fueled by development and implementation of optical fibers (Nobel Prize in Physics

in 2009). Interconnects between different components are one of the bottlenecks of current computing devices. Similar to how optical fibers have replaced cables, replacing metallic interconnects with photonic waveguides can potentially improve both bandwidth and losses. However, it will require compact and efficient optoelectronic sources and sensors. Plasmonics is very promising in this regard [18].

Another very prominent application for photonics and plasmonics is imaging, including both recording and reproducing light distributions. Active holographic displays are the ultimate type of displays, which would be able to recreate distribution of electromagnetic fields in space-time both in amplitude/intensity and phase (current displays only work with intensity). One of the concepts of such devices is to use arrays of photonic emitting elements (pixels), that can be addressed individually (using electric currents or voltage), each of which serving as a source of light with controlled amplitude and phase [19]. According to Kotelnikov's sampling theorem, to reproduce arbitrary field distributions one needs the spacing between pixels to be smaller than half the wavelengths of light, which unfortunately leads to an extreme number of such pixels, currently out of reach for both fabrication technology and information handling capabilities: holographic display of laptop size would comprise trillions of pixels, which must be coherent with each other. However, there are alternative concepts, which do not intend to fully recreate electromagnetic field, but rather to create the appearance of 3D images. One such device is the so-called light-field display, which can direct light into multiple angles, creating necessary illusion. There are many proposed ways of implementing such devices. Photonic waveguides and gratings allow for implementation of such technology, which is currently being actively pursued by several companies.

Even though general purpose holographic displays are very hard to implement, there are some very important applications that can benefit from their simplified versions, built for specific goals. For example, there is a lot of interest in development of active metasurfaces, which can steer the reflected beam into arbitrary angles electronically [20]. Such technology is important for the implementation of mechanically-free fully-electronic LiDARs for autonomous driving and other applications (Fig. 1.2). Such LiDARs potentially are much faster and smaller than currently used ones, in which a laser beam is mechanically shifted to scan surroundings and recreate their depth field. Metasurfaces can also be designed to manipulate light in other ways (like focusing), creating a real way for implementing flat optics.

An interesting and maybe somewhat surprising fact is that development of

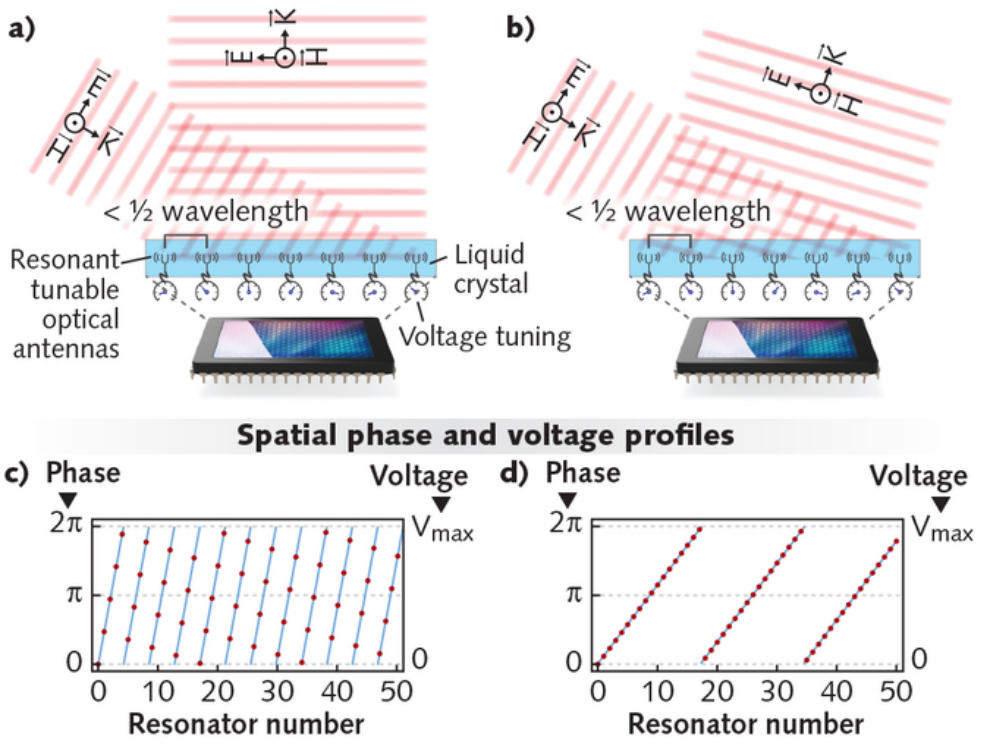


Figure 1.2: An illustration by [21] of liquid-crystal metasurface (LCM) beam steering depicts its operation. The incident light is TM-polarized. Shown here, the LCM steers to two different output angles (a and b), determined by the spatial frequency of the phase-modulation pattern applied to the array of tunable resonators on the LCM (c and d). Higher spatial frequencies steer the output beam closer to the incident beam.

photonic metasurfaces and photonic processors have many similar engineering challenges and goals. If one can develop a true active photonic holographic display, one can adapt such technology for the purposes of building nanophotonic circuits for linear optics quantum computer, as requirements for the coherence of light are similar. Of course, classical optical computing can also benefit from such technology. Because of versatility, photonics is a very interesting platform for such purposes.

Photonics and plasmonics can be used also for the implementation of next-generation cameras. One of the applications is compact filters for hyperspectral imaging. The hyperspectral camera is able to record images of the scene in several wavelengths of light (not only in red, green, blue as in traditional cameras). Such images convey more information about objects, allowing potentially for identification of materials [22]. Hyperspectral sensors find applications in astronomy, agriculture, molecular biology, biomedical imaging, geosciences, physics, and surveillance.

Plasmonic filters allow for compactification of such devices.

Here we have highlighted just a few applications of nanophotonics (e.g., others are solar cells and artificial photosynthesis for sustainable and renewable energy).

1.2 Scope of this thesis

In this thesis we explore different novel concepts and materials for the next-generation of nanophotonic and optoelectronic devices, that could be used both in classical and quantum settings.

- In Chapter 2, we study quantum decoherence properties of surface plasmons in the regime of extreme dispersion, which is achieved near surface plasmon resonance. By performing, experiment with a pair of polarization entangled photons, that propagate through plasmonic hole array, we can make conclusions regarding some microscopic details of surface plasmons and possibilities for further investigations and applicability in quantum optics.

- In Chapter 3, we systematically study different passivation schemes of sulfur vacancies in 2D MoS_2 using first-principles calculations based on density functional theory. The main focus is dedicated to hydrogenation and protonation of sulfur vacancies, as this is a possible mechanism of passivation of sulfur vacancies in 2D MoS_2 during treatment with TFSI superacid – popular approach to improve optical properties. In addition, we also study passivation with other elements.

- In Chapter 4, we study the application of advanced methods of optimization and machine learning to the design of different nanophotonic devices. We explore feasibility of using the novel multi-fidelity Gaussian processes optimization technique to optimize plasmonic mirror filters for hyperspectral imaging. We compare our results with other common optimization approaches. Then we apply deep-learning inspired techniques to optimize control voltages of individual pixels of active metasurfaces to achieve dynamic beamsteering. We obtain interesting results that pave the way for future experiments.

Chapter 2

PRESERVATION OF QUANTUM ENTANGLEMENT BY HIGHLY-DISPERSIVE SURFACE PLASMONS

2.1 Introduction

Understanding the quantum nature of light and matter is of central importance for advancing modern technology. For example, one approach to physical realization of a quantum computer is envisioned to be through the use of linear optical components [23], which can be arranged in the form of integrated photonic circuits. Since some of the building block elements for such a scalable quantum photonic system (phase shifters, modulators, directional couplers, etc.) and the coherence and fabrication requirements are very similar to present-day chip-based nanophotonic circuit elements, recent experimental advances [19] have inspired optimism about the technical feasibility of quantum integrated photonic systems, if suitable single-photon sources, memories, and detectors can be realized.

One of the important branches of photonics is plasmonics, which enables extreme light confinement utilizing surface plasmon polaritons (SPPs), the quanta of so-called surface plasma waves that are excited on the boundary between a metal and a dielectric [10]. SPPs are bosons, so their quantum statistical behavior is expected to be similar to that of photons. By confining electromagnetic energy in small modal volumes, plasmonics allows significantly enhanced light-matter interactions at the nanoscale, and has found interesting applications in classical photonics for sensing [11, 12, 13], sub-diffraction limit imaging [14, 15], and paving the way towards strong light-matter interactions by reaching, for example, strong coupling regimes [16, 17]. However, light-matter hybridization in SPPs has an important consequence. SPPs are collective excitation of electrons with a mixed electronic and electromagnetic character, while photons in free space are purely electromagnetic excitations, and light propagating through dielectric linear media is described by polaritons, mixing electromagnetic excitation and the motion of bound electrons that do not experience much interactions with the rest of the environment. Through the motion of free electrons in the metal SPPs are coupled to matter by many degrees of freedom. One consequence is that SPPs experience propagating losses due to the Ohmic losses of the moving electrons comprised in an SPP. In

a fully quantum optical picture, light-matter hybridization could be anticipated to make SPPs sensitive to decoherence from dephasing, which is often present in systems with Ohmic losses, leading to disappearance of their quantum features, such as entanglement.

Recently, several groups have performed plasmonic analogues to landmark quantum optics experiments using SPPs in lieu of single photons, yielding results such as single plasmon interferences [24, 25], plasmonic Hong-Ou-Mandel experiments [26, 27, 28], and entanglement experiments [29, 30, 31]. These successful experiments faced significant plasmonic absorption, manifested as Ohmic losses, but managed to preserve enough of the plasmons to highlight various quantum features of discrete SPPs. However, an apparently surprising result was that most of the experiments performed reported very good or excellent preservation of the quantum interference contrast, possibly indicating that pure dephasing processes are much slower than pure absorption. Notably in each of these experiments, despite the fact that the experimental conditions pre-selected and tested the coherence of the “surviving" non-absorbed plasmons, the strong plasmonic absorption observed is the undeniable sign of non-negligible coupling between the particles and their environment. The observation of such coupling is an indication that the degrees of freedom of the plasmons are likely to become entangled with the degrees of freedom of the environment, a description that is commonly used to explain the vanishing of quantum interference features, or in other words, decoherence. This has been verified, for example, in an experiment based on plasmonic waveguides [32]. Decoherence is one of the limiting factors for current and future quantum technology. Hence the question of how SPPs lose quantum mechanical coherence and if their quantum properties can be protected over long propagation distances or under strong light-matter interactions is of significant importance.

We note that until now, quantum plasmonics experiments used exclusively plasmons in a regime far from the surface plasma frequency. In other words, the plasmons exhibited a highly “photon-like" behavior with weak confinement, which intrinsically limits the decoherence processes. Indeed, in the case of “photon-like" plasmons, we expect the plasmon resonances to have only a weak admixture with the electronic degrees of freedom in the metal, leaving the gateway only ajar to significant coupling and entanglement between the SPPs and the metallic environment. Therefore, in this “photon-like" regime, one could argue that from the perspective of decoherence processes, some quantum plasmonics experiments are somewhat

analogous to other quantum optics experiments performed with photons all the way and exhibiting no decoherence. That would also mean that plasmonic losses here play a role that is not different from optical losses introduced by beam splitters, stray reflections, or neutral density filters. Thus, there is a need to investigate quantum plasmonics in other regimes of plasmon propagation, where the competition between absorption and pure dephasing could result in observable decoherence.

In this chapter, we report results of a quantum plasmonics experiment to investigate the robustness of coherence in a highly dispersive plasmon regime and the disappearance of quantum entanglement. More precisely, in a series of experiments inspired by [29], we measure the preservation of polarization entanglement between two photons after one photon is converted into a plasmon propagating on a hole array, which is then subsequently reconverted into a photon [33, 34, 35]. We performed polarization entanglement experiments in plasmonic hole arrays with circular holes which are designed to be in a highly dispersive regime, i.e., with single plasmons energy close to the surface plasma frequency. In this highly dispersive regime, SPPs are tightly confined and have a much stronger interaction with the electronic system (one manifestation of which is singificantly larger absorption), which in principle can lead to the destruction of quantum correlations. This experiment aims to build a better understanding of the robustness of quantum phenomena in quantum plasmonics.

In our experimental work, we generate pairs of polarization-entangled photons, propagating along two different paths, and interpose a plasmonic hole array in the path of one of the photons. This photon is thus converted into a plasmon, and the detected signal consists of plasmons reconverted into free-space photons after plasmon propagation over a few hundred nanometers in the hole array [29]. The quality of polarization entanglement between both outcoupled photons is measured and is representative of the effects of light-matter interactions during plasmon propagation. Whereas the work reported in [29] probed hole arrays with linear dispersion (close to light line), we use a plasmonic hole array that operates in a highly dispersive regime (close to the surface plasma frequency), to probe plasmon decoherence. We measure preservation of entanglement between photons even in this regime.

2.2 Generation of entangled pairs of particles

As a source of polarization-entangled photons, we used type-I spontaneous parametric down-conversion (SPDC), occurring in a pair of nonlinear bismuth borate

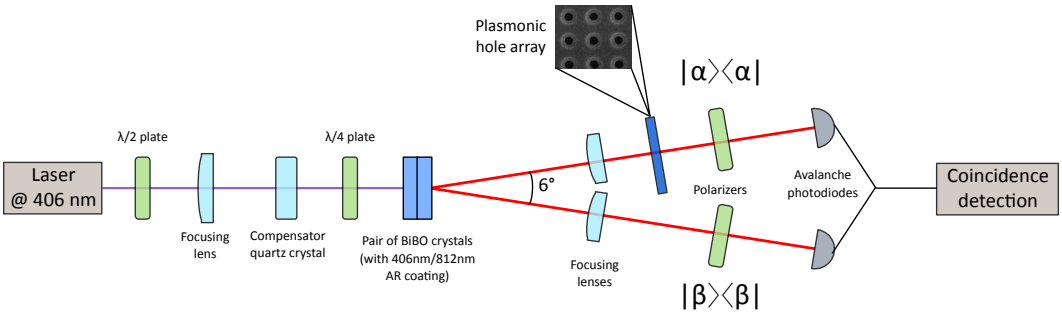


Figure 2.1: Experimental setup for the measurement of polarization-entanglement preservation. Pump photons at 406 nm are sent toward a pair of BiBO crystals and generate pairs of polarization entangled photons that propagate along two separate paths. Along the upper path, we can insert a metallic hole array, and measure the transmission of the entangled light that has been coupled to plasmons.

(BiBO) crystals. They are rotated by 90°with respect to each other and glued together [36] (Fig. 2.1), so that one crystal has its axis in the horizontal plane and the other one in a vertical plane. The pair of crystals is pumped by a laser diode emitting at 406 nm. The pump photons are linearly polarized at 45°with respect to the nonlinear crystal axis planes, so that type-I SPDC generates pairs of photons at 812 nm that are polarized parallel to either the horizontal direction or the vertical direction with equal probabilities. This setup generates polarization-entangled photons that, before their interaction with the environment $|E\rangle$ (plasmonic sample), can be described by the superposition state:

$$|\psi\rangle_{initial} = \frac{1}{\sqrt{2}}[|H,H\rangle + e^{i\Delta\phi_c}\,|V,V\rangle] \otimes |E\rangle], \tag{2.2.1}$$

where $\Delta\phi_c$ is a phase delay between the two polarizations, due to the birefringence of BiBO crystals.

The twin photons propagate in a horizontal plane, along the opposite edges of a cone whose apex angle is 6°. Each photon is focused towards a polarizer and a single-photon avalanche diode (SPAD). The detection of a photon by one of the SPADs is a projective measurement of its polarization state. A plasmonic hole array can be placed along one of the propagation paths, thus forcing one of the photons to be temporarily converted into a plasmon before eventually being detected.

In our experiment, in order to correctly estimate the influence of pure dephasing processes, we retain only coincident counts between the two SPADs, i.e., we consider only the case when both photons register counts at the detectors. In other words, when the hole array is in place, we do not record events in which a plasmon has

decayed through inelastic interactions with the electronic system – this is a well-understood mechanism for decoherence. On the contrary, we collect only photons from events in which the plasmon has survived. Such events, in principle, can be affected by elastic interactions or the inner structure of the plasmon quasiparticle.

In the general case, after propagation of the quantum entangled state and before applying any projective measurement, we can consider that the light has become entangled with the environment, and we can write:

$$|\psi\rangle_{final} = \frac{1}{\sqrt{|h|^2 + |v|^2}} \left[h\,|H,H\rangle \otimes |E_H\rangle + v e^{i\Delta\phi_c}\,|V,V\rangle \otimes |E_V\rangle \right], \tag{2.2.2}$$

where h and v are complex amplitude transmission coefficients for horizontal polarization $|H\rangle$ and vertical polarizations $|V\rangle$ respectively; $|E_H\rangle$ and $|E_V\rangle$ are environmental states, entangled with horizontal and vertical polarizations respectively.

By tracing over environmental states, one can obtain a reduced density matrix, from which a probability of a coincidence count can be computed:

$$\begin{aligned} P_{cc}(\alpha,\beta) &= \langle \alpha,\beta|\hat{\rho}_{reduced}|\alpha,\beta\rangle = \\ &= \frac{1}{1+\frac{|v|^2}{|h|^2}} \Bigg[\sin^2\alpha \sin^2\beta + \frac{|v|^2}{|h|^2}\cos^2\alpha\cos^2\beta + \\ &+ \frac{1}{2}\sin(2\alpha)\sin(2\beta)\frac{|v|}{|h|}\left|\langle E_V|E_H\rangle\right|\cos(\Delta\phi_E + \Delta\phi + \Delta\phi_c)\Bigg], \end{aligned} \tag{2.2.3}$$

where α and β are polarizers angles (see Fig. 2.1), $\frac{v}{h} = \frac{|v|}{|h|}e^{i\Delta\phi}$, $\Delta\phi$ being the phase difference between the complex amplitudes v and h, $\langle E_V|E_H\rangle = \left|\langle E_V|E_H\rangle\right| e^{i\Delta\phi_E}$, ϕ_E being the phase difference between the two environmental states, and α and β are the polarizers directions with respect to the vertical axis.

The first two terms can be obtained by classical analysis, whereas the last term is the so-called quantum interference term, which represents quantum mechanical nature of our system. Indeed, Eq. (2.2.2) describes a superposition state. The quantum interference term can be understood as the interference amplitude between the two terms of the superposition state when projective measurements are carried out on the two-particle state. The amplitude of this term depends on several factors. It depends sinusoidally on the polarizers directions, and is maximum for appropriate

choices of the polarizer directions verifying $\left|\sin(2\alpha)\sin(2\beta)\right| = 1$. This corresponds to the situation where the photonic parts of both terms in Eq. (2.2.2) are projected on a common state with equal amplitude. The amplitude of the quantum term is partially governed by the ratio $\frac{v}{h} = \frac{|v|}{|h|}e^{i(\Delta\phi+\Delta\phi_c)}$, which includes all perturbations inherent to the setup that affect the balance between the horizontal and the vertical polarization. Finally, we note here that the magnitude of quantum interference is also determined by the overlap between different environment states $\langle E_V|E_H\rangle = \left|\langle E_V|E_H\rangle\right| e^{i\Delta\phi_E}$, which represents quantum mechanical decoherence. The presence of $\Delta\phi_c$ in the last cosine factor of the quantum interference term shows that, in order to make judgements about quantum decoherence, one has to take a great care in eliminating or measuring phase differences between different polarizations. This can be done by inserting another birefringent element in the setup that will compensate the phase difference between the two polarizations. Optimization and alignment of our SPDC source included tweaking of a $\lambda/4$ plate (Fig. 2.1), which allowed us to experimentally eliminate $\Delta\phi_c$ in Eq. (2.2.3).

We now consider $h = v$, which represents equal probability of detecting horizontally or vertically polarized pairs of photons and is fulfilled when using a circular hole array. In the case of perfect coherence $\langle E_V|E_H\rangle = 1$ ($E_V = E_H$), we get the rather simple expression $P_{cc}(\alpha,\beta) = \frac{1}{2}\cos^2(\alpha-\beta)$. There is no entanglement between the photon state and the environment. This ensures the preservation of polarization entanglement between photons. By contrast, in the case of total decoherence $\langle E_V|E_H\rangle = 0$, we get a constant probability $P_{cc}(\alpha,\beta = 45°) = \frac{1}{4}$ regardless of α (if β is kept fixed at 45°). Both terms of the superposition in (2.2.2) are now incoherent, and quantum interferences vanish. The measured state can be considered as a statistical mixture of the two states $|H,H\rangle$ and $|V,V\rangle$ in equal proportions.

These considerations suggest a measure of quality of the entanglement, visibility, which we define simply as the visibility of the cosine curve described by $P_{cc}(\alpha,\beta = 45°)$ for the case when we keep polarizer β fixed at 45°(both polarizations contribute to the measurement) $V = \frac{P_{cc}^{max}-P_{cc}^{min}}{P_{cc}^{max}+P_{cc}^{min}}$, where P_{cc}^{min} is the minimum probability of coincidence count (rate in an experiment) and P_{cc}^{max} is the maximum count rate. For $h = v$ and $\Delta\phi_c = 0°$ visibility is equal to $V = \left|\langle E_V|E_H\rangle\right|\cos(\Delta\phi_E)$. From the above analysis, we get that $V = 100\%$ for fully entangled (quantum) light ($\langle E_V|E_H\rangle = 1$), and $V = 0\%$ for for a pure statistical mixture of polarizations ("classical" light, $\langle E_V|E_H\rangle = 0$). Note, that visibility of a cosine $P_{cc}(\alpha,\beta = 45°)$ is identical to the visibility of the cosine $P_{cc}(\alpha,\beta = 135°)$, hence we can use either one

of them, or use one versus another to validate the correctness of the measurement.

In addition to that, we performed Bell's inequalities violation measurements, where we use Bell's inequalities in CHSH form [37, 38]. We performed 16-point measurements in order to calculate Bell's parameter S, comparing our experimental measurement with the best possible prediction of any classical local hidden variable theory (LHVT). $S > 2$ indicates the impossibility of the explanation by any LHVT.

We characterized our SPDC source (Fig. 2.2) without plasmonic samples, measuring visibility on the order of $V = 99\% \pm 1\%$ and $S = 2.81 \pm 0.02$, which is just a standard deviation away from the maximal theoretical value $S_{max} = 2\sqrt{2} \approx 2.83$. From this, we conclude that we have high quality pairs of entangled photons. In the next section, we investigate the influence of the insertion of a plasmonic hole array on the quality of entanglement between photons, as defined by the previous measurement procedures.

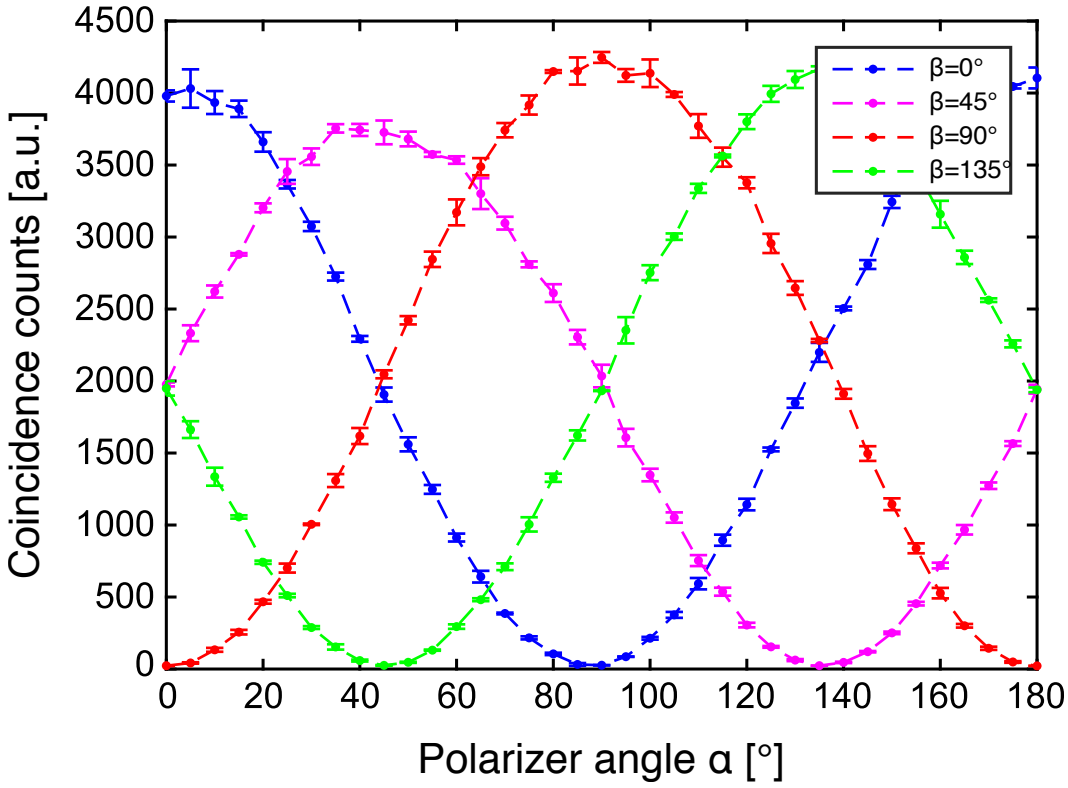


Figure 2.2: Calibration of the setup, and entanglement between produced pairs of photons. Number of coincidence counts as a function of polarizers angles without plasmonic sample (solid lines are fits to cosine). The visibility of the different cosine fits is nearly equal to one, indicating quasi-perfect entanglement between the photons of our SPDC source.

2.3 Elliptical hole array in the linear dispersion regime

In this part, we investigate the preservation of quantum entanglement when sending photons through a hole array when excited plasmons are "photon-like", i.e., when choosing an operating point on the dispersion relation close to the light line. To do so, we designed, fabricated, and characterized a hole array sample dedicated to the excitation of plasmons at the interface between gold and glass. As a warm-up experiment, we study the influence of the hole geometry on the preservation of quantum entanglement. This experiment gives us a reference point to initiate a comparison with plasmon excitation in the highly-dispersive regime.

To study geometrical effects, we used a plasmonic array of elliptical holes milled through a 200 nm thick layer of gold deposited on a glass substrate, by using a focused ion beam with purposely-introduced astigmatism. The dispersion relation of the gold-glass interface was computed from an analytical modal dispersion model (see Fig. 2.3). A first choice for the hole dimensions and the array periodicity was made after numerical simulations of the structure designed to enhance extraordinary transmission at 812 nm – the wavelength of our down-converted photons. The sample we used had a size 1 mm x 1 mm and was fabricated in a clean room environment. Its transmission was enough to detect a satisfying level of signal. The hole shape in these arrays is close to elliptical, with axes equal to 240nm and 190nm (see Fig. 2.4(A)). The difference between those two dimensions makes the optical transmission of this sample polarization-dependent (Fig. 2.4(B)). This sample has linear SPP dispersion (Fig. 2.3), so that we could focus on hole geometry effects only in this experiment.

First, we performed a measurement using polarization entangled light (Fig. 2.4(C)) (rotating a sample in such a way that hole array eigenmodes directions were along the vertical and horizontal polarizations). The transmission of the first channel (considered as an influence of the environment on the system) being now polarization dependent, we expect the two terms of the state in Eq. 2.2.3 to be differently affected by the insertion of the plasmonic array, and thus a reduction of the visibility. In both, we found the visibility of curves, corresponding to the mixture of eigenmodes ($\beta = 45°$ and $\beta = 135°$ on Fig. 2.4(C)) of hole array, to be $V = 86\% \pm 5\%$.

However, we need to determine to what extent this reduction is caused by the quantum mechanical decoherence (through a decrease of the overlap factor $V = \left|\langle E_V | E_H \rangle\right| \cos(\Delta\phi + \Delta\phi_E)$, provided $\Delta\phi_c = 0°$) or by the modification of the complex transmission ratio $\frac{v}{h}$, which is a purely classical effect. In order to do

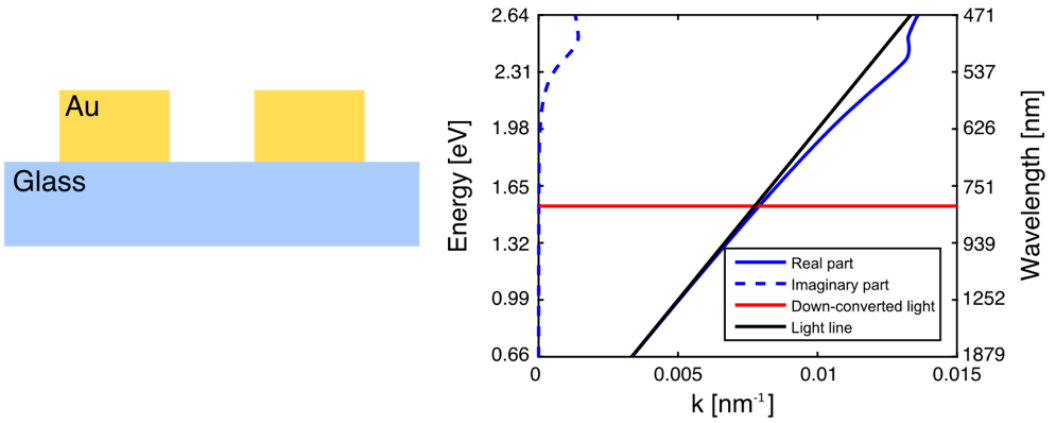


Figure 2.3: Design of plasmonic elliptical hole array. Cross-sectional schematic and dispersion relation of elliptical hole arrays for SPPs supported at the gold/glass interface. At the wavelength of the down-converted photons (812 nm, as shown by the red solid line), the dispersion is "photon-like", i.e., linear, and very close to the light line.

so, we performed a control experiment using classical light, in which we generated just unentangled polarized photon pairs, either in the pure $|H,H\rangle$ state, or in the pure $|V,V\rangle$ state.(Fig. 2.4(D)). To record Fig. 2.4(D), we first fixed the hole array at exactly the same rotation as used in Fig. 2.4(C) and transmitted horizontal and then vertical polarizations (H0, V0 curves on Fig. 2.4(D) correspond to $\beta = 0°$ and $\beta = 90°$ on Fig. 2.4(C)). The amplitude ratio between the two sine curves is an indication of the ratio of the transmission amplitudes of vertical and horizontal polarizations through the hole array $\frac{|v|}{|h|}$. Then we rotated the hole array by $\beta = 45°$ and repeated the measurement (H45, V45 curves on Fig. 2.4(D) correspond to $\beta = 45°$ and $\beta = 135°$ on Fig. 2.4(C)). In this configuration, the hole array behaves as a birefringent plate whose axis are at $\beta = 45°$ with respect to the polarization of the incident photon. The dephasing and the transmission ratio between the two eigenpolarizations can be related to the azimuth and the ellipticity of the output polarization of the photon. From the measurement with classical light and by fitting the different plots of Fig. 2.4(D), we determined $\frac{|v|}{|h|} = 5.2 \pm 0.3$ and $\Delta\phi = 48° \pm 4°$. Substituting these values to Eq. 2.2.3 and fitting 2.2.3 to curves on Fig. 2.4(C) gives $\left|\langle E_V|E_H\rangle\right| = 1.0 \pm 0.1$ and, $\Delta\phi_E = 1° \pm 4°$ so that within experimental error we did not detect quantum mechanical decoherence.

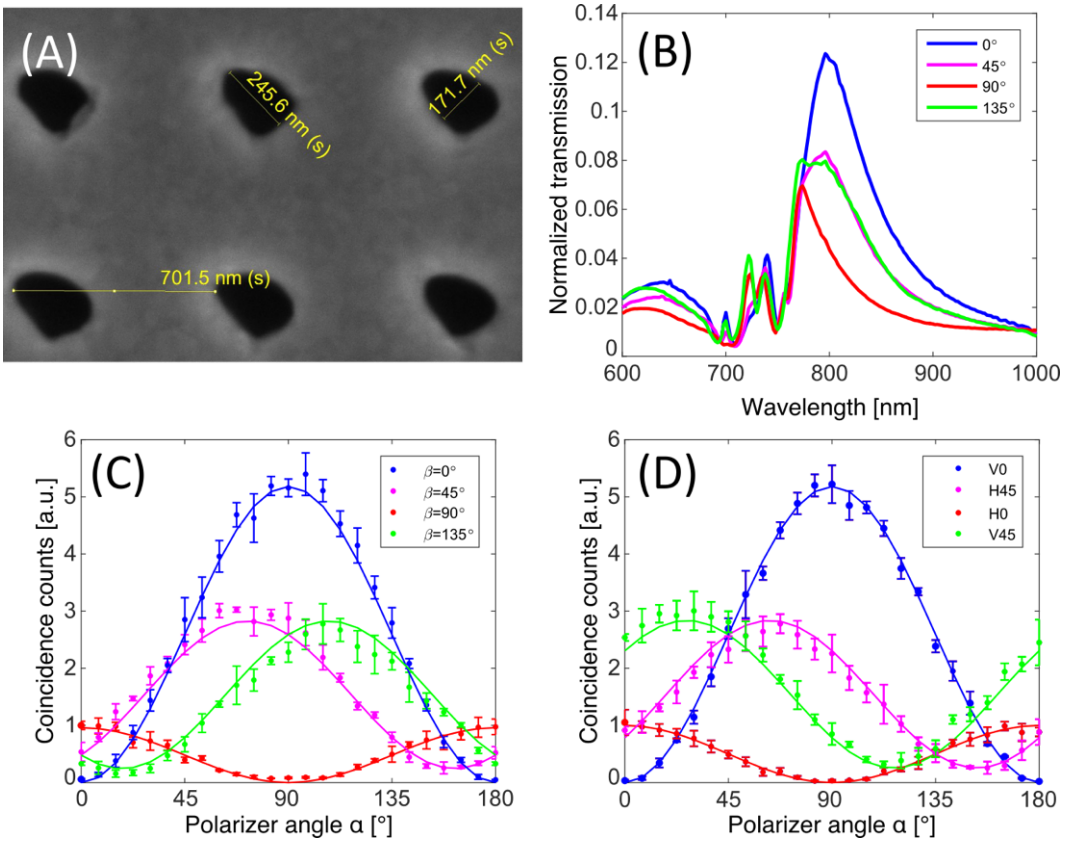


Figure 2.4: Study of entanglement preservation with the elliptical hole array. Elliptical hole array used to study the influence of hole geometry on the preservation of photon entanglement: (A) SEM image, where the orientation of the minor and major axis can be seen oriented at ±45°; (B) Transmission factor of the device for different polarizations of incident light. The holes have no rotation symmetry anymore, so that the transmission maximum varies between 12.5% for horizontally polarized light and 7% for vertically polarized light. (C)&(D) Normalized number of coincidence counts as a function of polarizer angles in the presence of an elliptical plasmonic hole array (solid line represents fit to the full model) in two configurations : (C) With entangled photons and for different fixed directions of the polarizer beta; (D) With classical light and similar choice for beta. The plots with $\beta = 45°$ and $\beta = 135°$ show a similar decrease in visibility in both configurations, indicating that this results from a purely classical effect.

2.4 Hole array with nonlinear dispersion

In this section, we investigate the same polarization entanglement process using this time single SPPs that propagate in a hole array, and most importantly in a regime of highly nonlinear dispersion, far from the light line, in an attempt to reveal effects of pure dephasing on decoherence through a decrease of entanglement visibility. In the highly dispersive regime, the quasiparticle confinement at the

metal/dielectric interface is much stronger due to the reduction in the plasmon wavelength. In other words, the plasmons' excitation wavefunction has a much larger overlap with the electronic degrees of freedom in the metal. As a consequence, plasmons in the highly dispersive regime have a generally higher rate of interaction with the electronic system than in the case of materials with linear dispersion. Hence, in addition to shorter total decoherence time T_2, one can expect a shorter pure dephasing time T_2^* (which is the relevant time scale probed by our experiment). The group velocity of highly dispersive plasmons can be an order of magnitude smaller than for plasmons in the "photon-like" regime, so that these plasmons propagate for a longer time t_p (even if the propagation distances are the same). Therefore, for comparable experiments (i.e., similar propagation distances), strongly-confined plasmons are expected to experience greater decoherence and thus exhibit weaker quantum interference than photon-like plasmons.

In order to probe the highly dispersive regime in an analogous situation, we excite plasmons at the interface between gold and amorphous silicon. Amorphous silicon has a higher dielectric constant than glass, moving the surface plasma frequency close to the frequency of entangled photons (see Fig. 2.5). The use of a higher index material leads both to a stronger SPP confinement and significantly enhanced light-matter interaction compared with plasmons propagating at the interface between gold and glass (in a "photon-like" regime): a 6-fold increase of the plasmon wave vector, a 12-fold reduction of their group velocity ($0.05c$ versus $0.59c$, where c is the speed of light in vacuum), and more than 100-fold reduction in the absorption length (200 nm vs 27 μm).

By performing finite-difference time-domain simulations with Lumerical software, we obtained transmission spectra of a variety of nano-hole array structures with different geometries and periodicities. The positions of the transmission maxima were considered as indicative of the resonance energies and of the dispersion relation of the surface plasmons. An initial choice for the range of geometrical parameters (hole dimensions, array periodicity, films thicknesses) in our experiment was made using the parameters of the simulated structures that exhibited enhanced extraordinary optical transmission at 812 nm – the wavelength of our down-converted photons.

However, additional analysis was still necessary to confirm that the fabricated structure correctly reproduces the initially simulated behavior and operates in the non-linear dispersion regime, far from the light line.

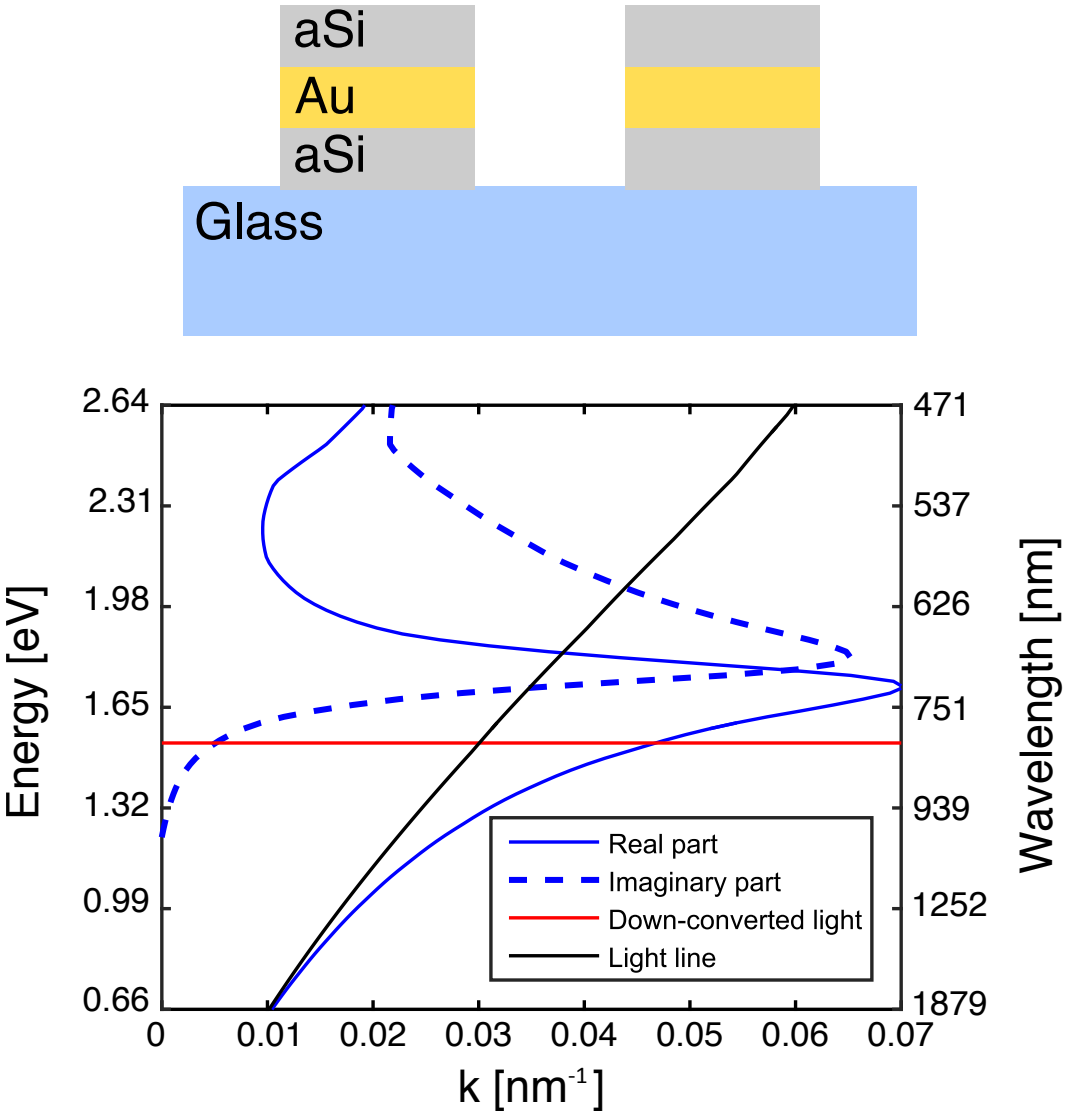


Figure 2.5: Plasmonic hole array design in our experiment: cross-sectional schematic (top) and dispersion relation of circular hole arrays for SPPs supported in the silicon/gold/silicon structure, which exhibits strongly nonlinear dispersion at 812 nm (bottom). The red line shows the energy of the plasmons excited in our experiment.

The direct comparison of the transmission spectrum at normal incidence of a single structure with the simulated data is made difficult by the overall extreme sensitivity of surface plasmons to geometric defects and roughness of the different layers (intrinsically related to the limits of the fabrication process), that were not included in the numerical model of the structure (performed using only perfectly flat layers). These effects result in a broadening of the transmission peaks in the experimental data, which complicates the identification of the different resonances.

In order to overcome this difficulty, we fabricated a set of hole arrays with different periodicity ranging from 600 nm to 1000 nm, following the same fabrication procedures (resulting in the same film thicknesses and hole diameters). The period of the structures defines the wavevectors of plasmons that can couple to the structures. Overall, and thanks to the reproducibility of the fabrication processes, the acquisition of transmission spectra at normal incidence for these different structures can be seen as an indirect measurement of the dispersion relation of the structure. The next step of our approach consists of comparing these experimental transmission spectra of the different hole array structures with the analytical dispersion relation displayed on Fig. 2.5, and to check if the evolution of their respective characteristics (in practice, evolution and spectral shift of the resonance maxima) are compatible with each other.

Fig. 2.6 displays an interpolation of a set of transmission spectra obtained experimentally for the nine different structures (color map). The operating point of our main experiment, with a structure of periodicity of 850 nm and a plasmon energy around 1.52 eV, is marked with a black cross. The map displays different branches, i.e., different sets of plasmon energies and wavevectors that fulfill a resonant condition, all excited in parallel when performing our experiment, and corresponding to different folded portions of the dispersion relation of the structure.

As a reasonable approximation, the resonance condition fulfilled by plasmons propagating in the structure can be expressed as:

$$\sqrt{2}p = d + \frac{2\pi n}{k}, \tag{2.4.1}$$

where p is the period of the structure, d is the hole diameter, k is the plasmon momentum, and n is an integer. The factor $\sqrt{2}$ takes into account the fact, that in our experiment, hole array plasmon eigenmodes propagate along diagonal directions. SEM images and measurements show that two values must be considered for the holes diameter. The diameter of the holes in the upper aSi layer is $d = 430$ nm and is associated to plasmons propagating along the upper aSi-gold interface. The FIB milling process produced smaller holes in the lower aSi layer, with a diameter $d = 300$ nm, this value being the time associated to plasmons propagating along the lower gold-aSi interface. Using the analytical dispersion relation of Fig. 2.5 to relate the plasmon momentum k and the plasmon energy, we plot the relation (2.4.1) on Fig. 2.6 for different values of n and different values of the hole diameter. We see that we have great agreement between our experimental data and the model of

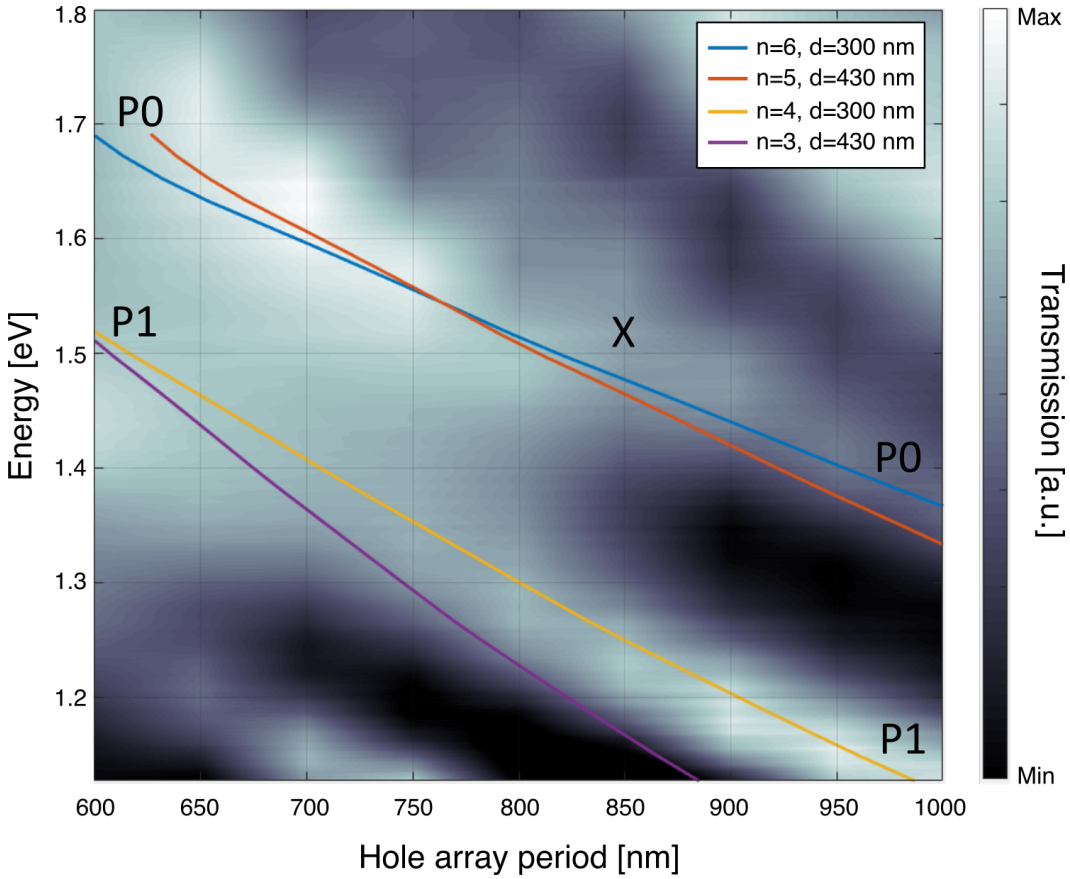


Figure 2.6: Investigation of the evolution of the transmission resonances for various structures. The transmission spectra at normal incidence of nine different hole array structures, with periods ranging from 600 nm to 1000 nm have been experimentally measured. The transmission amplitude is displayed as a color map, and as a function of both the energy (vertical axis) and of the periodicity of the hole array (horizontal axis). Experimental data have been interpolated between the nine sets of data points. The different white branches displayed on the colormap correspond to transmission resonances. The branch corresponding to the plasmon resonance exploited in the main experiment is labeled as P0 and the operating point of the experiment (structure periodicity of 850 nm, plasmon resonance at 812 nm) is marked with a black X. Another branch of plasmon resonance at lower energies is identified as P1. These branches were fitted using relation 2.4.1 for different sets of parameters (solid color lines). For both P0 and P1, agreement between experimental data and model is good and allows us to determine the plasmon wavevectors at resonance.

(2.4.1) for different branches, and in particular for the plasmon resonance that we exploit in our experiments (branch labeled "P0" of the colormap including the black cross) when considering $n = 6$ and $d = 300$ nm for the bottom plasmon (blue solid line) and $n = 5$, $d = 430$ nm for the top plasmon (red solid line). We emphasize that, while we can in theory expect two different resonances for the two plasmons

propagating in the structure, or in other words two sets of branches, in practice, both are significantly broadened and merge into each other. It is not possible to resolve them separately. and "P0" displays only one local transmission maximum although it contains the contribution of both top-propagating plasmons and bottom-propagating plasmons.

The parameters n and d allow us to identify the proper fraction of the dispersion relation that has to be compared with our transmission measurements. Using relation (2.4.1) and the parameters $n = 6$ and $d = 300$ nm, we now plot on Fig. 2.7 our experimental transmission measurements as a function of plasmon energy and $k = \frac{2\pi n}{\sqrt{2}p-d}$, the wavevector corresponding to the excitation of the plasmons belonging to P0, and we superimpose on the same plot the dispersion relation of Fig. 2.5. This choice of plot highlights the agreement between or experimental data and the initial numerical design of the structure.

We finally check the robustness of our model by unfolding further the data of the transmission spectra and comparing them with the analytical dispersion relation on a larger wavevector scale. We consider, for each of the nine transmission spectra obtained with the nine experimentally characterized structures (vertical cross sections of Fig. 2.6), the wavelengths of the two local transmission maxima corresponding to both branches P0 and P1. As previously discussed, each one of these transmission peaks merges the contribution of the two plasmons propagating in the structure, either along the top or along the bottom gold-aSi interface. On Fig. 2.8, we compare our experimental measurement of the different transmission maxima for the different structures with the analytical dispersion relation (black solid line). Each transmission maximum that was experimentally measured in the different spectra is represented as two data points placed at the same plasmon energy, but at the two different wavevectors corresponding to the top plasmon and the bottom plasmon. The values of the two plasmon wavevectors were derived using the relation 2.4.1 with the two sets of parameters (n, d) associated to P0 or P1. We can see that this unfolding procedure shows great agreement between our experimental spectra and the dispersion relation for the two major resonance branches that could be observed.

We can conclude that, taking into account geometrical defects and possible small discrepancy in optical constants between experimental and literature values used in simulations, the experimental dispersion relation is compatible with our analytical model, and that the chosen operating point at 812 nm for the structure with a periodicity of 850 nm corresponds with a surface plasmon following the

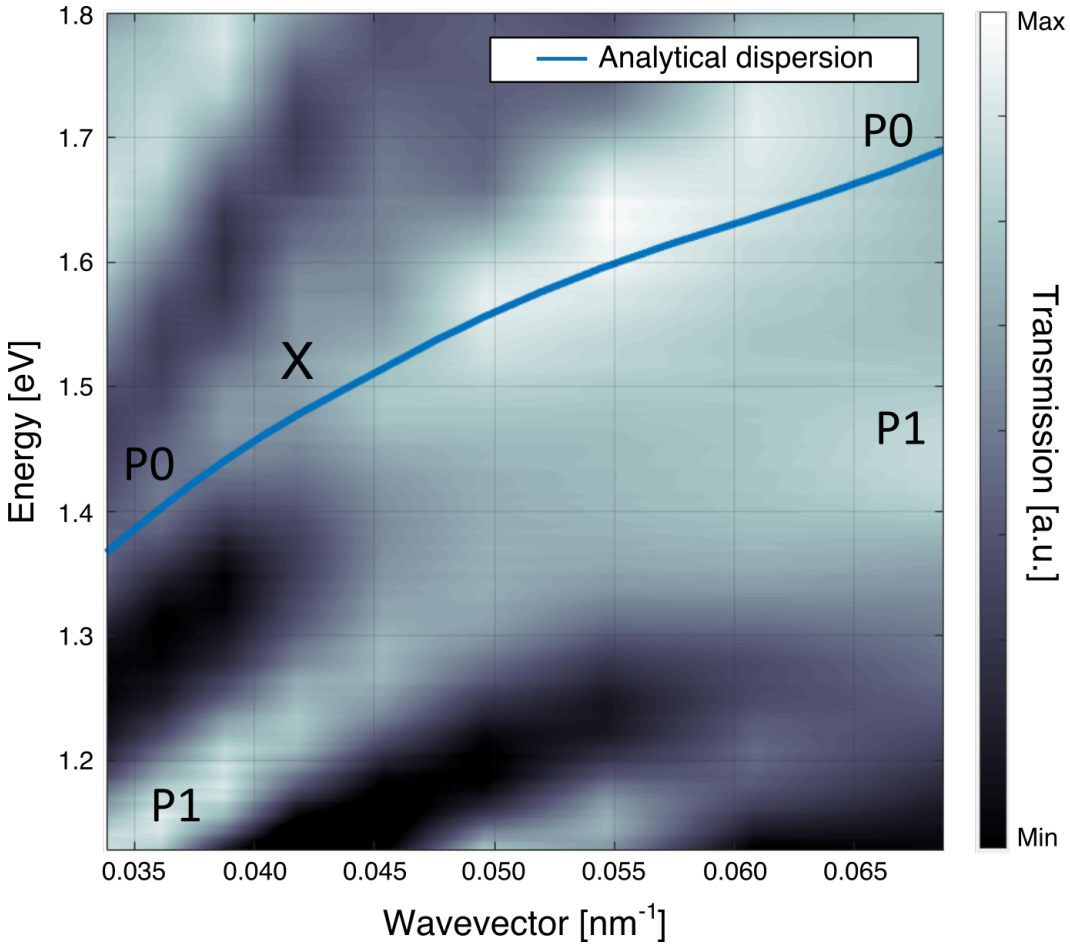


Figure 2.7: Comparison of the analytical dispersion relation with the transmission measurements. The colormap of Fig. 2.6 is represented this time as a function of the wavevector, using the conversion between periodicity of the structures and resonant wavevectors given by relation (2.4.1) with $n = 6$ and $d = 300$ nm. The two branches P0 and P1 are identified on the plot, the operating point of the main experiment is marked with a black X. P0 is well fitted by a segment of the dispersion relation that corresponds to a non-linear highly-dispersive regime for the plasmons (blue solid line), and the position of the operating point is in agreement with the initial design of the experiment.

intended resonance far from the light line, in the highly-dispersive regime.

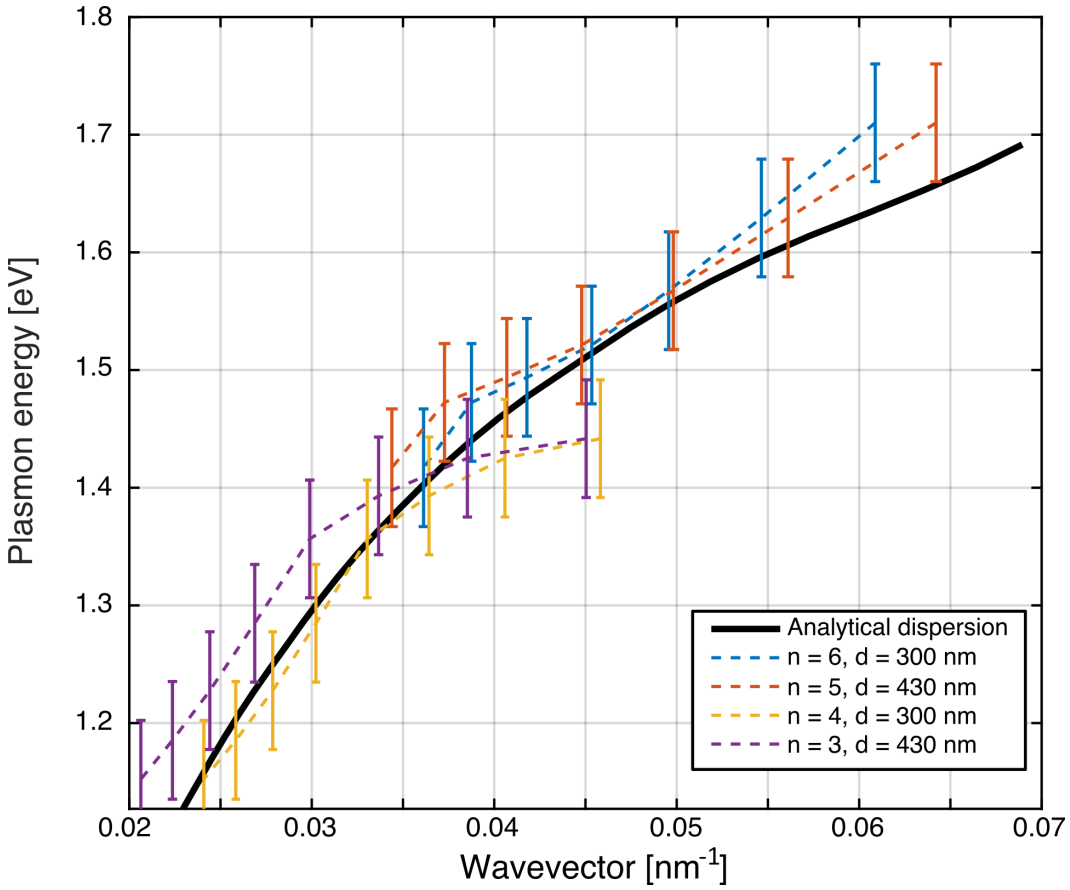


Figure 2.8: Unfolding the transmission spectra to measure indirectly the dispersion relation. We compare the analytical dispersion relation (black solid line) of the structure with the position of the local transmission maxima measured experimentally with structures of various periodicities (vertical cross sections of Fig. 2.6). For each measured transmission maximum, two values of associated wavevector are derived, by using (2.4.1) with two sets of parameters. If the transmission maximum belongs to the branch P0, the two wavevectors values are calculated using $n = 6$, $d = 300$ nm (dashed blue line) or $n = 5$, $d = 430$ nm (dashed red line). If the transmission maximum belongs to the branch P1, the two wavevectors values are calculated using $n = 4$, $d = 300$ nm (dashed yellow line), or $n = 3$, $d = 430$ nm (dashed purple line). This shows that experimental transmission spectra are in great agreement with the expected dispersion relation.

2.5 Results and discussion

All of the analysis above allowed us to identify and fabricate a sample with optimal design, which is a 2 mm × 2 mm hole array configured in a three-layer structure (50 nm of amorphous silicon – 100 nm of gold – 50 nm of amorphous silicon) and a periodicity $P = 850$nm (Fig. 2.9(A)). We find a plasmon-enhanced transmission peak at the desired 812 nm wavelength (Fig. 2.9(B)). In this hole array,

SPPs excited on the top and bottom gold surfaces are uncoupled and have the same dispersion.

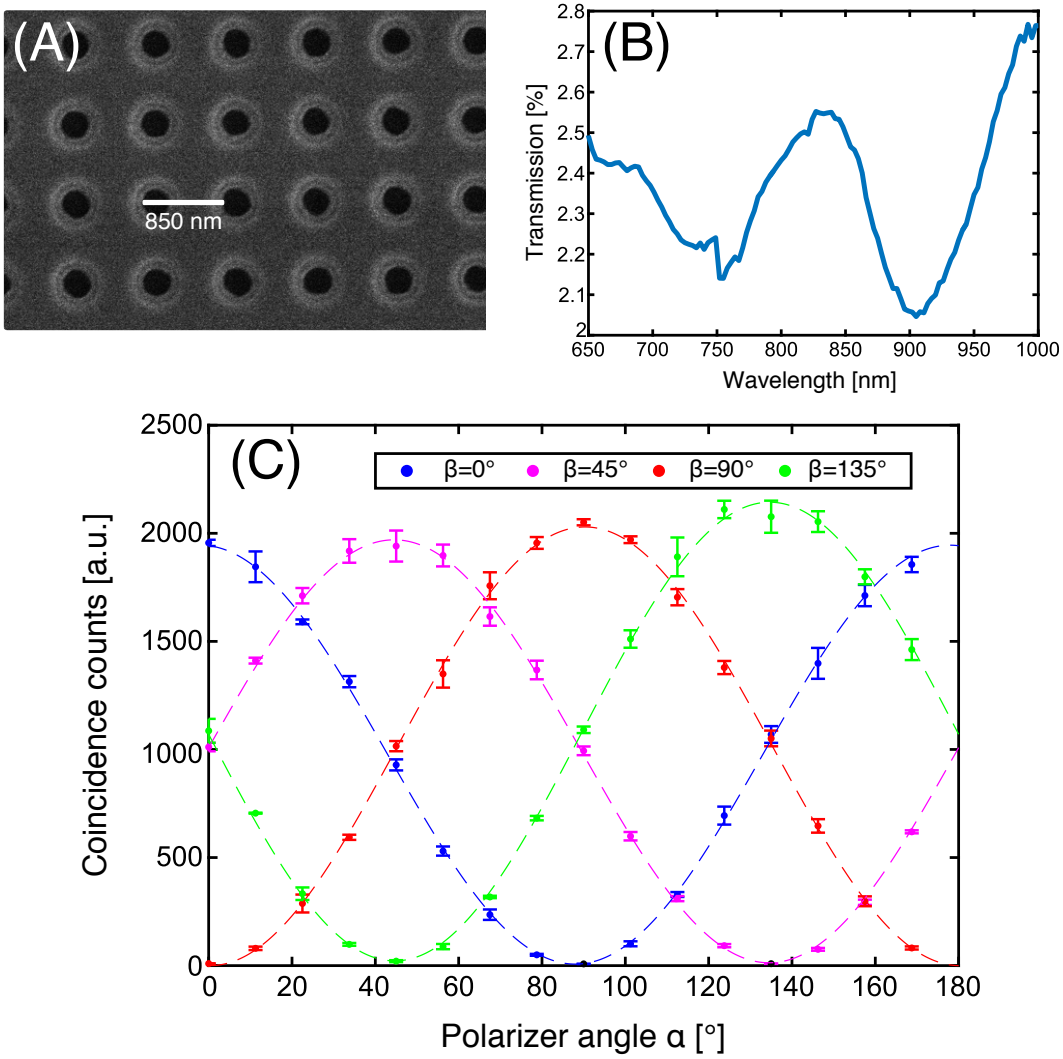


Figure 2.9: Hole array for plasmons in higly-dispersive regime. (A) SEM image of the sample. The period of the 2D array is 850 nm. The different material species experience different milling rates that slightly affect the shape of the holes' perimeter. (B) Transmission spectrum of the hole array. With the holes being circular, there is no polarization dependence. The broadening of the transmission feature around 810 nm can be attributed to the imperfect shape of the holes. Note that the plasmons experience a significant absorption. (C) Number of coincidence counts as a function of polarizer angles in the presence of the hole array in a highly-dispersive regime (solid lines are fits to cosine). Whatever the choice of β is, and even when placed at 45°or 135°, the visibility of quantum interference remains almost equal to one, indicating near perfect preservation of entanglement between particles.

Thanks to the rather large size of the sample, we were able to collect a large

portion of transmitted light and hence improve our statistics, even in the presence of considerable losses and with intrinsically low transmission (Fig. 2.9(C)). We recorded a visibility of $V = 98\% \pm 2\%$ and Bell's number $S = 2.83 \pm 0.04$: this measurement implies that even in the highly-dispersive regime, the entanglement is perfectly preserved and no quantitative signs of pure dephasing could be detected. Through numerical computation of the dispersion relation, we can estimate the propagation time to be on the order of $t_p \sim \frac{1.2\mu\text{m}}{0.05c} \sim 80\,\text{fs}$, based on the distance between two diagonally separated holes (this is the relevant distance in our case, since eigenmodes of our hole array are diagonally oriented, similarly to [29]). It is worth noting, that this time is much longer than the literature reported value ([32, 39]) of SPP total dephasing time $T_2 = 20 - 30\,\text{fs}$ (in agreement with our own estimations from either simulated, or experimental data). This confirms the fact discussed above, that in our experiment relevant decoherence mechanism is pure dephasing, caused by elastic scattering of SPPs, rather than dephasing due to inelastic scattering (i.e., population decay). Since, in our experiment, we measured visibility to be almost 100% (within the margin of experimental precision), we conclude that in our system, pure dephasing is a remarkably slow process compared to absorption. SPP pure dephasing time T_2^* must be much larger than propagation time, and we put a conservative order of magnitude estimate of 100 fs on its lower bound, which is similar to the value reported in [32].

We note, however, that this time could be in practice much higher, as our experiment remarkably reports no quantitative trace of quantum decoherence. Performing the same experiment in an even more highly dispersive regime could hypothetically allow us to make the decoherence process eventually visible at some extent. This would make the degradation of the fringes' visibility capable of being modeled, one of the parameters of such a model being the pure dephasing time, that could be more precisely estimated. The design of such an experiment is however fundamentally limited by the high level of absorption that coexists with the enhancement of light-matter interactions. It dramatically reduces the signal level at the output of the plasmonic path progressively to almost zero.

In summary, we have examined the influence of plasmon dispersion on the quantum decoherence properties of surface plasmons. The excitation of highly-dispersive plasmons did not result in the reduction of the quality of a single-particle quantum state for transmitted light. Plasmons excited in hole arrays are found to preserve quantum mechanical correlations, even in the presence of extreme disper-

sion near the plasmon resonance and strong absorption. Moreover, the focus of our measurements is on the elastic dephasing processes, which consideration is commonly neglected in quantum optics modeling due to its supposed insignificance. Our findings provide experimental proof for such an assumption, and also emphasize the difference between decay and decoherence. Thus, we conclude that despite being lossy, plasmonic structures may find applications in the realms of quantum technology, where the power of extreme light confinement can be effectively leveraged.

Chapter 3

FIRST-PRINCIPLES STUDY OF PASSIVATION OF SULFUR VACANCIES IN MoS_2

3.1 Introduction

2D materials have been fascinating scientists for a long time due the emergence of new physics when transitioning from bulk to monolayer materials. Discovery of graphene and the subsequent Nobel Prize in Physics have put the topic of 2D materials at the forefront of experimental and theoretical research since then.

One of the important categories of 2D materials is transition metal dichalcogenides (TMDCs). Their unique electronic, optical, and mechanical properties give them a wide variety of applications in electronics and optics [40, 41].

TMDCs are Van-der-Waals materials. In bulk, these materials are comprised of monolayers that have very strong bonding in-plane, but between each other, monolayers are attracted through weak Van-der-Waals forces, making possible the isolation of individual 2D monolayers.

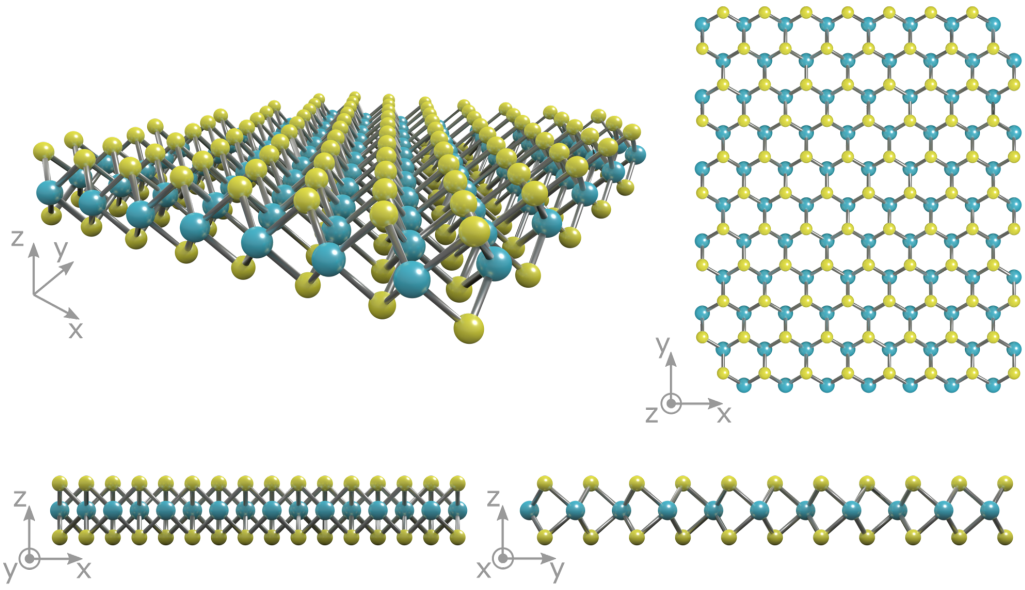


Figure 3.1: Crystalline structure of a MoS_2 monolayer, blue balls are molybdenum atoms, yellow circles are sulfur atoms. Illustration adapted from https://www.ossila.com/pages/molybdenum-disulfide-mos2.

In this chapter, we focus our attention on a very popular TMDC - molybdenum disulfide (MoS_2). It is a semiconductor with an indirect bandgap in bulk, that becomes a direct bandgap in a single monolayer, giving it significant advantage over graphene for optoelectronic applcations. A monolayer of MoS_2 consists of a layer of molybdenum atoms sandwiched between two layers of sulfur atoms (Fig. 3.1).

Without any treatment, exfoliated 2D MoS_2 usually exhibits low photoluminescence due defect-mediated non-radiative recombination. A popular method for alleviating this problem is passivation using bis(trifluoromethane)sulfonimide (colloquially TFSI) superacid, which results in an almost 200x increase in photoluminescence [42].

However, microscopic understanding of this mechanism is still lacking, and this creates obstacles for a controllable design of defects in many optoelectronics applications of TMDCs. Several explanations are proposed [43, 44], but a scientific consensus has not been reached yet. There are also other suggested ways of passivating defects in this material, like ambient annealing [45].

Superacids are defined by their extreme ability to donate a proton. This suggests that during TFSI treatment defects in 2D MoS_2 could be "healed" through hydrogenation or protonation. Most common defects in MoS_2 are sulfur vacancies [46]. Thus, the goal of this chapter is to study hydrogenation/protonation of sulfur vacancies in 2D MoS_2 to build microscopical understanding of this mechanism through first-principles calculations method.

3.2 Computational methods

In, current work, we are interested in understanding underlying mechanisms at a microscopic level. One of the most powerful theoretical tools are is computations from first principles (or ab initio calculations). In such a setting, we do not assume any empirical knowledge about the material (except for types of atoms and their rough geometrical positions (to initialize computations)), and we use the most comprehensive and fundamental model of interaction between particles/quasiparticles based on quantum mechanics. In our case, we are exclusively interested in electronic and optical properties of the material, hence, we can concentrate our modelling around electrons.

In non-relativistic quantum mechanics, the behavior of the physical system is governed by a Schrodinger equation. Unfortunately, wavefunction of a material is an extremely complicated mathematical object, since it is a function of positions of

all of the electrons in the system, and number of electrons is astronomically high (remember Avogadro number).

Positions of atoms in crystalline materials can be calculated by periodic translation of primitive cell. Hence, it is very convenient to introduce periodic boundary conditions on wavefunction of such materials. Even in such a simplified setting, it is still too complicated (and some may argue not absolutely necessary) to model full material wavefunction.

Density functional theory (DFT) provides an extremely useful framework for first-principles computations. In fact, papers that lay its foundations are among some of the most cited in physics [47, 48].

DFT is based on a variational principle of quantum mechanics. In DFT, we construct kinetic and potential energy functionals from the density of electrons, which is a function of only three coordinates, and hence it is much simpler than a wavefunction. We then vary these functionals and obtain Kohn-Sham equations for quasi-electrons, which are independent and interact with each other only through exchange-correlation energy, which is negative and a purely quantum mechanical effect. Since classical electrodynamics fails to explain stability of materials (see Earnshaw's theorem), this exchange-correlation energy is crucial for our understanding of solids, and sometimes it is called the binding glue of matter.

In principle, if we knew exact exchange-correlation functionals, DFT would the exact theory of ground state. Unfortunately, knowing them exactly is probably as hard as just solving the many-body quantum mechanical problem directly. However, efficient approximations have been under development since the beginning of DFT.

One additional approximation that is commonly done in DFT is the usage of pseudopotentials (PPs). Since, in most cases, properties of materials are determined by valence electrons, it appears to be reasonable to attempt to somehow exclude core electrons from calculations. The pseudopotential technique allows us to compute wavefunctions and energies for valence electrons only. PPs are built in such a way that wavefunctions computed for single atoms using PPs are equal to exact wavefunctions computed using all electrons outside of some cut-off radius. Our hope is then when we plugin such PPs into materials consisting of multiple types of atoms, we would still be able to compute properties of such materials accurately. There are many different flavors of pseudopotentials available for DFT practitioners [49], even though PPs is a very active area of research among theorists.

One should be careful in selecting appropriate pseudopotentials for their application, as there are many known failures due to inability to capture relevant portions of physics [50, 51].

In this work we use Quantum Espresso – a popular open-source software package for DFT calculations [52, 53]. It uses plane-wave basis for expanding Kohn-Sham wavefunctions, which makes it extremely suitable for calculating properties of periodic systems, such as crystals. In this approach, after substituting plane-wave expansions into the Kohn-Sham equation, we get a huge system of linear equations (under some assumptions, see the next paragraph), tying together coefficients of expansion with eigen-energies of the Kohm-Sham orbitals. Since we are only interested in occupied states (and probably a few unoccupied), we only need to find the lowest eigen-energies that correspond to these states. In Quantum Espresso, this is done via the Davidson diagonalization method, which was originally developed for such purposes [54].

Since the Kohn-Sham equation is extremely non-linear (exchange-correlation terms are functionals of density), Quantum Espresso uses an iterative method for solving it. Usually, wavefunctions are initialized as a superposition of atomic wavefunctions plus some noise. Then, density is computed from them, and finally we construct necessary functionals from density. These functionals are considered to be fixed for current iteration. Then, under this assumption we calculate plane-wave expansion coefficients for selected orbitals using the Davidson diagonalization procedure. We use these coefficients to construct new wavefunctions, density, and functionals, and we repeat this procedure several times until some convergence criteria are met (e.g., total energy and/or eigen-energies). Such a procedure is commonly called self-consistent calculation.

Quantum Espresso also allows us to perform optimization of the geometrical structure of the material both in terms of a unit cell size and relative positions of atoms within the unit cell. This is done through minimization of the total energy of the system using Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm.

In principle, the number of plane waves is infinite, which is, of course, impractical. To make calculations possible, we limit it to some finite value, which is defined by cut-off energy - maximum kinetic energy of a plane wave in our expansion. Determining an acceptable value for cut-off energy is done through varying this parameter until some target quantity (say, total energy, or positions of conduction band minimum and valence band maximum) stops changing significantly

(convergence test).

Similarly, we perform our calculations in reciprocal space, and we can only do that with limited number of k-points. In some cases (such as metallic systems), finite k-points' grids introduce significant irregularities into quantities that depend on occupations of partially filled states. In this case, we use Gaussian smearing of occupations. Acceptable size of k-points' grid and amount of smearing are interdependent and should be adjusted together in convergence tests.

Since we study 2D materials in this work, we have to introduce vacuum space between periodic images of 2D sheet in the z direction. The size of this vacuum is another parameter that must be carefully determined during convergence tests, so that interaction between images is so low, giving us effectively independent 2D sheets of material.

In this work, we are concerned with properties of defects, which are in principle some random perturbations in the crystalline lattice. Periodic boundary conditions pose a challenge to modeling them. Here we employ the supercell method, in which we construct a big unit cell (called a supercell) out of primitive cells of material and we put one defect in such a supercell. The size of supercell should be chosen in such a way that interaction of defects in different images is at an acceptable low level.

The overall procedure for performing accurate calculations using Quantum Espresso contains following steps.

1. Selection of pseudopentials.
2. Convergence tests in number of k-points, cut-off energies, occupations smearing, vacuum size, supercell size.
3. Geometrical optimization (atoms positions and/or unit cell parameters).
4. Self-consistent DFT calculation to obtain high-quality charge density (remember, in DFT charge density determines energy functionals).
5. Non-self-consistent calculation (we perform only one iteration of self-consistent loop) with charge density from self-consistent run and denser k-point grid. This calculation allows us to obtain more accurate Kohn-Sham wavefunctions and eigen-energies.
6. Postprocessing step using computed Kohn-Sham wavefunctions: band structure, density of states, dielectric function, etc.

In this work, we use norm-conserving pseudopotentials [55] of PBE flavor [48]. We determined that 25Å of vacuum in z direction is sufficient to simulate 2D sheets of material.

We investigated influence of the supercell size on properties of defective 2D MoS_2 (see Fig. 3.2). We determine that 4x4 supercell (in the xy plane) is sufficient to have qualitative understanding of influence of defects, even though dispersion of defect states is visible on bandstructure (Fig. 3.2a). If one aims at predicting material properties as accurately as possible, then the 5x5 supercell might be necessary as it provides enough spacing between defects so that interaction between them is minimal, as in reality defect states should be flat (Fig. 3.2b). However, DFT calculations scale as $O(N^3)$ with number of electrons N, so as one goes from 4x4 to 5x5 supercell, computational cost increases by at least 4 times, and in reality, even more depending on particular implementation.

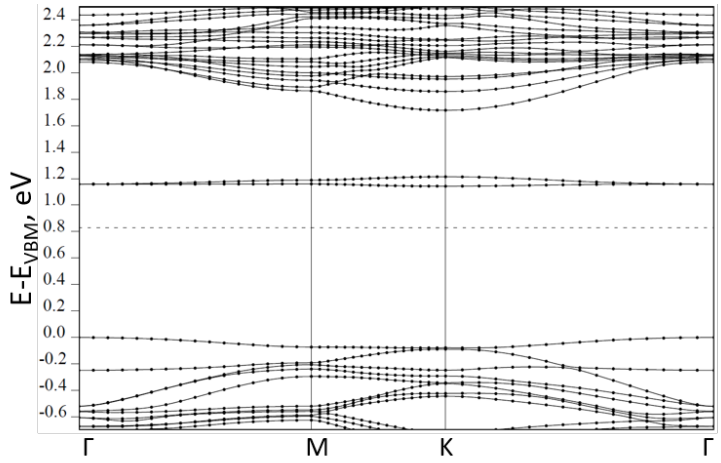
The experimental lattice constant for 2D MoS_2 is 3.16Å. Our DFT calculations predict it to be roughly 3.19Å, a remarkably close prediction. Nevertheless, we investigate influence of strain on bandstructure even further. As we see in Fig. 3.3, when we increase strain, valence band maximum (VBM) at a Γ point shifts higher relative to VBM at K point, and around 3.19Å, we see transition from direct to indirect bandgap. Because of that and the fact that there is still somewhat noticeable interaction between defects in 4x4 supercell, we decided to fix lattice constant at an experimental value of 3.16Å, which is a pretty common convention in computational DFT community.

We compute real and imaginary parts of dielectric function using the epsilon.x code of Quantum Espresso, which implements this calculation in random phase approximation (RPA). This code uses previously obtained DFT wavefunctions to compute matrix elements and joint density of states. Resulting dielectric function has only independent particle contribution (no excitons, and no phonons). We determined that we need to use Gaussian smearing with a width of 0.06 eV.

We used epsilon.x to also analyze the influence of spin-orbit coupling on dielectric function. As seen in Fig. 3.4, spin-orbit interaction leads to some minor modifications of the curves. However, for our purposes it is not significant enough to justify 10x increase in computational cost. Hence, for all results presented in this chapter, we did not include spin-orbit coupling in our calculations.

By performing careful convergence tests, we determined other DFT parameters

(a) 4x4 supercell



(b) 5x5 supercell

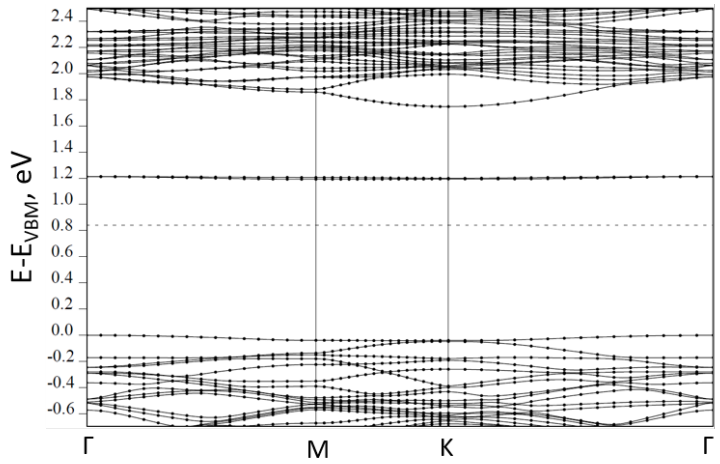


Figure 3.2: Influence of the size of supercell on bandstructure of 2D MoS_2.

sufficient for our calculations. We set wavefunctions' cut-off energy at 80 Ry. Our k-point grid is shifted and symmetrized: we use a k-point grid of 4x4x1 for structure relaxations (in x, y, z directions respectively), 16x16x1 for self-consistent calculations, 32x32x1 for non-self consistent calculations (used for computations of density of states and dielectric function). We use Gaussian smearing of occupations with width of 0.002 Ry to calculate projected density of states (PDOS) and density of states (DOS) using projwfc.x code of Quantum Espresso.

All calculations in this chapter were performed on the Cori supercomputer at the National Energy Research Scientific Computing Center (NERSC) – the scientific

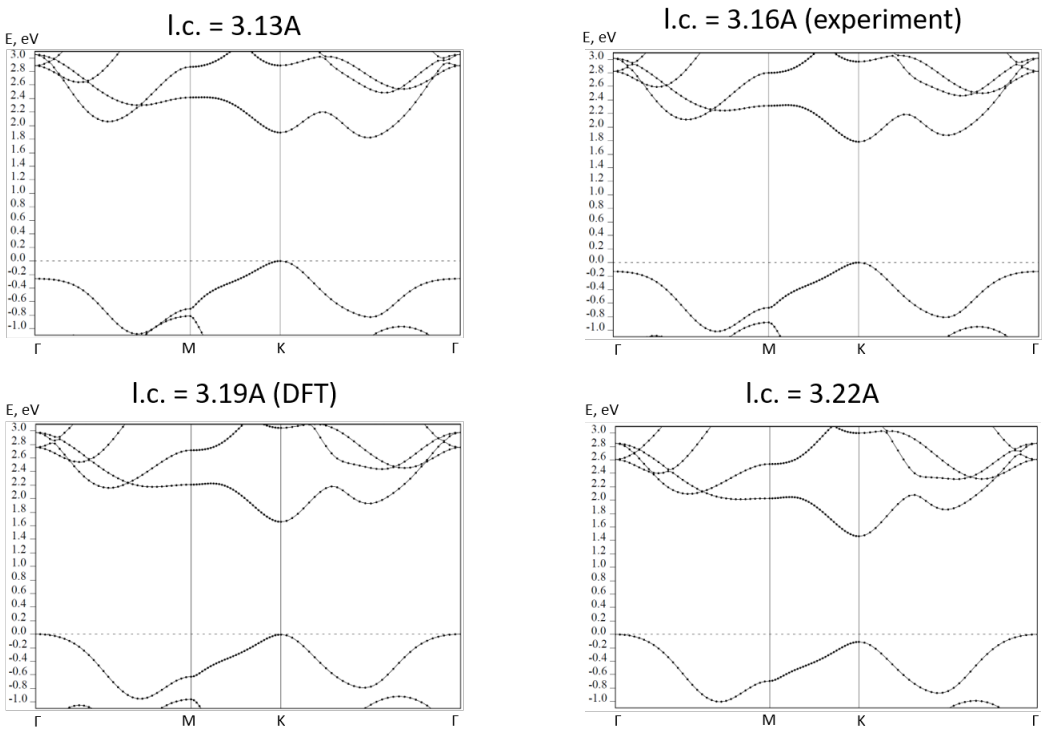


Figure 3.3: Influence of the strain on bandstructure of primitive 2D MoS_2 (computed on primitive cell).

(a) Real part (b) Imaginary part

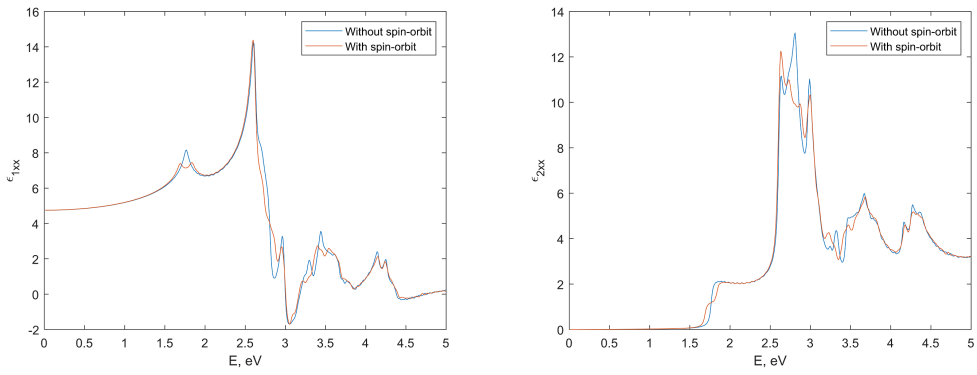


Figure 3.4: Influence of spin-orbit coupling on dielectric function of 2D MoS_2.

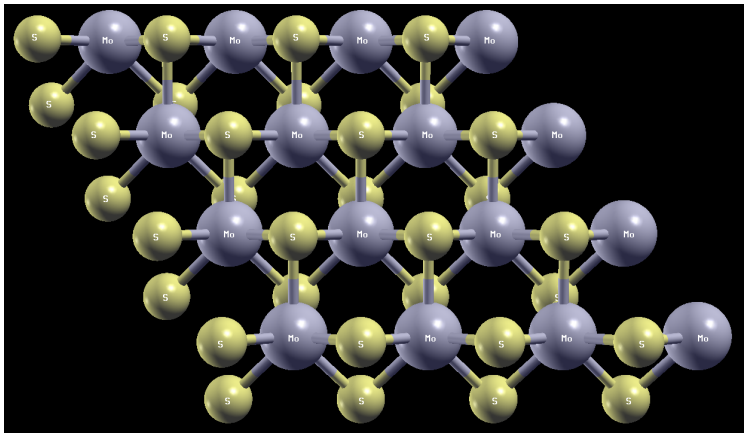
computing facility for the Office of Science in the U.S. Department of Energy.

3.3 Comparison of properties of pristine MoS_2 and MoS_2 with sulfur vacancy

The first calculations we performed were predicting properties of pristine and defective structures of 2D MoS_2. Even though in pristine case, we do not have to use supercells, which results in waste of computational resources and folding of bandstructure (making it less readable), we still perform all calculations using a 4x4 supercell (Fig. 3.5a) in this case in order to be consistent and to make a fair comparison with defective cases.

To simulate a defect under consideration, we remove one sulfur atom from pristine configuration and perform relaxation of atom positions, keeping the lattice constant fixed at 3.16Å (Fig. 3.5b).

(a) Pristine MoS_2



(b) MoS_2 with sulfur vacancy

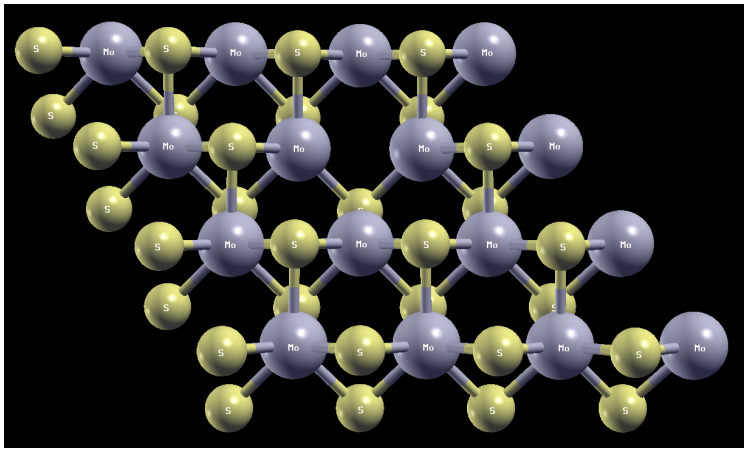


Figure 3.5: 4x4 supercells of pristine (top) and defective (bottom) MoS_2 (slightly rotated top view).

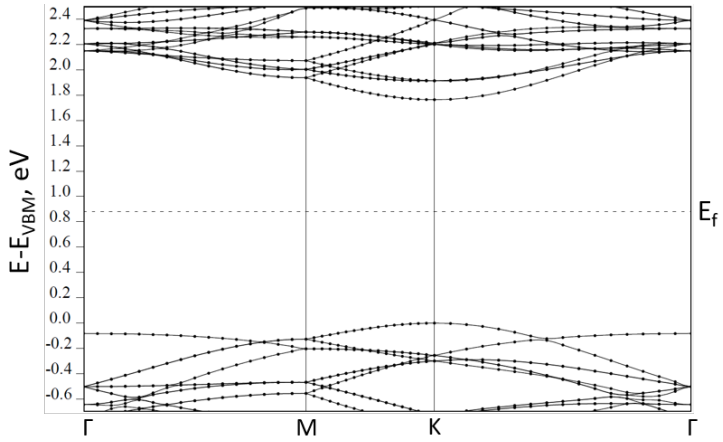
As a first step, we computed bandstructures for both pristine and defective cases. Computed bandgap in pristine material is roughly 1.8 eV (Fig. 3.6a), which appears to be close to experimentally reported values. However, this is an electronic (or quasiparticle) bandgap, not an optical bandgap. In reality, electronic bandgap is much higher than 1.8 eV (underestimating bandgap is a known problem in DFT), and it can be accurately computed using other methods such as GW approximation. Optical bandgap takes into account exciton binding energy, and since, in 2D MoS_2, it is quite large, optical and electronic bandgaps are different. It is a curious coincidence that optical bandgap corresponds really well to electronic bandgap computed from DFT in 2D MoS_2 (it is not generally true for other materials).

Introduction of sulfur vacancy significantly modifies bandstructure of 2D MoS_2 (Fig. 3.6). By comparing Fig. 3.6a with Fig. 3.6b, we see that sulfur vacancy creates three additional defect states: two within the bandgap and one within valence zone. This fact is not surprising since every sulfur atom is bonded to three neighboring molybdenum atoms. Apparent dispersion of defect states is due to a somewhat small size of supercell leading to interaction between defects in the crystalline lattice (see discussion of 4x4 vs 5x5 supercell in the previous section). Moreover, the defective case appears to have an indirect bandgap, which might be partially responsible for worsened optical properties.

Next, we calculate projected density of states (PDOS) for both cases (Fig. 3.7). PDOS shows us contribution of different kinds of atoms into density of states (DOS). We see significant disturbance of overall shape and appearance of an additional peak in DOS/PDOS within the bandgap that corresponds to defect states. Moreover, by looking at PDOS, we see that this peak is mostly dominated by contribution from d-states of molybdenum atoms.

Finally, we compare imaginary parts of RPA dielectric functions between pristine and defective 2D MoS_2 (Fig. 3.8). Again, there is significant difference between two cases. Because of defect states, we see significantly higher photon absorption within the bandgap, and there is a peak there that directly corresponds to excitation of electrons from valence band to defect states.

(a) Pristine MoS_2



(b) MoS_2 with sulfur vacancy

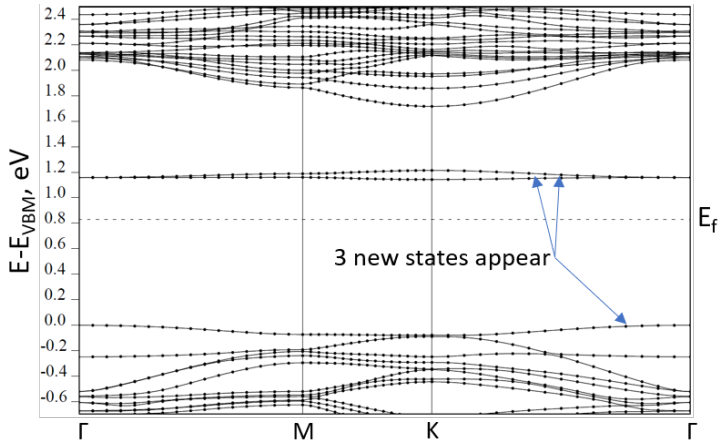


Figure 3.6: Comparison of band structures of pristine (top) and defective (bottom) MoS_2 (4x4 supercell).

(a) Pristine MoS_2 (b) MoS_2 with sulfur vacancy

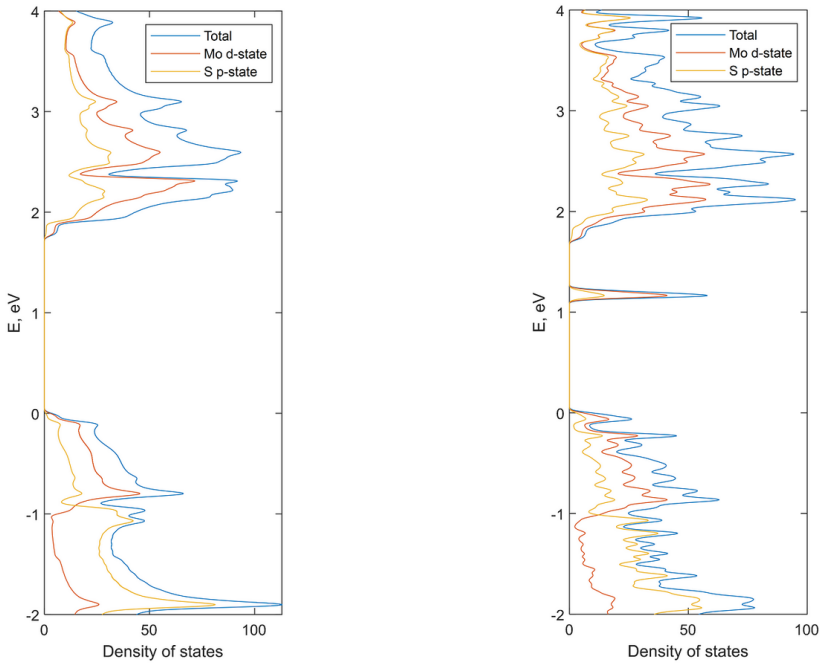


Figure 3.7: Comparison of densities of states of pristine (top) and defective (bottom) MoS_2 (4x4 supercell).

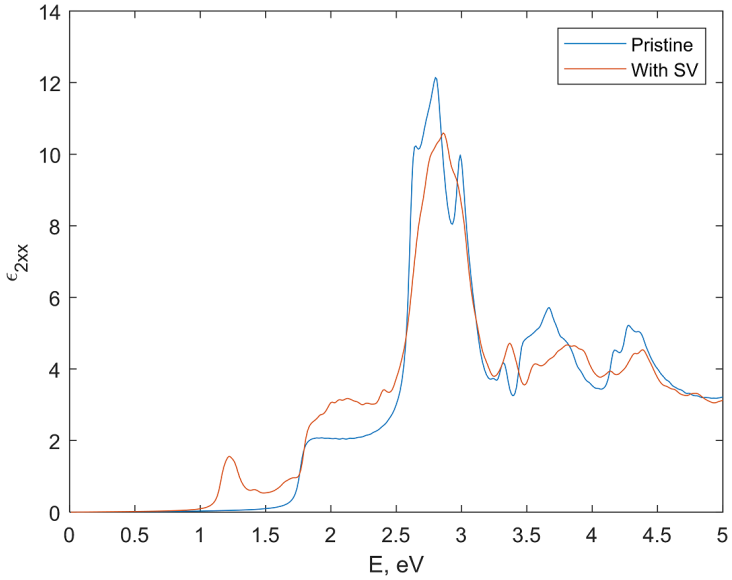


Figure 3.8: Comparison of imaginary parts of dielectric functions of pristine and defective MoS_2.

3.4 Hydrogenation of sulfur vacancies in 2D MoS_2

To study effects of hydrogenation of sulfur vacancies in 2D MoS_2, we considered four different structures corresponding to one, two, three, and four hydrogen atoms within the space of a sulfur vacancy. As usual, we performed geometrical relaxation while keeping the lattice constant at 3.16Å for all of them (Fig. 3.9). Interestingly, we were able to find energetical minima in all four cases, though the case of three hydrogen atoms is least stable: when relatively small perturbations into positions were introduced, the final configuration would be changed (in some of our runs we even observed one hydrogen atom stayed in the center of sulfur vacancy, while two other hydrogen atoms formed hydrogen molecule, that "flew" away from MoS_2). There are also other possible stable configurations for cases of two and four hydrogen atoms, though all of them have a higher total energy. On Fig. 3.9, we present configurations that we think are the most stable ones, given the number of hydrogen atoms.

We computed bandstructures, DOS, and dielectric functions for all four cases (Fig. 3.10-3.13, Fig. 3.14, Fig. 3.15). Our first word of caution is that cases of one and three hydrogen atoms have odd number of electrons in a unit cell. In non-relativistic DFT, every occupied state is actually doubly occupied (spin up, spin down). However, when we have an odd number of electrons, it corresponds to some states being partially occupied, which effectively leads to material being metallic. It is not necessary for the material under consideration to actually be metal (Mott insulators are exactly this kind of material), but it does point to some limitations of DFT. The case of one hydrogen atom has also one additional complication: the total magnetization is not zero, hence there is a difference in energy between spin-up and spin-down cases (Fig. 3.10).

Even with these cautions, we observe a very important effect. When we add one hydrogen atom, one defect state "disappears" from the bandstructure (Fig. 3.10). When we add two hydrogen atoms, we "remove" one more defect state from the bandgap. And when we add three hydrogen atoms, all defects' states effectively disappear. However, in this case, Fermi energy lies within the top of valence band, as if we introduced extreme p-doping. When we introduce four hydrogen atoms, we effectively "heal" bandstructure: bandgap is again around 1.8eV and Fermi level is within the bandgap (Fig. 3.13).

To make a comparison between pristine and passivated by four hydrogen atoms materials, we plot an imaginary part of dielectric function of both cases on the same

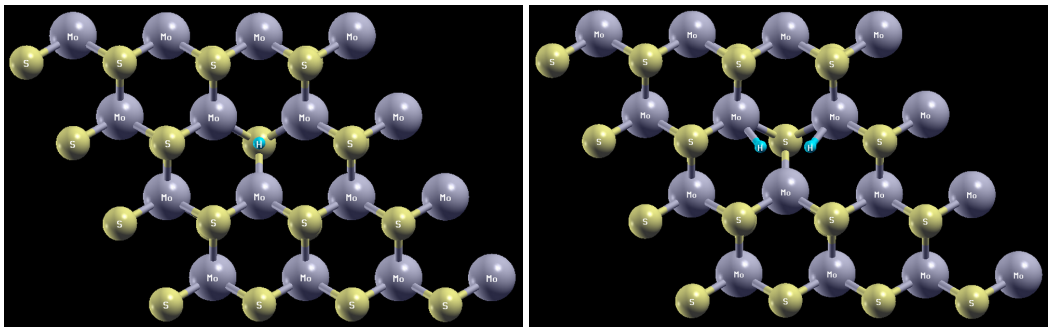
graph (Fig. 3.16). We see that the dielectric function of MoS_2 passivated with four atoms closely resembles the pristine one. This strongly suggests that four hydrogen atoms can fully passivate sulfur vacancies in 2D MoS_2.

This fact is actually surprising according to regular bonding theory. In MoS_2, molybdenum and sulfur form covalent bonds, and their formal oxydation states are +4 and -2 respectively, so that each sulfur atom accepts two electrons to fill two uncoupled 3p-orbitals in their outer shell. Every hydrogen atom provides one uncoupled 1s orbital, meaning that each hydrogen atom can accept one electron. So it appears that, to simulate effects of a missing sulfur atom in bonding and to fully passivate material, one needs to provide two hydrogen atoms instead of four.

One can argue that our hydrogen atoms also form bonds with each other, so that the overall hydrogen complex is actually accepting less electrons from molybdenum atoms in MoS_2. Studying protonation instead of hydrogenation may give us some insight into what is happening (see next section).

(a) Sulfur vacancy + 1H

(b) Sulfur vacancy + 2H



(c) Sulfur vacancy + 3H

(d) Sulfur vacancy + 4H

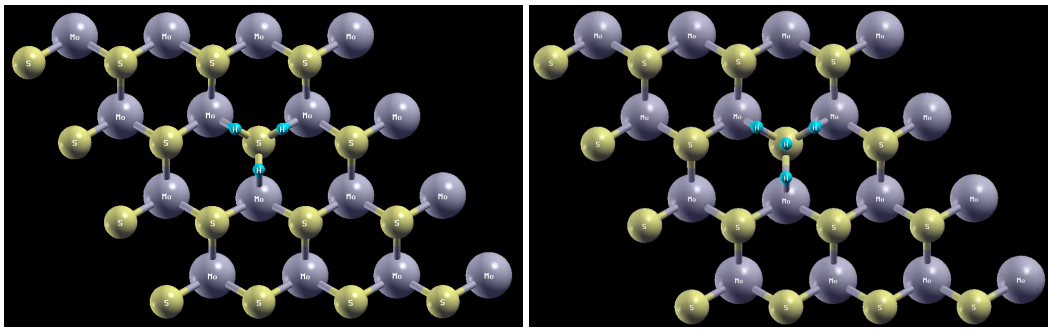


Figure 3.9: 4x4 supercells of 2D MoS_2 in which sulfur vacancies were filled with different number of hydrogen atoms.

(a) Spin up. (b) Spin down.

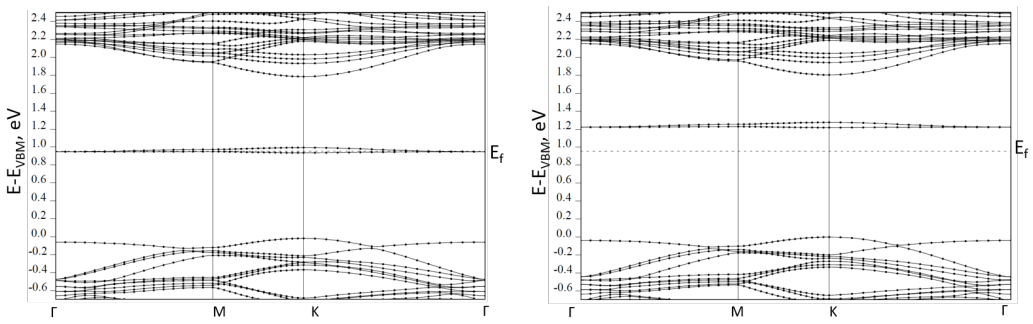


Figure 3.10: Band structure of MoS_2 in which sulfur vacancy is filled with one hydrogen atom (4x4 supercell).

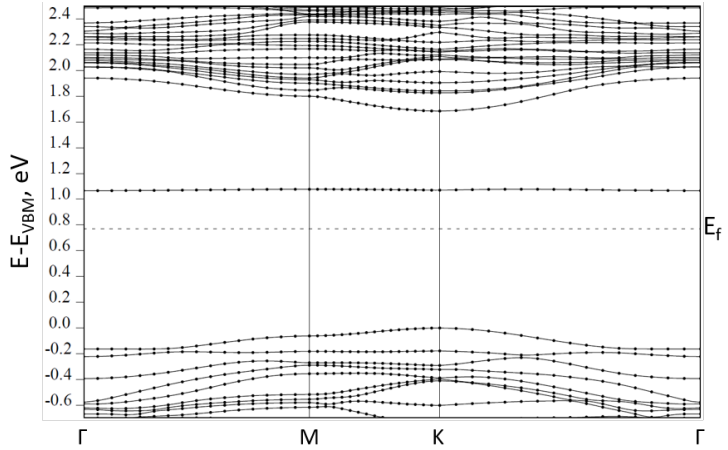


Figure 3.11: Band structure of MoS_2 in which sulfur vacancy is filled with two hydrogen atoms (4x4 supercell).

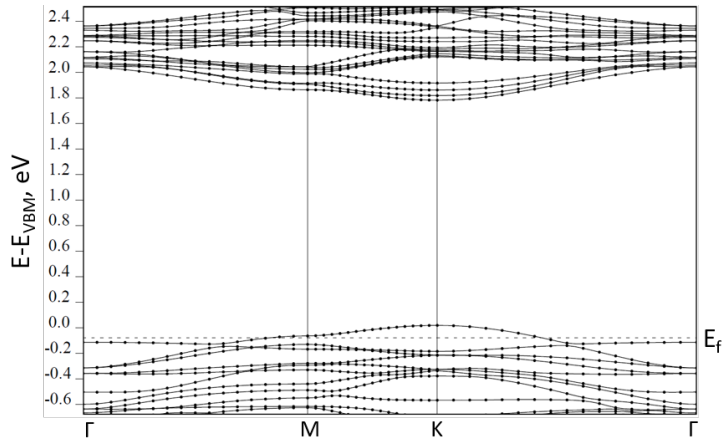


Figure 3.12: Band structure of MoS_2 in which sulfur vacancy is filled with three hydrogen atoms (4x4 supercell).

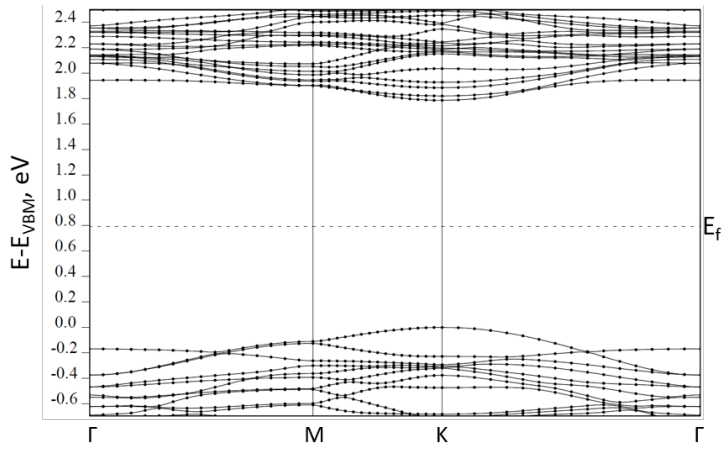


Figure 3.13: Band structure of MoS_2 in which sulfur vacancy is filled with four hydrogen atoms (4x4 supercell).

(a) SV + 1H (b) SV + 2H (c) SV + 3H (d) SV + 4H

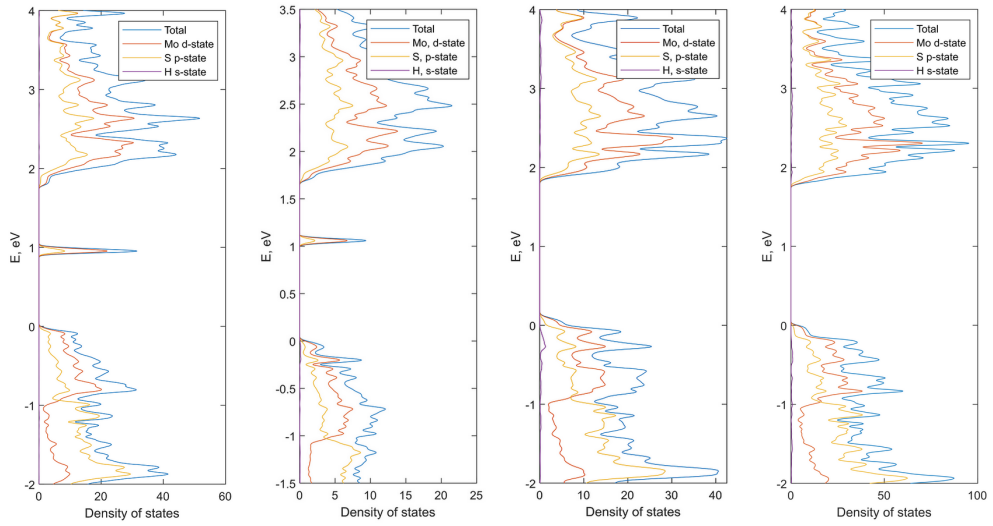
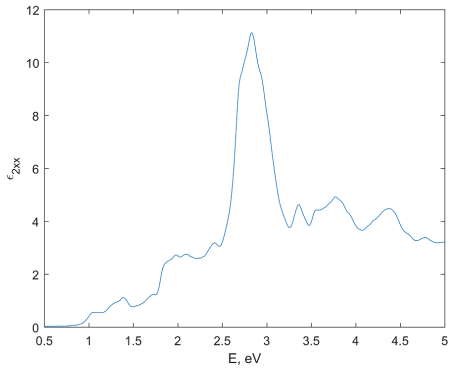
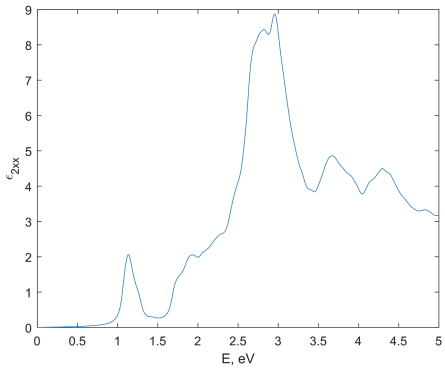


Figure 3.14: Projected densities of states of 2D MoS_2 in which sulfur vacancies were filled with different number of hydrogen atoms.

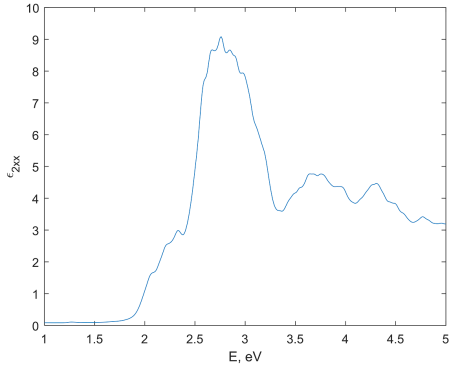
(a) SV + 1H



(b) SV + 2H



(c) SV + 3H



(d) SV + 4H

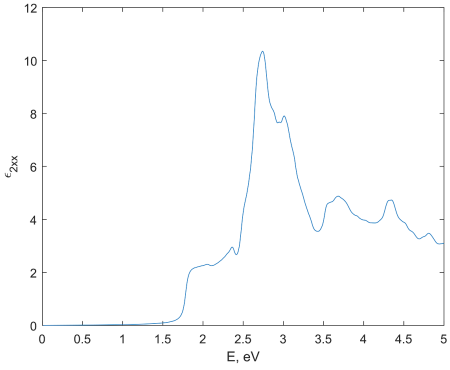


Figure 3.15: Imaginary parts of dielectric functions of 2D MoS_2 in which sulfur vacancies (SV) were filled with different number of hydrogen atoms.

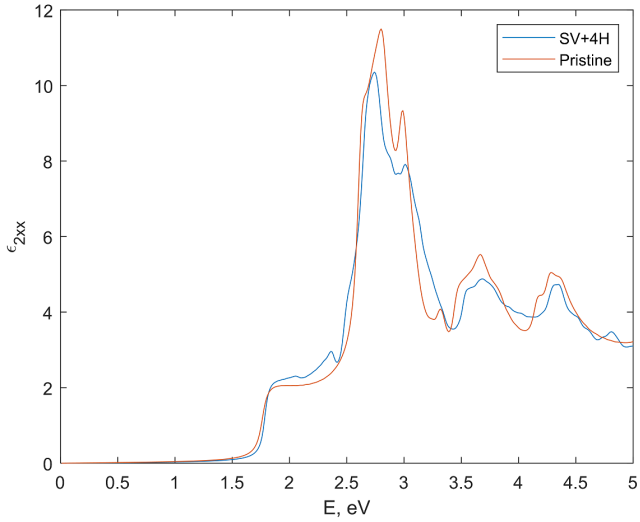


Figure 3.16: Comparison of imaginary parts of dielectric functions of pristine MoS_2 and defective 2D MoS_2 in which sulfur vacancies (SV) were filled with four hydrogen atoms.

3.5 Passivation by hydrohalic acids (simulating protonation)

As seen in the previous section, hydrogenation may provide a way to passivate sulfur vacancies in 2D MoS_2. However, this mechanism is very unlikely when we treat our material with TFSI superacid. By definition, acids are molecules or ions that are capable of donating protons (hydrogen ions H+), and their strength is determined by their tendency to dissociate into a proton and an anion. So, in acid solutions we do not have hydrogen atoms floating around, but rather protons.

Since TFSI is a superacid, its ability to donate protons is extreme in comparison with other acids. One of possible MoS_2 passivation mechanisms with TFSI superacid is protonation, filling in sulfur vacancies with protons. According to speculations based on conventional bonding theory (see previous section), even one proton could be enough, since it can accept two electrons similar to how sulfur accepts two electrons forming covalent bonds of MoS_2.

However, simulating a proton in DFT is challenging. It requires doing calculations with charged supercell, and in a periodic settings, it leads to an infinite macroscopic charge and a divergence. There are some tricks that allow us circumvent this problem to some extent (like introducing compensating background jelly charge), but they might lead to different kinds of physics being simulated.

One way of simulating a proton is to actually incorporate a full acid molecule into our DFT calculation. An acid anion pulls an electron from hydrogen onto itself, effectively stripping hydrogen of this electron and making it a proton.

The TFSI molecule is quite big, unfortunately, potentially requiring bigger supercell size than what was used in this work so far. In order to simulate protonation, we might try using other strong acids. In fact, there is a class of hydrohalyc acids, which molecules have only two atoms: HX, where H is hydrogen and X is halogen.

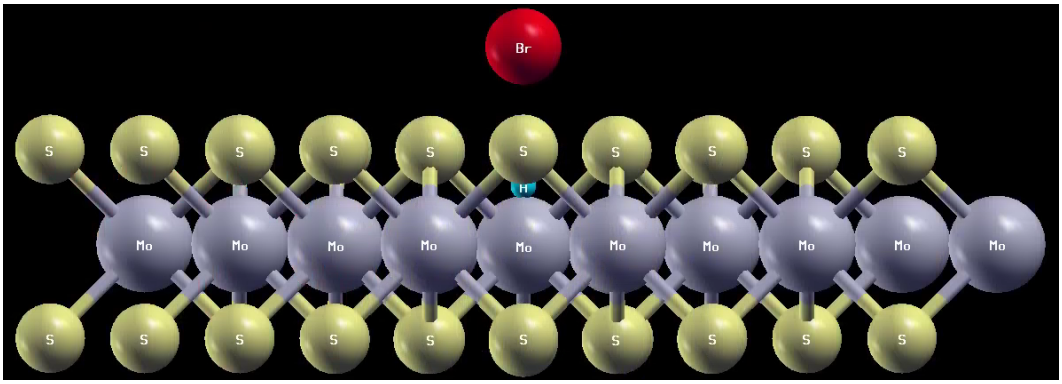
In this work, we studied effects of sulfur vacancy passivation using hydrobromic (HBr) and hydroidic (HI) acids. We also tried considering hydrocloric acid (HCl), but it would not bind to the material.

Both HBr and HI are very strong acids, making them ideal candidates to study protonation of sulfur vacancies.

As usual, our first step after introducing acid molecules into sulfur vacancies in 2D MoS_2 is to perform relaxation of atoms' positions (while keeping the lattice constant fixed at 3.16Å). Fig. 3.17 shows results of relaxation after introducing one

HBr or one HI molecule. It is worth noting that HI has a longer bond length as expected.

(a) SV + HBr



(b) SV + HI

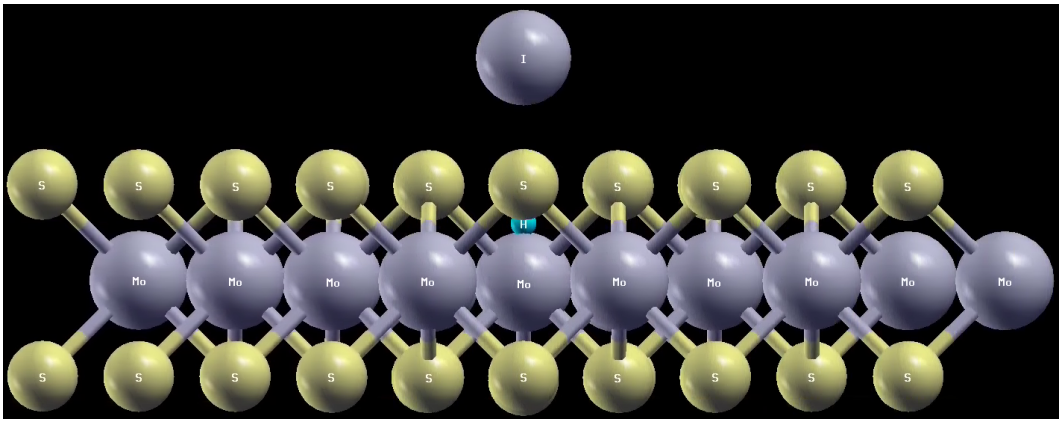


Figure 3.17: Cross-sectional view of 4x4 supercells of 2D MoS_2 in which sulfur vacancy is filled with different hydrohalic acids.

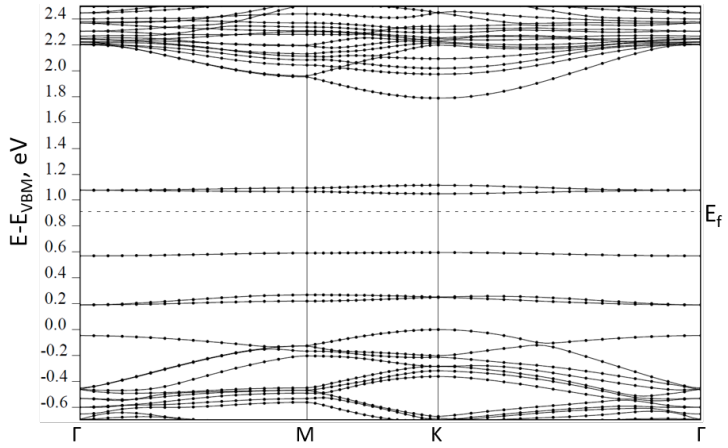
Computed bandstructures in both cases (Fig. 3.18) look very similar to a bandstructure of 2D MoS_2, in which sulfur vacancy was filled with one hydrogen atom (Fig. 3.10). We see also the appearance of additional flat states in bandgaps, but they are associated with atomic states of Br and I, as evidenced by their overwhelming contribution to the density of states at these energies (Fig. 3.19). Surprisingly, density of states in all three cases looks very similar (except for these localized atomic contributions of Br and I).

We also computed RPA dielectric functions of 2D MoS_2 passivated with HBr and HI. Comparing them with a dielectric function in case of passivation with one

hydrogen atom (Fig. 3.20), we see a significant resemblance between all of them.

From these results, to our surprise, we can conclude that effects of hydrogenation and protonation are remarkably close. Thus, we might expect that four protons will be able to passivate sulfur vacancies in 2D MoS_2 (similar to four hydrogen atoms). Such protons can be provided by acid solutions, clearly demonstrating that passivation by TFSI superacid is possible through this mechanism.

(a) SV + HBr



(b) SV + HI

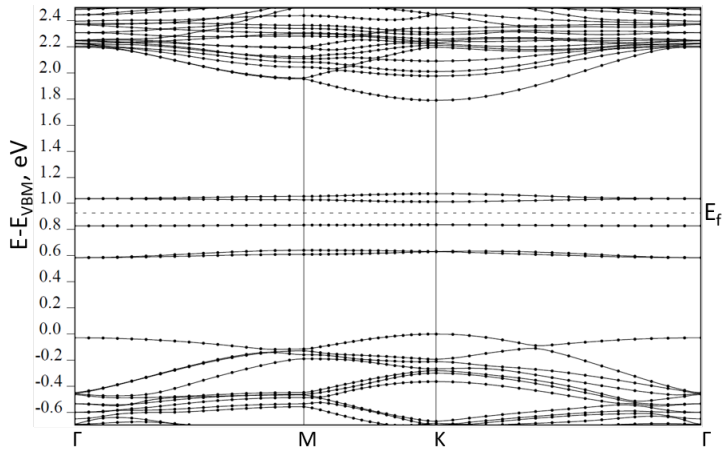


Figure 3.18: Bandstructure of 2D MoS_2 in which sulfur vacancy is filled with different hydrohalic acids.

(a) SV + 1H (b) SV + HBr (c) SV + HI

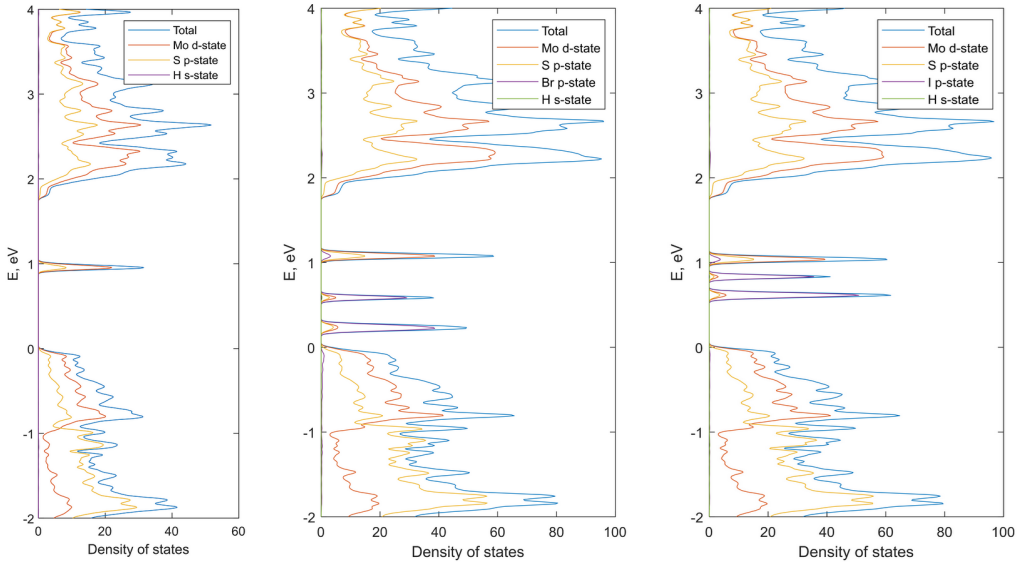


Figure 3.19: Projected densities of states of 2D MoS_2 in which sulfur vacancies were filled with different hydrohalic acids in comparison with case of passivation by one hydrogen atom.

(a) SV + HBr (b) SV + HI

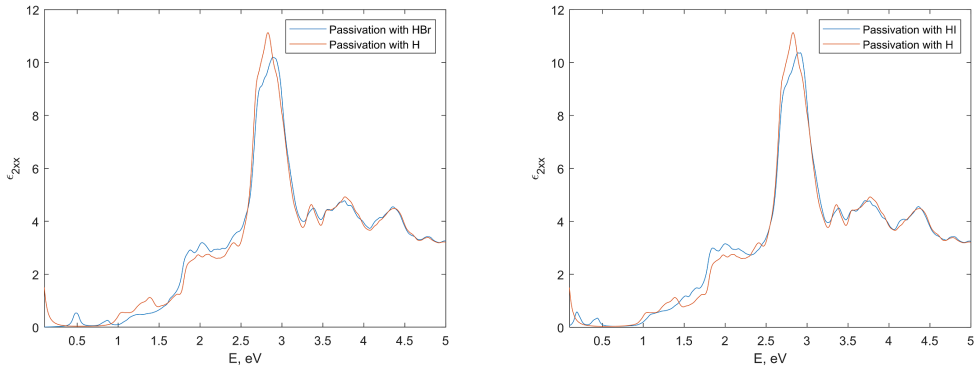


Figure 3.20: Imaginary parts of dielectric functions of 2D MoS_2 in which sulfur vacancies were filled with different hydrohalic acids in comparison with case of passivation by one hydrogen atom.

3.6 Passivation by carbon and oxygen

As found in the previous sections, it appears that we need to provide four valence electrons to passivate sulfur vacancies in 2D MoS_2. This fact contradicts intuition based on bonding theory considerations. However, we need to keep in

mind that this intuition is based on atomic orbitals. But, in reality, crystals have a different set of eigen states/wavefunctions, and energy bands of solids are the result of collective interaction between electrons from different atoms, hence our conventional chemical intiution should not exactly match real systems.

It is a curious observation that we have four p-electrons on the outer shell of sulfur. Even though two of them are paired, interactions in the crystal might still be peculiar enough so that all four p-electrons are somehow needed for bonds in MoS_2 (note that by convention we state number of atoms in primitive cell of chemical formula of a crystal, and it does not mean that one molybdenum atom is bonded to two sulfur atoms, which we know is not true). By putting four hydrogen atoms into the sulfur vacancy, we are introducing four electrons back into the system, and, hence, we are able to fully passivate the material. We should be extremely cautious with this reasoning, and further investigation is necessary.

To test this idea, we considered passivating sulfur vacancies with atoms that have different a number of valence electrons on their outer shells. Boron, carbon, and nitrogen have one, two, three valence electrons respectively, so one might expect that passivating with these elements could be similar to passivating with one, two, and three hydrogen atoms respectively. Unfortunately, by introducing boron and nitrogen, number of electrons in a supercell becomes odd, and as discussed earlier, this makes considerations within DFT more difficult.

Hence, we focus our attention on passivation by carbon. We considered introducing one and two carbon atoms into the sulfur vacancy. Fig. 3.21 shows the top view of a supercell with two carbon atoms in the sulfur vacancy after relaxation of geometrical positions (the lattice constant was kept at 3.16Å).

Introduction of one carbon atom removes two defect states from the bandstructure: one from the bandgap and one from the valence band (Fig. 3.22a). This is, in fact, very similar to passivation by two hydrogen atoms (Fig. 3.11). Even densities of states between two cases look alike (Fig. 3.23b vs Fig. 3.14b).

When we put two carbon atoms into the sulfur vacancy, we effectively fully passivate sulfur vacancies. We do not have defect states in bandstructure anymore (Fig. 3.22b). Moreover, there is an even stronger similarity of densities of states in this case and pristine material (Fig. 3.23) than in the case of passivation with four hydrogen atoms (Fig. 3.14d). We also computed RPA dielectric function in this case: it looks remarkably close to the pristine case (Fig. 3.24). Thus we can

conclude that two carbon atoms (or a carbon molecule) can passivate sulfur vacancy in 2D MoS_2 (similar to four hydrogen atoms, which might surprising).

As an additional exercise, we also considered passivation with one and two oxygen atoms (Fig. 3.25, 3.26, 3.27, 3.28). We see that in both cases, oxygen can fully passivate sulfur vacancies. This is not surprising, since oxygen is a chalcogen (like sulfur). Though, there are some discrepancies in the case of two oxygen atoms: we see the appearance of an oxygen state within the bandgap close to valence band (Fig. 3.26b), and dielectic funcion looks more dissimilar to a pristine one than a case of passivation with one oxygen atom (Fig. 3.28).

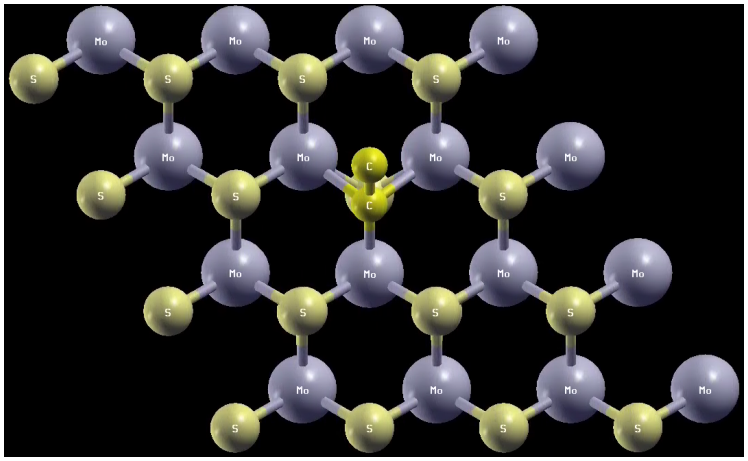


Figure 3.21: 4x4 supercell of 2D MoS_2 in which sulfur vacancies were filled with two carbon atoms.

(a) SV + C (b) SV + 2C

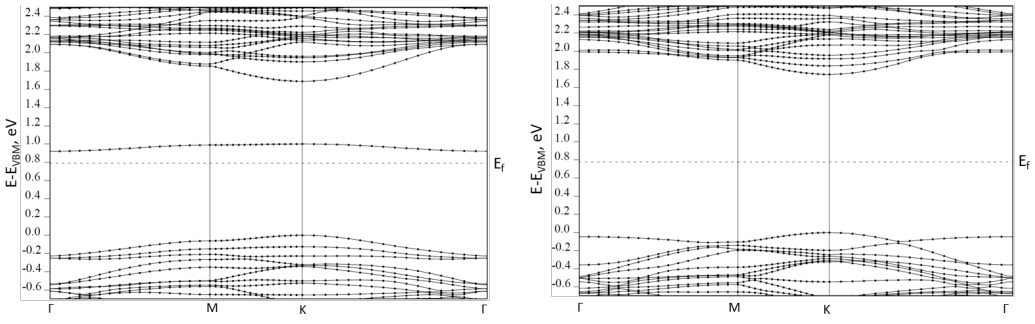


Figure 3.22: Bandstructure of 2D MoS_2 in which sulfur vacancy is filled with a different number of carbon atoms.

(a) Pristine (b) SV + C (c) SV + 2C

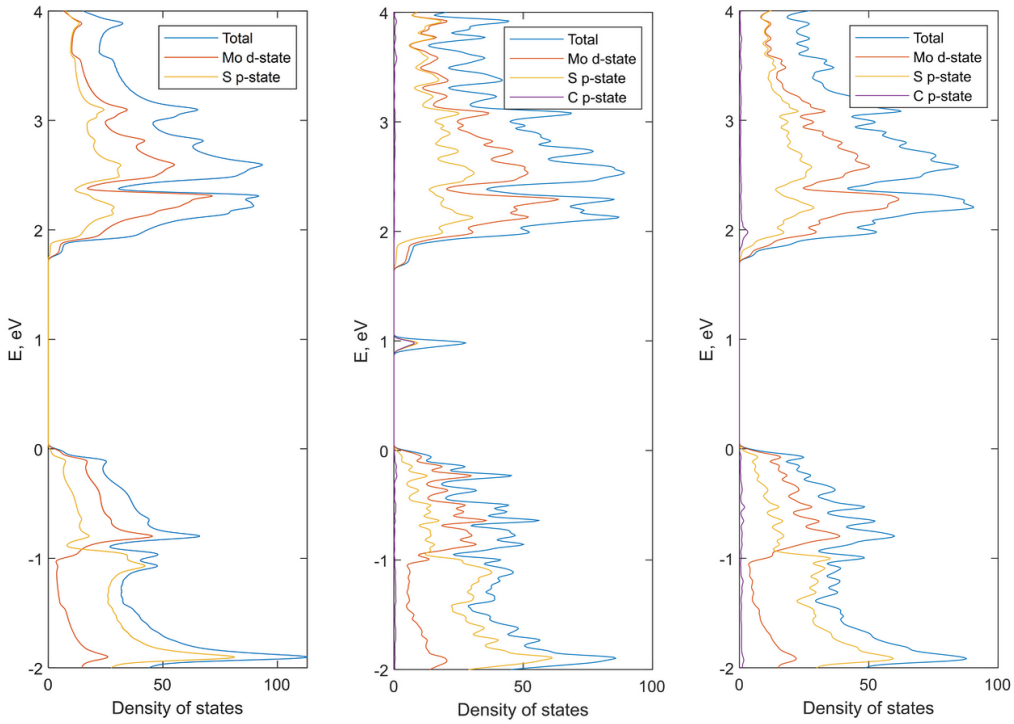


Figure 3.23: Projected densities of states of 2D MoS_2 in which sulfur vacancies were filled with a different number of carbon atoms.

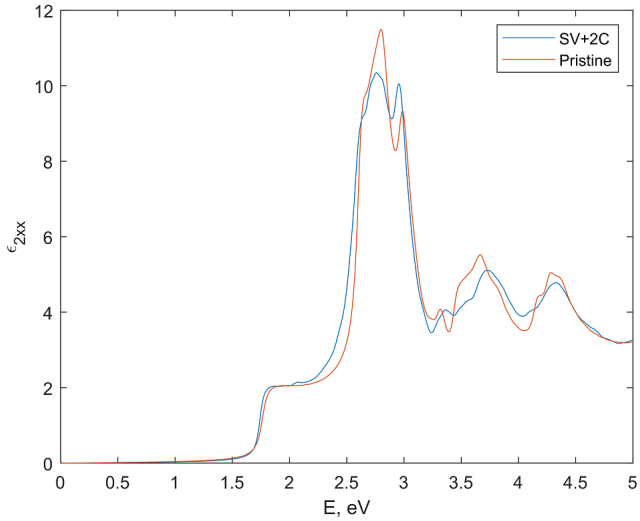


Figure 3.24: Imaginary part of dielectric function of 2D MoS_2 in which sulfur vacancies were filled with different two carbon atoms.

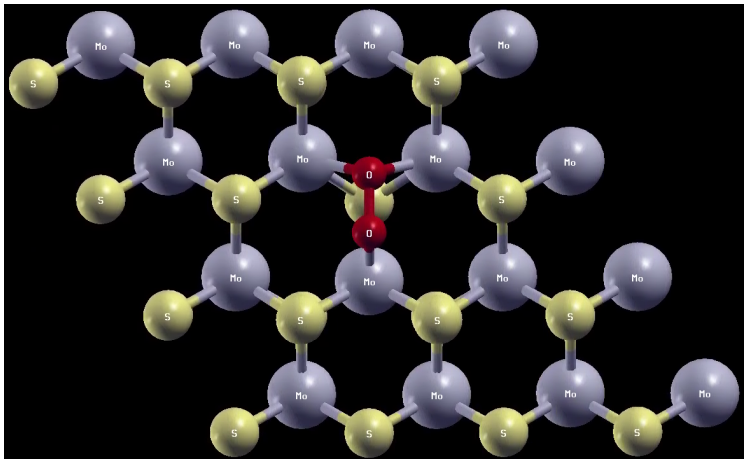


Figure 3.25: 4x4 supercell of 2D MoS_2 in which sulfur vacancies were filled with two oxygen atoms.

(a) SV + O

(b) SV + 2O

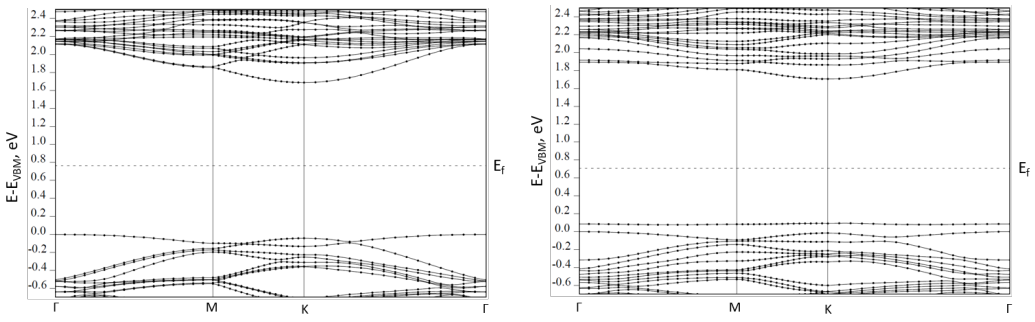


Figure 3.26: Bandstructure of 2D MoS_2 in which sulfur vacancy is filled with a different number of oxygen atoms.

(a) Pristine (b) SV + O (c) SV + 2O

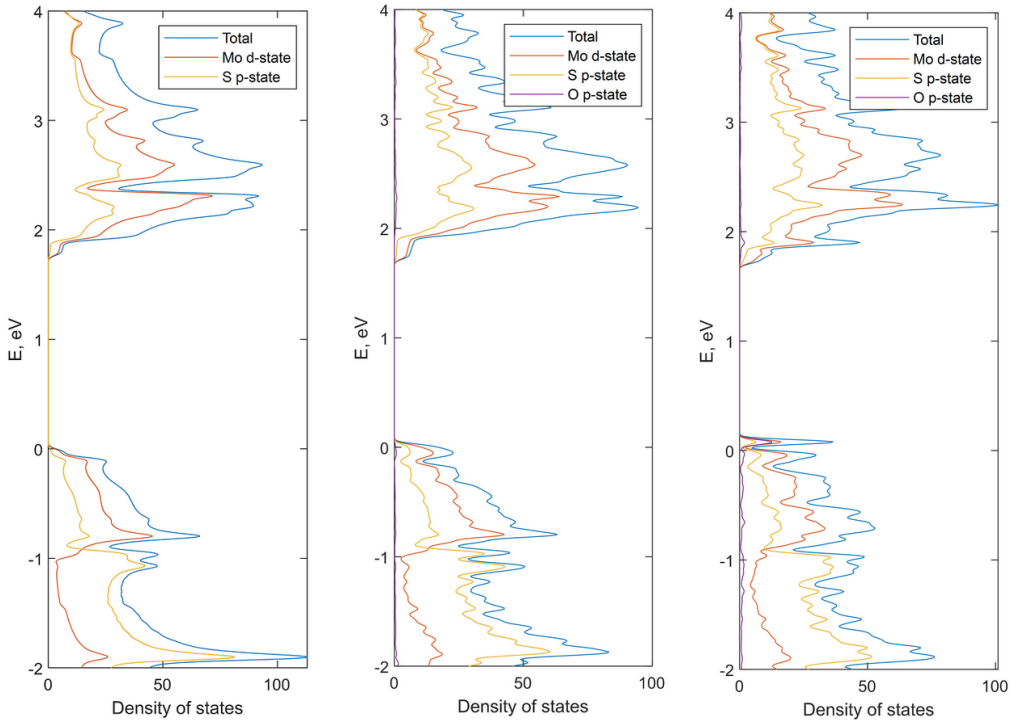


Figure 3.27: Projected densities of states of 2D MoS_2 in which sulfur vacancies were filled with a different number of oxygen atoms.

(a) SV + O (b) SV + 2O

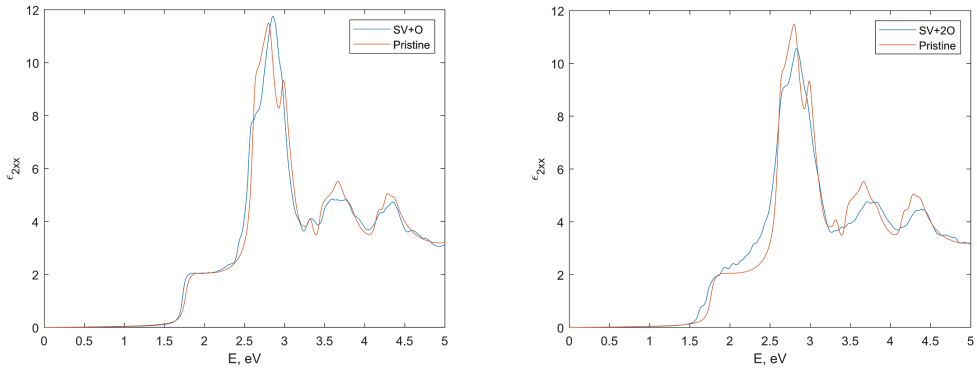


Figure 3.28: Imaginary parts of dielectric functions of 2D MoS_2 in which sulfur vacancies were filled with a different number of oxygen atoms.

3.7 Conclusion and outlook

In summary, we have performed a systematic study of different passivation schemes of sulfur vacancies in 2D MoS_2.

- We found that incorporating hydrogen atoms into sulfur vacancies removes defect states from bandstructure (hydrogenation). Adding four hydrogen atoms appears to fully passivate material.
- We performed analysis of protonation of sulfur vacancies by introducing strong acids into defect sites. We found that effects of protonation are very similar to hydrogenation. Hence, we expect that protonation is a viable mechanism of passivation in our study.
- To get a better insight into necessary conditions of passivation, we considered passivation schemes with other elements, like carbon and oxygen. We found that one carbon has similar effects to two hydrogen atoms, and two carbon atoms passivate material fully (similarly to four hydrogen atoms).

Overall, the work in this chapter provides a solid ground for explaining the mechanisms of defects' passivation in 2D MoS_2, treated with TFSI superacid. We showed that both hydrogenation and protonation can "heal" defect states, and since TFSI superacid is defined by its strong ability to donate protons, we conclude that this mechanism is possible.

As for future directions, we believe it would be really interesting to do DFT calculations with whole a TFSI molecule in a supercell. It will be more challenging, but it will give us a better insight into microscopic details of passivation. It is known that TFSI has Lewis acid properties, meaning that the TFSI molecule can potentially accept electrons from molybdenum atoms. Doing such DFT calculations will give us an opportunity to assess possibilities of different mechanisms of passivation of sulfur vacancies in 2D MoS_2.

Chapter 4

OPTIMIZATION OF NANOPHOTONIC DEVICES

4.1 Introduction

In nanophotonics, a common problem is to design a device that meets some target performance. There is a great demand for intelligent automatic methods that could potentially help experts in their task, or even find such solutions that are impossible to come up with using conventional physical intuition [56]. Many problems in the design of nanophotonic devices can be formulated as a problem of mathematical optimization of some figure of merit that measures how good current design is in achieving target performance. Finding optimal materials and geometries is often a very complicated task of discrete and continuous optimization. It is not uncommon that design of the figure of merit function has its own challenges and uncertainties. And even when such a function can be formulated, it might be rigged with the problem of multiple local optima, which puts the problem into the space of non-convex optimization, being by itself an area of active research in mathematical and computational communites. There is a tremendous interest from both academic and industry leaders for advanced methods of optimization that could assist with the design of state-of-the-art nanophotonic devices.

In this chapter, we explore application of advanced optimization methods to the design of different nanophotonic devices: one with small a number of parameters, and one with high-dimensional optimization space. It is worth noting that the work presented here is highly collaborative and multi-disciplinary, involving people from several research groups.

Sections 4.2-4.5 present optimization of plasmonic mirror filters for hyper-spectral imaging [22] using multi-fidelity Gaussian Processes optimization [57]. In Sections 4.6-4.9, we explore application of neural networks to optimize different aspects of nanophotonic phased arrays for universal metasurfaces [58].

4.2 Optimization of plasmonic mirror filters

In this part of this chapter, we use design from [22] to test a novel optimization algorithm.

Designing compact integrated color filters with ultra-narrow bandwidth is of

great importance for realizing practical multispectral and hyperspectral imaging. Each pixel of a hyperspectral imaging device records the spectrum of light from the environment, providing significantly more information in comparison with conventional imaging techniques, that can help with, e.g., materials identification or objects detection. Such devices can find a wide range of applications in different areas of science and technology, including medicine, material science, astronomy, and environment monitoring to name a few.

Surface plasmon polaritons (SPPs) allow for extreme miniaturization of integrated photonic devices via strong light confinement that can result in a very small wavelength of light (potentially dozens of times smaller than in free space) [59]. Hence, such plasmonic devices look very promising as a platform for designing ultra-compact integrated narrow-band photonic filters [60, 61]. Periodic arrays of subwavelength holes or nanoslits in metal films enable efficient excitation of SPPs by satisfying momentum-matching with the addition of a grating wavevector. The grating materials, geometry, and symmetry control the excitation efficiency [62]. In particular, periodic arrays of subwavelength apertures passing through an optically thick metal film exhibit enhanced transmission exclusively at conditions corresponding to constructive mutual interference between incident light and SPPs traveling along the surface between adjacent slits and acting as a band-pass color filter [63]. However, periodicity of slits can be effectively achieved by placing reflective mirrors around a slit.

Designing such plasmonic mirror filters sets up a non-trivial optimization challenge: the number of independent parameters can easily exceed a few dozen and the optimization landscape itself is non-convex with many local minima. Additional challenges arises from the fact that the derivation of an analytical model is nearly impossible due to near-field effects and complicated geometry, hence numerical simulations of underlying physical processes (governed by Maxwell's equations) are required.

We use Lumerical commercial implementation of the finite-difference time-domain (FDTD) method to simulate the transmission spectra of such devices. In post-processing analysis, we extract observable scalar figures of merit corresponding to the goodness of the spectrum such as transmission peak amplitude, its offset from designed wavelength, full-width half-maximum, and signal-to-noise ratio, and combine them into one weighted figure of merit (FOM). Using this FOM, the evolution of the design can then be pursued as a minimization problem over a geometric

parameter space, which can be driven using any of a variety of iteration schemes. Here, we will search over a five-parameter domain, describing the geometry of our devices.

The filter's center wavelength, linewidth, and amplitude are determined by the interaction of several physical processes including the amplitude and phase of photon-plasmon coupling at the slits, the strength of mutual coupling between the propagating waveguide channels, and the effective index of each participating electromagnetic mode. Due to the interplay of these several physical resonances, the FOM corresponding to our filters exhibits oscillations in parameter space that are likely to trap a local directed search method in a globally non-optimal local solution. Therefore, gradient descent or other local optimization modalities are excluded for this purpose. Conversely, the relatively large computational cost of the FDTD forward problem limits the applicability of purely stochastic approaches like evolutionary methods. Instead, we seek methods which execute a global, derivative-free search with an efficient iteration strategy that calls the forward problem solver a limited number of times (under a defined budget).

Here we present an application of an in-house developed optimization strategy based on multi-fidelity Gaussian processes [57] to this nanophotonics design problem. Our simulation setup allows us to easily control the fidelity of numerical calculations by changing geometric mesh size and total time duration of simulated physical processes. We compare it with a conventional Gaussian Processes approach and a commonly used algorithm, Particle Swarm Optimization, which is implemented in Lumerical commercial nanophotonics software[1].

4.3 Numerical approaches to black-box optimization

Classical approaches to single-fidelity black-box optimization

Heuristics There are many stochastic heuristics for finding approximate solutions of non-convex optimization problems, such as simulated annealing [64], genetic algorithms [65], particle swarm optimization (PSO) [66], and many others. In this chapter, we are using PSO as one of the baselines for comparison, as it has a wide use in nanophotonics community [67, 68], and it is implemented in Lumerical commercial nanophotonics software.

In the particle swarm algorithm, the potential solutions, called particles, are initialized at random positions and velocities, and then move within the parameter

[1] https://kb.lumerical.com

search space. The particles are subject to three forces as they move: spring force towards the personal best position ever achieved by that individual particle, spring force towards the global best position ever achieved by any particle, and a frictional force proportional to the velocity. At each iteration velocity of each particle is stochastically updated based on these forces and previous velocity values, then new particles' locations are calculated as the old ones plus the velocities, modified to keep particles within bounds. The algorithm proceeds until a specified stopping criterion is met.

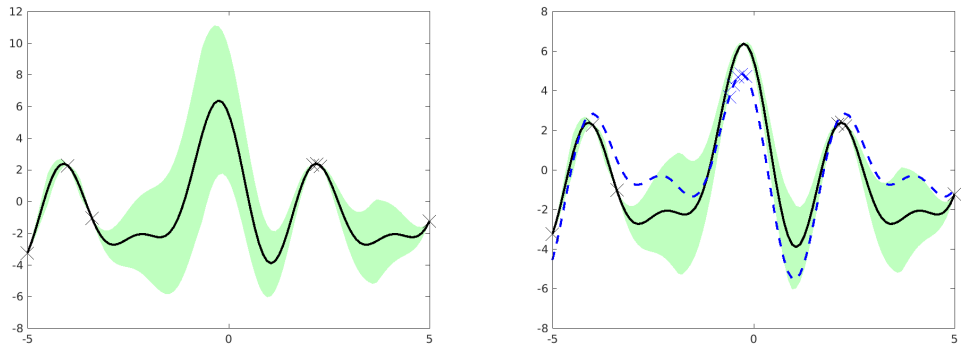
PSO is inspired by the behavior of animal aggregations like flocks of birds or insects swarming. Each particle is attracted to some degree to the best location it has found so far, and also to the best location any member of the swarm has found. After some steps, the population can coalesce around one location, or can coalesce around a few locations, or can continue to move.

Gaussian processes optimization Optimizing an unknown and noisy function is a common task in Bayesian optimization. In real applications, such functions tend to be expensive to evaluate, for example, tuning hyperparameters for deep learning models [69], so the number of queries should be minimized. As a way to model the unknown function, the Gaussian process (GP) [70] is an expressive and flexible tool to model a large class of functions. A classical method for Bayesian optimization with GPs is GP-UCB [71], which treats Bayesian optimization as a multi-armed bandit problem and proposes an upper-confidence bound based algorithm for query selections. The authors provide a theoretical bound on the cumulative regret that is connected with the amount of mutual information gained through the queries. [72] directly incorporates mutual information into the UCB framework and demonstrates the empirical values of their method.

Entropy search [73] represents another class of GP-based Bayesian optimization approaches. Its main idea is to directly search for the global optimum of an unknown function through queries. Each query point is selected based on its informativeness in learning the location for the function optimum. A predictive entropy search [74] addresses some computational issues from the entropy search by maximizing the expected information gain with respect to the location of the global optimum. Max-value entropy search [75, 76] approaches the task of searching the global optimum differently. Instead of searching for the location of the global optimum, it looks for the value of the global optimum. This effectively avoids issues

related to the dimension of the search space, and the authors are able to provide regret bound analysis that the previous two entropy search methods lack.

Multi-fidelitiy Bayesian optimization



(a) Only querying target fidelity function. (b) Querying both target and a lower fidelity.

Figure 4.1: Benefit from multi-fidelity Bayesian optimization. The left panel shows normal single fidelity Bayesian optimization where locations near a query point (crosses) have low uncertainty. When there is a lower fidelity cheaper approximation in the right panel, by querying a large number of points of the lower fidelity function, the uncertainty in the target fidelity can also be reduced significantly.

Multi-fidelity optimization is a general framework that captures the trade-off between cheap low-quality and expensive high-quality data (cf. Figure 4.1). There have been several works on using GPs to model functions of different fidelity levels. Recursive co-kriging [77, 78] considers an autoregressive model for multi-fidelity GP regression, which assumes that the higher fidelity consists of a lower fidelity term and an *independent* GP term which models the systematic error for approximating the higher-fidelity output. Therefore, one can model cross-covariance between the high-fidelity and low-fidelity functions using the covariance of the lower fidelity function only. *Virtual vs Real* [79] extends this idea to Bayesian optimization. The authors consider a two-fidelity setting (i.e., virtual simulation and real system experiments), where they model the correlation between the two fidelities through co-kriging, and then apply the entropy search to optimize the target output. Zhang et al. (2017) [80] model the dependencies between different fidelities with convolved Gaussian processes [81], and then apply predictive entropy search (PES) [74] to efficient exploration.

Although these multi-fidelity heuristics have shown promising empirical results on some experimental datasets, little is known about their theoretical perfor-

mance. Recently, Kandasamy et al. (2016) propose MF-GP-UCB (Multi-fidelity GP-UCB) [82], a principled for multi-fidelity Bayesian optimization. In particular, the authors consider an iterative two-stage optimization procedure and view each fidelity as an independent component, and at each iteration update the estimate of each fidelity *only* based on observations from the corresponding fidelity. In a follow-up work [83], the authors address the disconnect issue by considering a continuous fidelity space and performing joint updates to effectively share information among different fidelity levels. However, as elaborated on in [57], these approaches are likely to pick sub-optimal actions in some pessimistic cases, due to the modeling assumption and the two-stage query selection criteria. In this chapter, we focus on MF-MI-Greedy, a principled multi-fidelity algorithm as recently proposed in [84]. We describe the details of the algorithm in §4.4, and evaluate it against the MF-GP-UCB algorithm as well as other single-fidelity baselines in §4.5.

4.4 Multi-fidelity Bayesian optimization

Preliminary and Problem Formulation

Consider the problem of maximizing an unknown payoff function $f_m : \mathcal{X} \to \mathbb{R}$. We can probe the function f_m by directly querying it at some $x \in \mathcal{X}$ and obtaining a noisy observation $y_{\langle x,m\rangle} = f_m(x) + \varepsilon(x)$, where $\varepsilon(x) \sim \mathcal{N}(0, \sigma^2)$ denotes i.i.d. Gaussian white noise. In addition to the payoff function f_m, we are also given access to oracle calls to some unknown auxiliary functions $f_1, \ldots, f_{m-1} : \mathcal{X} \to \mathbb{R}$; similarly, we obtain a noisy observation $y_{\langle x,\ell\rangle} = f_\ell(x) + \varepsilon$ when querying f_ℓ at x. Here, each f_ℓ could be viewed as a low-fidelity version of f_m for $\ell < m$. For example, if $f_m(x)$ represents the actual reward obtained by running a real physical system with input x, then $f_\ell(x)$ may represent the simulated payoff from a numerical simulator at fidelity level ℓ.

We assume that multiple fidelities $\{f_\ell\}_{\ell\in[m]}$ are mutually dependent through some fixed, (possibly) unknown joint probability distribution $\mathbb{P}[f_1, \ldots, f_m]$. In particular, we model $\mathbb{P}$ with a multiple output Gaussian process; hence the marginal distribution on each fidelity is a separate GP, i.e., $\forall \ell \in [m],\ f_\ell \sim \mathrm{GP}\left(\mu_\ell(x), k_\ell(x, x')\right)$, where μ_ℓ, k_ℓ specify the (prior) mean and covariance at fidelity level ℓ.

Let us use $\langle x, \ell\rangle$ to denote the action of querying f_ℓ at x. Each action $\langle x, \ell\rangle$ incurs cost λ_ℓ, and achieves reward

$$q(\langle x, \ell \rangle) = \begin{cases} f_m(x) & \text{if } \ell = m \\ q_{\min} & \text{o.w.} \end{cases} \tag{4.4.1}$$

That is, performing $\langle x, m \rangle$ (at the target fidelity) achieves a reward $f_m(x)$. We receive the minimal immediate reward $q_{\min}$ with lower fidelity actions $\langle x, \ell \rangle$ for $\ell < m$, even though it may provide some information about f_m and could thus lead to more informed decisions in the future. W.l.o.g., we assume that $\max_x f_m(x) \geq 0$, and $q_{\min} \equiv 0$.

Let us encode an adaptive strategy for picking actions as a policy π. In words, a policy specifies which action to perform next, based on the actions picked so far and the corresponding observations. We consider policies with a fixed budget Λ. Upon termination, π returns a sequence of actions $\mathcal{S}_\pi$, such that $\sum_{\langle x,\ell \rangle \in \mathcal{S}_\pi} \lambda_\ell \leq \Lambda$. Note that for a given policy π, the sequence $\mathcal{S}_\pi$ is a random variable, dependent on the joint distribution $\mathbb{P}$ and the (random) observations of the selected actions. Given a budget Λ on π, our goal is to maximize the expected cumulative reward, so as to identify an action $\langle x, m \rangle$ with performance close to $x^* = \max_{x \in \mathcal{X}} f_m(x)$ as rapidly as possible. Formally, we seek

$$\pi^* = \underset{\pi : \sum_{\langle x,\ell \rangle \in \mathcal{S}_\pi} \lambda_\ell \leq \Lambda}{\arg\max} \mathbb{E}_{\mathcal{S}_\pi} \left[\sum_{\langle x,\ell \rangle \in \mathcal{S}_\pi} q(\langle x, \ell \rangle) \right] \tag{4.4.2}$$

The MF-MI-Greedy Algorithm

We briefly describe MF-MI-Greedy proposed in [57], a mutual information based multi-fidelity Gaussian process optimization algorithm. It consists of two components: an exploratory procedure to gather information about the target level fidelity function via querying lower fidelity functions; and an exploitative procedure to optimize the target level fidelity with the previously gathered information.

Exploration MF-MI-Greedy considers an information-theoretic selection criterion for choosing low fidelity queries. The quality of a low fidelity query $\langle x, \ell \rangle$ is measured as the *information gain per unit cost*, defined as the amount of entropy reduction in the posterior distribution of the target fidelity function divided by the cost of the query: $\frac{\mathbb{I}(y_{\langle x,\ell \rangle}; f_m \mid \mathbf{y}_\mathcal{S})}{\lambda_\ell} = \frac{\mathbb{H}(y_{\langle x,\ell \rangle} \mid \mathbf{y}_\mathcal{S}) - \mathbb{H}(y_{\langle x,\ell \rangle} \mid f_m, \mathbf{y}_\mathcal{S})}{\lambda_\ell}$. Here, $\mathcal{S}$ denotes the set of previously selected actions, and $\mathbf{y}_\mathcal{S}$ denote the observation history. As

Algorithm 1: Multi-fidelity Mutual Information Greedy Optimization (MF-MI-Greedy)

1 **Input**: Total budget Λ; cost λ_i for all fidelities $i \in [m]$; joint GP (prior) distribution on $\{f_i, \varepsilon_i\}_{i\in[m]}$

begin

2 $\quad \mathcal{S} \leftarrow \emptyset$

3 $\quad B \leftarrow \Lambda$; /* initialize remaining budget */

$\quad$ **while** $B > 0$ **do**

$\quad\quad$ /* explore with low fidelity */

4 $\quad\quad \mathcal{L} \leftarrow$ Explore-LF $\left(B, [\lambda_\ell], \text{GP}\left(\{f_\ell, \varepsilon_\ell\}_{\ell\in[m]}\right), \mathcal{S}\right)$

$\quad\quad$ /* select target fidelity */

5 $\quad\quad x^* \leftarrow$ SF-GP-OPT$(\text{GP}\left(\{f_m, \varepsilon_m\}\right), \mathbf{y}_{\mathcal{S}\cup\mathcal{L}})$

6 $\quad\quad \mathcal{S} \leftarrow \mathcal{S} \cup \mathcal{L} \cup \{\langle x^*, m\rangle\}$

7 $\quad\quad B \leftarrow \Lambda - \Lambda_{\mathcal{S}}$; /* update remaining budget */

8 $\quad$ **Output**: Optimizer of the target function f_m

shown in Algorithm 2, this criterion is used greedily to select queries for low fidelity functions. To ensure that the algorithm does not explore excessively, we consider the following stopping conditions: (i) when the budget is exhausted (Line 6), (ii) when a single target fidelity action is better than all the low fidelity actions in terms of the benefit-cost ratio (Line 7), and (iii) when the cumulative benefit-cost ratio is small (Line 8). Here, the parameter β is set to be $\Omega\left(\frac{1}{\sqrt{B}}\right)$ where B is the allocated budget.

Exploitation At the end of the exploration phase, MF-MI-Greedy updates the posterior distribution of the joint GP using the full observation history and searches for a target fidelity action via the (single-fidelity) GP optimization subroutine SF-GP-OPT (Line 5). Here, SF-GP-OPT could be *any* off-the-shelf Bayesian optimization algorithm, such as GP-UCB [71], GP-MI [72], EST [75], and MVES [76], etc. Different from the previous exploration phase which seeks an informative set of low fidelity actions, the GP optimization subroutine aims to trade off exploration and exploitation on the target fidelity, and outputs a single action at each round. MF-MI-Greedy then proceeds to the next round until it exhausts the preset budget, and eventually outputs an estimator of the target function optimizer.

Algorithm 2: Explore-LF: Explore low fidelities

Input: Exploration budget B; cost $[\lambda_\ell]_{\ell\in[m]}$; joint GP distribution on $\{f_i, \varepsilon_i\}_{i\in[m]}$; previously selected items $\mathcal{S}$

begin

- $\mathcal{L} \leftarrow \emptyset$; /* selected actions */
- $\Lambda_{\mathcal{L}} \leftarrow 0$; /* cost of selected actions */
- $\beta \leftarrow \frac{1}{\alpha(B)}$; /* threshold */
- **while** *true* **do**
  - /* greedy benefit-cost ratio */
  - $\langle x^*, \ell^* \rangle \leftarrow \arg\max_{\langle x,\ell\rangle:\lambda_\ell \le B-\Lambda_{\mathcal{L}}-\lambda_m} \frac{\mathbb{I}\left(y_{\langle x,\ell\rangle}; f_m \mid \mathbf{y}_{\mathcal{S}\cup\mathcal{L}}\right)}{\lambda_\ell}$
  - **if** $\ell^* = \mathit{null}$ **then**
    - break ; /* budget exhausted */
  - **if** $\ell^* = m$ **then**
    - break ; /* worse than target */
  - **else if** $\frac{\mathbb{I}\left(\mathbf{y}_{\mathcal{L}\cup\{\langle x^*,\ell^*\rangle\}}; f_m \mid \mathbf{y}_{\mathcal{S}}\right)}{\left(\Lambda_{\mathcal{L}}+\lambda_{\ell^*}\right)} < \beta$ **then**
    - break ; /* low cumulative ratio */
  - **else**
    - $\mathcal{L} \leftarrow \mathcal{L} \cup \{\langle x^*, \ell^* \rangle\}$
    - $\Lambda_{\mathcal{L}} \leftarrow \Lambda_{\mathcal{L}} + \lambda_{\ell^*}$

Output: Selected set of items $\mathcal{L}$ from lower fidelities

Practical Implementation

In Algorithm 2 and the algorithm used for SF-GP-OPT, we need to find the argmax of a function. For the photonic nanostructure experiment in §4.5, this optimization is over a discrete set of candidate queries. Naively, we would need to evaluate the function at each query point in order to determine the optimizer, which is a costly operation. Instead, we devise an approximate optimization step to address this computational challenge. We first directly optimize the function over its *continuous* domain and obtain an optimizer. Then we project the optimizer down to the candidate set by selecting the closest available query point based on Euclidean distance. This approximation scheme takes advantage of existing fast optimizers for continuous functions and becomes necessary for large candidate size.

4.5 Experimental setup and results

Datasets

Our nanophotonic structure is characterized by the five geometric parameters. For each parameter setting, we use a score, commonly called a figure-of-merit (FOM), to represent how well the resulting structure satisfies the desired color

filtering property. By minimizing FOM, we can find a set of high-quality design parameters. Traditionally, FOMs can only be computed through the actual fabrication of a structure and subsequent measurements of its various physical properties, which is a time-consuming process. Alternatively, simulations can be utilized to estimate what physical properties a design will have, e.g. using the Lumerical software. By solving a 2D variant of Maxwell's equations, we could simulate the transmission spectrum of a given nanophotonic device and then compute FOM from it. We could obtain different fidelity level data by controlling aspects of the numerical solution process.

We experiment with three design tasks for filtering light with wavelengths of 550 nm, 650 nm, and 750 nm. For each task, we vary the conformal mesh size and the time-domain solver's total time duration of simulated physical processes to obtain two sets of multi-fidelity data, each with three fidelity levels on 4983 designs.

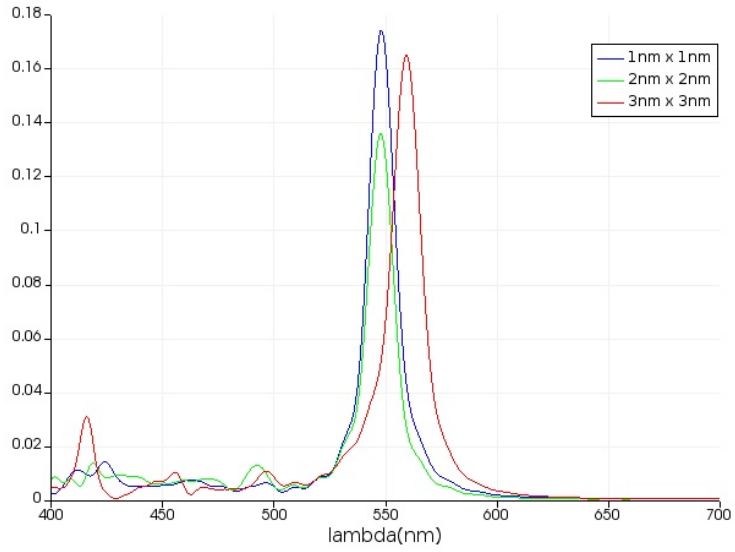
The first set of data is based on different conformal mesh sizes. The mesh size determines how accurate the final results are, with finer meshes leading to more accurate results (Fig. 4.2). We generated the lowest fidelity data using a mesh size of $3\text{nm} \times 3\text{nm}$, the middle fidelity $2\text{nm} \times 2\text{nm}$, and the target fidelity $1\text{nm} \times 1\text{nm}$. The costs, CPU time, are inverse proportional to the mesh size, so we use the following costs $[1, 2.25, 9]$ for our three fidelity function evaluation, respectively.

The second set of data is based on the different total time duration of simulated physical processes for the time-domain solver. Since the transmission spectrum is calculated through Fourier transform of the electromagnetic pulse, which is passed through the color filter, we expect more accurate solutions with longer physical simulation time duration. We generated the lowest fidelity data using a simulation time of 40 fs (femtoseconds), the middle fidelity 70 fs and the target fidelity 100 fs. The costs are proportional to the simulation time, so we use the following costs $[40, 70, 100]$ for our three fidelity function evaluation, respectively.

Experimental Setup

To model the relationship between a low fidelity function f_i and the target fidelity function f_m, we use an additive model. Specifically, we assume that $f_i = f_m + \varepsilon_i$ for all fidelity levels $i < m$ where ε_i is an unknown function characterizing the error incurred by a lower fidelity function. We use Gaussian processes to model f_m and ε_i. Since f_m is embedded in every fidelity level, we can use an observation from any fidelity to update the posterior for *every* fidelity level. We use

(a) Transmission spectra



(b) Electric field profiles at 550nm

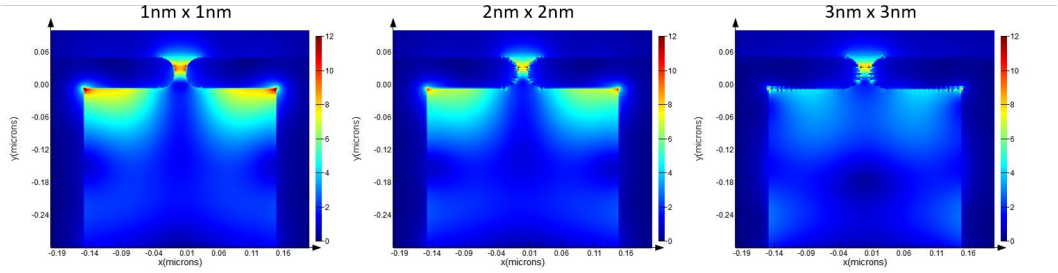


Figure 4.2: Influence of mesh size on the results of FDTD simulations.

square exponential kernels for all the GP covariances, with hyperparameter tuning scheduled periodically during optimization. Following prior work on practical Bayesian optimization [85], we use 10% of the total budget for initialization. For multi-fidelity methods, the initialization budget is spent on randomly querying the lowest fidelity function. For the single-fidelity method, it is spent on randomly querying the target fidelity function. For all experiments, we use a total budget of 100 times the cost of target fidelity function call f_m. Every method is run 20 times to compute its mean and standard error.

Compared Methods

Our framework is general, and we could plug in different single fidelity Bayesian optimization algorithms for the SF-GP-OPT procedure in Algorithm 1. In our experiment, we choose to use GP-UCB as one instantiation. We compare with MF-GP-UCB [82] and GP-UCB [71]. MF-GP-UCB relies on several hyper-parameters in the algorithm, and we keep the same approach to choosing them as described in [82].

Besides the Bayesian optimization based method, we also compare with a common heuristic called Particle Swarm Optimization, which is inspired by the social behavior of animals and is used for nanophotonic structure designs [67, 68]. We use built-in MATLAB implementation of this algorithm. We specify a population of 5 particles and run Swarm optimization for 20 iterations, totaling $5 \times 20 = 100$ evaluations of the target fidelity function. All other algorithm parameters are kept at default MATLAB values.

Optimizing Figure of Merit

Figure 4.3 and Figure 4.4 show the results of this experiment. As usual, the x-axis is the cost and y-axis is Figure of Merit, and smaller is better. After a small portion of the budget is used in initial exploration, MF-MI-Greedy (red) is able to arrive at a better final design compared with MF-GP-UCB, GP-UCB and Particle Swarm. MF-MI-Greedy tends to have a worse figure of merit at the beginning because the initial explorations in the lower fidelity do not yield FOM scores on the target fidelity, so essentially, it has a late start in all the plots because it starts querying the target fidelity late. However, the advantage of exploring lower fidelities becomes apparent once the exploitation phase starts in the target fidelity level, as seen by the rapid convergence to low FOM designs.

(a) 550nm wavelength (b) 650nm wavelength (c) 750nm wavelength

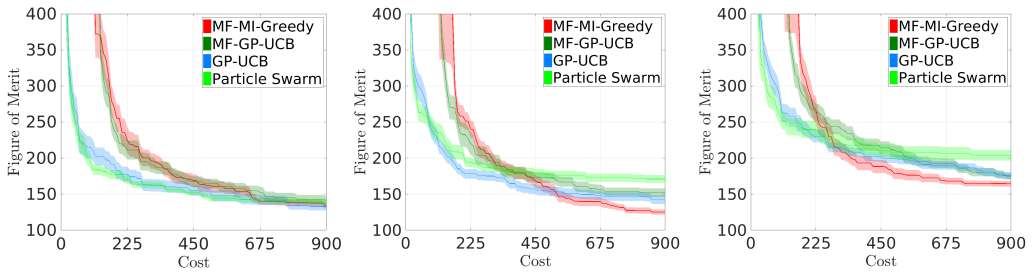


Figure 4.3: Multi-fidelity based on conformal mesh size. Every method is run 20 times, and we plot the mean plus/minus one standard error in the figures.

(a) 550nm wavelength (b) 650nm wavelength (c) 750nm wavelength

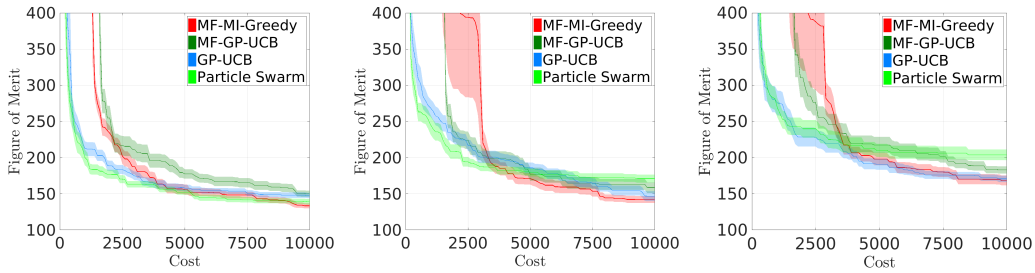


Figure 4.4: Multi-fidelity based on conformal simulation time. Every method is run 20 times, and we plot the mean plus/minus one standard error in the figures.

4.6 Optimization of universal metasurface

Rapid advances in chip-based nanophotonics technology have stimulated researchers to develop metasurfaces, which are low-profile, ultrathin, lightweight integrated devices with consumer electronics, medical, and aerospace applications [86, 87, 88]. Specifically, metasurfaces consist of arrays of artificially engineered subwavelength optical elements that have enabled revolutionary light manipulation by precise modulation of their local properties. In recent years, active metasurfaces, i.e., metasurfaces whose optical properties are dynamically reconfigurable after fabrication, have increasingly gained interest by offering a platform for dynamic control of light through application of external stimuli (e.g., voltage) to tune the constitutive optical properties of subwavelength antennas [89]. By independently addressing each nanostructured element, the wavefront of scattered light can be dynamically tailored through a pixel-by-pixel reconfiguration. In particular, electro-optic tuning mechanisms have been employed to demonstrate optical modulation of phase and

amplitude over wide ranges, as well as high switching speeds [20, 90].

Furthermore, the development of active metasurfaces has enabled the realization of a "universal metasurface" allowing for spatiotemporal control of fundamental properties of light and the combination of versatile functionalities in a single device. By employing external voltages, tunable metasurfaces can be realized by incorporating electro-optically active materials. Transparent conducting oxides (TCOs) that undergo a reliable and reproducible index change in response to an optical or electrical stimulus, provide high modulation speed, low energy consumption, robustness, wide tuning range, leading them to establish superiority over other active materials. Despite several studies conducted on actively modulating the response of metasurface devices, developing an active metasurface platform operating in the NIR wavelength range that dynamically tailors the wavefront of scattered light in an optimal manner through an element-by-element reconfiguration is still an outstanding research challenge.

In this chapter, we focus on the beamsteering application of a universal metasurface. Optimization of such devices poses a big challenge, as there are multiple design levels. On one side, we need to design individual elements (called pixels or antennas). On the other side, for every steering angle we need to find such values of control variables (voltages) so that beamsteering performance is optimal. This problem of co-design provides a rich opportunity for algorithmic development.

We focus mostly on the optimization of control variables, keeping antenna design fixed. We use design from [58] to test novel optimization approaches. In this design metasurface elements, which tune the phase and amplitude of reflected NIR light, are chip-integrated in 1D arrays of 96 pixels.

Far-field distribution of such a metasurface is a superposition of field, emitted by identical individual pixels, and in 1D case it can be written as:

$$E(\theta) = E_{pixel}(\theta) \sum_j \left[A(V_j) \cdot \exp(i\phi(V_j)) \cdot \exp(ikx_j \sin\theta) \right], \tag{4.6.1}$$

where θ - steering angle, E - electric far-field of the whole metasurface, E_pixel - electric far-field of an individual pixel, V_j - voltage applied to j-th pixel, x_j - position of j-th pixel, A and ϕ - amplitude and phase of reflection coefficient, and $k = 2\pi/\lambda$ - wavenumber of used light at wavelength λ.

All dependence on control variables is encoded in the term under summation in Eq. 4.6.1. In fact, this term represents the collective effect of amplification of

far-field distribution by array of pixels. Hence, it is commonly called array factor. We also note, that since there is one-to-one correspondence between phase and voltage, optimizing voltages and phases is equivalent, and we choose to optimize phases directly. In this work, we redefine the array factor in the following way:

$$AF(\theta) = \left| \sum_j \left[A(\phi_j) \cdot \exp(i\phi_j) \cdot \exp(ikx_j \sin\theta) \right] \right|^2, \tag{4.6.2}$$

which corresponds to the collective effect of intensity multiplication.

Our goal is to maximize steering into angle θ. A natural figure of merit in this case is directivity $D(\theta)$, which in 1D can be defined as the ratio of intensity $I(\theta)$ at some particular angle θ to average intensity over all angles:

$$D(\theta) = \frac{\pi I(\theta)}{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} I(\theta) d\theta} \tag{4.6.3}$$

Note that for the purposes of computing directivity, intensity can be taken as a square of the absolute value of array factor (conventional definition). Hence, for the rest of the chapter we use intensity and array factor interchangeably, and array factor is defined by Eq. 4.6.2, so that directivity is actually computed as:

$$D(\theta) = \frac{\pi AF(\theta)}{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} AF(\theta) d\theta} \tag{4.6.4}$$

Since θ is a continuous variable, for simplification we discretize it with some step. Thus we now have a defined optimization objective: for every θ find such ϕ_j's, that $D(\theta)$ is maximized. This is what we call inverse problem in this work.

We use the following parameters of a 1D metasurface array of 96 pixels: operating wavelength λ = 1510nm, pixels' positions are periodic with a spacing of 400nm (except when noted otherwise), amplitude-phase relationship is simulated in Lumerical for a pixel designed in [58], and it is fitted with 6th degree polynomial (Fig. 4.5), that is then used as $A(\phi_j)$ in 4.6.2. Note, that in this case, phase range is limited to roughly 272°.

For analysis and comparison purposes, we optimize three different devices, that have different amplitude-phase relationships. We define them in the following way.

- Ideal device - metasurface, in which pixels have constant reflection amplitude equal to one and full reflection phase range of 360°.

- Semi-realistic device - metasurface, in which pixels have constant reflection amplitude equal to one, but reflection phase range limited to roughly 272°.
- Realistic device - metasurface, in which pixels have constant covariant reflection amplitude-phase relationship defined by Fig. 4.5, reflection phase range limited to roughly 272°.

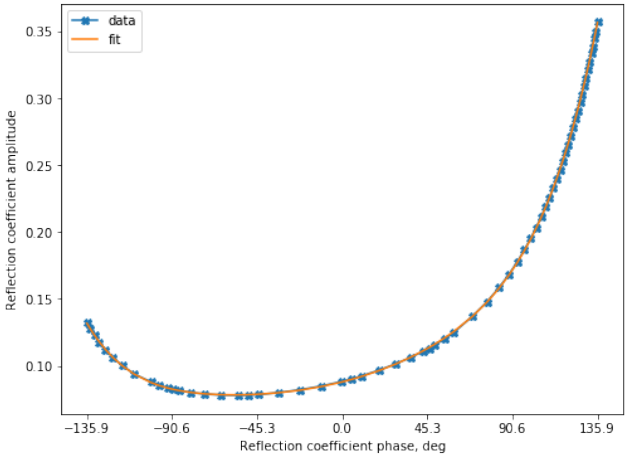


Figure 4.5: Reflection coefficient amplitude-phase relation.

4.7 Solving inverse problem using deep-learning

As stated in the previous section, we are interested in finding such a functional relationship $\phi_j(\theta)$ for all pixels, so that resultant directivity $D(\theta)$ is maximized. Since neural networks are universal function approximators [91], it makes sense to try using them in our case. Nowadays, rapid iteration and experimentation with neural networks is possible due to their flexibility and availability of high-quality software implementations. In this work, we use TensorFlow Python software package from Google. We ran our code on Nvidia GeForce 1080Ti GPU. All computations presented below ran for no longer than 20 minutes.

Solving inverse problem directly

In forward calculation, we compute outputs (array factor) using inputs (pixels phases). In inverse calculation, we change the order of inputs and outputs, so that we get pixel phases using given array factors. Reversing forward calculation in general is not easy, and it is commonly performed using optimization techniques.

As a warm-up exercise, we consider a simple case of an ideal device, driven by conventional linear phase profiles. We train neural networks on pairs of (array

factor, pixels phases), where we use array factors as inputs to neural networks, and phases as outputs (Fig. 4.6). We use mean squared error (MSE) between predicted and true phases as a neural network loss function. Our array factor curves are discretized with a step of 0.1°, so that every curve has 1801 points.

A dense neural network with a few layers is able to pick up an inverse relationship (Fig. 4.7). Fig. 4.7b shows comparison between true phases (orange curve) and predicted ones (blue curve). Fig. 4.7a shows a comparison between array factors that are calculated from true phases (orange curve) and from predicted phases (blue curve). The orange array factor in Fig. 4.7a was used as an input into neural network, which produced blue phase profile on Fig. 4.7b.

This is an encouraging result, showing that neural networks can indeed approximate an inverse relationship. However, it is very limited in usability, since only those phase relationships, that were in the training data, can be learned.

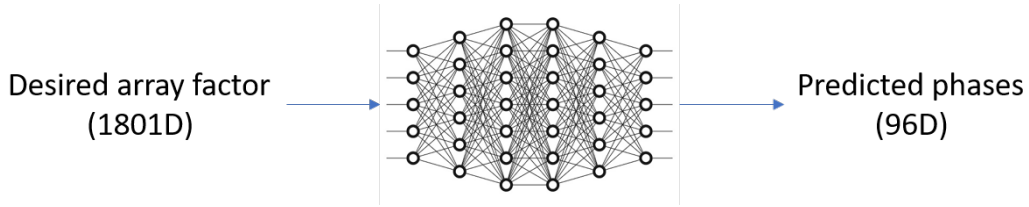
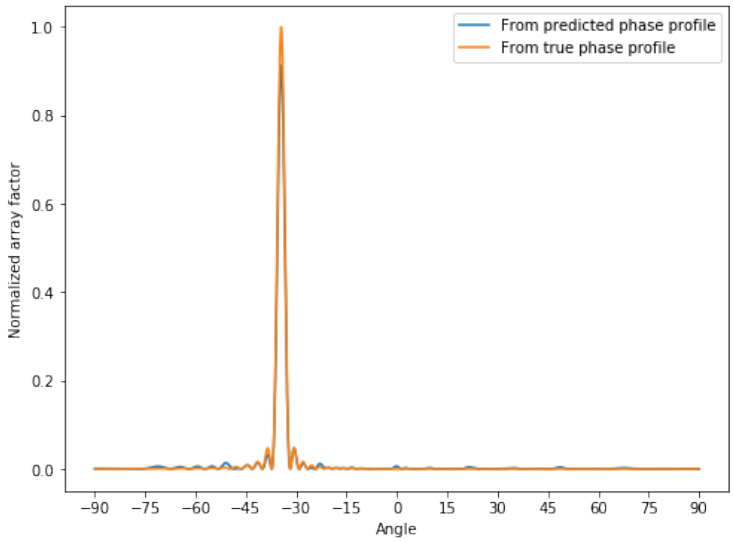


Figure 4.6: Solving inverse design problem directly using neural networks.

(a) Array factor



(b) Phases

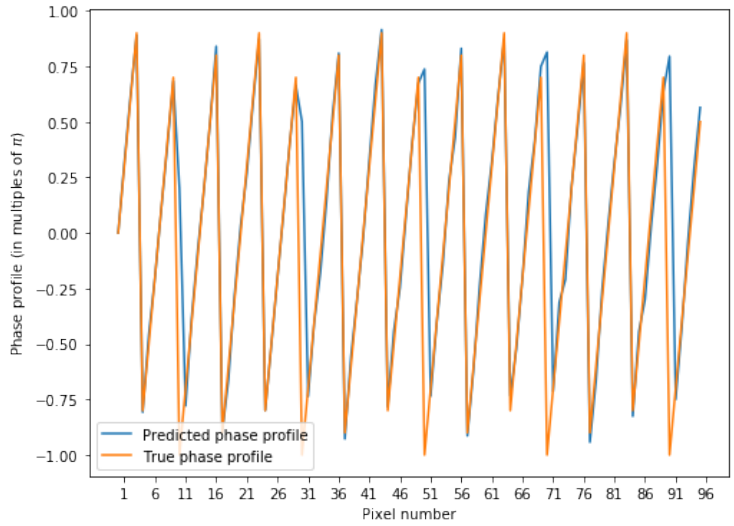


Figure 4.7: Results of neural network approximating inverse calculation for linear phase profiles applied to ideal device.

Solving inverse problem using autoencoder-like architecture

A more interesting neural network architecture, which can learn phase relationships on its own, is based on auto-encoders (Fig. 4.8). In this case, we have two parts of a network: the encoder, which takes array factor and converts it to phases, and the decoder, which takes this phases and reconstructs array factor. Fortunately for us, we have a relatively straightforward relationship between phases and array factors given by Eq. 4.6.2, that can be used as a decoder directly. So, we only

need to train the encoder part of the network, and we only provide array factors (without phase profiles). Loss function in this case is mean squared logarithmic error between true and reconstructed array factors. The encoder never sees "true" phases that generated training array factors.

Overall, this scheme can be viewed as an optimizer that tries to reconstruct given array factor. If we only wanted to reconstruct one array factor, we would not need a neural network at all, we could optimize phases directly. Neural network here plays a role of a universal approximator of the inverse relationship, keeping and sharing information that is learned during training, since we are actually trying to reconstruct all array factors in the training set using the same neural network. We can also incorporate ideal array factors into training data to find phase profiles that can reconstruct them.

This autoencoder-like architecture is able is find such phase profiles, that reconstruct given array factors (Fig. 4.9). Even though this architecture is more general than previously considered one, we still need to engineer training array factors.

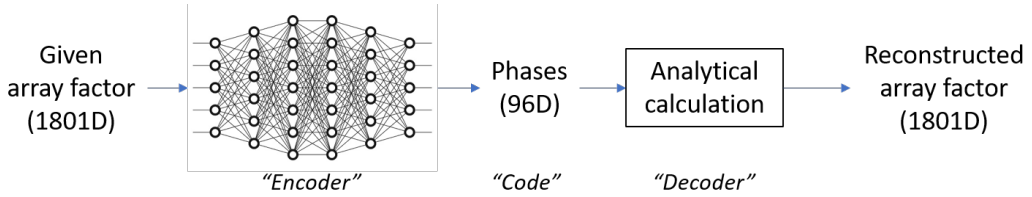
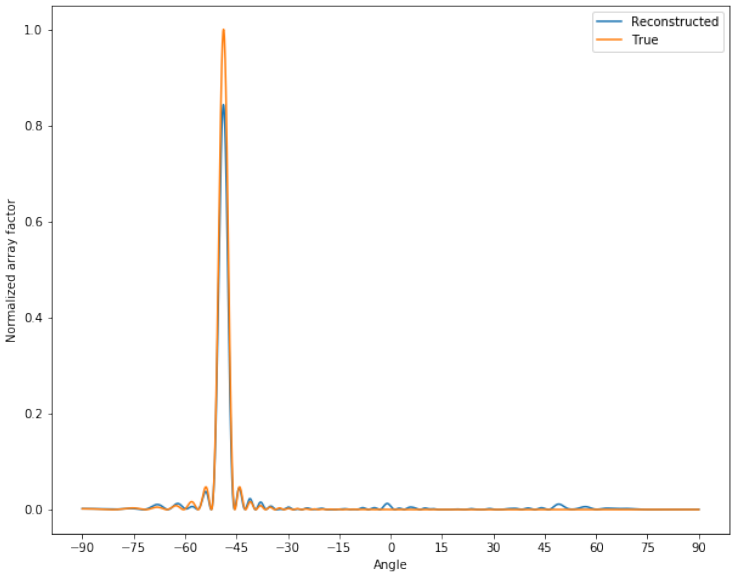


Figure 4.8: Solving inverse design problem directly using autoencoder-like network.

(a) Array factor



(b) Phases

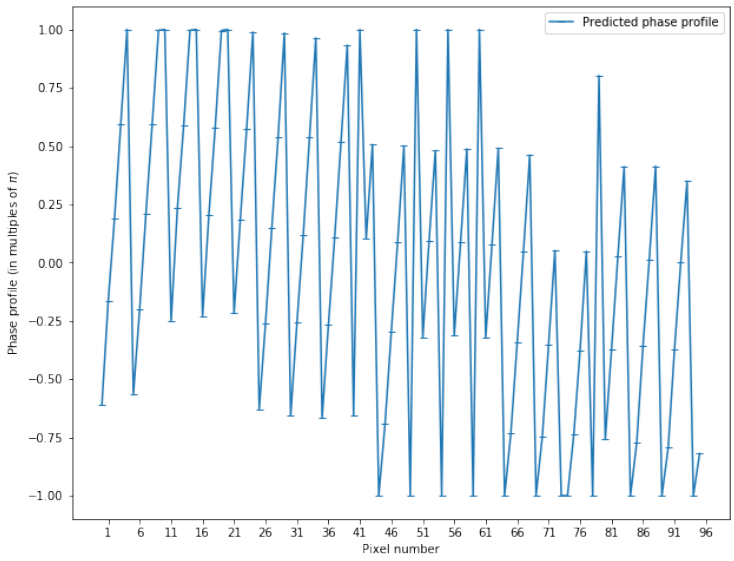


Figure 4.9: Results of autoencoder-like neural network.

4.8 Optimizing directivity with perceptron-like network

Inspired by the success of using neural networks for approximating inverse relationship in the previous section, we thought we could use directivity as a loss function during training of neural networks. We implemented custom layers and custom loss function in TensorFlow to realize this idea. Since optimizers in deep-

learning software are always minimizing loss function, we use negative directivity as a loss function for neural network training.

At the same time, our objective is to find such a functional relationship between phases and steering angles, so that directivities computed using these phases for different angles are maximized. So, we want to be able to input an arbitrary steering angle, and we want to have optimal phases as our output.

As an extreme limit, we considered perceptron-like architecture (Fig. 4.10). We call it perceptron-like because it does not contain hidden layers. In this approach, we represent discretized steering angles as one-hot vectors (all components are zero, except for one that corresponds to an index of discretized angle), so that every input neuron corresponds to a separate angle. We then directly connected input neurons to output neurons, which correspond to our 96 pixels. Weights of such connections are actually optimized phases.

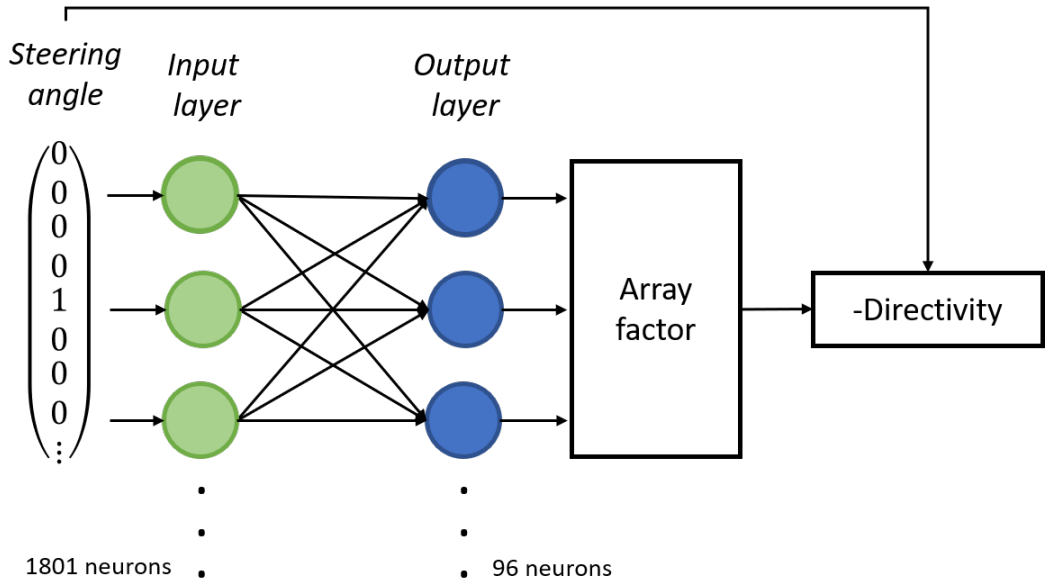


Figure 4.10: Solving inverse design problem directly using perceptron-like network.

In this setting, there are no hidden layers, and one can argue that we are performing independent optimization of phases for each angle. This can serve as a means of finding the optimal phases limit and judging how well architectures with hidden layers can approach this limit. Hidden layers allow the sharing of information between phases optimization for different steering angles, so in principle such optimization could be more efficient. We experimented with different architectures of such a neural network, and we indeed found that we could converge faster, though

solutions were worse. One consideration to keep in mind is that it does not make much sense to use such a neural network that has more weights than a neural network without hidden units (otherwise, it is a waste of resources). So we tried architectures with a smaller number of weights. Such an exercise can be considered as an attempt to compress information, contained in perceptpron-like network weights, so it is not surprising that quality of such optimization is lower.

Figures 4.11, 4.12, and 4.13 give us results of optimization runs using this perceptron-like network for different devices (ideal, semi-realistic, and realistic). We compare optimized directivities (orange curves) with directivities that are computed for the ideal device using linear phase profiles (conventionally used for beamsteering based on physical intuition).

In case of the ideal device (Fig. 4.11), we can noticeably beat conventional phase profiles, and in cases of semi-realistic and realistic devices, we can approach this performance for small steering angles, and can still beat it at higher angles (Fig. 4.12, 4.13).

Optimized phase profiles for the real device look significantly different from a linear phase profile (Fig. 4.14, 4.15). This is a demonstration that we can outperform conventional intuition is some cases using optimization techniques for inverse design.

Using this framework, we can optimize for different objectives. As an example, we considered steering into two angles. We chose the product of directivities computed for two angles as a figure of merit in this case. Fig. 4.16 shows an example of such an optimization: we successfully find such a phase profile, that gives us two sharp peaks in array factors at specified angles.

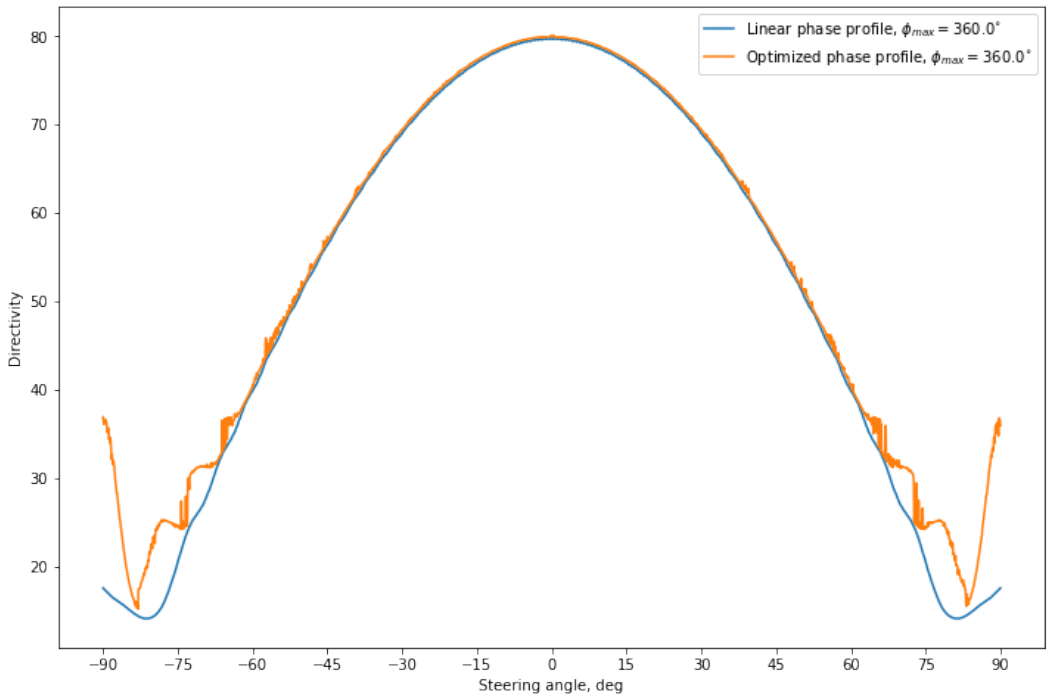


Figure 4.11: Optimized directivity of ideal device (orange curve) in comparison with directivities computed using conventional linear phase profile (blue curve).

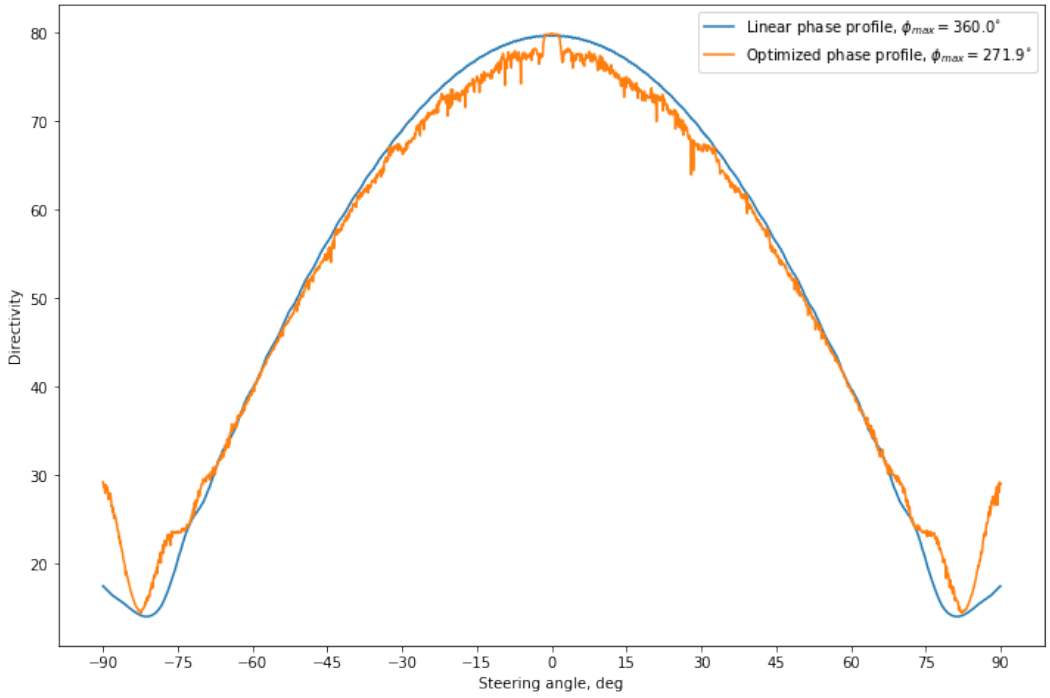


Figure 4.12: Optimized directivity of semi-realistic device (orange curve) in comparison with directivities computed using conventional linear phase profile (blue curve).

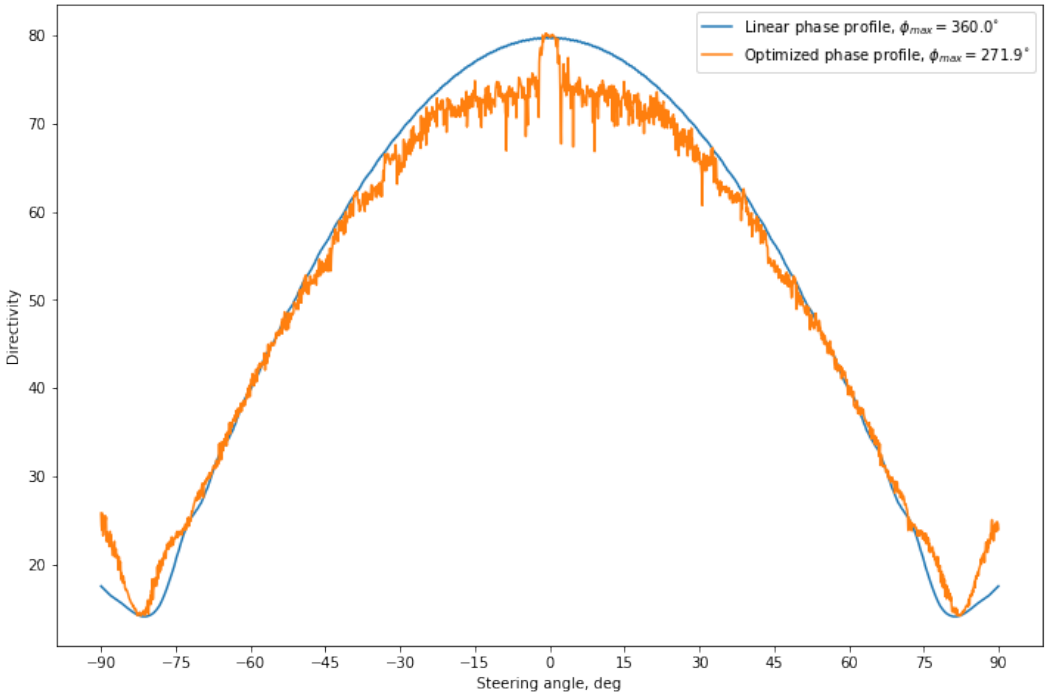


Figure 4.13: Optimized directivity of realistic device (orange curve) in comparison with directivities computed using conventional linear phase profile (blue curve).

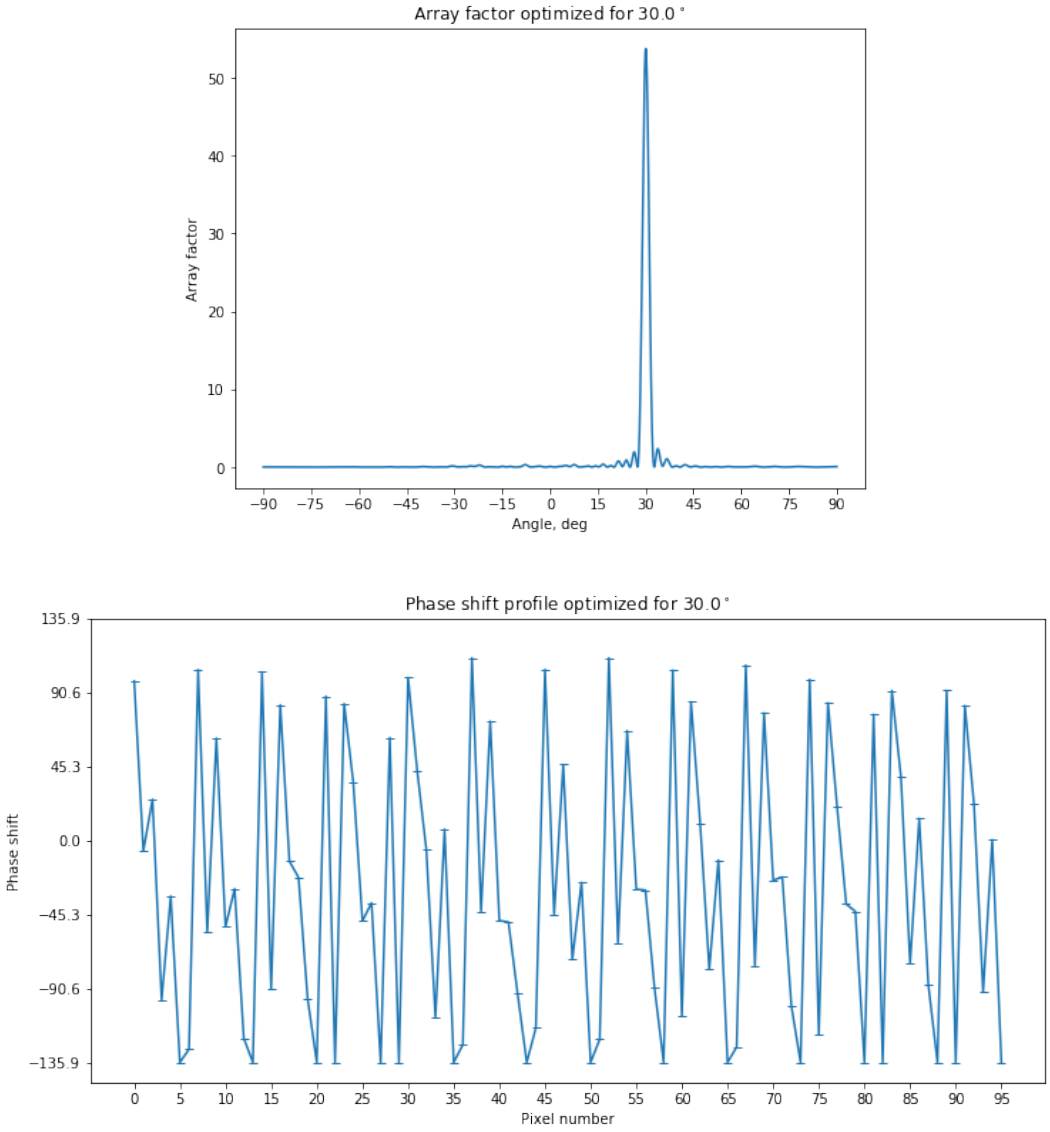


Figure 4.14: Array factor (top) and phases (bottom) optimized for steering into 30°using real device.

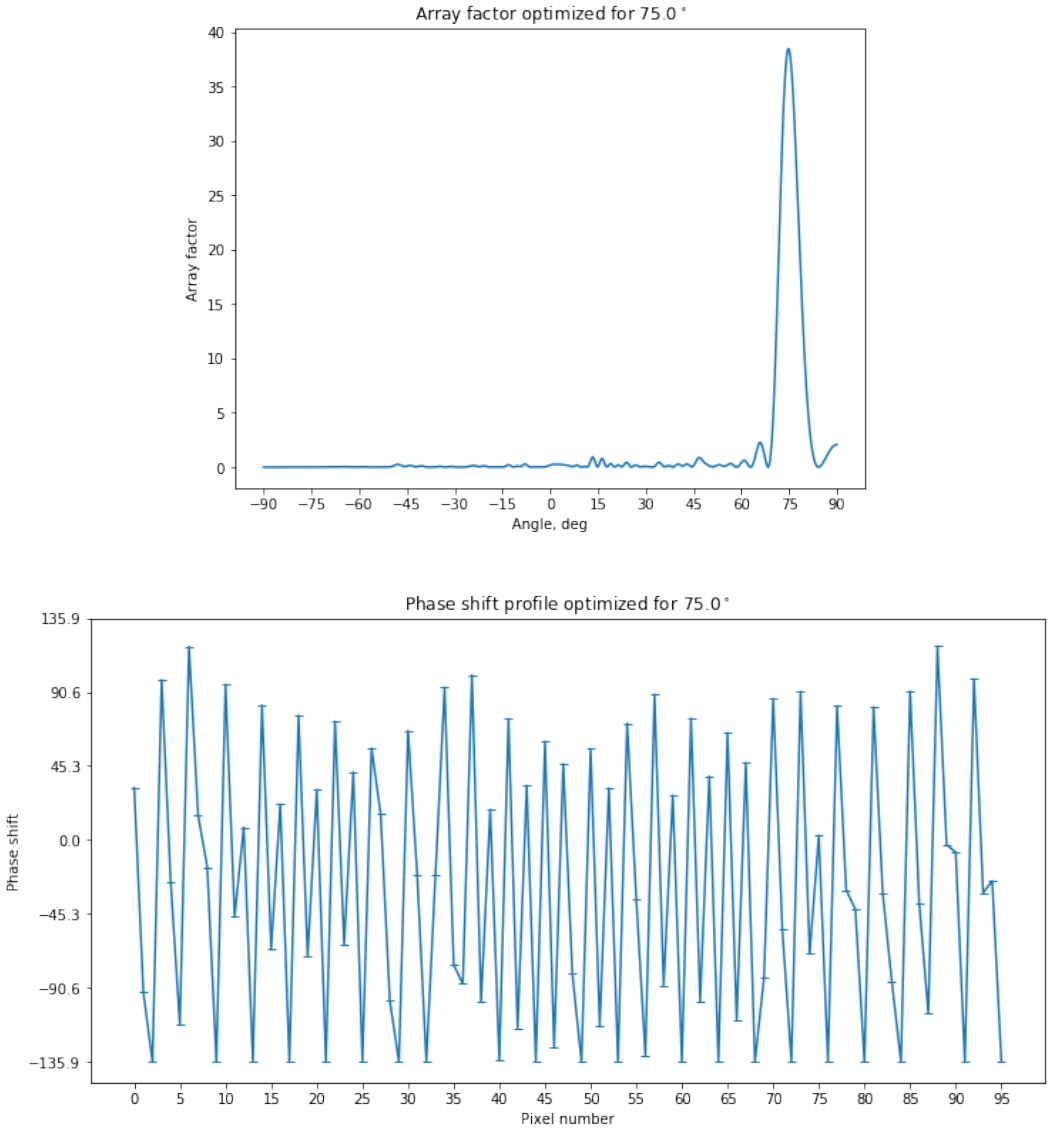


Figure 4.15: Array factor (top) and phases (bottom) optimized for steering into75°using real device.

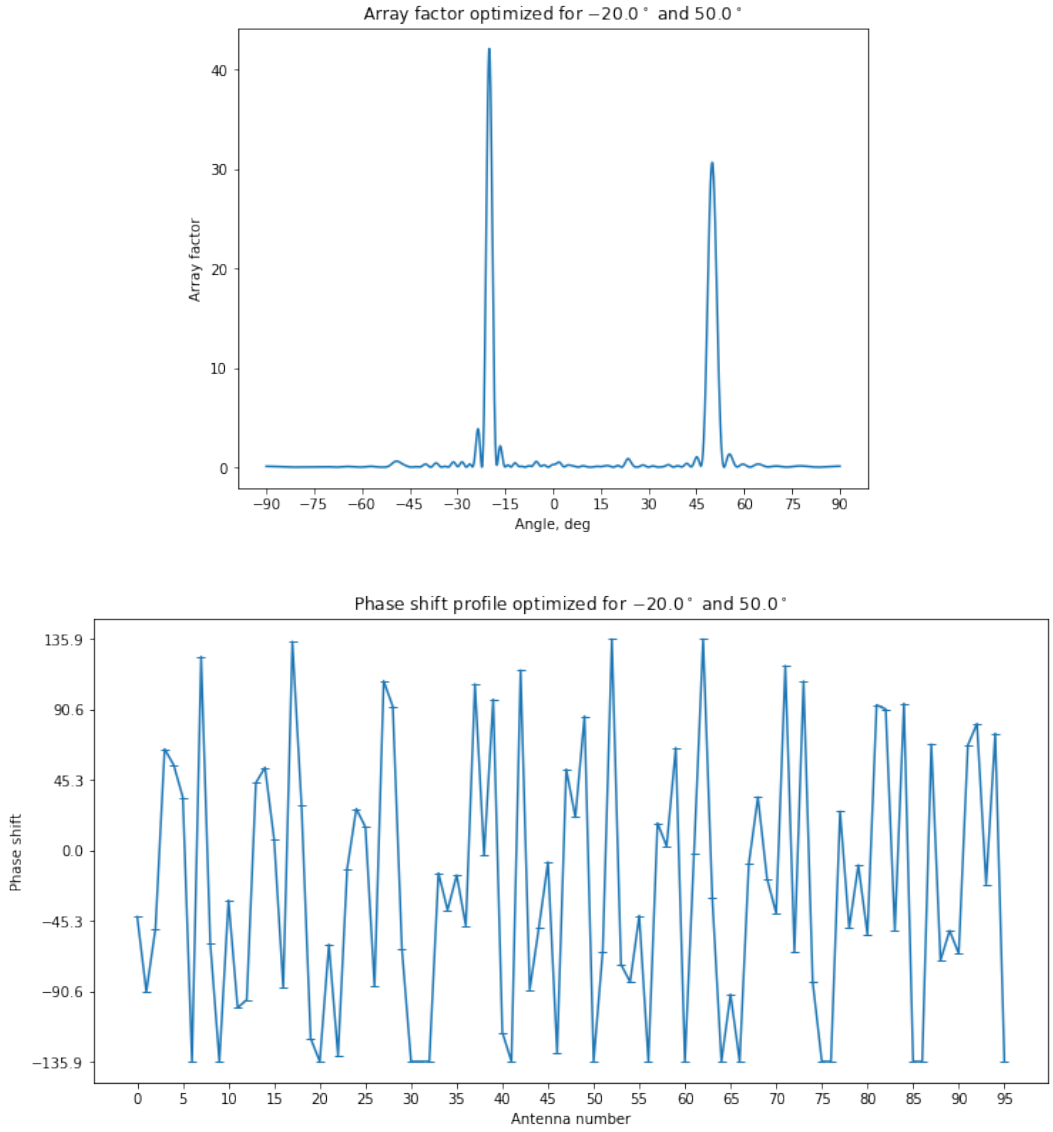


Figure 4.16: Array factor (top) and phases (bottom) optimized for steering into two angles using real device: -20°and 50°.

4.9 Optimization of positions of metasurface pixels

We now consider another important objective: optimization of positions of pixels. This task becomes even more important if we want to use large spacing between pixels. Results described in previous sections are dealing with the case of deeply subwavelength spacing. This puts limitations on types of pixels that we can use, increases fabrication complexity, and restricts size of our metasurface. Hence, it is actually desired to have larger spacing between pixels.

If we have a spacing of larger than half a wavelength, then periodic positioning of pixels will inevitably result in appearance of diffraction orders in far-field, which are impossible to remove using optimization of phases only. However, it has been shown previously, that by adjusting positions of pixels, we can suppress these diffraction orders [92].

We implemented co-optimization of phases and positions under the same Adam optimization algorithm in TensorFlow. Since positions are the same for all steering angles, we used figure of merit that is averaged across all angles. We noticed, that directivity is not a good figure of merit, when the average spacing between antennas is larger than a wavelength, so we used sidelobe suppression ratio instead. This figure of merit is defined as a ratio between biggest and second biggest peaks in far-field intensity.

To make a fair comparison between different cases that correspond to the same average spacing, we kept array apertures unchanged during adjustment of positions by fixing positions of first and last antennas. We then considered cases of average spacing of 2 μm, 10 μm, and 100 μm (there are 96 pixels in all cases).

Optimizing positions at an average spacing of 2 μm was the easiest exercise across the outlined three cases. Fig. 4.17 compares periodic positions (blue curve) with optimized ones (orange curve). We see that optimized positions appear to be irregular. This irregularity leads to suppression of diffraction orders (Fig. 4.19). Fig. 4.18 provides comparison of directivities in three cases: optimization of positions and phases of the ideal device (blue curve), optimizations of positions and phases of the realistic device (green curve), and optimization of phases only while keeping positions of pixels completely random (orange curve). Interestingly, randomization of positions allows us to suppress diffraction orders, while still being able to steer the light into a wide angle range using adjustment of control variables (phases). Obviously, random positioning is very unlikely to give optimal distribution of positions, hence, by allowing optimization of pixel positions, we are able to

improve directivity even further.

When we go to an average spacing of 10 μm (Fig. 4.20, 4.21), it becomes harder to optimize positions, especially in case of the realistic device (Fig. 4.21). It is even more pronounced in the case of average spacing of 100 μm ((Fig. 4.20, 4.21). Despite that, we are still able to suppress diffraction orders and steer the beam in a wide range of angles with a considerable directivity.

All of this indicates that adjusting positions is a very promising technique in controlling the properties of metasurfaces. Our findings suggest it is worth spending more time on development of proper optimization algorithms, capabale of determining optimal positions and phases simultaneously.

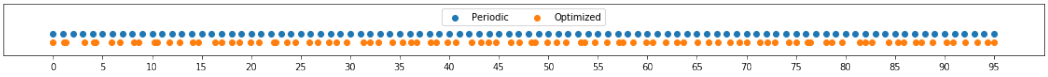


Figure 4.17: Comparison of periodic (blue dots) and optimized (orange) positions for array with average spacing between pixels equal to 2 μm.

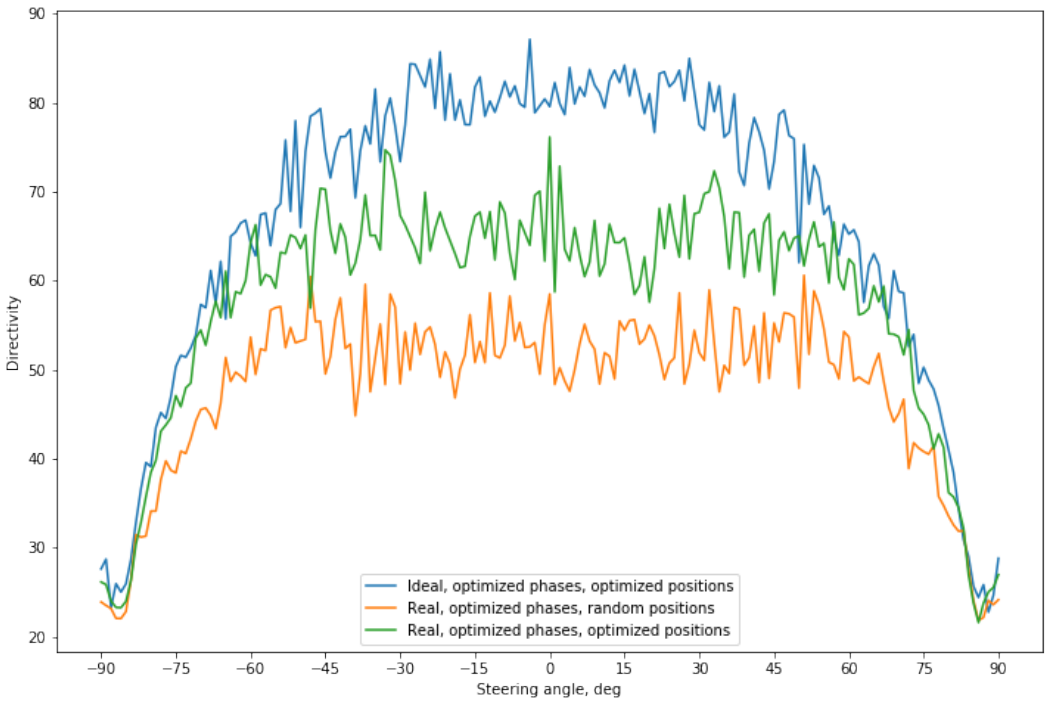


Figure 4.18: Comparison of directivities between: blue - ideal device with optimized phases and positions; green - realistic device with optimized phases and positions; orange - realistic device with optimized phases but random positions.

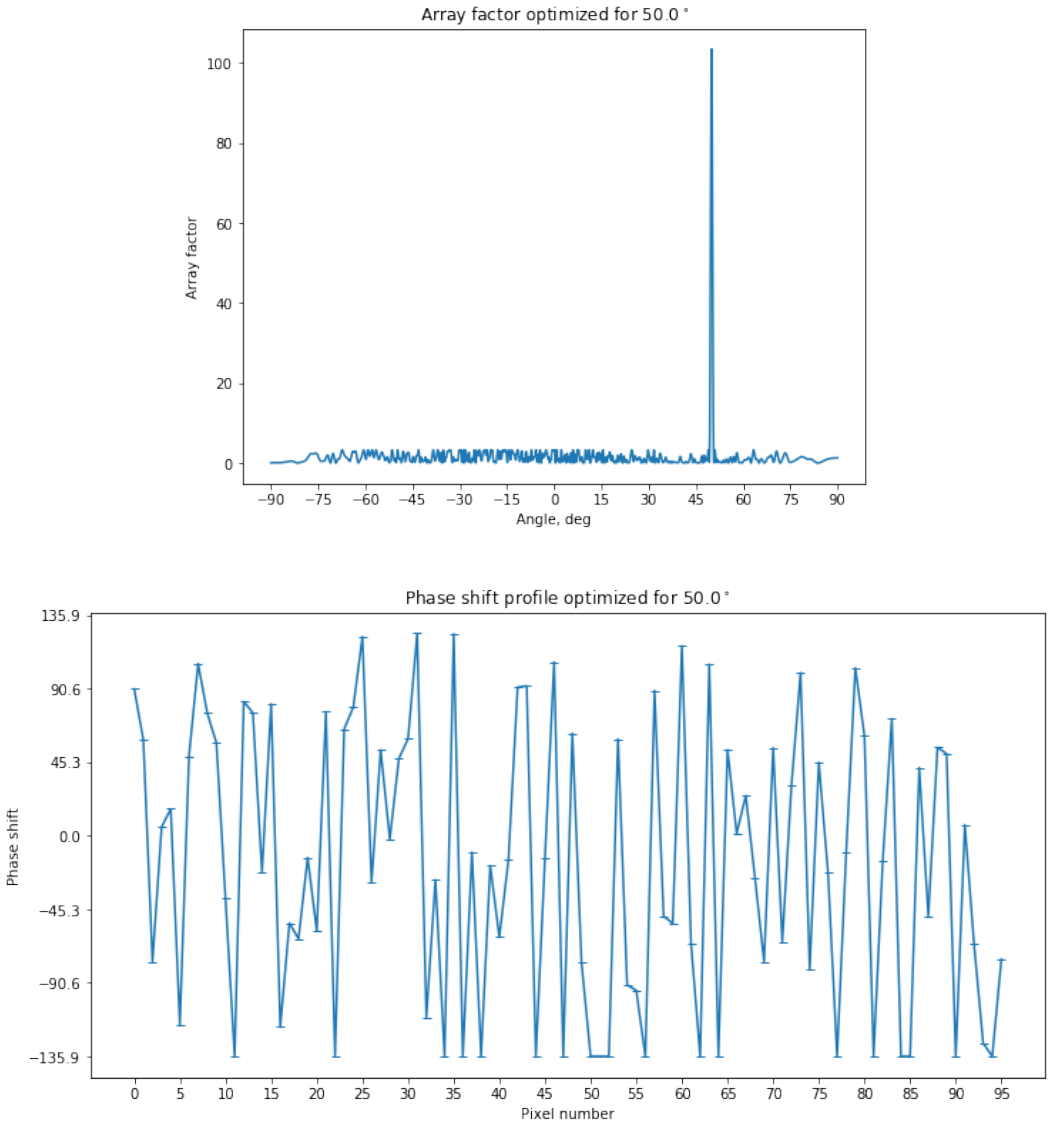


Figure 4.19: Array factor (top) and phases (bottom) optimized for steering into 50°for device with average spacing of 2 μm.

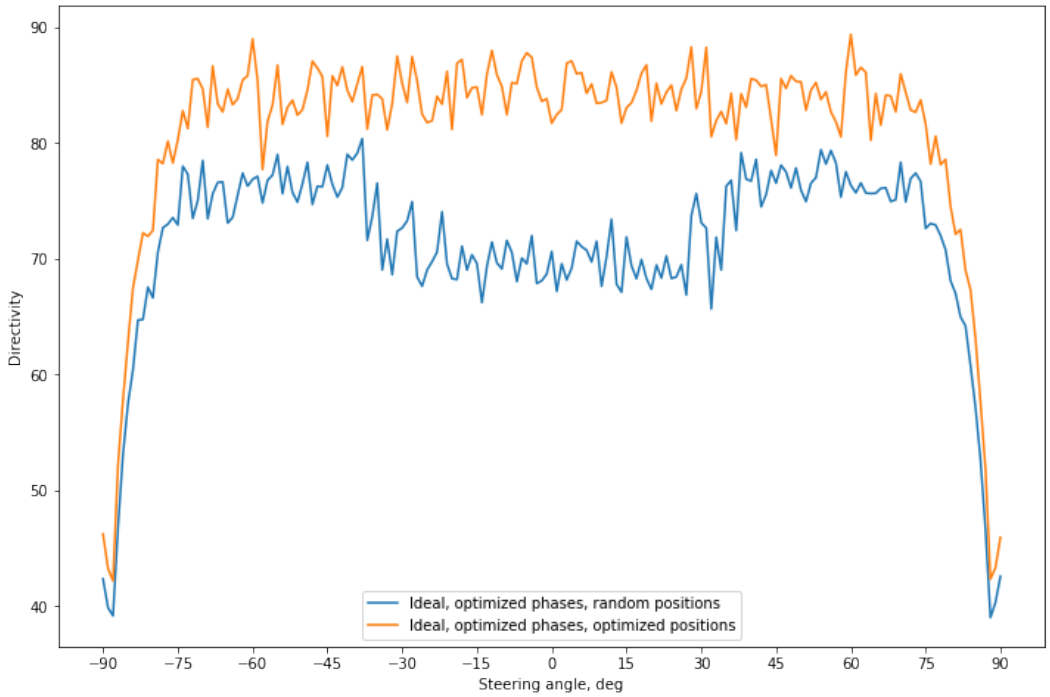


Figure 4.20: Comparison of directivities for cases of optimized and periodic positions for ideal device with average spacing of 10 μm.

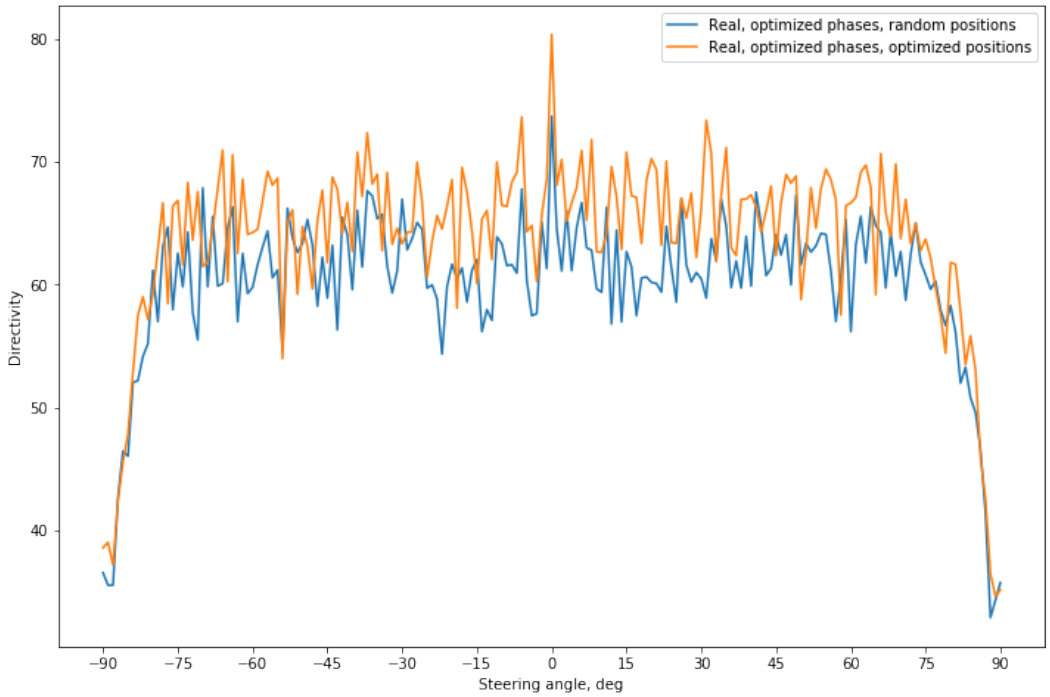


Figure 4.21: Comparison of directiviites for cases of optimized and periodic positions for realistic device with average spacing of 10 μm.

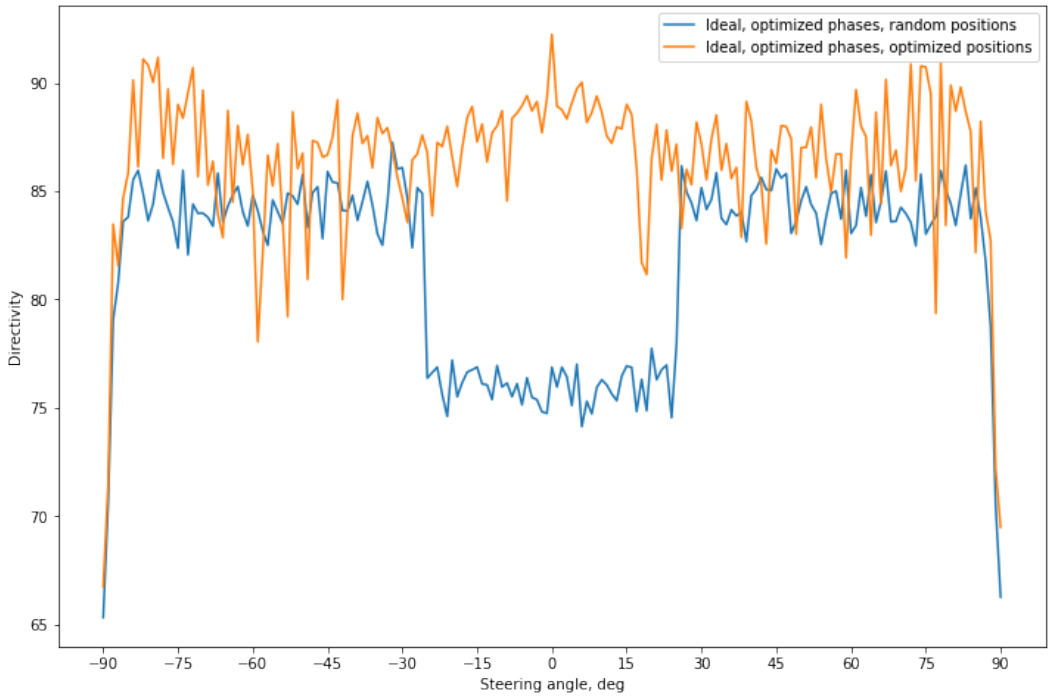


Figure 4.22: Comparison of directivities for cases of optimized and periodic positions for ideal device with average spacing of 100 μm.

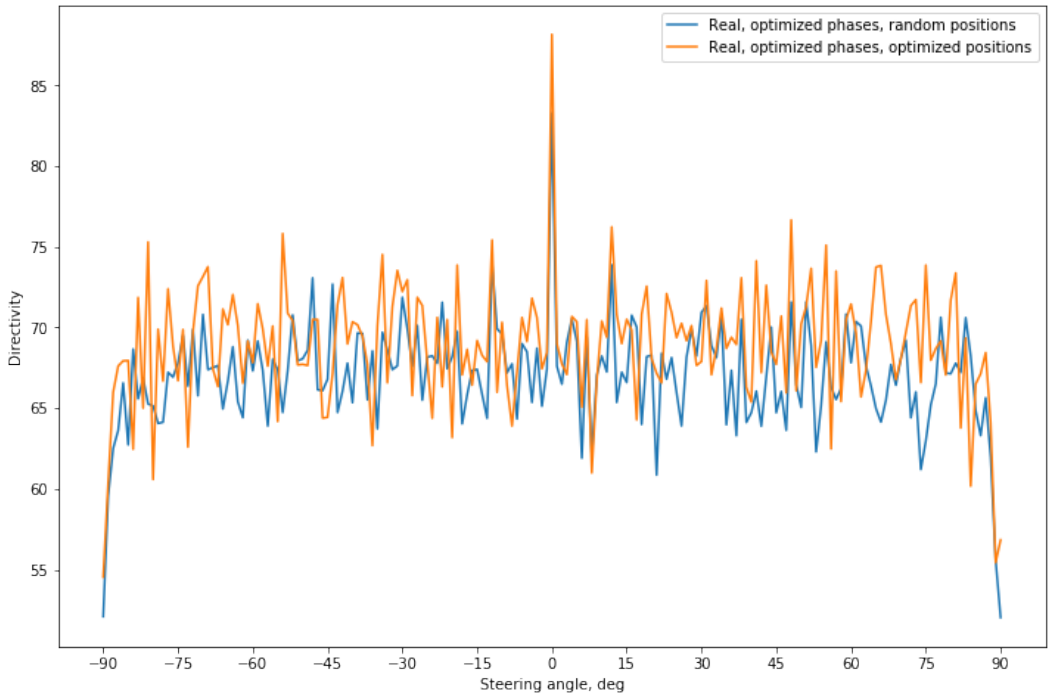


Figure 4.23: Comparison of directivities for cases of optimized and periodic positions for realistic device with average spacing of 100 μm.

4.10 Summary and outlook

In this chapter, we investigated applications of advanced optimization and machine learning techniques to important problems in nanophotonics design. We considered two problems that have different dimensionality of parameter space.

Firstly, we optimized the transmission properties of plasmonic mirror color filters, where we have access to multiple numerical simulators with different fidelity levels and computational costs. Since this problem has only five parameters, we were able to consider several derivative-free global optimization algorithms, including a commonly used approach in the nanophotonics community, and two recently developed multiple-fidelity Bayesian optimization approaches. Our results on several pre-collected nanophotonics datasets demonstrate the compelling performance of the multiple-fidelity Bayesian optimization approach. These experiments suggest that there is a great potential in utilizing cheap, multi-fidelity simulations to aid the discovery of the optimal photonic nanostructures.

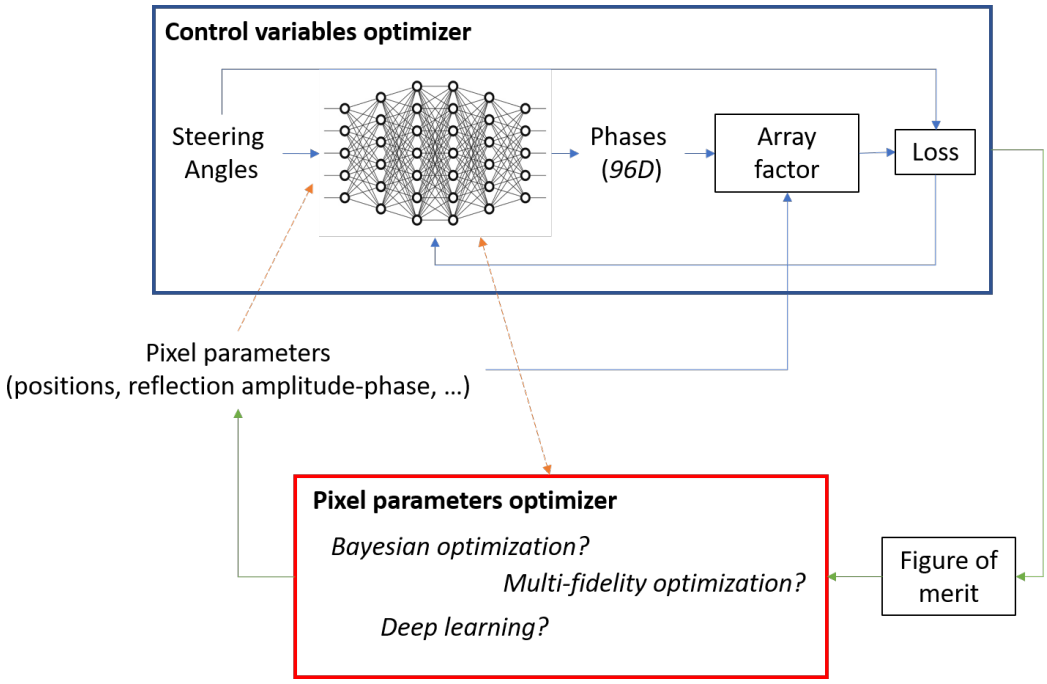


Figure 4.24: Possible schema for co-design of metasurface parameters.

Secondly, we considered the problem of optimizing dynamic and static variables of arrays of pixels, that comprise active metasurfaces. This is a high-dimensional problem with multiple layers of complexity and design. We showed the promise of using neural network inspired optimization techniques for determining optimal phases and positions of pixels in metasurface for beamsteering.

Our analysis allows us to formulate possible schemes for co-design of different parameters of metasurfaces (Fig. 4.24). In one setting, we can treat optimization of dynamic (phases/voltages) and static variables (individual pixels positions, geometry, materials) as a somewhat independent optimization procedure. We can use our fast optimizer for determining optimal phases while keeping all static parameters fixed. Then we can use the resultant overall figure of merit as a target for a separate optimizer that adjusts static parameters. This external optimizer could be the same multi-fidelity Bayesian optimization that we considered in the case of optimizing plasmonic mirror filters, or it could be some other schemes, or a combination of schemes.

In conclusion, we demonstrated how advanced optimization techniques can allow us to design novel nanophotonic devices.

BIBLIOGRAPHY

[1] Dror G. Feitelson. *Optical computing: a survey for computer scientists*. Cambridge, MA, USA: MIT Press, 1988. 393 pp. ISBN: 978-0-262-06112-4.

[2] Nicholas C. Harris, Jacques Carolan, Darius Bunandar, Mihika Prabhu, Michael Hochberg, Tom Baehr-Jones, Michael L. Fanto, A. Matthew Smith, Christopher C. Tison, Paul M. Alsing, and Dirk Englund. "Linear programmable nanophotonic processors". In: *Optica* 5.12 (Dec. 20, 2018). Publisher: Optical Society of America, pp. 1623–1631. ISSN: 2334-2536. DOI: 10.1364/OPTICA.5.001623. URL: https://www.osapublishing.org/optica/abstract.cfm?uri=optica-5-12-1623 (visited on 05/05/2020).

[3] Nicholas C. Harris, Gregory R. Steinbrecher, Mihika Prabhu, Yoav Lahini, Jacob Mower, Darius Bunandar, Changchen Chen, Franco N. C. Wong, Tom Baehr-Jones, Michael Hochberg, Seth Lloyd, and Dirk Englund. "Quantum transport simulations in a programmable nanophotonic processor". In: *Nature Photonics* 11.7 (July 2017). Number: 7 Publisher: Nature Publishing Group, pp. 447–452. ISSN: 1749-4893. DOI: 10.1038/nphoton.2017.95. URL: https://www.nature.com/articles/nphoton.2017.95 (visited on 05/05/2020).

[4] Andrea Annoni, Emanuele Guglielmi, Marco Carminati, Giorgio Ferrari, Marco Sampietro, David AB Miller, Andrea Melloni, and Francesco Morichetti. "Unscrambling light—automatically undoing strong mixing between modes". In: *Light: Science & Applications* 6.12 (Dec. 2017). Number: 12 Publisher: Nature Publishing Group, e17110–e17110. ISSN: 2047-7538. DOI: 10.1038/lsa.2017.110. URL: https://www.nature.com/articles/lsa2017110 (visited on 05/05/2020).

[5] Jacques Carolan, Christopher Harrold, Chris Sparrow, Enrique Martín-López, Nicholas J. Russell, Joshua W. Silverstone, Peter J. Shadbolt, Nobuyuki Matsuda, Manabu Oguma, Mikitaka Itoh, Graham D. Marshall, Mark G. Thompson, Jonathan C. F. Matthews, Toshikazu Hashimoto, Jeremy L. O'Brien, and Anthony Laing. "Universal linear optics". In: *Science* 349.6249 (Aug. 14, 2015). Publisher: American Association for the Advancement of Science Section: Research Article, pp. 711–716. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aab3642. URL: https://science.sciencemag.org/content/349/6249/711 (visited on 05/05/2020).

[6] Antonio Ribeiro, Alfonso Ruocco, Laurent Vanacker, and Wim Bogaerts. "Demonstration of a 4 x 4-port universal linear circuit". In: *Optica* 3.12 (Dec. 20, 2016). Publisher: Optical Society of America, pp. 1348–1357. ISSN: 2334-2536. DOI: 10.1364/OPTICA.3.001348. URL: https://www.

osapublishing.org/optica/abstract.cfm?uri=optica-3-12-1348 (visited on 05/05/2020).

[7] Norman P. Jouppi, Cliff Young, Nishant Patil, David Patterson, Gaurav Agrawal, Raminder Bajwa, Sarah Bates, Suresh Bhatia, Nan Boden, Al Borchers, Rick Boyle, Pierre-luc Cantin, Clifford Chao, Chris Clark, Jeremy Coriell, Mike Daley, Matt Dau, Jeffrey Dean, Ben Gelb, Tara Vazir Ghaemmaghami, Rajendra Gottipati, William Gulland, Robert Hagmann, C. Richard Ho, Doug Hogberg, John Hu, Robert Hundt, Dan Hurt, Julian Ibarz, Aaron Jaffey, Alek Jaworski, Alexander Kaplan, Harshit Khaitan, Andy Koch, Naveen Kumar, Steve Lacy, James Laudon, James Law, Diemthu Le, Chris Leary, Zhuyuan Liu, Kyle Lucke, Alan Lundin, Gordon MacKean, Adriana Maggiore, Maire Mahony, Kieran Miller, Rahul Nagarajan, Ravi Narayanaswami, Ray Ni, Kathy Nix, Thomas Norrie, Mark Omernick, Narayana Penukonda, Andy Phelps, Jonathan Ross, Matt Ross, Amir Salek, Emad Samadiani, Chris Severn, Gregory Sizikov, Matthew Snelham, Jed Souter, Dan Steinberg, Andy Swing, Mercedes Tan, Gregory Thorson, Bo Tian, Horia Toma, Erick Tuttle, Vijay Vasudevan, Richard Walter, Walter Wang, Eric Wilcox, and Doe Hyun Yoon. "In-Datacenter Performance Analysis of a Tensor Processing Unit". In: *arXiv:1704.04760 [cs]* (Apr. 16, 2017). arXiv: 1704.04760. URL: http://arxiv.org/abs/1704.04760 (visited on 05/05/2020).

[8] Yichen Shen, Nicholas C. Harris, Scott Skirlo, Mihika Prabhu, Tom Baehr-Jones, Michael Hochberg, Xin Sun, Shijie Zhao, Hugo Larochelle, Dirk Englund, and Marin Soljačić. "Deep learning with coherent nanophotonic circuits". In: *Nature Photonics* 11.7 (July 2017). Number: 7 Publisher: Nature Publishing Group, pp. 441–446. ISSN: 1749-4893. DOI: 10.1038/nphoton.2017.93. URL: https://www.nature.com/articles/nphoton.2017.93.

[9] Kelvin Wagner and Demetri Psaltis. "Multilayer optical learning networks". In: *Applied Optics* 26.23 (Dec. 1, 1987). Publisher: Optical Society of America, pp. 5061–5076. ISSN: 2155-3165. DOI: 10.1364/AO.26.005061. URL: https://www.osapublishing.org/ao/abstract.cfm?uri=ao-26-23-5061 (visited on 05/05/2020).

[10] R. H. Ritchie. "Plasma losses by fast electrons in thin films". In: *Physical Review* 106.5 (1957), p. 874.

[11] R. Alvarez-Puebla, J. P. Bravo-Vasquez B. Cui, T. Veres, and H. Fenniri. "Nanoimprinted SERS-active substrates with tunable surface plasmon resonances". In: *The Journal of Physical Chemistry C* 111.18 (2007), pp. 6720–6723.

[12] J. Homola. "Surface plasmon resonance sensors for detection of chemical and biological species". In: *Chemical reviews* 108.2 (2008), pp. 462–493.

[13] Mohammadali Tabatabaei, Mohamadreza Najiminaini, Kieffer Davieau, Bozena Kaminska, Mahi R. Singh, Jeffrey J. L. Carson, and François Lagugné-Labarthet. "Tunable 3D Plasmonic Cavity Nanosensors for Surface-Enhanced Raman Spectroscopy with Sub-femtomolar Limit of Detection". In: *ACS Photonics* 2.6 (2015), pp. 752–759.

[14] T. Taubner, D. Korobkin, Y. Urzhumov, G. Shvets, and R. Hillenbrand. "Near-field microscopy through a SiC superlens". In: *Science* 313 (2006), p. 1595.

[15] J. N. Farahani, D. W. Pohl, H.-J. Eisler, and B. Hecht. "Single quantum dot coupled to a scanning optical antenna: a tunable superemitter". In: *Phys. Rev. Lett.* 95 (2005), p. 17402.

[16] J. Dintinger, S. Klein, F. Bustos, W. L. Barnes, and T. W. Ebbesen. "Strong coupling between surface plasmon-polaritons and organic molecules in subwavelength hole arrays". In: *Physical Review B* 71.3 (2005), p. 035424.

[17] D. E. Chang, A. S. Sørensen, P. R. Hemmer, and M. D. Lukin. "Strong coupling of single emitters to surface plasmons". In: *Physical Review B* 76.3 (2007), p. 035420.

[18] Christopher J. Nitta, Matthew K. Farrens, and Venkatesh Akella. "On-Chip Photonic Interconnects: A Computer Architect's Perspective". In: *Synthesis Lectures on Computer Architecture* 8.5 (Oct. 31, 2013). Publisher: Morgan & Claypool Publishers, pp. 1–111. ISSN: 1935-3235. DOI: 10.2200/S00537ED1V01Y201309CAC027. URL: https://www.morganclaypool.com/doi/abs/10.2200/S00537ED1V01Y201309CAC027 (visited on 05/05/2020).

[19] J. Sun, E. Timurdogan, A. Yaacobi, E. S. Hosseini, and M. R. Watts. "Large-Scale Nanophotonic Phased Array". In: *Nature* 493 (2013), pp. 195–199.

[20] Ghazaleh Kafaie Shirmanesh, Ruzan Sokhoyan, Ragip A. Pala, and Harry A. Atwater. "Dual-Gated Active Metasurface at 1550 nm with Wide (>300°) Phase Tunability". In: *Nano Letters* 18.5 (May 9, 2018). Publisher: American Chemical Society, pp. 2957–2963. ISSN: 1530-6984. DOI: 10.1021/acs.nanolett.8b00351. URL: https://doi.org/10.1021/acs.nanolett.8b00351 (visited on 05/05/2020).

[21] Gleb Akselrod. *Metasurface beam steering enables solid-state, high-performance lidar*. Laser Focus World. Library Catalog: www.laserfocusworld.com. July 2019. URL: https://www.laserfocusworld.com/optics/article/14036818/metasurface-beam-steering-enables-solidstate-highperformance-lidar (visited on 05/05/2020).

[22] Dagny Fleischman, Katherine T. Fountaine, Colton R. Bukowsky, Giulia Tagliabue, Luke A. Sweatlock, and Harry A. Atwater. "High Spectral Resolution Plasmonic Color Filters with Subwavelength Dimensions". In: *ACS Photonics* 6.2 (Feb. 20, 2019). Publisher: American Chemical Society, pp. 332–

338. DOI: 10.1021/acsphotonics.8b01634. URL: https://doi.org/10.1021/acsphotonics.8b01634 (visited on 05/05/2020).

[23] E. Knill, R. Laflamme, and G. J. Milburn. "A Scheme for Efficient Quantum Computation with Linear Optics". In: *Nature* 409 (2001), pp. 46–52.

[24] M.-C. Dheur, E. Devaux, T. W. Ebbesen, A. Baron, J.-C. Rodier, J.-P. Hugonin, J.-J. Greffet, G. Messin, and F. Marquier. "Single-plasmon interferences". In: *Science advances* 2.3 (2016), e1501574.

[25] R. Kolesov, B. Grotz, G. Balasubramanian, R. J. Stöhr, A. A. Nicolet, P. R. Hemmer, F. Jelezko, and J. Wrachtrup. "Wave–particle duality of single surface plasmon polaritons". In: *Nature Physics* 5.7 (2009), p. 470.

[26] J. S. Fakonas, H. Lee, Y. A. Kelaita, and H. A. Atwater. "Two-plasmon quantum interference". In: *Nature Photonics* 8.4 (2014), p. 317.

[27] G. Di Martino, Y. Sonnefraud, M.S. Tame, S. Kéna-Cohen, F. Dieleman, Ş. K. Özdemir, M. S. Kim, and S. A. Maier. "Observation of quantum interference in the plasmonic Hong-Ou-Mandel effect". In: *Physical Review Applied* 1.3 (2014), p. 034004.

[28] B. Vest, M.-C. Dheur, É. Devaux, A. Baron, E. Rousseau, J.-P. Hugonin, J.-J. Greffet, G. Messin, and F. Marquier. "Anti-coalescence of bosons on a lossy beam splitter". In: *Science* 356.6345 (2017), pp. 1373–1376.

[29] E. Altewischer, M. P.van Exter, and J. P. Woerdman. "Plasmon-Assisted Transmission of Entangled Photons". In: *Nature* 418 (2002), pp. 304–306.

[30] M.-C. Dheur, B. Vest, É. Devaux, A. Baron, J.-P. Hugonin, J.-J. Greffet, G. Messin, and F. Marquier. "Remote preparation of single-plasmon states". In: *Physical Review B* 96.4 (2017), p. 045432.

[31] J. S. Fakonas, A. Mitskovets, and H. A. Atwater. "Path-Entanglement of Surface Plasmons". In: *New Journal of Physics* 17 (2015).

[32] S. G. Dlamini, J. T. Francis, X. Zhang, Ş. K. Özdemir, S. N. Chormaic, F. Petruccione, and M. S. Tame. "Probing decoherence in plasmonic waveguides in the quantum regime". In: *Physical Review Applied* 9.2 (2018), p. 024003.

[33] T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolff. "Extraordinary optical transmission through sub-wavelength hole arrays". In: *Nature* 391 (1998), p. 667.

[34] Shankar Balakrishnan, Mohamadreza Najiminaini, Mahi R. Singh, and Jeffrey J. L. Carson. "A study of angle dependent surface plasmon polaritons in nano-hole array structures". In: *Journal of Applied Physics* 120.3 (2016), p. 034302.

[35] Mahi R. Singh, Mohamadreza Najiminaini, Shankar Balakrishnan, and Jeffrey J. L. Carson. "Metamaterial-based theoretical description of light scattering by metallic nano-hole array structures". In: *Journal of Applied Physics* 117.18 (2015), p. 184302.

[36] R. Rangarajan, M. Goggin, and P. Kwiat. "Optimizing type-I polarization-entangled photons". In: *Optics Express* 17.21 (2009), pp. 18920–18933.

[37] J. Bell. "On The Problem of Hidden Variables in Quantum Mechanics". In: *Reviews of Modern Physics* 38.3 (1966), pp. 447–452.

[38] J. Clauser, M. Horne, A. Shimony, and R. Holt. "Proposed Experiment to Test Local Hidden-Variable Theories". In: *Physical Review Letters* 23.15 (1969), pp. 880–884.

[39] K. Puech, F. Z. Henari, W. J. Blau, D. Duff, and G. Schmid. "Investigation of the ultrafast dephasing time of gold nanoparticles using incoherent light". In: *Chemical Physics Letters* 247 (1995), pp. 13–17.

[40] Deep Jariwala, Vinod K. Sangwan, Lincoln J. Lauhon, Tobin J. Marks, and Mark C. Hersam. "Emerging Device Applications for Semiconducting Two-Dimensional Transition Metal Dichalcogenides". In: *ACS Nano* 8.2 (Feb. 25, 2014). Publisher: American Chemical Society, pp. 1102–1120. ISSN: 1936-0851. DOI: 10.1021/nn500064s. URL: https://doi.org/10.1021/nn500064s (visited on 05/07/2020).

[41] Marco Bernardi, Can Ataca, Maurizia Palummo, and Jeffrey C. Grossman. "Optical and Electronic Properties of Two-Dimensional Layered Materials". In: *Nanophotonics* 6.2 (Mar. 1, 2017). Publisher: De Gruyter Section: Nanophotonics, pp. 479–493. ISSN: 2192-8614, 2192-8606. DOI: 10.1515/nanoph-2015-0030. URL: https://www.degruyter.com/view/journals/nanoph/6/2/article-p479.xml (visited on 05/07/2020).

[42] Matin Amani, Der-Hsien Lien, Daisuke Kiriya, Jun Xiao, Angelica Azcatl, Jiyoung Noh, Surabhi R. Madhvapathy, Rafik Addou, Santosh Kc, Madan Dubey, Kyeongjae Cho, Robert M. Wallace, Si-Chen Lee, Jr-Hau He, Joel W. Ager, Xiang Zhang, Eli Yablonovitch, and Ali Javey. "Near-unity photoluminescence quantum yield in MoS2". In: *Science* 350.6264 (Nov. 27, 2015). Publisher: American Association for the Advancement of Science Section: Report, pp. 1065–1068. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aad2114. URL: https://science.sciencemag.org/content/350/6264/1065 (visited on 05/07/2020).

[43] Der-Hsien Lien, Shiekh Zia Uddin, Matthew Yeh, Matin Amani, Hyungjin Kim, Joel W. Ager, Eli Yablonovitch, and Ali Javey. "Electrical suppression of all nonradiative recombination pathways in monolayer semiconductors". In: *Science* 364.6439 (May 3, 2019). Publisher: American Association for the Advancement of Science Section: Report, pp. 468–471. ISSN: 0036-8075,

1095-9203. DOI: 10.1126/science.aaw8053. URL: https://science.sciencemag.org/content/364/6439/468 (visited on 05/07/2020).

[44] A. J. Goodman, A. P. Willard, and W. A. Tisdale. "Exciton trapping is responsible for the long apparent lifetime in acid-treated MoS2". In: *Physical Review B* 96.12 (Sept. 14, 2017). Publisher: American Physical Society, p. 121404. DOI: 10.1103/PhysRevB.96.121404. URL: https://link.aps.org/doi/10.1103/PhysRevB.96.121404 (visited on 05/07/2020).

[45] Xiaoxu Wei, Zhihao Yu, Fengrui Hu, Ying Cheng, Linwei Yu, Xiaoyong Wang, Min Xiao, Junzhuan Wang, Xinran Wang, and Yi Shi. "Mo-O bond doping and related-defect assisted enhancement of photoluminescence in monolayer MoS2". In: *AIP Advances* 4.12 (Oct. 6, 2014). Publisher: American Institute of Physics, p. 123004. DOI: 10.1063/1.4897522. URL: https://aip.scitation.org/doi/full/10.1063/1.4897522 (visited on 05/07/2020).

[46] Hongsheng Liu, Nannan Han, and Jijun Zhao. "Atomistic insight into the oxidation of monolayer transition metal dichalcogenides: from structures to electronic properties". In: *RSC Advances* 5.23 (Feb. 9, 2015). Publisher: The Royal Society of Chemistry, pp. 17572–17581. ISSN: 2046-2069. DOI: 10.1039/C4RA17320A. URL: https://pubs.rsc.org/en/content/articlelanding/2015/ra/c4ra17320a.

[47] W. Kohn and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects". In: *Physical Review* 140.4 (Nov. 15, 1965). Publisher: American Physical Society, A1133–A1138. DOI: 10.1103/PhysRev.140.A1133. URL: https://link.aps.org/doi/10.1103/PhysRev.140.A1133 (visited on 05/05/2020).

[48] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. "Generalized Gradient Approximation Made Simple". In: *Physical Review Letters* 77.18 (Oct. 28, 1996). Publisher: American Physical Society, pp. 3865–3868. DOI: 10.1103/PhysRevLett.77.3865. URL: https://link.aps.org/doi/10.1103/PhysRevLett.77.3865 (visited on 05/06/2020).

[49] John P. Perdew, Adrienn Ruzsinszky, Jianmin Tao, Viktor N. Staroverov, Gustavo E. Scuseria, and Gábor I. Csonka. "Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits". In: *The Journal of Chemical Physics* 123.6 (Aug. 8, 2005). Publisher: American Institute of Physics, p. 062201. ISSN: 0021-9606. DOI: 10.1063/1.1904565. URL: https://aip.scitation.org/doi/10.1063/1.1904565 (visited on 05/06/2020).

[50] Diola Bagayoko. "Understanding density functional theory (DFT) and completing it in practice". In: *AIP Advances* 4.12 (Dec. 1, 2014). Publisher: American Institute of Physics, p. 127104. DOI: 10.1063/1.4903408. URL: https://aip.scitation.org/doi/full/10.1063/1.4903408 (visited on 05/06/2020).

[51] Axel D. Becke. "Perspective: Fifty years of density-functional theory in chemical physics". In: *The Journal of Chemical Physics* 140.18 (Apr. 1, 2014). Publisher: American Institute of Physics, 18A301. ISSN: 0021-9606. DOI: 10.1063/1.4869598. URL: https://aip.scitation.org/doi/full/10.1063/1.4869598 (visited on 05/06/2020).

[52] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L. Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauzero, Ari P. Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M. Wentzcovitch. "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials". In: *Journal of Physics: Condensed Matter* 21.39 (Sept. 2009). Publisher: IOP Publishing, p. 395502. ISSN: 0953-8984. DOI: 10.1088/0953-8984/21/39/395502. URL: https://doi.org/10.1088%2F0953-8984%2F21%2F39%2F395502 (visited on 05/05/2020).

[53] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. Otero-de-la-Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu, and S. Baroni. "Advanced capabilities for materials modelling with Quantum ESPRESSO". In: *Journal of Physics: Condensed Matter* 29.46 (Oct. 2017). Publisher: IOP Publishing, p. 465901. ISSN: 0953-8984. DOI: 10.1088/1361-648X/aa8f79. URL: https://doi.org/10.1088%2F1361-648x%2Faa8f79.

[54] Ernest R. Davidson. "Super-matrix methods". In: *Computer Physics Communications* 53.1 (May 1, 1989), pp. 49–60. ISSN: 0010-4655. DOI: 10.1016/0010-4655(89)90147-1. URL: http://www.sciencedirect.com/science/article/pii/0010465589901471 (visited on 05/06/2020).

[55] D. R. Hamann. "Optimized norm-conserving Vanderbilt pseudopotentials". In: *Physical Review B* 88.8 (Aug. 19, 2013). Publisher: American Physical Society, p. 085117. DOI: 10.1103/PhysRevB.88.085117. URL: https://link.aps.org/doi/10.1103/PhysRevB.88.085117 (visited on 05/06/2020).

[56] Sean Molesky, Zin Lin, Alexander Y. Piggott, Weiliang Jin, Jelena Vucković, and Alejandro W. Rodriguez. "Inverse design in nanophotonics". In: *Nature Photonics* 12.11 (Nov. 2018). Number: 11 Publisher: Nature Publishing

Group, pp. 659–670. ISSN: 1749-4893. DOI: 10.1038/s41566-018-0246-9. URL: https://www.nature.com/articles/s41566-018-0246-9 (visited on 05/05/2020).

[57] Jialin Song, Yuxin Chen, and Yisong Yue. "A General Framework for Multi-fidelity Bayesian Optimization with Gaussian Processes". In: International Conference on Artificial Intelligence and Statistics (AISTATS). Vol. PMLR: Volume 89. Naha, Okinawa, Japan, 2019, p. 10.

[58] Ghazaleh Kafaie Shirmanesh, Ruzan Sokhoyan, Pin Chieh Wu, and Harry A. Atwater. "Electro-Optically Tunable Universal Metasurfaces". In: *arXiv:1910.02069 [physics]* (Oct. 4, 2019). arXiv: 1910.02069. URL: http://arxiv.org/abs/1910.02069 (visited on 05/07/2020).

[59] Stefan A. Maier, Mark L. Brongersma, Pieter G. Kik, Sheffer Meltzer, Ari A. G. Requicha, and Harry A. Atwater. "Plasmonics—a route to nanoscale optical devices". In: *Advanced materials* 13 (19 2001), pp. 1501–1505.

[60] Sozo Yokogawa, Stanley P. Burgos, and Harry A. Atwater. "Plasmonic Color Filters for CMOS Image Sensor Applications". In: *Nano. Lett., ACS Nano* 7 (2013), pp. 10038–10047.

[61] Jing Chen, Jian Yang, Zhuo Chena, Yi-Jiao Fang, Peng Zhan, and Zhen-Lin Wang. "Plasmonic reflectors and high-Q nano-cavities based on coupled metal-insulator-metal waveguides". In: *AIP Advances* 2 (2012), p. 012145.

[62] Thomas W. Ebbesen, Henri J. Lezec, H. F. Ghaemi, Tineke Thio, and Peter A. Wolff. "Extraordinary optical transmission through sub-wavelength hole arrays". In: *Nature* 391 (1998), pp. 667–668.

[63] Dagny Fleischman, Luke A. Sweatlock, Hirotaka Murakami, and Harry Atwater. "Hyper-selective plasmonic color filters". In: *Optics Express* 25 (2017), pp. 27386–27395.

[64] Scott Kirkpatrick, Charles Daniel Gelatt, and Mario P. Vecchi. "Optimization by Simulated Annealing". In: *Science* 220 (1983), pp. 671–680.

[65] John H. Holland. "Genetic Algorithms and Adaptation". In: *Adaptive Control of Ill-Defined Systems. NATO Conference Series (II Systems Science)* 16 (1984).

[66] James Kennedy and Russel C. Eberhart. "Particle Swarm Optimization". In: *Proceedings of the IEEE International Conference on Neural Networks. Perth, Australia* (1995), pp. 1942–1945.

[67] James G. Mutitu, Shouyuan Shi, Caihua Chen, Timothy Creazzo, Allen Barnett, Christiana Honsberg, and Dennis W Prather. "Thin film silicon solar cell design based on photonic crystal and diffractive grating structures". In: *Optics express* 16.19 (2008), pp. 15238–15248.

[68] Mehrdad Shokooh-Saremi and Robert Magnusson. “Leaky-mode resonant reflectors with extreme bandwidths”. In: *Optics letters* 35.8 (2010), pp. 1121–1123.

[69] Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. “Practical bayesian optimization of machine learning algorithms”. In: *Advances in neural information processing systems*. 2012, pp. 2951–2959.

[70] Carl E. Rasmussen and Chris K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. URL: https://bit.ly/2tYpBix.

[71] Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. “Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design”. In: *Proc. International Conference on Machine Learning (ICML)*. 2010. URL: https://bit.ly/2CNGPGc.

[72] Emile Contal, Vianney Perchet, and Nicolas Vayatis. “Gaussian process optimization with mutual information”. In: *International Conference on Machine Learning*. 2014, pp. 253–261. URL: https://bit.ly/2x7EEbw.

[73] Philipp Hennig and Christian J. Schuler. “Entropy search for information-efficient global optimization”. In: *Journal of Machine Learning Research* 13.Jun (2012), pp. 1809–1837. URL: https://bit.ly/2x5KMQC.

[74] José Miguel Hernández-Lobato, Matthew W. Hoffman, and Zoubin Ghahramani. “Predictive entropy search for efficient global optimization of black-box functions”. In: *Advances in neural information processing systems*. 2014, pp. 918–926.

[75] Zi Wang, Bolei Zhou, and Stefanie Jegelka. “Optimization as estimation with Gaussian processes in bandit settings”. In: *Artificial Intelligence and Statistics*. 2016, pp. 1022–1031. URL: https://bit.ly/2OeSOhp.

[76] Zi Wang and Stefanie Jegelka. “Max-value entropy search for efficient Bayesian optimization”. In: *arXiv preprint arXiv:1703.01968* (2017). URL: https://arxiv.org/pdf/1703.01968.pdf.

[77] Alexander I. J. Forrester, András Sóbester, and Andy J. Keane. “Multi-fidelity optimization via surrogate modelling”. In: *Proceedings of the royal society of london a: mathematical, physical and engineering sciences*. Vol. 463. The Royal Society. 2007, pp. 3251–3269. URL: https://bit.ly/2xkMXRr.

[78] Loic Le Gratiet and Josselin Garnier. “Recursive co-kriging model for design of computer experiments with multiple levels of fidelity”. In: *International Journal for Uncertainty Quantification* 4.5 (2014). URL: https://bit.ly/2PICVQu.

[79] Alonso Marco, Felix Berkenkamp, Philipp Hennig, Angela P. Schoellig, Andreas Krause, Stefan Schaal, and Sebastian Trimpe. “Virtual vs. real: Trading off simulations and physical experiments in reinforcement learning with Bayesian optimization”. In: *2017 IEEE International Conference*

on Robotics and Automation (ICRA). 2017, pp. 1557–1563. URL: https://bit.ly/2Oa4e62.

[80] Yehong Zhang, Trong Nghia Hoang, Bryan Kian Hsiang Low, and Mohan Kankanhalli. "Information-Based Multi-Fidelity Bayesian Optimization". In: *NIPS Workshop on Bayesian Optimization* (2017). URL: https://bit.ly/2N5CdjH.

[81] Mauricio Alvarez and Neil D Lawrence. "Sparse convolved Gaussian processes for multi-output regression". In: *Advances in neural information processing systems*. 2009, pp. 57–64.

[82] Kirthevasan Kandasamy, Gautam Dasarathy, Junier B. Oliva, Jeff Schneider, and Barnabás Póczos. "Gaussian process bandit optimisation with multi-fidelity evaluations". In: *Advances in Neural Information Processing Systems*. 2016, pp. 992–1000. URL: https://bit.ly/2Qngemh.

[83] Kirthevasan Kandasamy, Gautam Dasarathy, Jeff Schneider, and Barnabás Póczos. "Multi-fidelity Bayesian Optimization with Continuous Approximations". In: *International Conference on Machine Learning*. 2017, pp. 1799–1808. URL: https://bit.ly/2N9KgMq.

[84] Jialin Song, Yuxin Chen, and Yisong Yue. "A General Framework for Multi-fidelity Bayesian Optimization with Gaussian Processes". In: *arXiv preprint arXiv:1811.00755* (2018).

[85] Eric Brochu, Vlad M Cora, and Nando De Freitas. "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning". In: *arXiv preprint arXiv:1012.2599* (2010).

[86] Feng Feng, Ian H. White, and Timothy D. Wilkinson. "Free Space Communications With Beam Steering a Two-Electrode Tapered Laser Diode Using Liquid-Crystal SLM". In: *J. Lightwave Technol.* 31.12 (June 2013). Publisher: OSA, pp. 2001–2007. URL: http://jlt.osa.org/abstract.cfm?URI=jlt-31-12-2001.

[87] Hui-Hsin Hsiao, Cheng Hung Chu, and Din Ping Tsai. "Fundamentals and Applications of Metasurfaces". In: *Small Methods* 1.4 (2017), p. 1600064. DOI: 10.1002/smtd.201600064. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/smtd.201600064.

[88] Nanfang Yu and Federico Capasso. "Flat optics with designer metasurfaces". In: *Nature Materials* 13.2 (Feb. 1, 2014), pp. 139–150. ISSN: 1476-4660. DOI: 10.1038/nmat3839. URL: https://doi.org/10.1038/nmat3839.

[89] Amr M. Shaltout, Vladimir M. Shalaev, and Mark L. Brongersma. "Spatiotemporal light control with active metasurfaces". In: *Science* 364.6441 (2019). Publisher: American Association for the Advancement of Science _eprint: https://science.sciencemag.org/content/364/6441/eaat3100.full.pdf. ISSN:

0036-8075. DOI: 10.1126/science.aat3100. URL: https://science.sciencemag.org/content/364/6441/eaat3100.

[90] Yao-Wei Huang, Ho Wai Howard Lee, Ruzan Sokhoyan, Ragip A. Pala, Krishnan Thyagarajan, Seunghoon Han, Din Ping Tsai, and Harry A. Atwater. "Gate-Tunable Conducting Oxide Metasurfaces". In: *Nano Letters* 16.9 (2016). _eprint: https://doi.org/10.1021/acs.nanolett.6b00555, pp. 5319–5325. DOI: 10.1021/acs.nanolett.6b00555. URL: https://doi.org/10.1021/acs.nanolett.6b00555.

[91] Moshe Leshno, Vladimir Ya Lin, Allan Pinkus, and Shimon Schocken. "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function". In: *Neural Networks* 6.6 (1993), pp. 861–867. ISSN: 0893-6080. DOI: https://doi.org/10.1016/S0893-6080(05)80131-5. URL: http://www.sciencedirect.com/science/article/pii/S0893608005801315.

[92] David N. Hutchison, Jie Sun, Jonathan K. Doylend, Ranjeet Kumar, John Heck, Woosung Kim, Christopher T. Phare, Avi Feshali, and Haisheng Rong. "High-resolution aliasing-free optical beam steering". In: *Optica* 3.8 (Aug. 20, 2016). Publisher: Optical Society of America, pp. 887–890. ISSN: 2334-2536. DOI: 10.1364/OPTICA.3.000887. URL: https://www.osapublishing.org/optica/abstract.cfm?uri=optica-3-8-887 (visited on 05/07/2020).